US 20210130833A1

(19) **United States**
(12) **Patent Application Publication**
Zhang et al.
(10) **Pub. No.: US 2021/0130833 A1**
(43) **Pub. Date: May 6, 2021**

(54) **BACTERIAL DEFENSE SYSTEMS AND METHODS OF IDENTIFYING THEREOF**

(71) Applicants: **THE BROAD INSTITUTE, INC.**, Cambridge, MA (US); **MASSACHUSETTS INSTITUTE OF TECHNOLOGY**, Cambridge, MA (US)

(72) Inventors: **Feng Zhang**, Cambridge, MA (US); **Linyi Gao**, Cambridge, MA (US)

(73) Assignees: **THE BROAD INSTITUTE, INC.**, Cambridge, MA (US); **MASSACHUSETTS INSTITUTE OF TECHNOLOGY**, Cambridge, MA (US)

(21) Appl. No.: **17/085,937**

(22) Filed: **Oct. 30, 2020**

Related U.S. Application Data

(60) Provisional application No. 62/928,269, filed on Oct. 30, 2019, provisional application No. 63/051,161, filed on Jul. 13, 2020.

Publication Classification

(51) **Int. Cl.**

C12N 15/70	(2006.01)
C12Q 1/689	(2006.01)
C12N 9/14	(2006.01)
C12N 9/78	(2006.01)
C12N 9/12	(2006.01)
C12N 9/22	(2006.01)

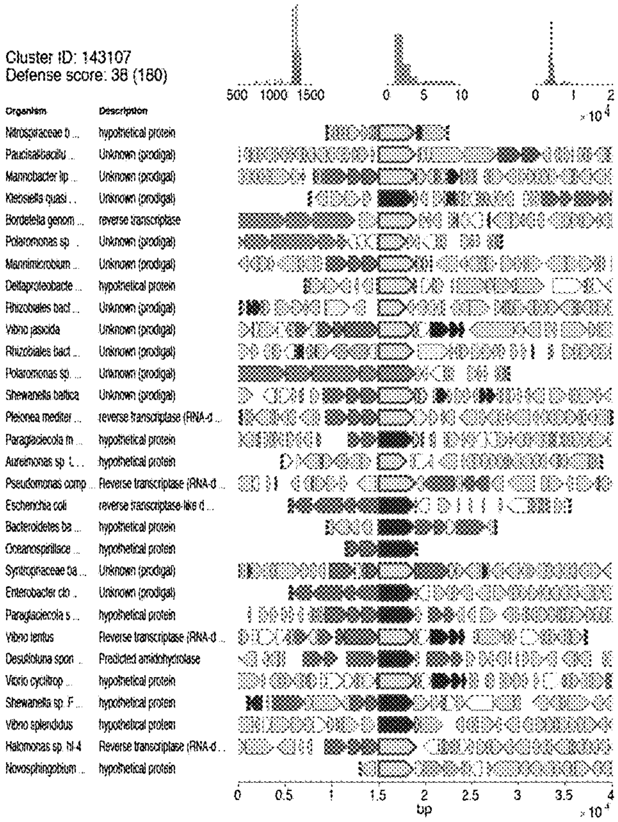
(52) **U.S. Cl.**

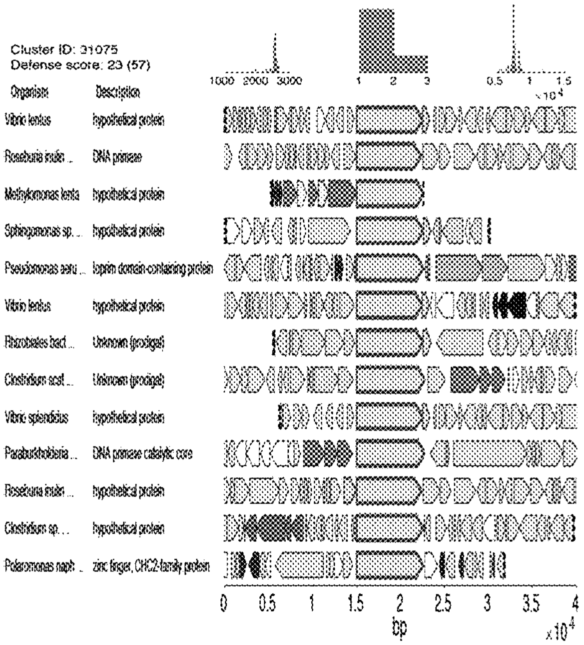
CPC ***C12N 15/70*** (2013.01); ***C12Q 1/689*** (2013.01); ***C12N 9/14*** (2013.01); ***C12Y 306/01003*** (2013.01); ***C12N 9/78*** (2013.01); *C12N 2800/80* (2013.01); ***C12N 9/1276*** (2013.01); ***C12Y 207/07049*** (2013.01); ***C12N 9/22*** (2013.01); *C07K 2319/00* (2013.01); ***C12Y 305/04004*** (2013.01)

(57) **ABSTRACT**

Engineered systems comprising components of defense systems identified in prokaryotes are provided.

Specification includes a Sequence Listing.





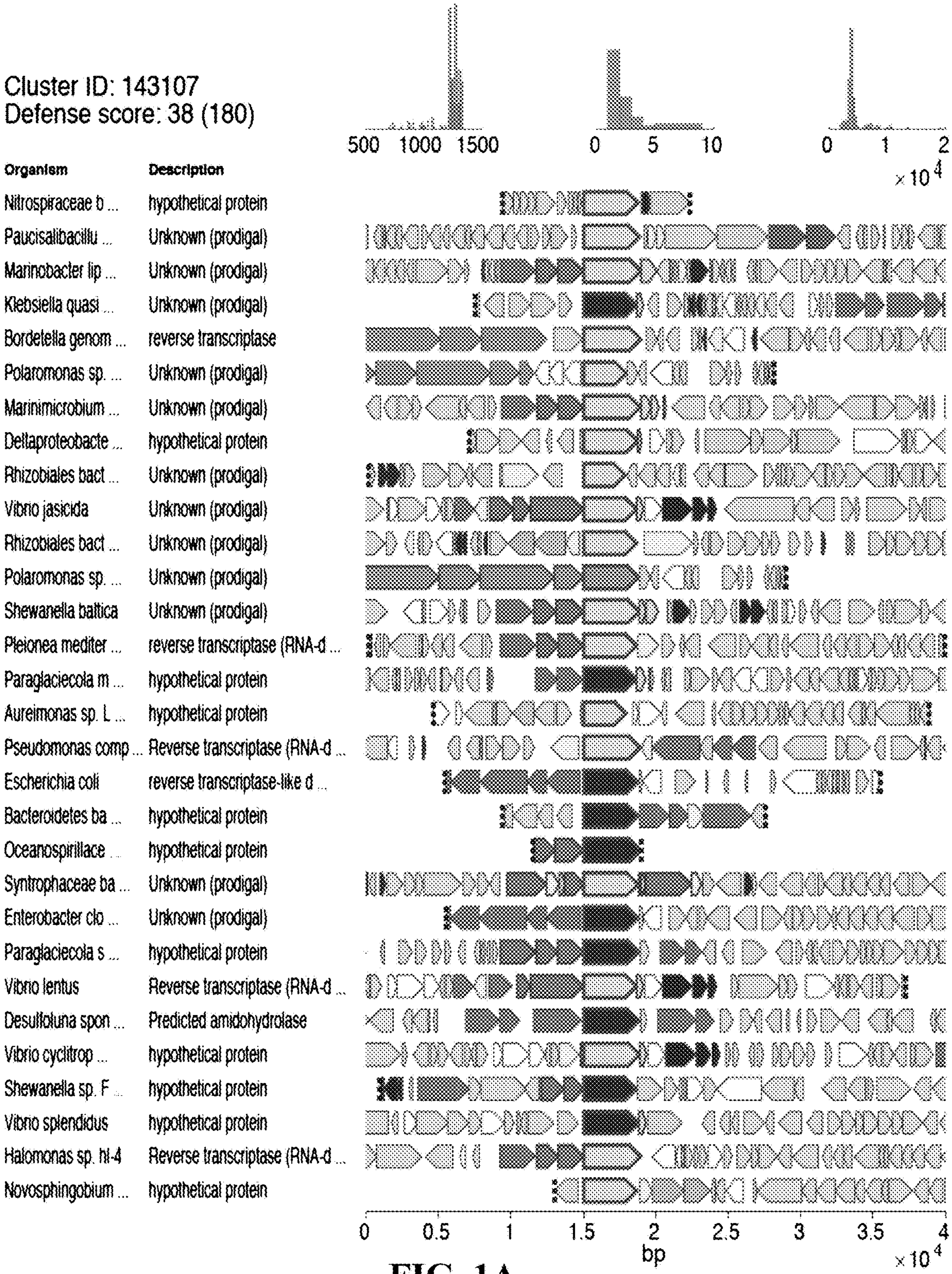

Cluster ID: 143107
Defense score: 38 (180)
500 1000 1500
0 5 10
0 1 2
×10⁴
Organism
Description
Nitrospiraceae b ... hypothetical protein
Paucisalibacillu ... Unknown (prodigal)
Marinobacter lip ... Unknown (prodigal)
Klebsiella quasi ... Unknown (prodigal)
Bordetella genom ... reverse transcriptase
Polaromonas sp. ... Unknown (prodigal)
Marinimicrobium ... Unknown (prodigal)
Deltaproteobacte ... hypothetical protein
Rhizobiales bact ... Unknown (prodigal)
Vibrio jasicida Unknown (prodigal)
Rhizobiales bact ... Unknown (prodigal)
Polaromonas sp. ... Unknown (prodigal)
Shewanella baltica Unknown (prodigal)
Pleionea mediter ... reverse transcriptase (RNA-d ...
Paraglaciecola m ... hypothetical protein
Aureimonas sp. L ... hypothetical protein
Pseudomonas comp ... Reverse transcriptase (RNA-d ...
Escherichia coli reverse transcriptase-like d ...
Bacteroidetes ba ... hypothetical protein
Oceanospirillace ... hypothetical protein
Syntrophaceae ba ... Unknown (prodigal)
Enterobacter clo ... Unknown (prodigal)
Paraglaciecola s ... hypothetical protein
Vibrio lentus Reverse transcriptase (RNA-d ...
Desulfoluna spon ... Predicted amidohydrolase
Vibrio cyclitrop ... hypothetical protein
Shewanella sp. F ... hypothetical protein
Vibrio splendidus hypothetical protein
Halomonas sp. hl-4 Reverse transcriptase (RNA-d ...
Novosphingobium ... hypothetical protein
0 0.5 1 1.5 2 2.5 3 3.5 4
bp
×10⁴

FIG. 1A


Cluster ID: 31075
Defense score: 23 (57)
1000
2000
3000
1
2
3
0.5
1
1.5
×10⁴
Organism
Description
Vibrio lentus
hypothetical protein
Roseburia inulin ...
DNA primase
Methylomonas lenta
hypothetical protein
Sphingomonas sp. ...
hypothetical protein
Pseudomonas aeru ...
toprim domain-containing protein
Vibrio lentus
hypothetical protein
Rhizobiales bact ...
Unknown (prodigal)
Clostridium scat ...
Unknown (prodigal)
Vibrio splendidus
hypothetical protein
Paraburkholderia ...
DNA primase catalytic core
Roseburia inulin ...
hypothetical protein
Clostridium sp. ...
hypothetical protein
Polaromonas naph ...
zinc finger, CHC2-family protein
0
0.5
1
1.5
2
2.5
3
3.5
4
bp
×10⁴

FIG. 1B

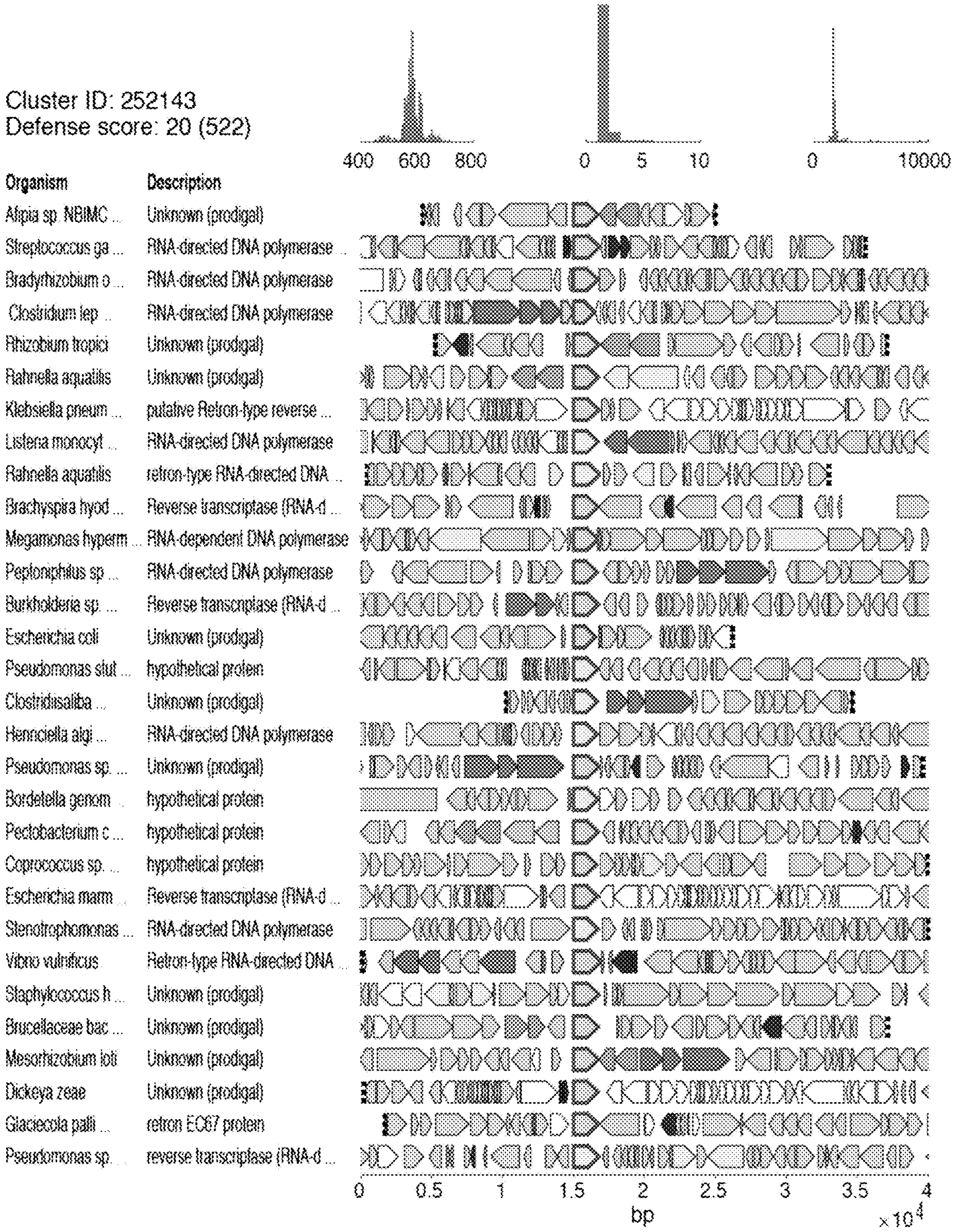

Cluster ID: 252143
Defense score: 20 (522)
400
600
800
0
5
10
0
10000
Organism
Description
Afipia sp. NBIMC ...
Unknown (prodigal)
Streptococcus ga ...
RNA-directed DNA polymerase ...
Bradyrhizobium o ...
RNA-directed DNA polymerase
Clostridium lep ...
RNA-directed DNA polymerase
Rhizobium tropici
Unknown (prodigal)
Rahnella aquatilis
Unknown (prodigal)
Klebsiella pneum ...
putative Retron-type reverse ...
Listeria monocyt ...
RNA-directed DNA polymerase
Rahnella aquatilis
retron-type RNA-directed DNA ...
Brachyspira hyod ...
Reverse transcriptase (RNA-d ...
Megamonas hyperm ...
RNA-dependent DNA polymerase
Peptoniphilus sp ...
RNA-directed DNA polymerase
Burkholderia sp. ...
Reverse transcriptase (RNA-d ...
Escherichia coli
Unknown (prodigal)
Pseudomonas stut ...
hypothetical protein
Clostridiisalibа ...
Unknown (prodigal)
Henriciella algi ...
RNA-directed DNA polymerase
Pseudomonas sp. ...
Unknown (prodigal)
Bordetella genom ...
hypothetical protein
Pectobacterium c ...
hypothetical protein
Coprococcus sp. ...
hypothetical protein
Escherichia marm ...
Reverse transcriptase (RNA-d ...
Stenotrophomonas ...
RNA-directed DNA polymerase
Vibrio vulnificus
Retron-type RNA-directed DNA ...
Staphylococcus h ...
Unknown (prodigal)
Brucellaceae bac ...
Unknown (prodigal)
Mesorhizobium loti
Unknown (prodigal)
Dickeya zeae
Unknown (prodigal)
Glaciecola palli ...
retron EC67 protein
Pseudomonas sp. ...
reverse transcriptase (RNA-d ...
0
0.5
1
1.5
2
2.5
3
3.5
4
bp
×10⁴

FIG. 1C


Cluster ID: 236518
Defense score: 35 (539)
0 500 1000
0 10 20
0 1 2
×10^4
Organism
Description
Salinibacter ruber Unknown (prodigal)
Helicobacter pyl ... Unknown (prodigal)
Helicobacter pyl ... Unknown (prodigal)
Actinophytocola ... reverse transcriptase (RNA-d ...
Pseudomonas sp. ... Unknown (prodigal)
Rheinheimera sp hypothetical protein
Candidatus Porib ... Unknown (prodigal)
Megasphaera sp. ... hypothetical protein
Pseudomonas fulva hypothetical protein
Pseudomonas putida Unknown (prodigal)
Staphylococcus s ... RNA-directed DNA polymerase
Campylobacter coli Unknown (prodigal)
Granulicatella e ... hypothetical protein
Corallococcus sp ... RNA-directed DNA polymerase
Rhodococcus rhod ... RNA-directed DNA polymerase
Campylobacter bl ... hypothetical protein
Helicobacter pyl ... reverse transcriptase
Helicobacter pyl ... putative uncharacterized protein
Lachnospiraceae ... Unknown (prodigal)
Pseudomonas sp. ... Unknown (prodigal)
Helicobacter pyl ... Unknown (prodigal)
Helicobacter pyl ... reverse transcriptase
Helicobacter pyl ... reverse transcriptase
Helicobacter pyl ... Unknown (prodigal)
Pseudomonas para ... Unknown (prodigal)
Spartobacteria b ... Unknown (prodigal)
Helicobacter pyl ... Unknown (prodigal)
Fusobacterium ul ... hypothetical protein
Ruminococcus sp. ... hypothetical protein
Campylobacter je ... hypothetical protein
0 0.5 1 1.5 2 2.5 3 3.5 4
bp
×10^4

FIG. 1D

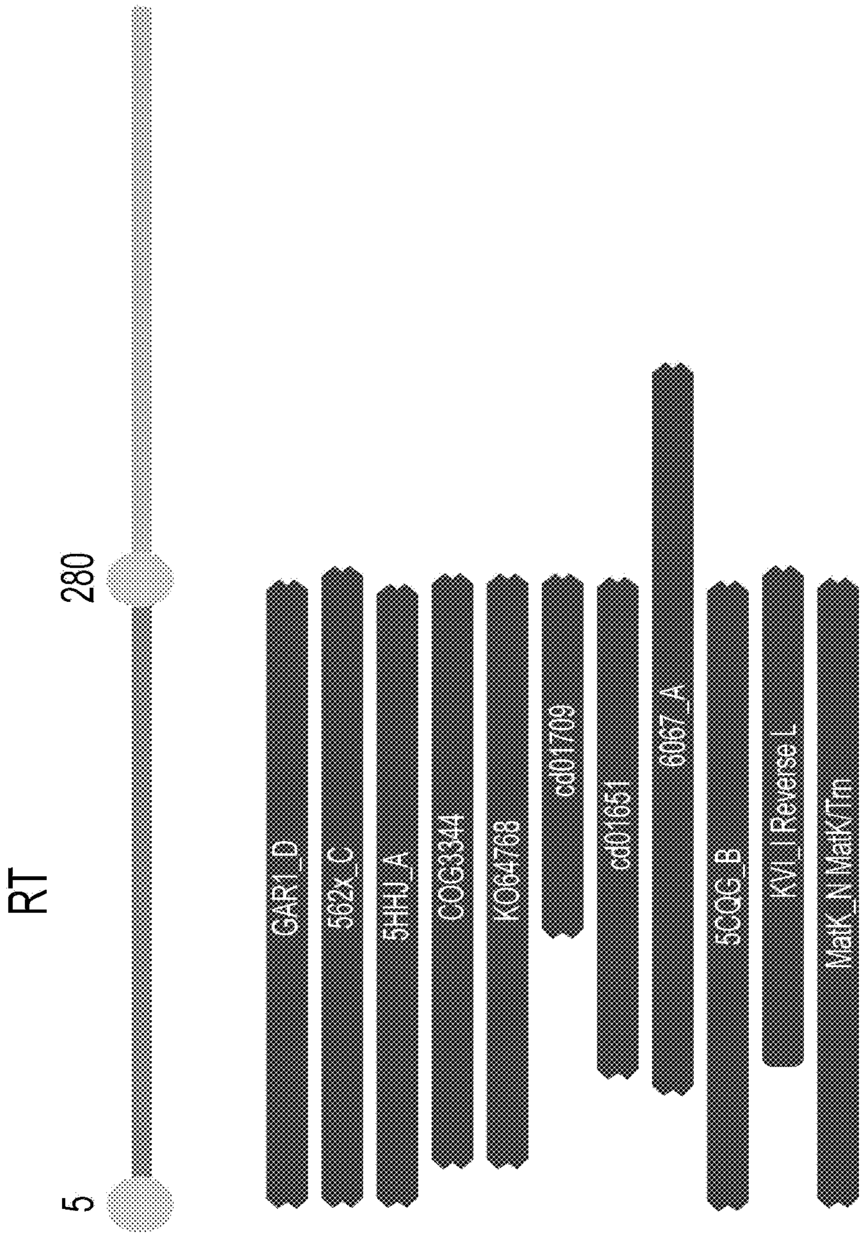
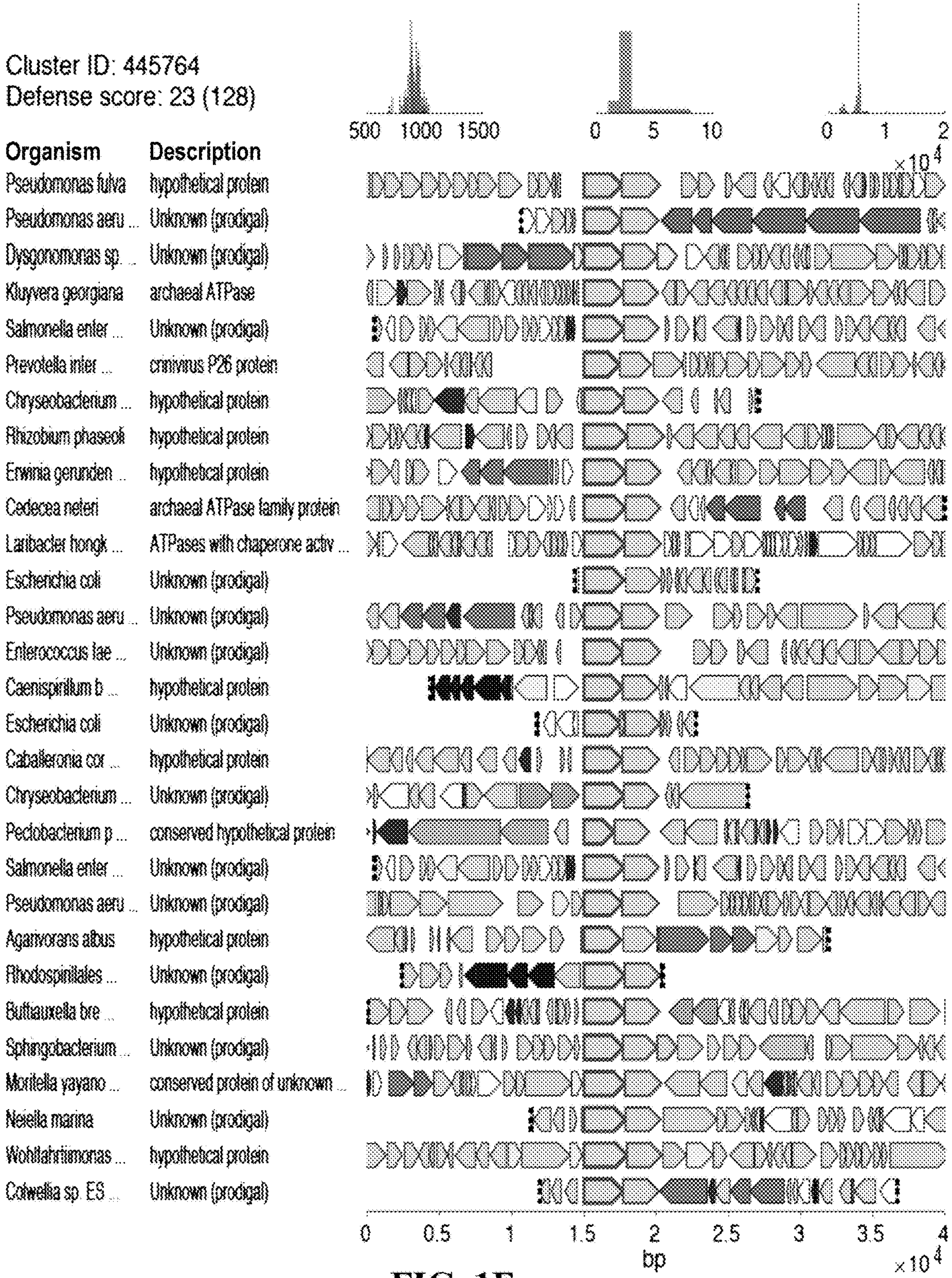

RT
5
280
GAR1_D
562x_C
5HHJ_A
COG3344
KO64768
cd01709
cd01651
6067_A
5CQG_B
KVI_I Reverse L
MatK_N MatK/Tn

FIG. 1D (Cont'd)



Organism	Description
Pseudomonas fulva	hypothetical protein
Pseudomonas aeru ...	Unknown (prodigal)
Dysgonomonas sp. ...	Unknown (prodigal)
Kluyvera georgiana	archaeal ATPase
Salmonella enter ...	Unknown (prodigal)
Prevotella inter ...	crinivirus P26 protein
Chryseobacterium ...	hypothetical protein
Rhizobium phaseoli	hypothetical protein
Erwinia gerunden ...	hypothetical protein
Cedecea neteri	archaeal ATPase family protein
Laribacter hongk ...	ATPases with chaperone activ ...
Escherichia coli	Unknown (prodigal)
Pseudomonas aeru ...	Unknown (prodigal)
Enterococcus lae ...	Unknown (prodigal)
Caenispirillum b ...	hypothetical protein
Escherichia coli	Unknown (prodigal)
Caballeronia cor ...	hypothetical protein
Chryseobacterium ...	Unknown (prodigal)
Pectobacterium p ...	conserved hypothetical protein
Salmonella enter ...	Unknown (prodigal)
Pseudomonas aeru ...	Unknown (prodigal)
Agarivorans albus	hypothetical protein
Rhodospirillales ...	Unknown (prodigal)
Buttiauxella bre ...	hypothetical protein
Sphingobacterium ...	Unknown (prodigal)
Moritella yayano ...	conserved protein of unknown ...
Neiella marina	Unknown (prodigal)
Wohlfahrtiimonas ...	hypothetical protein
Colwellia sp. ES ...	Unknown (prodigal)

FIG. 1E

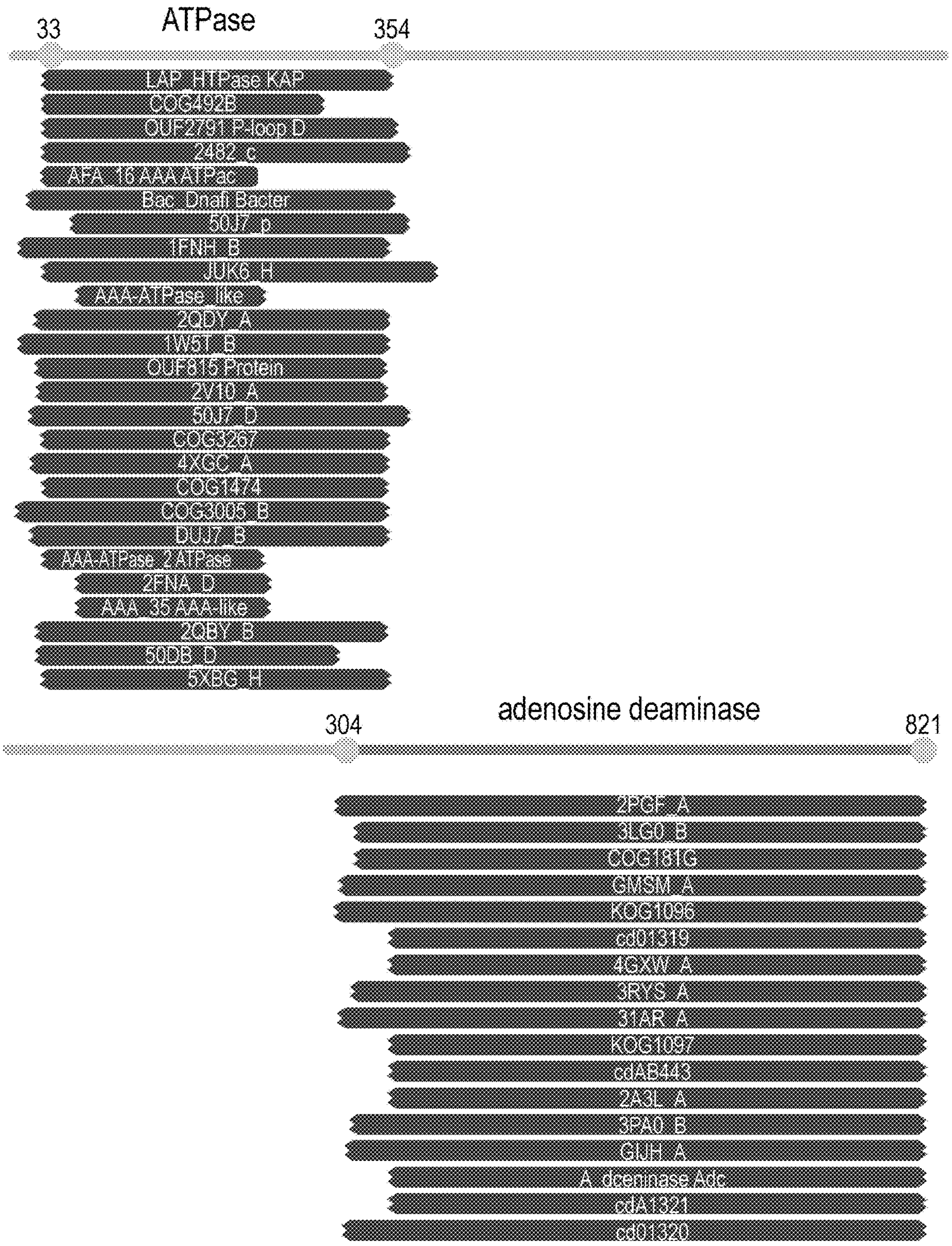

33
ATPase
354
LAP_HTPase KAP
COG492B
OUF2791 P-loop D
2482_c
AFA_16 AAA ATPac
Bac_Dnafi Bacter
50J7_p
1FNH_B
JUK6_H
AAA-ATPase_like
2QDY_A
1W5T_B
OUF815 Protein
2V10_A
50J7_D
COG3267
4XGC_A
COG1474
COG3005_B
DUJ7_B
AAA-ATPase_2 ATPase
2FNA_D
AAA_35 AAA-like
2QBY_B
50DB_D
5XBG_H
adenosine deaminase
304
821
2PGF_A
3LG0_B
COG181G
GMSM_A
KOG1096
cd01319
4GXW_A
3RYS_A
31AR_A
KOG1097
cdAB443
2A3L_A
3PA0_B
GIJH_A
A_dceninase Adc
cdA1321
cd01320

FIG. 1E (Cont'd)

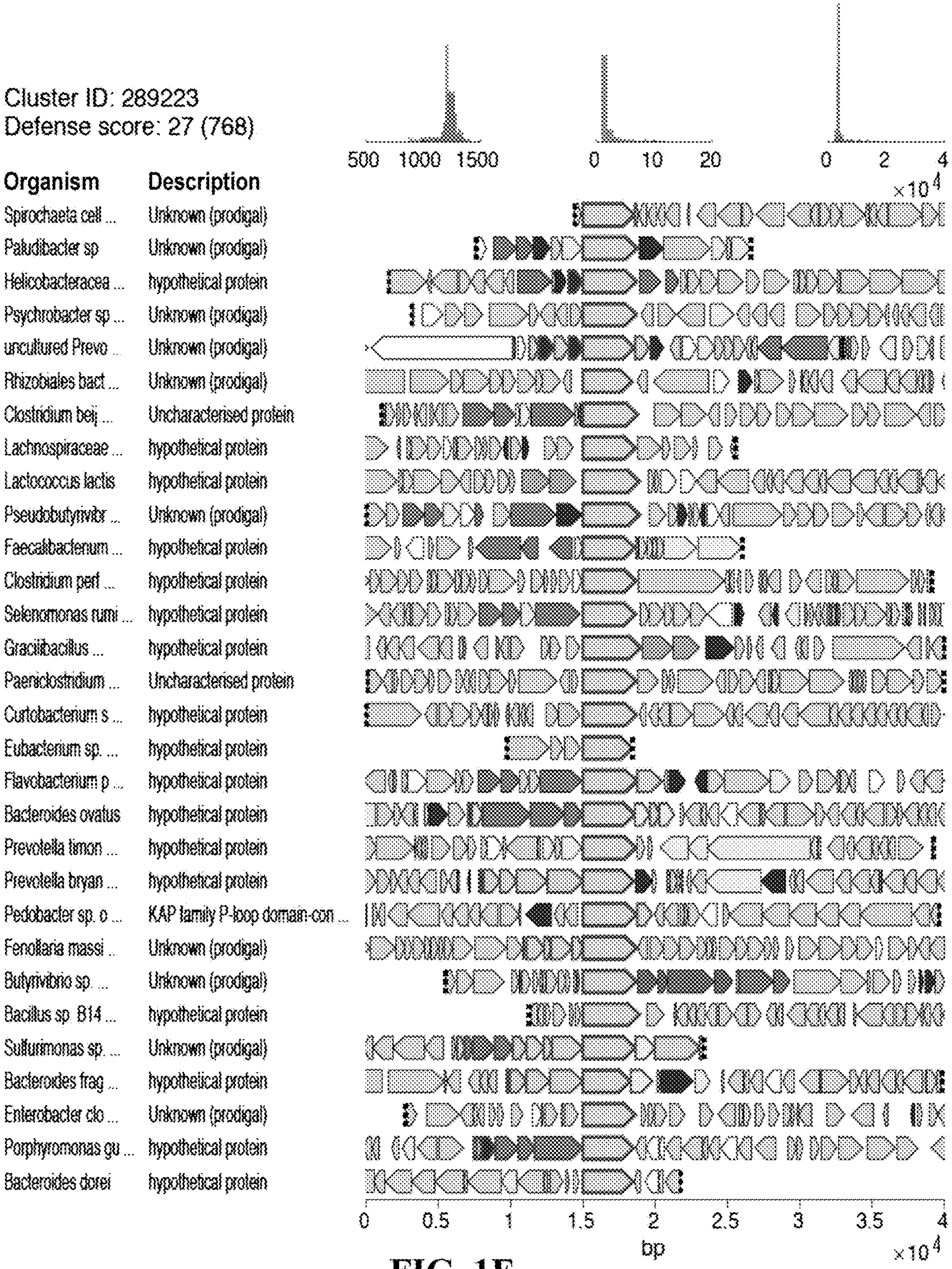

Cluster ID: 289223
Defense score: 27 (768)
500
1000
1500
0
10
20
0
2
4
×10⁴
Organism
Description
Spirochaeta cell ... Unknown (prodigal)
Paludibacter sp Unknown (prodigal)
Helicobacteracea ... hypothetical protein
Psychrobacter sp ... Unknown (prodigal)
uncultured Prevo ... Unknown (prodigal)
Rhizobiales bact ... Unknown (prodigal)
Clostridium beij ... Uncharacterised protein
Lachnospiraceae ... hypothetical protein
Lactococcus lactis hypothetical protein
Pseudobutyrivibr ... Unknown (prodigal)
Faecalibacterium ... hypothetical protein
Clostridium perf ... hypothetical protein
Selenomonas rumi ... hypothetical protein
Gracilibacillus ... hypothetical protein
Paeniclostridium ... Uncharacterised protein
Curtobacterium s ... hypothetical protein
Eubacterium sp. ... hypothetical protein
Flavobacterium p ... hypothetical protein
Bacteroides ovatus hypothetical protein
Prevotella timon ... hypothetical protein
Prevotella bryan ... hypothetical protein
Pedobacter sp. o ... KAP family P-loop domain-con ...
Fenollaria massi ... Unknown (prodigal)
Butyrivibrio sp. ... Unknown (prodigal)
Bacillus sp. B14 ... hypothetical protein
Sulfurimonas sp. ... Unknown (prodigal)
Bacteroides frag ... hypothetical protein
Enterobacter clo ... Unknown (prodigal)
Porphyromonas gu ... hypothetical protein
Bacteroides dorei hypothetical protein
0
0.5
1
1.5
2
2.5
3
3.5
4
bp
×10⁴

FIG. 1F

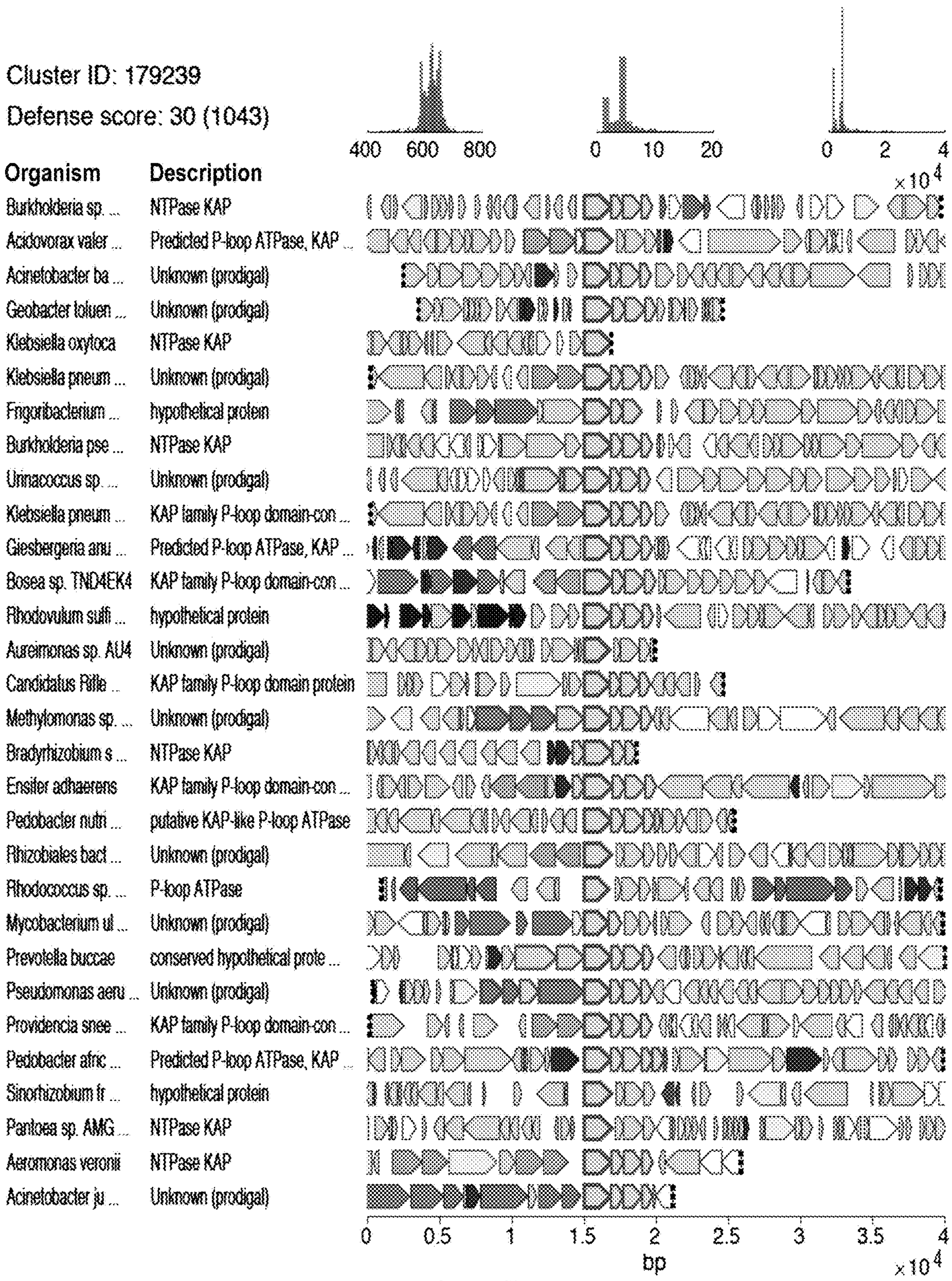

Cluster ID: 179239
Defense score: 30 (1043)
400 600 800
0 10 20
0 2 4
×10⁴
Organism
Description
Burkholderia sp. ... NTPase KAP
Acidovorax valer ... Predicted P-loop ATPase, KAP ...
Acinetobacter ba ... Unknown (prodigal)
Geobacter toluen ... Unknown (prodigal)
Klebsiella oxytoca NTPase KAP
Klebsiella pneum ... Unknown (prodigal)
Frigoribacterium ... hypothetical protein
Burkholderia pse ... NTPase KAP
Urinacoccus sp. ... Unknown (prodigal)
Klebsiella pneum ... KAP family P-loop domain-con ...
Giesbergeria anu ... Predicted P-loop ATPase, KAP ...
Bosea sp. TND4EK4 KAP family P-loop domain-con ...
Rhodovulum sulfi ... hypothetical protein
Aureimonas sp. AU4 Unknown (prodigal)
Candidatus Rifle ... KAP family P-loop domain protein
Methylomonas sp. ... Unknown (prodigal)
Bradyrhizobium s ... NTPase KAP
Ensifer adhaerens KAP family P-loop domain-con ...
Pedobacter nutri ... putative KAP-like P-loop ATPase
Rhizobiales bact ... Unknown (prodigal)
Rhodococcus sp. ... P-loop ATPase
Mycobacterium ul ... Unknown (prodigal)
Prevotella buccae conserved hypothetical prote ...
Pseudomonas aeru ... Unknown (prodigal)
Providencia snee ... KAP family P-loop domain-con ...
Pedobacter afric ... Predicted P-loop ATPase, KAP ...
Sinorhizobium fr ... hypothetical protein
Pantoea sp. AMG ... NTPase KAP
Aeromonas veronii NTPase KAP
Acinetobacter ju ... Unknown (prodigal)
0 0.5 1 1.5 2 2.5 3 3.5 4
bp
×10⁴

FIG. 1G

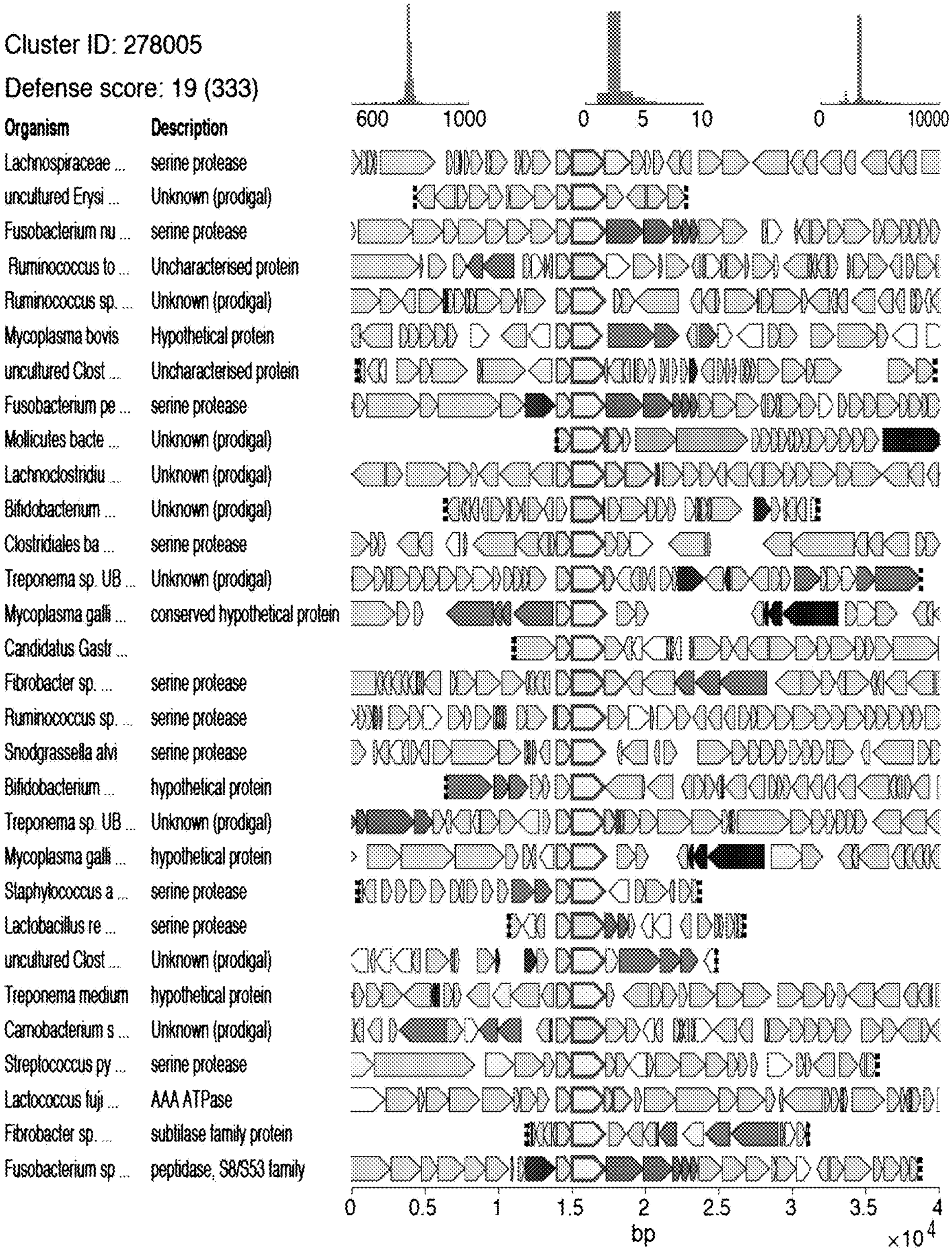

Cluster ID: 278005
Defense score: 19 (333)
Organism
Description
600
1000
0
5
10
0
10000
Lachnospiraceae ... serine protease
uncultured Erysi ... Unknown (prodigal)
Fusobacterium nu ... serine protease
Ruminococcus to ... Uncharacterised protein
Ruminococcus sp. ... Unknown (prodigal)
Mycoplasma bovis Hypothetical protein
uncultured Clost ... Uncharacterised protein
Fusobacterium pe ... serine protease
Mollicutes bacte ... Unknown (prodigal)
Lachnoclostridiu ... Unknown (prodigal)
Bifidobacterium ... Unknown (prodigal)
Clostridiales ba ... serine protease
Treponema sp. UB ... Unknown (prodigal)
Mycoplasma galli ... conserved hypothetical protein
Candidatus Gastr ...
Fibrobacter sp. ... serine protease
Ruminococcus sp. ... serine protease
Snodgrassella alvi serine protease
Bifidobacterium ... hypothetical protein
Treponema sp. UB ... Unknown (prodigal)
Mycoplasma galli ... hypothetical protein
Staphylococcus a ... serine protease
Lactobacillus re ... serine protease
uncultured Clost ... Unknown (prodigal)
Treponema medium hypothetical protein
Carnobacterium s ... Unknown (prodigal)
Streptococcus py ... serine protease
Lactococcus fuji ... AAA ATPase
Fibrobacter sp. ... subtilase family protein
Fusobacterium sp ... peptidase, S8/S53 family
0
0.5
1
1.5
2
2.5
3
3.5
4
bp
×10^4

FIG. 1H

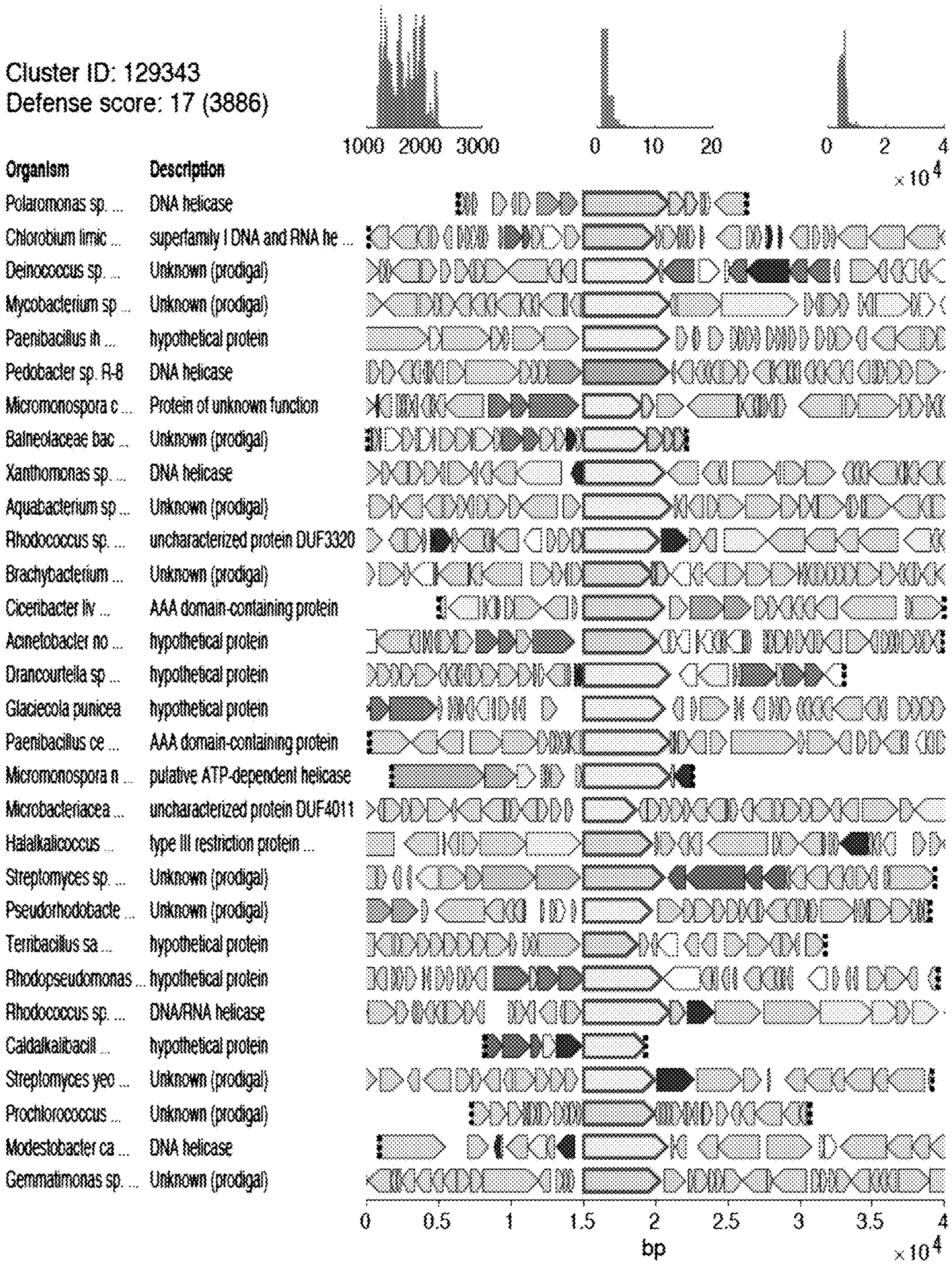

Cluster ID: 129343
Defense score: 17 (3886)
1000 2000 3000
0 10 20
0 2 4
×10^4
Organism
Description
Polaromonas sp. ... DNA helicase
Chlorobium limic ... superfamily I DNA and RNA he ...
Deinococcus sp. ... Unknown (prodigal)
Mycobacterium sp ... Unknown (prodigal)
Paenibacillus ih ... hypothetical protein
Pedobacter sp. R-8 DNA helicase
Micromonospora c ... Protein of unknown function
Balneolaceae bac ... Unknown (prodigal)
Xanthomonas sp. ... DNA helicase
Aquabacterium sp ... Unknown (prodigal)
Rhodococcus sp. ... uncharacterized protein DUF3320
Brachybacterium ... Unknown (prodigal)
Ciceribacter liv ... AAA domain-containing protein
Acinetobacter no ... hypothetical protein
Drancourtella sp ... hypothetical protein
Glaciecola punicea hypothetical protein
Paenibacillus ce ... AAA domain-containing protein
Micromonospora n ... putative ATP-dependent helicase
Microbacteriacea ... uncharacterized protein DUF4011
Halalkalicoccus ... type III restriction protein ...
Streptomyces sp. ... Unknown (prodigal)
Pseudorhodobacte ... Unknown (prodigal)
Terribacillus sa ... hypothetical protein
Rhodopseudomonas ... hypothetical protein
Rhodococcus sp. ... DNA/RNA helicase
Caldalkalibacill ... hypothetical protein
Streptomyces yeo ... Unknown (prodigal)
Prochlorococcus ... Unknown (prodigal)
Modestobacter ca ... DNA helicase
Gemmatimonas sp. ... Unknown (prodigal)
0 0.5 1 1.5 2 2.5 3 3.5 4
bp
×10^4

FIG. 1I

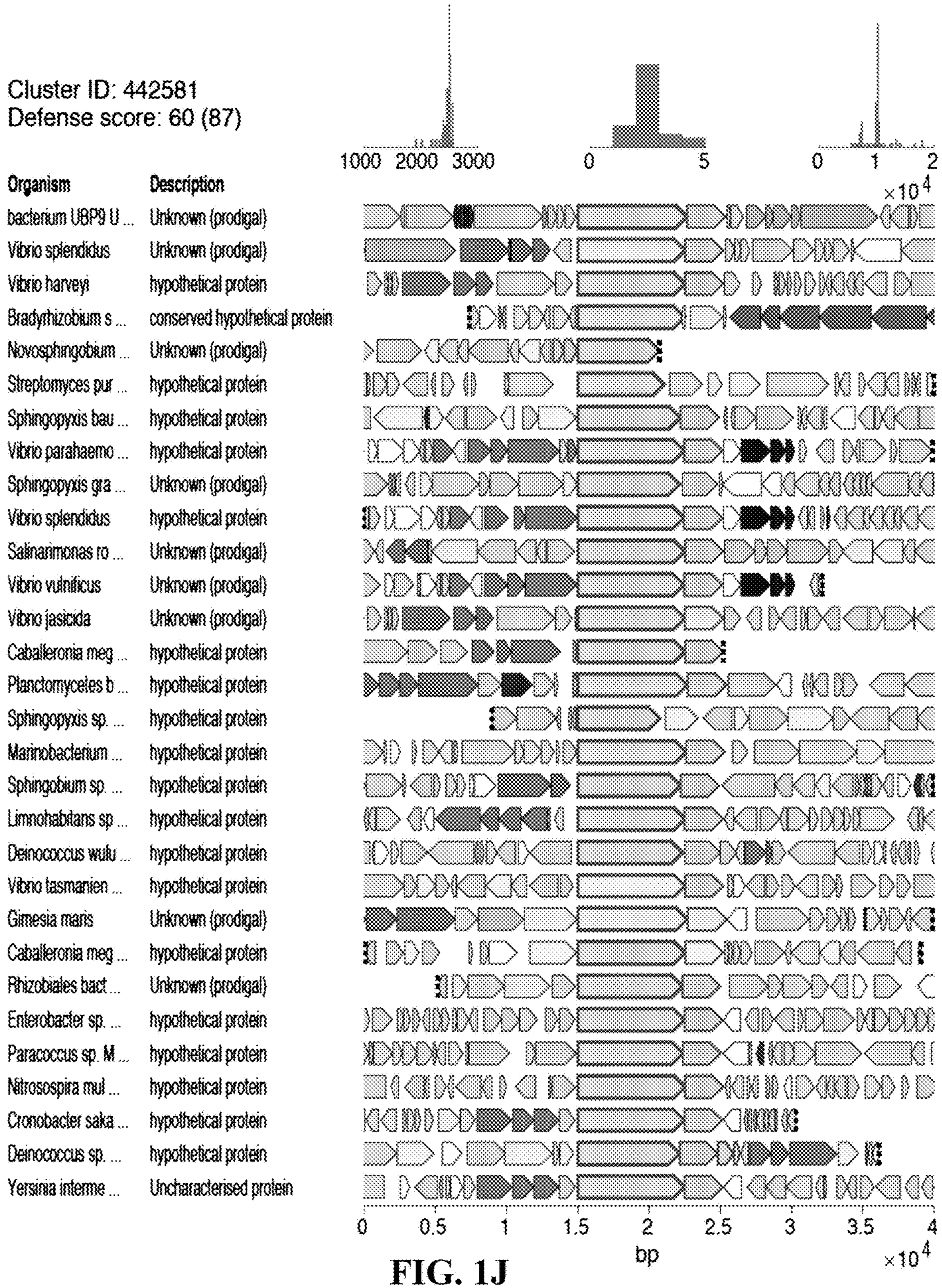

Cluster ID: 442581
Defense score: 60 (87)
1000 2000 3000
0 5
0 1 2
×10^4
Organism Description
bacterium UBP9 U ... Unknown (prodigal)
Vibrio splendidus Unknown (prodigal)
Vibrio harveyi hypothetical protein
Bradyrhizobium s ... conserved hypothetical protein
Novosphingobium ... Unknown (prodigal)
Streptomyces pur ... hypothetical protein
Sphingopyxis bau ... hypothetical protein
Vibrio parahaemo ... hypothetical protein
Sphingopyxis gra ... Unknown (prodigal)
Vibrio splendidus hypothetical protein
Salinarimonas ro ... Unknown (prodigal)
Vibrio vulnificus Unknown (prodigal)
Vibrio jasicida Unknown (prodigal)
Caballeronia meg ... hypothetical protein
Planctomycetes b ... hypothetical protein
Sphingopyxis sp. ... hypothetical protein
Marinobacterium ... hypothetical protein
Sphingobium sp. ... hypothetical protein
Limnohabitans sp ... hypothetical protein
Deinococcus wulu ... hypothetical protein
Vibrio tasmanien ... hypothetical protein
Gimesia maris Unknown (prodigal)
Caballeronia meg ... hypothetical protein
Rhizobiales bact ... Unknown (prodigal)
Enterobacter sp. ... hypothetical protein
Paracoccus sp. M ... hypothetical protein
Nitrosospira mul ... hypothetical protein
Cronobacter saka ... hypothetical protein
Deinococcus sp. ... hypothetical protein
Yersinia interme ... Uncharacterised protein
0 0.5 1 1.5 2 2.5 3 3.5 4
bp
×10^4

FIG. 1J

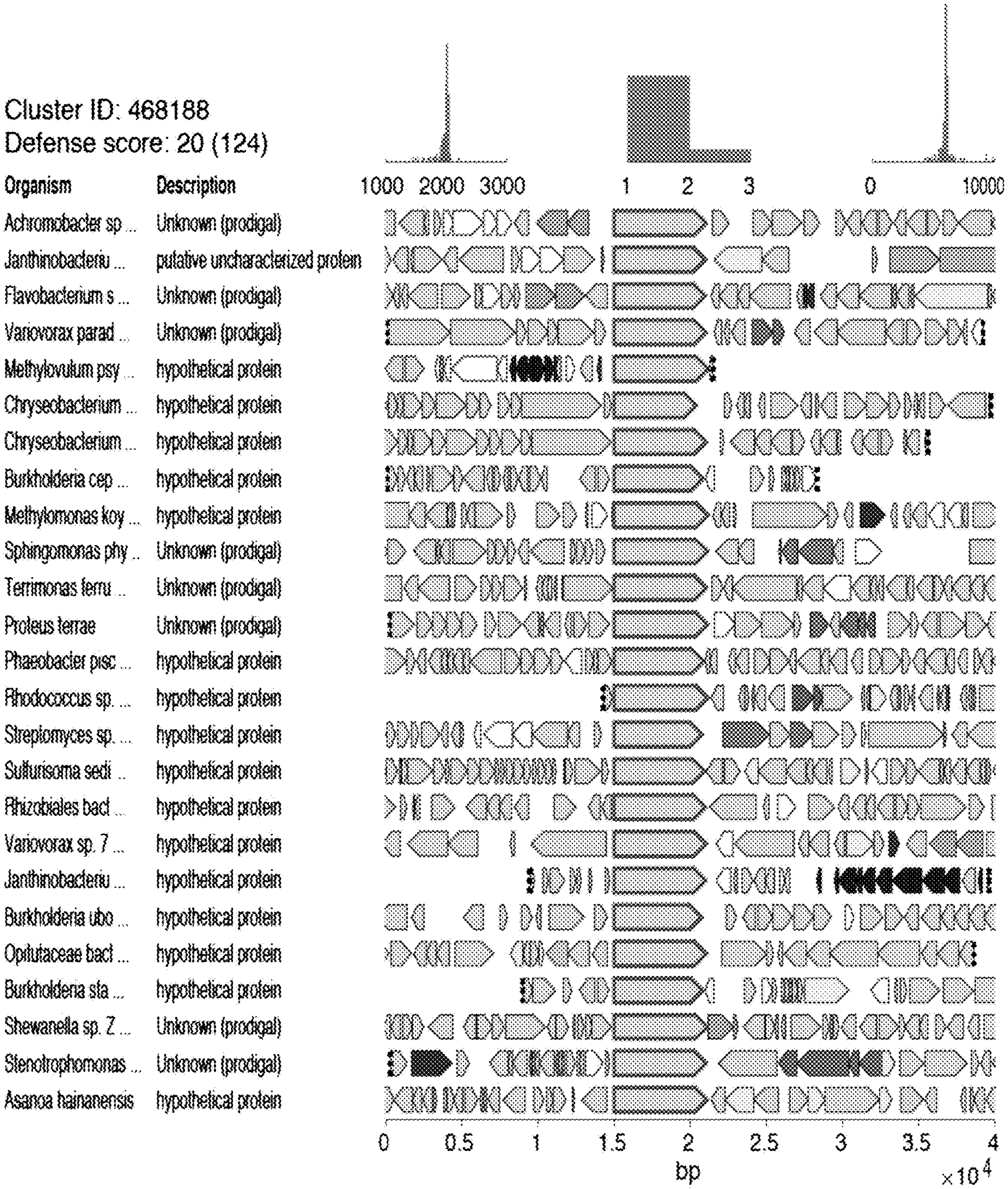

Cluster ID: 468188
Defense score: 20 (124)
Organism
Description
1000
2000
3000
1
2
3
0
10000
Achromobacter sp ...
Unknown (prodigal)
Janthinobacteriu ...
putative uncharacterized protein
Flavobacterium s ...
Unknown (prodigal)
Variovorax parad ...
Unknown (prodigal)
Methylovulum psy ...
hypothetical protein
Chryseobacterium ...
hypothetical protein
Chryseobacterium ...
hypothetical protein
Burkholderia cep ...
hypothetical protein
Methylomonas koy ...
hypothetical protein
Sphingomonas phy ...
Unknown (prodigal)
Terrimonas ferru ...
Unknown (prodigal)
Proteus terrae
Unknown (prodigal)
Phaeobacter pisc ...
hypothetical protein
Rhodococcus sp. ...
hypothetical protein
Streptomyces sp. ...
hypothetical protein
Sulfurisoma sedi ...
hypothetical protein
Rhizobiales bact ...
hypothetical protein
Variovorax sp. 7 ...
hypothetical protein
Janthinobacteriu ...
hypothetical protein
Burkholderia ubo ...
hypothetical protein
Opitutaceae bact ...
hypothetical protein
Burkholderia sta ...
hypothetical protein
Shewanella sp. Z ...
Unknown (prodigal)
Stenotrophomonas ...
Unknown (prodigal)
Asanoa hainanensis
hypothetical protein
0
0.5
1
1.5
2
2.5
3
3.5
4
bp
×10^4

FIG. 1K

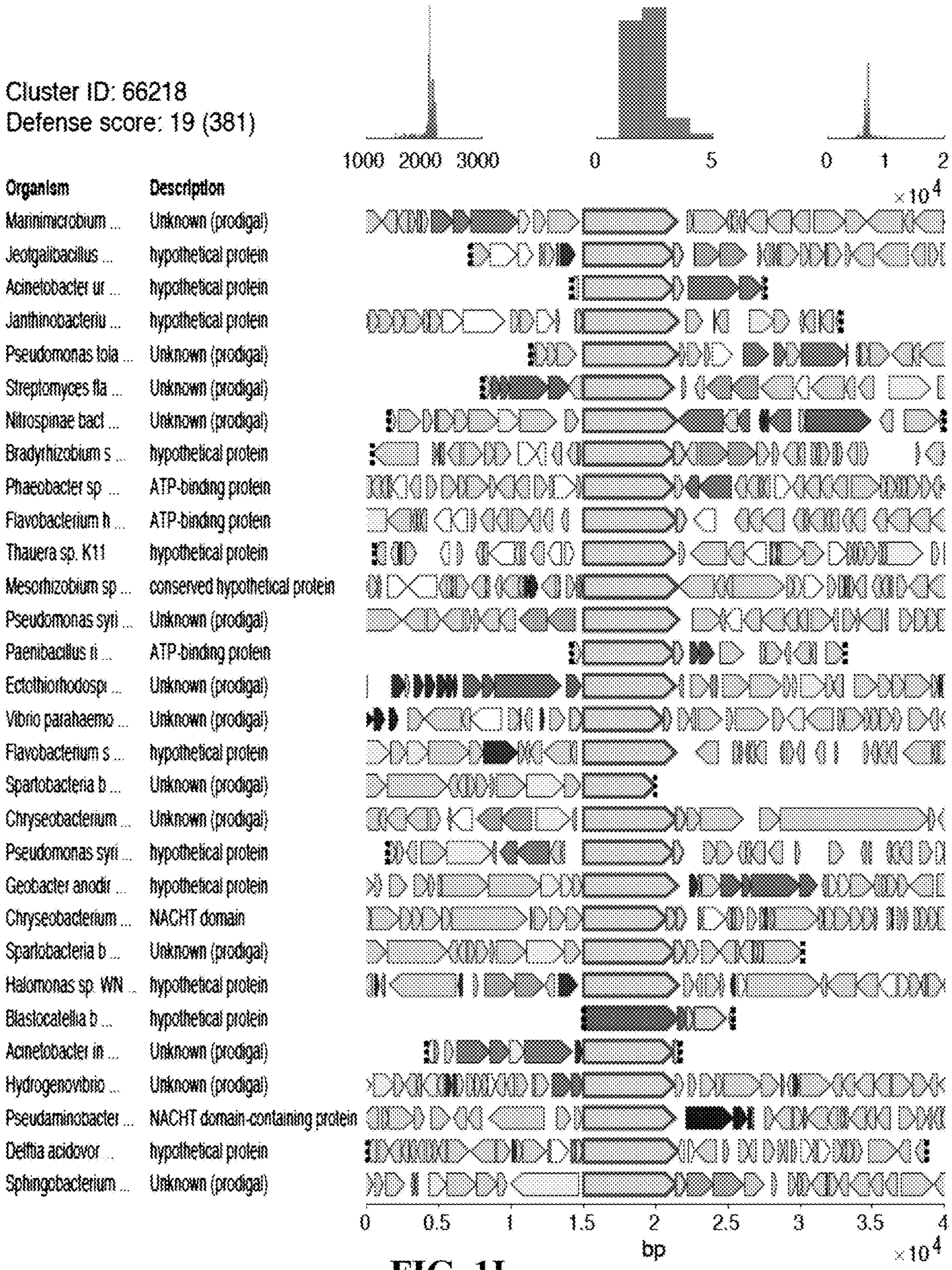

Cluster ID: 66218
Defense score: 19 (381)
1000 2000 3000
0 5
0 1 2
×10^4
Organism
Description
Marinimicrobium ...
Unknown (prodigal)
Jeotgalibacillus ...
hypothetical protein
Acinetobacter ur ...
hypothetical protein
Janthinobacteriu ...
hypothetical protein
Pseudomonas tola ...
Unknown (prodigal)
Streptomyces fla ...
Unknown (prodigal)
Nitrospinae bact ...
Unknown (prodigal)
Bradyrhizobium s ...
hypothetical protein
Phaeobacter sp ...
ATP-binding protein
Flavobacterium h ...
ATP-binding protein
Thauera sp. K11
hypothetical protein
Mesorhizobium sp ...
conserved hypothetical protein
Pseudomonas syri ...
Unknown (prodigal)
Paenibacillus ri ...
ATP-binding protein
Ectothiorhodospi ...
Unknown (prodigal)
Vibrio parahaemo ...
Unknown (prodigal)
Flavobacterium s ...
hypothetical protein
Spartobacteria b ...
Unknown (prodigal)
Chryseobacterium ...
Unknown (prodigal)
Pseudomonas syri ...
hypothetical protein
Geobacter anodir ...
hypothetical protein
Chryseobacterium ...
NACHT domain
Spartobacteria b ...
Unknown (prodigal)
Halomonas sp. WN ...
hypothetical protein
Blastocatellia b ...
hypothetical protein
Acinetobacter in ...
Unknown (prodigal)
Hydrogenovibrio ...
Unknown (prodigal)
Pseudaminobacter ...
NACHT domain-containing protein
Delftia acidovor ...
hypothetical protein
Sphingobacterium ...
Unknown (prodigal)
0 0.5 1 1.5 2 2.5 3 3.5 4
bp
×10^4

FIG. 1L

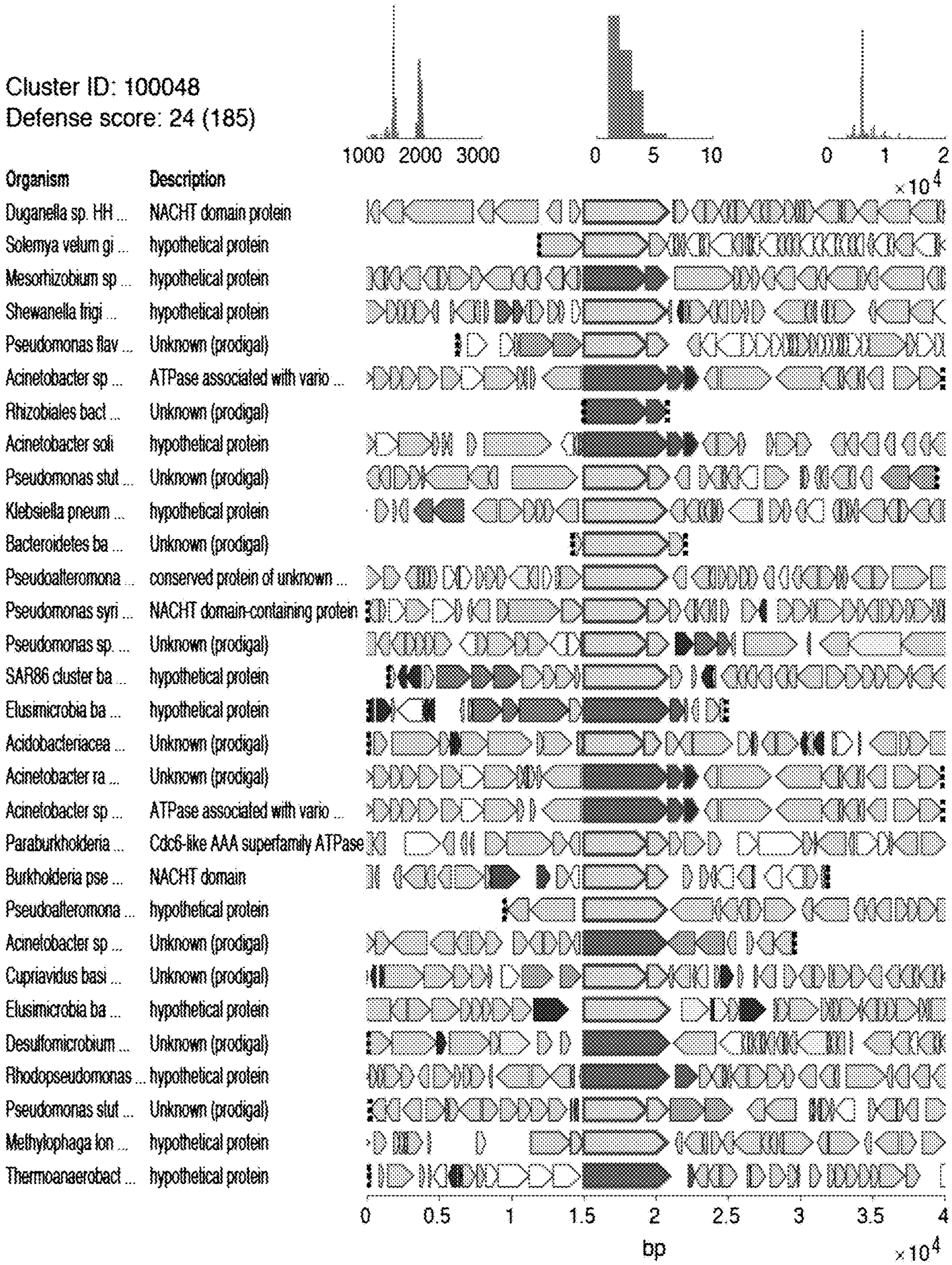
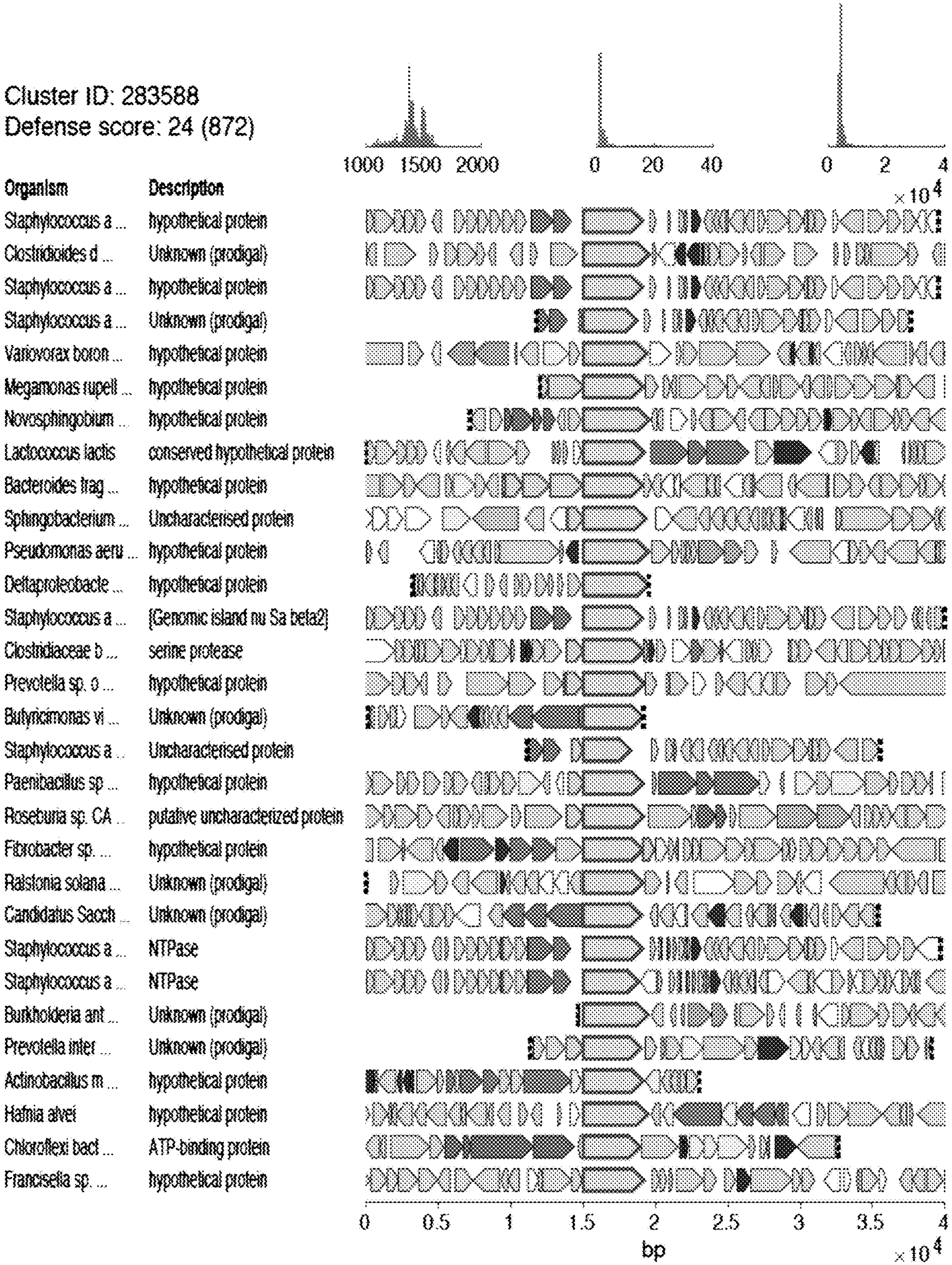
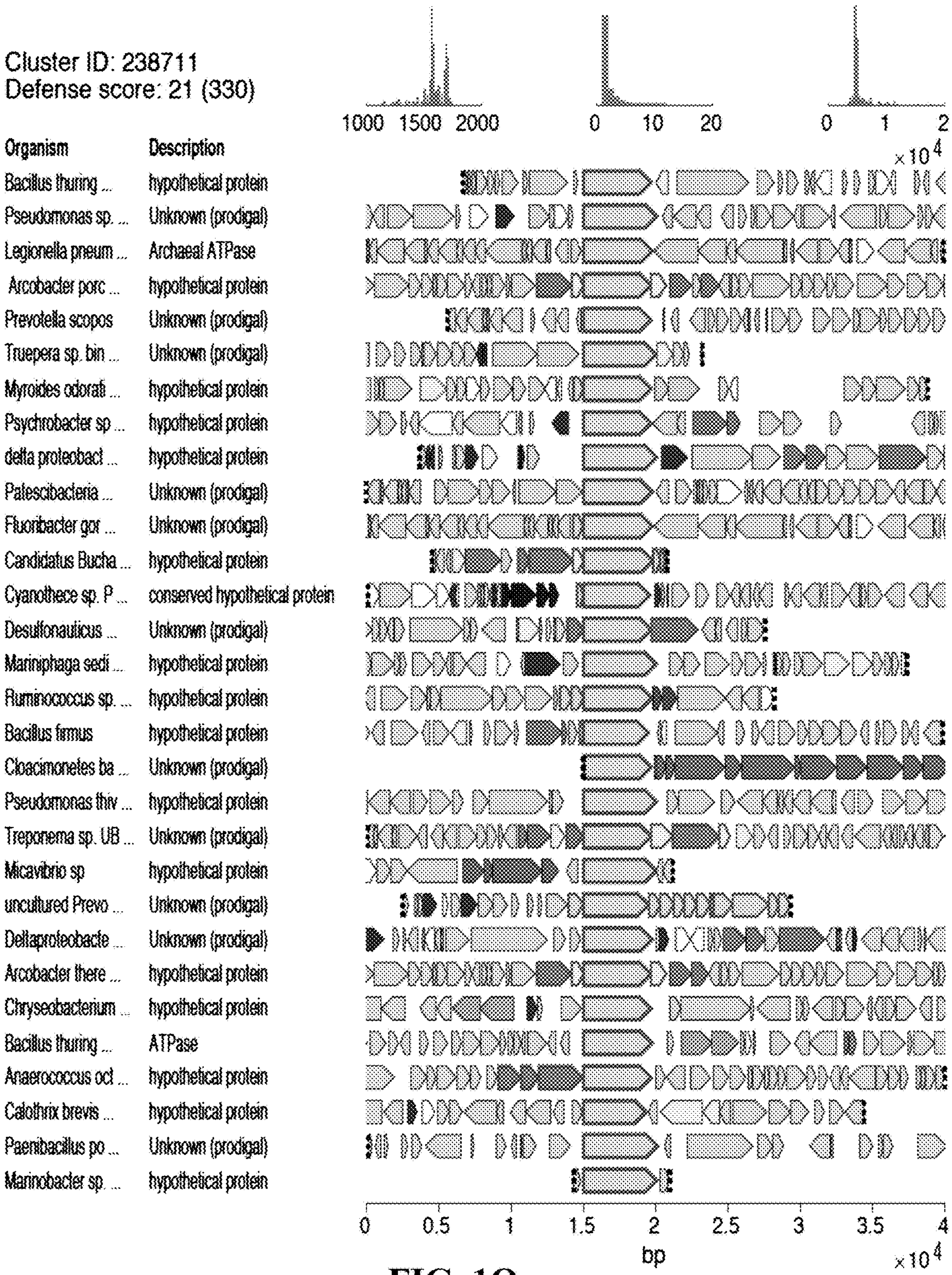

Cluster ID: 100048
Defense score: 24 (185)
1000 2000 3000
0 5 10
0 1 2
×10^4
Organism
Description
Duganella sp. HH ... NACHT domain protein
Solemya velum gi ... hypothetical protein
Mesorhizobium sp ... hypothetical protein
Shewanella frigi ... hypothetical protein
Pseudomonas flav ... Unknown (prodigal)
Acinetobacter sp ... ATPase associated with vario ...
Rhizobiales bact ... Unknown (prodigal)
Acinetobacter soli hypothetical protein
Pseudomonas stut ... Unknown (prodigal)
Klebsiella pneum ... hypothetical protein
Bacteroidetes ba ... Unknown (prodigal)
Pseudoalteromona ... conserved protein of unknown ...
Pseudomonas syri ... NACHT domain-containing protein
Pseudomonas sp. ... Unknown (prodigal)
SAR86 cluster ba ... hypothetical protein
Elusimicrobia ba ... hypothetical protein
Acidobacteriacea ... Unknown (prodigal)
Acinetobacter ra ... Unknown (prodigal)
Acinetobacter sp ... ATPase associated with vario ...
Paraburkholderia ... Cdc6-like AAA superfamily ATPase
Burkholderia pse ... NACHT domain
Pseudoalteromona ... hypothetical protein
Acinetobacter sp ... Unknown (prodigal)
Cupriavidus basi ... Unknown (prodigal)
Elusimicrobia ba ... hypothetical protein
Desulfomicrobium ... Unknown (prodigal)
Rhodopseudomonas ... hypothetical protein
Pseudomonas stut ... Unknown (prodigal)
Methylophaga lon ... hypothetical protein
Thermoanaerobact ... hypothetical protein
0 0.5 1 1.5 2 2.5 3 3.5 4
bp
×10^4

FIG. 1M



Organism	Description
Staphylococcus a ...	hypothetical protein
Clostridioides d ...	Unknown (prodigal)
Staphylococcus a ...	hypothetical protein
Staphylococcus a ...	Unknown (prodigal)
Variovorax boron ...	hypothetical protein
Megamonas rupell ...	hypothetical protein
Novosphingobium ...	hypothetical protein
Lactococcus lactis	conserved hypothetical protein
Bacteroides frag ...	hypothetical protein
Sphingobacterium ...	Uncharacterised protein
Pseudomonas aeru ...	hypothetical protein
Deltaproteobacte ...	hypothetical protein
Staphylococcus a ...	[Genomic island nu Sa beta2]
Clostridiaceae b ...	serine protease
Prevotella sp. o ...	hypothetical protein
Butyricimonas vi ...	Unknown (prodigal)
Staphylococcus a ...	Uncharacterised protein
Paenibacillus sp ...	hypothetical protein
Roseburia sp. CA ...	putative uncharacterized protein
Fibrobacter sp. ...	hypothetical protein
Ralstonia solana ...	Unknown (prodigal)
Candidatus Sacch ...	Unknown (prodigal)
Staphylococcus a ...	NTPase
Staphylococcus a ...	NTPase
Burkholderia ant ...	Unknown (prodigal)
Prevotella inter ...	Unknown (prodigal)
Actinobacillus m ...	hypothetical protein
Hafnia alvei	hypothetical protein
Chloroflexi bact ...	ATP-binding protein
Francisella sp. ...	hypothetical protein

FIG. 1N



Cluster ID: 238711
Defense score: 21 (330)

Organism	Description
Bacillus thuring ...	hypothetical protein
Pseudomonas sp. ...	Unknown (prodigal)
Legionella pneum ...	Archaeal ATPase
Arcobacter porc ...	hypothetical protein
Prevotella scopos	Unknown (prodigal)
Truepera sp. bin ...	Unknown (prodigal)
Myroides odorati ...	hypothetical protein
Psychrobacter sp ...	hypothetical protein
delta proteobact ...	hypothetical protein
Patescibacteria ...	Unknown (prodigal)
Fluoribacter gor ...	Unknown (prodigal)
Candidatus Bucha ...	hypothetical protein
Cyanothece sp. P ...	conserved hypothetical protein
Desulfonauticus ...	Unknown (prodigal)
Mariniphaga sedi ...	hypothetical protein
Ruminococcus sp. ...	hypothetical protein
Bacillus firmus	hypothetical protein
Cloacimonetes ba ...	Unknown (prodigal)
Pseudomonas thiv ...	hypothetical protein
Treponema sp. UB ...	Unknown (prodigal)
Micavibrio sp	hypothetical protein
uncultured Prevo ...	Unknown (prodigal)
Deltaproteobacte ...	Unknown (prodigal)
Arcobacter there ...	hypothetical protein
Chryseobacterium ...	hypothetical protein
Bacillus thuring ...	ATPase
Anaerococcus oct ...	hypothetical protein
Calothrix brevis ...	hypothetical protein
Paenibacillus po ...	Unknown (prodigal)
Marinobacter sp. ...	hypothetical protein

FIG. 1O

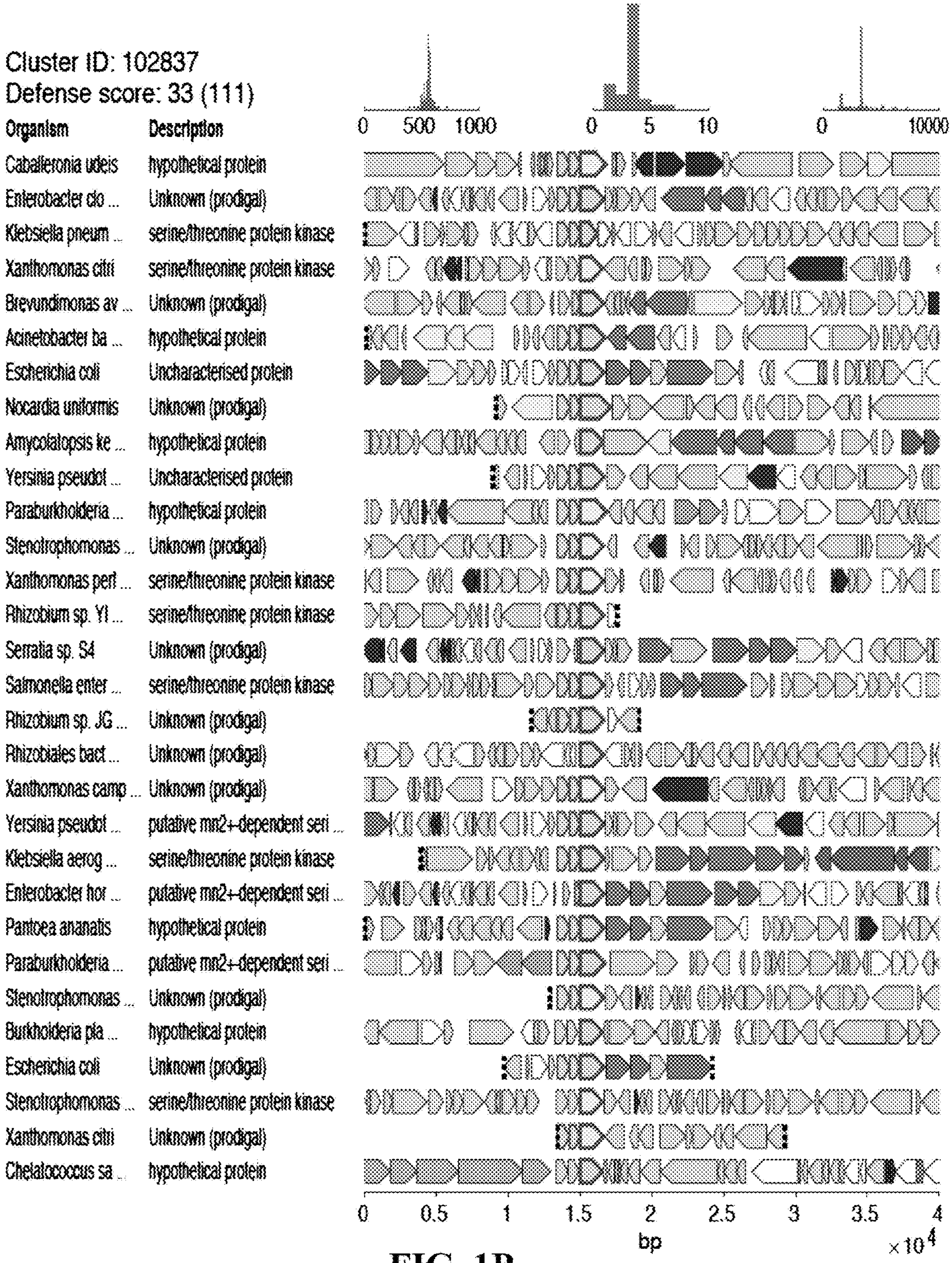
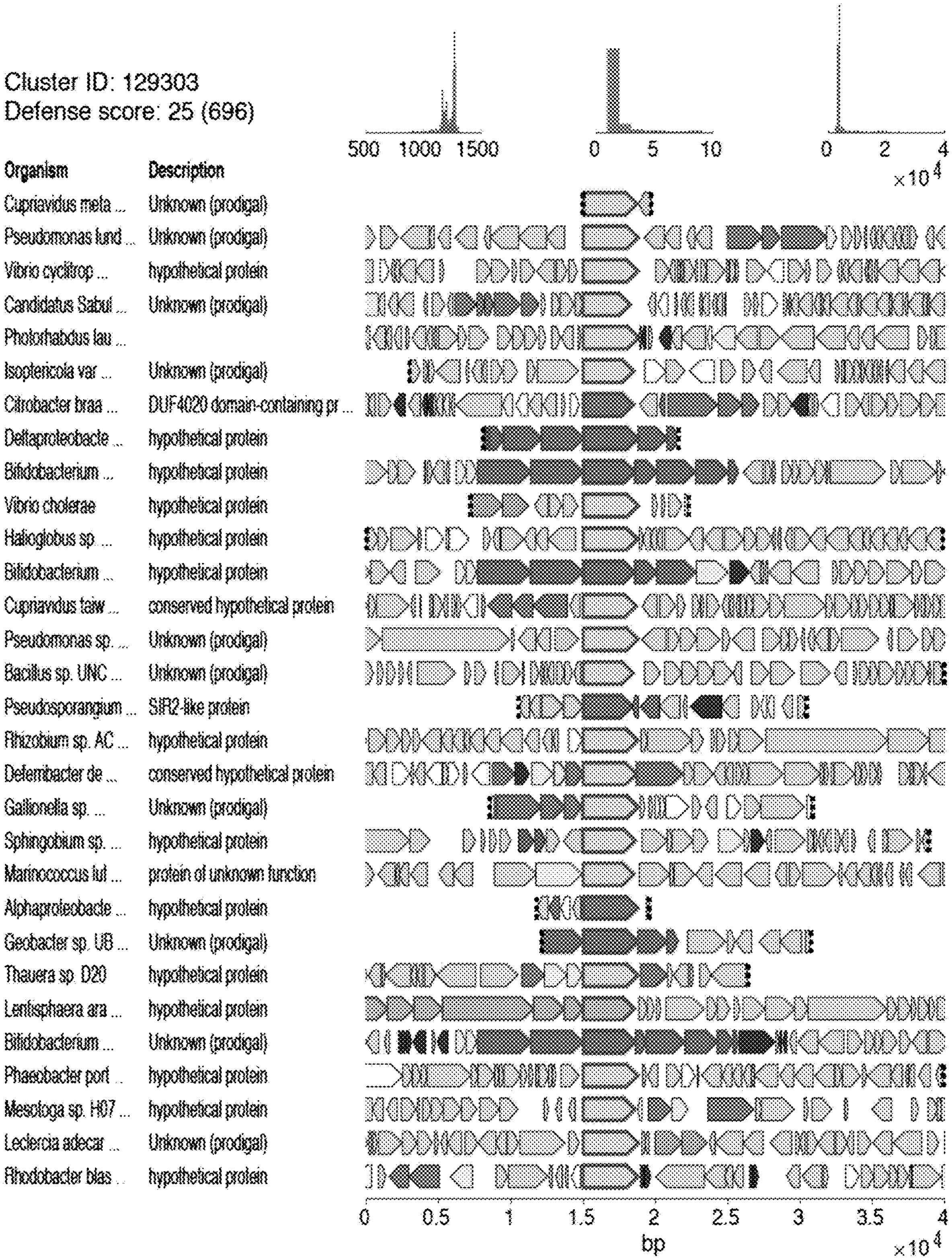

Cluster ID: 102837
Defense score: 33 (111)
Organism
Description
0 500 1000
0 5 10
0 10000
Caballeronia udeis hypothetical protein
Enterobacter clo ... Unknown (prodigal)
Klebsiella pneum ... serine/threonine protein kinase
Xanthomonas citri serine/threonine protein kinase
Brevundimonas av ... Unknown (prodigal)
Acinetobacter ba ... hypothetical protein
Escherichia coli Uncharacterised protein
Nocardia uniformis Unknown (prodigal)
Amycolatopsis ke ... hypothetical protein
Yersinia pseudot ... Uncharacterised protein
Paraburkholderia ... hypothetical protein
Stenotrophomonas ... Unknown (prodigal)
Xanthomonas perf ... serine/threonine protein kinase
Rhizobium sp. YI ... serine/threonine protein kinase
Serratia sp. S4 Unknown (prodigal)
Salmonella enter ... serine/threonine protein kinase
Rhizobium sp. JG ... Unknown (prodigal)
Rhizobiales bact ... Unknown (prodigal)
Xanthomonas camp ... Unknown (prodigal)
Yersinia pseudot ... putative mn2+-dependent seri ...
Klebsiella aerog ... serine/threonine protein kinase
Enterobacter hor ... putative mn2+-dependent seri ...
Pantoea ananatis hypothetical protein
Paraburkholderia ... putative mn2+-dependent seri ...
Stenotrophomonas ... Unknown (prodigal)
Burkholderia pla ... hypothetical protein
Escherichia coli Unknown (prodigal)
Stenotrophomonas ... serine/threonine protein kinase
Xanthomonas citri Unknown (prodigal)
Chelatococcus sa ... hypothetical protein
0 0.5 1 1.5 2 2.5 3 3.5 4
bp
×10^4

FIG. 1P



Organism	Description
Cupriavidus meta ...	Unknown (prodigal)
Pseudomonas lund ...	Unknown (prodigal)
Vibrio cyclitrop ...	hypothetical protein
Candidatus Sabul ...	Unknown (prodigal)
Photorhabdus lau ...	
Isoptericola var ...	Unknown (prodigal)
Citrobacter braa ...	DUF4020 domain-containing pr ...
Deltaproteobacte ...	hypothetical protein
Bifidobacterium ...	hypothetical protein
Vibrio cholerae	hypothetical protein
Halioglobus sp ...	hypothetical protein
Bifidobacterium ...	hypothetical protein
Cupriavidus taiw ...	conserved hypothetical protein
Pseudomonas sp. ...	Unknown (prodigal)
Bacillus sp. UNC ...	Unknown (prodigal)
Pseudosporangium ...	SIR2-like protein
Rhizobium sp. AC ...	hypothetical protein
Deferribacter de ...	conserved hypothetical protein
Gallionella sp. ...	Unknown (prodigal)
Sphingobium sp. ...	hypothetical protein
Marinococcus lut ...	protein of unknown function
Alphaproteobacte ...	hypothetical protein
Geobacter sp. UB ...	Unknown (prodigal)
Thauera sp. D20	hypothetical protein
Lentisphaera ara ...	hypothetical protein
Bifidobacterium ...	Unknown (prodigal)
Phaeobacter port ..	hypothetical protein
Mesotoga sp. H07 ...	hypothetical protein
Leclercia adecar ...	Unknown (prodigal)
Rhodobacter blas ..	hypothetical protein

FIG. 1Q

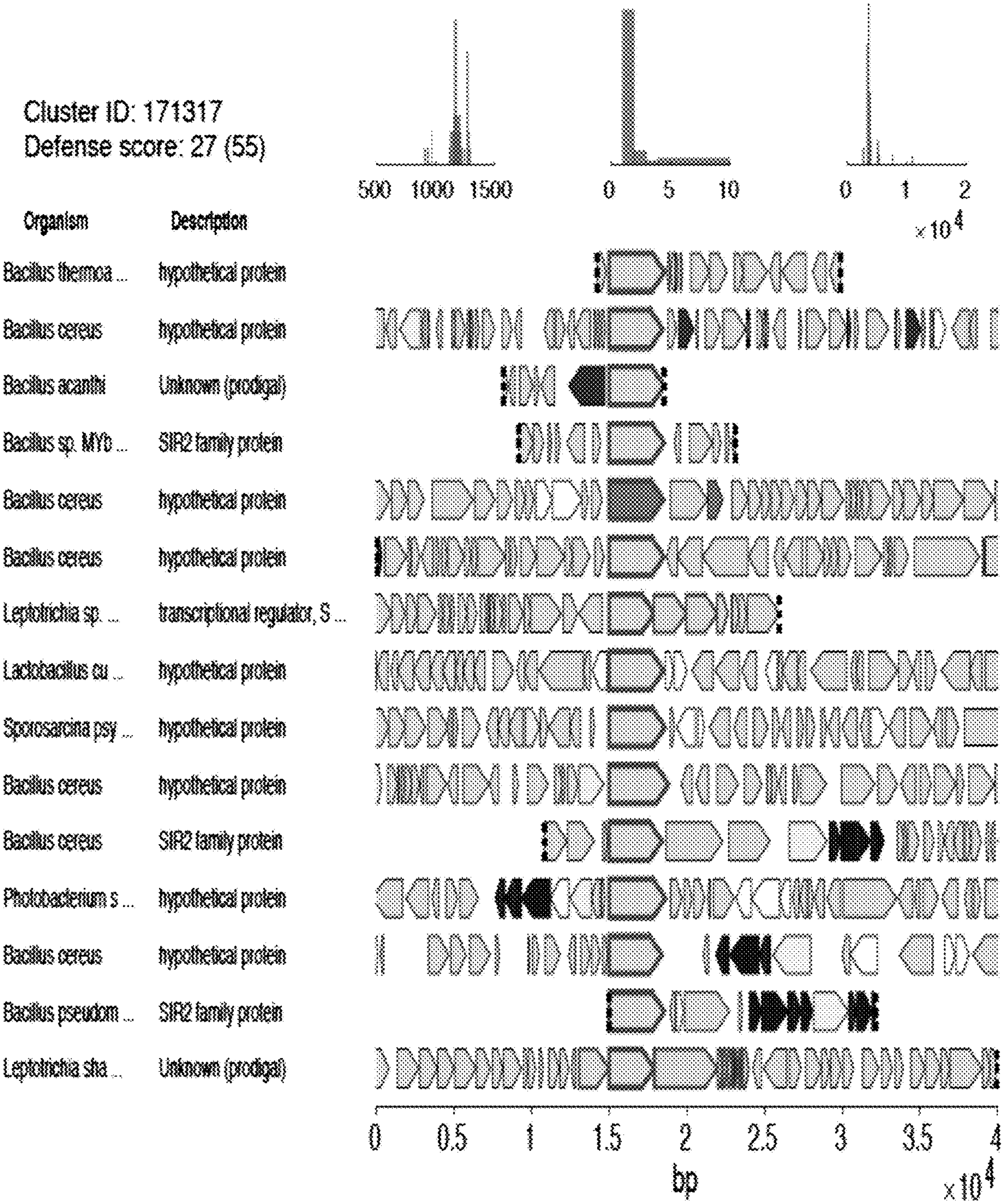

Cluster ID: 171317
Defense score: 27 (55)
500
1000
1500
0
5
10
0
1
2
×10⁴
Organism
Description
Bacillus thermoa ...
hypothetical protein
Bacillus cereus
hypothetical protein
Bacillus acanthi
Unknown (prodigal)
Bacillus sp. MYb ...
SIR2 family protein
Bacillus cereus
hypothetical protein
Bacillus cereus
hypothetical protein
Leptotrichia sp. ...
transcriptional regulator, S ...
Lactobacillus cu ...
hypothetical protein
Sporosarcina psy ...
hypothetical protein
Bacillus cereus
hypothetical protein
Bacillus cereus
SIR2 family protein
Photobacterium s ...
hypothetical protein
Bacillus cereus
hypothetical protein
Bacillus pseudom ...
SIR2 family protein
Leptotrichia sha ...
Unknown (prodigal)
0
0.5
1
1.5
2
2.5
3
3.5
4
bp
×10⁴

FIG. 1R


Cluster ID: 247472
Defense score: 21 (81)
600
800
1000
0
10
20
0
1
2
×10^4
Organism
Description
Arcobacter cryae ...
Unknown (prodigal)
Pseudomonas vero ...
hypothetical protein
Morganella morga ...
hypothetical protein
Enterobacter sp. ...
hypothetical protein
Leminorella grim ...
hypothetical protein
Vibrio lentus
hypothetical protein
Klebsiella vari ...
hypothetical protein
Candidatus Parcu ...
hypothetical protein
Lachnoclostridiu ...
hypothetical protein
Enterobacter sp. ...
Unknown (prodigal)
Anaerotignum lac ...
hypothetical protein
Clostridium sp. ...
hypothetical protein
Paraburkholderia ...
Unknown (prodigal)
Caballeronia gri ...
hypothetical protein
Xanthomonas nast ...
hypothetical protein
Lachnoanaerobacu ...
hypothetical protein
Alphaproteobacte ...
Unknown (prodigal)
0
0.5
1
1.5
2
2.5
3
3.5
4
bp
×10^4

FIG. 1S

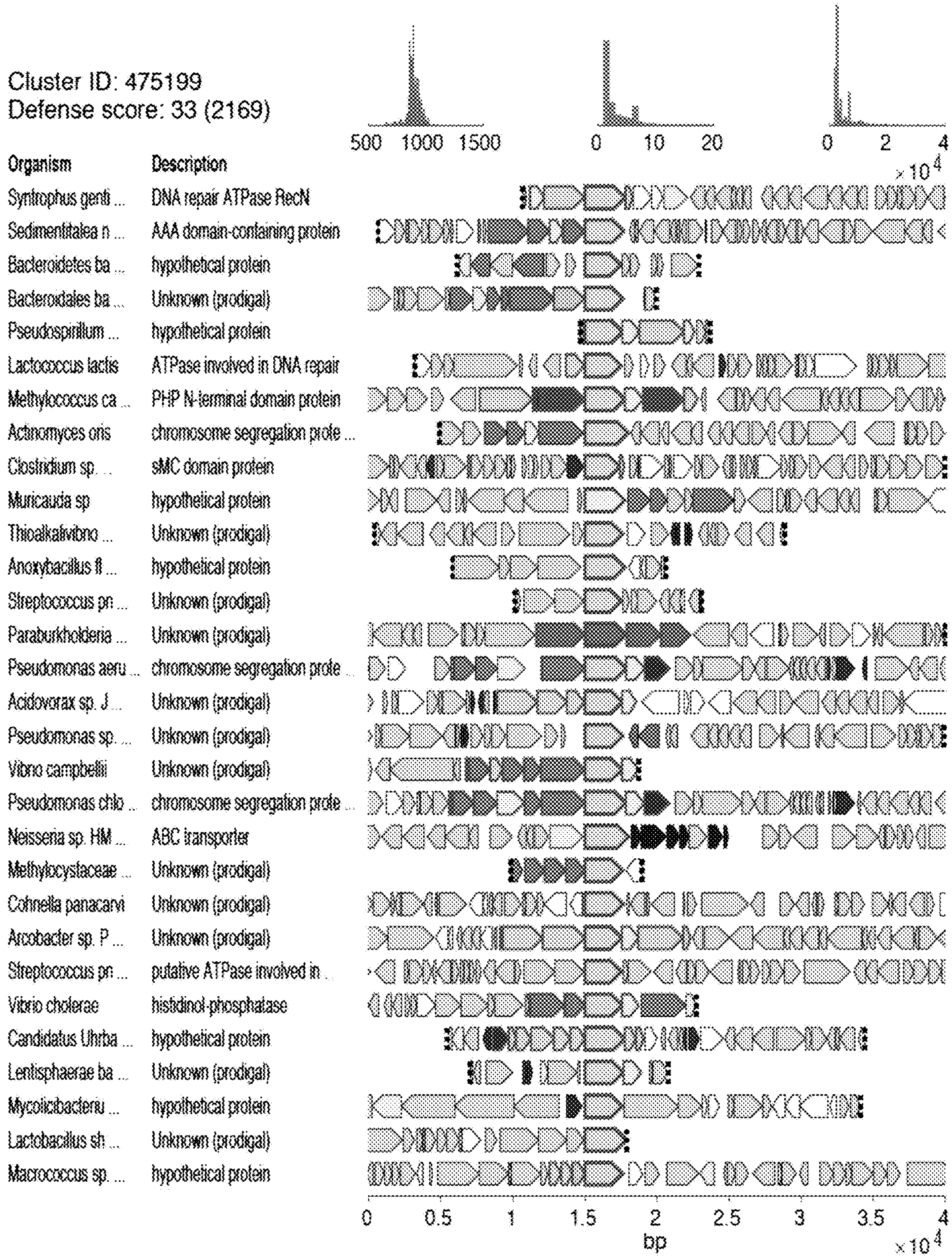

Cluster ID: 475199
Defense score: 33 (2169)
500
1000
1500
0
10
20
0
2
4
×10⁴
Organism
Description
Syntrophus genti ...
DNA repair ATPase RecN
Sedimentitalea n ...
AAA domain-containing protein
Bacteroidetes ba ...
hypothetical protein
Bacteroidales ba ...
Unknown (prodigal)
Pseudospirillum ...
hypothetical protein
Lactococcus lactis
ATPase involved in DNA repair
Methylococcus ca ...
PHP N-terminal domain protein
Actinomyces oris
chromosome segregation prote ...
Clostridium sp. ...
sMC domain protein
Muricauda sp
hypothetical protein
Thioalkalivibrio ...
Unknown (prodigal)
Anoxybacillus fl ...
hypothetical protein
Streptococcus pn ...
Unknown (prodigal)
Paraburkholderia ...
Unknown (prodigal)
Pseudomonas aeru ...
chromosome segregation prote ...
Acidovorax sp. J ...
Unknown (prodigal)
Pseudomonas sp. ...
Unknown (prodigal)
Vibrio campbellii
Unknown (prodigal)
Pseudomonas chlo ...
chromosome segregation prote ...
Neisseria sp. HM ...
ABC transporter
Methylocystaceae ...
Unknown (prodigal)
Cohnella panacarvi
Unknown (prodigal)
Arcobacter sp. P ...
Unknown (prodigal)
Streptococcus pn ...
putative ATPase involved in ...
Vibrio cholerae
histidinol-phosphatase
Candidatus Uhrba ...
hypothetical protein
Lentisphaerae ba ...
Unknown (prodigal)
Mycolicibacteriu ...
hypothetical protein
Lactobacillus sh ...
Unknown (prodigal)
Macrococcus sp. ...
hypothetical protein
0
0.5
1
1.5
2
2.5
3
3.5
4
bp
×10⁴

FIG. 1T

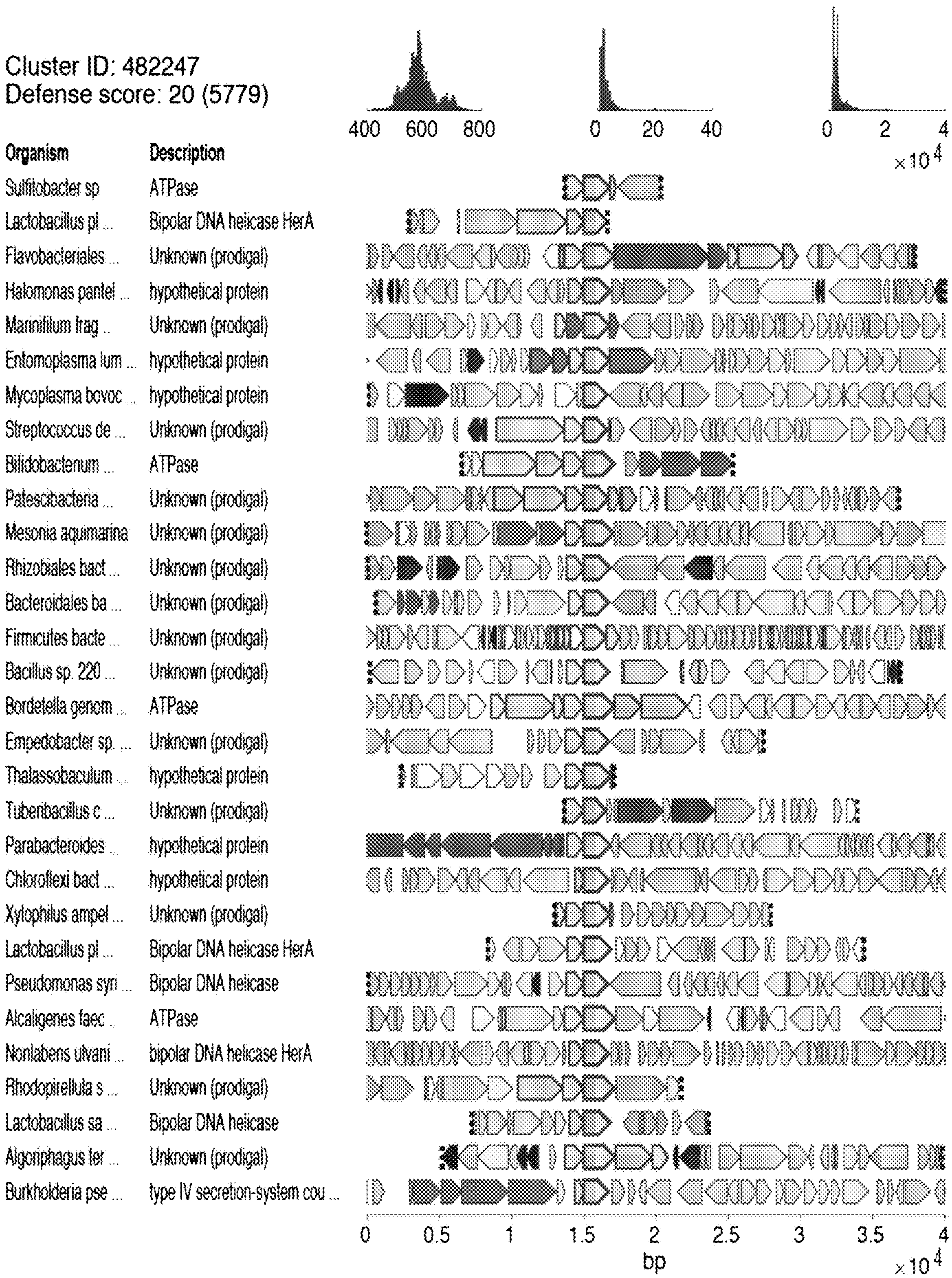
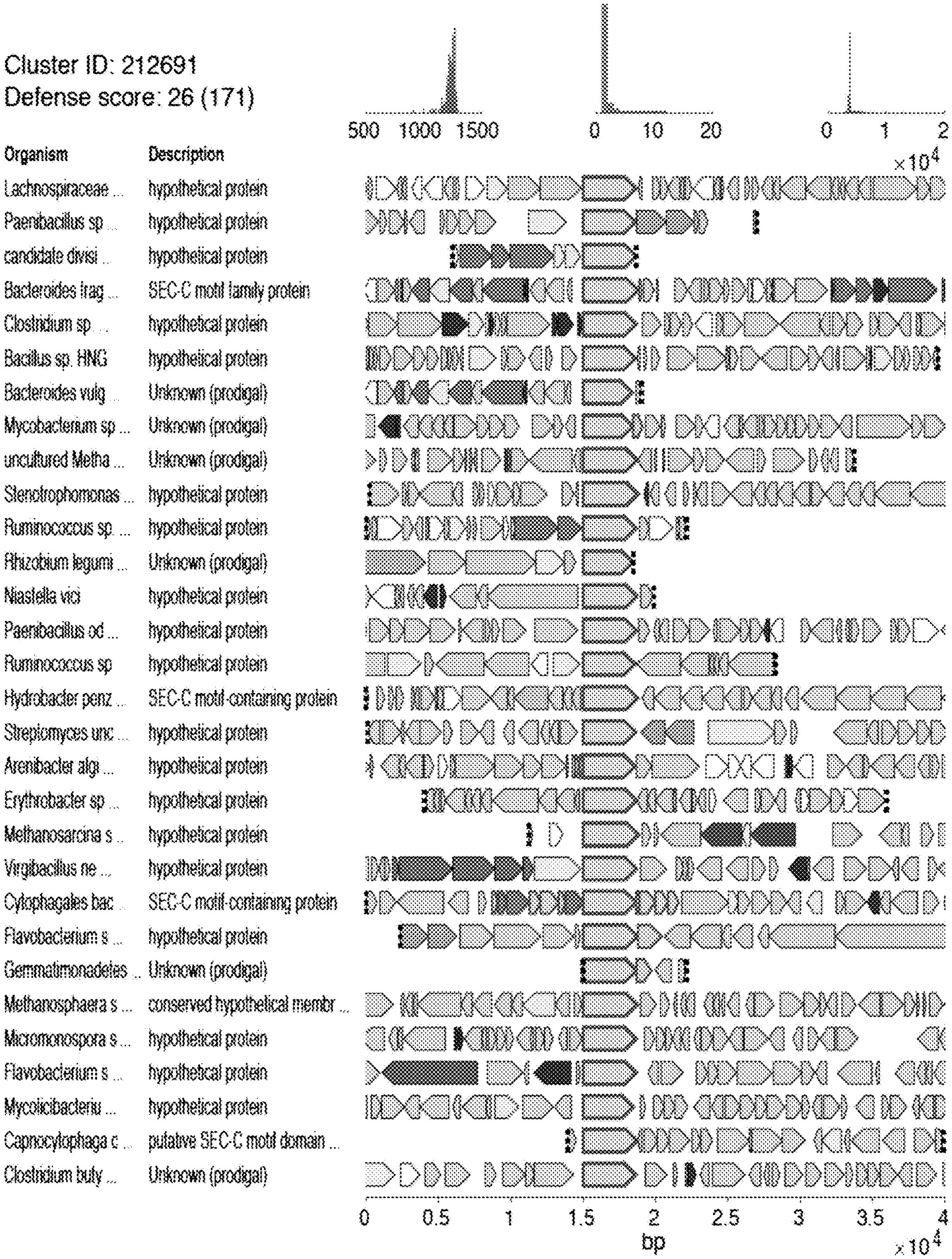

Cluster ID: 482247
Defense score: 20 (5779)
400
600
800
0
20
40
0
2
4
×10^4
Organism
Description
Sulfitobacter sp
ATPase
Lactobacillus pl ...
Bipolar DNA helicase HerA
Flavobacteriales ...
Unknown (prodigal)
Halomonas pantel ...
hypothetical protein
Marinifilum frag ..
Unknown (prodigal)
Entomoplasma lum ...
hypothetical protein
Mycoplasma bovoc ...
hypothetical protein
Streptococcus de ...
Unknown (prodigal)
Bifidobacterium ...
ATPase
Patescibacteria ...
Unknown (prodigal)
Mesonia aquimarina
Unknown (prodigal)
Rhizobiales bact ...
Unknown (prodigal)
Bacteroidales ba ...
Unknown (prodigal)
Firmicutes bacte ...
Unknown (prodigal)
Bacillus sp. 220 ...
Unknown (prodigal)
Bordetella genom ...
ATPase
Empedobacter sp. ...
Unknown (prodigal)
Thalassobaculum ...
hypothetical protein
Tuberibacillus c ...
Unknown (prodigal)
Parabacteroides ...
hypothetical protein
Chloroflexi bact ...
hypothetical protein
Xylophilus ampel ...
Unknown (prodigal)
Lactobacillus pl ...
Bipolar DNA helicase HerA
Pseudomonas syn ...
Bipolar DNA helicase
Alcaligenes faec ...
ATPase
Nonlabens ulvani ...
bipolar DNA helicase HerA
Rhodopirellula s ...
Unknown (prodigal)
Lactobacillus sa ...
Bipolar DNA helicase
Algoriphagus ter ...
Unknown (prodigal)
Burkholderia pse ...
type IV secretion-system cou ...
0
0.5
1
1.5
2
2.5
3
3.5
4
bp
×10^4

FIG. 1U



Organism	Description
Lachnospiraceae ...	hypothetical protein
Paenibacillus sp ...	hypothetical protein
candidate divisi ...	hypothetical protein
Bacteroides frag ...	SEC-C motif family protein
Clostridium sp ...	hypothetical protein
Bacillus sp. HNG	hypothetical protein
Bacteroides vulg ...	Unknown (prodigal)
Mycobacterium sp ...	Unknown (prodigal)
uncultured Metha ...	Unknown (prodigal)
Stenotrophomonas ...	hypothetical protein
Ruminococcus sp. ...	hypothetical protein
Rhizobium legumi ...	Unknown (prodigal)
Niastella vici	hypothetical protein
Paenibacillus od ...	hypothetical protein
Ruminococcus sp	hypothetical protein
Hydrobacter penz ...	SEC-C motif-containing protein
Streptomyces unc ...	hypothetical protein
Arenibacter algi ...	hypothetical protein
Erythrobacter sp ...	hypothetical protein
Methanosarcina s ...	hypothetical protein
Virgibacillus ne ...	hypothetical protein
Cytophagales bac ...	SEC-C motif-containing protein
Flavobacterium s ...	hypothetical protein
Gemmatimonadetes ...	Unknown (prodigal)
Methanosphaera s ...	conserved hypothetical membr ...
Micromonospora s ...	hypothetical protein
Flavobacterium s ...	hypothetical protein
Mycolicibacteriu ...	hypothetical protein
Capnocytophaga c ...	putative SEC-C motif domain ...
Clostridium buty ...	Unknown (prodigal)

FIG. 1V

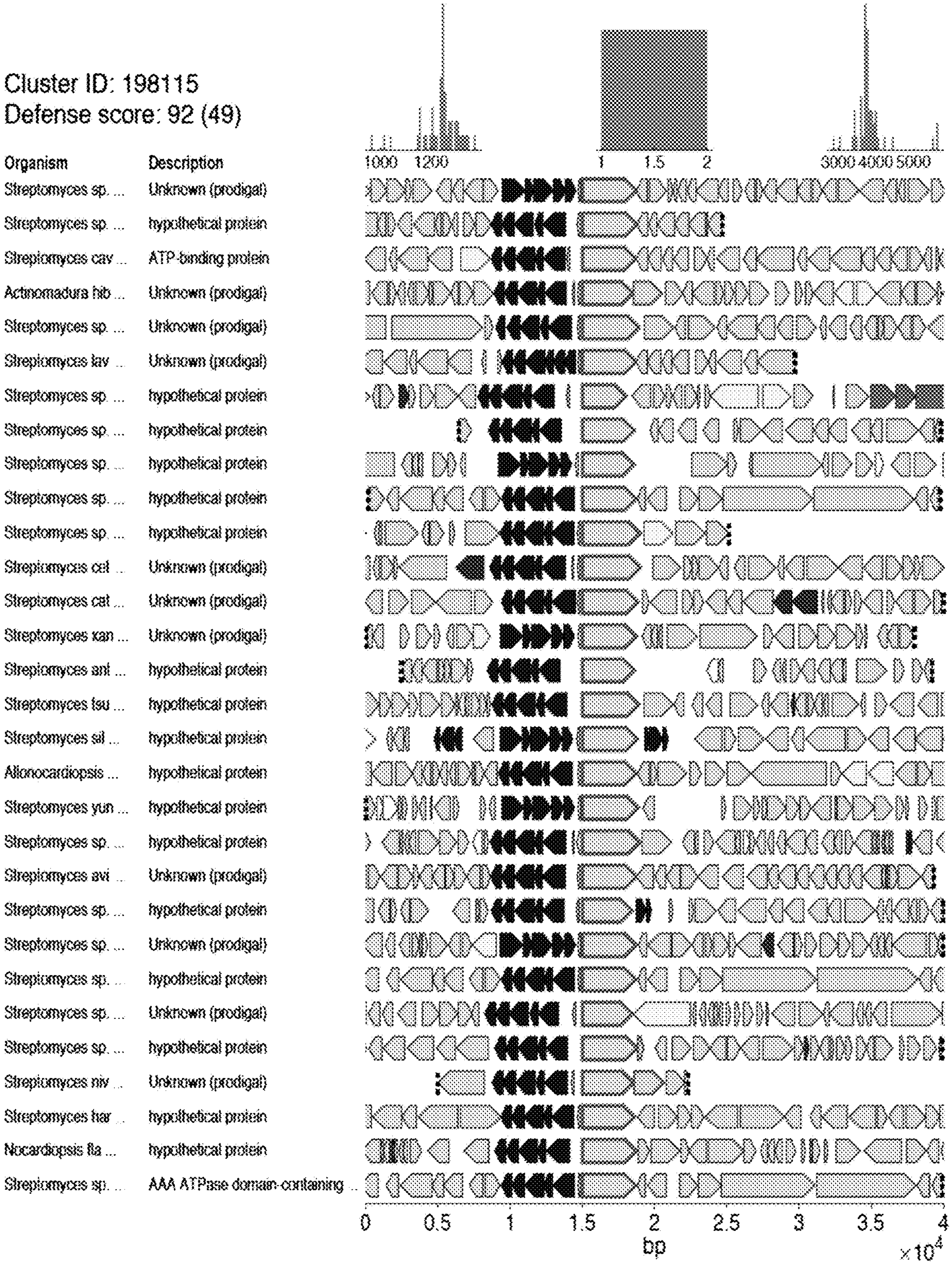
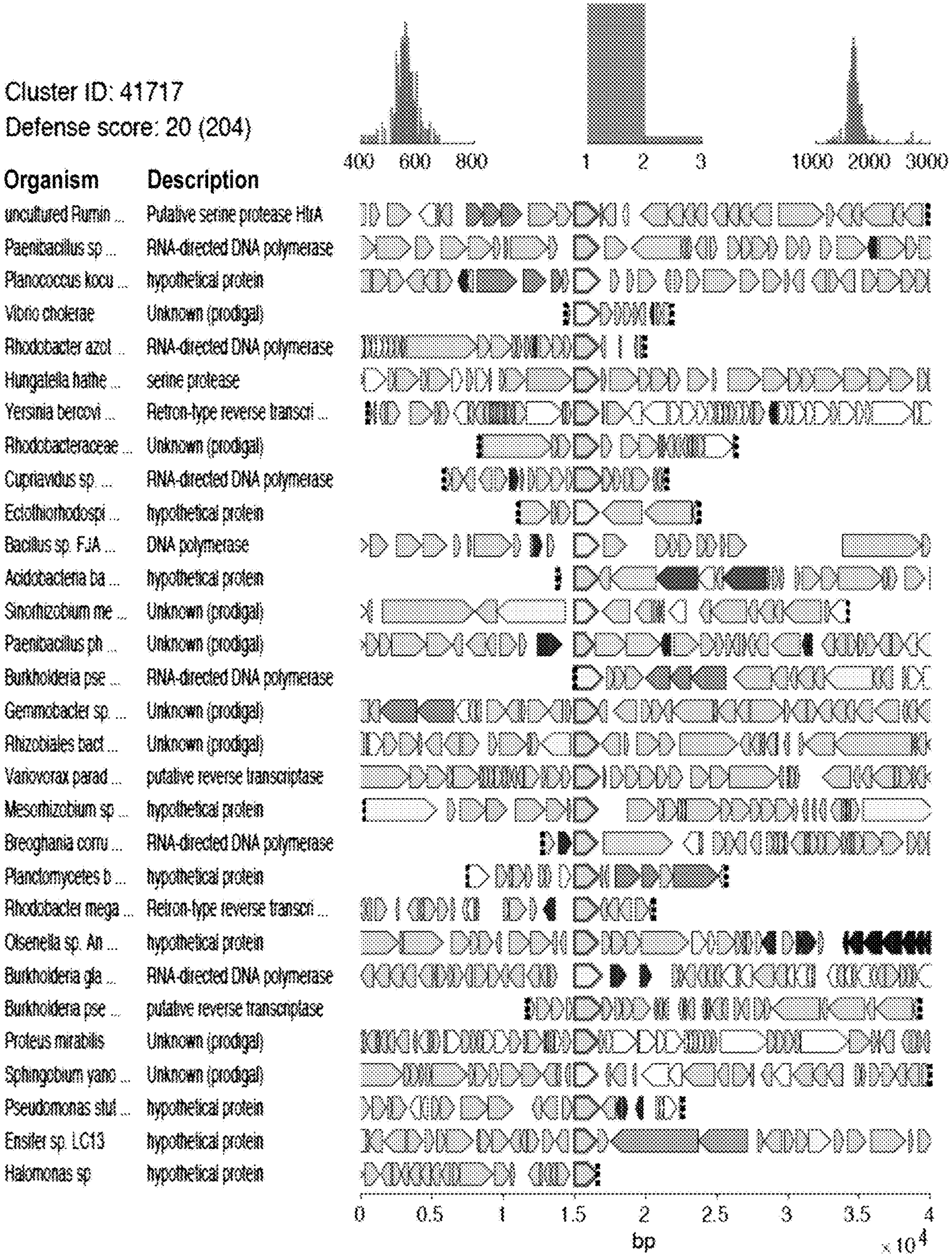

Cluster ID: 198115
Defense score: 92 (49)
1000
1200
1
1.5
2
3000
4000
5000
Organism
Description
Streptomyces sp. ...
Unknown (prodigal)
Streptomyces sp. ...
hypothetical protein
Streptomyces cav ...
ATP-binding protein
Actinomadura hib ...
Unknown (prodigal)
Streptomyces sp. ...
Unknown (prodigal)
Streptomyces lav ...
Unknown (prodigal)
Streptomyces sp. ...
hypothetical protein
Streptomyces sp. ...
hypothetical protein
Streptomyces sp. ...
hypothetical protein
Streptomyces sp. ...
hypothetical protein
Streptomyces sp. ...
hypothetical protein
Streptomyces cel ...
Unknown (prodigal)
Streptomyces cat ...
Unknown (prodigal)
Streptomyces xan ...
Unknown (prodigal)
Streptomyces ant ...
hypothetical protein
Streptomyces tsu ...
hypothetical protein
Streptomyces sil ...
hypothetical protein
Allonocardiopsis ...
hypothetical protein
Streptomyces yun ...
hypothetical protein
Streptomyces sp. ...
hypothetical protein
Streptomyces avi ...
Unknown (prodigal)
Streptomyces sp. ...
hypothetical protein
Streptomyces sp. ...
Unknown (prodigal)
Streptomyces sp. ...
hypothetical protein
Streptomyces sp. ...
Unknown (prodigal)
Streptomyces sp. ...
hypothetical protein
Streptomyces niv ...
Unknown (prodigal)
Streptomyces har ...
hypothetical protein
Nocardiopsis fla ...
hypothetical protein
Streptomyces sp. ...
AAA ATPase domain-containing ...
0
0.5
1
1.5
2
2.5
3
3.5
4
bp
×10^4

FIG. 1W


Cluster ID: 41717
Defense score: 20 (204)

Organism	Description
uncultured Rumin ...	Putative serine protease HtrA
Paenibacillus sp ...	RNA-directed DNA polymerase
Planococcus kocu ...	hypothetical protein
Vibrio cholerae	Unknown (prodigal)
Rhodobacter azot ...	RNA-directed DNA polymerase
Hungatella hathe ...	serine protease
Yersinia bercovi ...	Retron-type reverse transcri ...
Rhodobacteraceae ...	Unknown (prodigal)
Cupriavidus sp. ...	RNA-directed DNA polymerase
Ectothiorhodospi ...	hypothetical protein
Bacillus sp. FJA ...	DNA polymerase
Acidobacteria ba ...	hypothetical protein
Sinorhizobium me ...	Unknown (prodigal)
Paenibacillus ph ...	Unknown (prodigal)
Burkholderia pse ...	RNA-directed DNA polymerase
Gemmobacter sp. ...	Unknown (prodigal)
Rhizobiales bact ...	Unknown (prodigal)
Variovorax parad ...	putative reverse transcriptase
Mesorhizobium sp ...	hypothetical protein
Breoghania corru ...	RNA-directed DNA polymerase
Planctomycetes b ...	hypothetical protein
Rhodobacter mega ...	Retron-type reverse transcri ...
Olsenella sp. An ...	hypothetical protein
Burkholderia gla ...	RNA-directed DNA polymerase
Burkholderia pse ...	putative reverse transcriptase
Proteus mirabilis	Unknown (prodigal)
Sphingobium yano ...	Unknown (prodigal)
Pseudomonas stut ...	hypothetical protein
Ensifer sp. LC13	hypothetical protein
Halomonas sp	hypothetical protein

FIG. 1X

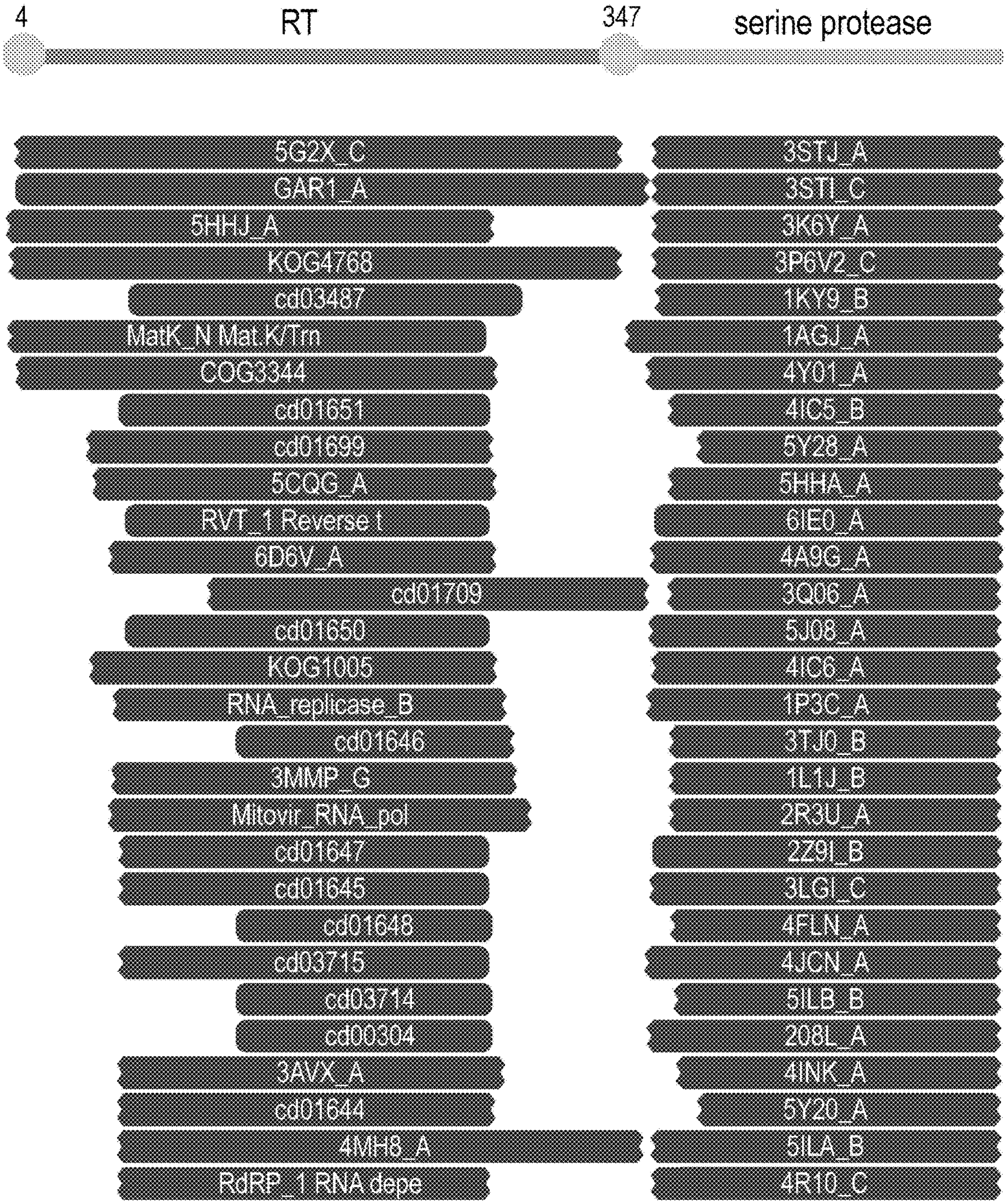

4
RT
347
serine protease
5G2X_C
3STJ_A
GAR1_A
3STI_C
5HHJ_A
3K6Y_A
KOG4768
3P6V2_C
cd03487
1KY9_B
MatK_N Mat.K/Trn
1AGJ_A
COG3344
4Y01_A
cd01651
4IC5_B
cd01699
5Y28_A
5CQG_A
5HHA_A
RVT_1 Reverse t
6IE0_A
6D6V_A
4A9G_A
cd01709
3Q06_A
cd01650
5J08_A
KOG1005
4IC6_A
RNA_replicase_B
1P3C_A
cd01646
3TJ0_B
3MMP_G
1L1J_B
Mitovir_RNA_pol
2R3U_A
cd01647
2Z9I_B
cd01645
3LGI_C
cd01648
4FLN_A
cd03715
4JCN_A
cd03714
5ILB_B
cd00304
2O8L_A
3AVX_A
4INK_A
cd01644
5Y20_A
4MH8_A
5ILA_B
RdRP_1 RNA depe
4R10_C

FIG. 1X (Cont'd)

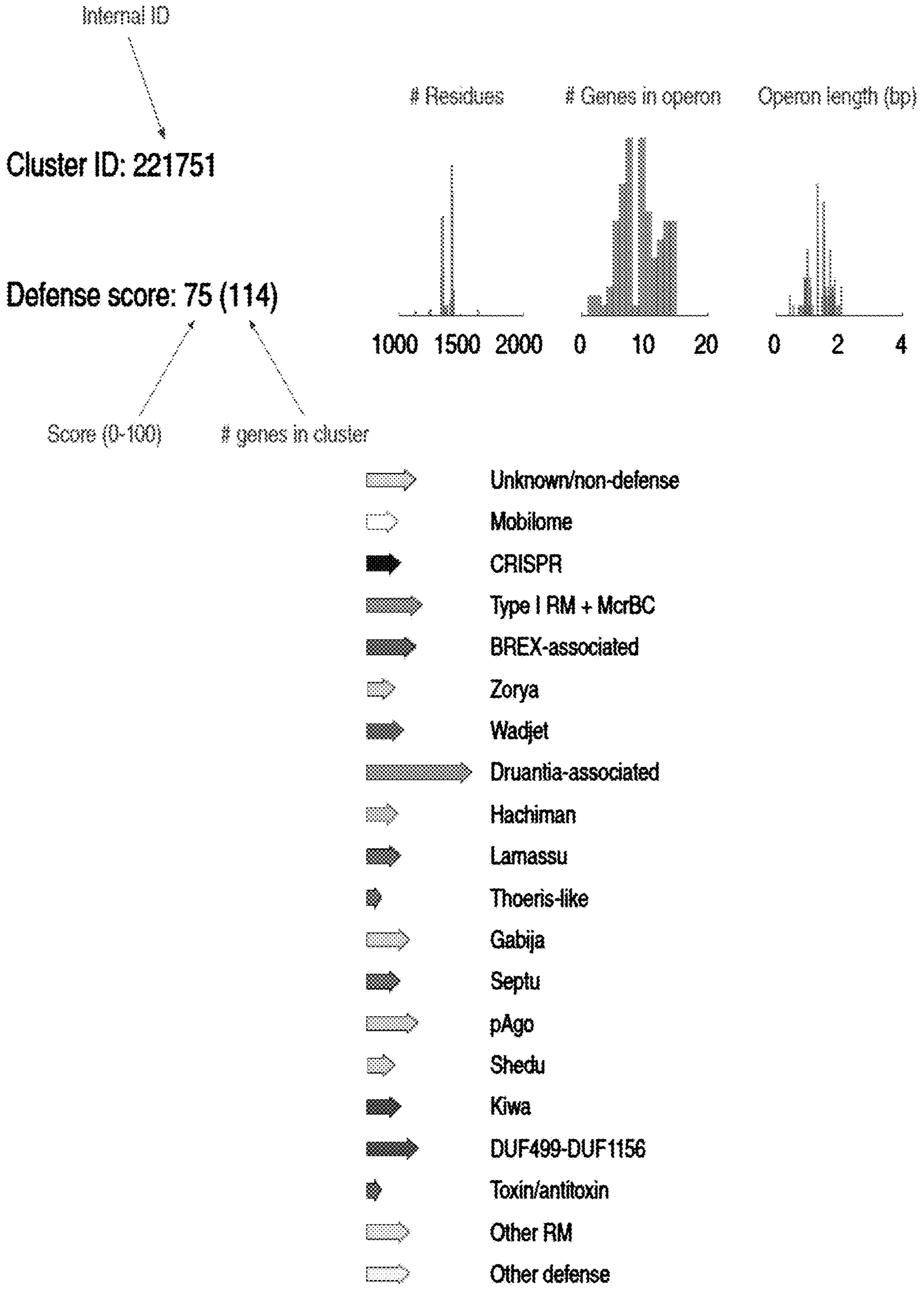

Internal ID
Cluster ID: 221751
Defense score: 75 (114)
Score (0-100)
genes in cluster
Residues
1000
1500
2000
Genes in operon
0
10
20
Operon length (bp)
0
2
4
Unknown/non-defense
Mobilome
CRISPR
Type I RM + McrBC
BREX-associated
Zorya
Wadjet
Druantia-associated
Hachiman
Lamassu
Thoeris-like
Gabija
Septu
pAgo
Shedu
Kiwa
DUF499-DUF1156
Toxin/antitoxin
Other RM
Other defense

FIG. 1Y

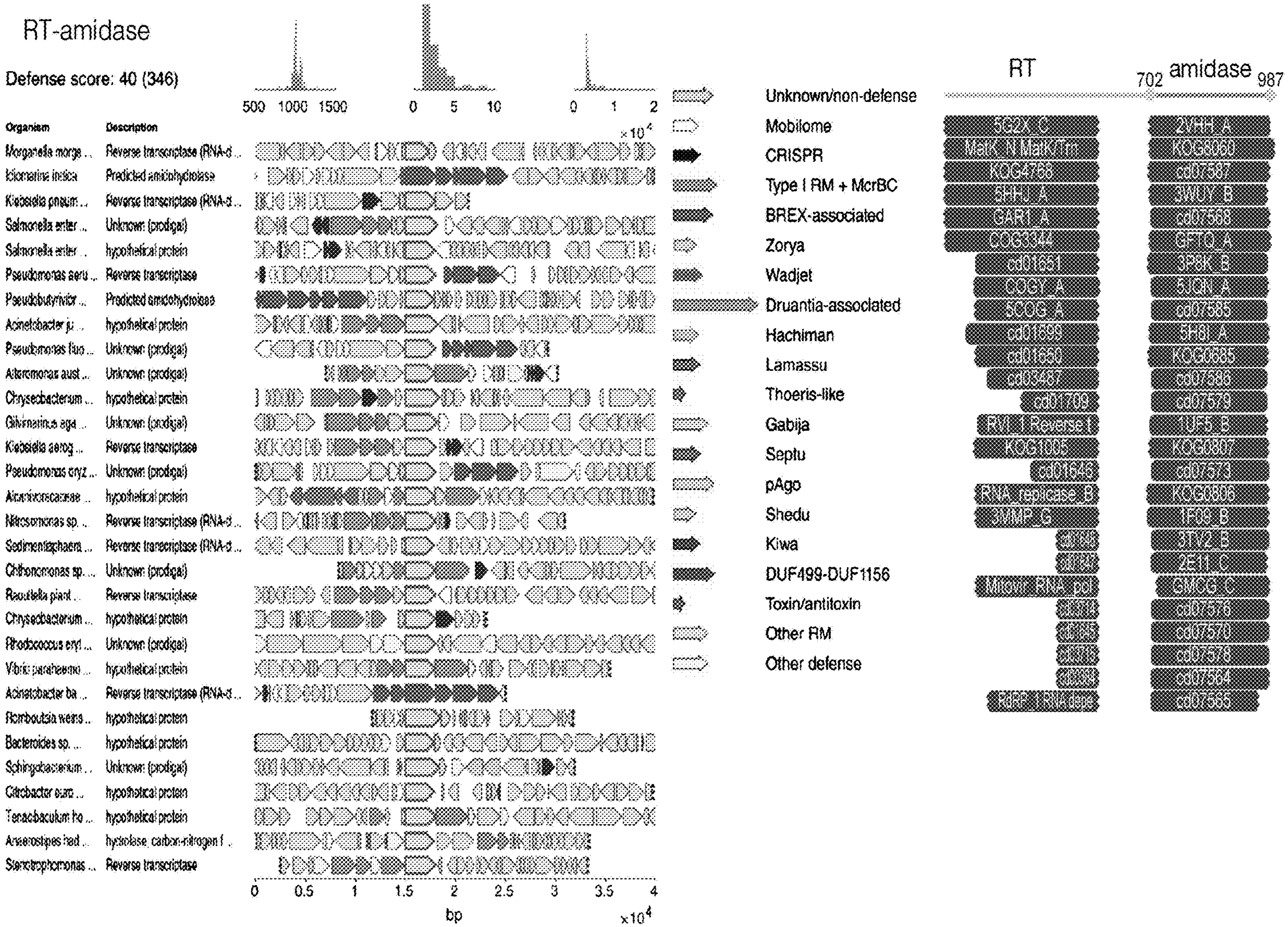

RT-amidase
Defense score: 40 (346)
500
1000
1500
0
5
10
0
1
2
×10⁴
Organism
Description
Morganella morgs ...
Reverse transcriptase (RNA-d ...
Idiomarina indica
Predicted amidohydrolase
Klebsiella pneum ...
Reverse transcriptase (RNA-d ...
Salmonella enter ...
Unknown (prodigal)
Salmonella enter ...
hypothetical protein
Pseudomonas aeru ...
Reverse transcriptase
Pseudobutyrivibr ...
Predicted amidohydrolase
Acinetobacter ju ...
hypothetical protein
Pseudomonas fluo ...
Unknown (prodigal)
Alteromonas aust ...
Unknown (prodigal)
Chryseobacterium ...
hypothetical protein
Gilvimarinus aga ...
Unknown (prodigal)
Klebsiella aerog ...
Reverse transcriptase
Pseudomonas oryz ...
Unknown (prodigal)
Alcanivoracaceae ...
hypothetical protein
Nitrosomonas sp. ...
Reverse transcriptase (RNA-d ...
Sedimentisphaera ...
Reverse transcriptase (RNA-d ...
Chthonomonas sp. ...
Unknown (prodigal)
Raoultella plant ...
Reverse transcriptase
Chryseobacterium ...
hypothetical protein
Rhodococcus eryt ...
Unknown (prodigal)
Vibrio parahaemo ...
hypothetical protein
Acinetobacter ba ...
Reverse transcriptase (RNA-d ...
Romboutsia weins ...
hypothetical protein
Bacteroides sp. ...
hypothetical protein
Sphingobacterium ...
Unknown (prodigal)
Citrobacter euro ...
hypothetical protein
Tenacibaculum ho ...
hypothetical protein
Anaerostipes had ...
hydrolase, carbon-nitrogen f ...
Stenotrophomonas ...
Reverse transcriptase
0
0.5
1
1.5
2
2.5
3
3.5
4
bp
×10⁴
Unknown/non-defense
Mobilome
CRISPR
Type I RM + McrBC
BREX-associated
Zorya
Wadjet
Druantia-associated
Hachiman
Lamassu
Thoeris-like
Gabija
Septu
pAgo
Shedu
Kiwa
DUF499-DUF1156
Toxin/antitoxin
Other RM
Other defense
RT
702
amidase
987
5G2X_C
MatK_N MatK/Trn
KOG4768
5HHJ_A
GAR1_A
COG3344
cd01651
COGY_A
5COG_A
cd01699
cd01650
cd03487
cd01709
RVT_1 Reverse t
KOG1005
cd01646
RNA_replicase_B
3MMP_G
cd01648
cd01647
Mitovir_RNA_pol
cd03714
cd01645
cd03715
cd00304
RdRP_1 RNA depe
2VHH_A
KOG8060
cd07587
3WUY_B
cd07568
GFTQ_A
3P8K_B
5JQN_A
cd07585
5H8I_A
KOG0685
cd07586
cd07579
1UF5_B
KOG0807
cd07573
KOG0806
1F09_B
3TV2_B
2E11_C
GMCG_C
cd07576
cd07570
cd07578
cd07564
cd07565

FIG. 2

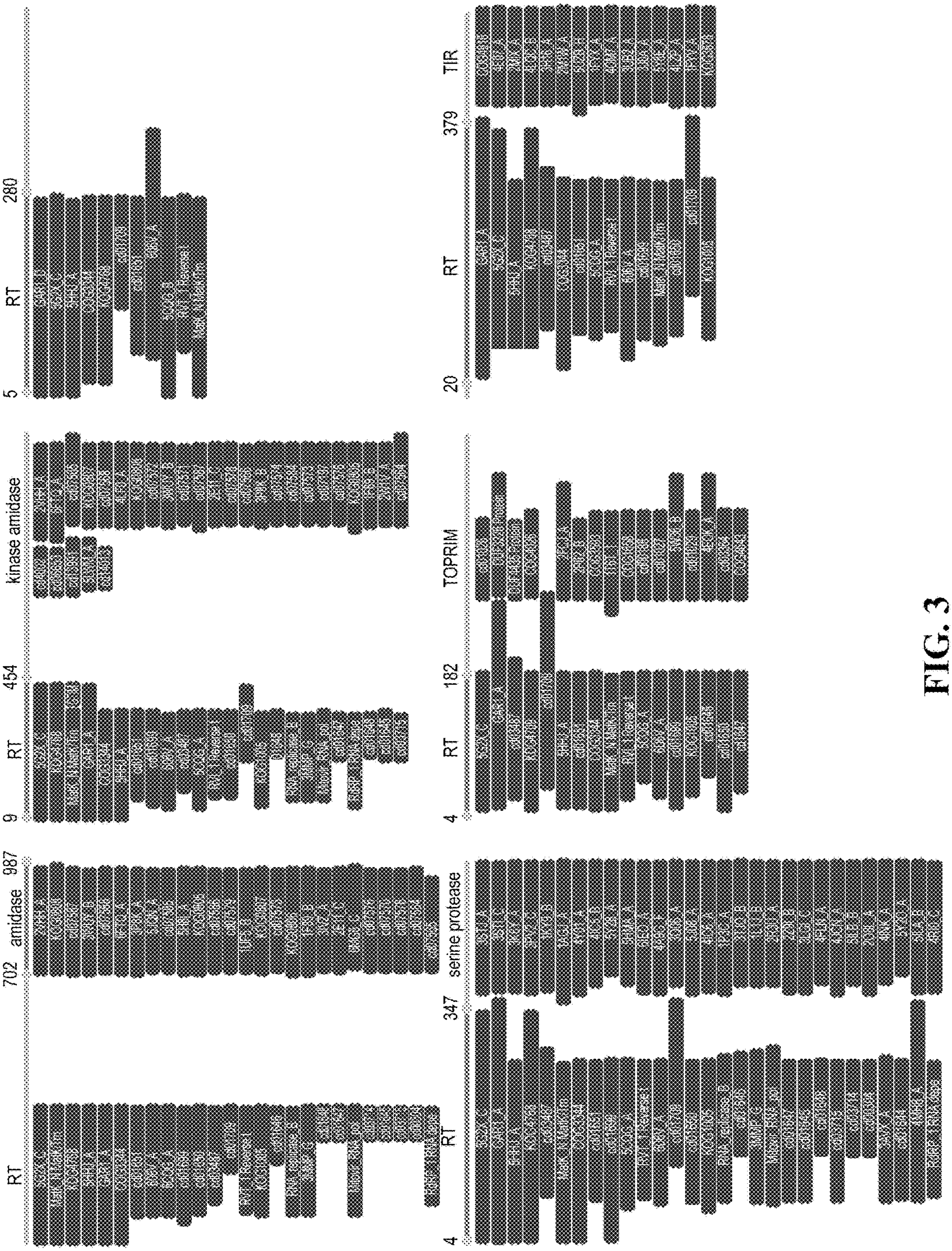

RT
702
amidase
987
9
RT
454
kinase amidase
5
RT
280
4
RT
347
serine protease
4
RT
182
TOPRIM
20
RT
379
TIR

FIG. 3

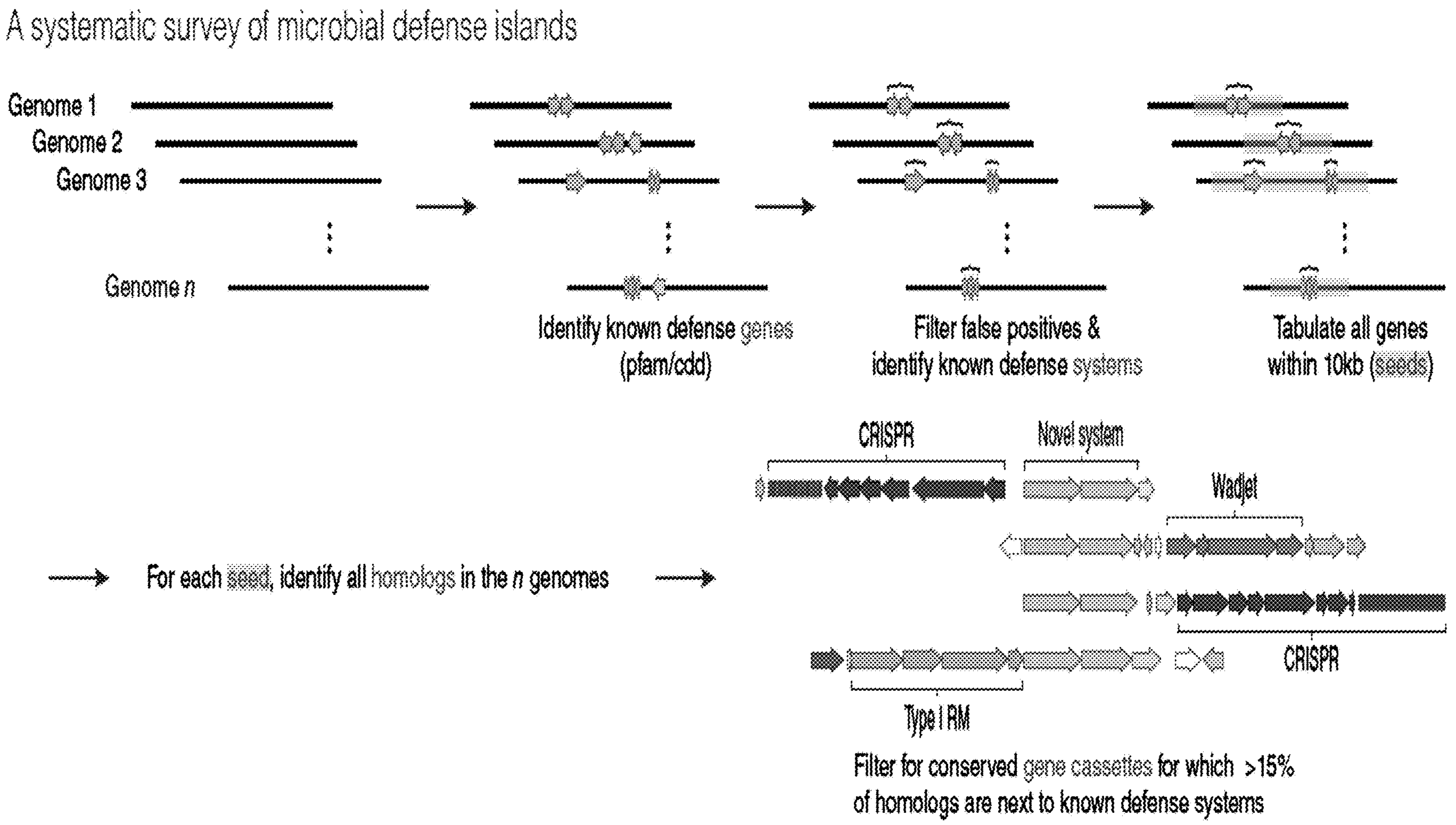

A systematic survey of microbial defense islands
Genome 1
Genome 2
Genome 3
Genome *n*
Identify known defense genes (pfam/cdd)
Filter false positives & identify known defense systems
Tabulate all genes within 10kb (seeds)
For each seed, identify all homologs in the *n* genomes
CRISPR
Novel system
Wadjet
CRISPR
Type I RM
Filter for conserved gene cassettes for which >15% of homologs are next to known defense systems

FIG. 4

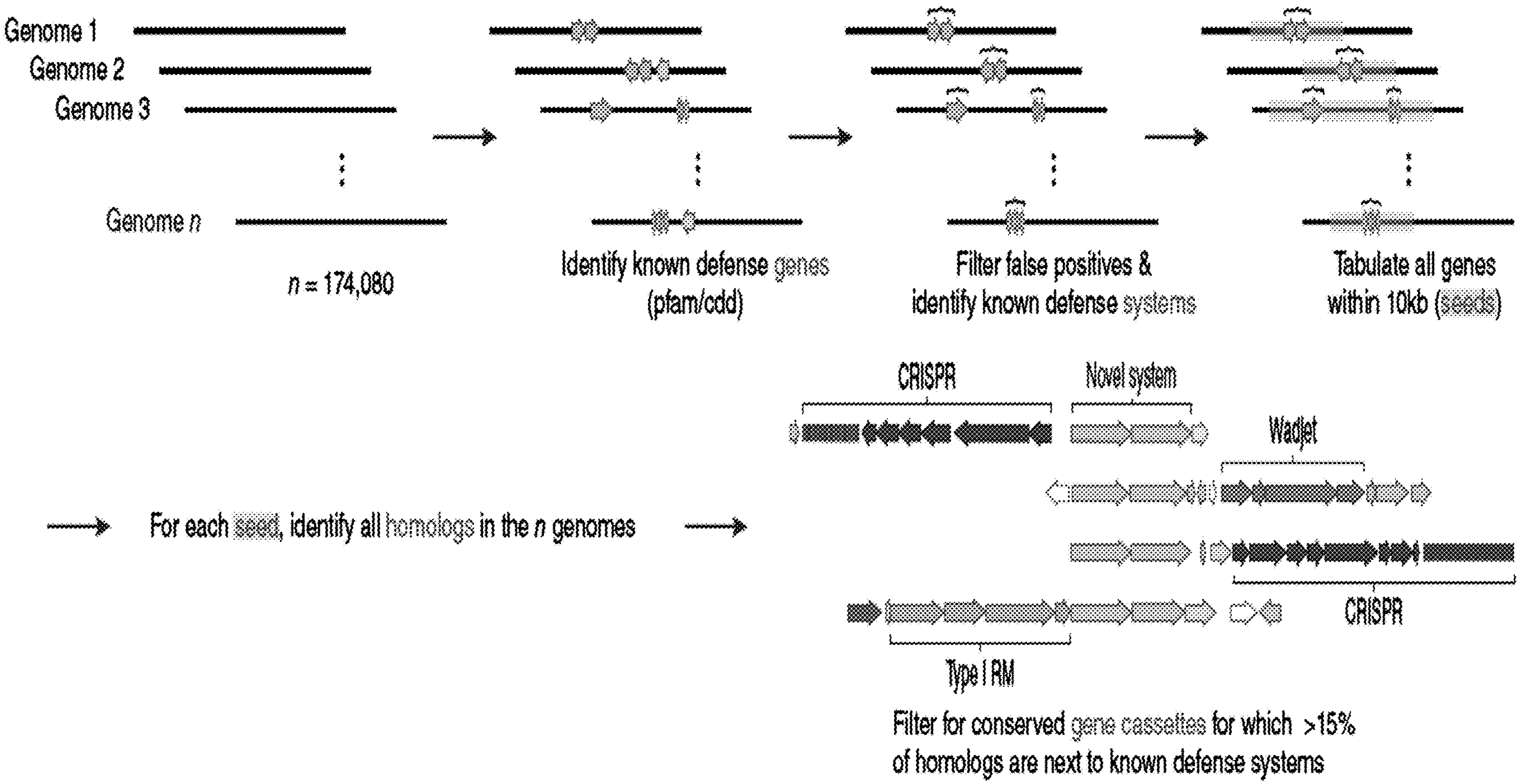

Genome 1
Genome 2
Genome 3
Genome *n*
n = 174,080
Identify known defense genes (pfam/cdd)
Filter false positives & identify known defense systems
Tabulate all genes within 10kb (seeds)
For each seed, identify all homologs in the *n* genomes
CRISPR
Novel system
Wadjet
CRISPR
Type I RM
Filter for conserved gene cassettes for which >15% of homologs are next to known defense systems

FIG. 5

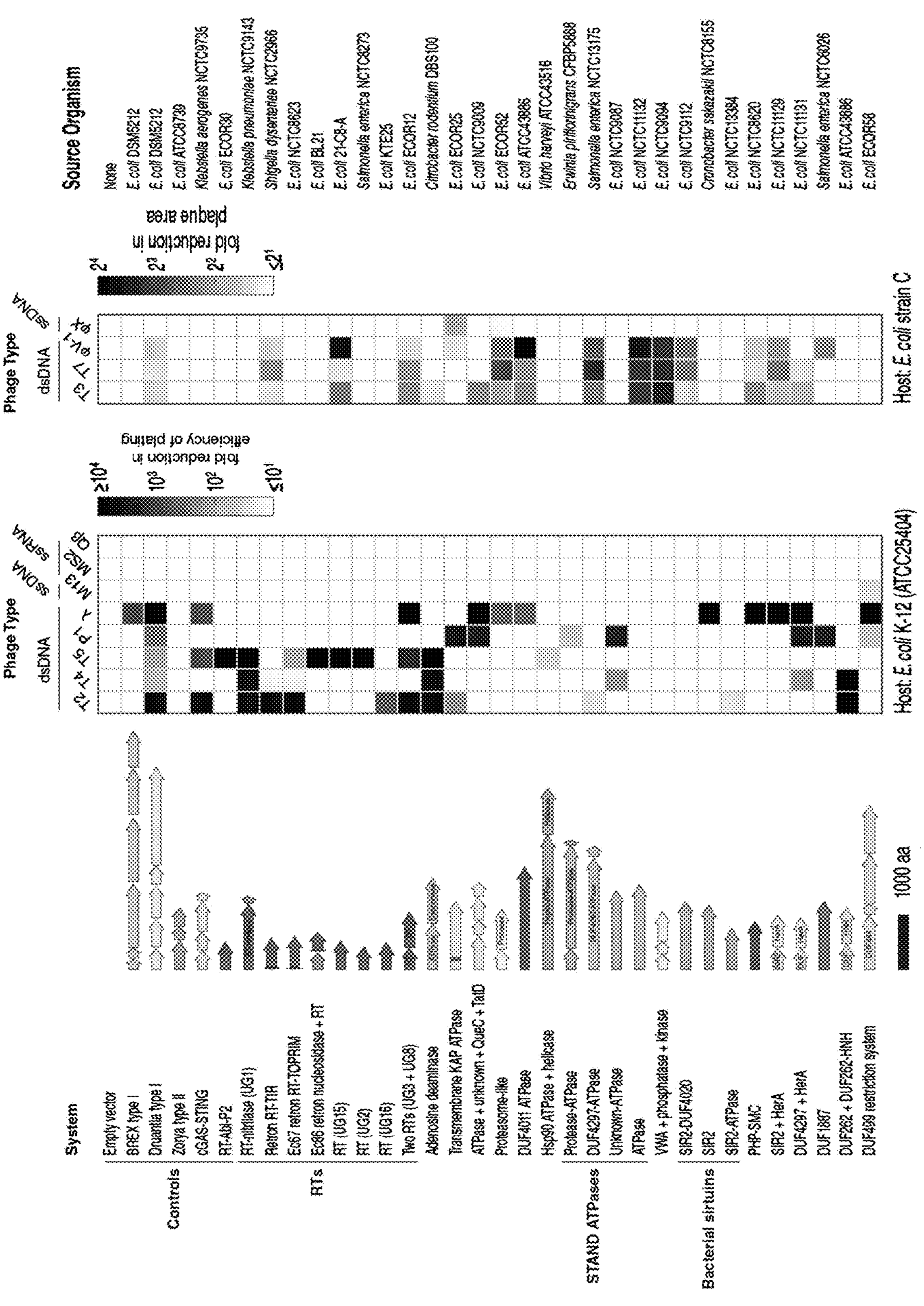

System
Controls
Empty vector
BREX type I
Druantia type I
Zorya type II
cGAS-STING
RT-Abi-P2
RTs
RT-nitrilase (UG1)
Retron RT-TIR
Ec67 retron RT-TOPRIM
Ec86 retron nucleosidase + RT
RT (UG15)
RT (UG2)
RT (UG16)
Two RTs (UG3 + UG8)
Adenosine deaminase
Transmembrane KAP ATPase
ATPase + unknown + QueC + TatD
Proteasome-like
DUF4011 ATPase
Hsp90 ATPase + helicase
STAND ATPases
Protease-ATPase
DUF4297-ATPase
Unknown-ATPase
ATPase
VWA + phosphatase + kinase
Bacterial sirtuins
SIR2-DUF4020
SIR2
SIR2-ATPase
PHP-SMC
SIR2 + HerA
DUF4297 + HerA
DUF1887
DUF262 + DUF262-HNH
DUF499 restriction system
1000 aa
Phage Type
dsDNA
T2
T4
T5
T6
T7
λ
ssDNA
M13
ssRNA
MS2
Qβ
≥10⁴
10³
10²
≤10¹
fold reduction in efficiency of plating
Host: E. coli K-12 (ATCC25404)
Phage Type
dsDNA
T3
T7
φV-1
ssDNA
φX
2⁴
2³
2²
≤2¹
fold reduction in plaque area
Host: E. coli strain C
Source Organism
None
E. coli DSM5212
E. coli DSM5212
E. coli ATCC8739
Klebsiella aerogenes NCTC9735
E. coli ECOR30
Klebsiella pneumoniae NCTC9143
Shigella dysenteriae NCTC2966
E. coli NCTC8623
E. coli BL21
E. coli 21-C8-A
Salmonella enterica NCTC8273
E. coli KTE25
E. coli ECOR12
Citrobacter rodentium DBS100
E. coli ECOR25
E. coli NCTC9009
E. coli ECOR62
E. coli ATCC43886
Vibrio harveyi ATCC43516
Erwinia piriflorinigrans CFBP5888
Salmonella enterica NCTC13175
E. coli NCTC9087
E. coli NCTC11132
E. coli NCTC9094
E. coli NCTC9112
Cronobacter sakazakii NCTC8155
E. coli NCTC13384
E. coli NCTC8620
E. coli NCTC11129
E. coli NCTC11131
Salmonella enterica NCTC8026
E. coli ATCC43886
E. coli ECOR58

FIG. 6A

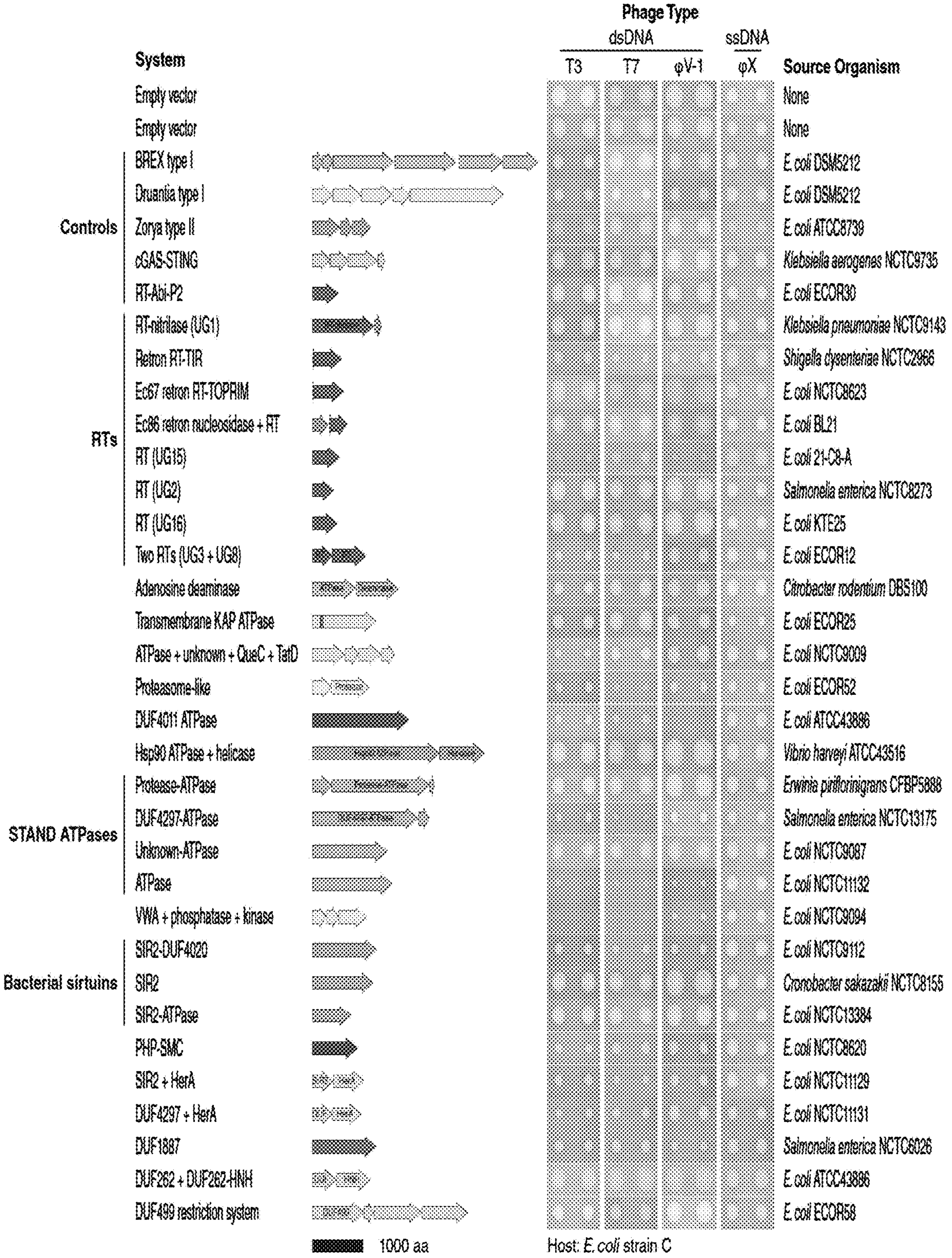

Phage Type
dsDNA
ssDNA
System
T3
T7
φV-1
φX
Source Organism
Empty vector — None
Empty vector — None
Controls
BREX type I — E. coli DSM5212
Druantia type I — E. coli DSM5212
Zorya type II — E. coli ATCC8739
cGAS-STING — Klebsiella aerogenes NCTC9735
RT-Abi-P2 — E. coli ECOR30
RTs
RT-nitrilase (UG1) — Klebsiella pneumoniae NCTC9143
Retron RT-TIR — Shigella dysenteriae NCTC2966
Ec67 retron RT-TOPRIM — E. coli NCTC8623
Ec86 retron nucleosidase + RT — E. coli BL21
RT (UG15) — E. coli 21-C8-A
RT (UG2) — Salmonella enterica NCTC8273
RT (UG16) — E. coli KTE25
Two RTs (UG3 + UG8) — E. coli ECOR12
Adenosine deaminase — Citrobacter rodentium DBS100
Transmembrane KAP ATPase — E. coli ECOR25
ATPase + unknown + QueC + TatD — E. coli NCTC9009
Proteasome-like — E. coli ECOR52
DUF4011 ATPase — E. coli ATCC43886
Hsp90 ATPase + helicase — Vibrio harveyi ATCC43516
STAND ATPases
Protease-ATPase — Erwinia piriflorinigrans CFBP5888
DUF4297-ATPase — Salmonella enterica NCTC13175
Unknown-ATPase — E. coli NCTC9087
ATPase — E. coli NCTC11132
VWA + phosphatase + kinase — E. coli NCTC9094
Bacterial sirtuins
SIR2-DUF4020 — E. coli NCTC9112
SIR2 — Cronobacter sakazakii NCTC8155
SIR2-ATPase — E. coli NCTC13384
PHP-SMC — E. coli NCTC8620
SIR2 + HerA — E. coli NCTC11129
DUF4297 + HerA — E. coli NCTC11131
DUF1887 — Salmonella enterica NCTC6026
DUF262 + DUF262-HNH — E. coli ATCC43886
DUF499 restriction system — E. coli ECOR58
1000 aa
Host: E. coli strain C

FIG. 6B

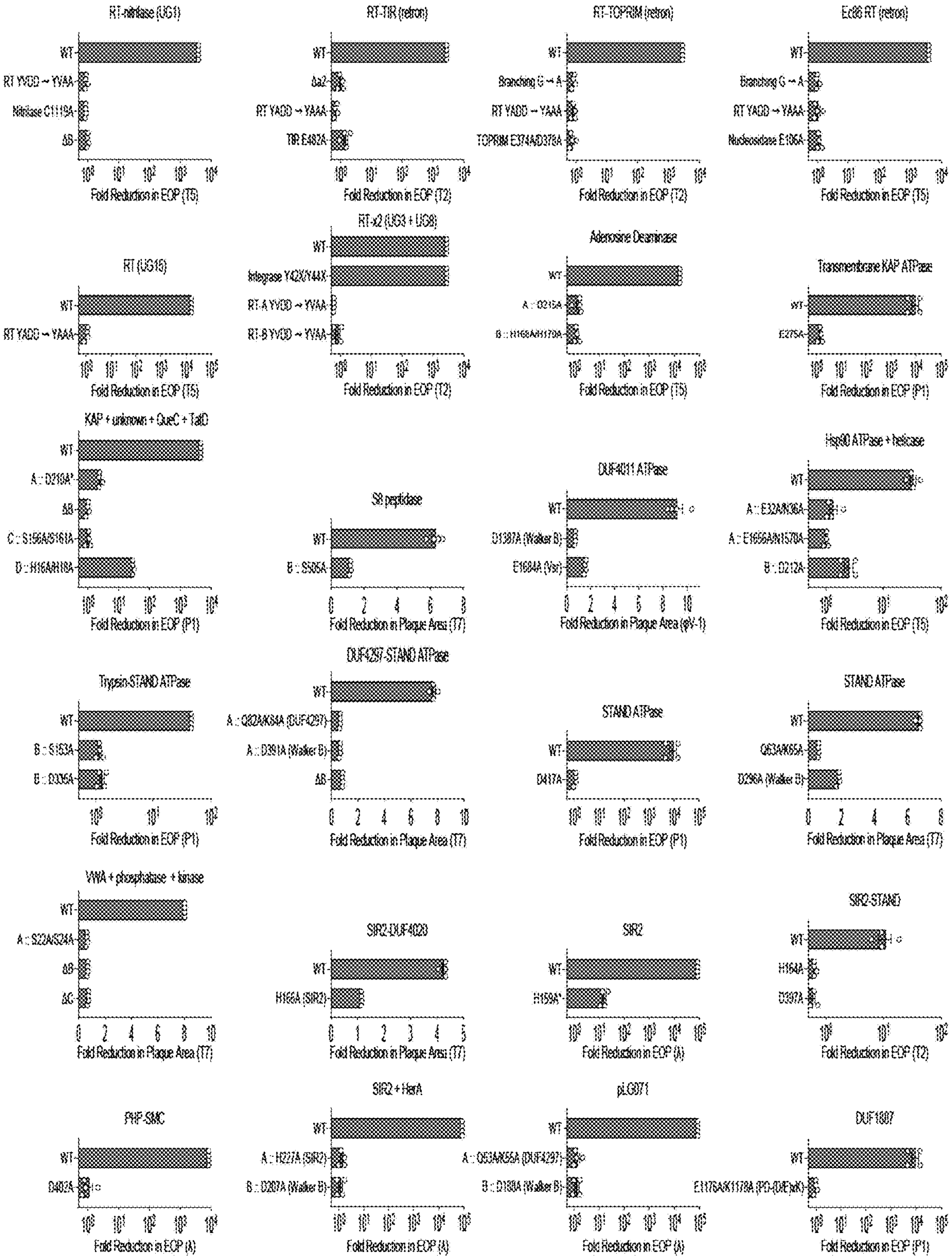

RT-nitrilase (UG1)
WT
RT YVDD → YVAA
Nitrilase C1119A
ΔB
10^0 10^1 10^2 10^3 10^4
Fold Reduction in EOP (T5)
RT-TIR (retron)
WT
Δa2
RT YADD → YAAA
TIR E482A
10^0 10^1 10^2 10^3 10^4
Fold Reduction in EOP (T2)
RT-TOPRIM (retron)
WT
Branching G → A
RT YADD → YAAA
TOPRIM E374A/D378A
10^0 10^1 10^2 10^3 10^4
Fold Reduction in EOP (T2)
Ec86 RT (retron)
WT
Branching G → A
RT YADD → YAAA
Nucleosidase E106A
10^0 10^1 10^2 10^3 10^4
Fold Reduction in EOP (T5)
RT (UG15)
WT
RT YADD → YAAA
10^0 10^1 10^2 10^3 10^4 10^5
Fold Reduction in EOP (T5)
RT-x2 (UG3 + UG8)
WT
Integrase Y42X/Y44X
RT-A YVDD → YVAA
RT-B YVDD → YVAA
10^0 10^1 10^2 10^3 10^4
Fold Reduction in EOP (T2)
Adenosine Deaminase
WT
A :: D215A
B :: H168A/H170A
10^0 10^1 10^2 10^3 10^4 10^5
Fold Reduction in EOP (T5)
Transmembrane KAP ATPase
WT
E275A
10^0 10^1 10^2 10^3 10^4 10^5
Fold Reduction in EOP (P1)
KAP + unkown + QueC + TatD
WT
A :: D210A*
ΔB
C :: S156A/S161A
D :: H16A/H18A
10^0 10^1 10^2 10^3 10^4
Fold Reduction in EOP (P1)
S8 peptidase
WT
B :: S505A
0 2 4 6 8
Fold Reduction in Plaque Area (T7)
DUF4011 ATPase
WT
D1387A (Walker B)
E1684A (Vsr)
0 2 4 6 8 10
Fold Reduction in Plaque Area (φV-1)
Hsp90 ATPase + helicase
WT
A :: E32A/N36A
A :: E1656A/N1570A
B :: D212A
10^0 10^1 10^2
Fold Reduction in EOP (T5)
Trypsin-STAND ATPase
WT
B :: S153A
B :: D306A
10^0 10^1 10^2
Fold Reduction in EOP (P1)
DUF4297-STAND ATPase
WT
A :: Q82A/K84A (DUF4297)
A :: D391A (Walker B)
ΔB
0 2 4 6 8 10
Fold Reduction in Plaque Area (T7)
STAND ATPase
WT
D417A
10^0 10^1 10^2 10^3 10^4 10^5
Fold Reduction in EOP (P1)
STAND ATPase
WT
Q63A/K65A
D296A (Walker B)
0 2 4 6 8
Fold Reduction in Plaque Area (T7)
VWA + phosphatase + kinase
WT
A :: S22A/S24A
ΔB
ΔC
0 2 4 6 8 10
Fold Reduction in Plaque Area (T7)
SIR2-DUF4020
WT
H166A (SIR2)
0 1 2 3 4 5
Fold Reduction in Plaque Area (T7)
SIR2
WT
H159A*
10^0 10^1 10^2 10^3 10^4 10^5
Fold Reduction in EOP (λ)
SIR2-STAND
WT
H164A
D397A
10^0 10^1 10^2
Fold Reduction in EOP (T2)
PHP-SMC
WT
D402A
10^0 10^1 10^2 10^3 10^4
Fold Reduction in EOP (λ)
SIR2 + HerA
WT
A :: H227A (SIR2)
B :: D207A (Walker B)
10^0 10^1 10^2 10^3 10^4 10^5
Fold Reduction in EOP (λ)
pLG071
WT
A :: Q53A/K55A (DUF4297)
B :: D188A (Walker B)
10^0 10^1 10^2 10^3 10^4 10^5
Fold Reduction in EOP (λ)
DUF1887
WT
E1176A/K1178A (PD-(D/E)xK)
10^0 10^1 10^2 10^3 10^4 10^5
Fold Reduction in EOP (P1)

FIG. 7

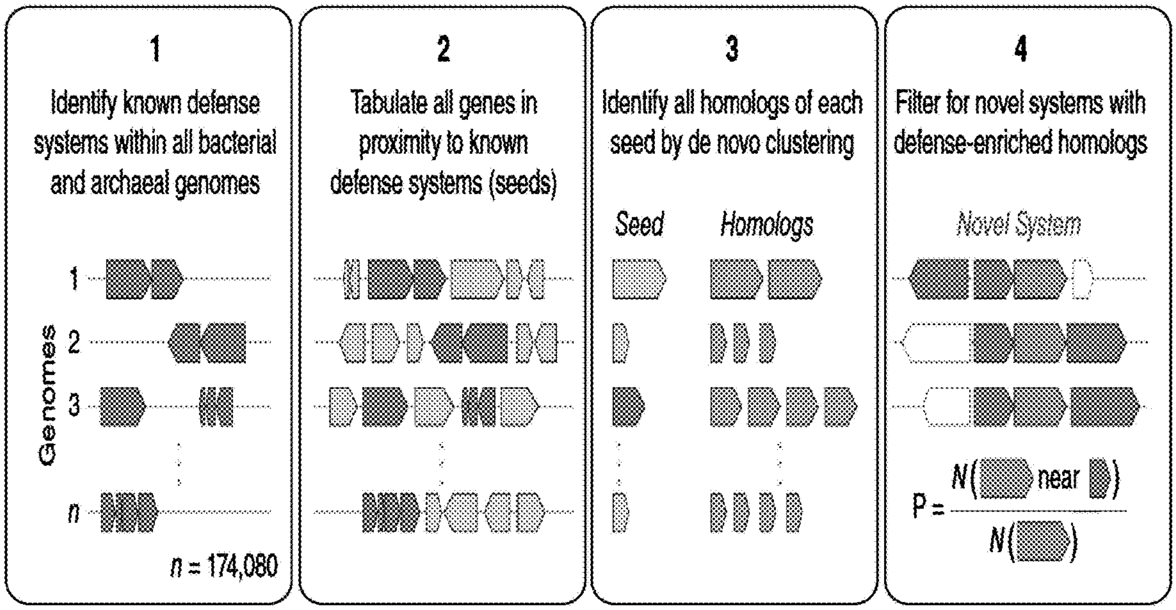
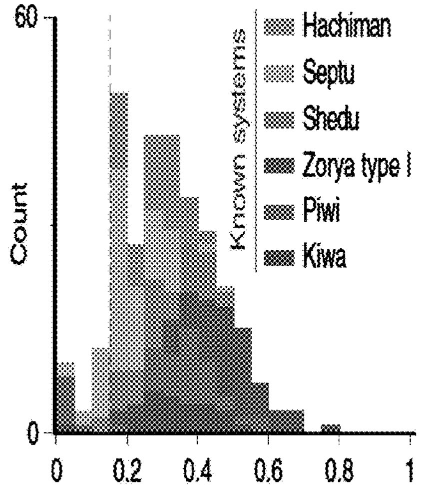


FIG. 8A



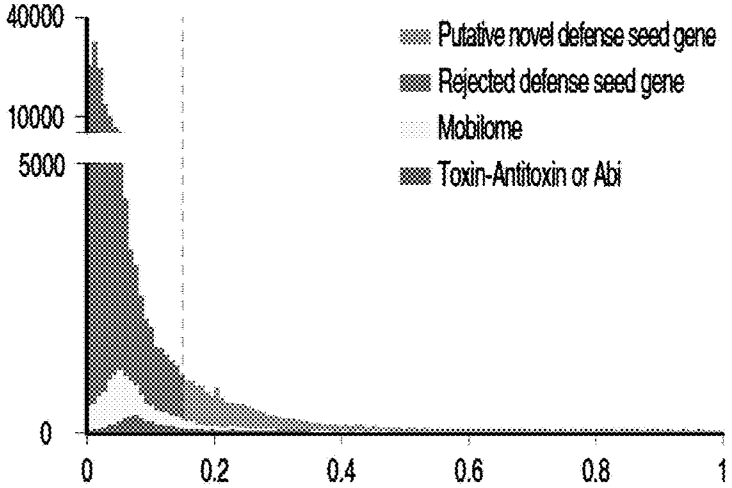


FIG. 8B

FIG. 8C

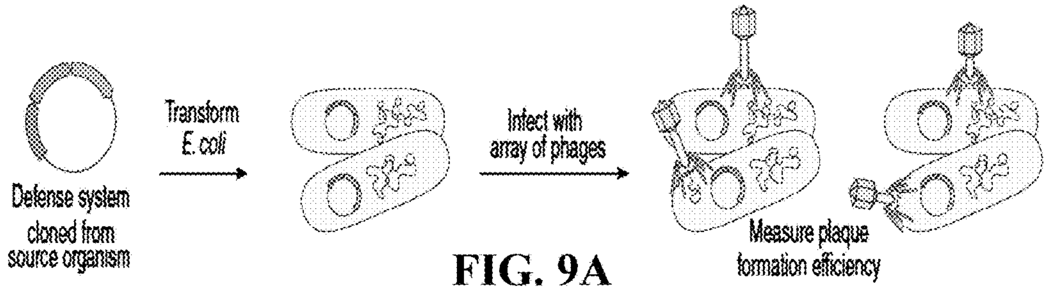

Defense system
cloned from
source organism
Transform
E. coli
Infect with
array of phages
Measure plaque
formation efficiency

FIG. 9A

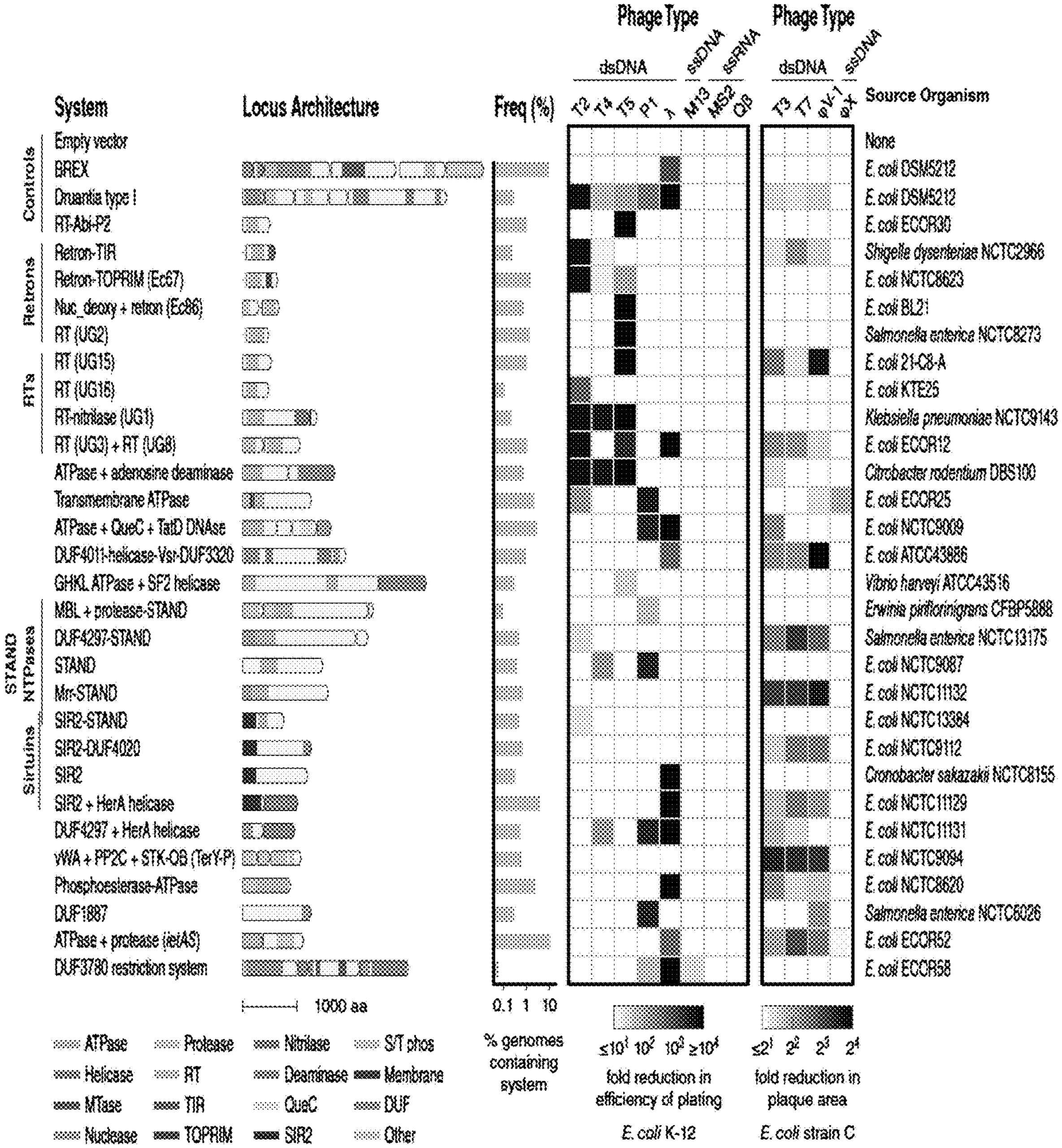

Phage Type
dsDNA
ssDNA
ssRNA
Phage Type
dsDNA
ssDNA
System
Locus Architecture
Freq (%)
T2
T4
T5
P1
λ
M13
MS2
Qβ
T3
T7
φV-1
φX
Source Organism
Controls
Retrons
RTs
STAND NTPases
Sirtuins
Empty vector
BREX
Druantia type I
RT-Abi-P2
Retron-TIR
Retron-TOPRIM (Ec67)
Nuc_deoxy + retron (Ec86)
RT (UG2)
RT (UG15)
RT (UG16)
RT-nitrilase (UG1)
RT (UG3) + RT (UG8)
ATPase + adenosine deaminase
Transmembrane ATPase
ATPase + QueC + TatD DNAse
DUF4011-helicase-Vsr-DUF3320
GHKL ATPase + SF2 helicase
MBL + protease-STAND
DUF4297-STAND
STAND
Mrr-STAND
SIR2-STAND
SIR2-DUF4020
SIR2
SIR2 + HerA helicase
DUF4297 + HerA helicase
vWA + PP2C + STK-OB (TerY-P)
Phosphoesterase-ATPase
DUF1887
ATPase + protease (ietAS)
DUF3780 restriction system
None
E. coli DSM5212
E. coli DSM5212
E. coli ECOR30
Shigella dysenteriae NCTC2966
E. coli NCTC8623
E. coli BL21
Salmonella enterica NCTC8273
E. coli 21-C8-A
E. coli KTE25
Klebsiella pneumoniae NCTC9143
E. coli ECOR12
Citrobacter rodentium DBS100
E. coli ECOR25
E. coli NCTC9009
E. coli ATCC43886
Vibrio harveyi ATCC43516
Erwinia piriflorinigrans CFBP5888
Salmonella enterica NCTC13175
E. coli NCTC9087
E. coli NCTC11132
E. coli NCTC13384
E. coli NCTC9112
Cronobacter sakazakii NCTC8155
E. coli NCTC11129
E. coli NCTC11131
E. coli NCTC9094
E. coli NCTC8620
Salmonella enterica NCTC6026
E. coli ECOR52
E. coli ECOR58
1000 aa
0.1 1 10
% genomes containing system
≤10¹ 10² 10³ ≥10⁴
fold reduction in efficiency of plating
E. coli K-12
≤2¹ 2² 2³ 2⁴
fold reduction in plaque area
E. coli strain C
ATPase
Helicase
MTase
Nuclease
Protease
RT
TIR
TOPRIM
Nitrilase
Deaminase
QueC
SIR2
S/T phos
Membrane
DUF
Other

FIG. 9B

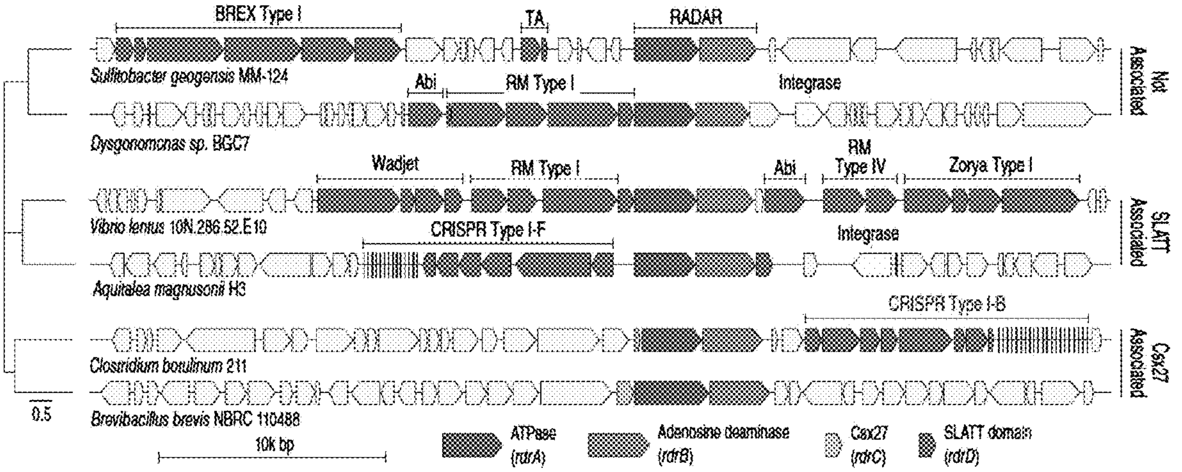


FIG. 10A

SEQ ID NOS: 473-476

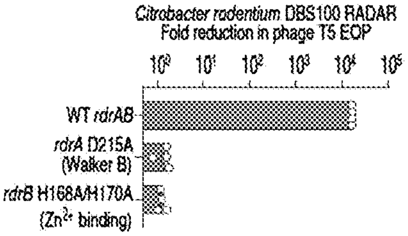


FIG. 10B

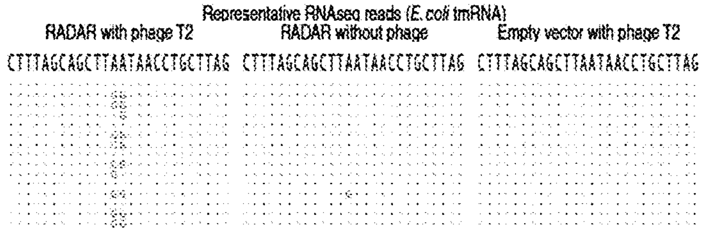


FIG. 10C

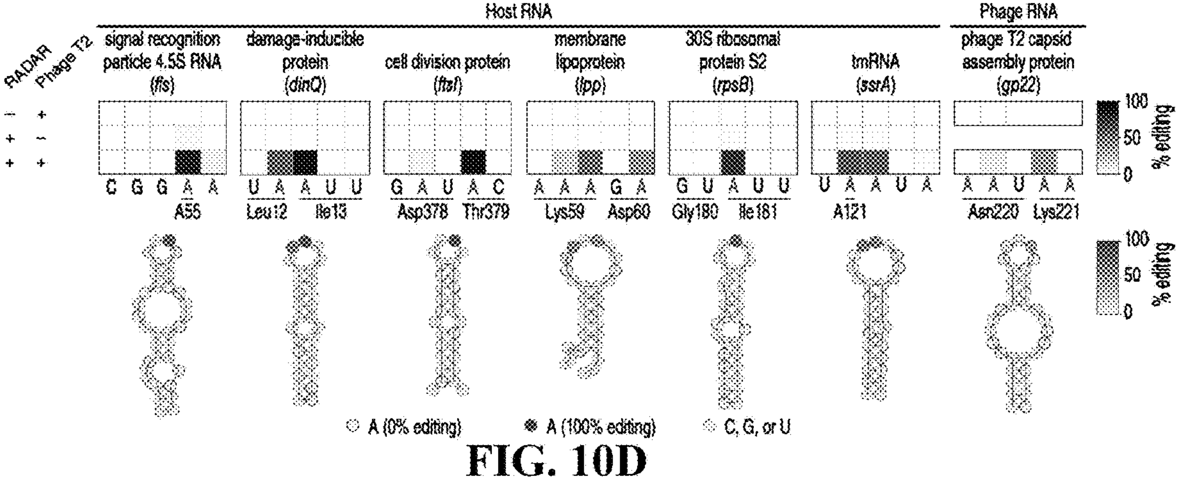


FIG. 10D

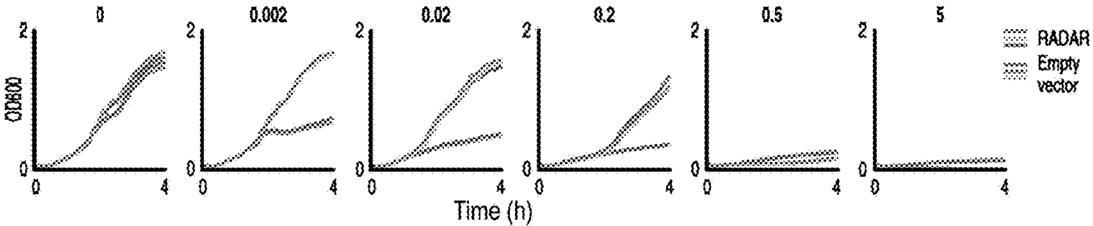


FIG. 10E

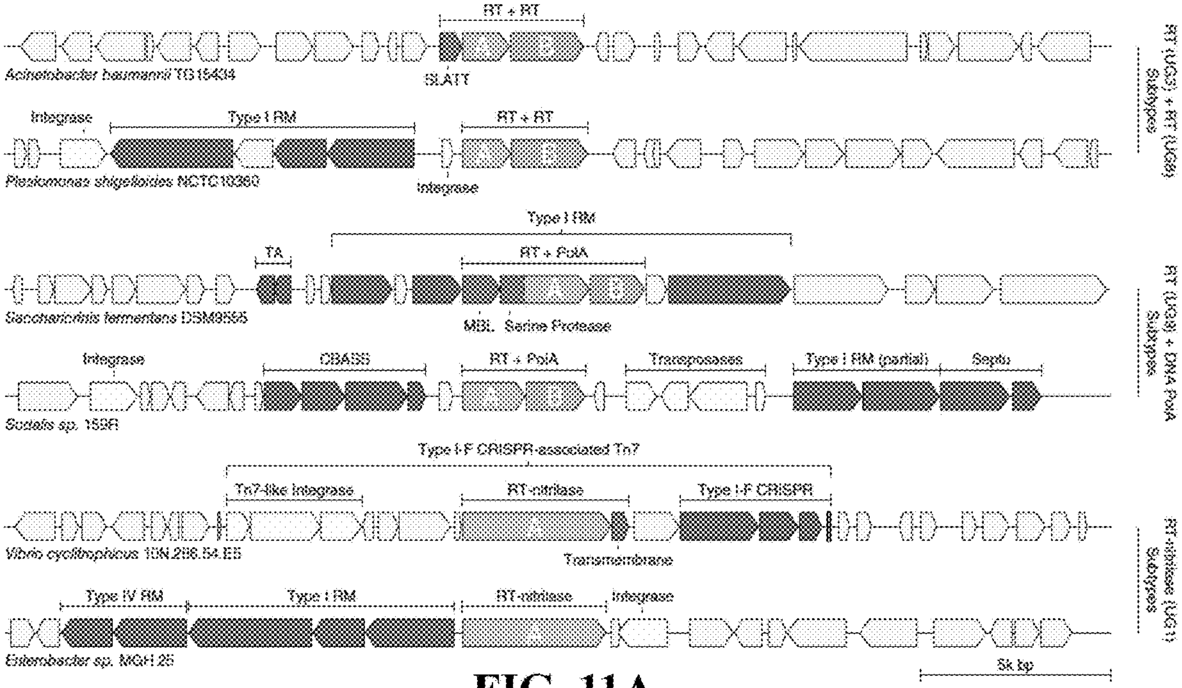


FIG. 11A

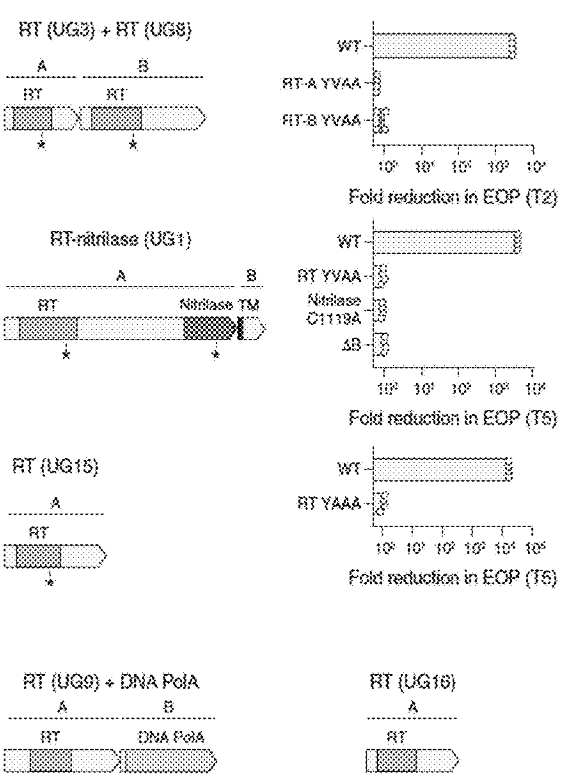


FIG. 11B

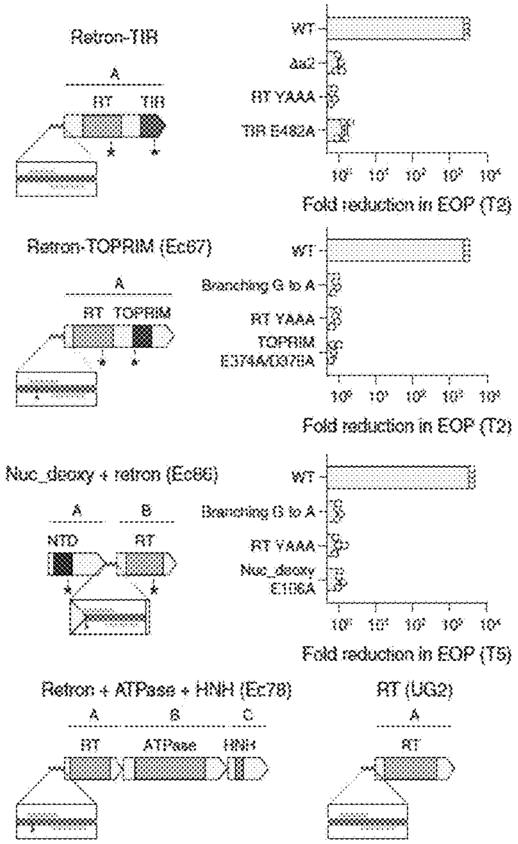


FIG. 11C

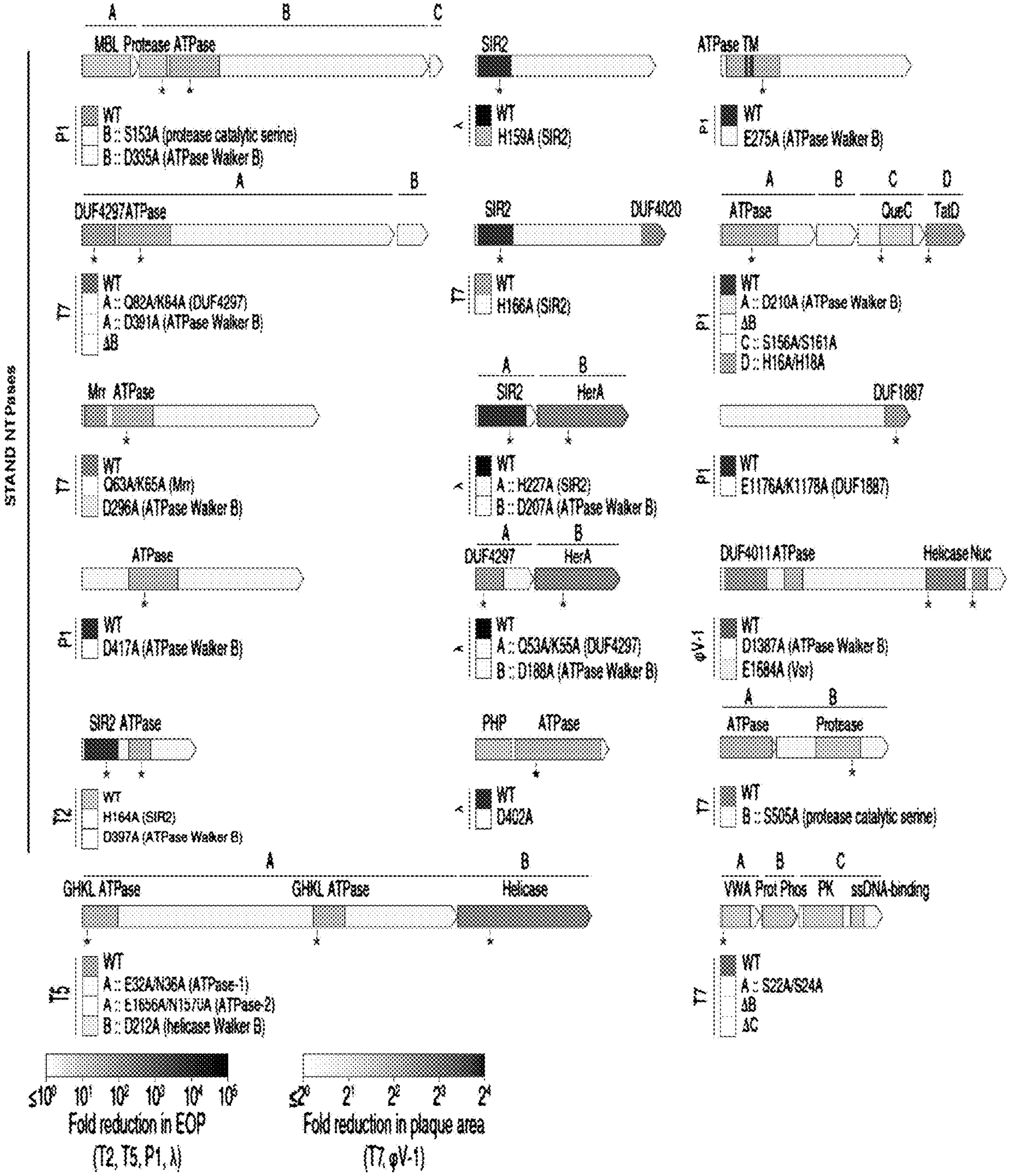

STAND NTPases
A
B
C
MBL Protease ATPase
P1
WT
B :: S153A (protease catalytic serine)
B :: D335A (ATPase Walker B)
SIR2
λ
WT
H159A (SIR2)
ATPase TM
P1
WT
E275A (ATPase Walker B)
A
B
DUF4297ATPase
T7
WT
A :: Q82A/K84A (DUF4297)
A :: D391A (ATPase Walker B)
ΔB
SIR2
DUF4020
T7
WT
H166A (SIR2)
A
B
C
D
ATPase
QueC
TatD
P1
WT
A :: D210A (ATPase Walker B)
ΔB
C :: S156A/S161A
D :: H16A/H18A
Mrr ATPase
T7
WT
Q63A/K65A (Mrr)
D296A (ATPase Walker B)
A
B
SIR2
HerA
λ
WT
A :: H227A (SIR2)
B :: D207A (ATPase Walker B)
DUF1887
P1
WT
E1176A/K1178A (DUF1887)
ATPase
P1
WT
D417A (ATPase Walker B)
A
B
DUF4297
HerA
λ
WT
A :: Q53A/K55A (DUF4297)
B :: D188A (ATPase Walker B)
DUF4011 ATPase
Helicase Nuc
φV-1
WT
D1387A (ATPase Walker B)
E1584A (Vsr)
SIR2 ATPase
T2
WT
H164A (SIR2)
D397A (ATPase Walker B)
PHP
ATPase
λ
WT
D402A
A
B
ATPase
Protease
T7
WT
B :: S505A (protease catalytic serine)
A
B
GHKL ATPase
GHKL ATPase
Helicase
T5
WT
A :: E32A/N36A (ATPase-1)
A :: E1656A/N1570A (ATPase-2)
B :: D212A (helicase Walker B)
A
B
C
VWA Prot Phos PK ssDNA-binding
T7
WT
A :: S22A/S24A
ΔB
ΔC
$\leq 10^0$ 10^1 10^2 10^3 10^4 10^5
Fold reduction in EOP
(T2, T5, P1, λ)
$\leq 2^0$ 2^1 2^2 2^3 2^4
Fold reduction in plaque area
(T7, φV-1)

FIG. 12

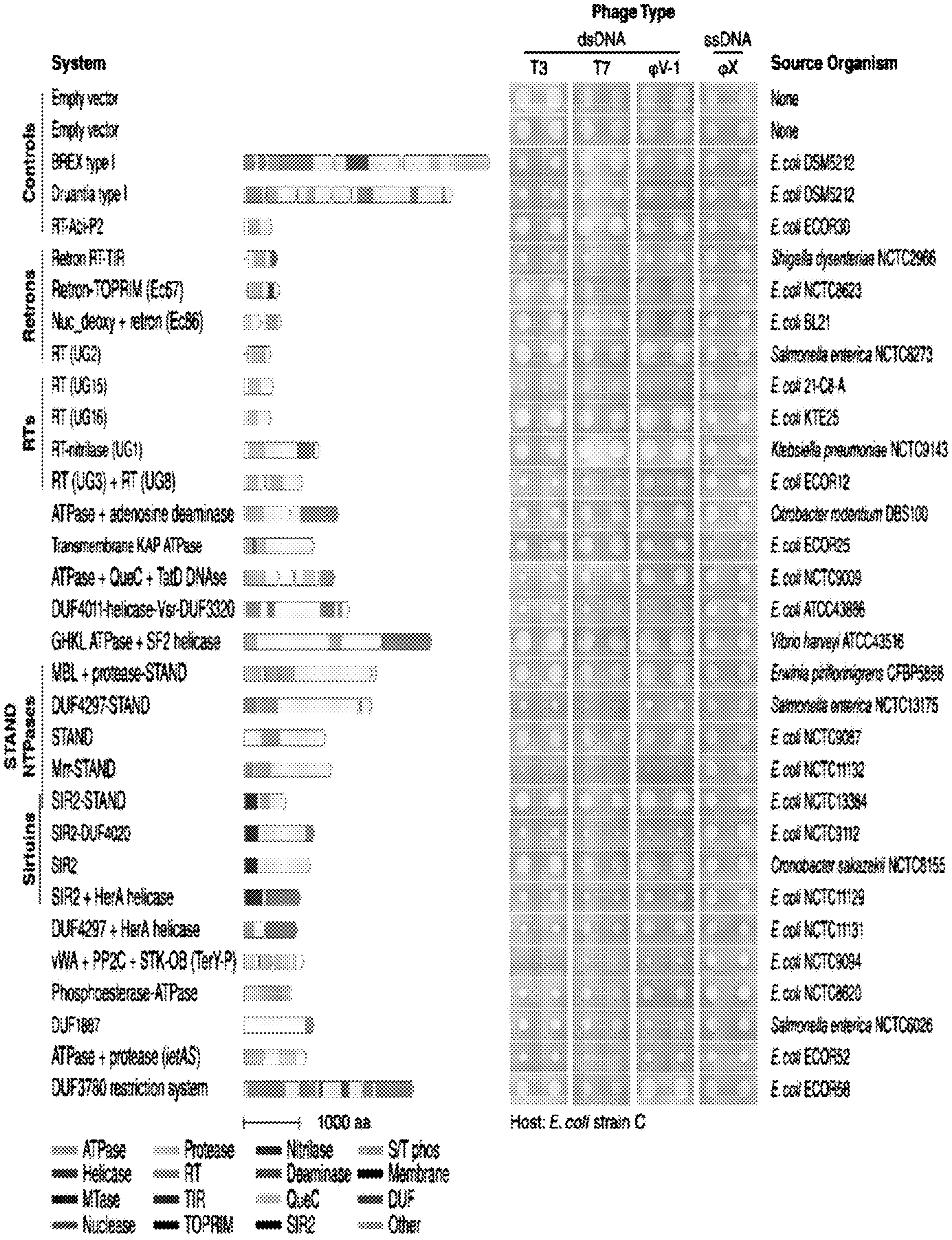

Phage Type
dsDNA
ssDNA
System
T3
T7
φV-1
φX
Source Organism
Controls
Empty vector
None
Empty vector
None
BREX type I
E. coli DSM5212
Druantia type I
E. coli DSM5212
RT-Abi-P2
E. coli ECOR30
Retrons
Retron RT-TIR
Shigella dysenteriae NCTC2966
Retron-TOPRIM (Ec67)
E. coli NCTC8623
Nuc_deoxy + retron (Ec86)
E. coli BL21
RT (UG2)
Salmonella enterica NCTC8273
RTs
RT (UG15)
E. coli 21-C8-A
RT (UG16)
E. coli KTE25
RT-nitrilase (UG1)
Klebsiella pneumoniae NCTC9143
RT (UG3) + RT (UG8)
E. coli ECOR12
ATPase + adenosine deaminase
Citrobacter rodentium DBS100
Transmembrane KAP ATPase
E. coli ECOR25
ATPase + QueC + TatD DNAse
E. coli NCTC9009
DUF4011-helicase-Vsr-DUF3320
E. coli ATCC43886
GHKL ATPase + SF2 helicase
Vibrio harveyi ATCC43516
STAND NTPases
MBL + protease-STAND
Erwinia piriflorinigrans CFBP5888
DUF4297-STAND
Salmonella enterica NCTC13175
STAND
E. coli NCTC9087
Mrr-STAND
E. coli NCTC11132
Sirtuins
SIR2-STAND
E. coli NCTC13384
SIR2-DUF4020
E. coli NCTC9112
SIR2
Cronobacter sakazakii NCTC8155
SIR2 + HerA helicase
E. coli NCTC11129
DUF4297 + HerA helicase
E. coli NCTC11131
vWA + PP2C + STK-OB (TerY-P)
E. coli NCTC9094
Phosphoesterase-ATPase
E. coli NCTC8620
DUF1887
Salmonella enterica NCTC6026
ATPase + protease (ietAS)
E. coli ECOR52
DUF3780 restriction system
E. coli ECOR58
1000 aa
Host: E. coli strain C
ATPase
Protease
Nitrilase
S/T phos
Helicase
RT
Deaminase
Membrane
MTase
TIR
QueC
DUF
Nuclease
TOPRIM
SIR2
Other

FIG. 13

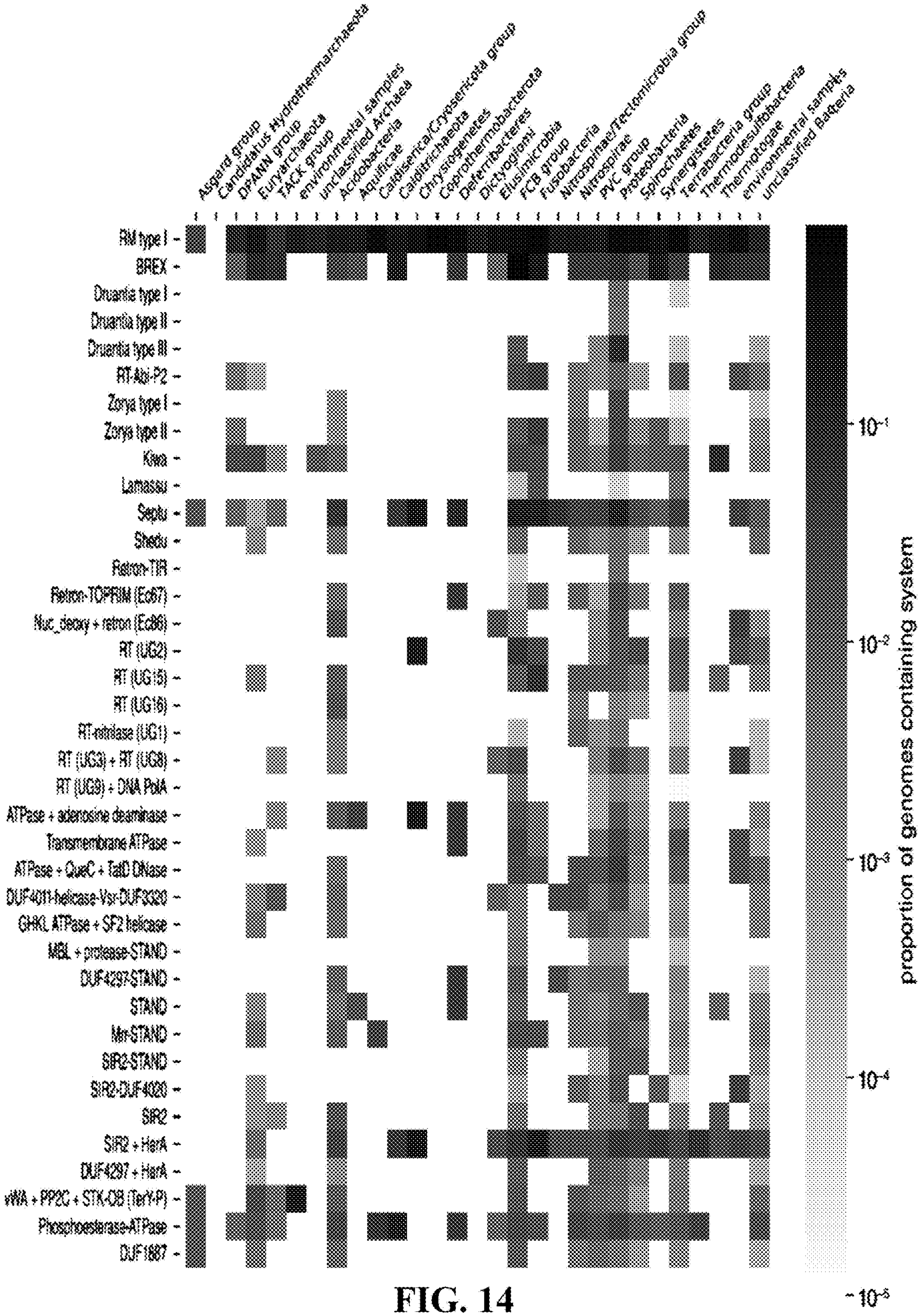

Asgard group
Candidatus Hydrothermarchaeota
DPANN group
Euryarchaeota
TACK group
environmental samples
unclassified Archaea
Acidobacteria
Aquificae
Caldiserica/Cryosericota group
Calditrichaeota
Chrysiogenetes
Coprothermobacterota
Deferribacteres
Dictyoglomi
Elusimicrobia
FCB group
Fusobacteria
Nitrospinae/Tectomicrobia group
Nitrospirae
PVC group
Proteobacteria
Spirochaetes
Synergistetes
Terrabacteria group
Thermodesulfobacteria
Thermotogae
environmental samples
unclassified Bacteria
RM type I
BREX
Druantia type I
Druantia type II
Druantia type III
RT-Abi-P2
Zorya type I
Zorya type II
Kiwa
Lamassu
Septu
Shedu
Retron-TIR
Retron-TOPRIM (Ec67)
Nuc_deoxy + retron (Ec86)
RT (UG2)
RT (UG15)
RT (UG16)
RT-nitrilase (UG1)
RT (UG3) + RT (UG8)
RT (UG9) + DNA PolA
ATPase + adenosine deaminase
Transmembrane ATPase
ATPase + QueC + TatD DNase
DUF4011-helicase-Vsr-DUF3320
GHKL ATPase + SF2 helicase
MBL + protease-STAND
DUF4297-STAND
STAND
Mrr-STAND
SIR2-STAND
SIR2-DUF4020
SIR2
SIR2 + HerA
DUF4297 + HerA
vWA + PP2C + STK-OB (TerY-P)
Phosphoesterase-ATPase
DUF1887
10^{-1}
10^{-2}
10^{-3}
10^{-4}
$\sim 10^{-5}$
proportion of genomes containing system

FIG. 14

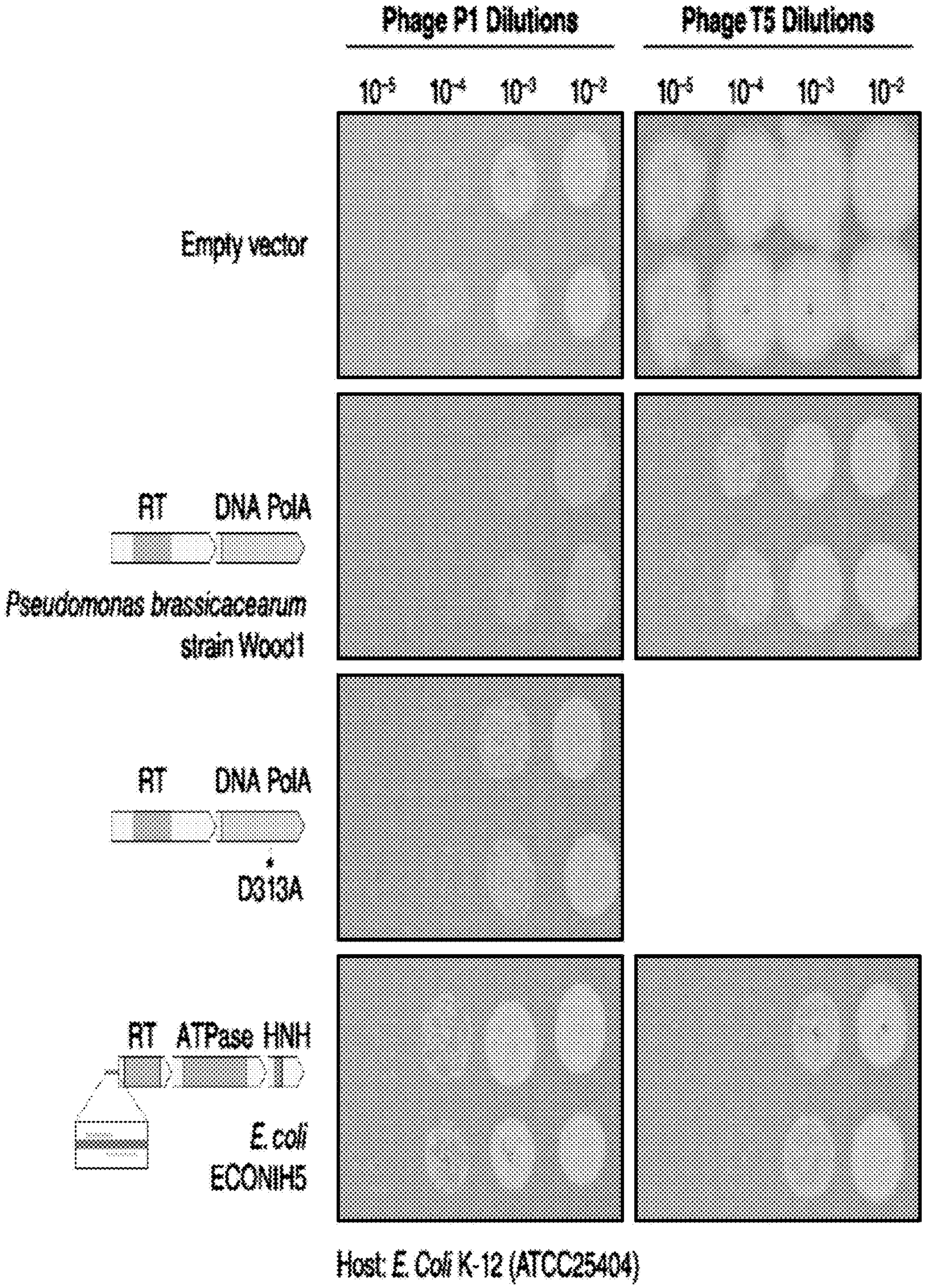

Phage P1 Dilutions
Phage T5 Dilutions
10-5
10-4
10-3
10-2
10-5
10-4
10-3
10-2
Empty vector
RT
DNA PolA
Pseudomonas brassicacearum
strain Wood1
RT
DNA PolA
D313A
RT
ATPase
HNH
E. coli
ECONIH5
Host: E. Coli K-12 (ATCC25404)

FIG. 15

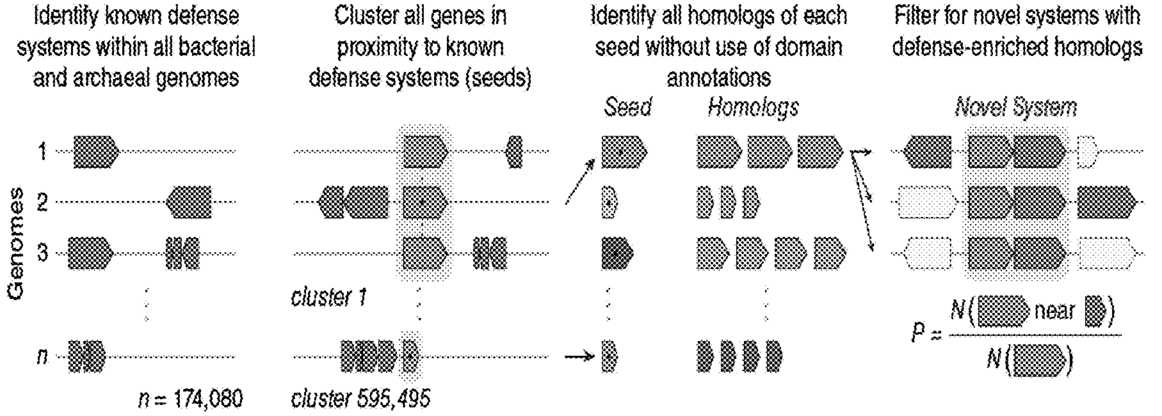


FIG. 16A

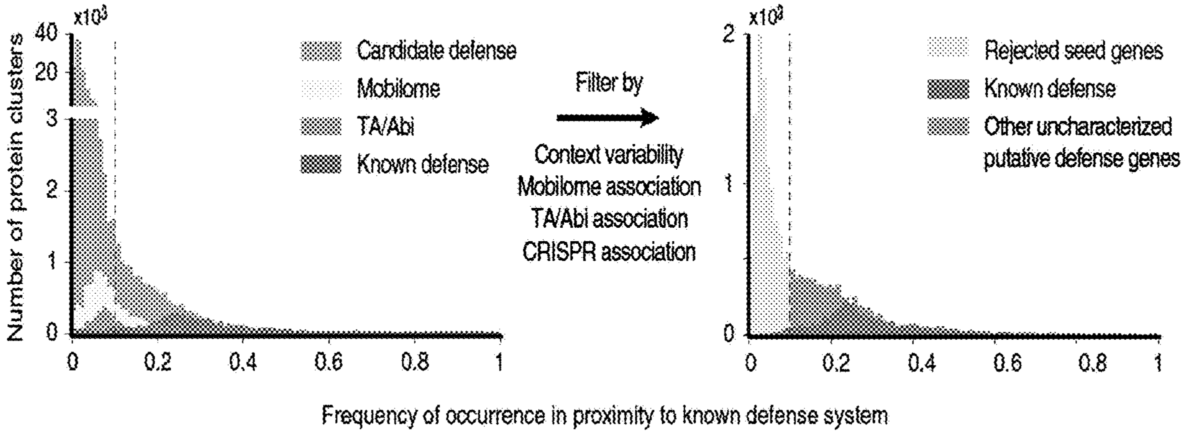


FIG. 16B

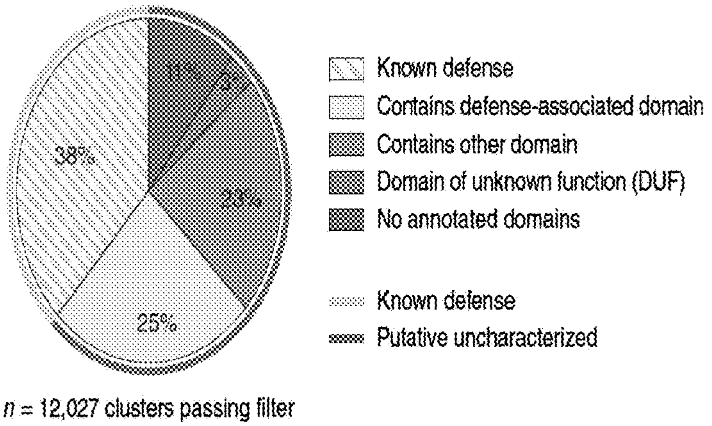


FIG. 16C

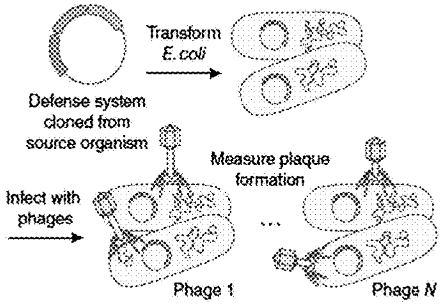


FIG. 17A

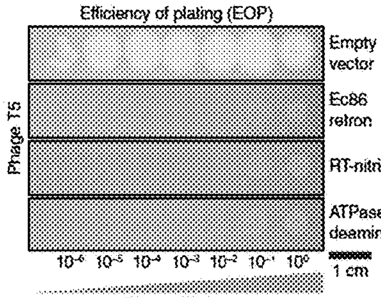


FIG. 17B

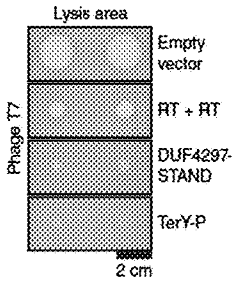


FIG. 17C

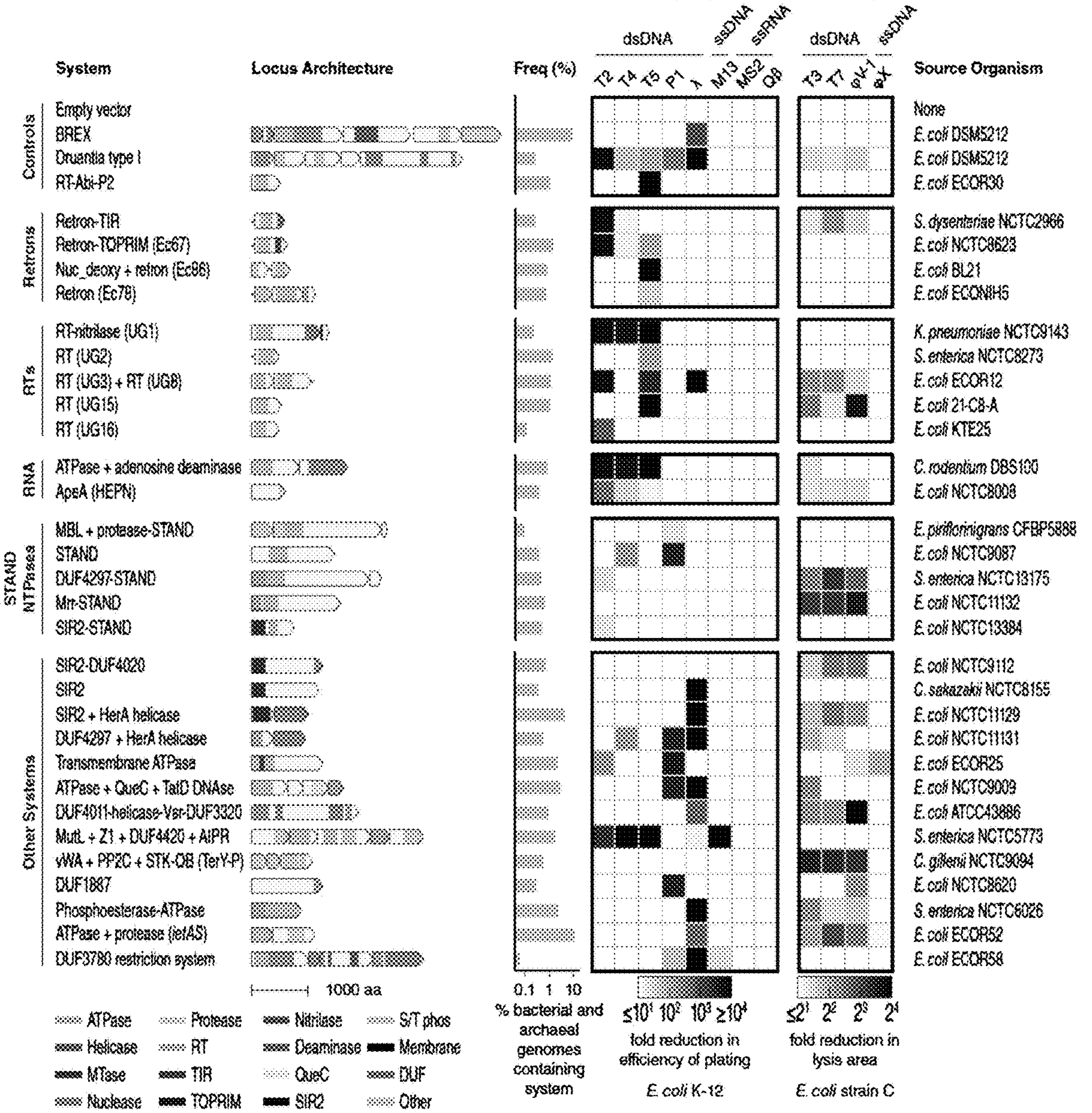


FIG. 17D

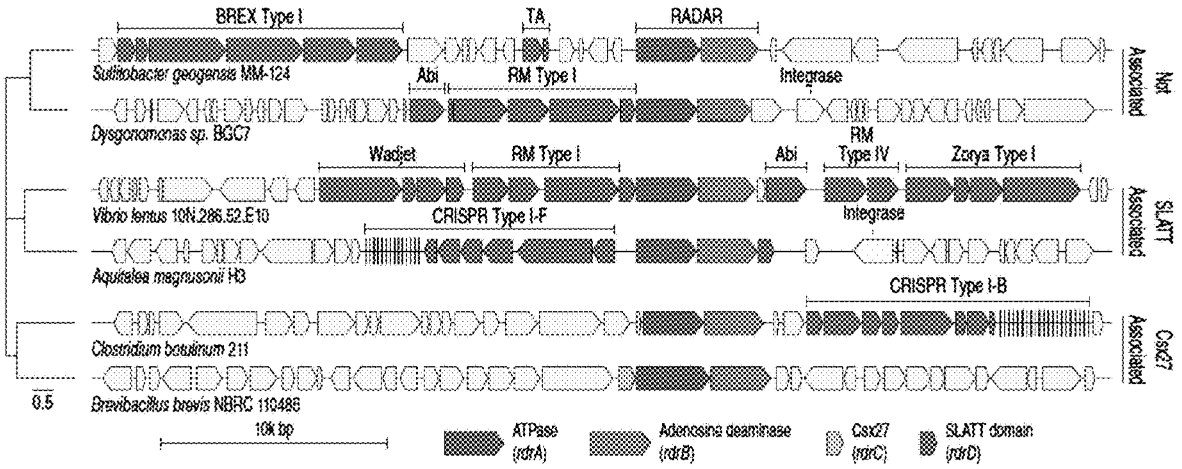


FIG. 18A

SEQ ID NOS: 477-480

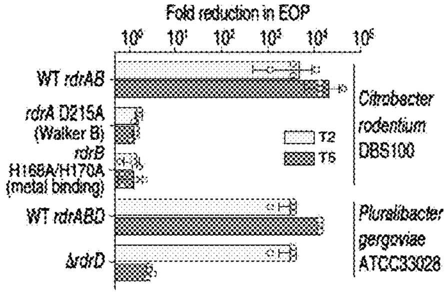


FIG. 18B

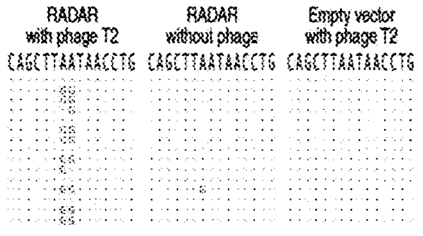


FIG. 18C

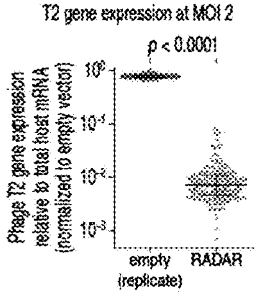


FIG. 18D

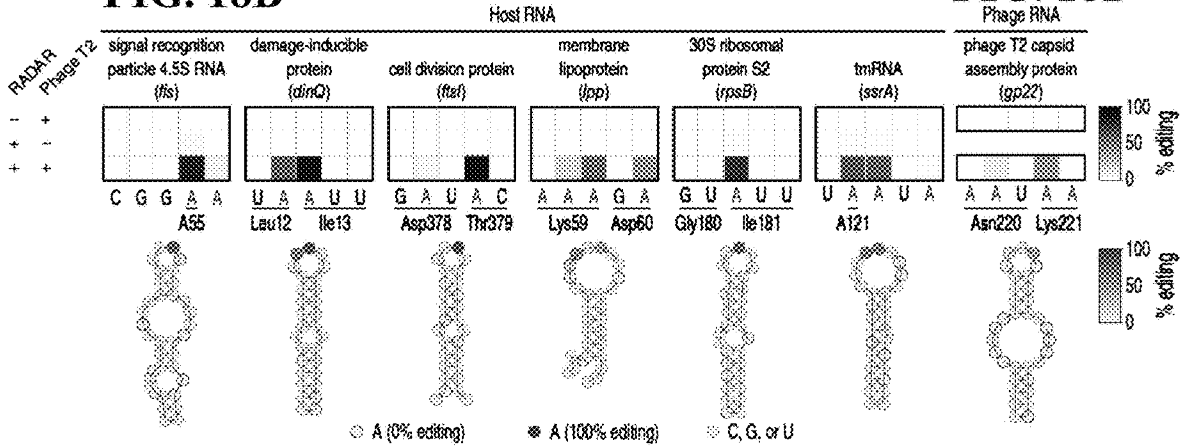


FIG. 18E

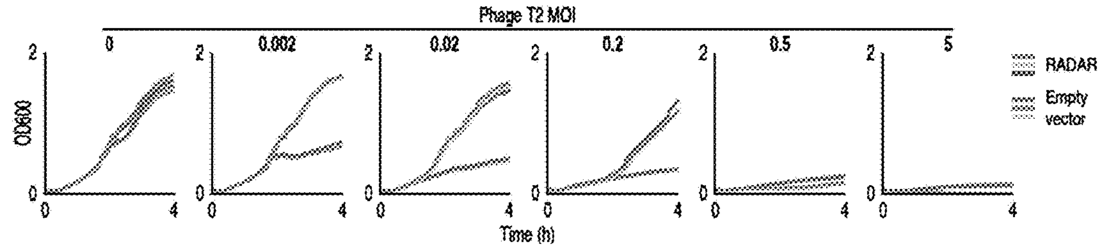


FIG. 18F

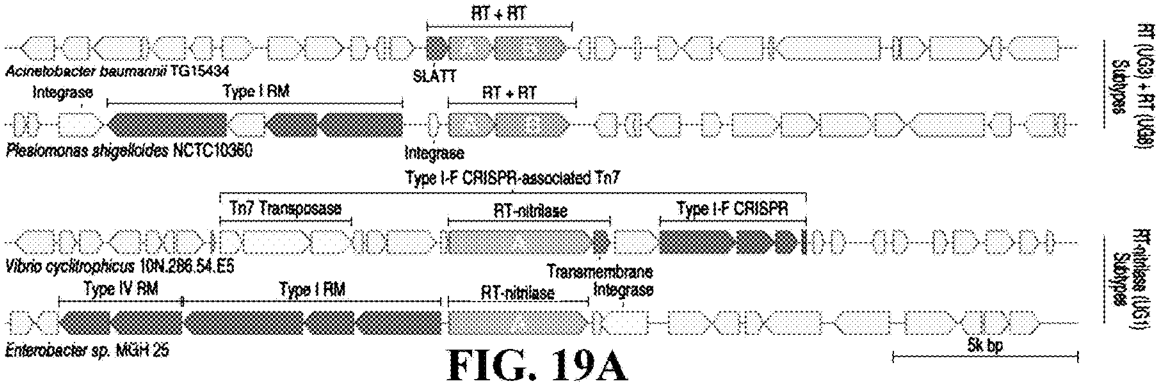


FIG. 19A

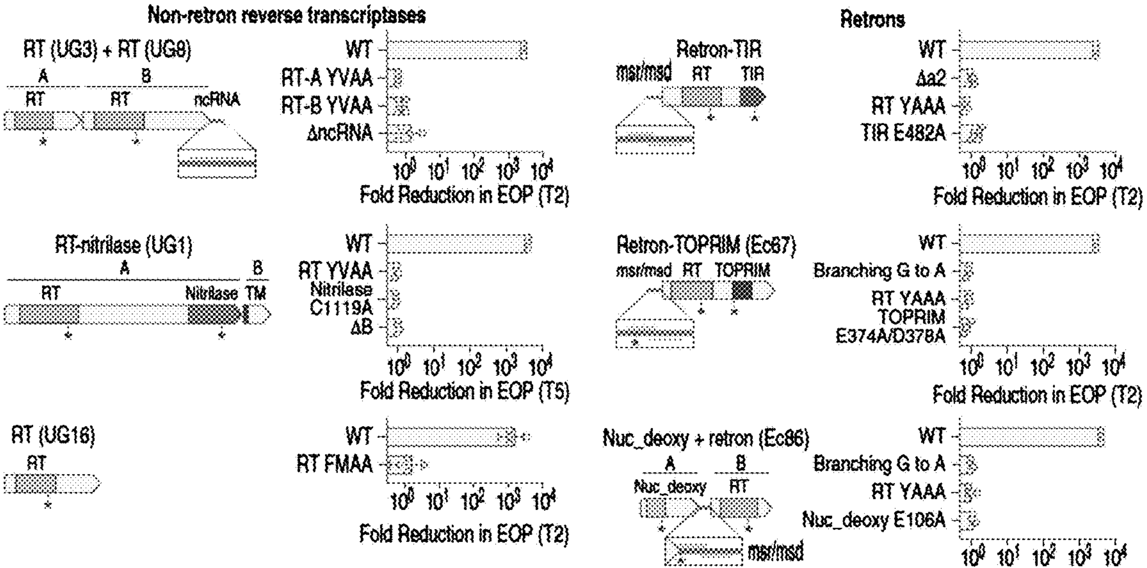


FIG. 19B

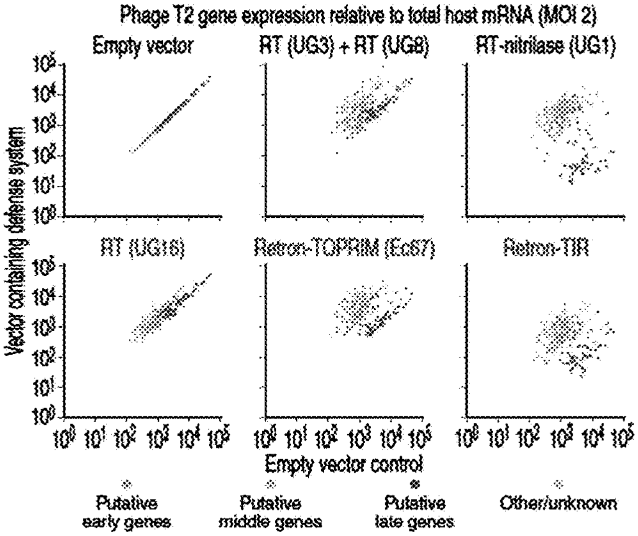


FIG. 19C

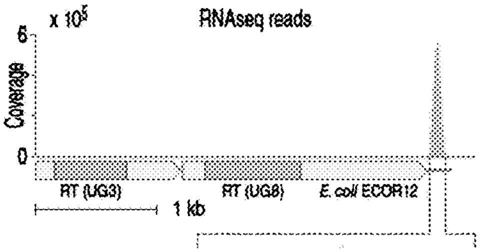


FIG. 19D

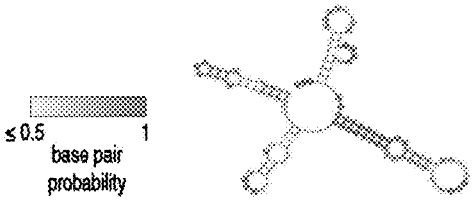


FIG. 19E

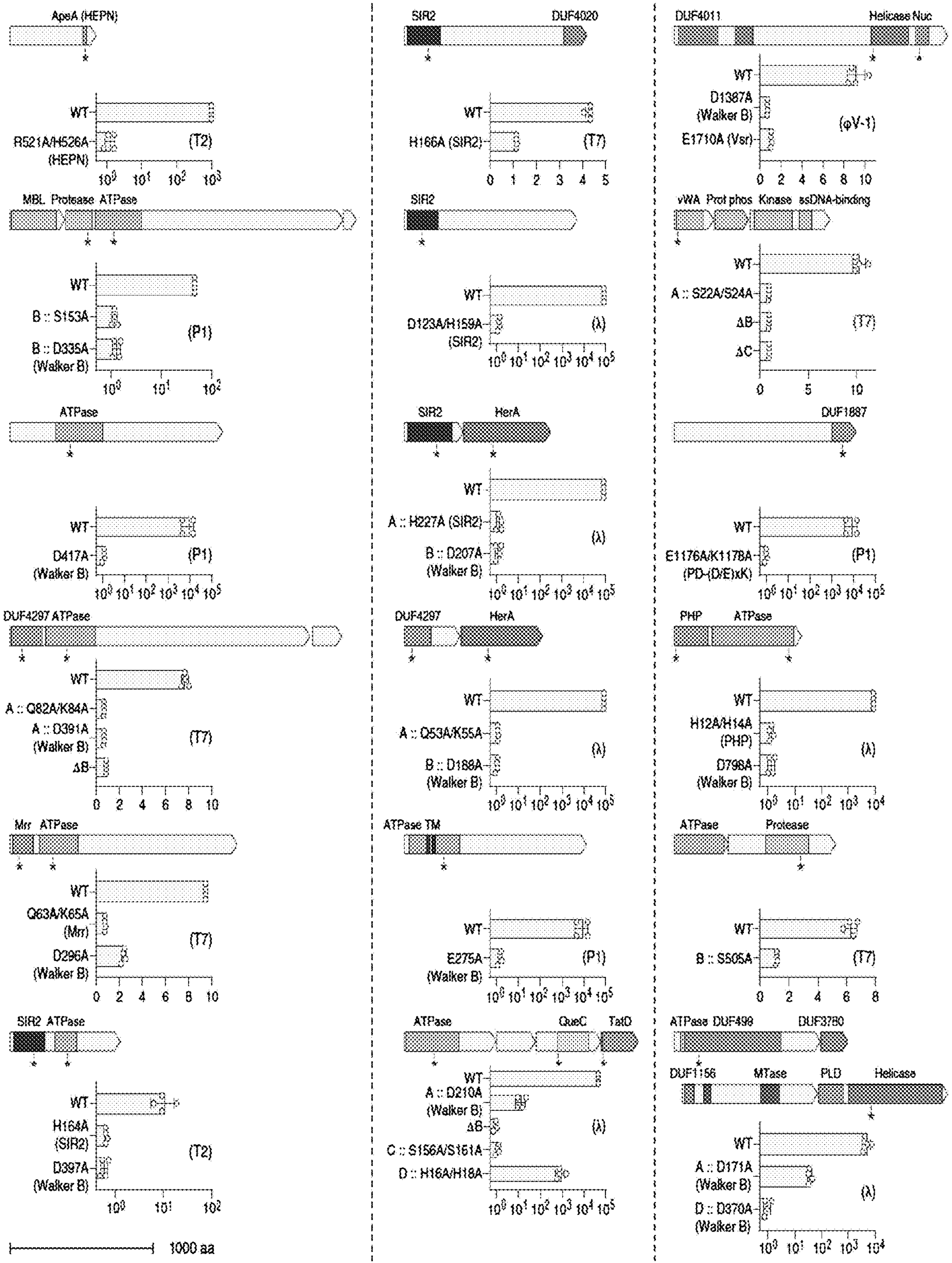

ApeA (HEPN)
WT
R521A/H526A (HEPN)
(T2)
10^0 10^1 10^2 10^3
MBL Protease ATPase
WT
B :: S153A
B :: D335A (Walker B)
(P1)
10^0 10^1 10^2
ATPase
WT
D417A (Walker B)
(P1)
10^0 10^1 10^2 10^3 10^4 10^5
DUF4297 ATPase
WT
A :: Q82A/K84A
A :: D391A (Walker B)
ΔB
(T7)
0 2 4 6 8 10
Mrr ATPase
WT
Q63A/K65A (Mrr)
D296A (Walker B)
(T7)
0 2 4 6 8 10
SIR2 ATPase
WT
H164A (SIR2)
D397A (Walker B)
(T2)
10^0 10^1 10^2
1000 aa
SIR2 DUF4020
WT
H166A (SIR2)
(T7)
0 1 2 3 4 5
SIR2
WT
D123A/H159A (SIR2)
(λ)
10^0 10^1 10^2 10^3 10^4 10^5
SIR2 HerA
WT
A :: H227A (SIR2)
B :: D207A (Walker B)
(λ)
10^0 10^1 10^2 10^3 10^4 10^5
DUF4297 HerA
WT
A :: Q53A/K55A
B :: D188A (Walker B)
(λ)
10^0 10^1 10^2 10^3 10^4 10^5
ATPase TM
WT
E275A (Walker B)
(P1)
10^0 10^1 10^2 10^3 10^4 10^5
ATPase QueC TatD
WT
A :: D210A (Walker B)
ΔB
C :: S156A/S161A
D :: H16A/H18A
(λ)
10^0 10^1 10^2 10^3 10^4 10^5
DUF4011 Helicase Nuc
WT
D1387A (Walker B)
E1710A (Vsr)
(φV-1)
0 2 4 6 8 10
vWA Prot phos Kinase ssDNA-binding
WT
A :: S22A/S24A
ΔB
ΔC
(T7)
0 5 10
DUF1887
WT
E1176A/K1178A (PD-(D/E)xK)
(P1)
10^0 10^1 10^2 10^3 10^4 10^5
PHP ATPase
WT
H12A/H14A (PHP)
D798A (Walker B)
(λ)
10^0 10^1 10^2 10^3 10^4
ATPase Protease
WT
B :: S505A
(T7)
0 2 4 6 8
ATPase DUF499 DUF3780
DUF1156 MTase PLD Helicase
WT
A :: D171A (Walker B)
D :: D370A (Walker B)
(λ)
10^0 10^1 10^2 10^3 10^4

FIG. 20

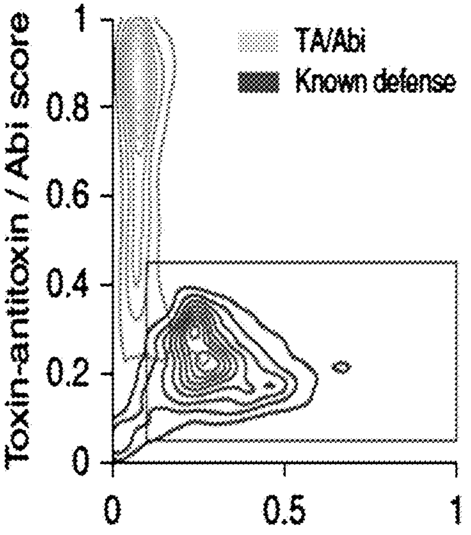


FIG. 21A

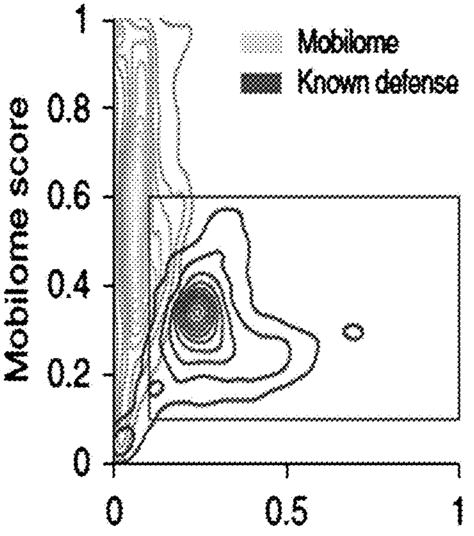


FIG. 21B

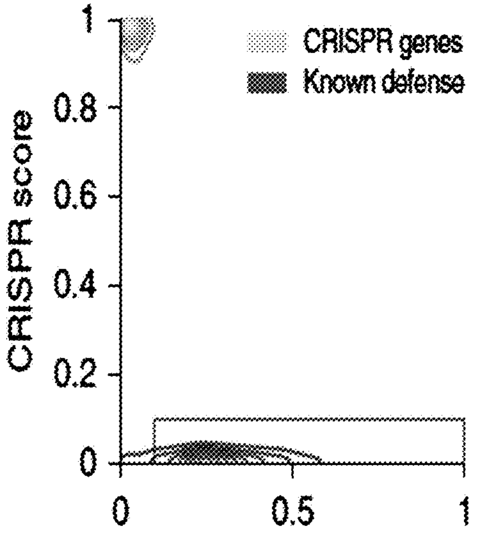


FIG. 21C

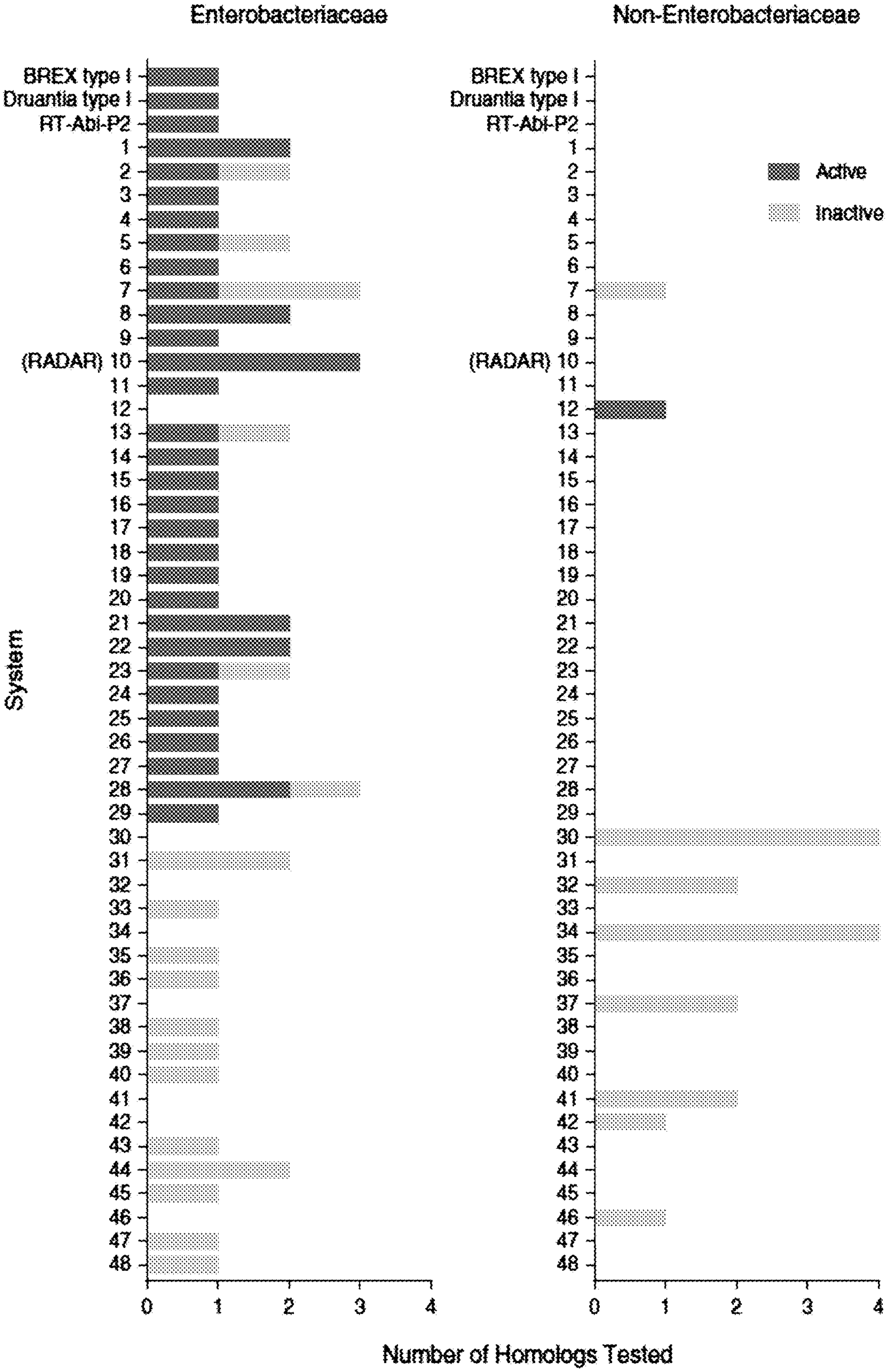

Enterobacteriaceae
Non-Enterobacteriaceae
BREX type I
Druantia type I
RT-Abi-P2
1
2
3
4
5
6
7
8
9
(RADAR) 10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
System
Active
Inactive
0
1
2
3
4
Number of Homologs Tested

FIG. 22

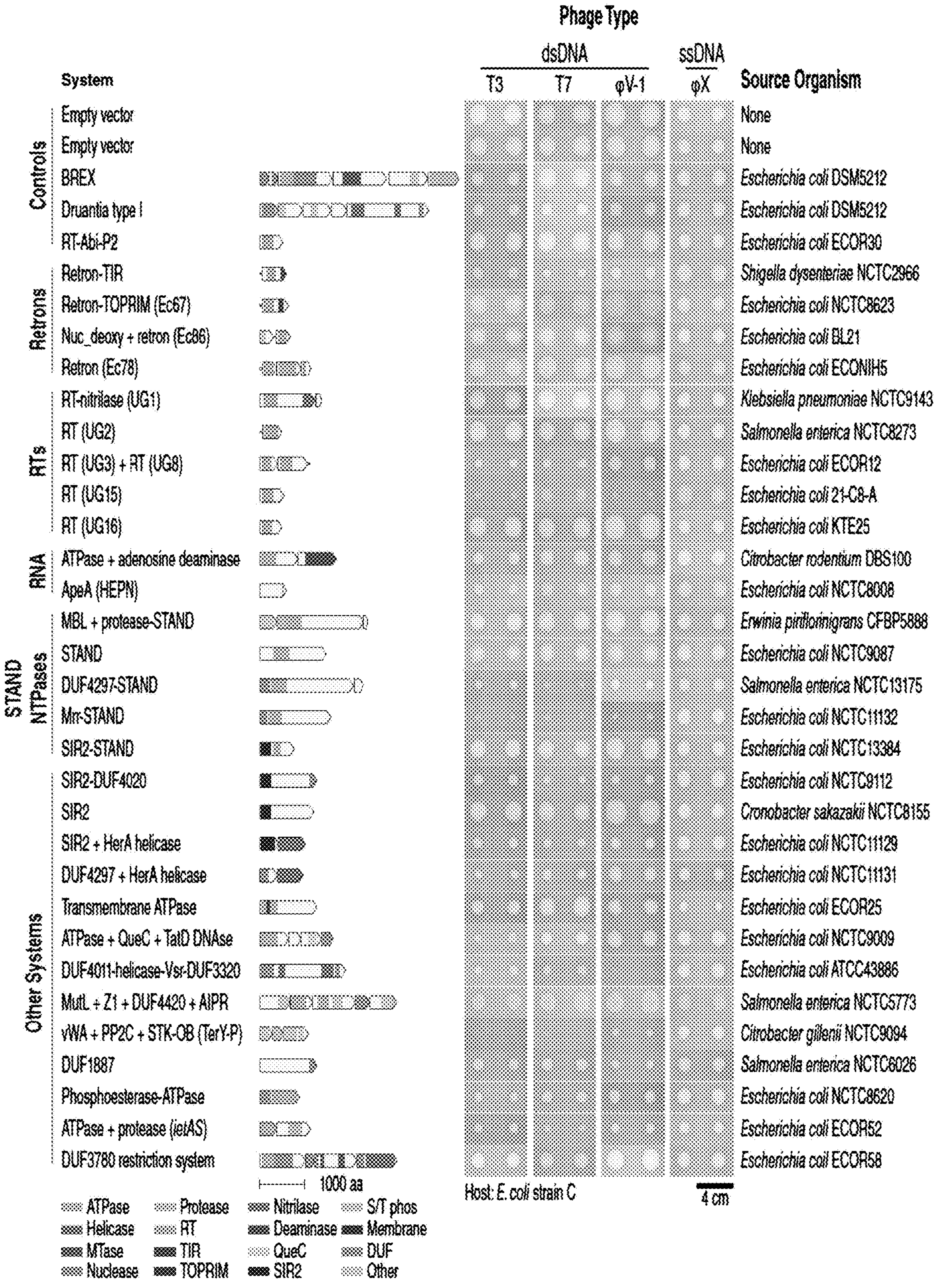

Phage Type
dsDNA
ssDNA
System
T3
T7
φV-1
φX
Source Organism
Controls
Empty vector
None
Empty vector
None
BREX
Escherichia coli DSM5212
Druantia type I
Escherichia coli DSM5212
RT-Abi-P2
Escherichia coli ECOR30
Retrons
Retron-TIR
Shigella dysenteriae NCTC2966
Retron-TOPRIM (Ec67)
Escherichia coli NCTC8623
Nuc_deoxy + retron (Ec86)
Escherichia coli BL21
Retron (Ec78)
Escherichia coli ECONIH5
RTs
RT-nitrilase (UG1)
Klebsiella pneumoniae NCTC9143
RT (UG2)
Salmonella enterica NCTC8273
RT (UG3) + RT (UG8)
Escherichia coli ECOR12
RT (UG15)
Escherichia coli 21-C8-A
RT (UG16)
Escherichia coli KTE25
RNA
ATPase + adenosine deaminase
Citrobacter rodentium DBS100
ApeA (HEPN)
Escherichia coli NCTC8008
STAND NTPases
MBL + protease-STAND
Erwinia piriflorinigrans CFBP5888
STAND
Escherichia coli NCTC9087
DUF4297-STAND
Salmonella enterica NCTC13175
Mrr-STAND
Escherichia coli NCTC11132
SIR2-STAND
Escherichia coli NCTC13384
Other Systems
SIR2-DUF4020
Escherichia coli NCTC9112
SIR2
Cronobacter sakazakii NCTC8155
SIR2 + HerA helicase
Escherichia coli NCTC11129
DUF4297 + HerA helicase
Escherichia coli NCTC11131
Transmembrane ATPase
Escherichia coli ECOR25
ATPase + QueC + TatD DNAse
Escherichia coli NCTC9009
DUF4011-helicase-Vsr-DUF3320
Escherichia coli ATCC43886
MutL + Z1 + DUF4420 + AIPR
Salmonella enterica NCTC5773
vWA + PP2C + STK-OB (TerY-P)
Citrobacter gillenii NCTC9094
DUF1887
Salmonella enterica NCTC6026
Phosphoesterase-ATPase
Escherichia coli NCTC8620
ATPase + protease (ietAS)
Escherichia coli ECOR52
DUF3780 restriction system
Escherichia coli ECOR58
1000 aa
Host: E. coli strain C
4 cm
ATPase
Protease
Nitrilase
S/T phos
Helicase
RT
Deaminase
Membrane
MTase
TIR
QueC
DUF
Nuclease
TOPRIM
SIR2
Other

FIG. 23

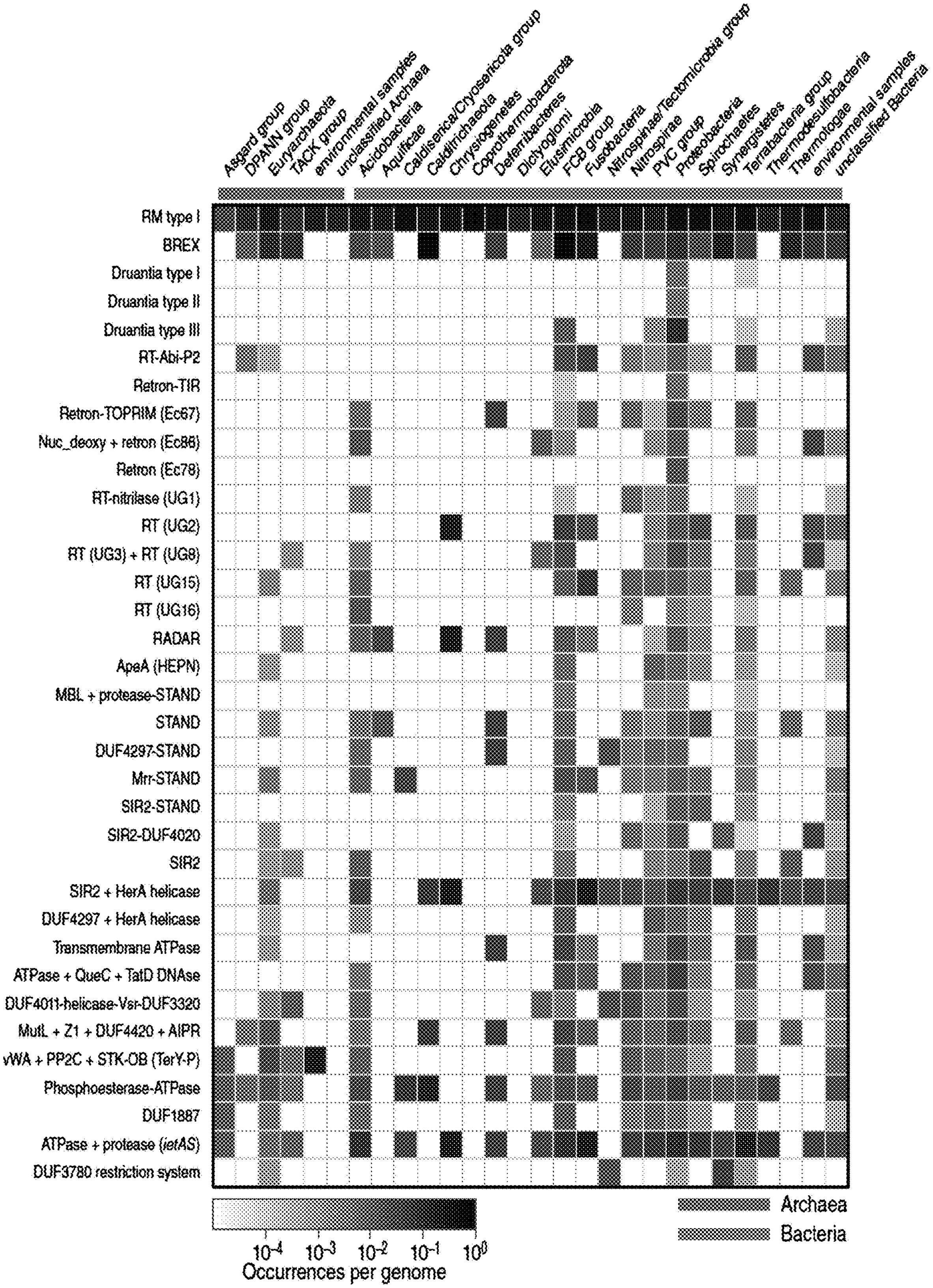

Asgard group
DPANN group
Euryarchaeota
TACK group
environmental samples
unclassified Archaea
Acidobacteria
Aquificae
Caldiserica/Cryosericota group
Calditrichaeota
Chrysiogenetes
Coprothermobacterota
Deferribacteres
Dictyoglomi
Elusimicrobia
FCB group
Fusobacteria
Nitrospinae/Tectomicrobia group
Nitrospirae
PVC group
Proteobacteria
Spirochaetes
Synergistetes
Terrabacteria group
Thermodesulfobacteria
Thermotogae
environmental samples
unclassified Bacteria
RM type I
BREX
Druantia type I
Druantia type II
Druantia type III
RT-Abi-P2
Retron-TIR
Retron-TOPRIM (Ec67)
Nuc_deoxy + retron (Ec86)
Retron (Ec78)
RT-nitrilase (UG1)
RT (UG2)
RT (UG3) + RT (UG8)
RT (UG15)
RT (UG16)
RADAR
ApeA (HEPN)
MBL + protease-STAND
STAND
DUF4297-STAND
Mrr-STAND
SIR2-STAND
SIR2-DUF4020
SIR2
SIR2 + HerA helicase
DUF4297 + HerA helicase
Transmembrane ATPase
ATPase + QueC + TatD DNAse
DUF4011-helicase-Vsr-DUF3320
MutL + Z1 + DUF4420 + AIPR
vWA + PP2C + STK-OB (TerY-P)
Phosphoesterase-ATPase
DUF1887
ATPase + protease (*ietAS*)
DUF3780 restriction system
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}
Occurrences per genome
Archaea
Bacteria

FIG. 24

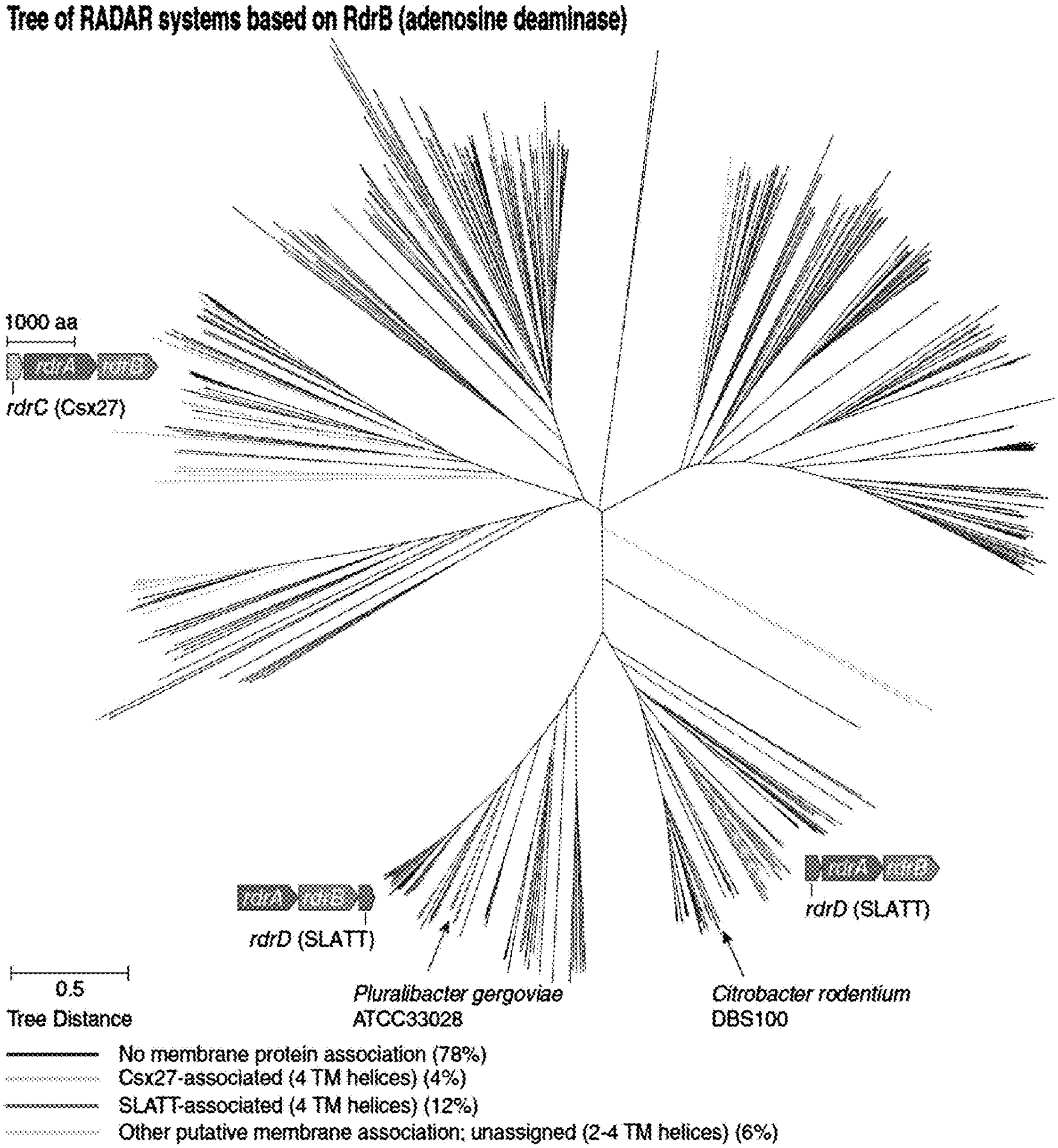


FIG. 25A

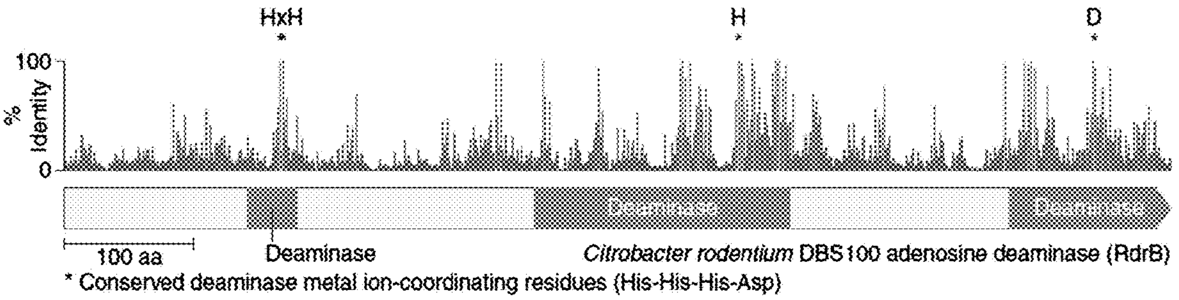


FIG. 25B

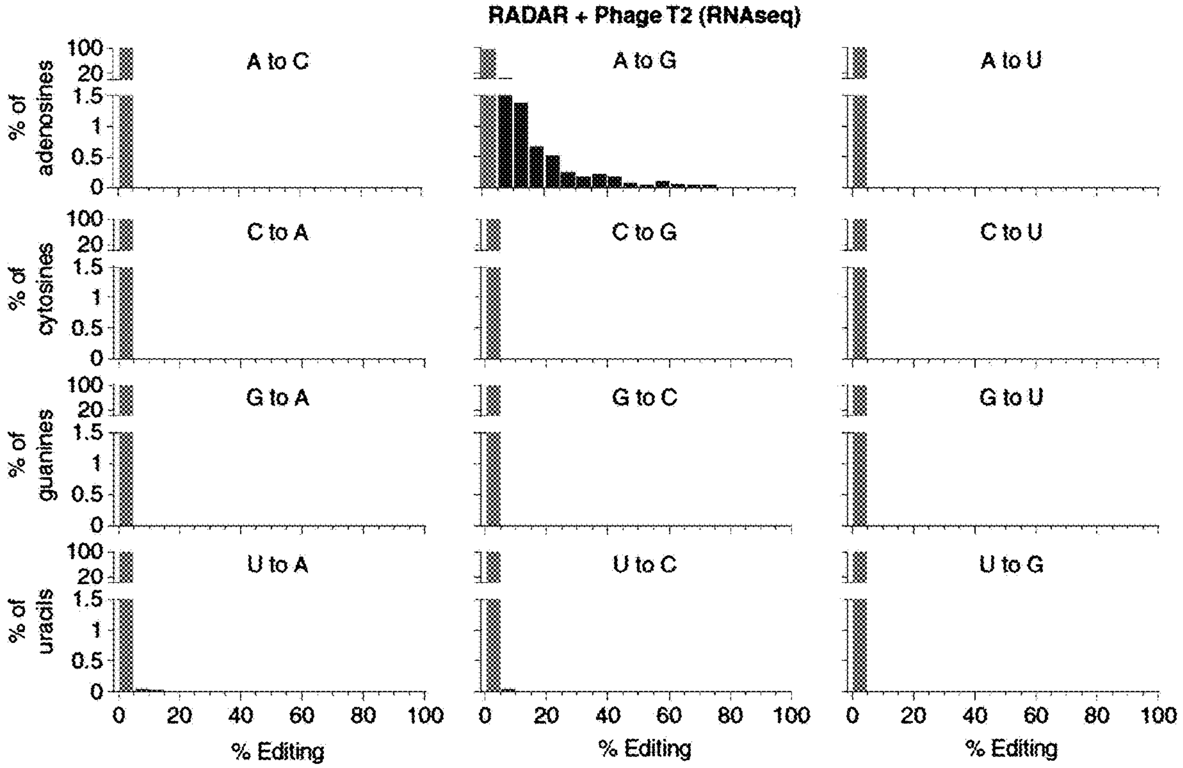


FIG. 26A

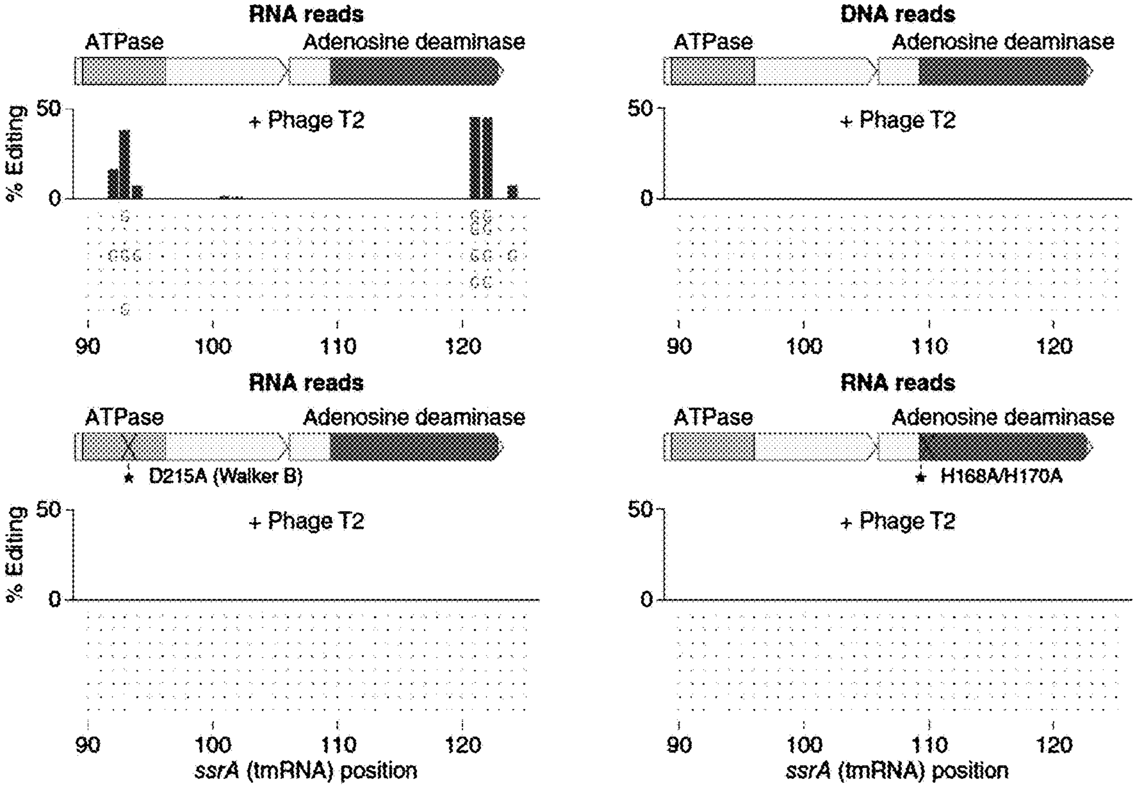


FIG. 26B

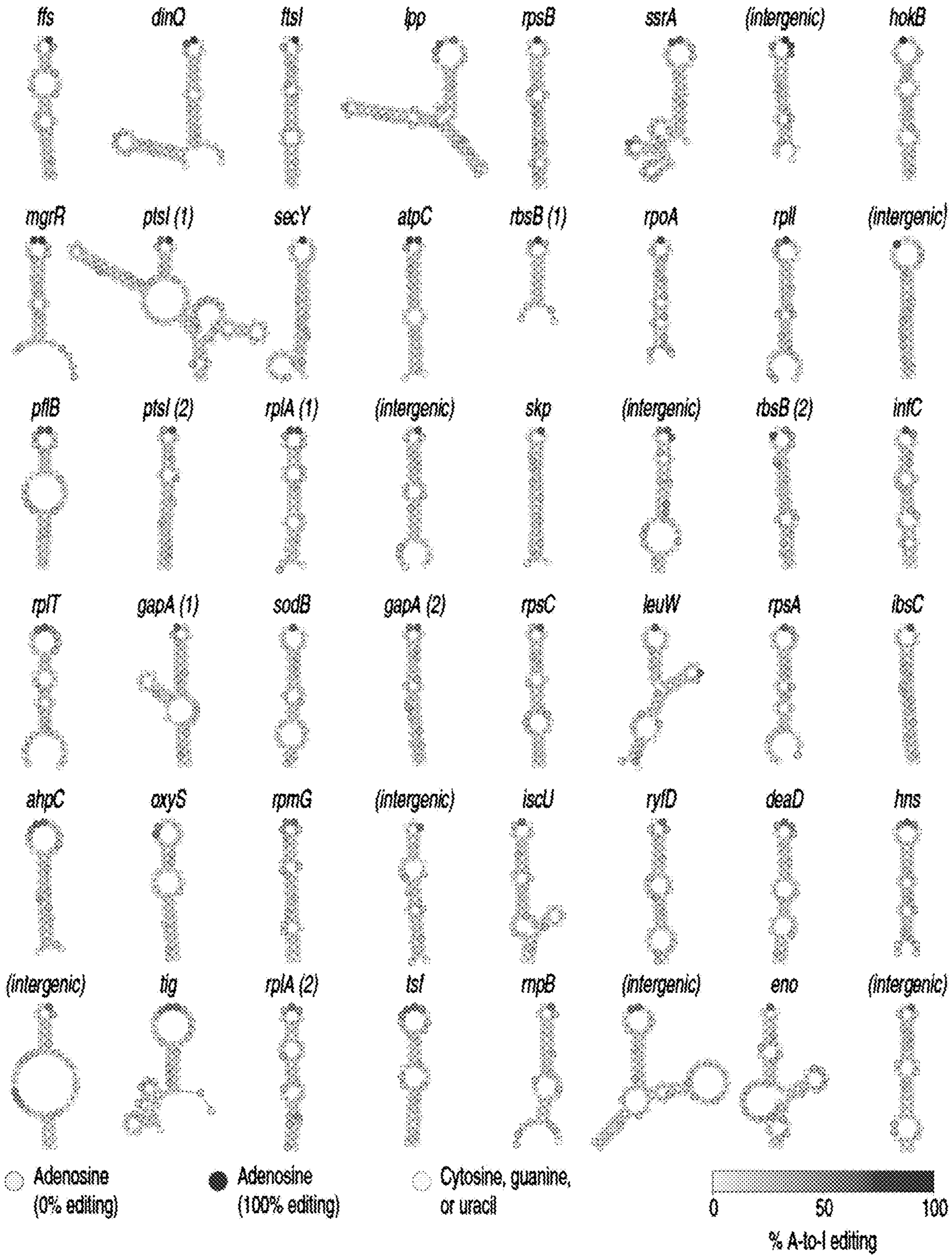

ffs
dinQ
ftsI
lpp
rpsB
ssrA
(intergenic)
hokB
mgrR
ptsI (1)
secY
atpC
rbsB (1)
rpoA
rplI
(intergenic)
pflB
ptsI (2)
rplA (1)
(intergenic)
skp
(intergenic)
rbsB (2)
infC
rplT
gapA (1)
sodB
gapA (2)
rpsC
leuW
rpsA
ibsC
ahpC
oxyS
rpmG
(intergenic)
iscU
ryfD
deaD
hns
(intergenic)
tig
rplA (2)
tsf
rnpB
(intergenic)
eno
(intergenic)
○ Adenosine (0% editing)
● Adenosine (100% editing)
○ Cytosine, guanine, or uracil
0
50
100
% A-to-I editing

FIG. 27

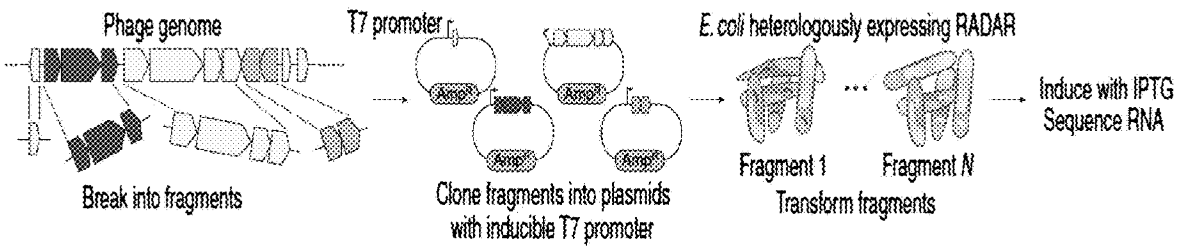


FIG. 28A

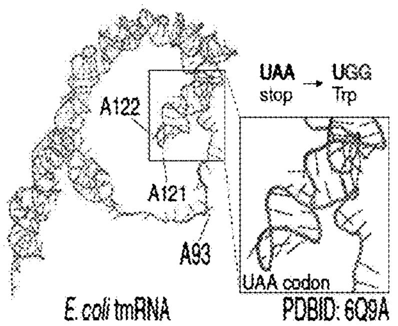


FIG. 28B

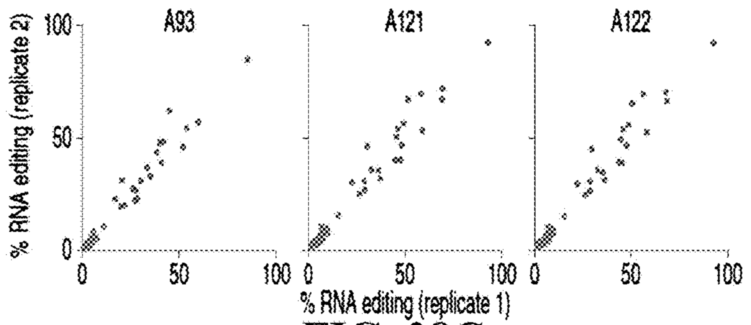


FIG. 28C

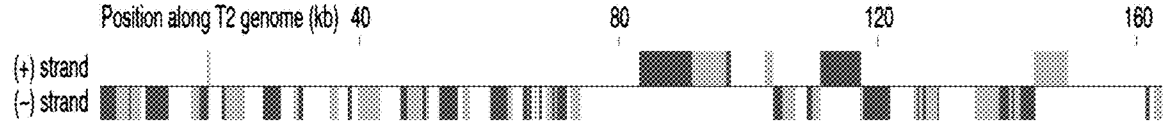


FIG. 28D

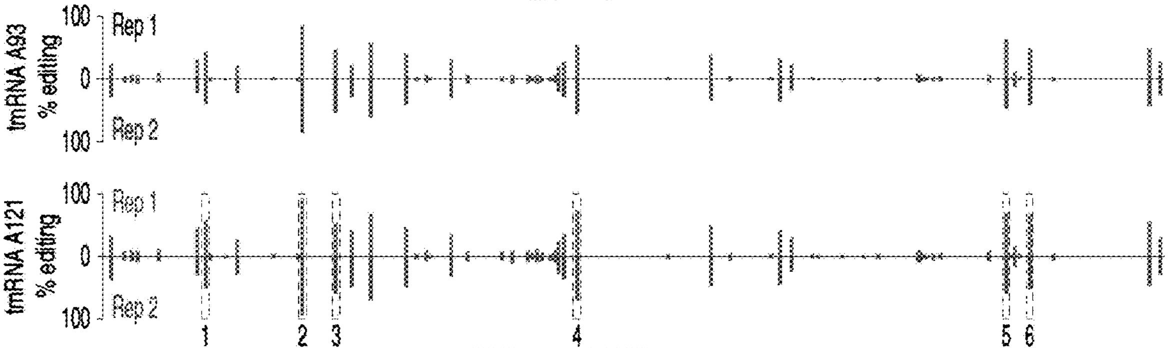


FIG. 28E

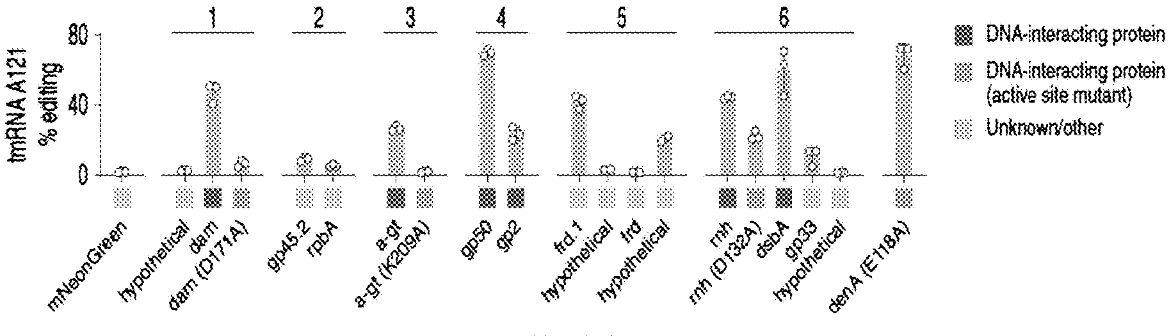


FIG. 28F

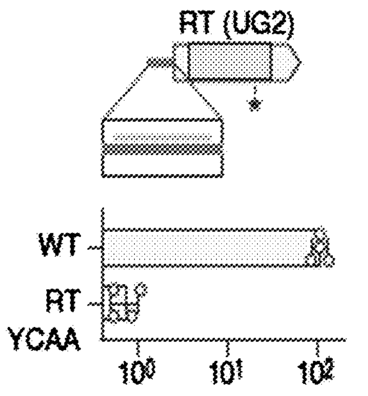


FIG. 29A

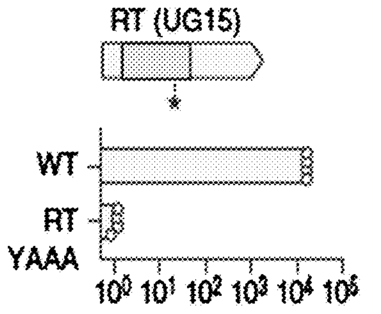


FIG. 29B

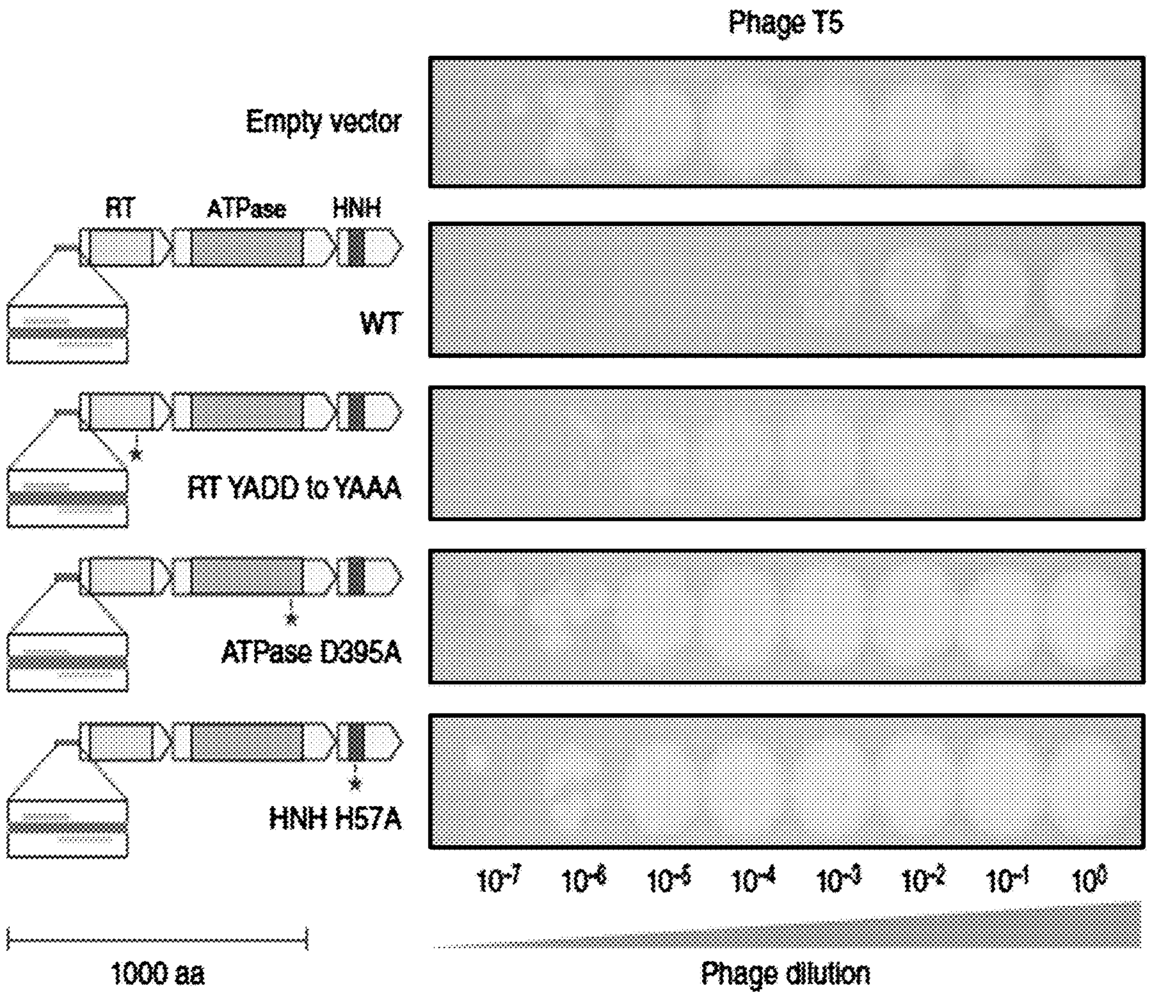


FIG. 29C

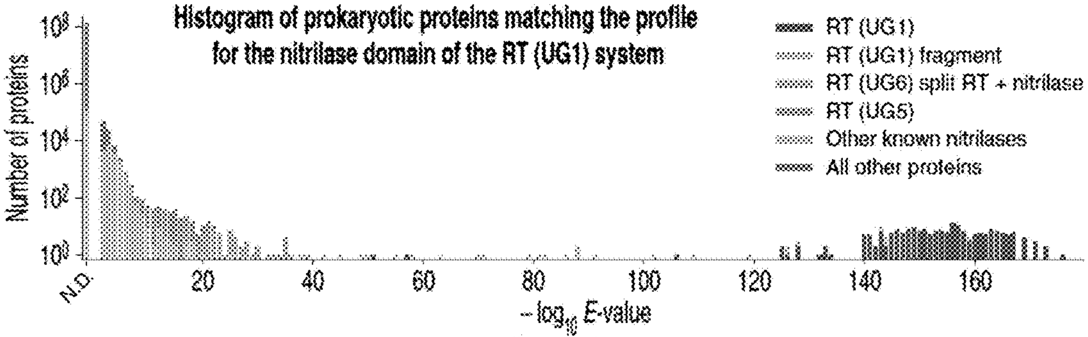


FIG. 30A

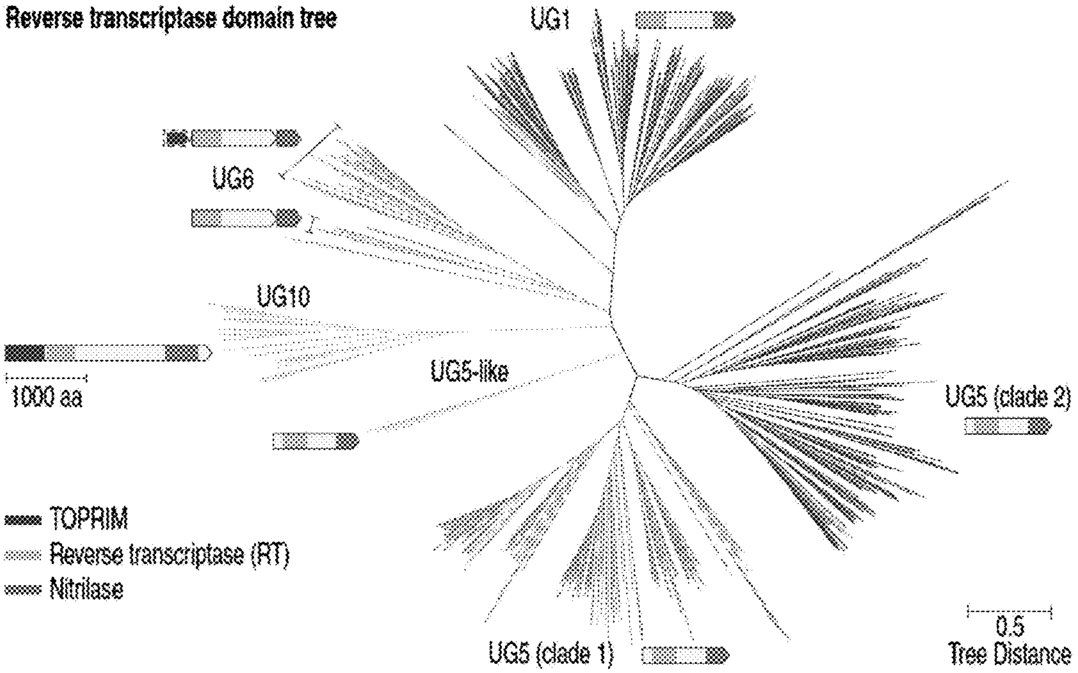


FIG. 30B

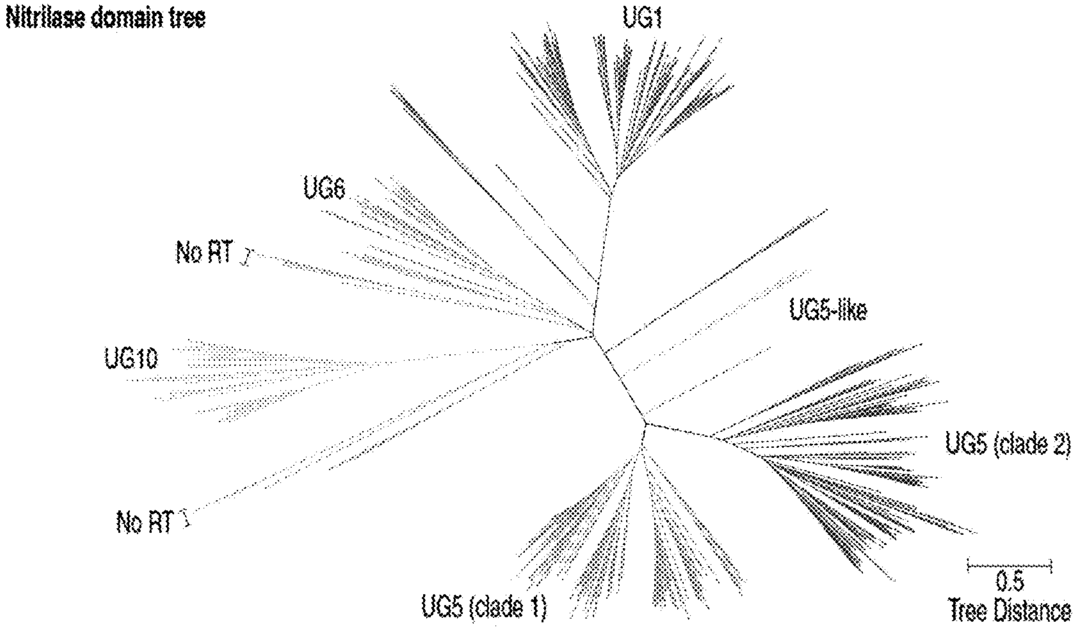


FIG. 30C

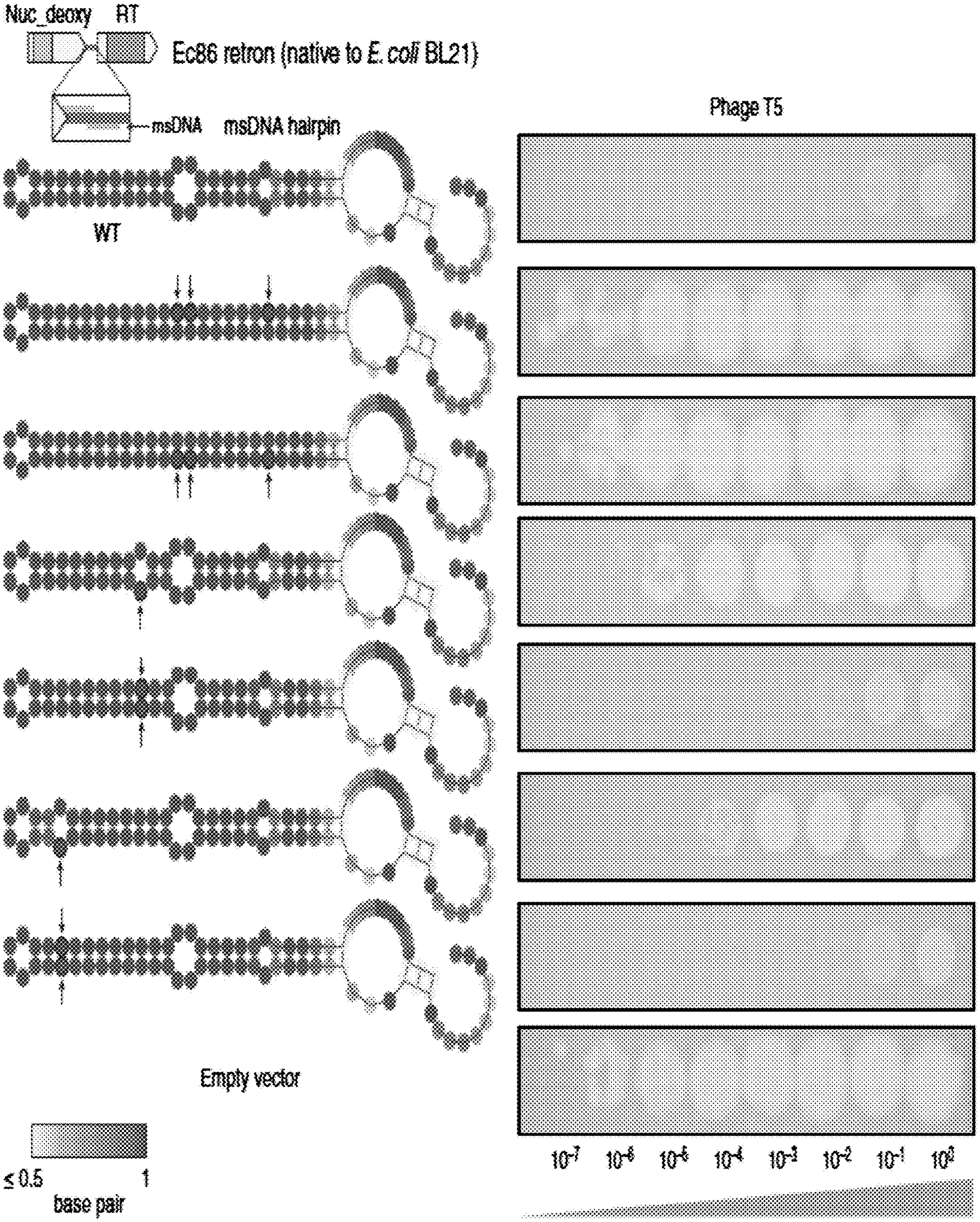

Nuc_deoxy
RT
Ec86 retron (native to *E. coli* BL21)
msDNA
msDNA hairpin
WT
Empty vector
Phage T5
≤ 0.5
1
base pair
10-7
10-6
10-5
10-4
10-3
10-2
10-1
100

FIG. 31

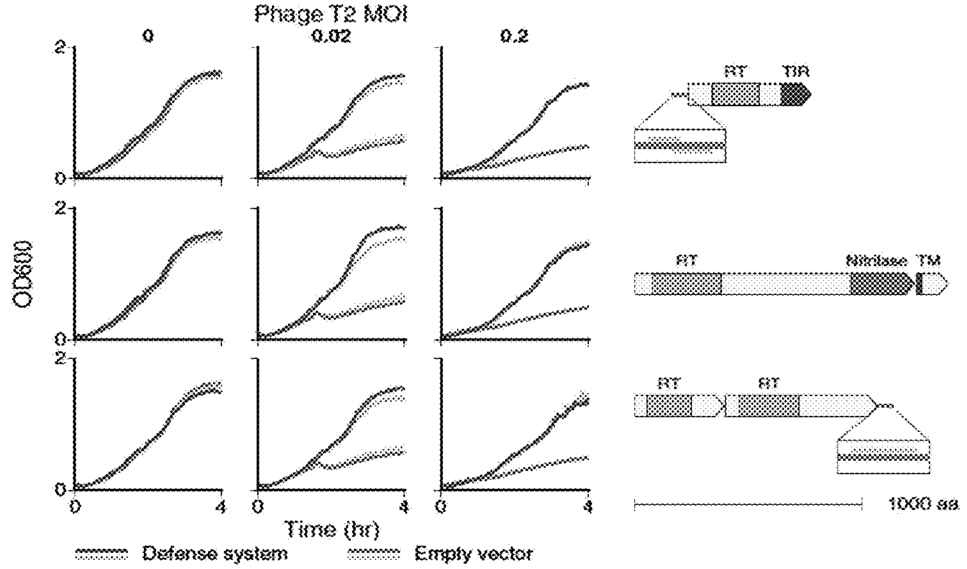


FIG. 32A

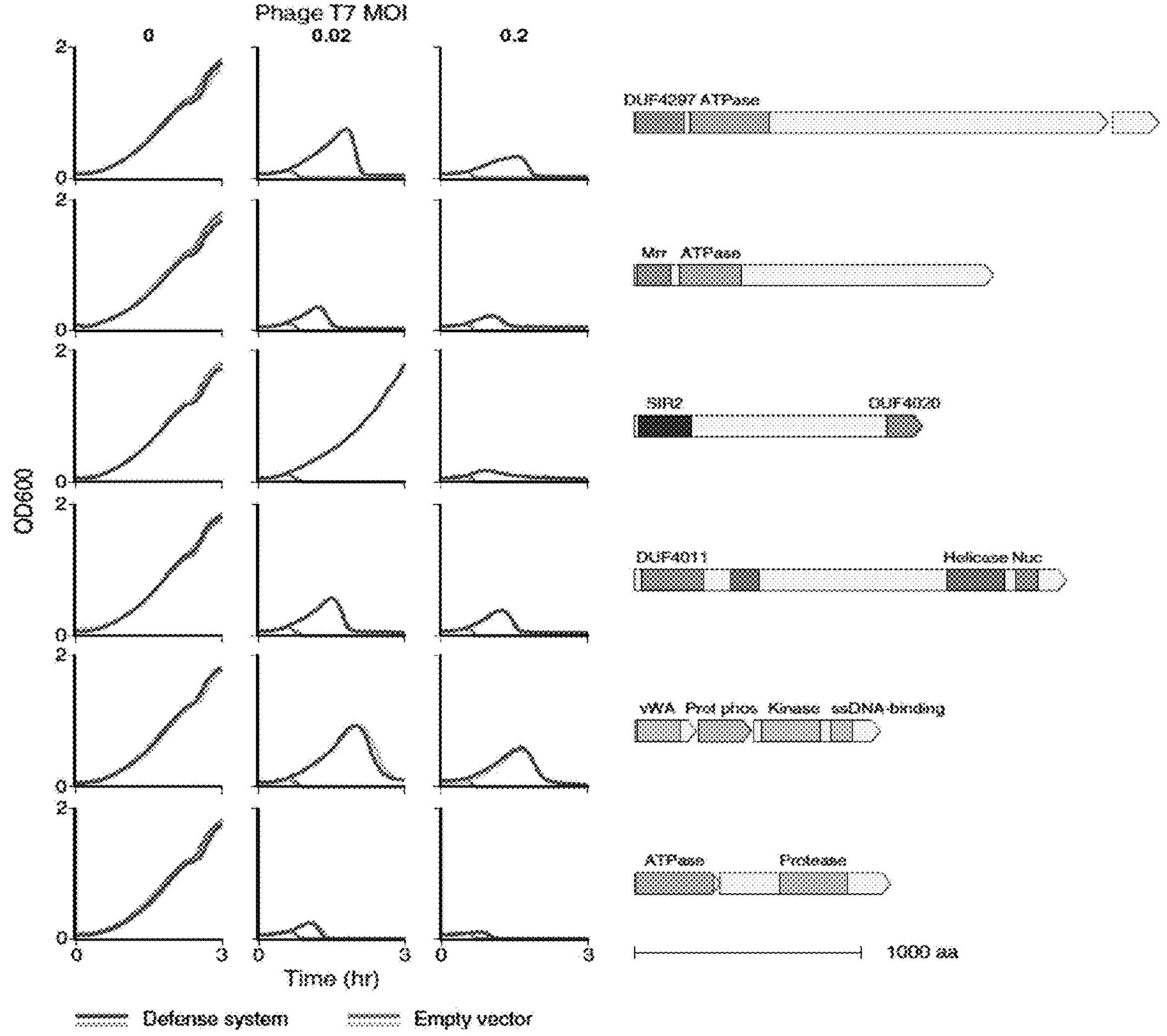


FIG. 32B

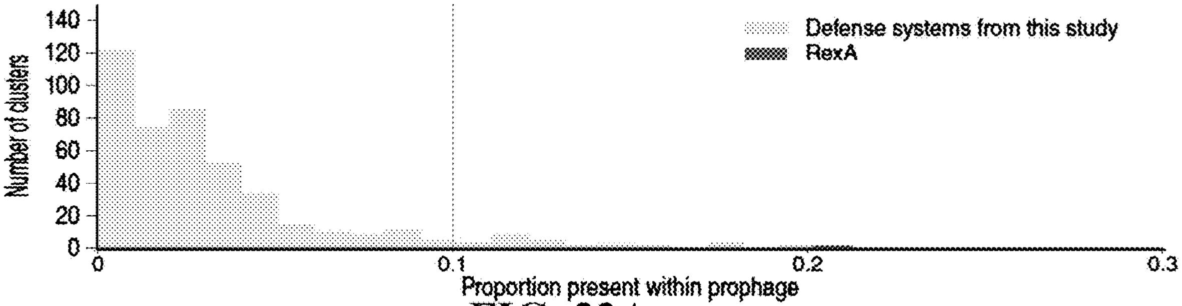


FIG. 33A

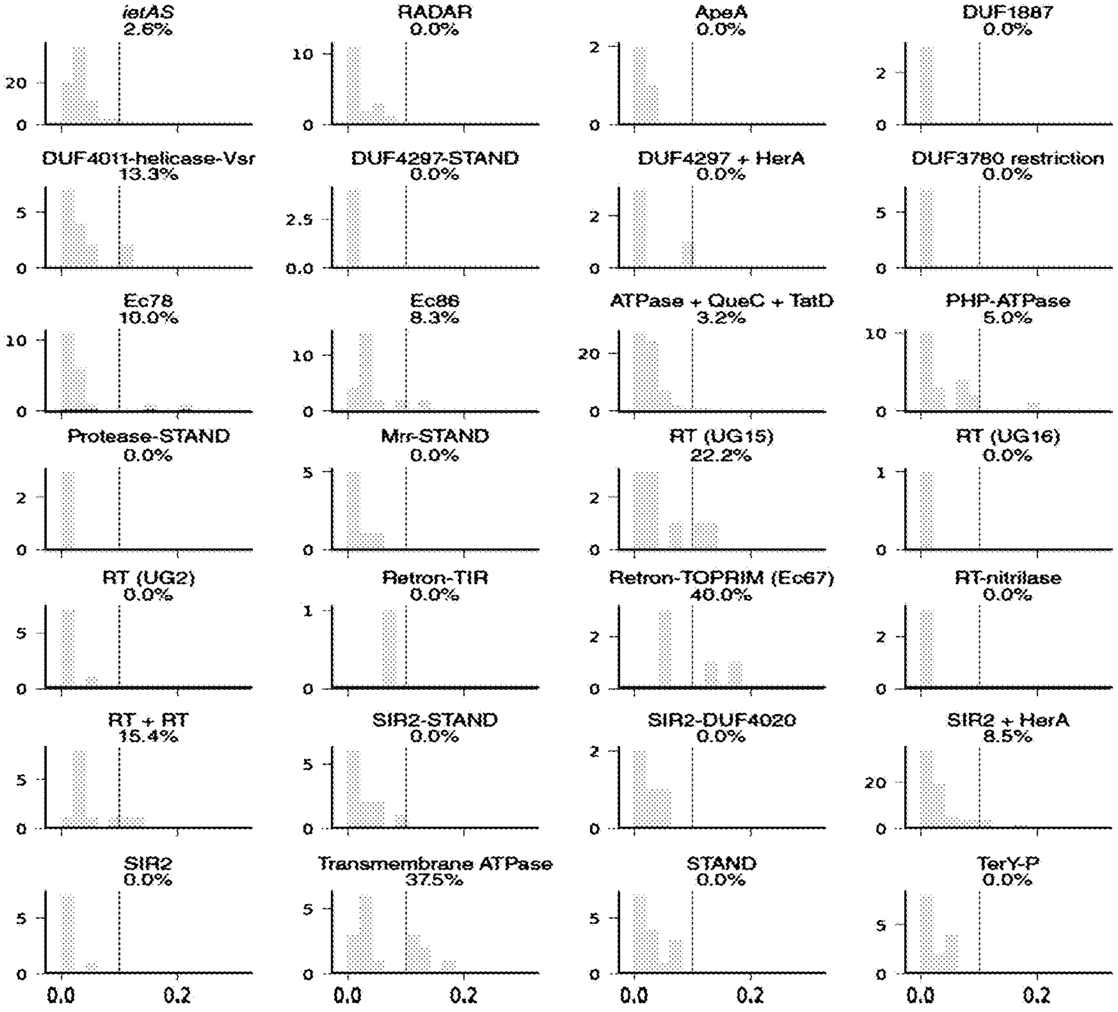


FIG. 33B

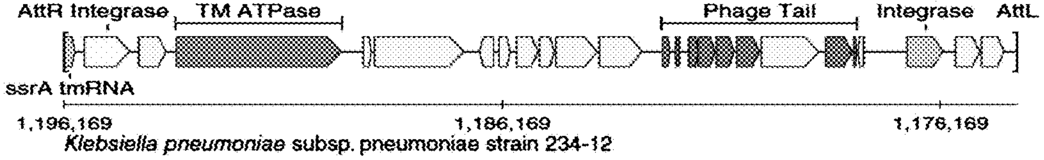


FIG. 33C

BACTERIAL DEFENSE SYSTEMS AND METHODS OF IDENTIFYING THEREOF

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 62/928,269, filed Oct. 30, 2019, and U.S. Provisional Application No. 63/051,161, filed Jul. 13, 2020. The entire contents of the above-identified applications are hereby fully incorporated herein by reference.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with government support under Grant Nos. HG009761, MH110049, and HL141201 awarded by the National Institutes of Health. The government has certain rights in the invention.

REFERENCE TO AN ELECTRONIC SEQUENCE LISTING

[0003] The contents of the electronic sequence listing ("BROD-4610US_ST25.txt"; Size is 2,039,992 bytes and it was created on Oct. 30, 2020) is herein incorporated by reference in its entirety.

TECHNICAL FIELD

[0004] The subject matter disclosed herein is generally directed to bacterial defense systems and methods of identifying thereof.

BACKGROUND

[0005] To survive from attacks by viruses (e.g., phages), bacteria have developed a variety of defense systems, including proteins and nucleic acids that help recognize and eliminate foreign proteins and nucleic acids, e.g., those from the infecting phages. A number of bacteria defense systems have been discovered, many of which have been adopted and engineered to tools in biotechnology. An example is the CRISPR-Cas systems, which recognize and cleave foreign RNA or DNA in bacteria and have been developed as a powerful gene editing tool. In view of the great potential of bacterial defense systems in biotechnology and new therapeutic or diagnostic applications, there is a need for identification of novel defense systems in a high throughput manner.

SUMMARY

[0006] In one aspect, the present disclosure provides an engineered system comprising an ATPase and an adenosine deaminase. In some embodiments, the ATPase comprises a sequence of WP_012906049.1 or WP_155731552.1, and the adenosine deaminase comprises a sequence of WP_012906048.1 or WP_064360593.1. In some embodiments, the ATPase comprises 1100 or less amino acid residues. In some embodiments, the adenosine deaminase comprises 1100 or less amino acid residues. In some embodiments, the system further comprises a membrane protein. In some embodiments, the membrane protein comprises a SLATT domain or Csx27. In some embodiments, the system is configured to modify a target nucleic acid. In some embodiments, the target nucleic acid is RNA. In some embodiments, the modification of the target nucleic acid comprises causing an A to G mutation in the target nucleic acid. In some embodiments, the system further comprises one or more phage proteins. In some embodiments, the one or more phage proteins are in Tables 18A-18B.

[0007] In another aspect, the present disclosure provides an engineered system comprising one or more reverse transcriptases comprising one or more UG1, UG2, UG3, UG8, UG15, or UG16 reverse transcriptase. In some embodiments, the system comprises a first and a second reverse transcriptase. In some embodiments, the first and the second reverse transcriptases are comprised in a protein. In some embodiments, the system further comprises a SLATT domain. In some embodiments, the system further comprises a DNA polymerase. In some embodiments, the DNA polymerase is a family A DNA polymerase. In some embodiments, the system further comprises a serine protease domain linked to or associated with the reverse transcriptase. In some embodiments, the system further comprises an MBL domain. In some embodiments, the system further comprises a nitrilase. In some embodiments, the nitrilase and the one or more reverse transcriptases are comprised in a protein, and the nitrilase is at a C-terminus of the protein. In some embodiments, the system further comprises a non-coding RNA element. In some embodiments, the reverse transcriptase comprises an active site, e.g., (Y/F)×DD (SEQ ID NO: 1-2), where X is any amino acid.

[0008] In another aspect, the present disclosure provides an engineered system comprising a retron or one or more molecules encoded by the retron. In some embodiments, the retron is an Ec67 retron. In some embodiments, the retron is an Ec86 retron. In some embodiments, the retron is an Ec78 retron. In some embodiments, the retron is a Tol/interleukin 1 receptor (TIR) domain-associated retron. In some embodiments, the TIR domain has NAD+ hydrolase activity. In some embodiments, the retron is a topoisomerase-primase (TOPRIM) domain-associated retron. In some embodiments, the TOPRIM domain has nuclease activity.

[0009] In another aspect, the present disclosure provides an engineered system comprising an NTPase of a STAND (signal transduction ATPases with numerous associated domains) superfamily. In some embodiments, the system further comprises DUF4297, Mrr-like nuclease, SIR2, a trypsin-like serine protease, and/or a helical domain.

[0010] In another aspect, the present disclosure provides an engineered system comprising a von Willebrand factor (VWF), a PP2C-like serine/threonine protein phosphatase, and a serine/threonine kinase.

[0011] In another aspect, the present disclosure provides an engineered system comprising SIR2 or a function domain thereof.

[0012] In another aspect, the present disclosure provides an engineered system comprising a transmembrane ATPase.

[0013] In another aspect, the present disclosure provides an engineered system comprising an ATPase, QueC synthase, and TatD endonuclease.

[0014] In another aspect, the present disclosure provides an engineered system comprising a S8 peptidase.

[0015] In another aspect, the present disclosure provides an engineered system comprising DUF4011, a helicase, an a Vsr endonuclease.

[0016] In another aspect, the present disclosure provides an engineered system comprising a silent information regulator (SIR)2-DUF4020.

[0017] In another aspect, the present disclosure provides an engineered system comprising a Polymerase and Histidinol Phosphatase (PHP)-ATPase.

[0018] In another aspect, the present disclosure provides an engineered system comprising SIR2 and HerA.

[0019] In another aspect, the present disclosure provides an engineered system comprising DUF4297 and HerA.

[0020] In another aspect, the present disclosure provides an engineered system comprising DUF 1887.

[0021] In another aspect, the present disclosure provides an engineered system comprising DUF499, DUF3780, and DUF1156 methyltransferase and a helicase.

[0022] In another aspect, the present disclosure provides an engineered system comprising a type I-E CRISPR-associated ATPase.

[0023] In another aspect, the present disclosure provides an engineered system comprising ApeA.

[0024] In some embodiments, any one of the systems herein comprises two proteins fused together. In some embodiments, any one of the systems herein comprises one or more components in a retrotransposon system.

[0025] In another aspect, the present disclosure provides a polynucleotide comprising coding sequences for one or more proteins in the system herein.

[0026] In another aspect, the present disclosure provides a vector comprising a polynucleotide herein.

[0027] In another aspect, the present disclosure provides a cell comprising the polynucleotide herein.

[0028] In another aspect, the present disclosure provides a method of identifying a defense system in a microorganism, the method comprising: identifying genes of known defense systems in a plurality of genomes of the microorganism; recording candidate genes located within 10 kb or 10 open reading frames from the identified genes of known defense systems in the genomes; identifying homologs of each candidate gene in the genomes; and selecting candidate genes, wherein at least 10% of homologs of the candidate genes are within 5000 nucleotides or 5 genes from one or more known defense systems on the genomes.

[0029] In some embodiments, identifying genes of known defense systems comprises identifying known defense genes and filtering false positive hits among the identified known defense genes. In some embodiments, the method further comprises validating the selected candidate genes. In some embodiments, the homologs of the candidate genes share at least 70% sequence identity with the candidate genes and/or the homologs have an e-value of 10^{-5} or lower. In some embodiments, the recorded candidate genes are within 10 kb from the identified genes of known defense systems on the genomes. In some embodiments, at least 15% of homologs of the selected candidate genes are within 5000 nucleotides or 5 genes from one or more known defense systems on the genomes. In some embodiments, the plurality of genomes comprises at least 100,000 genomes. In some embodiments, the known defense systems comprise one or more of a CRISPR system, Type I RM and McrBC system, BREX-associated system, Zorya system, Wadjet system, Druantia-associated system, Hachiman system, Lamassu system, Thoeris-like system, Gabija system, Septu system, pAgo system, Shedu system, Kiwa system, DUF499-DUF1156 system, and Toxin/antitoxin system. In some embodiments, the microorganism is *E. coli.*

[0030] These and other aspects, objects, features, and advantages of the example embodiments will become apparent to those having ordinary skill in the art upon consideration of the following detailed description of illustrated example embodiments.

BRIEF DESCRIPTION OF THE DRAWINGS

[0031] An understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention may be utilized, and the accompanying drawings of which:

[0032] FIGS. **1**A-**1**Y. FIG. **1**A shows diagrams of exemplary identified defense system comprising reverse transcriptase and nitrilase. FIG. **1**B shows diagrams of exemplary identified defense system comprising a reverse transcriptase and a nitrilase, and a topoisomerase-primase (TOPRIM). FIG. **1**C shows diagrams of exemplary identified defense system comprising a reverse transcriptase and TOPRIM. FIG. **1**D shows diagrams of exemplary identified defense system comprising a reverse transcriptase. FIG. **1**E shows diagrams of exemplary identified defense system comprising a deaminase. FIG. **1**F shows diagrams of exemplary identified defense system comprising a transmembrane ATPase. FIG. **1**G shows diagrams of exemplary identified defense system comprising an ATPase, QueC synthase, and TatD endonuclease. FIG. **1**H shows diagrams of exemplary identified defense system comprising a protease. FIG. **1**I shows diagrams of exemplary identified defense system comprising DUF4011 domain. FIG. **1**J shows diagrams of exemplary identified defense system comprising an Hsp90 ATPase and SF2-family helicase. FIG. **1**K shows diagrams of exemplary identified defense system comprising trypsin-STAND. FIG. **1**L shows diagrams of exemplary identified defense system comprising DUF4297-STAND and another protein. FIG. **1**M shows diagrams of another exemplary identified defense system comprising DUF4297-STAND. FIG. **1**N shows diagrams of exemplary identified defense system comprising a STAND ATPase. FIG. **1**O shows diagrams of another exemplary identified defense system comprising Mrr-STAND. FIG. **1**P shows diagrams of exemplary identified defense system comprising VWA, phosphatase, and kinase. FIG. **1**Q shows diagrams of exemplary identified defense system comprising SIR2 and a DUF4020 domain. FIG. **1**R shows diagrams of exemplary identified defense system comprising SIR2. FIG. **1**S shows diagrams of exemplary identified defense system comprising SIR2-STAND. FIG. **1**T shows diagrams of exemplary identified defense system comprising PHP-ATPase. FIG. **1**U shows diagrams of exemplary identified defense system comprising SIR2 and HerA. FIG. **1**V shows diagrams of exemplary identified defense system comprising DUF1887. FIG. **1**W shows diagrams of exemplary identified defense system comprising a CRISPR-associated enzyme and an ATPase. FIG. **1**X shows diagrams of exemplary identified defense system comprising reverse transcriptase and a protease. FIG. **1**Y shows figure legends used in FIGS. **1**A-**1**X.

[0033] FIG. **2** shows diagrams of exemplary identified defense system comprising reverse transcriptase and amidase.

[0034] FIG. **3** shows diagrams of exemplary identified defense systems that comprise reverse transcriptase.

[0035] FIG. **4** shows an exemplary method of identifying defense systems.

[0036] FIG. **5** shows another exemplary method of identifying defense systems.

[0037] FIGS. **6**A-**6**B show the examples of the identified bacterial defense systems, their domain structures, and their effects on phage growth.

[0038] FIG. **7** shows selected identified bacterial defense systems and mutated forms, and their effects on phage growth.

[0039] FIGS. **8**A-**8**C: Domain-independent identification of novel systems that were enriched in defense islands. (FIG. **8**A) Computational pipeline to identify uncharacterized putative defense systems across all sequenced bacterial and archaeal genomes. Defense systems were identified based on de novo analysis of amino acid sequences, independent of pre-existing protein domain annotations. Histograms of defense association probabilities for (FIG. **8**B) selected known systems used as control and (FIG. **8**C) novel seed genes (minimum 50 identified homologs). Seeds to the right of the dashed line (0.15) were selected for further analysis.

[0040] FIGS. **9**A-**9**B: Experimental validation of 29 novel defense gene cassettes. (FIG. **9**A) Experimental validation pipeline using phage plaque assays on *E. coli* heterologously expressing a cloned candidate defense system. (FIG. **9**B) Anti-phage activity across a diverse panel of coliphages with dsDNA, ssDNA, and ssRNA genomes (mean of n=2 replicates). Also shown is a bar graph of the abundance of each system within sequenced bacterial and archaeal genomes. See also FIGS. **12**-**13**.

[0041] FIGS. **10**A-**10**E: RADAR employs a divergent adenosine deaminase that edits RNA in response to phage infection. (FIG. **10**A) Examples of genomic loci containing three subtypes of RADAR (standalone, Csx27-associated, and SLATT-associated). (FIG. **10**B) Mutations at putative rdrA and rdrB active sites abolish activity against phage T5. (FIG. **10**C) Representative RNAseq reads from *E. coli* expressing either RADAR or an empty vector control. (FIG. **10**D) Examples of editing sites in the host and phage RNA, with identified RNA secondary structures. (FIG. **10**E) Growth kinetics of RADAR-containing *E. coli* in comparison with an empty vector control under varying multiplicity of infection (MOI).

[0042] FIGS. **11**A-**11**C: A diversity reverse transcriptases (RTs) mediate antiviral immunity. (FIG. **11**A) Examples of genomic loci containing novel antiviral RTs. Three validated RT systems are shown (with two representative subtypes for each system). Domain architectures and component essentiality of (FIG. **11**B) non-retron RTs and (FIG. **11**C) retron-like RTs. See also FIG. **15**.

[0043] FIG. **12**: Novel defense systems with diverse domain architectures. Graphics show domains identified using HHpred, with mutations at active sites.

[0044] FIG. **13**: Representative plaques for phages T3, T7, φV-1, and φX174 (n=2 replicates) on *E. coli* strain C, corresponding to the right panel of FIG. **9**B. A total of 5×10^6 virions were deposited per spot, and images were acquired after 68 h incubation at 37° C.

[0045] FIG. **14**: Abundance of defense systems within sequenced genomes stratified by phylum. Defense system homologs were predicted using a two-step HMM-based search across all sequenced bacterial and archaeal genomes in Genbank.

[0046] FIG. **15**: Anti-phage defense activity for two RT-containing systems 28 and 29 (see also FIGS. **11**A-**11**C). Ten-fold serial dilutions of phage were spotted on a soft agar overlay containing *E. coli*. D313 is the putative conserved active site aspartate for the family A DNA polymerase PolA.

[0047] FIGS. **16**A-**16**C: Domain-independent prediction of putative antiviral defense systems. (FIG. **16**A) Computational pipeline to identify uncharacterized putative defense systems across all sequenced bacterial and archaeal genomes. Defense systems were predicted based on analysis of amino acid sequences, independent of domain annotations. (FIG. **16**B) Histograms of defense association frequencies before filtering and after neighborhood context-based filtering (minimum 50 homologs). Seeds to the right of the dashed line (0.1) were selected for further analysis. (FIG. **16**C) Pie chart of the domain diversity among predicted defense genes, based on additional analysis using HHpred against pfam domains.

[0048] FIGS. **17**A-**17**D: Candidate defense systems exhibit antiviral activity in a heterologous system. (FIG. **17**A) Experimental validation pipeline using phage plaque assays on *E. coli* heterologously expressing a cloned candidate defense system. Example plaques (FIG. **17**B) and zones of lysis (FIG. **17**C) for six candidate defense systems. (FIG. **17**D) Anti-phage activity across a panel of 12 coliphages with dsDNA, ssDNA, and ssRNA genomes (mean of n=2 replicates). The bar graph shows the abundance of each system within sequenced bacterial and archaeal genomes. Domains: MTase: methyltransferase; RT: reverse transcriptase; TIR: Toll/interleukin-1 receptor homology domain; TOPRIM: topoisomerase-primase domain; QueC: 7-cyano-7-deazaguanine synthase-like domain; SIR2: sirtuin; S/T phos: serine/threonine protein phosphatase; membrane: transmembrane helix; DUF: domain of unknown function. Proposed gene names (underlined): DRT: defense-associated reverse transcriptase; RADAR: phage restriction by ADAR; AVAST: antiviral ATPase/NTPase of the STAND superfamily; drs: defense-associated sirtuin; tmn: transmembrane NTPase; qat: QueC-like associated with ATPase and TatD DNAase; hhe: HEPN, helicase, and Vsr endonuclease; mza: MutL, Z1, and AIPR; upx: uncharacterized (P)D-(D/E)-XK defense protein; ppl: polymerase/histidinol phosphatase-like.

[0049] FIGS. **18**A-**18**F: RADAR mediates RNA editing in response to phage infection. (FIG. **18**A) Examples of genomic loci containing three subtypes of RADAR (standalone, Csx27-associated, and SLATT-associated). (FIG. **18**B) Essentiality of the core RADAR genes rdrAB and the accessory gene rdrD against phages T2 and T5. (FIG. **18**C) Representative RNAseq reads from *E. coli* expressing either RADAR or an empty vector control. (FIG. **18**D) Expression of phage T2 RNA relative to total host RNA in *E. coli* containing RADAR. Each dot represents a phage gene. Cells were infected at a multiplicity of infection (MOI) of 2. The p value was determined by a Wilcoxon signed-rank test. (FIG. **18**E) Representative editing sites in the host and phage transcriptomes, with corresponding predicted RNA secondary structures. (FIG. **18**F) Growth kinetics of RADAR-containing *E. coli* in comparison with an empty vector control under varying MOI by phage T2.

[0050] FIGS. **19**A-**19**E: Diverse families of reverse transcriptases (RTs) mediate antiviral defense. (FIG. **19**A) Examples of genomic loci containing two validated RT systems (DRT type 1 and type 3), with two representative subtypes shown for each system. (FIG. **19**B) Essential components of non-retron RTs (left panel) and retrons (right panel). (FIG. **19**C) Effect of defense RTs on the expression of phage T2 genes in *E. coli* infected at an MOI of 2. (FIG. **19**D) RNAseq reads mapping to the DRT type 3 system.

(FIG. **19**E) Predicted secondary structure of the highly expressed non-coding RNA identified in (FIG. **19**D).

[0051] FIG. **20**: Domain architectures and mutational analysis of additional defense systems. Graphics show domains identified using HHpred, and stars indicate locations of active site mutations. Bar graphs (n=4 replicates per bar) show either $\log_{10}$ fold change of efficiency of plating (for phages T2, P1, and λ) or $\log_2$ fold change in the area of the zone of lysis (for phages T7 and φV-1) relative to the empty vector control. MBL: metallo β-lactamase; SIR2: sirtuin; HerA: helicase; QueC: 7-cyano-7-deazaguanine synthase-like domain; TatD: DNAse; vWA: von Willebrand factor type A; PHP: polymerase/histidinol phosphatase; MTase: methyltransferase; PLD: phospholipase D.

[0052] FIGS. **21**A-**21**C: Selection of filtering thresholds for prediction of putative defense genes. Contour density plots for predicted (FIG. **21**A) toxin-antitoxin/abi genes, (FIG. **21**B) mobilome genes, and (FIG. **21**C) CRISPR-Cas genes. Boxes indicated the parameter thresholds selected for filtering putative defense genes.

[0053] FIG. **22**: Summary of tested homologs of candidate defense systems, stratified by source organism (Enterobacteriaceae vs. non-Enterobacteriaceae). Systems 1-29 correspond to the numbering in FIG. **17**D.

[0054] FIG. **23**: Representative zones of lysis for phages T3, T7, V-1, and X174 on *E. coli* strain C (n=2 replicates each), corresponding to the right panel of FIG. **2**D. A total of 5×106 virions were deposited per spot.

[0055] FIG. **24**: Abundance of validated defense systems within sequenced genomes, stratified by phylum. Defense system homologs were predicted using a two-step HMM-based search across all bacterial and archaeal genomes in Genbank (see Methods).

[0056] FIGS. **25**A-**25**B: Domain and locus architecture of the RADAR deaminase. (FIG. **25**A) Unrooted neighbor-joining tree of RdrB homologs with the Jukes-Cantor genetic distance model. Distinct clades of RADAR incorporate accessory membrane proteins RdrC (Csx27) or RdrD (SLATT). (FIG. **25**B) RdrB contains a split deaminase domain (red) with uncharacterized insertions. Domain boundaries were predicted using HHpred. Percent identity was calculated from a multiple sequence alignment of 535 representative homologs with at most 98% pairwise similarity.

[0057] FIGS. **26**A-**26**B: Deamination by the RADAR system occurs only on adenosines within RNA and requires both RADAR genes. (FIG. **26**A) Empirical probability mass functions of editing frequency for each of the 12 possible RNA base changes, calculated using the highest-expressed mRNAs in the transcriptome of *E. coli* K-12 (ATCC25404) expressing the RADAR system from *Citrobacter rodentium* DBS100. Cells were harvested 1 hr after infection by phage T2 at an MOI of 2. (FIG. **26**B) Editing frequency at a selected site within the transfer messenger RNA (tmRNA) locus (RNA or DNA). Sequences below the graphs show representative reads.

[0058] FIG. **27**: RADAR preferentially deaminates adenosines within loop regions of RNA stem-loops. Predicted RNA secondary structures of the 48 highest-expressed strong RADAR editing sites (50% editing).

[0059] FIGS. **28**A-**28**F: Effect of expression of specific phage genes on RNA editing by RADAR. (FIG. **28**A) Phage genes were cloned after IPTG-inducible T7 promoter and transformed into *E. coli* heterologously expressing the RADAR system from *Citrobacter rodentium* DBS100. (FIG. **28**B) Structure of *E. coli* transfer messenger RNA (tmRNA) (PDBID: 6Q9A), highlighting adenosines strongly edited by RADAR. (FIG. **28**C) Scatter plots of RNA editing frequencies for two replicates. Each dot represents a different phage fragment. (FIG. **28**D) Locations of fragments on the phage T2 genome. Each colored box represents a distinct fragment. (FIG. **28**E) RNA editing frequencies of the fragments shown in (FIG. **28**D) at A93 and A121 of the *E. coli* tmRNA. (FIG. **28**F) RNA editing frequencies induced by expression of RADAR with individual genes within six of the highest-activity fragments identified in (FIG. **28**D). Purple squares indicate active site mutants created by site-directed mutagenesis. dam=DNA adenine methyltransferase; a-gt: DNA alpha glucosyltransferase; gp50: head completion protein; gp2: DNA end protector protein; frd: dihydrofolate reductase; rnh: RNase H; dsbA: dsDNA binding protein; denA: endonuclease II.

[0060] FIGS. **29**A-**29**C: Mutational analysis of three RT-containing defense systems. Active site mutations abolish defense activity against phage T5 for the (FIG. **29**A) RT (UG2), (FIG. **29**B) RT (UG15), and (FIG. **29**C) retron+ATPase+HNH (Ec78) systems. The ATPase and HNH proteins in Ec78 comprise the Septu defense system.

[0061] FIGS. **30**A-**30**C: The nitrilase domain of the RT (UG1) defense system forms a distinct Glade among nitrilase enzymes. (FIG. **30**A) Stacked histogram of E-values of sequence-profile matches (RPSBLAST) between prokaryotic proteins in Genbank against a custom position-specific scoring matrix for the RT (UG1) nitrilase domain (minimum 20% coverage). Proteins matching a known nitrilase PSSM from the CDD database (E-value -10^{-6}; minimum 40% coverage) are shown in green. (FIG. **30**B) Unrooted neighbor-joining tree of the reverse transcriptase (RT) domain in nitrilase-associated RTs (n=588). Colors indicate distinct clades (cutoff tree distance 0.15). (FIG. **30**C) Unrooted neighbor-joining tree of the nitrilase domain in proteins in (FIG. **30**B) with the same color scheme (based on RT domain Glade). Also included in the tree are the non-RT-associated nitrilases (green) that are most similar to the nitrilase domain in RT (UG1) among all prokaryotic proteins.

[0062] FIG. **31**: Effect of mutations in the multi-copy single-stranded DNA (msDNA) hairpin on defense activity for the Ec86 retron from *E. coli* BL21.

[0063] FIGS. **32**A-**32**B: Bacterial densities over time for (FIG. **32**A) retron-TIR, RT-nitrilase (UG1), and RT (UG3)+RT (UG8) defense systems infected with phage T2 and (FIG. **32**B) additional defense systems infected with phage T7.

[0064] FIGS. **33**A-**33**C: Phage and prophage association frequencies for validated defense system clusters. (FIG. **33**A) Overall association frequency for 28 defense systems in this study. The rexA immunity gene from phage lambda is shown in red. (FIG. **33**B) Per-system analysis of the distribution of phage association frequencies for each associated cluster in (FIG. **33**A). (FIG. **33**C) Example of the transmembrane ATPase located within an incomplete prophage.

[0065] The figures herein are for illustrative purposes only and are not necessarily drawn to scale.

DETAILED DESCRIPTION OF THE EXAMPLE EMBODIMENTS

General Definitions

[0066] Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure pertains. Definitions of common terms and techniques in molecular biology may be found in Molecular Cloning: A Laboratory Manual, 2^{nd} edition (1989) (Sambrook, Fritsch, and Maniatis); Molecular Cloning: A Laboratory Manual, 4^{th} edition (2012) (Green and Sambrook); Current Protocols in Molecular Biology (1987) (F. M. Ausubel et al. eds.); the series Methods in Enzymology (Academic Press, Inc.): PCR 2: A Practical Approach (1995) (M. J. MacPherson, B. D. Hames, and G. R. Taylor eds.): Antibodies, A Laboratory Manual (1988) (Harlow and Lane, eds.): Antibodies A Laboratory Manual, 2^{nd} edition 2013 (E. A. Greenfield ed.); Animal Cell Culture (1987) (R. I. Freshney, ed.); Benjamin Lewin, Genes IX, published by Jones and Bartlet, 2008 (ISBN 0763752223); Kendrew et al. (eds.), The Encyclopedia of Molecular Biology, published by Blackwell Science Ltd., 1994 (ISBN 0632021829); Robert A. Meyers (ed.), Molecular Biology and Biotechnology: a Comprehensive Desk Reference, published by VCH Publishers, Inc., 1995 (ISBN 9780471185710); Singleton et al., Dictionary of Microbiology and Molecular Biology 2^{nd} ed., J. Wiley & Sons (New York, N.Y. 1994), March, Advanced Organic Chemistry Reactions, Mechanisms and Structure 4th ed., John Wiley & Sons (New York, N.Y. 1992); and Marten H. Hofker and Jan van Deursen, Transgenic Mouse Methods and Protocols, 2^{nd} edition (2011).

[0067] As used herein, the singular forms “a”, “an”, and “the” include both singular and plural referents unless the context clearly dictates otherwise.

[0068] The term “optional” or “optionally” means that the subsequent described event, circumstance or substituent may or may not occur, and that the description includes instances where the event or circumstance occurs and instances where it does not.

[0069] The recitation of numerical ranges by endpoints includes all numbers and fractions subsumed within the respective ranges, as well as the recited endpoints.

[0070] The term “about” in relation to a reference numerical value and its grammatical equivalents as used herein can include the numerical value itself and a range of values plus or minus 10% from that numerical value. For example, the amount “about 10” includes 10 and any amounts from 9 to 11. For example, the term “about” in relation to a reference numerical value can also include a range of values plus or minus 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, or 1% from that value. As used herein, a “biological sample” may contain whole cells and/or live cells and/or cell debris. The biological sample may contain (or be derived from) a “bodily fluid”. The present invention encompasses embodiments wherein the bodily fluid is selected from amniotic fluid, aqueous humour, vitreous humour, bile, blood serum, breast milk, cerebrospinal fluid, cerumen (earwax), chyle, chyme, endolymph, perilymph, exudates, feces, female ejaculate, gastric acid, gastric juice, lymph, mucus (including nasal drainage and phlegm), pericardial fluid, peritoneal fluid, pleural fluid, pus, rheum, saliva, sebum (skin oil), semen, sputum, synovial fluid, sweat, tears, urine, vaginal secretion, vomit and mixtures of one or more thereof. Biological samples include cell cultures, bodily fluids, cell cultures from bodily fluids. Bodily fluids may be obtained from a mammal organism, for example by puncture, or other collecting or sampling procedures.

[0071] The terms “subject,” “individual,” and “patient” are used interchangeably herein to refer to a vertebrate, preferably a mammal, more preferably a human. Mammals include, but are not limited to, murines, simians, humans, farm animals, sport animals, and pets. Tissues, cells and their progeny of a biological entity obtained in vivo or cultured in vitro are also encompassed.

[0072] The term “exemplary” is used herein to mean serving as an example, instance, or illustration. Any aspect or design described herein as “exemplary” is not necessarily to be construed as preferred or advantageous over other aspects or designs. Rather, use of the word exemplary is intended to present concepts in a concrete fashion.

[0073] As used herein, when an enzyme is mentioned, the term also includes a functional domain of the enzyme. For example, a reverse transcriptase may refer to a reverse transcriptase protein or a reverse transcriptase domain.

[0074] A protein or nucleic acid derived from a species means that the protein or nucleic acid has a sequence identical to an endogenous protein or nucleic acid or a portion thereof in the species. The protein or nucleic acid derived from the species may be directly obtained from an organism of the species (e.g., by isolation), or may be produced, e.g., by recombination production or chemical synthesis.

[0075] Various embodiments are described hereinafter. It should be noted that the specific embodiments are not intended as an exhaustive description or as a limitation to the broader aspects discussed herein. One aspect described in conjunction with a particular embodiment is not necessarily limited to that embodiment and can be practiced with any other embodiment(s). Reference throughout this specification to “one embodiment”, “an embodiment,” “an example embodiment,” means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present invention. Thus, appearances of the phrases “in one embodiment,” “in an embodiment,” or “an example embodiment” in various places throughout this specification are not necessarily all referring to the same embodiment, but may. Furthermore, the particular features, structures or characteristics may be combined in any suitable manner, as would be apparent to a person skilled in the art from this disclosure, in one or more embodiments. Furthermore, while some embodiments described herein include some but not other features included in other embodiments, combinations of features of different embodiments are meant to be within the scope of the invention. For example, in the appended claims, any of the claimed embodiments can be used in any combination.

[0076] All publications, published patent documents, and patent applications cited herein are hereby incorporated by reference to the same extent as though each individual publication, published patent document, or patent application was specifically and individually indicated as being incorporated by reference.

Overview

[0077] The present disclosure provides various types of bacterial defense systems and the methods of identifying thereof. In some aspects, the present disclosure includes a number of newly identified defense systems. In some embodiments, the systems may be engineered, e.g., to have a desired activity or function. The engineered systems may be used as tools (e.g., to manipulate expression and/or activity of target genes or proteins) in biotechnology and medical applications. In one example, the system comprises an ATPase and an adenosine deaminase. Such system may be engineered to function as a base editor for gene editing applications. In another example, the system comprises one or more reverse transcriptases. In another example, the system comprises a retron or one or more molecules encoded by the retron. In another example, the system comprises an NTPase of a STAND (signal transduction ATPases with numerous associated domains) superfamily.

[0078] In another aspect, the present disclosure includes methods of identifying novel defense systems. In general, the methods are based on the fact that defense systems are often clustered in bacterial genomes. In some embodiments, the methods comprise identifying genes of known defense systems in a plurality of genomes of a bacterial species, identifying homolog genes close (e.g., within 10 kb) of the known defense systems, and selecting candidate genes among these homologs. For example, candidate genes may be selected when at least 10% of homologs of the genes are within 5000 nucleotides or 5 genes from one or more defense systems.

Defense Systems

[0079] In one aspect, the present disclosure provides defense systems in prokaryotes such as bacteria. The defense systems may include proteins and nucleic acids that play roles in the defense of virus and other foreign organisms' attack and invasion. The present disclosure also includes nucleic acids encoding the components of the defense systems and vectors comprising such nucleic acids. The functions and applications of the defense systems herein are not limited to defending bacteria from foreign organisms (e.g., virus). Rather the defense systems may be used in various applications, e.g., as research tools and reagents, therapeutic agents, and diagnostic agents. In some cases, a defense system may be engineered to have a desired function. Such engineered defense system may not have a function related to defending bacteria from foreign organisms.

[0080] The defense systems provided herein may be of various types. These defense systems may comprise one or more enzymes that can manipulate (e.g., cleave, eliminate, degrade, etc.) the proteins and nucleic acids from the foreign organisms. In some examples, a host cell with the defense system may be resistant to foreign organism attacks. The term "resistance" to, for example, foreign nucleic acid invasion, encompasses a decrease in activity (e.g. phage genomic replication, phage lysogeny, circularization of phage genome) in bacteria expressing a functional defense system in comparison to bacteria of the same species under the same developmental stage (e.g. culture state) which does not express a functional defense system. According to specific embodiments the decrease provided by such resistance to foreign organism invasion is at least 1.5-fold, at least 2-fold, at least 3-fold, at least 5-fold, at least 10-fold, or at least 20-fold as compared to same in the absence of the functional defense system.

[0081] In some embodiments, the defense systems have an anti-phage activity. The term "anti-phage activity" or "resistant to infection by at least one phage" may encompasses an activity providing increased resistance of a host cell to infection by at least one phage in comparison to the host cell of the same species under the same developmental stage (e.g. culture state) which does not express the functional defense system. In some embodiments, a host cell may comprise a microbial cell. In some embodiments, a host comprises a bacterium. Anti-phage activity or resistance of a host cell to infection by at least one phage may be determined by, for example but not limited to, bacterial viability, phage lysogeny, phage genomic replication or phage genomic degradation, or a combination thereof.

[0082] In some embodiments, the defense systems may provide a host cell with resistance to foreign nucleic acid invasion. In some embodiments, a defense system described herein, provides the host cell with resistance to a foreign nucleic acid invasion, wherein the foreign nucleic acid invasion comprises resistance to at least one phage infection, or resistance to plasmid transformation, or a combination of resistance to at least one phage infection and resistance to plasmid transformation. In some embodiments, it is the combination of defense systems that provides a host cell with resistance to a foreign nucleic acid invasion. One skilled in the art would appreciate that defense against a foreign nucleic acid invasion may encompass, defending against entry of a foreign nucleic acid into the host cell, as well as, defending against the actions of a foreign nucleic acid that has entered the host cell. In some embodiments, defense against a foreign nucleic acid invasion comprises defense from phage infection. In some embodiments, defense against a foreign nucleic acid invasion comprises defense from plasmid transformation. In some embodiments, defense against a foreign nucleic acid invasion comprises defense against entry of a conjugative element. In some embodiments, defense against a foreign nucleic acid invasion comprises defense against any combination of phage infection, plasmid transformation, and entry of a conjugative element.

[0083] In some embodiments, the components in the system may be heterologous, i.e., they do not naturally occur together in the same cell or an organism.

[0084] The components in a system herein may be derived from the same or different prokaryotes. In some cases, the components may be engineered to be optimized for expressing in eukaryotic (e.g., mammalian) cells.

Gene Clusters

[0085] In some embodiments, the components of a defense system may be in a gene cluster in a prokaryotic cell. The terms "gene cluster", "cassette of genes", "cassette", and "components of a system", may in some embodiments herein be used interchangeably having all the same meanings and qualities. In some embodiments, each gene of a "cassette of genes" comprises a nucleic acid sequence encoding a polypeptide component of the defense system. In some embodiments, a "cassette of genes" comprises nucleic acid sequences encoding components of the defense system including open reading frames encoding defense system polypeptide components, regulatory sequences, and non-coding RNAs. A skilled artisan would appreciate that a "cassette of genes" may encompass an operon. In some embodiments, a cassette of genes comprises regulatory sequences. In some embodiments, a cassette of gene comprises non-coding RNAs.

Host Cells

[0086] The defense systems may be from or originate from microorganisms such as bacteria or archaea. In some embodiments, the defense may be from or originate from bacteria. As used herein, when a defense system originates form a species, it may be the wild type defense system in the species, or a homolog of the wild type defense system in the species. The defense system that is a homolog of the wild type defense system in the species may comprise one or more variations (e.g., mutations, truncations, etc.) of the wild type defense system. The terms "ortholog" and "homolog" are well known in the art. By means of further guidance, a "homolog" of a protein as used herein is a protein of the same species which performs the same or a similar function as the protein it is a homolog of. Homologous proteins may but need not be structurally related, or are only partially structurally related. An "ortholog" of a protein as used herein is a protein of a different species which performs the same or a similar function as the protein it is an ortholog of. Orthologous proteins may but need not be structurally related, or are only partially structurally related. Homologs and orthologs may be identified by homology modelling (see, e.g., Greer, Science vol. 228 (1985) 1055, and Blundell et al. Eur J Biochem vol 172 (1988), 513) or "structural BLAST" (Dey F, Cliff Zhang Q, Petrey D, Honig B. Toward a "structural BLAST": using structural relationships to infer function. Protein Sci. 2013 April; 22(4):359-66. doi: 10.1002/pro.2225.). See also Shmakov et al. (2015) for application in the field of CRISPR-Cas loci. Homologous proteins may but need not be structurally related, or are only partially structurally related.

[0087] In some example, the host cells are *E coli*. In some embodiments, the bacteria may be gram positive bacteria. The term "Gram-positive bacteria" as used herein refers to bacteria characterized by having as part of their cell wall structure peptidoglycan as well as polysaccharides and/or teichoic acids and are characterized by their blue-violet color reaction in the Gram-staining procedure. Representative Gram-positive bacteria include: *Actinomyces* spp., *Bacillus anthracis, Bifidobacterium* spp., *Clostridium botulinum, Clostridium perfringens, Clostridium* spp., *Clostridium tetani, Corynebacterium diphtherias, Corynebacterium jeikeium, Enterococcus faecalis, Enterococcus faecium, Erysipelothrix rhusiopathiae, Eubacterium* spp., *Gardnerella vaginalis, Gemella morbillorum, Leuconostoc* spp., *Mycobacterium abcessus, Mycobacterium avium complex, Mycobacterium chelonae, Mycobacterium fortuitum, Mycobacterium haemophilium, Mycobacterium kansasii, Mycobacterium leprae, Mycobacterium marinum, Mycobacterium scrofulaceum, Mycobacterium smegmatis, Mycobacterium terrae, Mycobacterium tuberculosis, Mycobacterium ulcerans, Nocardia* spp., *Peptococcus niger, Peptostreptococcus* spp., *Proprionibacterium* spp., *Staphylococcus aureus, Staphylococcus auricularis, Staphylococcus capitis, Staphylococcus cohnii, Staphylococcus epidermidis, Staphylococcus haemolyticus, Staphylococcus hominis, Staphylococcus lugdanensis, Staphylococcus saccharolyticus, Staphylococcus saprophyticus, Staphylococcus schleiferi, Staphylococcus similans, Staphylococcus warneri, Staphylococcus xylosus, Streptococcus agalactiae* (group B *streptococcus*), *Streptococcus anginosus, Streptococcus bovis, Streptococcus canis, Streptococcus equi, Streptococcus milleri, Streptococcus mitior, Streptococcus mutans, Streptococcus pneumoniae, Streptococcus pyogenes* (group A *streptococcus*), *Streptococcus salivarius*, and *Streptococcus sanguis.*

[0088] In some embodiments, the term "Gram-negative bacteria" as used herein refer to bacteria characterized by the presence of a double membrane surrounding each bacterial cell. Representative Gram-negative bacteria include *Acinetobacter calcoaceticus, Actinobacillus actinomycetemcomitans, Aeromonas hydrophila, Alcaligenes xylosoxidans, Bacteroides, Bacteroides fragilis, Bartonella bacilliformis, Bordetella* spp., *Borrelia burgdorferi, Branhamella catarrhalis, Brucella* spp., *Campylobacter* spp., *Chalmydia pneumoniae, Chlamydia psittaci, Chlamydia trachomatis*, to *Chromobacterium violaceum, Citrobacter* spp., *Eikenella corrodens, Enterobacter aerogenes, Escherichia coli, Flavobacterium meningosepticum, Fusobacterium* spp., *Haemophilus influenzae, Haemophilus* spp., *Helicobacter pylori, Klebsiella* spp., *Legionella* spp., *Leptospira* spp., *Moraxella catarrhalis, Morganella morganii, Mycoplasma pneumoniae, Neisseria gonorrhoeae, Neisseria meningitidis, Pasteurella multocida, Plesiomonas shigelloides, Prevotella* spp., *Proteus* spp., *Providencia rettgeri, Pseudomonas aeruginosa, Pseudomonas* spp., *Rickettsia prowazekii, Rickettsia rickettsii, Rochalimaea* spp., *Salmonella* spp., *Salmonella typhi, Serratia marcescens, Shigella* spp., *Treponema carateum, Treponema pallidum, Treponema pallidum endemicum, Treponema pertenue, Veillonella* spp., *Vibrio cholerae, Vibrio vulnificus, Yersinia enterocolitica*, and *Yersinia pestis.*

Examples of Systems

[0089] A system provided herein may include one or more enzymes or functional protein domains, and/or polynucleotides encoding thereof. The systems may comprise one or more wild type proteins and/or polynucleotides. In certain cases, the systems may be engineered systems, e.g., comprising one or more mutations or variants compared to corresponding wild type counterparts.

[0090] In some embodiments, the systems herein may be configured to modify a nucleic acid, e.g., DNA, RNA, or a hybrid or duplex of RNA and DNA. In one example, the systems may be configured to modify RNA.

[0091] The systems and components thereof may be or share sequence homology (e.g., sequence identity) with the example systems and components herein. In some embodiments, the systems or components thereof may share at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 99%, or 100% sequence homology (e.g., sequence identity) with the example systems or components herein.

Systems Comprising ATPase and Adenosine Deaminase

[0092] In some examples, the systems comprise an ATPase and an adenosine deaminase. The ATPase may be a KAP-family ATPase. In some cases, the ATPase may comprise 1500 or less, e.g., 1400 or less, 1300 or less, 1200 or less, 1100 or less, 1000 or less, 950 or less, 900 or less, 850 or less, 800 or less, 750 or less, 700 or less, 650 or less, 600 or less, 500 or less, 400 or less, 300 or less, 200 or less, 100 or less amino acid residues. In one example, the ATPase may comprise 1000 or less amino acid residues. In certain examples, the ATPase may comprise 900 or less amino acid residues. In some cases, the adenosine deaminase may comprise 1500 or less, e.g., 1400 or less, 1300 or less, 1200 or less, 1100 or less, 1000 or less, 950 or less, 900 or less, 850 or less, 800 or less, 750 or less, 700 or less, 650 or less, 600 or less, 500 or less, 400 or less, 300 or less, 200 or less, 100 or less amino acid residues. In one example, the adenosine deaminase may comprise 1000 or less amino acid residues. In certain examples, the adenosine deaminase may comprise 900 or less amino acid residues.

[0093] In some examples, the system comprises an ATPase that is or share at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 99%, or 100% sequence homology (e.g., sequence identity) with the sequence of WP_012906049.1 and a adenosine deaminase that is or share at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 99%, or 100% sequence homology (e.g., sequence identity) with the sequence of WP_012906048.1. In some examples, the system comprises an ATPase that is or share at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 99%, or 100% sequence homology (e.g., sequence identity) with the sequence of WP_155731552.1 and a adenosine deaminase that is or share at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 99%, or 100% sequence homology (e.g., sequence identity) with the sequence of WP_064360593.1.

[0094] In some embodiments, the system comprising ATPase and an adenosine deaminase may further comprise one or more proteins or polypeptide domains. In some examples, the system may further comprise a membrane protein or domain. In certain examples, the system further comprises a SMODS and LOG-Smf/DprA-Associating Two TM (SLATT) domain. In certain examples, the system further comprises a CRISPR ancillary protein. The type VI-B CRISPR ancillary protein, e.g., Csx27.

[0095] In some embodiments, the systems may be engineered to function as a base editor in gene editing applications. For example, the systems may modify a nucleic acid. The modification may cause an A to G mutation in a nucleic acid. In some cases, the systems may modify RNA. In some cases, the systems may modify DNA.

[0096] In some embodiments, the adenosine deaminase may be those described in International Patent Publication Nos. WO2019071048, WO2019084063, WO2019126716, WO2019126709, WO2019126762, and WO2019126774; Cox DBT, et al., RNA editing with CRISPR-Cas13, Science. 2017 Nov. 24; 358(6366):1019-1027; Abudayyeh 00, et al., A cytosine deaminase for programmable single-base RNA editing, Science 26 Jul. 2019: Vol. 365, Issue 6451, pp. 382-386; Gaudelli N M et al., Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage, Nature volume 551, pages 464-471 (23 Nov. 2017); Komor A C, et al., Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature. 2016 May 19; 533(7603):420-4, or any variants, homologs, or orthologs thereof.

[0097] In some embodiments, the system further comprise one or more phage proteins. Examples of phage proteins include those in Tables 18A-18B.

Systems Comprising Reverse Transcriptase(s)

[0098] In some examples, the systems herein comprise one or more reverse transcriptases. A reverse transcriptase refers to an enzyme capable of synthesizing DNA strand (e.g., complementary DNA or cDNA) using RNA as a template. In some embodiments, the reverse transcriptase is error prone. For example, the reverse transcriptase may have low proof-reading ability. For example, the reverse transcriptase may introduce one or more errors (i.e., nucleotides that are not complementary to the corresponding nucleotides on the template). Examples of reverse transcriptases include the transcriptases from *Vibrio harveyi* ML phage, *Bifidobacterium longum, Bacteroides thetaiotaonicron, Treponema denticola, cyanobacteria*, such as *Trichodesmium* erythrism, the genus *Nostoc*, or *Nostoc punctiforme.*

[0099] As used herein, the reverse transcriptase may be full-length reverse transcriptase or a functional fragment thereof. A functional fragment of a full-length reverse transcriptase may be a polypeptide that is shorter than the full-length reverse transcriptase but has reverse transcriptase activity. For example, a functional fragment of a full-length reverse transcriptase may have at least about 50%, at least about 60%, at least about 70, % at least about 80%, at least about 90%, at least about 95%, at least about 99%, or at least about 100% of the activity of the corresponding reverse transcriptase. The reverse transcriptase activity may be measured as amount of cDNA generated with certain amount of RNA template.

[0100] For example, the systems may comprise a first reverse transcriptase and a second reverse transcriptase. The first and the second reverse transcriptases may be comprised in the same protein. The first and the second reverse transcriptase may be the same. In certain cases, the first and the second reverse transcriptase may be the different. The reverse transcriptase may be error prone.

[0101] Examples of reverse transcriptases include UG1, UG2, UG3, UG8, UG15, or UG16 reverse transcriptases. In some examples, the system comprises an UG1 reverse transcriptase that is or share at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 99%, or 100% sequence homology (e.g., sequence identity) with the sequence of WP_115196278.1. In some examples, the system comprises an U2 reverse transcriptase that is or share at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 99%, or 100% sequence homology (e.g., sequence identity) with the sequence of WP_012737279.1. In some examples, the system comprises an UG3 reverse transcriptase that is or share at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 99%, or 100% sequence homology (e.g., sequence identity) with the sequence of 087902017.1 and an U8 reverse transcriptase that is or share at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 99%, or 100% sequence homology (e.g., sequence identity) with the sequence of WP_062891751.1. In some examples, the system comprises an UG15 reverse transcriptase that is or share at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 99%, or 100% sequence homology (e.g., sequence identity) with the sequence of GCK53192.1. In some examples, the system comprises an UG16 reverse transcriptase that is or share at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 99%, or 100% sequence homology (e.g., sequence identity) with the sequence of WP_001524904.1.

[0102] In some examples, the systems comprising one or more reverse transcriptases may further comprise one or more proteins or polypeptide domains. In some examples, the systems further comprise a Cas protein, e.g., Cas1. In some examples, the systems further comprise Abi. In some examples, the systems further comprise a nitrilase-family C—N hydrolase. In some examples, the systems further comprise a DNA polymerase. The DNA polymerase may be a family A DNA polymerase. In some examples, the systems further comprise a nitrilase. In some examples, the systems comprise a protein comprising one or more reverse transcriptases and a nitrilase domain. The nitrilase domain may be at the C-terminus of the protein. In some examples, the systems further comprise a topoisomerase-primase (TOPRIM), and a nitrilase. In some examples, the systems further comprise a Tol/interleukin 1 receptor (TIR). In some examples, the systems further comprise a protease. The systems may further comprise a serine protease domain linked to or associated with the reverse transcriptase. In some examples, the systems further comprise an integrase. In some examples, the systems further comprise a transposase. In some examples, the systems further comprise an MBL domain.

[0103] In some cases, the system may comprise a polynucleotide encoding the reverse transcriptase. In certain examples, the polynucleotide comprising the variable region and/or the template region may comprise a coding sequence for the reverse transcriptase. In some examples, the polynucleotide encoding the reverse transcriptase may be different from the polynucleotide comprising the variable region and/or the template region.

[0104] In some embodiments, the reverse transcriptase comprises an active site, e.g., (Y/F)×DD (SEQ ID NOs: 1-2), where X is any amino acid.

Systems Comprising Retrons or Molecules Encoded by Retrons

[0105] In some examples, the systems herein comprise one or more retrons or molecules encoded by retrons. As used herein, a retron refers to a genetic element (e.g., a DNA molecule) which encodes components enabling the synthesis of branched RNA-linked single stranded DNA (msDNA) and a reverse transcriptase. Molecules encoded by retrons includes retron msr RNA that is the non-coding RNA produced by retron elements and is the immediate precursor to the synthesis of msDNA. Molecules encoded by retrons also include the reverse transcriptase and the corresponding RNA (e.g., mRNA).

[0106] In some examples, the retron is Ec67 retron. In some examples, the retron is Ec86 retron. In some examples, the retron is Ec78 retron. In some examples, the retron is TIR domain-associated retron. The TIR domain may have NAD+ hydrolase activity. In some examples, the retron is TOPRIM domain-associated retron. The TOPRIM domain may have nuclease activity.

Systems Comprising STAND NTPase

[0107] In some examples, the systems herein comprise one or more NTPases of a STAND (signal transduction ATPases with numerous associated domains) superfamily. In some examples, the systems comprising the NTPase may further comprise one or more proteins or polypeptide domains, such as DUF4297, Mrr-like nuclease, SIR2, a trypsin-like serine protease, and/or a helical domain.

Additional Examples of Systems

[0108] In some examples, the system may comprise a von Willebrand factor (VWF), a PP2C-like serine/threonine protein phosphatase, and a serine/threonine kinase. In some examples, the system may comprise SIR2 or a function domain thereof.

[0109] In some examples, the system may comprise a reverse transcriptase and a nitrilase. In some examples, the system may comprise a reverse transcriptase and a nitrilase, and a topoisomerase-primase (TOPRIM). In some examples, the system may comprise a reverse transcriptase and TIR. In some examples, the system may comprise an Ec67 retron. In some examples, the system may comprise Ec86 retron. In some examples, the system may comprise a reverse transcriptase. In some examples, the system may comprise two reverse transcriptases. In some examples, the system may comprise adenosine deaminase. In some examples, the system may comprise KAP ATPase. In some examples, the system may comprise KAP TatD. In some examples, the system may comprise a transmembrane ATPase. In some examples, the system may comprise an ATPase, QueC synthase, and TatD endonuclease. In some examples, the system may comprise S8 peptidase. In some examples, the system may comprise a DFU4011 domain. In some examples, the system may comprise a DFU4011 domain, a helicase, and a Vsr endonuclease. In some examples, the system may comprise a DUF3684 Hsp90-like ATPase and a helicase. In some examples, the system may comprise Trypsin-AAA35. In some examples, the system may comprise DUF4297-AAA3 and another protein. In some examples, the system may comprise DUF4297-AAA35. In some examples, the system may comprise AAA35. In some examples, the system may comprise RE-AAA35. In some examples, the system may comprise VWA and phosphatase and a kinase. In some examples, the system may comprise SIR2-DUF4020. In some examples, the system may comprise SIR2-STAND-TPR. In some examples, the system may comprise Polymerase and Histidinol Phosphatase (PHP)-ATPase. In some examples, the system may comprise PHP-SMC. In some examples, the system may comprise SIR2 and HerA. In some examples, the system may comprise DUF4297 and HerA. In some examples, the system may comprise Unknown-DUF1887. In some examples, the system may comprise DUF262 and DUF262-HNH. In some examples, the system may comprise DUF499, DUF3780, DUF1156 methyltransferase, and helicase. In some examples, the system may comprise Type I-E CRISPR-associated protein. In some examples, the system may comprise RT-protease. In some examples, the system may comprise ApeA.

[0110] Details of these systems are shown in Tables 1, 2, 5, 6, 9, 10, 12, 13, 15A, and 16A. Sequences of example systems are shown in Tables 6, 12, 15A, 15B, 15C, 16A, and 16B.

TABLE 1

Construct	# genes in operon	Short Description	Donor Strain	Diagram File Name	Note
pLG018	1	RT-nitrilase	*Klebsiella pneumoniae* NCTC9143	pLG018_RT-nitrilase	UG1/UG6 in Zimmerly & Wang (2015)

TABLE 1-continued

Construct	# genes in operon	Short Description	Donor Strain	Diagram File Name	Note
pLG022	1	TOPRIM-RT-nitrilase	*Vogesella indigofera* DSM3303	pLG022_TOPRIM-RT-nitrilase	UG10 in Zimmerly & Wang (2015)
pLG024	1	RT-TIR	*Shigella dysenteriae* NCTC2966	—	Novel retron
pLG026	1	Ec67 retron	*Escherichia coli* NCTC8623	pLG026_RT-TOPRIM (retron	Ec67 retron (reported in Lampson et al. Science 1989; function unknown until present study)
pLG199	1	Ec86 retron	*Escherichia coli* BL21	—	Ec86 retron (reported in Lim et al. Cell 1989; function unknown until present study)
pLG028	1	RT	*Escherichia coli* 21-C8-A	pLG028_RT	
pLG125	2	RT-x2	*Escherichia coli* ECOR12	—	Two RTs acting in concert; UG3/UG8 in Zimmerly & Wang (2015)
pLG032	2	Adenosine deaminase	*Citrobacter rodentium* DBS100	pLG032_Deaminase	ATPase + highly divergent adenosine deaminase
pLG034	1	KAP ATPase	*Escherichia coli* ECOR25	pLG034_KAP-transmembrane	Large transmembrane ATPase; described computationally in Aravind et al. Genome Biol (2004)
pLG037	4	KAP_TatD	*Escherichia coli* NCTC9009	pLG037_KAP	Described computationally in Aravind et al. Genome Biol (2004)
pLG039	2	S8 peptidase	*Escherichia coli* ECOR52	pLG039_Protease	Proteasome-like ATPase + serine protease
pLG041	1	DUF4011	*Escherichia coli* ATCC43886	pLG041_DUF4011	
pLG044	2	DUF3684 Hsp90-like ATPase + helicase	*Vibrio harveyi* ATCC43516	pLG044_Hsp90	Large gene (~2500aa) with large stretches of unknown regions; associated with a helicase
pLG046	3	Trypsin-AAA35	*Erwinia piriflorinigrans* CFBP5888	pLG046_Protease-STAND	STAND ATPase (these are not typically thought to be defensive)
pLG049	2	DUF4297-AAA3 + unknown	*Salmonella enterica* NCTC13175	pLG049_DUF4297-STAND	STAND ATPase
pLG050	1	DUF4297-AAA35	*Salmonella enterica* NCTC10718	pLG050_DUF4297-STAND	STAND ATPase
pLG051	1	AAA35	*Escherichia coli* NCTC9087	pLG051_STAND	STAND ATPase
pLG053	1	RE-AAA35	*Escherichia coli* NCTC11132	pLG053_STAND	STAND ATpase
pLG056	3	VWA + phosphatase + kinase	*Escherichia coli* NCTC9094	pLG056_VWA_phophatase_kinase	
pLG061	1	SIR2-DUF4020	*Escherichia coli* NCTC9112	pLG061_SIR2-DUF4020	
pLG062	1	SIR2	Cronobacter sakazakii NCTC8155	pLG062_SIR2	
pLG063	1	SIR2-STAND-TPR	*Escherichia coli* NCTC13384	pLG063_SIR2-STAND	STAND ATpase
pLG066	1	PHP-SMC	*Escherichia coli* NCTC8620	pLG066_Phosphoesterase (PHP)-SMC	
pLG070	2	SIR2 + HerA	*Escherichia coli* NCTC11129	pLG070_HerA	Modular system (HerA pump can be paired with SIR2, DUF4297, etc.)
pLG071	2	DUF4297 + HerA	*Escherichia coli* NCTC11131	pLG070_HerA	Modular system (HerA pump can be paired with SIR2, DUF4297, etc.)
pLG080	1	Unknown-DUF1887	*Salmonella enterica* NCTC6026	pLG080_DUF1887	~1200aa gene; first ~1000aa are unknown
pLG157	2	DUF262 + DUF262-HNH	*Escherichia coli* ATCC43886	—	Described computationally in Makarova et al. 2011
pLG078	4	DUF499 + DUF3780 + DUF1156 methyltransferase + helicase	*Escherichia coli* ECOR58	—	Restriction-modification-like system described computationally in Anantharaman et al. 2013

TABLE 2

Construct	# genes in operon	Short Description	Donor Strain	Diagram File Name	Note
—	6	Type I-E CRISPR-associated	—	CRISPR_ATPase	Described computationally in Shmakov et al. PNAS 2017; predicted to be non-defense
—	1	RT-protease	—	RT-protease	Retron; described computationally in Zimmerly & Wang (2015)

[0111] FIGS. **1A-1Y**, **2**, and **3** show diagrams of domain structures of exemplary defense systems.

Additional Exemplary Systems

[0112] Additional examples of systems are shown in Tables 3A-3B below.

TABLE 3A

Row No.	Vector	System	System details	# genes	Organism	Strain	bp	Note	Source
1	pLG003	Control	BREX type I	6	*E. coli*	NCTC9078 (DSM5212)	13703	Goldfarb et al. 2014	
2	pLG004	Control	Druantia type I	5	*E. coli*	NCTC9078 (DSM5212)	11823	Doron et al. Science 2018	
3	pLG005	Control	Type I RM	3	*E. coli*	NCTC13846 (DSM105182)	6946		bloodculture, human bacteraemia, UK
4	pLG006	Control	Zorya type II	3	*E. coli*	ATCC8739	3917	Doron et al. Science 2018	Feces
5	pLG007	Control	RT-AbiA	1	*E. coli*	ECOR30 (ATCC35349)	1921	Odegrip et al. 2006	Bison, Alberta, Canada
6	pLG008	Control	RT-AbiK	1	*Lactococcus lactis*	W-1	2102	Wang et al. NAR 2011	
7	pLG009	RT	RT-protease	1	*Stenotrophomonas maltophilia*	TG_2005			
8	pLG010	RT	RT-protease	1	Haematobacter massiliensis	KC2145			
9	pLG011	RT	RT-protease	1	*Sphingobium yanoikuyae*	ATCC51230 (DSM7462)	2029		clinical specimen
10	pLG012	RT	RT-protease	1	*Proteus mirabilis*	127_PMIR	2009		
11	pLG013	RT	RT-protease	1	*Pseudomonas aeruginosa*	PA-W9			
12	pLG014	RT	RT-protease	1	*Photobacterium damselae*	NCTC11646	2657		human, leg wound
13	pLG015	RT	RT-protease	1	*Paraburkholderia silvatlantica*	PSCR-88			
14	pLG016	RT	RT-protease	1	*Bacillus subtilis*	ATCC13952	2203		
15	pLG017	RT	RT-kinase-nitrilase	1	*E. coli*	N1	4154		
16	pLG018	RT	RT-kinase-nitrilase	1	*Klebsiella pneumoniae*	NCTC9143	5272	SLATT associated	Urine
17	pLG019	RT	RT-nitrilase	1	*E. coli*	NCTC4169	3679		human, excreta
18	pLG020	RT	RT-nitrilase	1	*Klebsiella pneumoniae*	KPNIH39	3479		uterine secretion
19	pLG021	RT	TOPRIM-RT-nitrilase	1	*Pseudomonas rhizosphaerae*	DSM16299	8446		rhizosphere of grasses
20	pLG108	RT	TOPRIM-RT-nitrilase	1	*Vogesella indigofera*	DSM3303			Garden soil, Pacific Grove California
21	pLG023	RT	RT-TIR	1	*E. coli*	NCTC9024	2393		
22	pLG024	RT	RT-TIR	1	*Shigella dysenteriae*	NCTC2966	2139		monkey with enteritis
23	pLG025	RT	RT-TOPRIM	1	*E. coli*	NCTC13441	2569		
24	pLG026	RT	RT-TOPRIM	1	*E. coli*	NCTC8623	2405		gastro-enteritis
25	pLG027	RT	RT-345	1	*E. coli*	STEC 66	1951		
26	pLG028	RT	RT-345	1	*E. coli*	21-C8-A	2141		
27	pLG029	RT	RT-x2	2	*E. coli*	NCTC9091	3648		
28	pLG030	RT	RT-x2	3	*Acinetobacter calcoaceticus*	NCTC7412	4236	SLATT associated	human, urine

TABLE 3A-continued

Row No.	Vector	System	System details	# genes	Organism	Strain	bp	Note	Source
29	pLG031	ADA	Adenosine deaminase	2	*E. coli*	NCTC11116	5533		
30	pLG032	ADA	Adenosine deaminase	2	*Citrobacter rodentium*	ATCC51459	5526		Laboratory mouse
31	pLG033	ADA	Adenosine deaminase	3	*Pluralibacter gergoviae*	ATCC33028	6689	SLATT associated	Urine, France
32	pLG034	KAP	Transmembrane KAP ATPase	1	*E. coli*	ECOR25 (ATCC35344)	4415		Dog, New York
33	pLG035	KAP	Transmembrane KAP ATPase	1	*E. coli*	NCTC8620	4037		human, diarrhoea
34	pLG036	KAP	KAP + unknown + QueC + TatD	4	*E. coli*	ECOR10 (ATCC35329)	4891		Adult human, New York
35	pLG037	KAP	KAP + unknown + QueC + TatD	4	*E. coli*	NCTC9009	5408		
36	pLG038	Protease	ATPase + serine protease	2	*E. coli*	ECOR12 (ATCC35331)	3678		Adult human, Sweden
37	pLG039	Protease	ATPase + serine protease	2	*E. coli*	ECOR52 (ATCC35371)	3676		Orangutan, Seattle Zoo, Washington
38	pLG040	Protease	ATPase + serine protease	2	*E. coli*	NCTC9008	3917		pathogenic to chicks
39	pLG041	DUF4011	DUF4011-helicase-Vsr-DUF3320	1	*E. coli*	ATCC43886	5958		Feces, human
40	pLG042	DUF4011	DUF4011-helicase-Vsr-DUF3320	1	*Citrobacter braakii*	NCTC9067	6502		
41	pLG043	DUF3684	Hsp90-like ATPase + SNF2	2	*Pectobacterium wasabiae*	CFBP3304 (ATCC43316)	10581		Japanese horseradish, Eutrema wasabi, Japan
42	pLG044	DUF3684	Hsp90-like ATPase + SNF2	2	*Vibrio harveyi*	ATCC43516	10687		Mouth of shark, Bahamas
43	pLG045	DUF3684	Hsp90-DUF3684-DUF3883-PDDEXK(CTD)	1	*Raoultella planticola*	NCTC9528	5918		butter
44	pLG046	AAA35	Protease-AAA35	3	*Erwinia piriflorinigrans*	CFBP 5888 (DSM26166)	7847		necrotic pear blossoms, Valencia, Spain
45	pLG047	AAA35	Protease-AAA35	3	*Pectobacterium fontis*	M022 (LMG30744)	7740		
46	pLG048	AAA35	DUF4297-AAA35-TPR	1	*E. coli*	NCTC9036	6514		
47	pLG049	AAA35	DUF4297-AAA35	2	*Salmonella enterica*	NCTC13175	7175		
48	pLG050	AAA35	DUF4297-AAA35	1	*Salmonella enterica*	NCTC10718	6261		
49	pLG051	AAA35	Unknown-AAA35-unknown	1	*E. coli*	NCTC9087	5109		
50	pLG052	AAA35	Unknown-AAA35-unknown	1	*E. coli*	NCTC10650	4781		
51	pLG053	AAA35	RE-AAA35	1	*E. coli*	NCTC11132	4964		
52	pLG054	Kinase	DUF2357	7	*Obesumbacterium proteus*	DSM2777	12191		ale yeast
53	pLG055	Kinase	Kinase-helicase_1600aa	2	*E. coli*	NCTC13919	6873		Clinical isolate. Human, rectum
54	pLG056	Kinase	VWA + phosphatase + kinase	3	*E. coli*	NCTC9094	3605		
55	pLG057	Kinase	5-gene McrBC-like	5	*Plasticicumulans lactativorans*	DSM25287	11931		lactate-fed bioreactor inoculated with activated sludge from a sewage treatment plant, Kralingseveer, Rotterdam, Netherlands

TABLE 3A-continued

Row No.	Vector	System	System details	# genes	Organism	Strain	bp	Note	Source
56	pLG058	GTPase	GTPase	3	*Pantoea cypripedii*	LMG 2657 (DSM3873)	4789		*cypripedium* orchid, California
57	pLG059	GTPase	GTPase	3	*Pectobacterium wasabiae*	CFBP3304 (ATCC43316)	5216		Japanese horseradish, Eutrema wasabi, Japan
58	pLG060	GTPase	GTPase	3	*E. coli*	NCTC10962	4577		faeces(arabian gulf)
59	pLG061	SIR2	SIR2-DUF4020	1	*E. coli*	NCTC9112	4212		
60	pLG062	SIR2	SIR2-TPR-HEAT	1	*Cronobacter sakazakii*	NCTC8155	4329		tin of dried milk
61	pLG063	SIR2	SIR2-AAA35	1	*E. coli*	NCTC13384 (ATCC11229)	3411		
62	pLG064	Misc	Dcm + unknown + unknown + HerA + Vsr	5	*Pseudomonas aeruginosa*	NCTC10727	11911		
63	pLG065	Misc	Dcm + unknown + unknown + HerA + Vsr	5	*Aquimonas voraii*	DSM16957	11635		water, Assam, India
64	pLG066	Misc	Phosphoesterase (PHP)-SMC	1	*E. coli*	NCTC8620	3066		human, diarrhoea
65	pLG067	Misc	Helicase-nuclease_unknown	2	*E. coli*	NCTC9033	7356		
66	pLG068	Misc	DUF3893 (possible pAgo)	3	*Pseudomonas syringae*	DSM10604	6714		common lilac
67	pLG069	Misc	RecQ	1	*Klebsiella oxytoca*	NCTC11696	5424		
68	pLG070	Misc	SIR2 + HerA	2	*E. coli*	NCTC11129	3308		
69	pLG071	Misc	DUF4297 + HerA	2	*E. coli*	NCTC11131	3419		
70	pLG072	Misc	Dcm + Hsp90-sensor histidine kinase + response regulator	4	*E. coli*	NCTC86 (DSM301)	7655		
71	pLG073	Misc	Dcm + Hsp90-sensor histidine kinase + response regulator	4	*E. coli*	NCTC11560	6042		
72	pLG074	Misc	Palatin + nucleotidyltransferase + UBCc/ThiF + ubiquitin-like	4	*Klebsiella aerogenes*	NCTC9735	4755		
73	pLG075	Misc	Sensor histidine kinase + phosphoribosyltransferase	2	*Pseudomonas aeruginosa*	NCTC13717	4088		
74	pLG076	Misc	PH-TerB-DUF726 (transmembrane) + Nup (transmembrane)	2	*Klebsiella pneumoniae*	NCTC11357	3637		
75	pLG077	Misc	TerB-DUF2791-Lhr	3	*E. coli*	NCTC9024	6037	Identified in Doron et al. Science 2018	
76	pLG078	Misc	DUF499 + DUF1156	3	*E. coli*	ECOR58 (ATCC35377)	9809	Identified in Anantharaman et al. Biology Direct 2013, 8: 15	Lion, Seattle Zoo, Washington
77	pLG079	Kinase	5-gene McrBC-like	5	*Yoonia sediminilitoris*	DSM29955	11425		tidal flat sediment, South Korea
78	pLG080	Misc	DUF1887 CTD; no other domains	1	*Salmonella enterica*	NCTC6026	4100		

TABLE 3B

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
1	pLG003	acagcaccacgttcatcttccttttttaactgattttacagagactttaatacagttaaaattttatttcctgagctgtaatcgattaagttgatgcatttaatgggaatgatatagggtcatttccagtctcacttatagaaatggctaaagcatgactctcgccaaaaccgtttatgtgttgtacataacgcgatcatccctctcacaaattgccttttctcatggcatctcgcccggtccccattacaatcactttttgtttttttgcgagctgcattccagtcttcagagggtttttcgatgattaaaaatgacaaggcatggataggagacttgctgggcggaccgctcatgagcagggaaagccgcgtcattgccgaactgttgctaaccgatcccgatgaacagacatggcaagagcaaattgttggccacaacattttacaagcctcttctcctaacaccgcaaaacgttacgcggcaacaatcaggcttcgcctgaacacgctggataaaagcgcgtggacattgattgccgaaggtagtgaacgggaacgccaacaacttctgtttgtggctctgatgctacattcgccggtagttaaggatttctggctgaagtggtgaacgatctgcgcaggcagttcaaggaaaagttgcctggcaatagctggaacgaatttgtgaatagccaggttcgcctacatccggtactcgccagctactcagattcatctattgcaaaaatgggaaacaatctggtgaaggcgcttgctgaagcgggttatgtggatacgccccgcagacgtaacctgcaggcagtttaccttttaccggaaactcaggcagtgttacagcgcctgggacaacaggacttgatatctattctggagggaaaacggtgatagatcccgttcttgaatatcgcctgtctcaaatccagagtcgcattaacgaagatcgcttcctcaaaaataacggctccggaaatgaaattggtttttggatctttgattatcccgcgcagtgcgaactgcaggtacgggagcatttgaaatatctgctccggcatctggaaaaggaccataaatttgcctgtctgaatgtcttccaaatcatcatcgatatgctcaatgaacgcggccttttcgagcgcgtctgccagcaggaagtcaaagtgggtactgagacgctgaaaaagcagcttgctggtccgttaaatcagaaaaagatcgctgattttatagcgaaaaaagtcgatctggctgcccaggattttgtcattcttaccggcatgggcaacgcctggccattagtacgcggtcatgaactgatgagtgccttgcaggatgtcatggggttcaccccactgctgatgttttatcctggcacctacagcgggtacaacctttccccgctcacagacaccggttcacaaattattatcgcgctttcagactggtaccagatacgggacccgcagcaacattgaatcctcaatgaagagcataacaatgaatattgaacagattttgaaaaacctctaaaacgaaatataaacggggtagtcaaagcagagcaaaccgatgatgccagcgcgtacatcgagttagatgaatatgtcatcacccgcgaactggaaaaccatcttcgccatttcttcgaatcctatgttcctgccactggcccggaacggatccgtatggaaaacaagatcggcgtatgggtttcaggcttcttcggttcaggtaaatcgcactttattaagattctttcttatcttttatctaaccgcaaagttacacataacggtacggaacgtaatgcttactccttctttgaagataaaatcaaagatgcattattccttgccgatattaacaaagcggtgcattacccgactgaagtcattctgttcaatattgattcgcgtgccaacgtagatgacaaagaagatgccattcttaaagtcttcctgaaagttttcaacgaacgcattggatactgcgctgattttccgcatattgcccatcttgagcgcgagctggataaacgcggtcagtatgaaacctttaaagccgcgtttgccgatatcaatggctcgcgctgggaagacgagcgcgacgcttactacttcatcagcgatgacatggcacaagcattaagccaggccacgcagcagagtcttgaatcctcccgccaatgggtggaacaactcgacaaaaacttcccgctggatatcaataatttttgccagtgggtaaaagagtggctggatgacaatggtaagaacatcctctttatggtggatgaagtcggtcagttcattggcaaaaatacgcaaatgatgctgaagctgcagactattactgaaaaccttggggtaatttgcggtggccgcgcatgggttatcgtgacttcgcaggccgatatcaacgcggcaatcggtggtatgagcagtcgcgacggacaggacttctccaagatccaggggcgcttctctacacgcctgcaactttccagctctaacacatcagaagttatccagaaacgtttgttggtaaagactgacgaagcaaaagcggcactggcaaaagtgtggcaagagaaagccgatatcctgcgtaaccagctggcttttgacactacaacaactactgcactacgtccttttaccagcgaagaagagttcgttgacaactacccgtttgtcccgtggcactatcagattctgcaaaaagtgtttgaatctattcggacgaaaggtgcagcgggtaaacaattggccatgggtgagcgttctcagctggaggcattccagacggcggcgcagcaaatctcagcgcaagggctggattctctggtgcctttctggcgcttctatgccgccattgagagcttcctggaacctgccgttagccgcaccatcactcaggcttgccagaatggcattcttgatgagttcgatggcaacctgcttaaaacgctgttcctgatccgctatgtggaaacgctgaaaagcaccctggataacctggtcacattgtctatcgataggatcgatgccgataaagttgagttgcgccgccgggtcgaaaaaagtctcaacacgcttgaacgcctgatgctcattgcgcgcgttgaagataaatatgtgttcctgaccaacgaagagaaagagatcgaaaacgagatccgtaacgttgatgtcgatttctctgcgatcaacaaaaaactggcatcgatcatctttgatgacattctgaaaagccgtaaatatcgttatccggctaacaagcaagactttgatatcagccgcttcctgaacgggcatccattagacggcgcagtgcttaacgatctggtggtgaagatcctgaccccctaaagatccgacttattcgttctataacagcgatgcgacctgtcgcccttatacgtcagaaggcgacggctgtatttgattcgtctgcccgaagagggccgtacctggagcgatattgatttagtcgtccagactgaaaagttcctcaaagataacgccgggcaacgtccggaacaggcaaccctgctctcagaaaaagcgcgtgaaaacagcaaccgggaaaaattactccgtgttcagttggaatcactacttgcagaagcagacgtctgggcgattggcgaacgcttaccgaaaaaatcctccacgccatcgaacattgtcgatgaagcctgccgttacgtgattgaaaacaccttcggcaagctgaagatgctgcggccttttaacggtgacatctcccgtgaaattcatgcattactgacggttgagaacgacaccgaactggatctcggtaacctcgaagagtccaaccccgacgccatgcgcgaggtagaaacctggatcagcatgaatatcgaatacaataaacctgtgtatttacgcgatattctgaaccattttgcgcgtcgcccttatggctggcccgaagacgaagtgaaactgctagtagcccgtctggcctgcaaaggtaaattcagcttcagccagcaaaacaacaacgtcgagcgaaaacaggcgtgggagttatttaataacagccgccgccatagcgaattgcgtctgcataaagttcgccgtcatgatgaagcgcaggtgcgtaaagccgcgcaaaccatggctgacatcgctcagcagccgtttaacgaacgggaagagccggcgctggttgaacatattcgtcaggtatttgaagagtggaagcaagagctgaacgtattccgcgccaaggcagagggcggaaacaatccgggggaaaaacgagattgaatccggtctgcgcctgcttaatgccattcttaatgagaaagaagattttgccctgatcgaaaaagtctcatcgctgaaagatgaacttctggatttcagcgaagaccgtgaagatttggtcgacttctaccgtaagcaattcgccacctggcaaaaactgggtgctgcgctgaatggcagctttaaatctaaccgcagcgcgctggaaaaagacgccgcagcggttaaagcgctgggcgagctggaaagcatctggcaaatgccggaaccttataagcatctcaatcgcatcacgccgttgattgaacaggtccagaacgtcaaccatcagttagtcgaacagcatcgccagcacgccctcgaacgcattgacgcccgcattgaggaaagccgtcaacgcttgctggaagcgcacgccacgtcggagctgcaaaacagcgttctgctgccgatgcaaaaagccagaaaacgcgctgaagtcagccagtcgattccggaaattttggcggaacagcaagagacaaaagcgctgcaaatggatgcagataaaaagattaacctgtggatcgacgagctgcgtaaaaagcaagaagcacaactccgggcagcaaatgaagctaaacgcgctgccgactcagaacagacttatgttgtggtggaaaaaaccgttatccaaccggtaccgaaaaaaacgcatctggtgaatgtcgccagtgagatgcgtaatgccaccggtggtgaagttctggaaacgaccgaacaggtggaaaaggcgctcgacacgttacgcacaacgctgctggccgtcattaaagcaggcgatcgcattcgccttcagtaactcccatttcagggcagcactctgctgccctttgcaggattttctatgaataccaataacattaaaaaatatgccccacaggcccgtaacgacttccgcgatgcggtgatccagaagctaacgacgcttgggatcgctgcagataaaaaggcaatttgcagattgccgaggccgaaaccattggcgagaccgtgcgttacggtcagtttgattacccgttatcgacccttccccgccgcgaacggctggtaaaacgcgcccgtgagcagggttttgaggtgctggttgagcactgcgcctacacctggtttaaccgcttatgtgcaattcgctatatggagctacacggttatcttgagcacggcttccgtatgttgtcccacccggagacgccgaccgcgtttgaggtgctggatcatgtgccggaagtggcagaagccctgctgccggaaaataaggcgcagctggttgaaatgaagctttccggtaatcaggacgaagccctgtaccgcgaactgctgctggggcagtgccacgccctgcaccacgcgatgccgttcctgtttgaagcggtagatgacgaagcggaactgctgttgccggataacctgacccgtaccgactctattctgcgtgggctggttgatgatattccggaagaagactgggagcaggtagaggttatcggctggctgtatcagttctatatttcggaaaagaaagatgccgtgattggcaaagtggtgaagagcgaagatattcctgccgccacccagctgtttacgccaaactggattgtgcagtatctggtacaaaactccgttggccgccagtggttgcagacctacccggactcgccgctgaaagacaaaatggagtactacatcgagcctgcggaacaaacgccggaagtgcaggcgcagctggcggcgattaccccagccagcattgaacccgaaagtattaaagtgctcgacccagcctgcggctccggtcatattttgattgaagcctataatgtgctgaaaaa

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		tatctacgaagagcgtggttatcgcgggcgtgatattccacaactgattctggaaaataatatttttggtcttgatatcgacgacc gcgcggcacagctttccggctttgcattattaatgatggcgcgtcaggatgaccgcagaatatttacccgcgatgtacgtctgaat attgtctctttgcaggaaagcctgcatctggatatcgccaaactctggcagcaactgaatttccaccagcaggtacaaaccggcag tatgggggatatgtttgctgaaaataacgcgttaacccaaactgacagcgcagaatatcagctgctgatgcgcacgctgaaacgct ttgtgaatgcaaaaacgctgggctcactgattcaggtgccgcaggaagaagaagcggaactgaaggtattcctggacgcgttgtat cgcctggaacaggaaggcgatttccagcagaagacggcggcaaaagcgtttattccgtttattcagcaggcgtggatttagcgca gcgatatgatgcggtagtggcgaatccgccgtatatggggggtaattatatggagacagaacttaagaatttcgtctcttcttact accctcaaggaaaggcggatctttattcttcatttatggtcagattacttttacaattaaaagataatcgcactttaagcctaatg acccccttacttggatgaatttatcatcatttgaagagctccgaaaaattatacttacaaatttcagcattcagtcattagtaca gcctgaatatcattcatttttttgagtcagcttatgtcccaatttgtgcttttagcatttcaaataccccattaagctggaatgcaa aattttttgatttatcagatttttatggagaaaaaaatcaagctccaaattttcagtatgcaattaaaaatgacaataaatgtcat tggaaatataacagaatcaccacggactttctatgtactcccggatatatcattgcttactctctgcctgattctgcgttatcttg cttcaaaacatccaaaaaacttcatgatgtttgcaatctaaaacaaggattaattactggtgataatgaaagatacctaagattct ggcatgaaatcagctataactctttcagtctcaatgaaaaaagaaaaaaaacaaaatggttcccatatcaaaaaggtggtgcatac cgtaaatggtatggtaataatgattatgttgttgactgggagaatgatggttattccattaaaaacttttataatgacaaaggtaa attacgctcacgccctcaaaacatacaatttttattgtaaagagggtttaacatggacaagtttaactatttcgtcactatcgatga gatatgtaccaaatggatatatttttgatgcaaaaggacctatgtgttttccgaaatcctctttggatatctggaatattcttggc tatgcgaatagcaaagtaatagatatatttctcaaacaattagcgcccaccatggattattctcaagggcctgttggaaatgtccc attcaaatttaacgatggtgatttgaacgagataataaaagaactcgtaaacattcacaaacgtgactgggatgaaaatgaaacat cttttgagtttaagagagatatgttggttcatttttcaagagatattaacactattaagggtagttttacactaaggcaaggggaa aataaaaaagcgattaacagaacaaaatttttagaagaaatgaataactctttctttataaattgctttaatctaactgatatttt atctccagaaattgaactaaacaaaatcacgttaacgcatgcaactattgaaattgatattcaaaaaataatttcatatgcaatag gctgccaaatgggacgttactcccttgatcgcgaaggtctggtatacgctcatgaaggcaataatggcttcgccgatcttgtcgcc gaaggtgcttataaaagcttcccggctgatagtgacggcattctgccgctaatggatgaagagtggtttgacgatgacgtcacctc tcgcgtcaaggagtttatccgcaccgtttggggcgaagaatatttgcgcgaaaacctcgattttatagccgaagttctcaagccca aaaaaggcgaatctgcgctggagaccattcgtcgctatctttccacccagttctggaaagatcatctgaaaatgtataaaaagcgt ccaatctactggctattcagctccggtaaagagaaagcgtttgagtgcttggtgtatctgcatcgctataacgatgccacgctgtc gagaatgcgtaccgaatatgtggtgccgctgctggcgcgttatcaggccaatattgatcgcctgaacgatcaacttgatgaggctt ctggcggtgaatccacacgtctgaaacgcgaacgcgacagcctgatcaaaaaattcagcgaactgcgcagctatgacgatcgcctg cgtcactatgctgatatgagaatcagtattgatctcgacgatggcgttaaggttaactacggcaagtttggcgatctgctggcaga tgtcaaagccatcaccggcaatgccccagaggtgatctaaaccagacggcacgttctcctgttgccgggttctgcccggtggcaaa taccaccgggaaacgcgccgctgctgacatttctccacctcacttcatgataaaatgcgccaccgtgtcaaaatctccttttcgcg ttttggcgctttcttattcatcgtaacaacatgggattgtgaacttgcaaaatcaggactttattgctggccttaaagctaaattt gccgaacatcgcatcgttttctggcacgatcccgataaacgttttattgaggaactggaacagctcaagcttgaaagcgtcacgct aatcaacatgacccacgagtcacagctggcggtaaaaaaacgcatcgagattgatgagccagaacagcagttcctgctgtggttcc cccatgatgcgccgcctcatgaacaagactggctgctggatatccgcctttacagcagcgaattccatgccgatttgccgccatc accctgaacacgctgggcattccccagcttggcctgcgcgagcatattcagcgacgcaaggccttcttcagcactaaacgcacgca ggcgctgaaaaatctggcgacagaacaggaagatgaagcctcgctggataagaaaatgattgcggtgatcgctggcgcaaagaccg cgaaaaccgaagacatttgttcaacctgattacccagtacgttaaccaacaaatagaagacgacagcgaactggaaaacacgcag gcgatgctgaaacgccacggtctggactcggtattgtgggaaatgctcaaccacgaaatgggctaccaggcagaggagccatcgct ggaaaacctgctcctgaaactgwtgtaccgatctctctgcccaggccgacccacagcagcgcgcctggctggaaaaaaatgtcctg ctgacgccatccggcagagcatctgccctggcatttatggtgacctggcgtgccgatcgtcgctataaagaggcttatgactactg cgctcagcaaatgcaggccgccctgcacccggaagatcattaccgactcagctcgccgtatgatttgcacgaatgcgaaaccaccc tcagcatcgaacaaaccattattcatgcgctggtaacacagctgctggaagagagcaccacgctcgatcgggaagcctttaaaaaa ctgctctctgagcgccagagcaaatactggtgtcagacacaaccagagtattacgccatctatgacgcattgcgccaggctgagcg gttgctgaacctgcgcaatcgccacatcgatggtttccactaccaggacagcgccaccttctggaaagcctactgcgaagaactgt tccgcttcgaccaggcttatcgcctgtttaatgaatatgccttgctggttcacagcaaaggagcgatgatcctcaagagcctggat gattatatcgaggcgctctacagcaactggtatctggcagagttaagccgtaactggaacgaagtgctggaagcggaaaatcgtat gcaggcgtggcaaatccctggcgtgccgcgtcagcagaacttcttcaatgaggtggtgaagccacagttccaaaatccgcaaatca aacgcgtgttcgtgataatttccgatgccctgcgttatgaagtggcggaggagctggggaatcaaatcaataccgagaaacgcttt accgcagaactgcgctcgcagctcggcgtgctccccagctacacccaactgggaatggcggcattgctgccccatgaacaactttg ctatcaacccggtaacggcgacatcgtttatgctgatgggctgtcgacctcgggtattcctaaccgcgataccattctgaagaact ataagggaatggcgataaaatcgaaggaccttctggagttaaaaaatcaggaagggcgagaccttattcgcgattacgaagtggtg tatatctggcataacacgattgatgccactggcgacacggcatccacggaagataaaaccttcgaagcgtgccgcacggcggtggc tgaactgaaagatttagtcaccaaggtgatcaaccgcctccacggcacacgcatttttgttacggcggatcacggtttcctgttcc agcaacaggcgctttcggttcaggataaaaccactctgcaaattaagccggaaaacaccatcaagaaccacaaacgctttattatc ggccatcagcttcccgccgatgattttgctggaaagggaaagtggcggataccgcaggcgtgagcgacaacagcgagttcctgat tccgaaagggatccagcgcttccatttctctggcggcgcgcgcttcgttcatggcggcaccatgttgcaggaggtttgcgttccgg tattgcagataaaagccctgcaaaaaccgccgcagaaaaacagccacagcgccgcccggtggatattgtcgcttaccatccgatg attaagctagtgaacaatatcgataaagtgagcctgttgcagacgcatccggtgggcgaactttatgaaccgcgtatcctgaacat ttacattgtcgacaacgccaacaatgtggtctcgggcaaagagcgcatcagctttgacagtgataacaacaccatggaaaaacgcg tacgcgaagttacgctgaagctgattggcgctaacttcaaccgtcgcaatgagtactggttgatactggaagacgcacaaacggaa acggggtatcagaagtacccggtcattatcgatctggcgttccaggatgatttcttctaagtgaggcgatatgcaaacccatcatg atttacctgtttcaggcgtatccgcaggggaaattgcctccgagggttacgatctggacgccctgctgaaccagcattttgctggt cgcgtggtgcgtaaagatctcaccaagcaactcaaggaaggggcaaacgtcccggtgtatgtgcttgagtatctgctcggcatgta ctgcgcctctgacgatgacgatgtggtcgagcaagggttgcaaaacgttaagcgtattctggctgataactatgtgcgcccggatg aagcggagaaagtgaagtcgctgatccgcgagcgtggttcgtacaaaatcatcgataaagtgtcggtgaaactgaaccagaaaaaa gacgtttacgaagcccagctttctaacctcggcatcaaagacgcgctggtgccctcgcagatggttaaagacaacgagaagctact gacgggcggtatctggtgcatgattaccgtcaactatttctttgaagaagggcagaagacctcacccttctcattgatgacgctca agcctatccagatgccgaatatggatatggaagaggtgttcgatgcgcgtaaacactttaaccgtgaccagtggatcgatgtgctg ctgcgctcggtgggtatggagcccgccaatattgagcaacgcaccaaatggcaccttatcacccgtatgatcccgttcgtggagaa caactataacgtttgcgagctggggccgcgtggcaccggtaaaagccatgtgtataaagagtgttctcctaactccctgttagttt

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		ccggcgggcaaacgaccgttgccaacttgttctacaacatggccagtcgccagatcggcctggttggcatgtgggatgtggtagcgttcgacgaagtcgcggggatcactttcaaagataaagacggcgtgcaaatcatgaaagattacatggcgtcaggatctttctctcgcggcagagattcgattgaaggtaaagcgtcgatggttttcgtcggcaacatcaatcaaagcgtagagactctcgttaaaaccagccatttgctggcaccatttccgactgcgatgattgatacagcatttttcgaccgctttcatgcctatattcccggttgggaaatccccaaaatgcgcccggaattctttaccaaccgttacgggctgattacggattatctcgctgaatatatgcgcgaaatgcgcaaacgcagtttctctgatgcgattgataaattctttaagctgggtaacaacctcaaccagcgtgacgttattgccgttcgacgtaccgtgtcggggttgttaaaactcatgcatcccgatggcgcgtacagcaaagaagatgtgcgagtctgcctgacctatgcgatggaagttcgccgccgcgtgaaagagcaacttaaaaaactgggcggtctggagttcttcgatgtgaactttagctacatcgacaacgaaacgctggaagagtttttgtgagcgtaccggaacagggcggcagcgaacttattcctgccggaatgccaaagccgggtgttgtgcatctggtcactcaggcagaaagcggcatgaccgggctgtatcgttttgaaacacagatgactgccggtaatggtaagcatagtgtatcgggtctgggttcaaatacctccgcgaaagaagctatccgcgtcggtttcgattacttcaaaggcaatttgaatcgggtaagcgcggccgcgaaattctccgatcatgaatatcaccttcatgtcgttgaactgcataatactggcccaagcaccgcaaccagtcttgctgcgcttatcgctttatgttcgatattgctggcaaaaccggtgcaggaacagatggtggtgttgggcagtatgacgcttggtggggtaattaacccggtgcaggatcttgccgccagtttgcagctcgccttcgacagcggtgcaaaacgggttctgttgccgatgtcctcggctatggatattccaacggttccggcagagttatttaccaagtttcaggtgagtttttactcagacccggttgatgctgtttataaggcgctgggtgtgaattaacgtagtaactattttaatgaac (SEQ ID NO: 3)
2	pLG004	ggtgaacgtttggttgatagggtagtaaaactagtaatcatcctataattagctatattcgtggttattagattgaaaacagataacattaacaaaatctataaatcgatttgaatgatttttttcatcaatactgttgtaagctcctgctatcaaaagttttgcacacaatctataagctcccagaattgcttgtataaatgctatcattggcgctgtcccgatcgagggagcaaggaggggactctcttgtgccatgcgattaatcactggggctctaagtgaaatttagtgggactaaatactaattggaacgtgagataaaaatgcacaaatatccctctataatagttaatatcaaccttcgagaagccaaactgaaaaagaaggtacgtgagcatttacaatccttgggttttacaagatctgattctggagcgctccaggcccgggaaataccaaagatgtaatacgggctcttcatagttctcaacgagctgagcggatatttgcaaaccaaaagttcataacgctaagagcggcaaagcttattaaatttttcgcatccggcaatgaggtcattccggataagatttcaccggtacttgaacgtgtaaagtcaggaacctggcaaggagatctctttaggttagcagcattaacttggtccgtacctgtttcaagcggatttggaaggcgtctccggtatcttgtatgggatgaaagcaacggaaaattgatagggctgatcgcaattggtgaccctgtgttcaaccttgcagtccgagataatttgattgggtgggatactcatgccagaagttcccggcttgttaatttgatggatgcatacgtcctcggtgctcttccccttataatgccctgctgggaggaaaattaattgcatgtctgcttcgtagccgcgatctttatgatgactttgcaaaggtctatggtgataccgttggagtaatatctcaaaaaaagaaacaagcacgtcttttggctattacaacaacatcgtctatggggcgctcatcggtatataaccgtttaaagctggatggaattcaatatttaaaatcgattggatatacaggcggttgggggcattttcatatacctgatagcttgttcattgaattacgtgattacttacgtgatatggatcacgcttatgcagatcattatatgtttggtaatgggcctaactggcgtttacgtacaactaaggcagctttaaatgcactaggatttagagataatttgatgaagcatggaattcaacgtgaagtgtttatcagtcagctagcagaaaatgcaactagtattctgcaaacaggcaaaggtgaaccagatctaacctctttgctttctgctaaagagatagctgagtgtgcgatggcacgatggatggttccacgatcaattcgcaatccagaatatcggctttggaaagcaagagatctatttgattttattagtaatgactcgctaaactttcccccgtttgacgagatagcgaaaacagttgtctaatcttaactgaaggggggagtaagtgaattacgctattgataagttcaccgggacactgatattagcagctcgagcaacgaaatatgctcaatatgtttgcccagtttgtaaaaaaggtgttaacctccgtaaagggaaggttataccccccatatttgctcatttgcccggacatggtacgtcagactgtgaaaatttgttcccggaaattctatcattgtcgaaactattaaaactatttcaaagcgatatatggatttgcgcttattgattcctgtcggaagtaatagtcgagagtggtcattagaattagtgttgccaacctgtaatttatgtagagcaaagataacgttagatgtaggaggcagaagccaaacgcttgatatgaggagtatggtaaagagtcgccagattggtgctgaattatcagtaaaatcttaccgtattgtttcatatagtggtgaaccagatccaaaatttgtaacagaagttgaaagagaatgcccaggtttaccttctgagggagcagcagttttcactgctttagggcgtggggcatcgaagggatttccacgagcacaagagttaagatgtactgaaacatttgcctttctttggcgacaccctgttgctccagatttcctgatgaattagaaataaaaagtttagctagtaaacagggatggaatttagctcttgttacaattcctgaagtcccttctgtggagagtatttcatggctaaaatctttacataccttcctgttgttcctgccagaacatctattacagcaatttggccgttcctaaatcaaaaaacaagtattaatcatgtcgaatgtgtttattctgacacaatattgttgtcaacaaatatggcaccaacatcatcagaaaatgttggaccaactatgtacgcacaaggttcctctttattactttcagcggttggtgttgaaacatcacctgctttcttcattctaaatcctggagaaaatgactttgtgggcgtttctggctcaattgagcaggacgtaaacttatttttttctttctataaaaaaaacgtttctgtacccagaaaatatccctcaatagatttggtttttactaagaggaataaagaaaagaccatcgtttccttacatcaaagaagatgcattgaagttatgatggaagcacgaatgtttggccataaattagaatacatgtctatgccttctggtgttgaaggagtggcaagaattcaaagacaaactgaaagtaatgttattaagttagtttctaatgatgacattgcagctcatgataagagcatgcggttactatctcctgttgcgttatctcaattatctgattgcttagcaaacttaacatgtcatgtagaaatagatttttttaggtcttggtaaaatattttttacctggttcttctatgctatcattagatgacgggaaatttattgaattatctcctaatcttcgctcacggatattaagttttatacttcaaatggggcacaccctccatggttttagtttaaataatgatttttttattagttgagaaattagtggatttgcagccggaaccacacttattaccgcattatagagcattggtaaaagaagttaagaccaatggatttgaatgtaaccgctttagataaggtgccttcgaatgagttaccaatatagccaagaggcaaaggaacggatctctaagttgggacaatccgaaattgttaactttatcaatgagatttctccaactttacgacgtaaagcttttggttgtttaccaaaagtaccgggattcagggcaggacatcccactgaaattaaagaaaaacagaaaagattgattgggtatatgttccagtcacatccttcctctgaggagagaaaagcatggaaaagtttttctctttttttggcagttttgggctgaagagaaaattgacaaatcatttagtatgattgataatttaggattaaaagaaaactctggctctatttttattagagagcttgctaaaaactttcctaaagttgctagagagaatatcgagcgcctgtttatctttagtgggtttgctgatgatccagacgttataaatgcatttaacctttttcctcctgcagttgttcttgcccgcgatatcgtgattgatactcttccaattcgtttagatgagcttgaagcacgtattagtttaattgccgataatgttgagaaaaaaaataatcatattaaagaacttgagttaaaaatagatgcttttttccgaacagtttgataattactttaataatgaaaagagcagtttaaaaataattaatgaactacaatctttgataaactcagagactaaacaatctgatattgctaataaagctattgacgagctttatcattttaatgaaaaaaacaaacagctaatattatctcttcaagaaaaattagattttaatgctctggctatgaatgatatttctgagcatgaaaaattgataaaaagtatggctaatgacatttcagaatttaaaaatgcattaacgatcttgtgtgataataaaataaagaataacgagttagattatgtcaatgaattaaaaaaactcactgaacgaatagatacacttgaaataaacacatctcaagctagcgaagtgagtgtcaccaatagatttacaaaattccatgaaatagcgcactatgaaaattatgagtatctttcatcctccgaagacatatctaatagaatttctttaaatttacaggctgttggattgacaaaaaattcagcagaaaaattggctagattgacattagctaccttcgtttctggacaaatcattcaattcagtggctctttggcagatattatcgcggatgcaattgccattgctattggtgcaccacgttatcacatatggagagttccagttggtattatttctgacatggatgcttttgattttatagagactatagctgaatcatctcgctgtctcctttttgaaaggggccaatctttcagcatttgagatttatggagcggcaattagagatatagttgttcaacggcaaatacatccaacaaattatgaccatctggcattgatagctacctggaaacaaggcccagctacattccctgatggaggaatgttggccgagttgggacctgt

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		tattgatactgatacattaaaaatgcgtggtttatcagctactttaccccaattgaaaccaggttgtcttgccaaggataaatggacaaatattgatggactacatcttgatagtgttgatgattatgtagatgaattaagagcattactggacgaagctggatttgatgggggaactttgtggaagagaatgattcatattttctatacttcactcataaggatccctaatggaaattatatttatgatctttattctgtcttgtctttttatactcttacatgggcaaaaattaaaggtggccccgtccaaaagatagaagatattgccaatcgtgaattaaaaaattatagtgcaaaaatatcttcttgaggaggtggttaatggagtggagagcagtatcacgagacaaagcactggatatgttatcaactgcattaaattgtcgatttgatgatgaagggttgagaatttcagcagtttcagaatgcttaaggagcgtattatatcaatattctatatctgaaacagaagaagctaggcaaactgtaacctcgcttcgactcactagtgcagtaaggcgaaaattggtacctttatggccagacattgctgatattgataatgctatacatccgggcattatgtctatattgaacagcttggctgaattgggtgacatgattaagttagaaggtggtaattggctaacagctcccccacatgcagtacgaattgacaataagatggctgtttttttttggtggagagccttcctgtacattttcaacgggcgtggtagctaaatctgctggaagagttcgcttggttgaagaaaaagtgtgtactggaagtgttgaaatctgggatgcaaatgagtggattggtgccccagcagaaggcaatgaagaatggtcatccagactactatctggaactatttccggctttatcgatgcacctggcaatatgagtgaaacgactgcatatgtgcggggaaaatggctccatttgtcagaactttcttttaataaaaagcaaatctacttatgcagaatgtccgttgataatcacttttcctattatttaggagaaattgaagctggacgcttatgtagaatgaattcgttagaatcgtctgatgatgtcagaagattacgtttttttctcgatacaaaagataattgtccgctaaaggtccgtatcaaaatatctaatgggctagcaagattaagattaaccagaagattaccaagacgagaaacgaaggtactcctgctaggctggagagaatcaggttttgaaaatgaacattcaggaataacacaccatgtattccccgaggaaatattacccatagtgcgtagcgcttttgaagggcttggtattatttggattaacgaattcacgcgacggaatgaaatatgattaataaaaataaagtaactgaacgttcaggtatacatgataccgtgaaaagccttagtgaaaatctgagaaaatacattgaggcacaatatcatatccgggatgaagggttaattgctgagcgacgagcgcttttacagcaaaatgaaactattgctcaagctccttatatagaagcaaccccaatttatgaacctggtgcgccatacagtgaattgcctattcccgaagcagcaagtaatgtgctaactcaactatcagaacttggaattggcctctatcaacgccctataaacaccaatcacaggcacttgagtcatttcttggcgaaaacgcttctgatctggtcattgcaacaggtacaggctccggtaagactgaaagctttctaatgccaattattggaaaattggcgattgaatcttccgagagacctaaatctgcatcccttccaggttgtagagcaattttattatatccaatgaatgcattagttaacgatcaacttgctcgtatcagacgtctttttggtgattctgaagcctctaaaatactgagatctggaagatgtgcccctgtacgctttggcgcttatacgggaagaacgccttaccctggtcgtcgtagctctagacgagacgagctttttatcaaacccctttcgatgagttttacaataaactcgcaaataacgcccccgtacgtgcggaactgaaccgcattggtcgctggccaagtaaagatcttgatgctttttatgggcaaagcgcatctcaggctaaaacctacgtctcaggcaaaaaaacgggtaagcaatttgttttgaacaattgggggagaggctaattacccagcctgaggatcgtgagctaatgacccggcatgaaatacagaatcgctgtccagaattactgataacgaactactccatgcttgagtatatgctgatgcgacctatcgagcgtaatatttttgagcagactaaggaatggctcaaagctgatgagatgaatgagcttatcttagtgcttgatgaagcgcatatgtatagaggagcagggggagcagaggtagcccttttaatacgtcgcctctgtgctcggttggatattccccgggaacgtatgcgctgcatccttaccagtgctagtctagggtccattgaggatggagaacgttttgcccaagacttaactggcttatcaccaacctcttcgaggaaatttcgaattattgagggtacaagggaatcgcgtcctgagtcacaaattgttaccagtaaagaagctaatgcactggctgaattcgacctaaattcatttcagtgcgtagctgaggatcttgaatctgcatatgcagcaatagagtctcttgccgaacgaatgggctggcaaaagccgatgataaaagatcatagtacactacgtaattggttatttgataatttgactggttttggtcctattgaaacgcttattgaaatagtttcaggtaaagcggttaagctaaatatcttgagtgaaaaccttttttccagactctccacagcaaatcgcagagcgagcaacagatgcattactcgcattgggttgctatgctcagagggcatccgatggcagagtgcttattccaactcgcatgcatctttttttatcggggattaccaggtctttatgcctgtatagatcccgattgtaatcaacgtttgggtaaccatagcgggccaactatacttggccgcctttatacgaaaccactggatcaatgtaaatgcgcttcaaaagggcgagtctacgaattatttacccaccgtgactgcggtgcggcttttattcgtggatacgttagttccgaaatggactttgtatggcaccagccgaacggaccattatcagaagatgaggatatcgatcttgttcccatagatatattggtcgaggaaacacctcatgtacatagtgattaccaggacagatggctacatatagcaacaggacgcctttctaaacagtgtcaagatgaggattctggttatcgtaaagtctttatacctgaccgagttaagtctggatcagaaattacatttgatgaatgccctgtttgtatgcgtaagacaagaagtgctcagaatgaaccgtctaaaattatggatcatgttacaaaaggggaagcaccttttacaacgttagtacgtacacagatatctcaccagccagcgagtcgtcctattgatggtaaacatcccaatggggaaaaaaagtacttattttttctgatggccgacaaaaagcagctcggcttgcacgtgatattcctagagatattgagcttgatttgtttcggcaatccattgctctcgcctgttctaaactgaaagatatcaatcgggaacccaaaccaacatcagtactttaccttgctttcctatcagtcctttctgaacatgacttgcttattttgatggggaagattcacgaaaagttgtaatggcccgtgatgaattttatcgtgattataatagcgatctggctcaagcttttgatgatagcttcagcccccaagagtcaccgtcacgatataaaatagcgttgcttaaacttttatgtagcaattactattctctttccggaacaacagttggttttgttgaaccatcgcagcttaaatcaaaaaaaatgtgggaagatgtgcagtccaagaagctcaatattgagagcaaggatgttcatgctttagctgttgcttggattgataccttactcactgaatttgcttttgatgaatctattgattcgacactacgaatcaaagcagctggattctacaaacccacttggggtagtcaaggacggtttggaaaagctcttaggaaaaccctgatacagtatcctgctatgggggagctttatgtggaagttttggaggagatttttcgtactcatctgacattaggaaaagatggtgtctactttcttgctccaaatgcactacgtctgaaaatagatctcttgcatgtctggaaacaatgtaatgactgcacggcactaatgccatttgctttagaacattctacttgccttgcttgtggtagtaacagtgtcaaaacagtcgagccgtcggaaagcagctatattaatgcacgaaaaggattctggcgttcgccggtagaagaagttttggtttcaaattcgcggcttctaaaccttagcgttgaagagcatactgctcaactctcacatagagatagggccagcgttcatgccactacagaactctacgaactgagattccaagatgttcttattaatgataacgacaagcccattgatgtacttagttgtacgacgacgatggaagtggggggttgatattggatctctggttgctgttgctttaagaaacgtccctccgcaacgagaaaattatcagcaacgtgctgggcgagcaggccgccgtggcgcatctgtttcaacggtggttacatattctcaaaatggccctcatgatagttattatttccttaatcctgaacgcattgttgcaggttctcctcgtacacctgaagtgaaagtaaataatcccaaaatagccagaagacacgttcattctttttagttcagaccttttttcacgagttaatggaacaaggaatttataatcccgcagagaaaactgccatacttgagaaagcacttggtactacacgagatttttttcatggagcaaaagatactggcctaaatctcgatagctttaataattgggttaaaaaccgtattctatctactaatggtgatttgagaacaagtgttgcagcatggcttcctcctgttcttgaaactggagggctttctgccagtgactggtttgctaaggtagcagaggaattttaaatacactccatgggctggctgaaattgttccacaaactgccgttcttgttgatgaggaaaatgaagatgatgagcagacttctggtggaatgaaatttgcacaagaagaattacttgagttcctgttttaccatggtttattaccaagttatgcatttcctacaagcctctgtagtttcttggtagaaaaaattgtaaagaatattagaggttcttttgaggtgcgaacagtacaacagcctcagcaatcaatttctcaggctctgagtgaatatgccccgggacgtttgattgttattgataggaaaacctatcgctctggtggtgtttttctaatgcattgaaaggcgaactaaaccgggcaagaaagcttttcaataatcccaaaaagtttattcattgcgataagtgctcttttgtccgcgatcctcataataatcagaatagcgaaaatacttgtccgatctgtggtggcattctaaaagtagaaataatgattcagcccgaagtctttggacctgaaaatgccaaggaacttaatgaggacgacagagagcaagaaatcacctatgtaacagcggcacaatatccacaacctgttgatcctgaagattttaagttcaataatggaggtgctcatattgtttttactcacgcaatagatcagaaactggtgacggtgaaccgagggaaaaatgaggggagtccagtggtttttcagtatgttgcgaatgtggtgcggcctccgtttatgattcctactcaccggcaaaggggcacatgaaagaccgtataaatatatagcaactaaggaaacgcctcgcttatgctctggcgagtataaacgcgtttttctcggacatg

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		atttccgtactgatttgcttttattacgaataaccgttgggtctccgcttgtaactgatacttcaaatgctatcgttttacggatgtatgaagatgcattatatacaatagcggaagcactaaggcttgcagctagtcgccataaacaactggatcttgatcctgctgagtttggctctggtttcagaattttacccactatagaggaagatactcaggcattggatctcttcctttatgatactttatccggcggtgcgggttatgcggaagtagcagcagcgaatctagatgacattcttactgcaacactcgcattgttagaaagctgtgagtgcgatacctcctgtacagattgtctcaatcatttccacaaccagcatatacaaagccgtctcgataggaaactaggtgcatctttacttcgttatgcactatacggaatggttcctcgttgtgcttcacctgatattcaggtagaaaaattgtctcaattgagggcaagtctggaattggatggttttcaatgcataattaagggaactcaggaggcacctatgattgtgagtttgaatgaccgttctattgcagtgggaagttatcctggtcttattgatcgacccgactttcaacacgacgtatataagtcaaagcatactaatgctcatatagcctttaatgaatatcttcttcgttcaaatctgccacaatcgcatcaaaatattagaaaaatgttgcgctgatagcagcagtattgagtgccctaaagccctgtagggcactcaaggttttcagtgcgtgagcgggctttaactgaagccataaatgtacgtatgggagaaaatgtgaccatttaactcgccagcaactattgcacaatgtaaaattatgcccattgag (SEQ ID NO: 4)
3	pLG005	acggtaatgctgagtttctccattaccattgcaaatgactcaccagagcagactgaacagcgcagaagtgggattgtggatacgtgaagtgagagtaaggggaaaatccacaataatcatctatcgaacagggaggcgaactttacacgatggttttccgggagtgcttacccggggttcctcacctctggctaatctctggattgagtcgcgatactccaacaaaagcaacaagctaacgcagcaagaagttaacgctcatcgagagtaaaatgcacacttttatggcttactcgttacaataacagccagtttgttcagaaaaccggattcagtatggccagaataccaaccaaaaaagctaaagcaaaaaaagggtttgaagaaacattatgggatgccgcaaatcagcttcgcggcagcgttgagtcctccgaatacaagcacgtggtgttgagcctcgtgttcctgaaattcatcagcgataagtttgaaacacgccgcaaaaaaatgattgccgatgggcaggcagatttccttgagatggaagtgttctaccagcaggacaacattttctacctgccggaagaggcgcgttggtcatttatcaaacaaaatgcaaaacaggacgatattgcggttcgtattgacaccgccctctcgaccattgagaaacgtaacccaaccctgaaaggtgcgctgccagacaactacttcagccgtcagaatctggaaaccaaaaaactggcatcactgattgataccatcgacaacatcgaaacgctggcacacgagactgacgttgaaacgttatcgaaagaagacctggtcggacgcgtttatgaatacttcctcggtaagtttgccgccactgaaggcaaaggcggtggtgagttctacacgccaaaatgtgtggtcacgctgttaactgaaatgctcgaacccttccagggcaaaatttatgacccgtgctgcggctcggcaggaatgttcgtgcagtcggtgaagtttgtcgagagccatcagggtaaaagccgtgatatcgcgcgtatggtcaggagctgacagccacgacgtataaactggcaaaaatgaacctcgctattcgcggtctttcagctaacctcggcgaacgcccggcaaacactttctttagcgaccagcacccggacctgaaagctgactatattctggcgaacccgccgttcaacctgaaagactggcgtaacgaagcagaattaaccaaagatccacgttttgccggttatcgtatgccgccaaccggtaacgccaactacggctggattttgcatatgctctccaagctgtcggctaacggcacagcgggttttgtgctggcaaacggttcgatgagttctaacaccagcggtgaaggcgagatccgtgcacagatgatcgaaaatgatctgatcgactgcatgattgctctgccaggtcagttgttttacaccacgcagatcccggtgtgtttatggtttatgaccaaatcgaaggctgccgatccggccaaaggttatcgtgatcgtcagggcgagacgctgtttattgatgcgcgtaacctcggcaccatgattagccgcacaactaaagagttaacagcggaagatattgccacaatcgccgatacttaccatgcttggcgtagcacgccagaagaactggctgcacggattgcgcgtggtgacagcaagctggaaaaatatgaagaccaggcaggcttctgcaaagttgcgaccctgcaagatattaaagataacgactacgttctgacaccgggccgctatgtgggtgcagccgagcaggaagaagacggcgtggcatttgagaccaaaatgcgtgaattgtcgaagacgttgtttgagcagatgaagcaggcggaagaactggatcgtgcgattcgccagaatctggaggcgctgggttatggggagtaaatgggagaaaataaaacttaaagaagttgtagatattatcactactaaagttgatgtatcgcaaattagtctttgcgattacatatcaactgaaaatatgcttaccaattttggaggtatatcaatagcaaatagtaaacctagcacagggaaaataacaaaatttcattctggagatattttattctcgaatatcagaacatatttaaaaaactatggcttgcagatcgaactggtggctgttctaacgatgtaattgtattccgtcccaaaaaacatattaattctaattatattttatcagtattaatggatcaaaaattcatcgaatatactgttttaacatccaaaggcaccaaaatgccaaggggtgataaaacagctatattagattatgaatttaatcttgcaccagataaatattgccaacatatcgcaaaaacaaacactcttatatttagtaagttaaaatccaatgaagtaataaataagtcattagaacaaatgtcccaaactctcttcaaatcctggtttgtggatttgatccggtgatttataacgctctggatgcaggaaatccaatcccggaagctctgcaatctcgtgccgaattacgtcaaaaagtacgtaatagtacagattttaaaccgcttccggcggaaatccgttcgcttttcccaagtgaatttgaagaaacggagttgggttgggtgccgaaaggatggagtattgttcgaactgaagatattgcattgaaaataggaatgggaccatttggttccaatattaaagtatccacatttgttaatgctggtgtaccaattataagcggccaacatctgaaagccctccttcttatcgatggggataataatttcattactccagagcatgctgaaaagctcaaaaactctgctgtatatagaaaagacataatttttacacatgcaggtaatattggccaagtttctttaattcctgaagattctgaatatgacagatatataatttcccaacgtcaatttttcttacgcgtaaatgaatcaaaatcatcgccgtactatttgattcattattttaggtcagaaaaaggacaacatgctctgctttctaacgcctctcaggttggtgttccttcaattgctcagccttcaacacatttgaaaaatatatcattcctaaatcccccaatggttttgcttaaagagtttgaaaaatttagcaccccctttattccatcgctttagtaaaaatagaaaatgtggagtctcactaacagccctccgcaacaccctgctcccgaaacttatctccggtgagctatccctggaagatcttccggatctcagcaccgatacagaagccgcataacgcatttgccccctgtaaaatcaggggctttctggtaaggttttctactgatacaggaatgcttaccagaaattagccagggttggagcgcgatatgagtctctctttcagtgaagcaaaattagaacaagcgatcattgaactgttacaggatcaggggtatcaacatctgatcggcgataatgtcccacgttcgagtctcgatcaggtcattatcgaagacgatctccgtcattatttagcggcacgctaccagcctgatggcattactgaagaagagattcagcgactgatcaaacagttcaccacgcttccggcttccgatctttatgaaagcaacaaaacatttgtcgctggctggcaaatggttttctgttcaaacgcgacgatcggcaacaaaaagatctctacattgaattgctcgacacccggcatctacctgccgcactgcgccagatatttgacgccgaagatgtcctgttgcaacaggctgcggaactcccgccctcctatattaatccgccgcttaacctgattaagattgttaatcagcttaaaatctccggcaaagataatcagagtcgtattcctgacggcattctctatatcaacggtctgccactggtcgtctttgaatttaaaagtgcggtgcgcgagcaggatgctagtattggcaatgcctggagacaactctgcaaacgctatcgccgggatattccgcaactgtttatctacaacgcgctctgcattattagcgatggagttaataaccggatgggcaacctgtttgcgccctatgaatatttttactcatggcgaaaagtcaccggtaatgaaaaccgtgaacaggatggaattccatcattgcactcaatgattcaggggctgtttcatccggtacgtctgctggatgtaattaaaaactttatctgcttcccggataaagccaggcacgaagtaaaaatttgctgccgatatccgcagtactatgccgcccgcaaactctattacagcatcaagcaagcgcgtaaacctttcggtaacggtaaaggcggcacttactttggcgcaacgggctgtggcaaaagttacaccatgcaatttttaacgcgtcttttgatgaagagcgtagagtttgccagcccgaccattgttttgatcaccgaccgcaccgatctggacgatcagctttctgcgcaaatgtgcaacgccaaaaattacattggtgacgacaccatccttcccgttaccagccgtgaagatttgcgtaatcaactggcgggacgcaatagtggcggtgtcttcctgacaacgatccataaattcaccgaagacaccgaactcctttctgaacgcagcaatatcatttgcatctcggacgaagcacatcgcagccaggttaacctcgaccagaaagtcatcatcgataaagaaagcggaaaagtgcgcaaaacttatggctttgcgaaatacctgcacgattcactgccaaacgccacctatgttggctttaccggcacaccgattgacgcgacgctcgatgtcttcggtgaggtgatcgacagctacaccatgaccgaagccgttcaggatgaaatcactgtacgcatcgtgtacgaaggccgtgcggctaaagtgatcctggactccagcaaactggaggaagtcgaaaagtattacgaagagtgcgcaaacgcaggcaccaatgagtggcaaatcgacgaaagcaaaaaagccaccgcaaccatgaatgcggttctgggtgatgaagatcgattaaaagccctcgcggaagattttgccaaacattatgaaaaacgcgtagccgaaggt

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		tccaccgtaaaaggcaaagccatgtttgtttgtgccagccgtgaaattgcctgggatttctaccgccagcttaaagctattcgccctgcctggtttgaagtgaagcaagcccccgatggcgtcttcctgacagaacaggagcaaaaagagttaccgccttctgaaatggtgaagatggtcatgacgcgcggtaaagatgacgacgaggcgctttatgatttactgggcacaaaagaatatcgcaaagagctggataagcagttcaaaaacgctaaatcgaatttcaaaattgccattgttgttgatatgtggctgaccggttttgatgttcctgaactggatactatctatattgataagcccttacaaaaacataaccttatccagactatttctcgcgttaaccgtaaactggaaggcaaaagcaaagggttagtggtggactacatcggcattaaaagtcagatgaaccaggcactggcaatgtattcccgcattgatgccaccaactttgaagatattcagcaatcggtgactgaagttaaaaaccatctcgatttgttggggcaagtcttttacgactttgacagtcgggattattttagtggtgagccacaagcgcaattatcctgcctcaaccgcgcggcggaattcgttctgcgtacccagaaagttgaacgtcgttttatgggactggttaaacgcatgaaagccgcctacgacgtctgctgcggcagtgaagcactatcacagacagaacgtgatcatattcactattatcttgctgttcgttcaattgttttcaaactgacgaaaggtgacgcaccggatgttacccagatgaatgcacgcgttcgtgaaatgattgcagaagcgctaaaagctgatggcgtagaagaaatttattttcttggcgataaaaaagcggaatccatcgatatttttgacgaagattatctggcgcgaattaacaagatcaaacttccggcaacgaagatccagctattacaaaaattactggaaaaagcgatcagcgacttcaggaaagtgaaccagttgcaagggattaacttcacccgccgcttccaggctattatagatcgttataatgagcggcgagaagatgatgtactcaacggtgaagaattcgatacattcagtcaggaaatgaccgatattatctatgatattaaaacagaaatgggcacctgggccgatttaggtattgatattgaagaaaaagcgttcttcgacattcttgctcatatgcgcgataaatatcagttcacctatgacgatgaaaaaatgctgtcgctggcaaaagagatgaaaagcgtggttgacaacacatcgaaatatcctgactggagtaaacgcgatgatattaaagcgaaactgaaagttgaacttattctgcttctacacaagcataagttcccgccagtagcgaatgatgatgtttatatgggggtactggcgcaagcagagaactttaagaaaaatcacatgagttgagtctgtcataatggagtatctcatcagatactccttctttatctattttgtaagagccaaaatagataaattatgttacgcataaccagctcatttaaactatctggtctgtttcctccggttctacaaaaatagataggggtgcacctacgttaccaatactggcatcatggctacatacggtggtcagtttacgcttactcaccattctttactttttataagcgtcaataggtttgtaagcgactcgtcagaaccgtattgatat (SEQ ID NO: 5)
4	pLG006	acctgccttcctttgatacaattcgtaacaggttactatcatcataaaaaagctcaacccgatgaactcgctaaaaatgagacaaatcatttatatctcgaaaaaacttgttacaatcatgagcgctacaccgaacttaaccatataaattatgtgtgttttgtttattttttaaacgattacaactatccattatttacacaggtatcaaaatgttagcgcagctttttgagcagttgtttcaatcgatagactctacactgatcaccaatattttcatctgggctgttatattcgtatttttatcagcgtggtggtgtgacaaaaaaaatatacatagtaagtttagagaatatgctccaaccttaatggggcattaggtattctgggtactttcattggtattattattggtttactcaattttaataccgaaagtattgataccagcatccccgtattattaggtggcctaaaaacagcattcattacaagcattgtaggtatgtttttgccatttatttaatggaatggatgctttctttttgccaataaacgaagtgcgttagctgaaaataaccctgaatctgttacacctgaacatatctatcatgaattaaaagagcagaaccagactctgactaaattagtctcgggtattaacggtgatagtgaaggttctcttattgctcaaataaaattactacgtactgagattagcgattcctcgcaggcacaattagctaatcacactcatttcagtaataagctttgggaacaacttgaacaatttgcagatctaatggcaaaaggtgctacagaacaaattattgatgctttgcgacaagtcattattgattttaatgaaaatttaactgaacagtttggtgaaaactttaaagctcttgatgcctctgtaaaaaaacttgttgagtggcagggaaattataaaacgcaaattgagcagatgtcagaacaatatcaacaaagtgtcgagtccctggttgaaacaaaaactgcggttgcagggatttgggaagaatgtaaagaaattcctctggctatgtctgaactgcgtgaagtgcttcaggtgaaccaacatcaaatcagcgaactctcccgccatttagaaacctttgtcgccatccgcgataaagctacaaccgtattacctgaaatacagaacaaaatggctgaagtgggtgaactgctgaaatccggagctgcaaatgttagtgcatctcttgagcaaaccagccagcaaatacttcttaatgcagattcaatgcgcgttgccctggatgaaggtaccgaaggattcagacaatcggttacccaaacacaacaagcatttgcctcgatggcgcatgatgtcagcaattcctccgaaaccctaaccagcacgttaggtgaaacaattactgaaatgaaacaaagtggtgaagaattcctgaaatcactagagtcgcactcgaaagaattgcatagaaatatggaacaaaatacgacgaatgtgattgatatgttcagtaagactggtgaaaagattaaccatcaactatccagtaatgccgataatatgtttgattcaatccagacatcatttgataaggctggtgcagggctgacttctcaagtcagagaatcaattgaaaaatttgctctatccatcaacgagcagttacatgcttttgagcaagcaactgaacgtgaaatgaaccgtgaaatgcaatcattaggtaatgctctgctttcaatcagcaaaggttttgtcggtaactatgaaaaacttattaaagattaccaaatagttatggggcagttacaagcattaatttctgctaataaacatcgagggtaatcgatcatggataagattatagggaaacaattacctaaaaaagatcaagataatgaacattgggtatccatgtcagacctaatggcagggctgatgatggtttttatgttcatatctattgcttatatgcactacgtacgtattgaaaaagaaaaaattaaagaagttgccgtagcctacgagaatgctcagttacagatttataatgctctggatattgagtttgcaaaggatttacaagactgggatgcagagatcgataaacagactctggaggttcgatttaaatcaccggatgttttatttggcttaggaagcacagagctaaaaccaaagtttaaactcattcttgacgacttctttcctcgctacctaaaagttctagataattatcaggaacatattactgaagtccgcattgaaggtcacacaagtactgactggacaggaacaacgaatcctgatattgcttattttaataatatggcactatcgcaaggtcgtacacgtgcagtattacaatacgtttatgacataaaaaatatcgcgacacaccaacaatgggttaaaagtaaatttgccgcagtaggttattcatctgcacatcccattcttgataaaaccggcaaagaggaccctaatcgctctcgtcgtgtcaccttcaaagttgtaacaaatgccgagttgcagattagaaagattattcaggagtaagagatgaaattatctatcgacatttcagaacttattcaattagggaagaaaatgttaccagaaggagtcgatttttttctggatgaatcccctattgactttgatcctatagatattgagttatccacgggtaaagaagttagtatcgaagatcttgaccctggtagcgggcttatctcttatcatggccgccaggttcttttatatattcgggaccattcagggcgttatgatgcggctatcgtagatggcgaaaaaggaaaacgttttcatattgcctggtgcagaactcttgatgaaatgcgccataaaaatcgatttgaaaggtatcatgcaactaaccgcatagatggtttattcgaaattgatgatggttcaggtcggagccaggatgttgatttacgggtatgtatgaattgcctcgaacgacttaattataaaggaagtattgataaacaacgaaaaagagagattttttaaatcattctcattaaatgagtttttttcagattatagtacctgttttcgtcatatgcctaagggtatctatgacaaaacaaatagtgggtatgtcgaaaactggaaggaaatatctaaagaaatacgagaaaaggcaaattatgtttgtaatgattgtggcgtgaatttatcaaccgccaaaaacttgtgccatgtccatcataaaaatggcatcaaatatgataatcaccatgaaaaccttcttgttctgtgcaaggattgccatcgaaaacagcccctccacgaaggtatattcgttacccaagcagagatggctatcattcaacgtttacgttcccaacaagggttattaaaagcagaatcctggaatgaaatatatgacctgactgatccatcagtgcatggtgatattaatatgatgcaacataaaggctttcaacctcctgttcctgggttagatcttcaaaactcagaacatgaaattattgcaaccgtagaagctgcatggccaggccttaaaattgcagttaaccttactcccgccgaagtcgaaggatggagaatatataccgtgggtgagctggttaaagaaatacaaaccggagcctttacgccagcaaaattgtaaattctaaaactccgtgaaagttaaggctttcacggaagataaataaagtttccctgatttgtgactcaaattacaaaagtagtttatggcataacttgtctgattttttatggtgtaacaggtataaaagcatatgctatggttcgcctcatacttaaaacttccctcatatgggtgaaggttaaagcttggtagacagaagacagtcacaatgaataaagcaataaattga (SEQ ID NO: 6)
5	pLG007	acatcccgtcatcatgccatcacgacgcgctgagacgctgaaaaaataaaatcagcaccaccgtcagcgcgcagtgctttccccgcctcgcccgcccgcttcatgagacggttttaatgcagttgcattatgtcccgctcctcagtgctgcgctccatcctgattacaaaaaccgttatcaaaaacacatgcaaatagacgcagtcaaatgcgctaccgcctctcgcaataccttcaatttcatgataaaaaacatca

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		tccctaacaagagcattatcctcatgaaaaaagtatatgaactaaccagtgaagaagcactgtcatattttcttcgccatgactcctacacaacattagaattaccggcttatattaatttcaccacattattaaatgatattaattcatctatccataacaaaaaaattaaaattgaaccaaccgccaaggagctgatgggtaaagatatcaattatgaggtgcttgtcagtaaagatggtctatatagctggcgtaggataacacttatcaatcccctttattatgtctacttctgtagaaaaatcacagcaccagcaacctgggaaatcataacagaaaaattcaaatcttttgaatcaaacgacctttttacatgttcaagcatccccgtcagaaaagacaactcgtcaaacattgctgcgtctgtaatgaattggtgggaagatttgaacaaaaaagccttgcccttgctcttgaatacgaattcatgttcagcactgacatctcaaacttctacccatcaatatatactcatagttttgaatgggtattcatatcaaaagaagaggcaaagaagaaaaaaagcaaaaataacccagggggattaattgacagccacattcaaatgatgatgaacaaccagacaaatggtattccactcggcagcacattgatggatacatttgctgagcttatcttgggtcaaatcgatatagaattaagaaaaaaaactaacgaactcaaaataataaactacaaggtagtacgctaccgtgatgattaccggatcttctctaatagcaaagatgatttagacataatatcaaaatgtttagtcaatgtattgggcgattttggtttagatctaaactcaaaaaaaactgaactatatgaagacatcatacttcattcgttgaaacaagctaaaaaagactacatcaaagaaaaaagacataagtcactccagaaaatgctctattcaatatatttattttcacttaaacatccaaactcgaaaacaaccgttagatatctaaatgatttcttaggaatttatttaagcgaaagacaattaaagataacggccaacaggttgatgctatgcttggtattatttcaagcatcatggcaaaaaccctacaacgtacccagtaggaacggcaattttctcaaaactcctcagttttctttatggtgatgacacccaaaaaaaattaacaaagctagaacaactccataaaaaactggataaacaacccaatacagaaatgcttgacatatggtttcagcgaactcaagcaaaaataaacctagagtggaataaatcttataagtcagctctatgcgtccgtataaatgatgaactcacaaaagagaaaacattttctgtaaataatttatggaatattgactggatccaaggaaaagaaacaagccccaataaagccaaaatattatccttgctaagaaaaacaaaaatcgttgacacagataaatttgataaaatggatgacaatataacacctgaagaagttaatctattctttaaagagcacagcaattaatatcccaaagccatgttagtaacataacatggctttttaaatcactcattatcagttatcaagaacgaacataacattctattccgaggag (SEQ ID NO: 7)
6	pLG008	agtttttaaaggggttattttctaattatagtcccttaatttccattttcgtgtctaattatttgacattagtccatacaatagtgactctaagatttaaggataacatcaactttcaacataagcacaataactatttttttattataattgaaaagagaattgaattattacctataaaacttaaaggagtataattatgaaaaaagagtttactgaattatatgattttatatttgatcctatttttcttgtaagatacggctattatgatagatctattaaaaacaaaaaaatgaatactgcaaaagttgaattagacaatgaatatggaaaatcagattctttttattttaaagtatttaatatggaatcctttgcagattatttaaggagtcatgatttaaaaacacattttaacggtaaaaaacctctatcaacagacccagtatatttaatattccaaaaaatatagaagctagaagacaatataagatgcccaatttatacagttatatggcattaaattattatatatgtgacaataaaaaagagtttatagaagtatttattgataacaaattttcaacgtcaaaatttttaatcaattgaatttgattatcctaagacacaagaaattacacaaacattattatatggaggaataaagaaattacatttagatttatctaattttatcatactttatatacacatagtataccatggatgattgatggaaaatctgcatctaaacaaaatagaaaaaaagggttttctaatacattagatactttgattacagcttgtcaatacgacgaaacacatggcattccaactggaaatctattgtctaggattattaccgaactatatatgtgccattttgataaacaaatggaatataagaagtttgtgtattcaagatatgtagatgattttatatttccgtttactttgagaatgaaaagcaagaattttaaatgaatttaatctaatctgtcgagaaaataacttaattattaatgataataaaacgaaagttgacaatttcccgtttgttgataaatcgagtaaatcggatatttttctttttttgaaaatattacttcaactaattccaacgacaagtggattaaagaaataagcaattttatagattattgtgtgaatgaagaacatttagggaataagggagctataaaatgtattttcccagttataacaaatacattgaaacaaaaaaaagtagatactaaaaatatagacaatatcttttcgaaaagaaacatggttaccaattttaatgttttcgaaaaaatattagatttatcattaaaagattcaagattaactaataagtttttgactttctttgaaaatattaatgaatttggattttcaagtttatcagcttcaaatattgtaaaaaaatattttagtaataattcaaagggcttaaaagaaaaaatagaccactatcgtaaaaataattttaatcaagaattatatcaaatattgttgtatatggttgtctttgaaatagatgatttattaaatcaagaagaattactaaacttaattgatttaaatattgatgattattctttaattttagggacgattttatacctaaagaatagttcatataaattggaaaaattattaaaaaaaatagatcaattatttattaatactcatgccaactacgacgttaaaacttctcgtatggcagaaaaattatggctatttcgttatttcttttatttttttaaattgtaagaatatttttagtcaaaaagagataaatagttattgtcaatctcaaaactataattcaggacagaacggatatcaaacagaacttaattggaattatattaaaggtcaagggaaggatcttagagcgaataactttttttaatgaattgatagtaaaagaagtttggttaatttcttgtggtgagaacgaagatttcaaatatttaaattgataagtatttgaaatctattattagttcctgaaaaaatagctgtgtcttgtcaatataaatgacaagacacagctattttttttaattttgaaatttataatt (SEQ ID NO: 8)
7	pLG009	
8	pLG010	
9	pLG011	gcccatcattgcattaagtgatgggcggagcctttggcctctaatctggaactagctgcgattttcagactcgaatgctaaaaggtcgtttcgcacctgaaatcaagctgctagagttctcttacggggttctcccctcgcatacgcgctgtagtaactgcggcgtaagagtaaatgtctgcacatatcatgcccgccatgatcattcggtaattcctggcgtgactggaagggagaccccgtgccacctatgggccatatttttggaccagtgagtttcgtgaagttgccgccggagttgatgagtgaggccagtcttcttgctcatcttggcgttggccgtgccgaacttaatgtcattagttggtacgccggtaggatgtaccataaattcgacattaaaaagaagtctggcaaggcgagggtgattaatgcgccggatcgtcggctgaagatgttgcagaggaagatcgccgatttgctgacgcctctctatcggaggcgcaaccctgttcacgggttcgtgatcggtcgttctgtgaagaccaatgctcagtcccatctgggcagcaagttcatcgtcaacttggatttgaaggatttcttcccgtccatttcgtacggacgcgtgacgggcgtgctgcgttcgcttggcatgaagcgcgaggtcgcggaagctattgcgacaatttgctgcctcaatgggacgttgccccaaggcgctccgagcagtccgatcttgtccaatatggtttgcttccgcttggatcggaggctgcgggagttagccaaggacgcccgttgcatttacacccgctatgcggacgacctgagcttttccagctaccagccgctaatgggattgttcgaaacgacaccaccggcttcagggcatttctcaccggatctgttgtcggaaaaacttaagcagattttcagcggtaacgggtttgtgctgaacccggacaaggctcactatgctgacaagcattcgcgccgcaccgtgacaggcatccggattaacgaggctctcaatgtcgaccggcggtttgtgaggaatttgcgggcagccctttactctgttgaaactttgggactggccgccgcccaggcaaaattcaaatccttgcatggtggtaaagctgacgtcggccagcacctgcaaggcaaggtatcgtggttggggtacatcaaaggcgcatctgacccagtctttcggagtgtcgcatcccgtttcaacgctgcattcccgccgctcgcgctcgatattttgcccagtccccaagaaatacgagaacgatcagtgtggctgattgagcactgggaaacaggggggtgaccaaggcacggcgtttttcatgaagggtgtcggtctggtaacggcagagcattgcatatcgccgtccggtatagttgagttgtatcacccgacgaagccgtcgaataaattcgcggcgtccgtgaagcatcgatgcccagatcgcgatctggccgttctcgaccatgcaatccccaacaacgaattctatgagctcgaaaccgccggcaaggcagccgcgacaggcgatgccacgaccgcgatcgggtatcccggttatggacccggcgacagactgaacatccgacctggcgcagttacgtccctgccaactaagagtgcggtgaagatggtcgaggtccagcagatgctgacgccgggcatgtcaggagggccattgctggatgtggatgaccgcgtcgttggcgtcgttcacaagggcggccatgatcatggtcggcaactcgctattgccatatctgaactgcatgcttggctgccctgacctg

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		attagccgaaccggctaatcgcgcaggcgccgaaccagccgtttccagcttgcttcactgttcatccagtcaggccggtccggttgtcgaggcgttggagcaaatcgttcaggatgtccccgacagcgcgtgcagcgcaggtgcgatccgacggtttccatagcggtgttccagcaatgcgcgaggaaccagcggttgagttt (SEQ ID NO: 9)
10	pLG012	tctatctaaaagtatacatatagtatttcaatgaaggttatattatattttgtggctgttttctaattttatcaataagattattgcaaaaggctgataaatataatagctttattatatcggaggagttgatttaactttcctatactatctgtataggctaataccaatggcaattttgccctcaaattggtctccttaatgtttatcaacgtgttatacggtagtgataaaacctcctccgatatttttctcatgaattgggatattttaaatatgttttgctcagtaaccaagttgcatgaatgtaaaaatgttgaacaattatactatttttttaggatgtgaagaggctgaaattagtaggtttttatatagtggagtaattaaataccgctctttttccatacttaaaaaaaatggtaatttta gaaatataagagcacctgtaaagtatttaaaagaaattcagtataagataaaggatgagctcgaaaaatattataccccgaaatcatgtactcatggttttatagctggaaggaatataatcacaaatgcgaaacctcatataagaaaagaatttattttaaatatagatttaaaggatttttttgattcaattaattttggacgagttagtcgtttatttcaaagccaacctctaaacttgccagagaatgttgcccatgttttggcacatatttgttgctataatagagccttacctcaaggtgctcccacatccccaattatatctaatatgatatcttatcgtttagacagacaattgaaggagttggcaagaaataatgcgtgtacttataccagatatgcagatgatataactttttcttttactaaaactaaaaagtatcttccaaaatcaattgtttctttaagtaaagataataacattatactaggccatgaattaaaaaaggtaattgaagataattggtttgaaataaatgaaggaaaagtaaggttacaacataaaacacaaagacaatcagtaacaaatattacggttaacactaaaattaatataagtagaaaatttaaaaaacaaacttcagctatggttaatgcattatttaaatatggagcatctaaagctgaaagagaatattttagtaagtatcacaagggttatatagcagaaaggcaatataataagattaaagaaaaaccaggtttattatttacacaaaaagtaagaggaaggttgaattatatccgattagtttgtggtaagaataatgaaagctggagaaagctcatgtataaatatactgtggcaataggacaacctaatgaggagtacaatagaacattgtgggatattgctggtgattcaacgttcattctttggtcgaattcctcacaaggaagtggtttttttcttgaaaatattggtttagttacaaatgagcatgtaatcgaaggaatagaaaacagcaatattaataatgatctaataatactttggttaccaaatgaaagaaaagaatatattgagttacacttagcttggaaagatgataatactgatttagctgtaattacttctaatatatcttttcttgacataaagcctttacaagtagagccagttcctatttatgatataggaacagaagtatatgcagttgggtatcctaattatgacgccagaggctcaattggaaaacctactattattacagcaaaaataacgagtataattactcgagaaaggcaagaaagaatcgttatagaccaaccaatagtacatgggcatagtggtggggtcgttttaaatgctgatggacgtgtaataggcattgttgcaaatggaaatgccgagggggaattaagagtagttcctaatgcttttattcctattgaaatattattaaatgagcacaagttacgaactaaatcataaaattattattcttaaaataattaaatatttttttaaaaccactagtttgataactagcggtttttttatttttggagtacat (SEQ ID NO: 10)
11	pLG013	
12	pLG014	ttataacaagcatttatagtttaaagatactttttctaatcaagtagaacctttgggtggcatcggcctatctcgcttttgtccaaatgtgggctgatggggcatgaaaaatggaaatgccccattcctacttagtgctattactcattcatacctcgttaacgtgattttggattagttttattcactgtatatatcaacagttataatgaagcgcggtgattttatcgctttagttctgtttttaataagaaatatttcttgttaaaaacagaagtgaaatcataactaattgaaaattatatcgtttaacatttcagtttgtatttaataagactgattaaatacatttcttacttttcacaccctctttcaaatcggtgagtataagaaagtgccagtaagctcataatatttaacgattatatcgagtataatatctatcttttataagtatatttttgcgtaaaagtaagaatgcttattaatatactgttagttgcatcaagtgatgcattgcattctgtttagtattgttatagattctgccgcaagaggcgagagtttaactttctgctgttaatctgcggcggtcataagcatgtttctttttaccggttttcagctagtctgatgatgccgttacgctgtacaagagaaaacaaaatcgcctcgttctttaagggtttgttactttggtagacatttcattaatttcccaaattgcagctaaagctgcattctcatccaatattcaagtacctctacctaataaattgaaagattgctcatgcgttgaagggctgactcaatatctgggttttacgaattatgatgagctgaaaaaactgatatacccctcagttgaccacctatataaaggctttagcattcctaaaaaaaatggcgagtttcgaacgattgatgcgccaaaaaaggagctaaaaacaatacaaagtttcctttcgaaggaattggttcaagtttactctcctcgtaatgctactcatggttttgtaaaagatcgaagtatagttacaaatgcgtcgaagcatgtagacaaaaaatacgtactcaatttagatcttgaggacttcttcggctcaattcattttggtcgcgttcgaaacctgtttcaatcgcatcctttgaacttacaccattcggtggcgacggttttatctcacctatgctgccacaatggcaagttacctcaaggcgctccaacatccccgatcatctcaaatatgatcgcttatcgtttagacaagcaactgcagacattggcttctaaaaatagatgcacatatacacgctatgctgacgatataacattctctttcacacaaactcgtgggcgcttgcccaaatctattgttacgttaactcgcgatctacaactctctttgggtaatgagctaaaggagcttattactgagaatggttttgttatcaattctgataaaactagaatagctgcgcgaagtaataggcaagaggtcactggtgtgatcgtcaatgagcgtatgaatgtgtctcgaaagtacattaaacaaacacgttccatgctatatgcatggaaaaagtttggtctcgaagatgctgaagaaacctacttgagaaagtttcatggaaaaacagtgtttgagaagcaccagcggcgaattgacgaaaagaaagggcagtttttttaagaaagttgtaaaaggcagaattaactttattaaaatggttcgtggtgctgaagatttaatatacagaaaaatagcttacgaattctctgtattaattagcaagcctaaaccagagcttgtgcaaaccccattggataaagcgtgtgattcaatatttatcgttgaaaatatggtggagaagagccaagggacagcgttttgctgaagggaattggtatcgttacaaatgaacatgttgtgcgtggaatcgatgaggaactgtcagatcttttggagctatttaggtatcatgagcaggaaactaagcgtccagttaaatttcaaaagtcatgcagatctagggatttggctattctaaaaccaactacaagctacaacggtattaagcgcttggatgttggtgatgatagtcagatcggtattggttcggttgtaaccgtcttaggttttccccagtattcgcctggtgaaacgccttatatcaatacaggcaaaattatccaatctaaagtattgtttggtgaacgcgtctggttgctagatatacctgtaatccatggaaatagtggtggccctgttcttaatgaccgtcaagaagttatcggcgtagctgcaataggttcgccaacacatgaccactcaacgaaactccatggcttcataccaatttccacgttattagcgtatgtggaagaatgcaactaacaaataaggatatgtgtcgcgaagccgacacctatccgaagtgttggacaagcccaagccaccttatataagtaaataccatcaagagtaatgtcaaatccttacttttcctaatctctaaaagcctaaatagaacgaacggtctaagaagcttttgtccaacaacgagctagcttatgtgatagctagtttgtgatcaaactttagatttttacactctacaaatagcttgaaaagtcacatttccgatcagactta (SEQ ID NO: 11)
13	pLG015	
14	pLG016	cgttaataattatgttgttagcttaccacatttcattatcataaatacttacagtaggtaagataatgtaaaacatcgcgattaaatataaacttttcaaaaatgctgttaatattgatgaatatatatagtataatttacactgacagcaagggtaagaaaaaattgactttatggcggtgaaatcgccgtctgttatttaaagggtatacttaatttacacgcttattttatcttcgaagttttattcgatttgtctaatcgctattaggagaagggtagaattttaacccttgctgttgtaaataggaggggattgctatggtttataagttaaattttgaattacagagcaatctagaggatattaaacaaaatttcaagaatttatcttgttttgaagatgtagctctccttttagaggtaccaaaagaattattgtggaaagtacttataaaaaataaaggagctaattataaggcgtttaaattaaaaaagaaaaatggttcagaacgt

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		gttatttttcgcctactttaagtttatctattctgcaaaaaaagctagcttatatttggagtctaactataaaaaccataggcaatcatatggttttgtaaaaggaagaggaatagttgataatgctcaaaagcatttaaataaaaaatatgtactaaattttgatatagagaatttttcgaaagtataacctttagaagagttagatcaatgtttatgacatattataaatttaatgaaaaagttgctacaaccttagcaaatatatgttgtcatccgaatggttttctgccacagggagcagcaacatcccctatcatatcaaatattatatgtaatagaatagataaagagttttctaaattggccaaaaacaacagatgtcaatatactaggtatgctgacgatataacgttttctacaagcaggagggttttccctcatgatattgcatatataaaagaggggtctatttttctgaatagtaatgtaattagtattgtggaatatcaggggtttaagattaataaagaaaagacaagacttcagaattatagacaaaatcaaactgtaacgggaattacggtaaatgaaaaattaaatgttaaaagaagctatgtaagaagaataaggtcaattcttcactgtattgaaaaaaacgttgaagatttacagaaagcagaacaaattttcgaagaaaaatacccatttcgtcaaaagaaatatcttgataatattaatatgtttgctattttaaaaggtatgatttcacatgttgggcatgtaaaaggaaaagatgaccctttatatttgaaattagcaaagagatttaataaaatatcttatcttagtgaaactatatctccttttaaattagaatctttaaagaaatttcatgaaacttatacatatataattgattatgatgataaagttcctttagtttgttttgaaaacgataaaatggaggaaatattatacggtcaaggaacgggctttttattaaagggagttggcttaatcactaatgctcacgttatagaagatgcaatagaagctattaaggacaataaaaaatttaacaatgagtatggtatctcattttttagaggtaattatcctgatttaaaatataaagcgaaagtatccaaatatgacctagataaagatattgcaattttagatataaaaggttttaatatagacaatcaaggatatgaatataacattgacatgaaagatgggcagaaaattgaattaatagggtatccagactacaaaatagggcaagaaataaaaatcgaaactggccacctaaaaggtattagaaaacatagagattcaaccggaacgttccattcacgacgggaaatatcggcaatcatatacggaggaaacagtggcggacctataataaatgaaagtaatgaagtcataggagttgcagttaaaggtgctacccttcatggtgtttccccaagcgagattattccaattgaagatgtaattaatttaaactccagtaactcagaggtcagctccaagattgcaactaagcctcattaaaagatttaatattttaatgcgaaaagtcgatttttaatcaatctactttttttatttttcattttaagttgtaaatatctcttacaatttattttatttcaacgacatatttgggtatc (SEQ ID NO: 12)
15	pLG017	gtggcaagattataccccatcaggcataagatgctttgacttataacgcatcagtttgaaacacaatggtgatgggggtcacagggagctgacatgtacttttaagattaaaaagcattaacatctacttttgaagaaaacagaaaaaaacaatcacaaacctttaaaaacaaaaactatgccaattattaataaaaagtatcaagagcttcagttaacagatgagtacattaccgatccactgctcatggccctagcctggaagaaaagccatcactacatacgtaccacaaattggtatgctgacaactttgaactagacctgtcggctttggacctaatgcagcactgtaaagattgggtcaagagaatgcaggacaaaaaagaatttaaattttcagagctacaacttgttcctgtaccaaaagcctgtaaatgggagtttaagactgtcgaaaataaggttctatggcaaccttgtgatgaaaaagaacttaccctacgccccccttgcccatatacccatagctgaacaaaccatcatgacattagtcatgatgtgcctagccaatacaatagaaaccaagcaaggaaacccagacaccagctatgacatcgtccaccagaaaggtatcgtcaattacggaaatagactttattgtcagtatattgacgataaagcagagcacagcttcggtgcaacagtgacatatagtaaatacttcactgattatcggaaattttaaataggccttatcattttgcgtcaaaagcgcaaggtgaaatttcgccggacgaagccgtttacatcatagaactagatcttgcgaagtttttcgatttagtaaacaggaagactctaattcaaaagataaaaaaccatatcagtgagtcaataaacaataaagaaaacccactcgccaatcatttatttaaatgttttgcaaactgggactggactgcatctagcataaaaaattatgacatatgcaagtcagacgaagtaacagaaataccaaaaggcatccctcaaggattggttgcagcagggtttctatcaaatatttacttacttgaattagatcaattcttgcataataaaattaacacagacataactgatgacattaaatttgttgattactgtcgatatgtcgatgacatgcgatttgtggttaaggttaaaaaatcaaaaaataataataccgcattcataaatgatgtaataaccaatcttcttaaaaatgagatagataatcttggactgataattaatcctaaaaaaacaaaagtagaaatttttagaggcaaatccgcaggcatctcgcgtagcttggaaaacatccagaccagattaagcggcccaatatcaatggatagcgccaacgaacaacttgggcatcttgagtcattattaagtctgacaaaaaccgatttgaaccaccgaaaaatggtaaatcaaatagattagctgagattgaaaaagaccgtttcgatgtcagggaggacactcttaagcgcttttctgccaataaaatcagtaagatactaaaagagttaagacatttcatctcgcaggatatagatactgatggggaggttattgccggggaatgggattatctgcaagaacgtttggcacggcgttttattgtctgttggagccatgacccgtcactggcactgctactcaagaaagggctggaacttttccctgatcctaagctattagaccctatacttgaacagctttgctcactcattgaaagcgataatgaaaaacaaagtgcagtagctacttattgccttgctgaaatatttcgacattcagcaatgactattcataaaaaagacacctatgcattccctgcacaagccaatgtggatgggtactttgaaaaaatacaacattgcgccgcgacattcattaataagcgcagcgcctctgacaacgaaacttggaacctgttaattaatcaggctagttttctgttgcttgtgcgtttagataatacattagaaaaaaatggcactgatgccaggcatgatcttatcttaaaactggcatcaggctttagaacaattacacttcccactaaaatggatagcaagactatagcctcatgtatttgttggctagtcaattagttaaagataacaaaccatttattcgctcctgcgcttctttgtgcgaaagaatttatgacaaagaacacgtcataaaattgaagaaaatagttagcataatatcacatcaaaacttatcattgtttaaatccttagtttatcattcacgacctttacaacagaagtggctaaactcagactccgtgaaaataataattaatgaatgccatatagatatacaacctttggcgacttctttaggcatgataaaaagtagtcactcattacttagaatcatatcaagacctgataacccatttgccaatgagataatggcattaaaactgatgcaagcccttttattggacaggattgtttgcctggataataaaaaagattatcaaataagtgtagcaaacaccaaagtgacgtttcataactactccaacccttccaacatcgaatgtcttcgatgcaggaatggatatggatgcaaaattattcaaatcatcgggatgggtcgattctattttcacggatgatgcagacactcaaatattgtatagagttgccatgtgcatccgttcagtactactcggcaaacaagactggacagattttggtcaagcaatitcccccaaacagggttatcggggiattaaaactagtagagacaaacgtcaattggggatgatgacaacacctgagtccattgccggtgagaactctcaggtttctggttggcttaccacactcttatccaagttgcttgcctggccgggaatttcagtgggtgataatggatatcaatggccagcaatttttacagtagatgctgtcagaaaactagttgatgctcggctgagtaaacttaagcaggattactgcaaactatcaggaactccgggacttacagaaaaaatacagttcaactggtctgactcgaaaaaagccctaacagttgctatggtccagtcaaaactgcctgcaacgaaagatttgtcagccatggacttcttttaaactccgcaaagtatagagtgattcatcgcagacatgttgctgaagtggctgatttagttgtaaaacacacgcttgcacaaaaaacaactcaacgaactcatggtgaaaaaatagagaacattgatttaatagtatggcctgagctcgctgtacatagtgacgatttggatgtactcatcgccttatctagaaaaacgaatgcaatcatatactcgggcctgacatttattgagcaacctggaatcaaaggaccaaataattgtgccgtttggattgtcccacctaaaagcaatagcagccagaaagaaatgataagacttcaaggcaagcataatatgatggaagatgagaaaggccgggttgaaccctggagaccataccaattgatgcttgaacttgttcacccccaatttactgataaaaaaggatttgttctcacaggctccatttgttatgacgcaaccgacatcgcgctaagtgcagatctcagggataaatcaaatgcttatcttgtagcagcattaaacagggatgttaatacattcgattccatggttgaagcactgcattatcatatgtaccagcatgttgtgctcgttaactcaggggaattcggaggatcttacgctaaagcaccttacaaggagccgtttaatcgtttgattgctcatgttcatggcaatgatcaggtagctataagtacgtttgaaatgaacatgtttgatttccgtcgtgataatataggaaaaagtatgcaatccgggttagataaaaaaactgctcctgcaggaatcataatgtaataaatattagatatttttatattagaggtgaggagatggcgtcacctctaatattttcgctgattgtatttagcatcaaataataaaggtacaattaatttaagtgactatcatgaaaaaattagttccgccatatcaagtaaccccggcacaaatctatcgttccgttgccagttctacagccattgaaaccggaaaac (SEQ ID NO: 13)

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
16	pLG018	gcttatcccctccctactggtaacagcgttatcgaacttggaataccatcatttatacctatatctgttggtagatgtgcattgaagtgggttgaccttgagagagccagtatcgcgggcgcaggaatgacaggtaagcactaaatttcaggcacaaaaaaagctgcccttaagcgacttgattgtatcttttggtgcgaaggccggactcgcacataaaacttaacctcatgatttaaaaaagataacaaaaaacagtttaattttataccaacacagataccaacacgaaaattcattgttcttgggtatcgaacccggacaaacatgactgagttgtattagctcagatttgacctgacacagttatggcacagatctcaacctaatctgacaggcagctccgtatcagaagcggaagtgatgaccaagtttaagcatcattcttggcttgtatgagaatggcactgatctagcgatcagtaaaacttcatcgcttcatcgaaatgccctaaaactttagattaggagaaagttctatttatgccagctacaatttttcgggggagttaccttaccgctaaataaaccgaaaatcgatgctggacaatctctaactcggtggtcaattttcgttgaccactacataatggtcctcctgatgcatctgatgtatcaggaggaccgtccttaaacacgacaaaacctgtgatacttaccatggattcctctatgaaggaaaggtagtatagccattttgggtgatacatacagtgaatgtcattgctgtagttgaagtgagtaagagcgcttaagattaagttgagagaaaatgaaactacttgataaaaagtattacaacctcgagcccaaatatgagtaccttaaggactcatttattttaggactggcatggaaaaaaacagatagttttgtaagaactcacaattggtatgcagatattttagagctggacaagtgtgcgtttgatattagtgatgaagtcactaattggtcaaacgagatctcaaagaacgctctttccaaaagtgatattgaattgataccggctccaaaaggagcaagctggttcattaatcaaggtaaatggactaccaataaagataatagaaagataaggcctttggctaacatatctattagggatcagtcttttgctacagcagtaacaatgtgccttgctgatgctatagaaacaagacagaaagactgttcgttgagcaatcttggctatgctgagcatgtaaagaacaaggttgttagttacggaaataggcttgtctgcgattgggacaatgaaagggcaagatttcgttggggaggaagtgaatattataggaagttctcttccgattatcgaagctttctacaaagacctatctatataggcagggaaacagtaaataaagttagcggaattgatgatgtatatatcatcagtttagatctgaaaaatttttcggttctataaaaataaaccttctgttagaaaaaatcaaaaaaatatccgctgatcattatgcagctaaattcataaatgataatgaattttggactttggcgaatcggattttaagttgggattggcctgaagaatctttatctttacttgagagtttggatataaaagaaaaaaatgttggtcttccccagggattagcttctgctggtgctctggcgaatgcatatctcattgagtttgatgaatctttaatttctaagcttcgtactaagatagaagacagccaaataatactgcatgattattgtcgatatgtcgatgatattagattagtgatttcaggagaagcactagaaagtaataagattaaggaatctattcatgcattagttcagggcattcttgatgagacattggctcaaaatccgtcagataatgaaccatatttaaaaattaacgatagcaagacttatattcttgagctttcagacattgacaacggaagtgggcttacaaatcgaatcaatgaaattcagcatgaagtaggagcttcgagtatcccagagcgtaacggactcgataataatatcccggcacttcaacaattattactgaccgaacaggataatttttccgaggatgttgatagtttatttcccgggtttaaaaatgataagtcgataaaggtagaatctgtacgtagattttctgcccataggctggaaaaaagtttggctaaaaaaagcaagctaatttcacctgaggagaggaaacaatttgataatgaaacctcactgattgcaaaaaattattaaaagcttggctaaaagatccatcaattatggttatcttccgcaaagcgatagctatcaatcctaatctagatgcttatagcaccattcttgaaattatttttcaagaatacaacgcaatcgtgataaacgagataaatatataatgctgtatcttctttctgatatatttcgtagcgtcattgatgtctatcgaaacctagaatcagaatacgtcgacgattatcaaaaattgatgggtgaagttacattgtttgcccaaaaaatactttcctgcaaatcttttattccaaattacgcatatcagcaagcattattttatctcgcagtgatcaataaaccatttatagctagtaataaagcttcttttgatcttgcaaggcttcaatgcgtcttaattaaacagcatttagaaccgttgaatagtagtgatggatacctatttgaggtatctgctcaaatcagtaaagactaccgagcaaatgccgcttttctactttctcatacaaatagtaacaaagtagtagacttaattatcgaaaaaatttgctttccgaggaggtgaattctggaatgcaatttggaaagaaattgttaggatgcaagataaagataggattaacgaatttagatgggccatatcaaaatatgagtcaaagccaaatagttcggagcactatctttcatcagtgatcagtttcaaggaaaacccatttagatatgaacatgcgcttctcaagctaggtgtagcattagttgaactctttgatgatacagagaaaaacgtatggcaacctgatggtaagcagtattctccacatgaaataaaagtaaaattagaaggtaactcaacctcatggggtgaattatggcgtccaaattttagtatttcatgctcgatagataagaaaggtgaacctggtaaagacccacgctatataagccctgagtggttggcaaattatccacagactcaaaatgatgaacaaaaaatctattgggtttgcagtgtgctaagaagtgctgctttaggcaatgtagattatactcaaagaaatgatttaaaacttgataaagctaagtatgatggtatccattctcagttttacaagcgacgtatgggaatgttacatacaccagagtcaattgttggttcatatggaactataacagattggtttgcaagttttcttcagcatggattgcaatggccaggtttttcttcttcgtatataagccaagaagatatattgtcaattactaatattattgagtttaaaaactgtttattggaacggctaggctacttaaataagcagatatgtatttcatcgaatgttccaaccttaccgactgttgtcaacaggcctgaattagcatctaaccattttagaattgttacggttcagcagttatttcctaaggatactaatttccatccttctgacgtgactttggctaatcccgatgtgcgctggaagcacagagagcaccttgcggaaatctgtaagctaacggagcaaactttaaatgcaaaacttaaaactgagtctagggaacatacaagcacagctgatctaatcgtttttctgagttagcagttcacccagaagatgaagatatagttagagcactggcatttagaaccaaagccatcattttttccggctttgtcttctgtgaacaagatggccgaatagttaacaaagctcgttggattattccagactcttcagagtctgggacccaatggcgtgtccgtgatcaggggaaacatcatatgaccagtgatgaagtggctcttggcattcaaggatatagaccatcccaacatattatttcaattgagggtcaccctgagggaccatttaaattaactggtgcgatttgctacgatgcaacagatataaagcttgcggcagatctgagagatttgactgacatgtttgtcattgcagcatacaataaagatgtagacacatttgataatatggcttcagcactacaatggcatatgtatcagcatattgttattacgaatacgggagaatatggaggctcaactatgcaagccccgtacaaagagaaatatcataaattgatttctcatgctcatgggactggtcaaatagcaattagtactgctgatatagatttagcagcattcaggcggaagctacaaatatataaaaagaccaaaacccagcctgctggatacaatagaaaacattaaggatttttatggatactttagttaagttagctacaattatttctccattaattagtgctggagtagctatttgggcaattttggttgctaaaaaaccatcagtgaaagcaaagaaattgccaagaaaaccatcgctgatacggcctaccaagcatatttgcaattagccatggagaacccacaattttcgaaaggctacagcgcagattgtagacaggagcgagaccctatgtatgatcaatatgtttggtacgtggctaggatgatattctgctttgagaaaatcatcgaggttgaagtaaacttaaaagatagttcttgggcaaatacgttggaaaaacatttgaagtttcattctgaacattttaagaaaacgaatgttgtcgaagaggctctctatattccccctattttggatctcataagatgtgcagctaactaataacttatcccaataggattatattccacacgataagcccactggaaaatgtaacatcccaagatagtttttgggattgtttcccagtgggcggaaagtatcatgatagttgtcacccccggtggagctgcaaagatttttatggggtgggtgttacattgcgcgataaatttgaaatcgtggctttaatttctgcttcttgctcaaaagcagactgtcagatttgattgtgtgctgccagtgagaagcgtcagatcaagtctgagctaatacaactgagttaagatgccgaaatctg (SEQ ID NO: 14)
17	pLG019	agggatacgccacagcaagaaatagtttacttattcctcattttgtcgactaaaaatcgacattaaacaaaaaattcaacttaatcactttcgggaaaaatgtgacaaatatatgctcggactggttgcggggagcgtgtaacatggatacaaatcaaaattattgccagcctcactgatggattactggtgtcaagagccccccttcgggcatgaaacggctggctaattctgtacagactgtaatctaaggacgataacgcatgacatatcaggcaattttcactggctgggatgatctgacgattgaagaccttctggtcgcttaccggaaagcaaaagccgatagcttctttgagaatacatttcctgttgctatcaaatttgccgagtatgagcaggaattacttgaaaacctgcaaaaactcttagatcttttgcagagcgaagatggattcagtagcaataagaagttgattggcaaatttcgtttgttaccgaaaaaattaaccacaaagaaaaaaacatgaatcccaaaatggacacgtccacttttctaatcctaaacgagcagccgaccatttatttaataattttgatctgataccagagtttcgtattattggtgacttcccggttgatagtcacattatctctgcactatggattaacatggtcgggcataaatttgatgccagcttagataactgttgctatggcgcgcggctaaagcgtattcgtaatgatgaattatttagcaatgagcaggataatccattccatatcagtgccgtgggttcttttagcccctacttccagccctaccaaaaatggcgtggtgatggcttaaaagctatacg

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		tgacgagttggaaaaagatcgtgacattatcgccgcctcactggatttaaaaagttactatcattttattgatccactggctataacctctgatgatctctataacacactaaacataaaactgactgaggatgaaaaagcgtttactgcacagttagcagtattcttaaagcactggtctgacggcgcagcggcatttggaaagaaaatagcgtacaaaacacctgttattaatggtggtctggtcattggattaacagccagtcggatcatttcaaatatattgctacaccattgggataaattagtcattgaaaaactatcaccaattcactacggtcgttatgtcgatgatatgttccHgtaatacgcgatacagggacaattactaataatcacgaatttatgttattgctgcaagataggcttggcaatgattgcgtttatttgaaaaacgagcaaaaacaaatatggcaaatacagcagggcgagcatttccagggtaagaccaccatccagttacaatccgataagcaaaaacttttcgtgcttcaagggagggctggaatagacctgctcgacagtatcgaaaaggagatctacgagctttctagtgaacaccgcttgatgccttcaccggatcaactggaacactccaccgcagctaaagtcctttccgctgccggtagtgtaggtgaaaatgccgatactctgcgccgtgcggatggattaaccattcgtcgtttgggctggtcactgcaattacgctacgttgaaacactggcacgagatctgcctccaagtgaatggaaagaacagcgggaagagttttatcagtttgcctacaaccatattcttagggctgataatctatttgcacattttagttatctgccaaggctgcttggctttgctatcagtatgaatgaatggcagcacgcggaaaaaattgtacttaaagcttacgaatccatcaacctgttggcatcggtgattacttcaggtaaggaagtgaatataaatggttgcaaaactcgagcagtaaatgatctttggcgctgtataaaaggcacattaagctggctatttgttgatgcagcgacacgatattacagtcctgacagattatttcttgataaacgttcaaagaaagaagagtgccttgcggatacatttttaatcatatttcacaaagtctgacgaatctaaaggatttactggatcttcgctttgattcagcagattttatttaaaagcgccattggtagctcgagctgatttagcaaaggaaccttataaacagatcgtaaagagtcagtcggcagaaaaacttgttaatcagcgtgatagtaaaaaagaagttaaaatactgaaattaatgagcgactcatcgcttattgatattgacgttattaagctattttgaaatcaaccaagaatacccgactggaaaaagtggctaaaggaaatcgtaagaacgaaagttacctaccttacattttccctacacgtcctttaacacccgctgaaatatcagaactggcccccgaatgtgttggattaccctccacatccgacaaaaaccagatgagagaccgtccaccatttgggcaaaatatactcaagcattacgcggagtatggatcaaaccgacgttgctagcatcggagcaggactcagatgaagcgacaaaaaaagctcggcctaagaaattcattcatattggcacagacaggaaacataaagttgtcgttgcgctaaccagcattaaaacagaggaggacgactgggctaaaatggcctgcaataaatctaacttgtcccgttcaaggtaccagcggatttctgaactggttaatgcaacattgaaactatctcctaaacctgattatgttttattccctgagctttcaatcccgttacgctgggttaacagtattgctgatcgtttgagttcggcgggtatcagtctaattgcgggaacagaataccgccacttagacgataatcaactgaagagtgaggccgtacttgtcctttcagataacagactcggctatccagcgagtgtcaaaatatggcaacccaagctggaacccgccgtaggtgaagatgaggcattattttcaatttatggtaagtcttgggattcgacacttaatgttaaacaacgtaagccggtatatattcatcacggcgtcaattttggcgttatgatttgctctgaactccagaatagtaaagcgaggatccgttttcagggcgcactcgatgcattaatggtattgagctggaataaagatctagatacgtttgcatcgttgattgaatcagcagcgctggatattcatgcctatactattttagtgaataaccgaaaatacggcgatagtcgcgtacgttccccggcaaaagaaccctttatgcgtgatattgctcgtgtgaagggcggtgataatgactttgtggtcgctgcaacgctggatatcgactcgttaagggcatttcagagcagggcaaaacgctggcctaaaggcggcgataaattcaaaccgttacctgaaggattccagttggcaaagaaccgcaaaaagctaccgccaaaataagaaactgatttcgctattaataatcagggtattttgcgtgagatgttggtaaacatgatgtagcccttgccactcatgaccaatcgcagtatctttctcccgcgcctgcaaaatcaggcgtcgggattagcctcctgaagaaatcttatcggcgacacatgacgcgccagcgtctttttttgtgttgttcgcacggttacatc (SEQ ID NO: 15)
18	pLG020	ttttcaaaggagtttcgctttccaaatatacaagaaatcattatttctaaaggtatctataagtggatgattcgttttattggaacagttgcattctcgttaattaaagcggctgcttccgaccggcgaatggtcattcagaagctgagaatgtggttatttttaaagaggaattggcatgattattagccttgaagagcttggccttgcctaccgaaaagcaaaagtcgatctgtactattcatcccatgtttcgctggaagcaattgcgtcttacgaagagtccctacatacgaatctgacggttctgcaggaaaaaatacaaggtgacgacgaatcatgggtggaagagaatgagttcactggcaactggtttctggccacaaaatctgtagacatgtcttgctgggaacagcagcgagaaccgcaagctaacggtctcatattttcctcacctgctgaaaagtgggcatatgcttgcaacccaatggctgataaaaacgaacaaaaaaaaatcaaagccgagtttcgagtaatggctcaatgcagtctggattttcatgttctctcgactctttggatgttaaaagtcgggcatctttttgatgccaaattatctacctgtgcttacggtaaccgcctgcgccgtactctagatggaaaagacatcaatgcactttcaattggttcttttcaaccttacctcagaccttttcgtgattggcgtgacaatggcattaacgccatgcggagcgcgctaagtgaaagcaaaaaaatcgtggcactcactgctgatgttagttctttctatcacgaactgaatcccgggtttatgcttgatccaaccttcgtcaaagatatttggagttggaactcactgctgaacaaagcaagcttaatcgattattcattaatgcgttaaaagcatgggcaattgagactccgttgaagaaagggttaccagtaggtctccctgcttcagctgttgttgccaacgtagccctgatcgagctggatcgcgttattgagcagcaagtcgcacctatatattacggacggtatgtagatgacatcattctggtcatggaaaatggtgcgaatttccgttccatggcagagctatggcaatggttgttcgcccgttcttccggcaaactggactgggtaaagggcgaggaaaacaaacagatcagttttcaaccaaactacctgcatgacagccagattcgttttgcaaatgcgaagaataaagtgtttatccttgcgggtgactccggaaaaaccttagtggaagctattgctcatcagatttatgaacgagccagcgagtggcgagccatgcctcggttaccgcattcctcgaacaatgttggaactgatttgcttgctgcaactcaaagtaatggcgaagtcgctgacaatttgcgtaaagcagatgcactgactatgcgtagggctggttttgccatcaaactacgcgactttgaagcctatgagcgtgacctgcaaccgggcacatggaaaggccatcgccaggcatttttcgggcatttattgatcatgttgtggtgctgccacaattctttgatttatcagtctacctaccccgagtgatccgactggccacggcctgtgaggactttgtcgaactgcgcaaacttatcttagcgctcgagaatatttgcgatgaagttcgagaaaattgcctccttaccatcaaggcgtgtcctgatgatcacctcccttttgaagcagagattattggcaaatggagggctcagctttttagcagtgtgcttgaagctatcgttgcggcatttcctccgcgtatttccaaggtgggtaagcaaacctggaatgaccatttaaaaaactggcacgcccggtgtgggctagacattcaatattcgggtcgtgatttttcattaaagggctaccaagaacagcaggcgagattattctctttcgacttagcgcacatgccattccgctttattggtctaccaaaagagatgattgctcaacggggcatacccgctccgaaaacagtagcccactgtgcggaagcagcagaattactgcctgatattgtcgttttgggtaatcaggttgtagcaaaatggtgcaaatttaaaatcattccacatggactgctatttgccacccggcctttcagcctgccggaactctttatcctaaacaatgaggcttatacagcttcagctcagcaagaaatgcgagctattattttcgctgttcgcggttttgtactcggtaataaaacaccttgtgtcgataaacaaggcatattgcaaatccctgacggccaatctgctggaaaatatggggttgccatatctagctggaaaacgtccatgtcaagctggactgcggcggtcatgcgttcagccgatccggatgcaaaccgttacgctcgcttatgtcgcttgcttgatggtgtgatagcccaaccacataacagtcgttacttaattctgccggagctctcactccctgcgcactggtttattagaattgcccgtaagttacaaggtcgcgggatttcacttgtcaccggcattgaatatttacatgccagtaaagcaagagtacgcaatcaggtatgggcttccttgtctcatgatggattgggttttccttcactaatgatttaccgtcaggacaaacaacgcccagcactgcatgaagagcaggaattacaacgaatagcagggctagaaatgaaaccagaaaagaaatggacaacgcctcccatcattcaacacggtgatttcgtttttccttgttgatttgtagtgagctgaccaatattagttatcgcgcagcgctgcgtggcaacgttgacgcgctgtttgtgccagaatggaatcaggatactgaaactttcaatgccttggtcgagtctgctgcgctagatatccatgcttacatcatccaatgcaatgaccgccagtatggcgatagccgcatccgaggccctttcaaagatagctggaagcgtgatgtattgcgagtcaaaggtggtattacagattattgtgtaataggcgaaattgacgtacattctttacgacaatttcaaagtagctatcgttctcctggtaaaccctttaagccggttccggatggatttgagatagagcactctcgaaaaatgttgccagaagcataagtaaaattggaaaaaaatatcgatgcaggttattaaagatgaggcaacatgccatagtcaatcataacctg

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		cagatgtaatttgaaactgcatgttgagaattacggatttatttgtgtattcaccctcgcataaaaatgaagtagctttcatattccacactactgatacccccctgaaaatatataactaaaaaaaacaattttaaaacatgaggtaggaatagcaatctgactgtgatgtagttatttttttgatgaagataattaggtgctcgttgttc (SEQ ID NO: 16)
19	pLG021	ccactacaccggtgaccatgatttattgatcgttcctccttagtgaaccgattctgcccgcttaaccttacccccctgggggtagatgtaagcaacggagttctgttcgccgccaggtcaaaccacgatgacttgatcggcaggacagggaccacaatagaccttcaggtcggaatcagggatagaaggggacatgggcgaccgacagatatgaagatatgatggctatggcggcatctctgcccaccctcaggtccaaagcgaaaggaatcggaatgccccgtatcaacgttgagaaactgctgcttgagatcgaaatcgacaaggtggcagagcgattgggtatggcgcttaggagcgaatcagctacgcgcaagctcacgctgtgcccgttccatgacgataaaactccttcccttctaattgatacgagcagagataattctggacagcattaccactgctttgcctgcggtgaacatggagatgcaatcgatctggtgaagggagttcttcatatcgatttcaaaggtgcattagagtggctgtcaccaaactctactaccaccccctgtaaatagggcgagaaaacagaaggctatgcagcctgagcagccagaaggctcagggcttgcgcaagcttataagttatacctgttaagcaatgacaagcaacgactagctaactgggtgactgatcgcaagcttgatatttttttgatggaagatgcaggattcatatacgcacacaaaaactcactatctaaacaggtttcctcaagaaaagatttggaacgaagcgtgaattagcagcaacattggaagaagcgaacctaatacgcaaaatccttccaagctcggggttccaaaactactatttaaatctacagtcaatccacgacaacaactatatagactttttttcaggggatcgaatcgtattcccgataagagacgatcagaaaaaactactaggccttgccgcccgggcggtagatgagcaaccagcaaaatacctattctcaaaaaacttttccaaaatccaaagctatttttagaatagagcaagctacaaccactctacgagcattggctaagcgaggcgaaacagatctacgcttatatatctgcgaaggatttttttgacgctctaagattggaaagcttgggatttcctgcagtagcagtaatgggaacatcaattagcaaagaacaaattaagattatgaaagggcttagcgacacgctcccttcaaagctagcctctttgacaatctgtatttgttttgatcgcgatgaagcgggattaagaggagcatccgaggctgtactaaaattcttaggcgctaatctcgacgtggtatttgtatggcctactactgctcagcttacaagcgcagaccattcaaacacaagcataaaagatcctgacgaatatttgagaaatttgtccgcgccgcaggccaagtcacttatcgatgtttccacctatggacctgtagtagcagtactagcaaatcagtttggtgtgcatgccgacgaactgcttgaaaatctaaagtggaacagtgccagtcgctctcgaaaatacaggtcatttgagaaaactcgtgctgaactcaggaaagttgtagccaacccccatctccaatcaagcgacctttttttaaatggccgaacagatcttgactcggcggctcaaatagaatggattgatttttttaagtgtcgacattgcgactgaagccgctccatcggaatgttatcttaccaactcaggcaccagactaaaccacgcccgactgctcgcctatatgggctcacgaagaggagagttgccctgcgaagaatcaaaatgggagcggttagatattgcggcaagtgcattcaatgtgttgctcgctgaacgattggctaatgaaatacatggacccatcgacccgttcgaggccgtatgggtgccgaggtccttcggcgcagaagagccgagattaaaggtgatgcctcaacctgaggatttaatagcgcatcagtacttactaaatgagctacttacagaacgctgggatgcttccgctctcggtgttacagcattcagccagtgcataccagctgtccgctattaccgcgaagaaagaaaaactgttacgacaggaatatctacccccctcagataacacccaacctattatacttgaacagacgctaagtttcgcctatcaaattgatatggaggttattgagggcaggcagccagcttcagatcagggaatgtttcgtccgttcctagactgctggcgagactttatgcagtcccttaaaaatcaagccaaatctataaattacgtgcatgttatccgcctcgatgtcagtcgatattacgaccgcatccgcagacacgtcgtaagagacagcattcaaccatttatacaacaagctctggaaactgtcgctgataatgcaccggcgtttgctgaactgatgaaaatacaagcatctgcggatgaagcagcggacaaatccgcaataattgtcgagcaattatgcgacatgctctttggctacccataccttagccctgataacgggagaattaataaatcagatcccttacgcggtattcctcaaggcccagtaatctcagcatggttaggctcagtggctttgtttccagtagatctcgcggcactggaaatgatgaacaaatacaatgtagacggggaaactcatctagggtatgcaaggtatgtagatgacatagttttactagctagcagctccgtacttcttgaggaactgagagagctagttgatcaaaaaactcggagcttagacctggcgttggtcgcgaaagctgacgctattccgccaatgtctgctgaggaatttgcagattatgcaaatcaagggcgagctttagaagcatctggtccagcgtgggaaccaccgttggctggcgatggtgaagcggggtgggagttttggtcaggcactccccccctcagatagacaatctgccctgcaactgctatcaaattgggagatatacaaaagcccaatagaaataatcttgcaaacagtgaaaacgtccttcctagctatggatttacgttctagcgagcttgcaaagggagcaaggctaatatggtacgttgtagcatccgacctcctctcagctgacattgatccaagcgatgcggcagatttagcgtgggaaatttatgatcgctattggaaggaatgtactgaggagtgtgggtggcagttaaacccggatagtttcggatgggaggcaccgaatctgttcgcacttgagggactggaaaagcttatagatcataaaaatagcctccaatcgggtttaactgctttagaaaataccgttcggcacaaacgcatctctttcctagctagaaccgtgcttggggagcggttcaaactgcatgctcttgaaagcagctctacgcttaagcaccagatagataaaagactagatctcctcgaatggaaagcgtcaaaatcgtgcggaatgcccgttcgtagaactaaatcctacgcagagcgatcaatgtatattcgctcctggcaacccttcaactggttccatgccgcagtagaagatttcatgctcgcggatcagtccagcggatccgacccattgagttcatatgtcactcagttccaatctatagaaaagagcatcagacctaatcacgccgcttcttatgagttcttccggtatttactgccatccgatggcagcgatagcgatcttgagtttttctcaaaaacagagaatcgatactccggcttagcaattcagattttggttgcattagtccctcgggaaagcataatacagattctctcaaatagagcgcgcttactttgtcctctagaagctggtaaaaaactattagtcatgccccctcttcctggcgtcaatcagcaacgtatagttgcttgccagatcgatagctcctcagaaaacaaaatcaaaaaaatcagctcgtttgagtgctatgaaatagattcaactaaaaccaataccacatctctagactttttttggtgcaaactctgcgggcgtagttgtgcttacacccacatggaacaccgaagcccaacctcaatccgccatacttcgatcaaactcagaagtcccgaaaaatcttttgttggaggtatttgagaaaccgtcaaccggtttcccttccgctattcagggattgaagcacgtagcctcactatatagagccattgtggtaataatggctgaatacgagaggcaaaatgatggtttagagcttatacccgcttggccataccttgccacagatatgacctctgggaactgctacctaatttgtgagggcgtaacgaaaggagaagtaggaaaccgagcatttgtaagagacggtgggcgggccctaagaaccattgagataccgatatacgaagcccagttgtggcgagccggggttgcgctaagcgattacataggcctgcacgacgatattgctaaatttagctcctccgaatccgaaatacctttggatgcgacaacgcttgccgccccgtcacagtacgtgctacgaagccaacttcgtaaactgaggggtgcctttgctaactcacaaatagggcggcgcgttatgcccccaagttttcttccggcaagtgttgaacgtgcgcttgagttattggagcattttccggaagactcagatagtacaaagatgcagctaatgcatctgcttgccactgaaaccgaaactgcgggaatgcgcgtccgctatgagaaaaatattgaggtcacagagctcacggtatttctacgtgcggtcgccgacagggttctaacgaaactacccttaagcataggtgaggtcattgctgcaccgactacagcagtcagtggcctgaggagagacctgagtggggtcttgacccttgccagaagcatatggtcgatggatgaagaagaaaaactctctccaattttgcgtggaagatttttcgagctggaattgtaggtattggtatcgctgttgctctacggggattatagcttcactaagaagccacgggggtttgcacgctttgagggatttgattttccagcggaatgggagcttcccccctgccacagcagttttatccgaaccggcgacaacagataaaaccactgatgaaaatgtaagcctcctcgaccatttccgggtactcgtatcacatctcggacaccgaatgaggttggacgacaacggcgagccacaaatcccagaagaaatcagcacagaaataagaaaatacgctacagcattagcgggcctcactactaaagactcaactgcggtggacgcaagcgactggcctttctttgatatcagcgaaaaagtttttgataccctaaatatagaattattagagaacgtcagcaatctaatcaaaaacttagattccgcgcttggtctccaggtaattttggttacgcaacaatcatacggcttcaatgctcaaaccaaacgcttcactgactcaagaggacttgcatgggatataaagccatggatgatctcgcaatacccattgcgtgctcgccacgttgaggagtgttttgatcaagaccgtagaatcgtacgtgtatggagcgagatttacgaaaaaaacagtcaacgcctgctttctatatcagtactaggcgagcctttcgcatcaattgcactatgtaaggacttggaatcgccttatgccgagactaaaaatgtagacagcaagcacaacactgtattaggtcctagcgagcagggttctgaaagcgcacccatagatatttcaccgattcttga

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		aactgctgagcctgaggccgagactgccttagcagacacacaattaataccaaccccaaaccaaactagcactgaagacagctttg ataaaatagatactgagcgtaatacaacacacaataaaaaactaccgcttaccgacgcaacactcaacgcccgaaagaattcattt agaaatagccagctaacagcctggagcgataggaagtccaataaaaaccctgcccatgttcgggtagctctatttcagtgggacca agagctgagctatgcacaccctatggtggaggccaccccacaaaaatggcctttcagttccgtctgtaaaccagcagttttaaaag aacttaaacgcctatataactctccctatcaagcccttttgaatgcaactgaatctgccggtcaacaccacctatggaaaaacgaa aatatttccctacccagctggggtgagcttcgtcgtcggcgattattgctcaacgcagtgaacgcatgccagtcatttggcgtgga cttattgatacttcctgaatactcagtccgtgcagaaactgttaagtggttaaaagaagagtgcttacccggaaagacggtagcgg ttttagcaggaacattttagctttcgactccggtccgccccccctaaaacaaagcgcgagcctcaacctcttgtggcccgtaccg cgtgatattgccgaatgcctcaaaccgcttgcacccaaaacaaatgaagatgctatgtccttgagtgacaagattgacaagggcat tgtattgcaatggggcagatcaaagaaataccgatcagtagctctaaatgagttcatccggcctggaactgatcctctcacccccc tgttcatgcccggaaaaataatagatgaattgagacgtgcaaattgggatctggacgctgatggtgttgttaagttgctagccaac acagagttgccacttgcgaatttcatggagctgatatgctctgagattttcctgttcacgagcccaaccaacattccagagatggc aagagattatgtttcaatgtgtgcaagatttggcttcggcgctgcagaagctcaagtctgggcggatctcaaactactatctaaat ggctttcggtctgttccaagcctggtggtgccgactctagacgatcaattttgatcgtacctgccgcgaccactcgtactgctgat tattggatagcaggccaagctggcttgcttgccgccggcactacaactgtatttatcaatggcgtaggatctgggcttaagggtgg cagttgttttattggcagagagagctggaaaacaggggctggttctcacggttacattgagaccattacgccataccatggctggt caaaaggaatttactataatagcaaacatgacccactgagcgaaattgatcaagcattggtgatcgcagatatcgatcctcataac atgcttgaaggcaaacctagacctcagatgctgccagttcccttacagctagtggcatacctaccaatcgttgaaactgtcgacga aacaagcttggaccaaactctctgtgacgcagttcaggttgaccataacaatattgcaagaattaatcagggtcagcgattgggtg gacgacttaaaagtcgaaatgagttctggcaacttatcacgcaaagtataaataatgatgtcgacaacgactttatcattaacttc agtaaatactttactgatgggaaagcgattcttgagcgagcaaactctttcttcaacaatggacaccaacagcctttttcatcggt agttaagctagacctgctctgctctccggcactttacgactggctagaggccgatatgacgttgcgggagggtgaggcgttaccca acatctcagtcccttcatggaccaaataacttcggatagattacgagcccctaggataaagcctgtcgataggggctggtcacatt ccccgcagcagggcggtgccgataatagctgctcacatagcttagagagcagtcaccgcttggcactttggagctgggagagcgtt ggcatcgtagaatcgtcggcagtgaaaattcggtacagctacggtacggcacctagcttctgtcaactaattcaaactacactcaa caccatatactacggtgcctccagctatgccaacctacgttcagctaagaacgacttcactaggcatacatggtcgcccagcaact cataatcccttggtcgcaggttcgagtcctgctgggcccaccaagctttgagagccgcgctttgcgcggctttttttgtgaagcca agcactcagtttggtccgaacaccacgccaaagtgtttttcaagatcgcacatcccagaccacacgatgcacagacttcatgttga agcgccgtcttcagaaataagctgggaaaaggtcaatagctttcaatttgtagcagccaaccgtgatcacaggtagagcacgggtc gatttgatcttgcaatcctttgggcagcaagacccttgggctgttcaccggcgttgctgcacaaccagccacgctggaatcattac tgtcatcaaggttgagaa (SEQ ID NO: 17)
20	pLG108	
21	pLG023	atccctgaattccccgaaggtgaacaatccactgttcacccttcaccgtatattaacccgttatcacactgaaattaaaagagaaa aatgaaaggtgaacagtgtgaacaatcaaatcaaaaaaactttctactcccactatagcctgactggtcgtctccaaaacgagcgg aaaagcatcaacaatgaatagttaactgttaactccgcgccaactcattaccacttaactcaatgatattaaatggaaaactatcg aaatgaatactctgcaaaattaaatgcaaaaaaatatatgccagtcaaatttcgttacgcactctcttccaagaaagagataaatg ctttatacgtccaccatactatgttattttttaatacggctctgccttaaatctgtgaggttgtttcgcctcgaagtatcttatg ttagcacatcacgctaccaatcagcggttagttacttgacgtaactgttaattggctaaagtttgcatagagtgattgggcggagc cgtaaatttagtccataaatacagtaacgaggtagagagtgtctttacatgacaagctactgatgcttagtctcaattcggcgaat aaagaagaagatgagacaatcccggagttacctaagttagagcctcagccctatcaagctggaaataagttgaaatgggataataa agagctgaaaaatcagcccatcacttcaaagaatgacattaatgtaatatgcaaaaaaattgaaaacaaaagcattgtaattacat cagcaaacgatgtagccaatctgttagaagtcccggtcggacaattattatttattttatataataaaaaagataactatagaact tttgaaataaaaaagaaaaatggaaaaagtagaatcataaatgcacctcaaggcggtttatcaattctgcaagagaaattaaagcc agttcttgagtacttttatcgccccaaaaaaccagcacatggatttattaaggataaaagtatattaacaaatgcagaaaaacata caaagaaaaaatatgttgttaatgtagatttagaaaattattttggttcagtcactttcgctagagtatatgggatatttaaaagt aagccatttaatttctctcatcctgcggcgagtatattagctcaactatgtactaaggatggaaaattacctcaaggagcatgtac ctcccctgttctagcaaatttagcatcagcctcactcgataaacacctaacccaactggcacgtagaaaaaacatcacatatacaa gatatgcagatgatattactttttcatttaatcaacgacaagtcagagaaatcataacgctagataatgaaaataatttttgaattg ggcgaggcgattatctctgtgatagagaaaagtggcttcagcataaacacaagtaaattcagagttcagaaaagaaatgaacgtca aaaagttactggtctagtggtaaatgaaaaagtaaatgttgagcgtaaatatcttagagttactcgttcattagttcataaatgga gagaagacaagttaacatcagcattgttgtttgttactaaaaaggttttaaggcaacaaataacgaacatgctatatcaattttt cgcaatcatatttatgggcgattgagttttataaaaatgatccgtggtgaggacttcccgttatatcttaaattaatggctgaaat gagtcatcatgatcctttaaaaacaaaagaagggcttagagcaatgaaagaaactgaaacttacgatgtatttatttgtcatgcaag cgaagataaaacatccatcgcaattccaatttacgaagaattaattaaattaaatatatcaacattcatagatcatgttgaaataaa ttggggcgattcattaatccaaaaaattaactcagctcttgtaaagtctaaatatgtaattgccattctttcggctaattctgtag ataaacattggcctaagaaagaattgcattctgtgcttgcaagagaaatcactgaaggtgaagtaaaattacttactcttgtaaaa gaagcagatgaagcaatagttgctgaatctttgccgctcttaagtgataagctttatatgacctataaagataatccggcagaagt tgcagataaggttcgtgcgcttttaaacaagtgacagctactgtcaaatgtgtataaagtcattgatattttatataaaatcaatg gattgcaatccatataagattccttatgcatcagtgacccggtgctcgcccggtcactgcttcagtcccagcagaactcagacgag gcgcttaacatctaacgggatgccaacccgacgtttggttttatcggctatctagcctatatagaagca (SEQ ID NO: 18)
22	pLG024	ctattgtgagcgagaaacgcgctactactatatatagacagacaagatgcacttactgaataaatactcataacggagaaaccagc tgtatagtgaacaatagatttccagtagcatatttttacttcacttttagttattaatatgataatcataaactacggctctgcct taaatttgtgaggttgtttcgcctcgaaggaactaatgttaggacatacgccaccgttcagtcgatggtaacgcttcttaactagt ggtccgctaagtgatgcgcaaagtgattgggcagagccgaaacgtttacaatccgataggagttggttttgtcgctacatgataaa ttattaatgcataacttcgcattagccaataaaaaaagccctgacttcatatctgaacttcctcaaattgaacctaaaccatacag caatggacataaaattaaatggataaaccacacacttactagcactgaagttactcccccctgataacctgattaaaatatgcatat tgattgagtcaggggaaattgctataacatcagtaagtgatattgccaatttacttggagttcctgctggccaattactttatata ctatatcgtaaaaaagataattatcgtacttttgaaatagaaaagaagaatggtaaaaaaagagtcattaatgctccttgtggcgg tctatcgatactccaaacgagactaaagcccgttcttgaatatttctacaggccaaagaaatctgctcatggttttataaaaggaa

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		agagcatcattactaatgctgggatgcatattaaaaaaaatttgtcgtaaacattgatctagaaaactatttcgaatcaataagttttgctagggtttatggaatatttaaaagtaaaccttttaattttgctcatcctgcagctactgttttagctcagttatgtactcacaatggaaaattacctcaaggtgcgtgtacatcgccaatattagcaaatattgcatcagcttctctagacaaacagctcacccaatttgcaggaagaaaaaaaatatcttattctaggtatgctgacgacataacttttttctttcaatcagagaaatattgatataatcaaaaaaaacgacgacggaagttatagtcttagtgaaactatagacaatattatttcaaaaaatggctttaaaataaattatgataaatttagagttcaaaccagaaatacaagacaaagtgttactggcttagtggttaatgataaagttaacattaacagaagatatataagaattacacgttcaatgattcatagatggacagatgataagctaaagtatgcacttctctttgctacagaaaaaggatatcaggcaaaggataataaccacgcaattcaaattttccgaaatcatatttatggaaggcttagctttataaaaatggttagagggaaagactatccaggatatttaaaactgatgtcatacatgagtcataacgatccattaaaaacccaagaaggattgcgagcaatgaaagaaacagaaaactttgatgtttttatatgccatgcaagcgaagacaaaaaagacattgcaattccaatatatgacgagttaactaaacttaaaatttcagccttcatagatcatgttgagataaaatggggcgactccttaattgataaaataaatgcagcactagttaaatcaaaatatgtcatcgctattttatctgctaattcagtcaataaggaatggcctcaaaaagaattaagagcagttttagccagcgaaatatcgagtggcgacgtaaaacttttgaccttattaaaaaaagaagacgaggaggtcgtaaacctatcattacctttacttagtgataagttttatatggtctatgataataatcctgaagtagtcgccaacaatattaaatcactcttacaacgataattctctcacaaaagaaaatgtgcagattgatgcgtattaagtattaatctgcacatacaaaaaaaataataaaataatacatttttcataacttgtaggtaacaacaatatatgtcgtaacgaatatttggataacctctataccctattaaccaaccaattaactctatgtaatctcgcagcc (SEQ ID NO: 19)
23	pLG025	cacgtaaatatgaaaactgttagcccacatagcccaacaaaaatatttgatagttaaccttctgttactaaagaaaacaggaaagtaaaagtgggctaaagcttatgcgccctcgatgttgggctagccccaaaaacggtaaatttagcttaagtgcataattggttagctcaaaagcattatttttcatttaaataaattagttaattggtcttgtttagatgattcaactgggctgactactttctttgtatatactccggataaattttcccagctaacttgcctaatcatcactctgatgccagaaatgaacagaacgcaaaccatctataacttattgaggattttgaaaaaaattgattgggggcttgagttatatgatgactatgctaatttaatacggcacatgcaggtagatttgttggttgtggtatcgcaatcagtgttaacaaggtcgggagtattcgccctctgactgccgtcaagtcatcttggcgtcaccgttaaatgcgtaagagtacctgcatgtgcattaacataatcaataatggaatttactgttatgtttaaacctacctatctggcaaggctgcaggcttgttgtaacaaatttgaactggctgatttgcttcagattaaagttacatttctgactaatgttttgtatagaataaggccagaaaatcaatacaaaaaaatttactataaagaaaaagtctggaggagagcgggagatctttgctcctgatgaaaaactgaaagatattcaacaacgactttctgaacttctatatatatgccaggaagaaatttgggcaaaaaataatattaaacaaaatgtatcacatggttttgagaagaataaaactataattacaaatgctgagaggcatcgagataaaaatattgtatttaatattgatattgagaatttcttcccatcctttaattttggtcgcgtgcgaggatattttattgcaaaccaaaatttcaagttacatccaaatgttgcaaccattattgcgcagatagcctgcctggatggatcgcttccgcaaggaagcccttgttctccagtaataactaatcttatttgtaggattttagatttcagattatcaaagctagcagtcacatatggttgtagttacagccgctatgcagatgacattacgttttcaacaaacaaaaaaaacatccctgatgcattagtttctaatgagaaagaaaacgaaccaggtaagatattggtagaagaaattcatcgtgcaggcttcactttaaaccataataaaaacagagtgtctaggtgtacatcaagacagcaagttacaggtttaactgtaaataaaaaaataaatgtaagcagagagtatataaagaatacaagagcgatggcgcattctttatactttgaaggttcgtatacacttattgagaaagatggaaaacatagaaagggcacccttagtgaattagaagggcgatttgcatttatcgatatgcttgataaatataataatgtggaagcaaagaaaaatgcgcgtcctgagagatatgtggttaaaggatttgggttggattttaagcagagacttaactccagagagaaagcatacagcaaattcctatactataaaaatttctatggaaatgagcaaataacaatcttaacagaagggaaaactgacccggtttatcttaagtgtgcaattgattctttgtttttggattaccctcagttagttagagaggaaaaaaacacaaagaatagagtgttaaaagttaatttatttaaaaccaatgacaagaaaaaatatttctcgatttgtctggtggagctgcagactattcgaggtttttcagacgacatggtttactttgtaaagcgtatgaaaaacagcctcctaaaaatccagtgataattttattagataatgacacagggccatctgacttcataaatcaaataataaaggattattcgcatctaccaaaaaaagcggaggatgttagaaaaggggcgttttatcacttagagagtaatttatatgttcttttactccgttattaccagggataactattcttcactagaggatttttttgaaccaaaagttttgcaaatgaagtataatggaaaaagcttcgataaaagcaataatcatgacagttctactacatttggaaaagatagatttgctacttatatagtaagggaaaatagaaaaactatcgatttttcattattcaaacccatacttgattcaattattgaaatcaaaaaacattttatcaatctacacccatcaaagtgatggttatgaaaagagataaaaatgctgatgtcaaaagaggcttatgctcggcacagtggagtgagctgccaaactgtcgatgactgggtagccggtggggcggaagtagttatgtcccgtagcaaggttaagatttgctcttgtgtgtggggaaccttagtcaattactttcctggcgcactgtgttagattttgtaaaattttaaaagactaaagatttaatatcacttctccatggaggttgtg (SEQ ID NO: 20)
24	pLG026	ctatacgccgttatagctgaattttccggtgatttcagggcacattaaccaatttagataatactatagtaatggttgggctgattttttcaagaacaaaagtaattttcaagctttgtaacatgttgattttccgcttttcgctcaagcgagctttcatctttgcaagcccatatgttcgtttttcaagcgattattcagatacgttaacttcccatggcagtgcatgactatgctgcatgaaatcgcatgatcgatcgaggatcgtctatgcttagaccagccagaaatggcgggcttttgctcatgtcatgcagctgcatgaaaaccactgcataaagtgggcaggcgtggcggggatacgagggcgcgctatcacgtaaaataggcaaaatacttctggaaaacagaaagttgaagtgatatgttcataaacacgcatgtaggcagatttgttggttgtgaatcgcaaccagtggccttaatggcaggaggaatcgcctccctaaaatccttgattcagagctatacggcaggtgtgctgtgcgaaggagtgcctgcatgcgtttctccttggcctttttcctctgggatgaagaagaaatgacaaaaacatctaaacttgacgcacttagggctgctacttcacgtgaagacttggctaaaattttagatgttaagttggtatttttaactaacgttctatatagaatcggctcggataatcaatacactcaatttacaataccgaagaaaggaaaaggggtaaggactatttctgcacctacagaccggttgaaggacatccaacgaagaatatgtgacttactttctgattgtagagatgagatctttgctataaggaaaattagtaacaactattcctttggttttgagaggggaaaatcaataatcctaaatgcttataagcatagaggcaaacaaataatattaaatatagatcttaaggatttttttgaaagctttaatttcggacgagttagaggatattttctttccaatcaggattttttattaaatcctgtggtggcaacgacacttgcaaaagctgcatgctataatggaaccctcccccagggaagtccatgttctcctattatctcaaatctaatttgcaatattatggatatgagattagctaaactggctaaaaaatatggatgtacttatagcagatatgctgatgatataacaatttctacaaataaaaatacatttccgttagaaatggctactgtgcaacctgaaggggttgttttgggaaaagttttggtaaaagaaatagaaaactctggattcgaaataaatgattcaaagactaggcttacgtataagacatcaaggcaagaagtaacgggacttacagttaacagaatcgttaatattgatagatgttattataaaaaaactcgggcgttggcacatgctttgtatcgtacaggtgaatataaagtgccagatgaaaatggtgttttagtttcaggaggtctggataaacttgaggggatgtttggttttattgatcaagttgataagtttaacaatataaagaaaaaactgaacaagcaacctgatagatatgtattgactaatgcgactttgcatggttttaaattaaagttgaatgcgcgagaaaaagcatatagtaaatttatttactataaatttttttcatggcaacacctgtcctacgataattacagaagggaagactgatcggatatatttgaaggctgctttgcattctttggagacatcatatcctgagttgtttagagaaaaaacagatagtaaaaagaaagaaataaatcttaatatatttaaatctaatgaaaagaccaaatatttttttagatctttctggggggaactgc

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		agatctgaaaaaatttgtagagcgttataaaaataattatgcttcttattatggttctgttccaaaacagccagtgattatggttc ttgataatgatacaggtccaagcgatttacttaattttctgcgcaataaagttaaaagctgcccagacgatgtaactgaaatgaga aagatgaaatatattcatgttttctataatttatatatagttctcacaccattgagtccttccggcgaacaaacttcaatggagga tcttttccctaaagatattttagatatcaagattgatggtaagaaattcaacaaaaataatgatggagactcaaaaacggaatatg ggaagcatatttttttccatgagggttgttagagataaaaagcggaaaatagattttaaggcattttgttgtattttgatgctata aaagatataaaggaacattataaattaatgttaaatagctaatgaacagccctaacgttatgaacgctaaggctgatttttcg (SEQ ID NO: 21)
25	pLG027	aattccccgaaaatccgcccgtttttactgaaaaaagccatgcatcgataaggtgcatggctttgcatgcgttttcctgcctcatt ttctgcagaccgcgccattcccggcgcggcctgagcgtgtcagtgcaactgcattaaaactgccccgcaaagcgggcgggcgaggc ggggaaagcactgcgcgcaagctatgtgaggtgatgtgtaatacatatcacgaatagcgtaggtagctgttggctttgcctgatca aggtgacagtatacatatcttaaaatataaatatttatgattatttatttgaaagaggttgaataatgatttttgatgaaaaaaga catttatatgaagctctgctgcggcataattattttccgaatcagaaggggacgatttcagaaatcccaccatgtttttcttcaag aacttttacaccagaaatttgtgaattaatagtttctaatgagccggggaaaagaaaattacatggatacgattgtgtcgaatact catcgactaggtataataactttcccagagtattatccttaattcacccaagagcatatgcacagttagcaaagcatttgtatgag tcttgggatgagattcgaaaaatcaaagaaaataaaaacagtatgattaaacctgaaatgcatcctgacggtagactttttatcat gaattatgaggatgcagaaacaagaactgtaagggagttaaacgatggatttggaagacgatttaaagttaaaactgatatcgcag gatgttttaacaatatatattcacactcaattccttgggctgttgtcggtgtgaataaggcaaagacatcaatgaataagcataaa aatagccaagatgttcattggagtgatagattggattattatcaaagacaaacaagacgaggcgaaactcatggtgtccctgttgg acctgcaacgtcaagtattgtatgtgagataatattaagttccatagataatattcttgagaataaaggattcttattcagacgtt acattgatgattatacatgttattgtaaaactcatgatgaagcgaaagagtttctccatgttttaggtactgaactttctaagtta aagttatctctaaatttgcataaaactaaaattaccagtcttcccagtacattgaatgatgattgggtgtcgttgcttagtattaa ctctccatccaggagagtattcaggaataatgactcggatatattatctgcatctgaggttataagctttttggattatgcggtac aacttcatctgacgaatgggggcggtagtatattaaagtatgctatatctttaattattaataaagtagatgaggcgtcagcaaga gagatgtacgactacgttttaaatctgagttggcactatcctatattaattccatatttagatgtattgcatccaaagattaacat taatgatgaggtcaggttaaaacttaatgaggttttgaattcctgcatagataataagttttctgatggcatggcttgggtgttgt attattgcttaaaatattccattgatattgacagttgtctcattagtaagattttgaaaacggtgattgcctaagtatttgtatt ttggataaaactggaagatatgataaggaaatagaagaattttctaaaaatataatttcattggattatttgtatgaggttgataa atattggatattgttttatcagcgattctattcagggaaaggatataatccttacaatgatgattgttgtttcgatataatgaaaa catatggagttaattttatgcctgatgatggttatcaaacgaaagctgaacactattgtaatatagtaaatagtccatttcttgag aatgatgaacaagtaataagttttaacgattattgttcataatttataattagcctccg (SEQ ID NO: 22)
26	pLG028	cctgtcaaaaaatccccgtaaatcccgctatttttaacgaaataagccatgcatccataaggtgcatggttttgcatgcgttttcc cgttcctgtactcccgaccagcgtcagtcccggcgcgacctgaggtcacctttgcacctgcattaaaagcggccccttaagcgggc aggcgtggcggggagagcattgcgcgccaaagcgtattgatatactgccagcatttttttgatactcacacccatctacaggagtag gtcactaccgatgtagagcttttccggattcagataaaaccacttagcatcggagcaaagtaactcaataccgaacaataaatatg agcccttcgtgaaaccgggtaaggtcaaactcataaaccaacaaaaggggaaaagtgggatatgtgaggcgtgtatgatttttatt tattgggcttcgttaaaaatggtgatttaatagccctttaaatttatcacttttttaactaactccgagggtttatggttattttttg atgaaaaacggcatttgtatgaagccttactgaggcacaattatttccctaaccaaaaaggttcaataagtgaaatacctccgtgc ttttcttccagaacattcacaccggaaatagcagagctaatttcatctgatacatcagggcgcaggagtctacaaggttatgattg cgtggaatattacgccaccagatataataacttcccaagaacgctgtcaatcatccatccaaaagcgtactcaaagctagccaagc atatacatgataactgggaggaaatacggtttataaaagaaaatgaaaacagcatgatcaaaccagacatgcatgctgacggtcgc atcataatcatgaattatgaggacgcagaaactaaaaccataagagagctaaatgatggttttggacggcgatttaaagttaacgc agatatatcaggctgctttacaaatatctactcacactctatcccgtgggcagttataggggttaataatgcaaaaatagccttaa atactaaagtaaaaaaccaggataaacattggagcgacaaacttgactactttcagcgtcaagctaaaagaaatgaaacacatggt gttcctattggtcctgcaacctcaagcattgtttgtgagattattttaagtgctgtggataagcgtcttagggatgatggatttta tttagacgttatatagatgattacacatgctattgcaaaacacacgatgatgctaaggagtttttacatttactcggtatggagttg tctaagtataagttatcactgaacttacataaaactaaaataactaatctcccaggaactttgaatgataactgggtttctttgct taatgtaaattcaccaacaaaaaaacgttttacagatcaggatttaaacaagctaagttcttctgaagtaattaatttcctagatt acgctgtacaattgaacactcaggttggtggtggaagcatactaaaatatgctatttccttggttataaataatttagatgagtat acaatcactcaggtgtatgactaccttctaaacttatcatggcattatccaatgctcatcccatatctaggcgtacttatcgaaca tgtctatttagatgatggtgatgaatataaaaataaattcaatgaaattttgagtatgtgtgcagagaataaatgttctgacggca tggcctggactctttatttttgcatcaagaataacattgatattgatgatgatgttatagaaaagattatatgtttcggcgactgc ttgagcttatgcttgctagatagctcagatatatatgaagaaaaaattaataattttgttagcgatatcatcaaactagattatga atatgacattgacagatattggctcctttttttatcagcggttctttaaagataaagccccaagcccttataatgacaaatgctttg atattatgaaaggttatggcgttgactttatgccagatgaaaattacaaaactaaagctgagtcatattgtcatgtcgtcaataac ccatttctagaagacggagatgagattgtaagctttaatgattatatggcgatagcgtagcttttaggcctcatt (SEQ ID NO: 23)
27	pLG029	gcgttgaatggtataactatggcacggttaccgcatgttttgagctgtaatcgaagttatgaaaattgctatataaagcggtcgct gttgtggagatacgattgcgggaagtgatggaaagagctataaaaagtacagaggatagtttaatgagggtattatgaaccgtcag ccgtttacttcagcagcacttaaacgaaacttaagtgaaagtgagaaggcttattattttaaaaaaaataatgttgctgagttaga atcattaattagtgatgccgttttaattgctaatgagaattttcgctctggtgtgagtgtaaagaaactaaatattaagggacgct gcgtttacactgcttcatgtttgaaggaaaaaataatacttagacattgcaatgcaaatttaaaatgccttgaatcgcttcgtccc aaacaacgaaatacaataattagtgagcttaaaatttatttggaagaaggtactccattcaaaatatatcgtttggatataaagtc tttctttgaatcaattgatttaccgcagctttttcagctcttacataacgaaacacgactgtctagacatacaaaaaatttgctag aatggtatcttaaatcgtgtgaaaggcttcactcttcgaaaggattacctagagggttagaaattagtcctatgttatcagaattg tacttggcacaatttgataatagtattcataggcatccagaagtattttattattcaagatttgtagatgatatggtaatcgtttc aagtggttgtgaatgtgaagcgtcctttatggaatttatacaagatgtattaccaaagggattggcwaaataaaaataaattaaaa atatctccatgcataccaaagagaagtaagggtttaaataaacaggataaattgcttcatgaatttgactttctagggtactcgtt ttctataatagacacacctttgagcaaagatggtgagattaatagctgttacagaaaggttgttgttaatttatctaaatctcgcc tgaagaaaattaaaacaagaatagctaggtctttctactcttatcatattaatggtgattttaaactattgctagacaggatttct

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		tttttgactagtaacagggatttaaatcgcaaaataaaatcgttaagttctttagaaaaaagcaagataagtacaggtatttattacagtaatgcgaagttagatgttgactccatatccctaaaaaaattagatgacttttgctatattgtgtgcaatctaatactgggcgtttgaatagtgttgcaaaaaaaccttttaatttgaagcaaaaaaaagaactgctaagaaatagttttagaaaaggctttgtggatagagtatatagaaagtataactttaagcgctatactgagattacaaaaatatggttataaagaaaaacattaaacttgataagaaagattatctcagggctttactatgtgatacactgcccggtgattgtccaattatttttcaaatgatggcttatatataaacttaacagaatatgatagagtttgtaatgatttgttacattttactccggtttcttctttcttaaaaaaaatagttaaccctaatttagactcttctattagtgtcgcagatcgccaccgagaaaagaagaaacaaagctccccatttggctattgtatagtaaaagatgcctttagccaaagacatctttctttaattcacccaagatctcaaattaattattcggaattttataaaacatactcatccgttatcacattaaatactttaaaaagtaatttttctattcgctacccacgtaaggtcgctaactctttctttttatatgaaaataatgctttggaaaaatataaaggggaagatatcgaaacaacaaaggatgagttaatgaggaaatattcatcctcttatttagttatggcggtttcaacaggatatataaactatttcaaagtaagatgtttattgagcttgagaaaagattctcggtgatgtggatgttagatgtatcacattgttttgatagcatatatacgcattcggtttcttgggcattaaaaaataaatcatatatcaaaaaacatgttaaacacagcaatcaatttggacaagaattagatacactgatgcaacgtagcaataataatgaaacaaatggaatacctattggttcagagtttagcagggtttttgcagaattaatatttcagcgaattgattgcaatattgagtcatgccttcttagtgaacatggatgggttaataataaagattatgttatattgagatatgtagatgatttattgtttttgtaatggtgagtcaagtgccgaagttattacaaaaataattaatgtgaagttaaatgaatataatctacaattaaatgtaaacaagcttaagaagtattctaggccattttgcactagcaagacaagtttgattgtcaaagttaatgaattaattcgcaatttagaaattaaactgtatgaaaaacgtgatagtggctttactttaaataaaataagaagtaagcatgatttaaagatatatgtaattaatcatgtcaagtctatatgcattgaaaatcaagtgtcttattctgatgtttcatcatatataatatcatctctttccaaaagattaatatcaataattgatatattacgagttcaagaaaatgaagatgatgtagatgtaaaaaaaaggattaaggacttaattttcacaataaccgatattatgttgttctttttcagtgttaacccaactgtttcatcatcttataaattatcaaagacaatggttgttgttaataactatttgaatgaaatatctagtgactatagtagtatttttatgactacgttagtgaatgctgcggaaaacattaattttggtgagaatgataatgggctgtttattgatgatttcatttcaattgaaaaggttaatttaatcttggctgctactttttttggagataattatcttataagtgacagtttttttcatggagttatacataaaaagaaattggactactttactataatctcactgctattctatttagaaacagaagatcattccgaaaattgaagtgtataatagagggtgaaataaaggaaatattaagttctaatatggatttgctgcaatcatcggaaaaggcacatttatttttggatgtcatgtcatgtccatttgtctcaatagagacaaggcgtttttatatagaaaatatctcaagagctatgagccaaagctgaacagaagtcatctggagattgagaatgatttgcaatctctgcttcaaacatattggtttgtcaagtgggatgagttagatattgtgaaaatgattgagaaaaaagaattgaaagaaagctattaatttgataaatatgagtcgtggtcagtttcaaaatacttacgtcatcgtcgtcggtgtattttatatcgattatgaagacgatttcgctggaactgaaatcggcttgaatgcttaaacttaagctaaaaaaacagtttgagaccaaagcctaaattattaggctttggattttcaggttcagttgagagtaattgctgtctg (SEQ ID NO: 24)
28	pLG030	cttgagtttgcgtaagataatttcgtgaaaattaaagcaattaatataaaaaatgtaattactagtgtgtacagatatgaaaaatgatagttataaaaccatatgaaaattgaagaaagagttcaattttgccttgtcagtaacaaataggtagcttattgaaaaaagataaaaaattaacaaaaaatcaataaattcatatagaataaaaatattaaagaaatgaaataagtgtttgcttcatcagttttagggatacattaaagtggttgataaagaaaaatattatactggattaataaaagatataaaaatagtagcttatgcaagattcaataaaatacgtcgtttaaagagaaataatttttaggattgttatctatttcggtagtttctatcttagttattatattatcaattgtagaaaaaatttataatataaaaacaatgagtttaattccattgtttgaaccaaatatagaaatatggttcttttgtatacttgcttcaataattattctttgtatatctattgcactctctactatgaagattgatattgaaatagaaaggttaaataaaagtgcagttgaacttaatgaagtaaggcggaaaattgaatttaatattgagaatagtaattatcaaaatagtacattgtttgataaatatcttgaaataataaagtcagacttaataaatcatgatgaggttgattataaaataaataagtatttagtcagtaaagttggtagtaagtttgcttattatcgaatgtattttattgatcagaattttacatcaatattttatctttttataacatttttaagcttttcttcaattatttcaattattttgcaggtaatgttgaagtgataagacaagattttagtgtaaattccctgttgagaatcacaactaaaaatgaaattgttaaatttaacttgggtcgtaataaggaagagtatgctattgcattatctcaagtttctaattatctattagagggcaatgaaataatagataatttaagctgtagaatagaaagaaataaagttatatttagtactaattcaattaatacttttatgctttaaaaaaaatttctaaagatttaagccgattgtataaaattgagcctcctaatagagatgatatttctgaacaaatttatagaatttttgaacactctacaagctatagtattgtaaggttagacattaaaagtttttatgaaaatattcaatataatgaggtaattaaaaagctggatagagataaaatactagttgcaaaatctattaaaattcttaaggatttatataactttattgataatggtttaccacgaggtttatctataagtcctattttgtcagaaatatttatgaaagaagtcgatcaacaaattagaaatatagatcatgtatactattatgctagatatgttgatgacataatagtaatttcaacagataagagtgattctatatatgaaaaaacaattaaagttttagagaaatatgatttaaatgttaatagtaagagatatataaaaaatattcctgctgtgaacaataatgaaatctcaactttatataagtttgattacttaggatataagtatattatagatacaatttcatataaaaataaacgaatagttaaagcggaactgtcagatgataaaaaaagaaaaattaaaactagaataatacatagtcttttagatagagtttataatacaacgcattatgatcgggaggagttgttaattaagcgattaaaagtgttatcctctaactactcaataacatataatgaattgtcaaaaactaatttaaaagctggtatgttttatagtcataggttagtaaataattatggtatttttagtgaatttaataaatttttatctaaagctatctactgtcaacaaaacaatttctttggtaaagctatgtcgcagattcctagtaaagaaaaagaaaatattattaaaagtatttgttttgttagtggatttaaagataaaaactttattgagttagagagggttgaaatggaacgagtaaaaaagtgttggaaaaataaacgatataagaagctttgaggtaaaaatgaaaagtaagatttatttagataaaaaggatttttatagagtattgttaactgatgtattaccctatgaagtaccttttattttaagtaatgaaggtttttatagaaacttaaaaagcaactcatttcattcagttactaaaaaaatattagaattaactttatttacttcacaagtaaacactaatccttttaattttaaaatctctaaagatgatagtaattttaggaagttatatttagttcacccaagttcacaaataaaaatatcaaatttatataaaaattattatcaattaattacgcatttgtgtagtagaagttctttttcacttagatatccaacttatgttgcaaaagctttttatagtatagaaagagatagatctaattccgaaaattataaagatgaagatattgaattactgtcacaaaaaagccctaaatatgcaagtacttattttgtatataaagatatcagtttttttatataaattctatgattcttatagatttcaccgtattgaaaaaaagtttaataaactattaaagtttgatattgctaaatgttttgactcaatatcaacatttcaattacctagatcagttaataaaaattgtagctttgaaagtcatacagatatacatagttttgaacatttattttcttcaattatgaaaggtgcttatcatggtaatacacatggtattgtaataggaccagagttttctagaattttcgctgaaattttattgcaatctatagatgtagcaataaaaaataagttaagaaatgaaatgggaattaaggagggtgttgattatgttataaaaagatatgtagatgattatttttttattttataataatgagcaaacttcaaatttaatttttgaatgtattgttgaagaactttctaagtatagactattttgcaatgaatcaaaaagtattaggactactattcctttttattacaggtattactattgctaaacatgaaataaggaagagattagaaactttttttgaattatttgagtcaataaataataaagatgattatattgggctaaaattaaatcattattataaaatatcaatcaattaattagtgatattaagtgtattgtttttaataataatgtaagttattcaagtatttctggttattttttttactttaatgaaaaatcatgttttgcatataaaaaatagtttttcttttgaggataaatctaaagttgaaaatttaagtaagttatttcttattattcttgatgtttcgtttttttgtttactgtatgaattttaaagttagaagcacatatttaatttctcaaattatagttttgattagtactattgctgaatcatttgatttaaatttgatagatttaat

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		taataaaaaaatatatgatgaggtggatttggttttaaagataaagtcaaattcaaacttattgaataatattgaaattttaaatc tattaattgctgttagagatattgatcttaattatcagatcttagtagatgatcttatgttattgttttcttcagaaaggattaat aagtataattatttctctttaatgactttttttattttatgttcaaaggaaaaaacagtatcagcctatcagagatagaatttatgc aataataattcaaaaatttaatcagaataatctaaatgtctcaaatgattctgagttaattcacattttttttgactcacttagct gtccttatttaactaaaaatcaaaaaattaatataactaactctgcattaaattctattattaaattaaatgataatgaaattgat gtttttgtagaagaaatgagcaaaactaattggtttattgactggaacttgcaaacaaaagatgcaattcagcgtttgctgatgaa aaaagaattgaaatcaccctatgaaaattgagataattaagctagaaactagatatacctccgacatttgttggttgattttacac actatataactcctagtttctataaaaggatgtttctaacatccttttatttttttttgagatttaatttttcttttagtgacaact aagttttactataactaatagc (SEQ ID NO: 25)
29	pLG031	actgctcgacaaaacgaaccgttcattcgcgaggatggtggcagtgaatgaggtggtcagttttatcagcgcttcaaggtagcttt ataggatggattgtagcgaagtgcccaacaaattgattgaagctaagggcattgagcattgcatgcatcatgctcagactgacaaa aaatcaaaataaatggattgatacggacatgacagacagcgtacagactgaaactaccgagggaaaaatcatcatcaacttgtttg ctcccaatcttcccggaagtaccaaagaagatgatctcattcagaaatctctgcgtgaccagttggttgagagtatccgaaactcg attgcttatcctgacaccgataagtttgctgggctaacacggtttattgatgagtccggccgtaatgtatttttttgtggatggtac tcgcggtgcgggtaaaactacttttatcaatagcgtggtcaaatctctgaacagtgatcaagatgatgtcaaagtcaacatcaagt gtttgccgaccatcgaccccaccaagttgccgcgtcatgagccaattttggtcactgtgactgcccgtctgaataaaatggtgtcc gacaaattaaaaggatactgggcgtcgaatgactatagaaaacaaaaagaacaatggcagaatcatcttgcacaacttcagcgtgg tttacatctgctgacagacaaggaatataagccggaatatttcagtgacgctttgaaactggatgcccagcttgattactccattg gtggtcaggatttgtcagaaatctttgaggagctggttaaacgcgcgtgtgaaattctcgactgcaaagccattttgattactttt gatgatattgatactcagtttgacgcgggttgggatgtacttgaatctattcgtaaattctttaacagccggaaattggtggtggt agcgacaggtgacttgcgtctatattcccaattgattcgcggtaaacaatacgaaaattacagcaaaactttgctcgaacaggaaa aagagagcgtccgcttagcagagcgaggctatatggttgaacaccttgaacagcaatatttattaaaactttttccggtacaaaaa cgtattcaattgaaaacaatgttgcaattggtcggcgaaaagggaaaagccggtaaagaggagatcaaggttaaaaccgagccagg catgcaggatattgacgccatagatgttcggcaagcaattggcgatgctgttagggaaggccttaatttgagagagggatcagatg ctgacatgtatgtaaatgaactgctgaagcagccagtgcggttgttgatgcaggtgcttcaggatttctatacaaaaaaatatcat gccacatcggtaaagcttgatggtaaacaaagcagaaatgaaaggcctaatgagttatcagttccgaatttacttagaaatgcctt atatggctcgatgctaagcagcatttatcgtgcagggttaaattatgaacagcatcgatttggtatggattcgctctgtaaggaca tttttacctatgtaaagcaggatcgtgattttaacactgggttttatttacggcctcagtcagaaagcgaagcattaagaaattgc tctatttacttagcgtctcaggtgagtgaaaactgtcagggcagtctgtcaaagttcctacagatgcttttggttggttgtggctc tgtcagcatattcaaccaatttgtgaccgagttagcacgagctgaaaatgatagagaaaaattcgaacagcttattagtgagtatg tagcttatatgtctgttggcagaattgaaagtgcctcacattgggctaatcgatgttgtgcggtggttgcaaacagccctaatgat gagaaaattggtgtttttcttggcatggtgcaattaaatcgtaaatcacgacaacacatgcctggggttacaaaaaatttaacat tgatactgagaatggcctagcaaaagccgcaatggcgtcttccttgagtacggtagcttcaaataatcttatggatttctgtagtg tttttaatctgattggtgctattgcagatatctcagcatgccgttgtgaaaggtcagccattactaatgcttttaataaagttata gctcagacaacatgtattgttcccccatggagcgaggctgctgttcgtgcagaaatgaaaggctcaagtaaaagtgcagataacga tgctgctgttttggatgtagaccttgatcccaaggatgatggcgtgattgatgaaagtcagcaggatgacgcaacggaattttctg atgccattactaaagttgagcaatggcttaaaaacgtaaacgaaatcgagattggaattcgtccgtcggcacttttgattggtaaa gtatggagtcggttctatttcaaccttaataatgtagctgatcaacataaaaccagactctatagaaatgcagagcatggacgaat ggctagtcaatcaaatgccgcgaaaattatgcgttttaatgttttagcatttcttcatgcggtattggttgaagagagtttatatc attcggttagtgatagggaatatatcggtgaggggttaagactaaatccagttacttcagttgatgagtttgagaaaaagataaaa ataattggtgagaaattaaaagcggataataaaacatggaaaaatacccatccattgtttttcttattaattagctgtccaattct acatccgttcattttcctgttggtgggattaattgttcagtcaaagcactgaacaaagaaacaagtttcaataagctgattgatg aaattgttggcgataaattactttctgatgaagaatgggactatctgactaaaaataatgatcaaaaaacaaacactagacaacaa atttttcaaaatactataacatcgctgaattcctccacaatcgtcggagcatcatacgataaggatacaccagccaggaaaaccaa gtcacctttattaggtgatagcgaagaaaaatgataatggccttcgtataaggattgggtatggaaaggtttcttcttaactcaac agttctgttatataggctaagcacagtctctttggatgaggtatcacttgatgagagagtggagtcatctgtattccttgctcaat acgaacaggctcgtagtttacctgatcatgtagctaaatctgcttggtcatatttagtgcaacaaatcaaacagcggaatatgaaa ctcggcccagtagcaatcttacgcctgatagctgaaaagtttattaaaaacgagaaaggtggccccaaaatcgatctacctatgtt ctcggaatggcaaacgctgatgagtcgagtatcgtgtctaccaattatagcgtgtcatcaggtatttaatccagggccagccagtc aggaatatagtttcgctggcctttatacccatatcacccgacggttgaagactacattacccgtgaatgcttacatgaaactcac caacacctaaatggcagtaccagtgcagaagagtgttggctggatgcactcaaacacccagaagcatgcctcagagattttgagaa gggctgggcatctcaagagatgaaacaactctgcgcccagattgatccatctctgacacctagaatcttcaaggatcgtttgcaaa tcgcctgtaatattcgcgaaattctttgtcgggttgctcagggcgtggaattgccagagtggatagcatcaatgcaaaatccgcag caactggcgaatagcacaattctgcataatggccgggagtatgggtttgcgacagtttggccaattgacgacaaatacagtcagga gtctgagttttgctggctaaccggattgttggaaaaatggcggtttaatgcgccagaagggttagaacgattgctttggatttacc tgctgattcaaaatcagtacttgaccttactggttcagcgagacgatttttcggatttgaacagttccagaattacaccatgacg gagttgagggaggaaacagagaaatcttatttgtctcgttttaaacatgctcatggtgcaggagtgtattctcaggtgcgttatct ggaaggacgttttgctccgaagagcgaccccaacaaaatgcaaaagctgctcttcagtgtgttaagaggatattgggaatatctga gtgctcatatgtccatggaatgggtgcatgaaaagcctctgactatatcgcaagtgctcgataacctcgaactggttgaacctcat ggcaagtgtgtagagctggcgctagtgccgcactttatcaaaagaaagcccaaaaatggtgaggcctatcctcacgcattactatt caaagacctgaaaaatcaggcagctattctgatggacatgctgaagtctgaaccgcgtctgacaggctggattcgaggagtagatg ccgcagctaatgagatgcacgcaccacctgagttattttgccccttgttccgggtactagccaaatcaggtattgctcattttacc tatcatgttggcgaggactttccgcatctgatcagtggtattcgctccattgatgatgccttgagatttttaccattgcgtaatgg cgatcgtcttggtcactgcacggcgattggtattacacctagcatctggaaacgctctttgccattgtccttatccatgaccaaag agacgagattgctcgatttggtgtttatctggcgggaacttcgaagtcatccggaactgctgcgUacgctagtgatgcagcgattg aagctgttcgcttggctcataaagtgttttcgctggaagaggaagtctcgattaccacccttgatcaggtatttgaaatgcggggg ctgttggccgaatcggaaggcctactgagtgagctaaatgaaccattaaaacccaaatccctctggttggaagagtatgagcgcgc cagagagttggttaaaacaacgggtatgaaaaggccgttgaagttgtataagcaatggctaacatctgacaatgtgcgaaagcagc gtgctgaatatgttgaagttgccctagaatatttgccggatgaagcagttgttgcattacaacaagctgtaatggcaaaaatggca gaccgaaacattgcgatagaatgcccaccgaccagcaatacacgtatcagtcagtaccgaaacgtcagcgagcatcatatctttcg ctggatgggcttgccgggtgaggcgattgaaggtgatgttcctatgtctatttgccttggctctgatgatccggggatcttcgctg

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		cggacttgaaatccgagttctatcatctgttcgttgtgttaacccgaaagttcggtttgtcgccagcagatgctttgagaaaggtagctgaggtgaacgagaatgggcgcatttatcgctttcatgatgtcagctagcctgtatacattgaggattctgtaattgttcaagaccagcagtgctcattgctaactatctat (SEQ ID NO: 26)
30	pLG032	gaggatttatgcacaaaatcctgatgcgaaatgttttcaaaaattgtcaggttaacgttcctgcagatctttgcgttacatgtcatttctggatcctttcccgacaggttaggttgtgattgatatgatgcccatctctcatttagtgatcgttatccctttataaacaggagtttatatgttatctatatgcaatagacttaaatcgatatacgtgcgcagcttacgattcacctctctacttactatttaaggaaaagagtgaggggagaattgattttcattaagatattatgagagaattatgactagtgaaatagtgttaaatcttgatttcccagaatataaggatgattttgtactgatagcattgatgagcaagataatgagttgtggcagcaacaggccaataaaaagctactttcgtttctcgaggtgatgggggaggaagcaagacgatataaagaaaataattcccgtagtacgcatccacattataagacattgagtagttatcaccatgcaatctttatcagtggcgcgcgggggcggggaaaactgttttcatgagaaatgccagatttagctggcaaaaacattataataaagatctaaaacgccctaagctatattttattgatgtgattgacccgacgctattgaatattgatgaccgtttttctgaagtcattatcgcttcaatatatgctacggtagaaaagcggatgaagcaacctgatattgcgcagaatatcaaagataattttattaattcgcttaagacgttgtccggtgcattaggtaaatcaaaagattatgatgaatataggggcattgatcgtattcaaaaatatcgttctggaatccaccttgaaaaatatttccatcagttcttgatttcaagcgttgagttactggattgcgatgcgctggttttgccgattgatgatgttgatatgaaaatagataacgcttttggtgttctggacgatattcgctgcctgttgtcatgtccattagttctaccattagttagtggggataatgatctttatcggttcattgccaaaagtaaatttgaggaattattaaatcgtaaagcaaactctaattatgctaaagaaggcagcgagatagcagaaagattatcagaagcatatattactaaagtattccccagccatgtgaagatacccctccaaccgatagatgagttgttgccatatctttatatacattctaatgaagatgaaaataaacaacatacaagctattctgaatttatcaaacttgtacaacaaaaattctactttctttgtaatgggcaagaacgaagcacaaattggccgcagccgagaagcgcacgtgaagttacgcaactaatccgttctttacctccgtctactcttagtaaggaagatgattcgggaactgatttatggcaacgcttcgctgtctgggcggaagaacgtcgcgatggattagcattaaccaatgttgaatcttatctgtttattaagaatgcgaaagcagtagaagatttaaatctgtcaaatcttattgcttttaatcctttactgcaaaaaggaaaatatccctgggcagaaaaggatttttataaacagcagtcccaacgtcggaaagagctcaatgcccccgaaacaaattcaggtatccttaataccgtattttccgaacaaaggaaagattttattttaagaagtatgcctgcgctggaactcattatggagcctatgtatgtcactaagacggtagcagaaaaaaatgataattctgcgcttatagcgatctatacccattctgattattacagccagcagcagaacagacgatgtcatatatttttggcagagcttttgaaataatgttctggtcagtattagcgaaaactgaaaatcttccacaagaattttatgaaaaagataagtttaaatctttatttggtaatattttcaaaaaagtaccattctactcaatattttcaatgaaccctacaaaggttgttgatgaagaaaatgacgatggcagtgaacctgatttttcgcaaaaactggacgatagcattaatgaactggtggaagatatatatatctgggcaaccagtaataaattgcgagccttcaaaaataaaaatttaatacccttaatgacgtgcgtttttaataaggtattttcacagatcaatgtactgagaaaaaacgtgcaggacagagttaaatttagagatgaacatttgtcagatctggctaagcgatttgagtatatgtttattaatgctatctttactttcatcagagaaggggtagttgtcaataccaatgtggcaacaggcgcagctcctgccagagtacgtaatttatcagagtttaataggtatgataaaacattatccaggaatatgtccgggatttttatccgtgaaagaggataatggcttaacgatagtcaaagagagtgagggcgatatcgcagatctgttatttgaaatttggcatagcccattatttaaattaacaaccaggacatgttacccaataggtaaaataaattcgcaaaatacggcccaggaaaatttatcatcagattttaattcatttttgaaaatggtatcaacttcgaattgataaaacaatattattggcaaacttcaaatcatgataatatcaggacagcagacgttagggaatgggcaacttcacgtcttaatgaagcaatcatcctttttcatggatgaaagaaagcaagtctattaaagcgaaaattgacggacagagctacgagggtcggctctttcgcgggcttcagcaggcgctggaaggttatgaggaggtctgagtatgtttaatcaggatccttattggctcattcctaccctttgtctggcatcagaccgaattttttatgcacaattgcgagaccacttaggccagaaaagtagcggtgaacgcaaaaaagaaaaaaatggatatatactggtacaggcggcacaagactatcaattctatttggcggccgtattcggaaagaggatgtgcaaaataatgccttaatgtggcagatagaaactggtaatgaaaattgcttatcgatgcttgatagtttgtcagcatatttcctcacatggcgcggcaattgttttgaggtcaggcgtgagcgacttgaaccctggctgatgatctgttccgtgatagatcccgcatggattattgcctatgcataccaacaattgattaaacaaaatgttgtatgtgatagtgagcttatttctttgctgacagaacatcaatgtccatttgcctttccaaaaggcagaggggacatttcctttgctgataatcatgtccatcttaatggtcatggttatagttcaatttcaatgctgaactttatagatggaaattataaggttaaaaaagggataaaatggccctatcggcaggaatacaccctctttgaaagtggtcttctggataaaaatgatcttccccgctggctgtccgcttatagctcttgcttacttaaaaatgtatataattcatttcaacaaggaaaaagatccgaggtagatttcacatgtctgaaggatgcggtcgaaacggtgcttgcggatgaggataaatattattttttagaggtagcttcgctatatgatgttgtcaccttgcagcaaagagtgctttatgaagccgcccagcagaaatatcactcacatcaacgttggttactgtatacttgcggaataatgttaggtacagaatctgaagattatgcgaatgcgctggctaacctgatccgaatcagcaatattctaagaaactatatggttgtatctgcggttggattgggacaatttattgatttttcggcttcaactatcgtcgaataacaaagccagctgatacaaacaaccgagttcattatgattcttctgctggtatttccagagaatatcgtgtctctcctgattttgtactgggtagcggcgtaatgcctgatatatatgccaggcaacttttcgatttttattgtacccaagcacgcaagggcgtacccgaacaaggacatattgttgttcattttacacgttcctttcctgacaaaaaatcaacatatgataaattgctaaccgagtgtcgcgaacggttacgttctcagtgtgattatttggccgtttttaacatcgcttactttgcagtcgatagaatataaaaatttatctactgatgaagatcgaagcatagacattagaaaattagttcgtggctatgatgttgctggaaatgaaaacgagctacaaatagaggtatttgccccggttctccgggtactgcgtgctgctaaatttaaaggggagggggtgaactttaaaaggctacagcgcccttttattactgtacatgctggtgaggattattgtcatatactcagtggccttcgggctatggatgaagccgttgaattttgtatgttaggagaaggcgatcgtatagggcatggattagctctgggagtagatataaaactatgggcgaatcgccaaaagcgagcatacctgacggttggacaacatcttgataatttggtttgggcatatcatcaggcagtattactttctcaacatattgtcgagcatataccagtaatgcatgaattaagggataagatccattattggtctcatcaattatatagtgaaacttatacgccagatttactctttaaagcatggctgctccgccgtaactggccggattataagtcaatcatatctgatccagcaaatatcaatgaatgggtgcctgaccaacatattttagtcagtacagatgagactacagctaaggccagaaaaatttgggaacgttatttaaatagcggtctggcagaaaatgatgtttttaacagaataatttcagtaaattgtgcgcccgatacagcgcaaaatttttcaatgaccttaatgaaaatgaagatattttatccaaaggggaattattattgtatgaagctatccaggatttcttaatcgaaaaatatagtaggttgggtttagtcatagaagcttgtccaacctcaaatatttatattggcagactggagaaatatcatgagcacccattattccgttggaatcctcctgactcccaatggattaaacctggtgggaaatttaatcgctttggattgcgcacaggacctttatctgtctgtataaatacagatgacagtgcattgatgccaaccacaattgaaaacgaacatcgcttaatgagagactgcgccatacatttttatggtattggaacatggatggcggatttatggataaactcaatacgcataaaaggtattgaaatattcaaaggtaatcatttaagtcaggatttagataatttaatctaaatgtaaacaagaaatccacgcaaatgcgtggattttaagtcaacttattattctctgaaacggtttaaccgttcggaacaacagattaaatc (SEQ ID NO: 27)
31	pLG033	tgtggttagttatcacagcactaacctatttcgagcttttgattgaccaataccatttcttttaattatgaataatgatgcgtcaaccgatggcgaacgggccaaatccactcttctacaactgcccattgtcacggtgtggaataattaaaaatttagattttgaga

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		ttattctcattaccatcttgattttatttggttttgcatcaaaattcatagttcacaagcttttctcactccaaaaacaactgtaa agggattattgtgaacacgatatacataccattagacagcggagagtctgcggttcttaaggatccagataccttacttccccgaa atatttacgaacagcttactcgatttattgaaaaggctgttaatgaagtaccgaagcctcacgaagcgcttaatgaaacccgtagc cataaggctatatcgattgacggcgcaaggggacaggaaaaacgtcggtgctagtgaatttgaacgactatctgcagagtaatgc tcagcaactggcggggaaaattcatatccttgatcctatcgatccgactctacttgaagatggtgagtcgctgttcttgcatatta ttgttgctgccgtgcttcatgataaagagatcaaaactgcccaaagcagagacctcgataagtccagagtgtttacccagaagctt gagaacttggcacacggactggagtccgttgatttgcaacagaatcaacgtggaatggataaaattcgctccttatatggcagcaa gcatctggcaaattgcgttgaagagttttaaaatctgcgttggagttgatcggaaagaaattattgatactaccgattgatgatg tggacacttcactaaaccgggcatttgaaaatctggaaatattgcgtcgttatcttacctctccgtatgttttgccggtagtgagc ggcgatcgccgtttatatgatgaggtctgctggcgagattttcatggaaggttgaataaggattcagcatataatcgcaagaacac atatgatattgctagagatttggcaattgagtatcagcgtaaaattctgccgctaccgcgcagactgagtatgcccgatgtaagtg attactggcagcaagatggtatcgaagttacgctagataaaaatggcattcctctgcgtaattttatggcatggttgaaaatattt attactggccccgtgaatggccttgagggtagtgatttacctctaccgataccttcaatacgtgctttaacccagttcatcaacca ttgcagggatttaattcgtgagcttcctgaaccattcagaaagaaagtcagtacgctggccttacgtcgtatgtggcaaatgcctg atgttcctcttgatgttcttgaaagttttgctgaaaaacatcgggaattgagtaaagaagctaagcgtgaatatggggaggcttac aagctattttatgatggactaaagaattttactgcttgggatagtaaggcttatctagaagatgataaacaatctgcatggctcga taggttgtgtgagtattttcgttttgaacctaaggctggggctgtgtttttaacgcttcaggcaaaacagttctgggtctcatggg cgcagggtgacaatcgtaatcaatcgattcttgcgactccgctttttcaacccttattgcataattttcgtgaatacgatgtcttt gaaaggtatgatgatctttctgattgggaatctcagttaagaacaaggttaccggagagttggttgactgccattaaagggcaaaa aacgcttttaccctatcctgtagcagaagcgggaattaataccagtttaaagtggaggtattgggaagaattagagaactatgggt ttgatcctgctttggaaagcaaggcaaatttcctttttgtccacgttgatgcagaggaatttttatacaaactctaaacagtcagtc gtgataaatattggtagagtttttgaaataattattgctagtcttgtttcggatttagagttggccgacttgcagagaattagaca acgttctccatttactctgctagcgcgcttgcacctaccaaaacgttagatttggaagaggattttacgaaaaagaatacaagat ttatgaataacagaagtgaaactgacagagacatttctgatgatattcttgttgatgtgccggataaaaatgaggacgcatggaaa aaaatttgtgatgaaataaaccattggagaaagacacacaatgtggctagtacaaacttatcaccttggctggtttataaggtctt taataaaacatatagtcaggttgctaataatgtgtttgttcccagtggaatgcaaaatgttgatgcggctctaaatgtttttggta gggtttttatgcagtttggtcagcatttggtagttttgaaaaaggcgaattgttcggactatccgatgtggttgctacaactaat attatttcggcaaaaaattttataatcatgataacttccgagtgaatgttggaccgtttacgcctgagcaaaaccaaaattctga cagcgatcgtgaggcatatcagcatcgcaaaatgtatggtgaaaaaaccagagcggtaagttatgtattagcaactcatccgctga aaaaatggatcgacgaggtattacgcactgagtttaaacaaaaacagaatgctcagattcagaccgagagaaaaatgccgattcag gctgagaaaattatagatatcagcccggcaagagagtttatcacaagaaaactttcattaaattcacactcccggttggttaaaac acgtataataaaacagcttaagatgttatatccaaactacgataaggctaaggacttcattgatgaagttacaaaccacttccctc agaatgatcccgcaattaatacgcttcagaaagcatttgcagaactttaccccgatggtgacaaataatgttaactcggtctctaa gtgaacatgctgcagggtgtttttcactgatgagcgtctgtcacaacgctttctagatatccttttatcgccacccaaggatttt gaaacgtggtcatcattgcaggaggaatctttcaagctgctcgttaagagcatcgatagccgatatccacgcacttaccggttaac cgacgtacgccagcttgtggggaacatatgtgacaacgggttactgacgagtccgacactaccttggctcgatgtcattgcggatc agttactgttgcggaatggcgacttactctattaccgcgaaaataaggttcaagactacgtgcgaatagctgcggaactcgaccct gcccttctagtgggatggcgtcttggcgactggcttttgcaaagcccaccgccgcgattgacggacataacccgtgtggtgatggc gcagaatccgtttttgctccacctgctaatgcaggtaaaccttttgccgaggggcacgtacatctcgggggagtgacggctggag atactattttggatggctatctttttgaagagattgaactacccaaaagcaaagatatgttgttgtgggcgcacaaagagcatgat gagttaacaccgttgataaatcgagcaaagtctttgcttacagttctactttctgccccccctcaaacggtttctgagcaaactca aaatggttttgatcagcgtaaaactgtatctgagaagtacaaggcattacagaacccaatggatagcatccatcgtctcccagact ggttattgcttgctaaaaagaatcgcggaactgaaagcgtcagccccggctggtttttaaaccaactggcgcatgcctccgaaaaa aaacatccctcgcgctggctgtggctgcagctatacctttgccactcttatcagcttaaagacactcatccactggagcgcacggc aatactctgttttggcttacggtaaatgcgctacggcgtcacattattatggacggacaggggcttgcgtgttttaccgagcgtt attttaatggtgctttacgtgcgggtaagaaagctgacagtagcaatatgcgctacctgtttgccggtaaagacgatgtggccgaa gtgaaagcatccccaaaggctttcgatcatgagatggtcactggattttcctcgacattgctgaaaaccctcggcattccagctgt ttttccaccgtatatttttggtgagcatgagattaagccagatgaacgcgtgctgcgctatattggagcactggagcgctggcagt tttgtgggcactttctcgctctaaaactgcaagtcgcggcaagcgagcaaaggctgatttgcaggctaactggacagaagcggag cgattgttacagaaactgtacagtcataatggctggaatcatcccgtcttcttaggggtaaacgtaacccacatttcattttca gccgtcgaactggtttcgggggcttgatgttgcaggggatgaaaacgtactaaaaattgcaggctttgccccgatgctgcgctggc tacgaagtggattatatcccgtaccagaagggcttcgcgccagtatgagttttcatttcagtattcatgccggggaggattacgca catccggcgtcaggattgcgtcatattgatgaaacggttcgcttctgcgaaatgcgggagggagaccggctaggacatgctctggc tctcggaattgaacctgcgctctgggcgaaacggcatggtgaaatgatactacctctggatgaacatttagataatcttgtctggc agtggcactatgctacgcttttatcggcttcattgcctctcgctcaggcggtattaccgctgcttgagcgtagaattgcacgcttt attgcacggtgcgaatggtgcaaaaagagacctccgcaaatagataacagtgtggtggggaaacaggcctgtagtgatgataaacc tctggaaaatattacacctgatacgctctaccgggcctggctactgcggcgtaattgttcatatcgactccagcaactccacggcg gttcccctttgacctcgcaagagaaatgtgcgctgccggattgggccacgctcagcgataaaggcaatgtggcggcgcagctttat cagcaaagacactcgagtctccttgacgatatgccgccgcaactggtagttgtgcgtgtagcggacgaatggggaactcaggagct tattggcttgggaaatcctggtaaactgcgtcagcaggctcttgacggtaaagatatcctccaagacattgatacgccggtagagc tgcaatttatgcatgctttacaggactatttgctagatcactatgatcgtaaagggttaattatagaaaccaacccaacatcaaac gtatatatcgcgcgattcaaaaagcacgtagagcatcctatttttcgttggaatcctccggatgaagaactgttgaaaccaggcgc tgaatttaatcgttatggattgcgccgtgggccagtcagggttctggtcaatactgacgatccagggattatgcctacgacattac ggacggaatttttactactgcgagaggctgcgattgagcgtggtgtcagccgaacgatggcagaatattggctggaaaggctgcgc ctgtacgggctggaacagtttcagcgtaatcatttaaatgtatttgaagttattgaatagaggattttatcgtgagtggtacattc ccttacttgcaatatacggatgtcaatgggctacaacctaagctcaaagaagagttgaaaaatttacggagaaaagagtatttgtc ctactggcctcgttttctgatacgtagaatttcgctttatgctcttccattcctcatgttcttcacttttttcttttgtctgagtc tgacgaagaaagttggggcagaggaagtgactaatattcttggaaccgtgagtatatccttcagtagttgcctgctgctggggatt attatttctggtgtcgtgttactcttgcagtggacgtgcttcaactgtaaatacagtccgcaggatacgaatggagttgttggggc tcgtaagttaaattataaattacttgctcatgttgtatttgttattgcatgcgtgcttttatttgtttttatttattgcaccaata ataaagtgttttatggttttatcgtgtttcttggtttgacattattaccattggtaattgaccgtaccttgggggtgactcgtcaa aatgaacgtcacaaactctatatcagaaggttagagcgcctcgatgaattgaatattctccgggagaaaatgaatattaaattcga

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		agaatcccatttcatcgagtatatgaagcttgttgatgaagctgatcacggaaaaaaccaggatacagtaagcgatacatcctattttatgacgttgatagaaaataagctaaaagtgtaatcggttttaatatgatgctgtataaaaaactacgcaattgcgtggttttttgtcggactatgagggcaaggttgccctaaaacagaggttaaacgttgggatgtgatttattgcacatcatgccgtgcccatccagtagaatccggttcgaaatgtgtataggattgtgtatatgtttctgttcggtctcggattcttatacac (SEQ ID NO: 28)
32	pLG034	accgtgctggcatgtttttacggagtgacgctttcattaacctgtacacgaacttctattccggcatcatgacaggcctgcagccactgcgccacttccagcggatcgccctcccggcgtaccactctgccttctttattccataactgcagacaggtgctgccgtcgagacgcaccacaaaatccccacggcaggcctgataggggtttgagggccaaccgtacgaaaacgtacggtaagaggaaaattatcgtcttaaaaatcgatttatgctatcacagtcgtctcttcaggtaagtacggttgcctttgcctgctttcttctcgtctggttaagttaagaaattcagagatccatgcttgagataaaagcggaataaaaccagtaaaatgtaactaaaacaacaacggaattgtatcaatgataatgtccacaccgtggctgacaccgatcgttgccgatagtgatcatgctgaggcaaatgcagtgagctatgaagcactgactccgacagaactcgactcagataaagcaggctgttatatcagcgcgcttaattatgcttatgaacatccggatatccggaatattgctgttaccgggccgtatggggcagggaaaagctcagtattaaaaacatggtgcaaagctcacaatgggacactgcgggtgttaaccgtttctcttgctgattttgatatgcagagacatgtggatgaaagtaatggggacagcagtagtgacgaagggacgaaaaatactggtagtgttgaaaaatctattgaatacagtattctgcaacaaatactctacaaaaataaaaagcatgagcttccctgttcccgcattgaccgtatatcagatgtgactgcgggacaaatattgcggtctgcgtcttttctgacaggaaccattttactgagtggagctgctttatttttccttgcgccggattacgttacaacaaagctatctttgccgggagcattcgcccgttaccttcttgaatgcccgtttggggtgcgtgtgtccggtgcagtggcatctgtgatgggatcgttatgcctgcttttgaaccagttacatcgtatcggtatatttgacaggaaagtaagtcttgataaagtggaccttctgaaaggcgctgttacaacccgggcatcatcaccttctttacttaatgtctatattgatgaaattgtctatttttttgattcgactaaatatgatgtagtgatattcgaagatcttgaccgttttaacaatggccggattttcgtgaaattgcgggaaatcaatcaaattattaataactgcctttctgacagaaaacctgtaaaatttatttatgctgtcagagatggtattttcaactcagcagagtcaagaacgaaattctttgattttgttatgcctgttattccagtgatggataaccagaatgcttatgagcattttgttaaaaaattcaaagaagaagagataaataataacttaagcgaatgtatttctcgtattgcgacatttattcccaatatgcgtgtaatgcataatattacaaatgagtttcgactctatcagaatttagtcaatagtcgggaaaatctggccaaactacttgccatgatagcatataaaaatctctgtgcggaagattatcatggtatagatagtaaaaaaggtgttctttatcattttattcaaagctacttagaccatgaaattcagaatgaattattacattctgcaaataacgaacttgaggatatggcacagtcacttgtagcgataacaaatgaaaaactcgcaaaccgggaaaatctgcgcgaagaactgctcatgccttaccttagtaaaaattatagcggcgcgcttgtttttatacagaaggaaggcaaataagtcttgatgatttgatacaagatgaagatgaatttctcatgcttttagataaggaaaatattcaggtcgttacccccctataacagacaaaattttctcatgataaatcagcgggatacagaaaaactgaagcagcagtatgaaaaacgatgccatttaattgaaactaaatctgttgataatataaccagagtgaaaaataatatttccagtctggagtcattgaggaccgaaattctttccggaactgtagctgatatagcagaaaagatgacaaatgaaggctttgttgcctggataaagaagaaagaggatacaggtgtcctgacgattcagtcggaacatgaacagattgattttatattttttctgttatcaagtggttatttatcaacagattacatgtcctatcgctcaatcttcattcccggagggctgagtgagacagataatttatttcttaaggatgttatgtctggtaaaggtccggaaaaaacattctcattccatcttgataacgttaataatattgttgaacgactcaaaaagctgggggttctgcagcgtgacaatgctcaacatcctgctgttatcagatggctgattgataatgaccctgataccctgaaaaacaatataatggcattactgagtcagacgggtagccagcgtgtggttagtttgctgatgttgatgcagaacgatttcacaacgtatgttcgcctgcgttacctggagattttatgtcagatgaacatatactgaacagattgctggcacatttatgtgcgtcagaagaacgcacacccgagcaaaagttttttgttcaggaaatagcggcacacctgttatgcctgactgaaaaatcaaatatctggcaatcggttgagattaataaacgtatcggtgagcttatagattcctccccaattcttattactgctgtgccaaaaggatatggtgatgcgttttttgaagtgttgaaagataatacactttcagtttcatatattccaggtgatgtgggagacgagaagtgttctgttatcaggaaaattgcgggtgcaggattattcaaatattccgtcagtaatcttaaaaatgtttatctttgcctgacgcaagacaagaatgaagaaagaatgtcattctctctttatccgtttcattgtctcgagtccctggctatttctgaattaacagaaattctgtggactaacatagaagattttatttatcggtatttattgaatcggaagagattgatcgtattcctgaattgctgaattcttctgaagtctcaatgactgttgttgaacagattatagccaaaatggattttgtataaataatctggatgatattattaatcgttcagagtgtgcggacaataatgcttcagggagaaatatctatagcatgctgttgcagcatgacaggatttttccatcctttgataatattattcatttattgcatgatacatcaattaatacttccggtgaacttgttcagtgggtaaatgagaaacactttgaatttgaaccatctgatatagtcataaatgatacaggaatatttaataatttatttctgaattaatttgctcgccagtcatttcagaagaagctttactgaaagtactgagtaatttaaacgttgttattatcgatgtgcctgaaaacattccattgcgaaatgctgaactgttatgttcagagaaaaaactggcaccgacagttaatgtctttacggtgttgtttaatgctctcagtgaaaatgttgatgatattaacaggatgaatactctgcttggtaaccttattgcccagcgtcctgagattattacccaggagccagaagatatttttatatcgagggtgactttgatgaagaactggcaagcgaactttttcgtcacaagctaatcggtatgaatataaaagttgccgctttacgctggttgcgtgataacaaaccgggaattcttgataagagctacctgctgtcattagatattctggcagaactgagtccctggatgggtgacgatgatctgcgcctgacactgcttaaacgttgtctggttgccggggatgctggcaaagacgcgctttgcgtggtgctgaacagttttgctgatgagagctatcatggactgttaccacatgacaggttcaggaaaatccctcactccgtggatttgtgggaagtggccgaattaatcagcaatcttggatttattcagccgccaaaaatggggtcagggcgtgatgaacacaaaattgttattactcccgtacgctatgtccgtgatgttgagttttatgactgagcatcattgatacggtgttttaattgccttaaatacaaaaataaaaacagattaatgcttaatgtgcattaatctgttttagttatcaatggctgttaattattgttaattttacattaatctttctttttcttcaggaagatccgaaaactcctggtcacggatcttcct (SEQ ID NO: 29)
33	pLG035	attatctgccaaccgataagatggctgcctaagtcgtagcgattcagcactgttttagcggcgctcgattgcaaagtcgtgctttgctgacttgcgattgtgctctttacgagcaaagctttcaggtatagtaagtgctaactgtagtgtaaaattatagggatagatgaagaaaacaacgaggctttagctaatctttgcagttgtgtctgctataataaggcgaaattttatctgcatgattttgtttgattaactccgaaagccagctctctcggtgaagattgggaagggatatcaatgagtgatgatagctataaatttcaaaagttaacgccgttcagcgatgttgagctgggtgtatataaaaatgcgatagattttgtttttgccaataacgatctaaaaaatgttgcgatatcagggcaatatagcgcaggaaaaagtagtcttatcgaatcctataagaaaagtcattcaaatataaagtttgttcatatctcacttgctcatttcagatcgattgaggaagctgaaactaatgaaccaagtaaagatataaatgaaaccgcgttagaaggtaaagttcttaaccagttaattcaccaaattaatgctgatgatattccccagacacattttaaagtaaagaaaaaaataaaaactaacaacattgtgataaacaccatctttacggtgttatttatcgccatgatactacatatcacgctatttaataagtgggaaaagtttgtttcacttttatctgaaggtaatataaagacactacttacattatcaactaaatacgatacgcttttaattagtgggtttatatgtactatcctatcttgtattttcatttacaagttaataaaaacccaaaagaatcgtaatgttcttaagaaaataaatttacagggtaatgaaatagagatttttgaagaaagtaacgagtcttatttcgatagatatttaaatgaagtattgtaccttttcgagaacgttgatgctgatgccattgtttttgaagacatggaccgttttaatagtaataacatctttgaacgtcttcatgaggttaacagactggttaatattcaacgggacacagcagggcacaagaaatcgacgttacgttttatttacttgcttcgtgatgatatcttcatttcgaaggatagaaccaaattctttgattatatcattccagttattcctgttgttgatagttctaactcttacgatcagtttatcacacattttgatggtggtggtattctcaagttg

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		ttcaatgaaagatttctacaaggatgtctttatatattgatgatatgagaatattgaagaatatttataacgaatttcaaattta ttataacaaattaaacacgacagaacttgactgtaataaaatgttggccattattgcctataagaatattttcccaagagatttta gtgagttgcaacttaatcaaggtatggtttataccatatttagtgaaaaagacaaccttattattgaagaaataaagaaaatagaa aaagatattagagatagaaaaaaagagattgaggcaatcaatgatgaaatactcaactctagtcaggaggttgatgctatatacga taaggaattatctagatataataatcatcctcactataatcaggctgagaaagctgatatagcaaagagaagggcggctagaaaag aaagtgttgaaaataaatttaatggtaaaatagaagaaattaatgagcttatatcaagatcaagagaaagtttggttgattctaga aacaaaagacttaaagaagtaataactagagaaaacattgatgaaatatttaaactcacctataccaatgaaattggagaggaaag agactttaatgaaataaaaagcagtgagcattttgacttgcttaaataccttattcgtgatggttatattgatgaaacctataccg actatatgacctatttttatgaaaatagcctgagtcgaattgataagatgtttttacgcagcattaccgatcaaaaaggcaaagag ttcacttatcaactcaagaaccccaagctggtcgttgcccgccttcgagaagtggattttgaacaggaagaggcgcttaattttga tttattagcttatctgcttcaaacgccagcccaggtaaacttaataaaacgtttattcaaacaactaagaaaagatagaagagttg agtttattcgtggttactttgaaactgagagggctcagcctgtcttcattaatcgattaaatacacagtggcctgagtttttttct tatgcgctgacagagagtgaattttctgctgattgggttaaactctactctataggcacgttttattattctgccaatgacgccat cgaggccattaatattgatgattgtctgactgattacatctctgattcggcaggttatttagcaatatcagaaccgaaggttgaca aattaattagtggttttaagttgcttaacgtctcttttgtcagtattaaatUgaaaacgcaaataaagtactctUgatgcggttta ccagcattcactttatgatattaatttttccaacctgaccttaatgctgagtaaggtttacacgcttaatagtgaagatgatattc gccataagaactatacactagtgatgtcacaacctgattctcccttggctagttatgttaataaccatattagggactatctggat atggttttatctagttgtgatggttcaatcgtggatgatgaatccattgttttatccgttcttaataatgagggaatatctgatga acaaaaaggccagtatataaacgctttgcaaactttcgtgacatctctgagtgaggttgagagcgaatctttatggtcatctttgt tggataaagatagagcagtgtgctctgaggaaaatattgtctcttatttgaacatgttgatggactggatgactcacttatcgaa tttatcaatagaactgatgtagacctgaattttcaaaatattaatattgataacgagcttaaaggtaaattatttaaatcgattgt tatctgtaatgatttatcaaatgataaatatgaaaaattaatttgctcactaaatattatttgtaaaacatcctttagcgctagta atatcgcgagtgataagttcaaaatattagtggataaaaatattattcgtatgaatgttgcgccacttaatttcatacgagataac tattcagagcaactttcctattatattcataagaatatcagggcatacgttgaattaatgacgattgataactttatttggatga ggctatatcaatactttcttggaaagttgatgatgatttgaaagttaagctactcgagtttgttaaaactccgttggctatttata gtaagaattactctcaggtcgttaatgactatatttagaaaataatttaaaccagatgaacttctaatcttgacgtcatcttat aaaacttggggaacctctactcagtcgctcatcttgagtcgagcaatacaggatatatcagcattgatagcaagtcctaatgatgt ttctgaaccgttactaaaaaacctgtttgtcgcagagggactgaatatgcagaataaaatagcactgctaatcgctttgttgccgg gtaaggatttgagtaagacgacttgcaaagagtatcttgatctgcttggtttatcggagttcagtaaaattttggggcgaggcaaa cctaaaattgaagttgattcaactaatcaaagtttattaacagcattaagagataaccacttcttctctgattttgaggtggataa tgaaaatcccacttattataaaataacaaggcggcgctctatgtttggctcagatacatagcattatgtatttttctacagtttgg gcacttttatagtgcccaatttttacgctgaaacttacgcagataatctgacttttcccagttgacgagtacacctag (SEQ ID NO: 30)
34	pLG036	atctatagcagtcatcatattggattattggtgaagtggtacactgaatttgcccacctgaacagagttggttttatcaaacctgt agtttactcaatgacgtaaaaattggtgatgtaaaggatataaaaatgtggtcagacaaagagtcatcagaagactacctaaattt tggtgaagtatctcagttagccgtggatgtacttaccacgaaagatatgttaccagtatctatcggaattttggaaactgggggg caggtaaatcctctctgttaaaactgatagagcaaaaacttgagcaagacgacaaagattggattgttatcaattttgactcttgg ctctatcaggggtacgacgacgcccgtgccgcacttcttgaagtcatcgctacagaattgacaaaagctgctgaaggtaattctac ccttatatcaaaaactaagagactccttagtcgagttgatggttttagagctatgggattactagctgagggtacagctttaatgg caggattacctactggcggtttgctttctaggggattggtgcattaagaaatatcaccgatggcatccagagccaggaagagtat gaggctttaggcaatatagctaaagaaggtaaagaaactgcttgtggtttgattaaaccacaaacaaaaaaaagccccccctcagca gattgatgcctttcgtaaggaatatggggaaattctagaagaacttggaaagccactcattgtggtaatagataacctagaccgct gtctccctgccaatgctatccatacacttgaagctatcaggctattccttttcttgactaatacagcctttattattgcagcagat gaggacatgattcgctcttctgtggctgattacttcaaaggggcatcacagcgccatcaaatagattatctggataagctaatcca ggttcctattcgggtgcctaaggctggggtccgtgagatccgttcgtatctgttcatgctttatgccattgaacatggcttagaag gcgaaaaaataactatgctccgtgagggcttagagaaggcgttacagcaatcctggaaagatgaaccaatctcacgtcaggaggcc ttaaaaatgactggtgaagcggatgatagcaacctcgcgctggcgtttgcgcgtgctgaccgtattgctcccattttagccaactc tccaattattcatggtaatcccaggatcgttaaacgcttgttgaatgttgtgaaaatgcgatctcaaattgcgaagcgacgagcaa tgcctttggatgaagcaattattactaagctagtaatttttgaacgctgtgttggagtggatggcaccgctgatttatatcatctc gtggatattgaacaaggtgttccccagatacttaaacagcttgacgataatggcggtcaaatacctactgatgcaccaaagacatg gactgatagtccaacgactaaatctttcatcagtcaatgggcccaacttgaacctcgtcttggtgggattgacttaagggccgcca tatatctgtcccgagaaactatgccaataggtgcatatgtggttggtttatcgccatctggacgggaagtactaaatgcactaatt gaattgaaaaacactagttctcctacagcagaaaacctttgaaagcacttcctcgtgaggagcaaatacctgtaatggaaggttt aattaaccagttacggcaggtatcagattgggatcgtaagcccagaggcttttccggcgcatgtctgttggcccgctactcaacag atgcagccagcatattaattcgttatctacaggaattacagttggggatgaaacgaccagcgtggatgactgcagcattaaaagat gaacaatggaataaggacgcttaatgggaacatcacaatcaagtaaaggtccaggaggtggctctccgctggttccaccatgggct gatgatcagccacagcaaccgttaccctcgccgcaagaaaggaggtttgcgccatttcgagaatcgttgggaaatgcggtatcaaa tggaaatcgagcagatttcagaaaagccatagggcactacgcgcgaaaagcctccggagggagcagtaacgctgctcggcgattag ggagtgtcacgcaagctggggccgaattatttggggctttagtgggaatgccttcggctcccggagaaccaagcatcgatttgggc agtttggcaggccttccatgcgaaatagcaatatcaactattgctcaagctttaacatcacaggatggtgactcagaaaagatctg tgcggccatgaaccatgctttagtggaggctcttgatggcgtagaaattttcgatcctcaaaaaataactgatggtttgattgttg acacaatgattggttatctagcggaaagtattttccttcagatggtaatggattctaatagggcatggaacaaagcagatacacct tcaaaggcaattcatgcagaaattgaactccgggaattgattaaagttgttgttgataaacatatggcaccaaaacttgccggtaa cataagatcgttcacacgaaaccaaatggtaaaaattgaacgtcaggccattattgaggcctggcaagaatgggaggcataccagt gacacaattagttttccatcataaacatcaccatttgccgccagcaagtgagaaagtgttacctgttcagctatatggattaagtg gtcagaggcgcggagatatatctgttatcgggaatcctgcgattgatcggatcagacgtttgggagtacagcttccagctaaggtc atggattttctgagtgttgcattagcagtaactgcagcagatactttcgttcagcgtgaaagttccgaggatggttggacccgcca attgtcgttacgactccccccttcatgaaccatccagatggattagtctaaagaaagaacttgagagtgctttgcattttcttagtg gagacatctgggatttcgaattttgtgacgatggttatgcaccgccagagccttatagccagcattcaaggcatcgtctgattaag ctaaaagggcttgactgtgtcagcttattttcaggaggtctggattcagctattggtgcaatagatcttctggctgcagggcgcgc tccacttttggttagtcatgcttataaaggggataagtctcgtcaagatcagattgctgaaaaattaagtggccaattttcgcgct

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		ttgagattaatgctgacccacacatttatcaaggcgtgactgatattacgatgcgaactcgtagcctcaattttcttgcccttgcggccgtaggtgcttgtgccgtacaagagatatctcaacaagaaaagattgatttgttcgtacctgaaaatggatttatctcattaaatgcaccacttactccacggcggataggttcgctgagcacacgaacaacacatccacattttattacgagcatacaaaagatctttgatgcgctcggtatttcttgtcaaataatcaatccatatcagtttaagacaaaggaaaaatgatctccgaatgttcaaataagcagctcttatctaaaattgtggaaagtacagtatcctgcagtcattggaaacgaatggggcagcaatgtggggtatgtataccgtgtatcattcgacgagcatcacttcatgcaggggaattagtagagatgttgaatatattttccagtccttagctaaagtaatgaatgaaatagatcgcagggacgacctgatcgcccttaggattgcgatcacgcagaaatcgactttgaaaataggtacatggattgccaaaagtggccctttgcctacggcagaatttgataatttcaagcaagtatttaaggatggcctagatgaggttgaaagctatttactgagtgagaacatagtatgagcatcgatatgcactgtcatctagacttatatcctcggccagacctcgtggctgaagaaagtaaacgtcgagggacttatattctgtcggtgacaacaacacctaaagcatggcatggtacttctttattggctaaagaaagtcaacgaatccgaactgctcttgggctacatcctcaaatcgcgcatcaaagatcgcatgagttagacctgtttgattcattgctttcggaaactaagtatgtaggggaaatagggcttgatggtggacagggatttaaagaacattgggatattcaattgaaagtgttccgacacattctcaacagtgtaaatcgggctggtggcaagattatgactatccatagtcggggaagtgcatcagcggtgcttgatgagattgaaaatatcgatggggtggcaatattgcattggttcactggaacacctaagcagcttgaaagggcaattgatttaggatgctggttctcagtggggcctgctatgctcgatacaataaagggtaaggccttagttttgaaaatacccaaatcacgcattcttacagaaacagatgggccatttgctaagtttcgtaatgacccactaatgccatgggatagtgggattgcagagaaacagttagccgcattatggggattagtcagatggaggttaatgctcagctagttgataattttaaggtattatgtacatcataagaatgaaaaacttagatatgcatttacagttcaattcatttttcgtcatcagttaattacacataaaattaaaagtaagaatatatctaccctgtgaatgagcaaggcggatttatatagtttgtaattagtttaaatgtaagcagttcgtcagagtgcgtattccgctctattcgatcacggattggccgttatgaccc (SEQ ID NO: 31)
35	pLG037	gaaattatttggaatggatgatggcgcttgattactggaacaggtctatgacatgaaggttatgatttgttcactgctatgaggttaacactttaacaatttcccttactattcttgtactaattccttccaaatacttctgcttgagattaggatttatcctcttgtagtgttatttacaataaagattgtgatgctgatttaacccaacgtgttgtcagttgccttgctgaactaagttcagtatctagaaattagctcttgatacatgagcgaatcagcgaaaattttcatcccgaccaattaatgaccgtaatggataggatgttgctgctatttggcttccatgagggaacatatgtttttaaacgatcaagaaacgtccactgacctgctgtactacaccgctatcgccagcacagtggttaggcttgttgatgaaacgtcagatgcacccattacgattggtgtgcatggtgattgggggcgggaaaatcaagcgtactaaaaatgcttgaggctgcctgcgagaaaaaggataaaacgcactgtatctggtttaacggatggacgtttgagggattcgaagatgctaaaactgtaatcatcgaaaccatcgtcgaggatcttgttgcctcgcgcccgatgagcaccaaggtggcagaagcagcaaaaaaggttcttcgtcgaattgactggttgaaaatggccaagaaagcgggggactggcgtttaccgcatttactggcatacccacatttgatcagattaaggggatgtacgaactggcatccgactttctaagtgctccgcaggacaagctttctgctgcagatttcaaagcgtttgctgaaaaagcaggaggcttcatcaaagaggccgatactgatagtaatacgctacccaaacatattcatgctttccgtgaggagttcagggcgctgcttgatgctgctgaaattgaaaagctagtggtgatcgttgacgatcttgatcgctgcctgcctaaaaccgcgattgaaacgctcgaagctattcgcctttttcttgtttgtagagaaaactgcatttgttatcggtgcagatgaagccatgatcgaatatgcggtaaaagaccatttccccgacctgcctcaaagcaccgggccggtaagttatgcacgcaactatcttgaaaagctcatacaggttccatttcgaatccccgcactgggaactgcagaaacgcgtatatataccacgttgttgcttgcagaaaatgcgttgggttcggaggacgacaattttaaagcattgctcaataaagcacgggaagagatgaagcgtccttggatcagccgcgggcttgacagagaggcagtgatggcagcgttaaatggaaagattccggaggttgtggaaaacgcgctgctattcagcctacacgttacccctatgcttagttcggggacacatggtaatccaaggcagattaaacgctttttgaactcaatgatgttacgccaggcgattgctgatgaacgcgggttcggtagtgacattaagcgtcctgtactggcaaaaattatgcttgctgagcgtttttaccccagcgtatacggaaagcttgttcagcttgtatctaatcatccagagggaaaaccggaagctttggcggagtttgaagccttggtcagagggggaaaactgctccgaagagtcgcgctgacagcaaagagaattcctcagagtctgaagacgtccaaaactggctgaagattgattgggcgatcggttgggcaaaagcagagcccgcactttctggagaggatcttcgtccatatgtgtttgtcactcgtgacaaacacagtactttgagtaatctggtcgtatcaagccatctcattcctataatggagaaacttcttggtccgaaaattgggatggtgaaaatcaaaggggatttagagaaactgagtccaccggatgctgatgaattattcgaaatgcttagcgataagcttttccaagaagacagtttcaatcgaaaaccaagaggatttgacggcctcgaatatctcgtagaaacacaacctcaccttcaaaggagattgattgattttgcacggcgcattcctgtaaaaaaagcaggggggatggcttgctacccgtattgcgcaaagcctagtggaccctacgttaatagaagaatatacaaaactgatccaagaatgggcgagtcaggacgaaaatctgtccctctctaaatcagcaaaagcaaccctccagttatcgggatatcaacattaatgggaacctcaaaagcttacggggggcctgttcatggcctaatccccgatttcgtggagaatccatctccaccgaccctgccgcctgttgaccctgcggatgatagcacgctggatacgccgctcattccaccggattcgagtggctcagggccacttagcacaccgaaagcaaactttactcgatactcccgttcaggaagtcgtagttctctgggtaaggcggtcgctggatatgtccgcaatggagtgggggcgcaggcagggccagccgccgtatgggggcctcacgcgctgcagcaggggactgctcggtctcatcagcgactatcagcagggaggtgctactcaggctcttgagcgcttcaatcttggtaatttggcagggcagtctgcatcgactgctcttctctcccttgttgaatttttatgccctccaggtggttctgttgacgaggggttgcgcggcaggctatgctagagaccatcgccgatatgtctgatgtaggagaggagaattttgatgagctcactcccgatcaattaaaagaagtctttattggtttcgtggttcactccattgaagggaggctcatggcggatattggtaaaaatgggatcaagttaccagacgacatagacgctatcgtcagtatccaggaggacctgcatgatttgttgatggagctactcgtacacagctccgtgaggagctgaggaatcttacagggctttcaggggatgctatagacagaaaagtggaggagatttacaccgtggcatttgaattacttgcccgagaaggggagagattggaatgagccatcataccttagttgcccgtttgggcactgacgataactccgatttacagctcagccgccaaagcacgcatctgacagaaattaattttctcaaagagaacggtaaactggatttcggtctcgggcaggcgctgaatggtttgagtgatcttggtttaacgccaatggatgtctccgtggatctggcactactggccgcaacggtgactgcggcggacacccgaatctcacgtgggcataacgctcaagatctgtggacgcgcgaaattgcactttatatcccggtagcttccccgacattatggaatagtcagactggattgctcagcaggatgttgaattttcttaccggcgaccgttggacaattcatttccgctcgcgccctgttattgagcacgggctcattcagcgatcctctaaggaacgttcggtgaaccctacttctgtttgcttgttttccgggggctcgacagcttcatcggtgccattgatttattatctaatggggggaacccccccttctgatcagccactactgggatacgactaccagcgtttatcagcagaagtgtgctcagctgctgtcggagcgatatggacaatcgttcagccatgtgcgagctcgtgttgggtttgaaaaaacaacgattgagggagaagatggagaaaacacccttcgtggccgctctttcatgtttttctcgctcgcgacaatggccgcagacgccctcggcgggccggtcacgataaacgtccctgaaaatggtttgatctctctcaacgttcccctcgatccgcttcgtgtcggagcgctaagtactcggacaacccatccgttttacatggcgcgttttaatgagctgctgggcaaccttggcatcagtgcacatctggaaaatccctacgcctacaaaaccaaaggtgagatggctatccattgccatgaccatgcttttctaaggcaacacgcggctgacaccatgtcatgttcgtctccgcaaagtacgcgttggaaccctgcgctgaatgagcagcaatcaacacactgtggccgatgtgttccatgcttaatcaggcgagcatcattgtttacagctttcggcacggacgatacgatttaccgtatcccggatctccgtagccgggtactggacagctctaagcctgaaggtgaacacgttcgggcatttcaatttgctctggcaagattggcgcgatcaccgagtcgagcaaaatttgatattcacaaaccagggccgctcagcgact

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		atcccgactgcttagctgagtatgaaggtgtttatctgagaggaatgaaagaagttgaacgcctgctgagtggagtcataacgaggccccttacatgaaattagcaggacagaagcccgctccacaatgggtcgattttcactgtcatctggatctataccccaatcactctgcactcatccgtgaatgtgacatttcacgtgttgccacgctagcggtgacgacaacccccaaggcatggatgcgtaaccgggagttaacttccgattctccttatgttcgtgtcgcacttggtctacatccccagctgattgcggaacgtgagcatgagatagcgttactggagcactatctcccttctgcacgttacgttggggagatagggcttgatgccagcccgcgcttttatcgcagctttgaagcacaggagcggatttttcccgtattctgaatgcctgtttcgagcaggggataagattctcagcatccacagcgttcgcgctgcagccaaagtgttgggacatttggaaaacaccagacttactgaaaattgcaaggctgtcctacactggttcactgggagtatctccgaggctcgacgagctgttgaacttggatgctatttctctattaatgaagagatgctacgttctcctaaacatcgaaagctggtgtcctttttgccttcgaacgtatcttgacggagaccgatggaccttttgtgtttcacgaagaaaaagcgatacaccctcgtgatgtgcagcgtacggttcatgaaatcgcgcagatccaccacgtatcggacacagatgctgctatgagaatactttataatcttcgaagtttagtcaccaatagttctcacagtgagaatagttcatgaatctaattagttggattaatacaggggaatagttgaatacttcagtcccctaaaagctaatatgctctatgtcatctaatgataagtggctccaaagagccacttatcattaacttttctaaagggaggtagaagt (SEQ ID NO: 32)
36	pLG038	ttaatgcaaacgcatcaggaagggcagacctagtcacatgtagaatacgatagcaataaaaaagtctaattagaatgcaaattgatgcaactctatgccctccaagaactccaaacctgaaagatttatgtaaaacatagtgttcgtttcaccaaaatacatataaactacattaaaatagaaatttgtctcacctataagccatttagacaacagattaatgaggtttgtatcacaaatgaccacaaacgagatacttttcgcagcttatcagtcttggactcaaaggggataaagttgcttttgttcggcaggcttcgaaactcgcgcgttcctatgattctatggggctgcctgagcttgcttcagccattagaggtagtattcaagataaaaacacgtttaacttgcagaaagtatcacgcagtacatcacctatttttgaacgtcttgatacattacctgtagataaagaaactaaatttgatttagcagacgtaactcaaccgtcttctgaaattcaactcccattgttgaaagatagcactctgaaaaaaattaaagaatttttgactttcactgaacgagctaaagaattaaaggatgccggtcttggcgtgacatcctctatgattttatatgggccaccaggttgtggtaaaaccttgacatcaaaatatattgcatcctgtctaaatttaccgcttcttactgcaagatgtgactccttagtctcatcatatctggggtctacttctaaaaatatcaggcagctatttgagtatgcaagtaaagcaccatgtgttttatttctagatgaactagattctctagcaaaggctagagatgatcagcatgagttaggtgaactgaagagggtggtggtttctttattgcaaaatattgacaatctacctgaagaaacaatattgattgctgcaagtaatcatgaaaatcttctagatagcgcagtttggaggcgctttgagtatagaatatctattggattgcctgattttgaagtcagaaaacaactatttgaacaatattcaaacataaaagctacatatgacgattttgttgatgaccttgcggaaatatcatcagggctaaactgctcatttatagaacaatgctgcttaagatctgagcgacatgctctggtttacaataataaacaaatcgatacccgattttagtcgaggctatcttagaagcgaagggagttacatttgatgaagaagataatttacttataaagattgtgaccactctcagagaatacaatcccaaaagatttacaatacgaaagatagcaaaaatactagggctttcaaatgctaaagtgtcaaggctaactaagaactatagagagatattatgagtaacaaagaaagaccaataaaaataattgaggcgacacctcaagattttactgaaaaacatataatttcggaaagaaacaacctatccgaacagtaacaactagtctaaaaaatagactcaaacaagaagtcgatgacgttaaaaatttttttccagagctcatttaaaaaatggcccaatataccggcggtggctagagttactcttcatgaaaaagctcttgctaagtcacatcgcccatcaagcctattaggtgataatacatgtcccgtaataggcagtgataattttggagaattacttataagtgttactgaaaaagggttagcacaacttcgcaaaaaaattgaaaatagcactaattctcataatgggacagtacatattgctgtaattgaaaagatcgaaccttttagtcttaaccatgatgttatagataaaaataaatcagatagtttttcttctgaaactctttgaccataaagatagaacaactaaccgcagtatcgacaaagaattaatggaatttgcagatgaactaggaatacaaaaacccaaaaagtatgatatcagttcagatttgagtatatatgaagtaaaagggaatgataacatcgcccaactggcaagttttattggcatacgaaaattagaacctatgccaacatttggtcttactcatacagtatcgcaatatattcctgctgaaactctagacctagatgattttcccttacctcaagaggataaacattatccactactcggaattatagatagcggagtcgatcccaataacaacatacttaggccatggatttgggatagtttagatttagtaaaaggagaacacgactattctcatgggaacatggttgcaagtttagcaattaatggaagatggttgaataactatgctggtttttcctcaatgccaagctgaaattgttgatgttgcagcctttcccaaagatggtacgctcaaattgccacaattaatgaaagctatccgagaggctgtgaccacctatccagaagtacgtgtatggaatctgtcattaggttgtcaatccccatgttctgaagacagcttctctgaattggggcattttttaaatgcacttcatgatgagcatgattgtcttttcgtcgtagcatccggcaactacatttatgatcctcaacgaacctggcctcctcaagaattaggtgggcatgacagaatatcagcccccgcagattctgttcgttcattaactgttggctcagttgcccatttagaatcgtctgactctgtggtcaaaagatttgaaccttcatctttttctagaagaggtcccggcccagcctttatacccaaaccagagataaatcactttggaggtaattgtgacagtaaattaaactgtgaacataccggaatcatagctattggcgaggacaatgctctttgcgaaagtattggcacaagtttatcagcaccgttaatctcaagtttagcggcatcactgtggcatgaactagatgttaatggttctatttcaccatcgcctgaacgtatcaaggcactattaattcattctgcgttaaaaaactcaccagccaaaacggagcattatgcgtttaattatcaaggatttggacgcccaagcgatcatataaatgatattattggttgcaataaaaatgagattacatttctatttgaaatagatacccgagaaggtattgaattcagtagaacgccatttgtaataccacagtcattacgtactgaggatggaaaattcacaggtgaaattattatgacactcgtttattctccaccgcttgattatgactacccatctgaatattgccgttctaatgtggatgtgtcattcgggacttacacttatgatccagttaacgctaaatggatacatagcggaaaaattccacaaataaaagaaaagagtgaattatttgaaaaggtactgatagaaaatggcttcaaatggtctccagtcaaagtttatagaaaacaatttccgcaaggtataaatggggagcaatggagacttaaacttgatgttcagagacgagcagagcaagagcctctatcttcacctcaacgtgctgtattggctattacgttaagatctcttgccaattctactacagtctacaacgaagccgaggttgaaataaataatcttggttggaaagaaactgatattgttgttcgtgaacaaccaaaaatcaggattcgtcaaaaataagcattatggtcaccttttataggtgaccattta (SEQ ID NO: 33)
37	pLG039	atagaacgatgaaggatggaagctacatattctcggtactaagatttatttttctgacacaaaatgaccatttggcgttacataatcccaaaaaaaacgtatcaaaaatctcaaaatgcgttacgattagagagtattttgattctgcgtgctcatttttttgattgctgtggctttttgttgtgggagtgttgaatggattatttatcagaagtgttaaaaatcattgaaggtgcaacaaaggcaaatgcttcgatggctagtaattatgctgggttgctggcagataagctcgaacaaaaaggggaggtcaagcaagccagaatgataagagaaaggttgcttagagctccccaggcgttggcaggagctcaaagggctggaggtgggatatctctgggctcattaccggtagatattgatagtcgactcaacactgttgatgtcagttatcctaaattagacagttcagagatttttctgcctgcagcaatcagtacccgtgttgaagagtttatcactaatgttcaacgttatgatgagtttgttaaagctgatgcagcattgccgagtcgtatgctcgtgtatggaaagccaggaacaggtaagactatgttatctaagtacatcgctacccgcttagattttccacttcttacagtgcgttgcgatactttgattagtagtttattgggacaaaccagcaaaaatcttagacaggttttcgattatgtaatgcagaggccatcagtgctttttttagacgaatttgatgctttagctggagcaagaggtaatgagagagatataggtgagcttcagcgagttgtcatttcactattgcagaatatggatgcggcatcagaggatacggtaattattgcctcaactaaccatgagcaacttctggatcctgcaatctggaggcgatttagcttcagaattccaatgcctctgcctgacatacatcagagagagttaatttggaaaaatcgtttaaagaatatgatatgtagcgatctagatttaagtgatttatcaagaaaatcggaaggattatccggagcaataattgaacaggtgagcttggatgcacgtagggatgcagttattgaaggtg

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		caagtgtgataaatcaccataaattgtataggcgtttgtatcttgctcaatcgcttatggaaggtgtaaatttaagcacttacgaagatgaaattcgttggttacgttctaaagataaaaaattatttctatcagagttcttgctaatttgtacaaacttacatcaagagtaatttcaaacattctgaaggagtcaggagcatatgagcagaaggggtacacagtttagtaacgcaaaagttacaaacccaatgttaagaatccctttttccagtagtgacttgggtgcaatagtaaacgctggcggtggggcaaaggtattggttgatgtaacagccgaatatagacaagggctagtaagaaatttaacaaccagtaaacattatttagaatccaaactttcagagtaccctggaagcttgggtacttttggttttcaaattaagagaccagggaatagccaaaacgcataggccgaacaaaattgctcaagaggctggattgcaaaatgccggtcatgccaaaatagatgaaatgttggttgctgctcatgccggctgttttgacgtattagagtcagtcattttacatcggaatattaaagcgattttggctaatctaagcgcgattgagcgcattgaaccttgggatgagaataggaaggttccaggaggcactgatggtttgtttgaatcatcaaacatccttgtacgactatttgagtacacaggtgaagatgcaacttacaacaactatgaaaacgttatttctatattagaacaacacggagttaaatatgatgagattagacaaaaatgtggtcttcccttattaaggataatggatttatccccaaatgatagatatatattagacattctcattgattacccgggtataagaacgttaattcctgaaccaaaatattcagcattcccggttagtgtaagtgattctgttggcattgaaacaaatagctttcccgtaccatcagaagaattacccattgttgctgtatttgacactggggtaagccccatcgcggcaacaattactccttgggtagtgagtagggaaacatacgtaattcctcctgatacgagttatgaacatgggactatggtgtcttcattgatatcaggcgctcatttttaaatgacaatcatccatggattcctgatacaaatctaaaatccatgatgtttgtgccttagatgaaaatggatcttatatatcagatttaattctgaggctagcagatgctgtaaataaaagaccagatataaaagtctggaatttgtctttgggaggcggaccatgtaatgagcagacgtttagtgattttgcgatggagttagatcggctcagcgataaatttggtattttgtttgtagttgctgcaggtaattatgtagatgaacctatacgtacatggccaaatcctgatccgcttggaggtgctgatttaatttcctctcctggagagtcagtccgagcactaacagttggttcagtttctcatatggaagctaatgatgctttaagtgaaattggaacaccgacaccatatactcgtcgtggccctgggcctgtatttactccaaagccagatataatccatgctggcggtggggttcatagaccttggaatgtaggagcaagcagtttaaaggtcgtagggccagataataggctttgctctaattttggtactagttttgctgctccaattgtggcaagtttagctgcgcatacatggcagagaatagccactaatacagactttaatgtttcaccatcattgattaaagcattattaattcattccgctcaattatcttctcctgattactcgccaagtgaaagacgctatttgggagcgggaattcctaatgaagttattgagaccttatatgatagtgatgataggtttactctgattttccaaacattcttggttcctggggtgaggtggagaaaggataactatcccataccatcggcacttattcaaaatggaaaatttaaaggtgagattgtaattactgctgcatatgcaccaccactgaaccctaatgccggcagtgaatatgttcgcgcgaacgtagagctaagttttggcttaattgagaataatactataaaaggaaaagtgcctatggaaggagaaaacggtcaatctggatatgagagagctcaaattgagcatggtggaaagtggtcaccagtaaaaattcatcgcaaggcatttaataaaggaattacttcgggtaactgggctcttcaggctaaaacaacgttgagagcgaatgaaccggccttaatggagcctttacctgtaactattgtagtaactttaaaatcattagatggaaacacacaagtttatgctgatggcgtaagagctttaaatgctaataactgggctcactatccattgcctgctcgtgtgccagtttccgtataacaactatataaatcaaacccgctgtagcgggtttgatttatttgtgggtgtgttttataaaaataccgcccatacacaacaaaatacaa (SEQ ID NO: 34)
38	pLG040	gggacactcaggttacataacaatgagtgatacagttcacgtagtgaaggtactatgcctaggtgtttgattacactttgatcattgatgatacgctcatgaaggtattactttcctgtaatgagcaggtaggtaacgatgtcgaactaaatgaatttatagtaaactttgcaacaagagaacaagggagtatgaggggttatggctactgcagagcagatcaaagctttattgaaaagccacgttgatcgtgatgatcagcgtttcttttctattgctttgcaggtggcagctaaggaagcaaggcaaggtcatcataagcttgctaatgatataaaaaacttagttgataaaaatcagaaaacaacgagttctgtaggtttagttgaaaaacgacttacaccatttgttaagcagcctgatggtgatcttaaggggttacttgagcaaacgaacaagccagtacatcttcaagatctggtgatttctggaagcgttagggaaagattgaatcaggttctgcttgaacaaaaacagaaagataaactttctgagtttgggcttattccaagaagaaaaattcttttcactggtcctcccggtactggtaagacaatgtccgcatcagtcattgctacagagttaaagctaccactttatacagtcgtcttagataatctaatcactcgctatatgggtgaaactgcagctaagctgcgtttaatttttgaccacatacggcaaacaagagctgtatatttttttgacgagttcgatgctataggaactcagcgtggcgctcagaatgacgttggagaaattcgtagggtcttaaattctttttaatgtttgtagagcaggatgattctgagagcatagttttagctgcaaccaatcatccagagcttttagatcgcgccttatatagacgatttgacgatattataccgttcacaaggcctgaggataatctaatcaggaatcttattgaacagagactcgctgtctttgacctcggtaatttattttggagtgagatcattgatagtgcttcaggtctaagtgcagcggagatcacgcgagcaagtgaagatgctgccaaagaatcagtgctttataatgcaaacaatattacaaccgatttgttagtaaaggctataaagcgtaggcaagaaagtagacaataaggatgaaatgactaccaacaagaggcatatttattaaacggctatgtttcccccgaaaactatcgctctaggagcaatggtcgtagtccccaagtcccagctcgtgatcgagcggtacatggtatatcattactaaatcagtatagccgtatattgaatcattatgatgaaagaccgaggcttccccctgttactgatgaaaaagggatttatgttaggctaatcagttttgaacaatgcgatcttcctatagataaaatcgataatacttatttcaagctttgttctttagttaaatcaaataatcgtgaaactgcgattatatacattaatgaaaatgacagaactaaattcactaaaaaataaatgactatttgaatccatcgaaggatggtatcgagttccctagaaatcatttgttaattgatagcatacaaaatatcgagttagcagatataacttctttctggacagataaaaaagatcttattccggatgatcacggtgttgaaaagtggtttgagctttggcttaagggtaataaggaggatgtgctaaatattgctcggcgtttatgcgaaagaattaatggaaggctcgggaatacttctattaattttttcgatactactgttgttcttatccgtacgagtctatcgagattaaaagtttgtcctgaattaatatctaatttaaaagagataagatcagcgagggatgatatatcagttatagttaattccttacctacagaacagcatcagtgggcagaaaatgttgctgcaagaattacgcgtaacaatgaagctgatgtttctgtttgtatattagatacaggtgttaactacaataatccactattatctagatttactaactcatcactggcagctgcttgggacatatcttggccacttttcgatgattataatcaaaggccttataatgaccacggttccagacaagcaggactatgtgtttatggagatttcctgtctgttttattgaacgatcaggacatttcgattccgtacaatatcgaatcaggaaggatactacctccaagagctactaatgatcctaatctttatggagctattactacaggaacgtcaagtcgtctggagctggaaaacccgaactggcgcagagtttattcgcttgctgtgacagcagagcctaatactcttggaggccaaccgtcctcatggtctgcagagattgacaagtttagttttggtttagaggatgatatccgcagattatttataatttctgcgggtaactctcaacctacaaatttagaattagattattgggattcagtgactcttgctgaaattgaagatcctgctcaatcttggaatgcattaactgtaggggcgtatactgataaaacaacccatacagaccgcgaatatgatggttggtctcctttcgctatgtcagaagatattgcaccgtcatctcggtcatcggtatcctggggatggaaaaagcatgccccatataagccagatttagtagaggaaggcggaaacaaacttatatcacctagccgtgatgaaatcacaaatacaattgaattatctttgctcacaacctctggcagggcaacaaatcaattgtttgaagttaattcagatactagcgcagcctgtgctctagtatcaaaacatgctgctatgctaatggctcagtacccagaatattggcctgaaactattaggggattacttgttcatacagcaagatggactagtcgtatgcacgaacgatatagaacagaacgtgcacaggggacaccaaaatcggctaaagaaagcttattaaggatggttggttatggagtacctaatttaaatcgagcaatgcatagtgcggaaaatgcacttacattaatatctcagtcggaaatcaccccatttaaaagagatggttctactgatcctacattgaatgaaatgcatctgttttcactcccttggcccgtagaagctcttcgcttactaccaccagaaacaaatgttattttaagaatcacattgtcgtattttattgaacctaatccaagtcaaaaaggattcagacgacaatattcgtatcaatctcatggattgagatttgcagttattagacctaatcagacccttgaaaatttccgtgcttcgataaaccgtaatgcgaataatgaagaatacaatggacctgaaggagatgcgtcaggatggtttctggggcctcaactcagagttagaggttcattacactcagatgcttggaaaggcagtgctgcagatttaacagagatgaatactatcgctgtctatcctgt

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		tggtggatggtggaaatatcgtactgcgcaggatcgctatattaacaatgttaaatatagtttattggttagcatagatgtaccagatgagaacattgatatttacagtgagattcaaaacattattcaaattgataatcaaatagatattgaacattaaggttttatgcctaaggtttaatgagtttgaaatgaaaaatcctttactaattggctgggtcgatgataaagacctggccatctttttatacggaaatgatttatgttttattttactaaatttatattagaaccatcgtgcagattgtgataattccttcatactgatttttacctattatagttgattttgttgcttgatatctctctttaatacaacggcgtagtac (SEQ ID NO: 35)
39	pLG041	cggattgaatctgtttatgaaatttggctgctatcaactaatgggcgttaagttgattgtatgatctgattgataaagaaggggctaaaaatctcctcttctttgcagcagtttactgcggtcttttgtgatgcatcagcataaaacgttttacttgtggaccctaagaaatggagaacattatgtcgactgtagatacctctacagcagaggaactcaatcaaggaggctcagattttattctgacttccctcgaggctatgcgtaagaagttattggaccttacgtctcgaaatcgacttttgaatttccctatcactcaaaaagggtcttcactacgtattgttgatgaattaccagaacagctttatgaaaccctttgctcggaaatcccgatggaatttgctcctgtgcccgatccaactagagcgcagctgttagagcatggctatctcaaagttgggccagatggtaaagatatacagttaagagctcatcctagcgctaaggattgggcgcacgtcttaggaatccgtacagatttgatttaccagatagccataaaacggttgtttctgattcagatagagagttgctggaaaaagcccatcagtttatcttgcaatatgcccaaggccagaatggaaaattaacagggattcgttctgaatacgttaatcaaggtatagctttgtcagcgttgaaggaggcgtgctgcttagcaggctatgaagggcttgaggattttgaacgacaggcaaaggctgggaatgagattagtatatcttcttccaatccctctcatgacgataatcggatacaggctctgctttatccaaatgaactggaagcttgtttgcgcgccatctatggtaaggctcaaactgctttggaggagagtggcgccaacatcttgtatttggcgttagggttccttgagtggtatgaaagcgattcctctgaaaaggcacgttatgcaccgttatttacaattccggtgagatgtgaacgaggaaaattagatccgaaggatggtctttacaagtttcaactttattacacgggtgaagatatttgcccaatctctctttgaaggaaaaacttcaggctgactttggcctcgctcttcctttgttcaatgaagaggaaactccagagtcttatttgcttcggtgaagaaggttgtagagcagcacaaacctaaatggtctgtgaaacgttatggtgcacttagcttgctcaattttggcaagatgatgatgtatcttgacctcgatcctgcccgctggccttgtgacaagcgcaatatattgtctcatgaagtaattcgtcgctttttcaccagtcagagctgtggtcaagagaattccggcttacctggtggcttcggtcagcatgagtactgcatcgatagttaccctgatattcatgacaaggttccactaatcgatgatgcggatagctcgcagcacagtgcgttgatcgatgctatccgtggtcaaaacttagtcattgagggccctcctggtagtggcaaatcacaaacgatcaccaacttgattgcagcagctctgctcaacggtaagaaagtcctgtttgtggcagagaagatggctgcactggaggttgtcaaacgtcgcttggatcgtgcggggctaggtcaattttgcttagagttgcacagtcataaaactcataagcgcaaggtgctggatgatattaatgctcgcttggtgagtcaggcgaccatgcctactatggaagagattgatgctcagattttgcgttatgaagatcttaagcagcagctcaatgaatatgccgcattgatcaataaccaatgggcgcaaacaggcaaaacgatccatcagattttgagtggtgcaacccgttatcgtcacaaattagatattgatgcaacagcacttcatatcgaaaacctttccgggaagcagttggataaagtgacccaattacggctgcgtgaccaaatagtagaatttagccgcatctacaaagaggttcgtgagcaggtgggggctaatgcagaaatatatgagcacccttggagcggtgtgaataacacacaaattcaattgtttgacagcgctcgtatagtcgatttgctacaaacttggcagacatcaattatcgactttcaacatagctatcaagaatatgtagataagtgggcgttagaaggcgaaagccttaatacgcttcaatatattgagcaatfggtagaagatcagtcgaatcttccagtgttgtgtggttcagagcatttcccagcacttagtgagctagattcacccgatgccattgcacgggtgcgtcactatttagataggttcgagttgctacaaggtcattatgtggccttgagccaggttatcgagcctcaaaagctacgacttttagaacaaggacaatcgtgtgactttcctcgtgaagagctggaaaaatatggtgcagcagaggatttcactttacgtgatttggtcaggtggcttgaatccatccaatcaattcatgatgagttatcatctatttatgcgcaattaaacgatttcaaaaatgctttgccagatggtattgcttcgtatatcgatgattcgcaagctggattgctattctgctctgagttgttgtcgattctgggtgctttaccgactgagcttattagagttcgagatcctctttttgatgatgatgatatcgatgcagtattgcgcgacttaatgtgtcaaatcgaaacattgcgtcctttaagagatggtctatctactttgtatcaattggaccagttgccttcccaagagatgctcgcgcatgccgttgctgttatccagcaaggggattatttgcatggtttaagagtgattggcgtagtgccaaggcactgctcatggcgcaatctcgaaagcctgacactaagtttgctgagttaaaacgctgctcagctgatttgctcaagtattcggagctgttacaacggtttgaacaaagtgactttggtaatcaacttggtaatgcattccgagggttggacaccgactgtgaacaactcatgttattgcgtgattggtacaagaaggtccgagcttgttacgggataggttttggaaagcgagttgcgataggctctggattatttaacctagatggtgagattatcaaaggtgtgcatttaatcgagaaatcgcagattagctcaagattaatgactttggttaaacgggtcgagcacgaggctaagttattaccgcgtatttctagcttgttggaagaacatgcatcttggttaggtgagcaaggtgtattgatgcaatcttaccgacaggtgcggaatactctcattgccttgcagggatggtttatcaatccagatatatcattagagcagatgactcattcctccgagattttgcaaaacataaacgatcttcagatatcccttgaaaatgactcgttacagttaggggcgtttttacaattaaccccattggcttgcggtgcgtataaaaataatcaactgacgttagacactattaacgacacgctgaatttgccgagcaactggttgataagataaattgcgtatccttggctacccagatcagacatttggctagtggtagtgattacgatttactatgtcgtgatggtggagaaatagtttcgaaatggaatgaacagattaaaaatgctgagttatatgcgctagaaacaaagttagagcggagtcagtggctcaagtcgactgatggttctcttaatacattaatcgagcgcaacgaaagagcaatacagcaaccccgttggttgaacgggtgggttaactttattcgttgttacgagcagatgcatgaaaatggattgcagcgaatctggagtgctgtacttgcgggctcgctcccgattgaaaaagttgaattgggtttagcattagcaattcatgaccagctggcgcgggaggttattcacatccaccctgaattgatgagagtttccggctcacagcgcaatgctttgcagaagtcatttaaagagtacgacaaaaaactgattgaattacaacgtcagcggattgcagcaaaaattgcttgccgaaatataccagaagggaattctggtggtaagaaaagtgaatatacagaactagctttgatcaaaaatgagttgggtaaaaaaaccagacatattccaattaggcaattggttaaccgtgcatgtaatgcgctggttgcaattaaaccttgtttcatgatggggccaatgtcagcagctcattacctagaacctggacgaatggaatttgatctggtggtgatggacgaagcgtctcaggtgaagccagaggatgcattgggtgtcatcgcgaggggcaagcaactagtggtcgttggtgacccgaaacagctaccaccaaccagtttctttgatcgaagtgccgacggagaagatgacgatgatgccgcggctttaagtgatactgacagcattttggatgctgctttgccactgtttcctatgagacgtttgcgttggcactatcgttcacgacatgaaaagttgattgcatactctaaccgccattttataacagtgatttggtgatattcccttccccaaatgctgagtctccagagtatgggattaaatttacctatgtgtcaaaaggtcggttctccaatcaacacaatattgaagaagcccaagcagttgctgaggccgtacttcatcatgcgcatcaccggccgggtgagtcactcggggtagtggccatgagttccaagcaacgcgatcaaattgagcgcgctatcgatgaattgcgccgaaatcgccctgaatttaacgatgcaatcgatggcttacatgccatggaagagccactttttgtgaaaaaccttgagaacgttcaaggggatgagcgtgatgtaatctttatttcctttacctatggaccttctgagcatggtggaaaggtttatcaacgctttggacctatcaattccgatgttggctggcgtcgcttgaatgtgcttttcactcgatcaaaaaaacggatgcatgtgtttagttcaatgcgttctgaagatgtattgacgagtgaaaccagtaaacttggtgttatttcgttgaaaggttttttacagtttgccgaaagtggcaaactagattccctcacaacgcataccggcagggctccagatagtgactttgaggttgctgtaatggaagcactcaatcacgctgggtttgagtgtgaacctcaggtaggggttgcaggattctttattgatctagctgtgaaagatccaggttgtcctggccgttatttaatgggcatagagtgtgatggtgcggcttatcactcagctaaatctgctcgtgatcgtgaccgtttgcgtcaagaggttctggagcgtttgggttggagaattagccgcatttggtccactgattggttcagtaatcctgatgaggttctatctccgattatccgtaaactccatgagcttaaaacattggctccagacgttgttgtaccttcctatgaatatgtcgaaacgattgagtcaagcgctgaagtggcgtctgactcaattgattctcttatgcccaatttggggcttaaggagcaacttaagtattttgccacacat

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		gtcattgaggttgagcttcctaatgttgatgctgatcgtcgtttgttgcggcccgcaatgcttgaggctttgctggaacatcagcc tttatcacgttccgagtttgttgaacgaatacctcattatctgcggcaagcaacagatgtatacgaagcacaacgctttcttgacc gagtcttggcattaattgatggcgcagaggctgaagcgaatgatgcagcgtttgagtctgaattggcataattagttaaaggtaat aagaacagtgacaactgtcgg (SEQ ID NO: 36)
40	pLG042	gctatcctacctcagattactgggctgacctaatctatagatcaggttctctttatactttatgttagcgaaatactaagatgctt cttagtgacgacctcttgacggtagaggacgcgtgcatagattttacaatcactgcctttcgccccctaacctaatccgcgaatga tgcatcctgaacttgcgcgccagttcttatactcgccgtcagagcaatcaaattgctgatgctttctgcctgttcaaggcatctcc tgtcgtcagcaatactgtgcatatttgattgatttcctcttaaggagaattagtttcatgggtattaaagcgcaggtgagtatcgc gcacaagctggggttcacatcacaccaaaatgcagttccgctgttacgtgagcttatcttgcataatgagtccgaagagacatttc aggatctgacactgcatctgaggaccgtgccagctgtgctcgaagaaaaaaaatggaatatcgatcgcctgcttcccggtacttca cttgatatcagagatcgggatatcaaacttaatgctgaatggctagccgaactgactgaaagcgtactctgcgaagtcacgctaag tttgcgccagggtgaggaagaactcttcattacccattacccgcttgaggcactggcgaaaaatgaatggggcggcagtgcaatga ttgaattgctcccttcatttattattcctaatgatccggctgtggatcgtgtactcaaggcaacctctgatgtccttcgccgtgca ggcaaggatgacgctcttaatggttatgaaagcaagtcgagaactcgtgtctgggaaattgcctcagctctctggactgctgtttg caacctcaatatcagttatgcccttcccccagccagttttgaacgcaatggccagaaaattcgcactccaggagccattctggaag gaaaagtcgcgacctgtctggatacaacattattatttgcttcagcactggaacagattggtctgaattcactgctaatgctcagt gaaggtcatgcgtttgctggtgtctggttacaaccgcaggaattttcgcagctagtgacagatgacgtctctgcggtgcgcaaacg tgtcgacctgaaagaaatggtcgtatttgagacaactctcgcgaccagagctcacccgccttcatttactcaggcatctgatgaag cgttaaagcatcttaacgaggatgtttttcacgcagccattgattcccgtcgcgcgcgtatgcagaaaattcggccactggctctg gggggcactcgccttgaagaccagtcggatgcctgcgaggttattttgcatgggtttgaggaagcccccctatatccccgatgttga tattgatatcgagacaactggcgaaaaagaagccgggggcggctggtacagtggcaacgaaaacttctggacttaaccacccgta accgcctgttacacctgtctgaaagcgctaaaggcattcgtttgatctgtgcgaatccgggccatcttgaagataaactggctgaa ggcaaacgcattcgcattgtcccgctccctgatctcgaaagcggcggccgcgatgccgaactttatcagcagctcacaaatgagaa cctgcaggaagaatacgctcagattgcgctggaacgcggtgaagtcgtctcctcaatggaaaaataccgcctcgagtcatccctga tcgacctctatcgaaaatcgaaaagtgatctcgaggaaggtggtgccaacactcttttcctcgctgttggcttccttaaatggaaa aaatctgctgatgaccccaaaagttactctgctccactgatactgctgccgattcaacttgaccgtaaaagtgcactttcgggcgt gaccatgcgtttgctggaagaagagccccgcttcaaccttacactgcttgagctgctgcataatgactttgctctgacaatcaacg gcctcgatggtgatctacccaccgatgaaagtggtgttgatgtggatggtatctggaatatggtacggcgtgctgtacgcgacata cccggtttcgaagtcacccgcgatgtcgtgattggcacattctcttttgccaaatatctgatgtggaaagatctcatcgaccgggc acctcagctgatgcaaagtgcgctggtaaagtatcttatcgaacgcggccaggaaaatgccgttctggataagagcggagaagtca tcaacgctcatgaactcgatgacaacatcaatacgcaggatcttttcttgccgttgcctgcagattcctcgcaaatcgccgctgtt gtagcctctgcaaaaggcagggatttgttctggatggcccacccggtaccggtaagtcgcaaaccatagccaatatgatcgcgca taaccttgcgctaggcaggcgcgtactttttgtcgctgaaaagaaagcggcgctggatgtggtctatcgtaggcttgaggcccagg gactcggtgaattttgtctggaactgcactcgagcaaaacgtccaagatggattttctgaaacagctcgagcgggcatgggatgcg cgtgatctactaaccaccgaggagtggaaggaagaagcggccaaggtgcagcacctgcgtgacaaactcaatgaggttgtccgttt gctccatcggcgctggcccaatggcttaacactccatcaggcaatgggcacagttatcagggatgcaagtagcgccacgccgcact ttagctggcctgcatcgactttgcattcttctgcagagatgacacagttcagagagatagtaaaacgtctggagctgaaccgtgat gcatggaaacagcacggcgatcattttgaactcatcgcgcaggctgactggaccaatggatggcagtcctctctcattgctgcagc aaactcattgcctgcaaccatcgatcaccttgaagacgcgaccgaggcgttactgaaggcgacgggagttactctgctctctaccg agccggagagactgtcgcagttaacttcattctgtgaattattgtcggaagcttacggcattgatctgagtttcatgttcgcaccg gatgccgcaagccgtatagagtcagcgaataaagccgttcacctcctgaaagagattgaagcgacaaaggctaatctgtcagttac ctacccttgtaacagttggcagcacgttaatgtcccacagatcagaaacgcacttgacgtcgctgacaaaaaattctggttctttg cgaccagtgcccgcaagaaagtcattggtgaagttatccgacaacactcgctaacgtcagcccccgacttatccgttgatctcccc attgctgaaactctgcagacattgctgcaacgtctgaccgagcttaactctgctactgtatctctgccgggatgggttggactgga taccaacgttgcacagttgcagaccaccctgcaacttgccgaatctatccgcaattcgcttggtggtttcgcttcttcgccacagc agttggccgagatccgcactgcggtaaaaaacctgattgttgatgccaatgaccttctcggttcgcagggcgttatctccgcacta acccggaaactgcgcacagcgatcgccgatttcaatgatgcacaggttagcttctgcaatctgataaaaccatctgaggataaacc atcgctcccggcactgcgtgactgcgcactcaatatcctgcaacatcagtccgctcttaaagcctggagtgactggagccgtgtgc gtgaggaagcgatttcacatggcctgcaaccagtgatcaacgcgctggtccatcttgactcaggagacatcagcgcggcagagatt tttgaaactgcctattgccgctggtttgcatcgtggatgatcgattcagagccgctgctgcacaattttgtgccggctgagcacat gagtgatattgaggcttaccgtacgcaaaccgatcgtctgtccaaactggcagtacgctacatccgtgcccgtttatgtggcgtca ttcctgcaaaaaatgaggtcagcaagcagggtggttttgctctgcttaaacatgaactacagaaatcccgtcgtcataaaccggta cgtcagatggcagcagaaatgggagatgccatggccaaacttgcccccctgcatgcttatgagtccgctttcagtcgcccagttcct gccctcggaccaggacttgtttgaccttgtgattttcgatgaagcatcgcagattgccccgtgggatgctatcggcaccatggcgc gtggcaaacaggtggtaatcgctggcgatccccgccaaatgccgcctaccagctttttaatcgtgcagccaatgacactgacgat gatactgaagaagatatggaaagcattctggatgagtgtcttgctgccggcctgtataaccacagcctgagctggcattaccggag ccgtcatgaaagcctgattaccttctccaaccatcgctactatgacagtagcctgattacgttccccgcttcggaaacaaagcaaa gtgctgtccagtggtgcaaggttgcaggcgtctactctaaagggaaaggacgtcataatcaggccgaggcagaagcgatcgtcgct gaaacggtgaagcgactgactgataaagagttcgttgcatcaggcagatcgataggcattatcacgctgaataccgaacagcaaaa gctagtcagcgatctgctggaccgtgccagacagcaacaccctgaaattgaacccttcttccagtctgaactggaagaacctgttg tggttaaaaacctcgaaacggttcaggggatgaacgcgatttgatcatactctgcatcgggtacggcccgactgaaccgggcgca aatacaatgtcgatgaattttggaccgcttaatcgcgagggaggctggcgccgactgaatgttgccgtcacacgtgcgcggcagga aatgatggtcttcagctcgttcgatccttcctttatcgaccttaatcggaccaacgcccgcgcggttgctgacctcaaacacttta ttgagtttgcccagcgcggccctgtagctcttgcccaggcagtacgtgggtctgtaggcggttatgactcaccgtttgaagaggca gtggcaaatggcctgagaagaaaggctggcatgttgtcccgcaaattggcgtatcccgtttccgtattgatttggggatcgttca tccggataagcctggcgactatcttgtcggtgttgaatgtgacggcgccacttaccatagcgcagcaacagcacgcgatcgcgata aagtccggagctccatcctgcagggcctgggctggaaattactgcgcctctggtcaacagaatggtggattgataaagaaggcgca ctcgacaggctggatgcagcaataagtcgcctgctggaggactccagagcagcggaagccgcactgattgctgaagcagaaaaaca aaagcagattacgccagtcatcgctcccgtaaccaatgatgtcagtgatgacatactggtttctgaaactacacctgtcgctaatg atgcggaaatatccgcgtcagtaacccctgtcatcccgcttactgccaaagtaagcgaagatgatggtaacactgggctgaggtat gcatctttagcttctcagaataacgacaagccagtgaatgtcggtaagtatgtcgttaacgatcttcaggaatggtgcgacaggac

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		agatgcagaacaattctatatcgctgaatatgatgagacacttaaaaccctcattgaagcggtggtgacaagtgaatcaccggtcc tggatacaacgcttgtgcaacgcatcgcacgtatacacggcttcactcgcgccggcagactgatacgtgaacgcgtaatggaaatt gtggatcaacactatcaccttgcaaccgatcactcaggtgaagacttcgtctggctgtccgcagcgcaacgtgctgactggaatgt gtttcgtttgccagccacggataacgacattcgtcaggttgacgcgatccccagtgaggaattacgcgcactggcgctgagtattg aaggtgacaataagatacaggaaatgacccgctcgcttggcattaaacgcctgactagtcaggcaaaaaaaaggattgaatcagta cttgatgttgtttgaaggtcaaccgtgtggaaaacctcttttagagactaacagtctgaaatatagagtcttattcgatcatcttg agaccgaatgtattagagtcgatttctgacacctcttatcgtggttttctgcatcaccaacatcgaccagttgggcgtaatcaagg aggacgtctggaaaacgaatctatggtcactcccgtttttgcaacaccgattttgacaataagttggtttgcttgaatctattcgg catcagaatggaatttttttccacgcctcgatgagttccgcgcctgatgaa (SEQ ID NO: 37)
41	pLG043	aatcccaccctgacaaaaggcctgaaaaggtcttttgtcatttcttcacagttagagccctatcgagacgcgcaaggaagagtcgc gccagcctgtttttacgctagcgctctgctagtgacagccagctcacagggagtgagctggcagtgtttaacgtcctaccgagggg cgtaaattgcacacagaggttaatgatggctaaagcgcactccacgccgctcaacgatattgcgattatcgctgcgaatttaaaag accgttataaaaatggcttccctgttctgaaagaaattgtgcaaaacgcagatgacgcacaagcgtcatcattaatctttggctgg agccctggtattgctggggcagatcaccctttattgggcgatcccgcgcttttctttatcaataatgcgccgctgacactcgaaga tgtagaggggatcctctccattggcattggcactaaaccgggtgatgaaaatgcggtggggaaatttgggctcggtatgaaaagcc tgttccatctcggtgaagtattttttaccagtcctttgactggcatactgcttcggccaaatcagacgtttttaaccccctgggac agttacagatcttcttgggccgaggtgagcgagcaggataaagttcgtattgaggatgaagtccgcgcaattacccaaaatgcgtg tgatgattatttcgttgtctgggttccgctgcgttcagagagtatctatcaggcgcgccaggatgatgaaaactttattattgtcg gcgaagactatcgttatgaggtgcctgattttatttcagacccgggactcggggataagctcgccagcctgttaccgctgatgaaa accttgcaggacattgagctggtcgtgaaaacagggcaggggtatcagcgtcaaatacatatctcgctgcctgaaaaggcaactcg cccacaatttaccaatcttaatggtgctggggaatggcaaggccacattaccgttcagcgtgctggattgccggaccctcagcaaa aattctacgtcgggcatgaggttttgctgaatgctcctgagttttctgccctgaaatcacaacgcgcctggccattcagttattca cgagaaggtaagaagactgcggataaagcgctgcctcatgccgctgtggtgatgctggcggagaaagtaccagaaggagaggcaac gctggcggtggaatgggcggtgtttttacctttgggtgagcaggacaccgcgcagcatgcgcagaaacaaacattctctatttctg gtcagtactcgtatcaaattattctgcacggttactttttcatcgatgccgggcgagtgggtatccaggggctggctacactcacc agcgccacgccgttattcaatgccccagattctccaggccaggaacaactggttcaggaatggaaccgctgtcttgctactcaggg aacgttgccgctattaccgaaagcgcttgcctctcttatgtcgcttattcacgccagggatgcggaaaaagcggcaatttcggatg gtgtgcgtagagctttacgcaacaataatgcctggttccactgggtaacgttgtaccatctgtgggtatgcgaactaacgcgggat ggaagtcagtggtgtttagttgatgcgaacactcccgttcgtcgattgcctgccacaccttcaggtgaagcgcatcgcccctggga agtgctgcccgctctggaaagtctgggtgtaacgcaccgatttatcgatgaaacgcagcagaatatctacaacgaatttaaaagta agtggcagttgtcggagattcaggtgttgctgcatagcgtacccgaaatggtgttcactagcttaaagcttacaaattatctcaat caattgctgaaagaactgccgattcagtcagacagctttgtgcttgacctgattgcattgctcagaaaaacgttatttagcgtgcc gctggttgagctctcacgtaaccaggcggcgatcggagaattgatggcgttcattcgtccgacctggcgttacaggattgccattg accgtcaggagcaggccctgtgggaaacgcttgggcgtaccgctatggataggttgttggttcctgcttttctcgataacagtaaa gaacctgccagcgcatctctgaattgggagacggttggcagcctgctgcaagcgatgcagaaacaggcttctgccagcgataactt tgaaaaattggtgcgggattttattggcaagctctcatctcccgatcgtcaggagctataccgtcggtttgataccttgaaggtct ttaaggtttcacagccaacggggatatcttacctggagacgcgctgtcacttgcttgaactaaaacaaaagcgaaggatattcaaa cttggcgggagcgctaattttggtatgggtttaagcgcattgttgcagcaggcattgcttgaaaaagaaatcgtattgatcaccaa tgatattaaccagaccttatttggtggttctgaatattcagaagcaaaggagtgtgacagcgaaggggttatccatctgcttgagc ttcaccctcgtctggattcgccgacaaaacgtatcgatttactcaataaaatggctgcggacggggacaaatttagcgccggagat cggcttgtctatcgctatctgatgcacggtaattcggatgatactggtgaagctgaattgtggaaggcgggtaaagcgcatcccgt atgggcaaaaattctttctgatgccgattcggagcaggtcaagtggactattatttcgccagaaattgagcagaatcttggactga ctcccggattcgagaaggcgcttaggcttgatagtgtaacgccggatcatgtgatccaccgcttcaaagaaagccttgaatatctg gagtttgatgacttatctgcagaagatgcggaagaagttctgatgcacattggccgctctatgggcgaaacaatgtggcggcagat ggctcttcatcgtagggaaggcaaagaggggtatatatcccttgatgatcgttgtttcttgcgtggggggcgcattgaactgccca ctgaattgaatgacaacgtgacgttcatccaacccgccagtcagccagagatgcaggatcagcagcgcaaatatctgacaatggtg aacgccgaacatgcggtcatgctggctttatccgggccgaacccggaacgttactgcgactttatcctgcaattgttaatgcaacc gacgaatgatttgtcttcagagagagcattcaataacctgcgccgccaaaaatggctattgcaccgcggtgtggcgatggcaccag aaaatattctggatattagcgcggcagactatccggagatcgcgaagctgacagaagcgacgccgctcatcgctctgcttgaggat attgctctcccagatgaggctaactgtgcgctgagttcattggtcgtgcgaggcaaggctgcgttttacaaggcgctcactgtagc aggtacacttccactttatgcaatcggtagcagcttacgtctcactgatacgattattcttcaggccagtgacaggtcgtacgcgt ttgagagctttgacggttggttgctcttaattgagtgtctcaaaggtgctgagtcgcttgagggtaatgaggctatcaatgcgctg agtttttcgcatccggttacagacaagatagttgctagctaccggcatctcgttgacagcatgaatccaacccaaagtggtgaatt gcgtaaagcactgttaagcacgctgtgtcatacccattcagatcccgccagcgtactgcgttcaatcccgctcagaacggctgctg atacctgggcgttagccaccaatctctgttatggcgtaacgggagcagaacgtagtgctgtcctacatgacgacgactgggcgtat ttgtccccttggctgcaggctaatgacttgtcggtagacagtactgagtccgaagggcatctcagtcatgttgagcattctgccaa tgtcttaagggaatactttgcgccctgggaacgctgggttccacgtaaggcaattgctgcactgctggctttgctggcggggaatc gtaaggttcataagctatgtgagagctacctggggttgcaaagttatgccctgttcgtgaatgaactgtcgcaagacagcaaaccc ttaactaaccatgacgctcactttgcagagttaacgctcttacagtgcattgagaaatatgcctttgccgtgaaggtttacgaaga aaacacgttgcaggttcattctctgttccaggaacgtttgaccgtggcgctggcaactgacctggatacgatctttgtgggtcagc acggctacgcttttttataccggtcaggcaccgcaaatcttcattcgccgattttccccagaccagtatacgcctcagcaactttg gcgattctgaaacgcagcaccagctggctgcaggaaggtatttatctgcagaaggcaaggctagacacgctctggcaatcctttga gcaggccgagcagttggatgtgaatatcgcgcgcgtcactatcctgaacagcattgttgagcgcctgaaaacactgggccttaaaa actctcagcttaacgttttaatgagagcctatgagagtgagcttcactctcttgctgaaagtagtgacggcaagttgctccacagc tcgaggctcactgaaattgtctatgacattgcaaatgctatccaggatcgccctgaactgcaggctgaaatattaacggcggtcag aaagcgtatagaggatgctcagtatcagccatcaagcgttccttttgagctgttccagaatgccgatgatgcagtagaagagttgt tcaagctggatagcgatgcccgtcatgagcgggtacaccagaaatttatggtgaaagagcaaaacggcggattgtcattcttcaac tgggggagagaaattaaccgctttcagagcgtgaaaaatgagcaagtcgagaatgtacatgatggctacaaaaacgatctgaaaaa aatgctggcgctttaccagtcggataaagagcagggcgttaccggcaagttcggtctcggcttcaaaagctgtctgctggtgtctg atcatccttacctattgtcggggcggctggcgactaaaatagcgggtggaattgtgcccgaatcctgtgatgctgaaagttataaa caactaaaccaactcactgaaagtgccgcgacaaatggcctgtcacctactcttgtgtatttgccactgcgccagcatatgcaagc

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		ggaagtggtgttaaaagatttactctgtatgcaggtttgctaagtctttatgcacgtaacttgtgccagattgtcattgatgagcatgaatggcgctgggagcctgttcagtatgcacgtattcctggtctgtcattgggcaaggttatgctgcctaacggcaagggtgctcagtcgccagtgcgggtggtggtttaccagactgaaatcgatgatgagcgctgccatctggttttccaggtcacgcgtaggggcctgagaagttttgatactcatattccgcgattgtggaacttgtcgccattgatgagtgatacccggcagggctttttgattaacgctggatttgaggttgatattggtcgacgccagttggctattgaagctgaccgtaatcggggcattatccagaaagcgggagcaaaagttcattcgctgctggaattactttggtgggaaacggagcataactgggaggagctggttgttgagtgggaactgagccctgaattgacccatactcagttctgggaaagcttctgggacgtgatgtctacaggcattagtaacgatattaacgcgatggaaaacgaaaaattgctacagcagctttacgaaagcgaaaatggcatcatgagcttctatcgctcatatcccgcgctgcctaacggatttaaagagcaggctgccggactgataacgtggagcgacagagtgcgtagcgcggatgaactggtttctcgtctggcgagttcactgattcatctccctgcgtttcaggcattgcacagtgcacagtgcctggtggcagacacgacgggaagcaaacttaaagtcgaaagtaaactgtcgcttgaatcattaataagctcgtcgttgccggataaacagggtgttgatatccagcatctgtcaccgcgggatgctgaaaagctggcagtcgtatttaacgaagagttcgacaagcgactgggtgaactgacaggctggcaggacaaaattgaggctttcagaaaacagctgataaacctgcatgtgcaaacacaagcaggctctacacgcccgattagccaaatttgctcggtaacactccttgtgccgaaaaaaatgaacggatgatctctgggtttgcacctaccgatgccatcatttcatcatcatattctaagcaggcctgtgaatttattgtttattgcaaacgcagaagtcagggatatgtttttgaggatttagtcaaatgggcaaagcgcaaaggcctggcggctgataatcaaaagcggcaggcattttgtcgttttctgattgaaggactggaaggggagaaactggcgggtatgctgatggaagagataccaccggactggttgcttgaacttaagctgcgcccaggcgccttcccggcagactggcactggagcaataatgatattgcctctctcctgcaggggcggttactgactaacattgacagaacaaaggcatgggagcgcgagattcgggagacaccggaagaatacgaaccgttggtgacaccaggtgaagccgtacaaaaaatacacacctggtgggagaggaaccagcaggaagagttggtgaaatacaatgctcggctctaccctgaaggctggtttgactgggaagctttaagaaatgcctctgacgatcagcgttcacgcctggcgttattgaaactcctgtatctaggctcatgccagaccattgggcggactcaggaagagcaacacagtgccgcaattgagtattttgaggacaaaggctggtgggaaacctttatcaaccctgatgcagcgcagcaatggctggatgtgatggacaattatctggaggattctttgtacggagatacctaccgtatctggctgcaaatattgcctctgtatcgtttttcaaagcatttagattcctatcgcaaactactggatatgtcggaagcgttccttgaggatattggggatttgctgcgaccggcatccagtttcaatctttcgggaacgggcgtgggaactgtagtcccggagttacgtgcaactctgggtactggggtgaacttcatcttccgtgaattggtgcgtaataacgtatttatcgattccagcattcatcgatattgtttctctgcgccggaacgcgtcaggcgtctgttactggcgatggagttcgacgaaatggatgttaagcaatccactgccagtgactcgcttctgctgtggacgtttttccgcgaacatctcggtgaggaagatgcgacctttaatcattgtttcgacataccgctgcgcattttaaccagcgaaggaaacgctcacttcgtattgagatatttggacaggatcccctggattacgtatgaaaatgatctttcagcagggccagcaggtacgacatgaacgctttgggctggggacgattgaactcttgcgggaaaacactgcactcattcgtttcgagtcgagttttgaagaacgtccactttccgaactggagccggtgcgcagtgctcaggatgctttggcagaaggaaattatgacgatctgcgtgaagttctggcgcgcagtcaggcgcttgcgatccgctccatcaatgatagttggggggtgttctctacttcacgtatcaacctgctgccgcatcagttatgggtatgtcaccgcgtgttacggcaatggccggtacaaaagctgattgctgatgacgtagggttggggaaaaccgttgaggcggggctaatcctttggccgctgctggctaaaaagcgtgtgcagcgtctgttggttttagcgcctgcatcgttagtaccgcagtggcaggagcgtttgcggcagatgtttgatattcgtttgtccctctactccgcggaaattgatactgagcgatcagattactggaatacgcatccctgggtggtcgcttcattgccgacactgcgaaaagatattaatggcaggcacgagcgaatgctcaaagcagacgactgggacttgctgatcatcgatgaagcacatcaccttaactcgctagaagattcgggggcgactcagggctatcgatttgtgcagaagcttatcgatcacggaaagttcgcctcacggctttttttcacagctacccccatcgcgggaaaaattacggcttctttgctctgttgaggcttttacgtccagacttatttgacgtgaataagccatttgaaactcagcagcatcatgttcgggatgttgtgattcgcaataataagcaaaccgtcacgaatatggacggtgagcgtttgttcaagaccgtcaacgtgacctcacagacctatcatttttctgaggctgaacagtcattctatgaccggctcacacgatttattctttcagggcaggcctacgcttcgtcgctaagctctgcaaaccagcaggccgtgcaactggtgttaacggcaatacagaaactggcggcaagttcggtagcggcaatttatgccgcaataaatgggcgtatcgccaggctcggggaaaatcagaaaaagctgcaggcgctgaatgatgaaatgaatgccatcatgagtgattctcaggccccggatctcgatgatgcctacattgcgcttgaaagcgaatatgttgaaatgtctgcttcggttcaacttatgcaaaatgagctgcccatgcttgaagagctgcaggcgcttgcggggaatgtggaatcggaaacgaaaatccagaccttgcttcatgtgctggaaaacacgtttcttaatcgcaccgtcgtattctttactgaatataaagcgacacaggccctgctaattaatactctgaatgctcgctttggctatggttgcgtcagctttatcaatggcgaaggacgcctggaagggatttacaataaacagggcgtcaaaacgtcatggagtatggatcgctaccatgctgcggagcaatttaaaagcgggcaggtacgctttattgtttgtactgaagccggtggtgaaggtattgatttgcaggacaactgttattccatgattcatgttgatctgccgtggaatccgatgcgtcttcaccagcgtgtagggcgactcaaccgctatggtcaaaaaaatcaggttgaagttattactttacgcaaccccgatactgtagagtccagaatatgggacttgttaaacagcaaaataaccacagtcatgcgttctttgggcgacgcgatggaggaaccggaagatctgttgcagcttattcttgggatgagtgataaagtttttttcaattcacttttgctgatggcctgacacaaaagccagaaactctaaatacgtggttcgattctagagcagggaccttcggtggtcagtcagccgtcagcgtggttaaaggtcttgtaggccatgcggataagttcgagtatcagaacttagatgaggttccgaagcttgatcttatccatatgtatggtttcctcgagaacatgctgaaattgaatggacaccgtctggacaatgataagggtgttcttagctttgtcactcccaaagactggatcacacagtttggtatcaagaagaaatataacaatatgacttttgaacgtgttcctacagagaaatcgttagaagtgcttgggatagggcatgtgattattaataatgctattaatcaggctgagaaatttaacgcctctacggcagtagcaaggggtatttcctcagctttactgatttacacattgagagaccagattactggcgatagtaatgtacaatcattttcagttgttggagtggtactggaagataatattcaaattttggtcaacgctgagttagtcaataaactggcttttatatatgacaacctacctaaaggttcgacggtgattaagcttgacagtgcattccatgttaattttgagagggatataaagcgtgctgaggccgcattagatctctttattcctgggttgaatttaccctatgagcaagtagtatggcaacatacagcaacttttttgccacagtaaatatagcagtgttcaggatagcattgggaatgagaaaaactatatgaaaatatggtgctgataaagtattagtactatggtcgaacggctatgcgcttatgtcatggagctgattccagagagccttgaaaacgaaagatttaattttccccccagcgtcatccgctctggcaggtgagtcgcccgagtccgagtgcccagcattttcaaatcaccat (SEQ ID NO: 38)
42	pLG044	tgagaacttacacaattaacgccaattttcttattccatcacgcatacgataaccgtgatcaactttttctttttgcagcaccctataatgcaaccagtttaatttctttggatgcgtaatagtcagtgtgctgctcttgataaacagtagtcaataggcatagtccatatccgaaatctaacttttattaacgtacaaatagcaaaagaataaataacttagagcataggtcctcgaaaaatttttctaatgttcgatagtcttgcttttggcgtaatgtggtaagtccaataggtgataatgtgtatagttgcattgacctagtcttgtgagattgcatttaggatctccatcatcaattcatctttcgattcaatttcaaaaaaggtttctaaaatggcgggtgcttcaatagacgctattggtgtgattaaccaaatcaaagacaacttaacagaccgatacgaggatggctttcctgtccttaaagagatcattcaaaatgctgacgatgcgggtgcgaacgaattaactattggttggagtaaaggtttctgcaatgcagaaaatgaactactcaatgcgccagcgctgtttttatcaatgatgcaccactggcagaggaacaccgtgatgccattttatcgatagcgcagagctcgaaagctacatctaaggcatcagttggaaagtttggtttgggaatgaaaagtttgtttcatatgggtgaggcattcttctttatgtccgatcaatggcgaattgagcattgg

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		gcgtcagatgttttcaatccatgggataagtatcgtgatgcatggaatgaattcggtgaaaatgacaaatgccagatcgcaacaaagttaaaagggtttttaagtaccgataagccttggtttgttgtttgggtcccgttgcgtacaaaagcgctagctaaagcacacaataactacattatcatcaacaactttagtggtgatgaaaaactccctagtttctttaatcaggctcacttatcagagaaaacttctgagattttgcctcaactcaagaatctcaaagacatcggctttttctgcgagtctgacaagggtgtgtttgatgaagtgacctccatacagttacatgaagattcgtctcgaagctctttttgcggtgaaccgcgattaaataatggagactcttttgcagtcttctcagggaaaatctattcaaattcgaatgaagagcgttgtgcactggactatgcaggatgcgagcgagtcatctttgatgagcgtttaaatcaattaaaagacgaaaatatggggtggcctaagagttatcagttcgacaagaaagcgaacttgcctgttgaggctctcgacaaagctgaacagcatgcttctgtaacattttcgcgttttaaaacaaaggggcaagcgtacctcaaagccaactgggctgtttccttcccttaagccaaaccaaggaacttgttgctgtgcctatcgagggggagtacgactacaatctctatttacacggctacttctttgttgatgctgggcgtaaggggttgcatggccacgacaatcttgggtttctacctccctagagcatgtaaaaaatgatgagaaaaagctgcgtgaggtttggaacatcattctagccagtgaggggacattcaacctcgttttaccggctctaaatgagttttgtcagaagttaaggctgccacatcaaataaaaactgttttgaccaaggctttgtacgatctcctcatagaaagatatagaaaagaagtatccaagagcgccaattggataatcaatatcgatgacaaggggggctgcttggtctttacttgataagaatgcccaatgcttaccgatccctcgtccagagaatagtgattactctcgaatttggtcaacgttgcctggtttgagtaagttactggataaaaagtcactgtatgaagccacgggtaatgaatttttaaccgagcagaatcaacgtgatagttggaatattacgctcctggaagaagcgttaggaagtggtgttgtcaacgcattttacagatcaatcaatattgaatatctgcttcagttccttcaactagctaaggagcagtgcacgacggaagatttgataacctgattattccacagttccgagaggtattgtctactcataagcttgctgaactttcattgaacaaggctcttaacacgcaagtttttgagcttgttagcgcacctaaaaccgtcgtactaccaattgataaagatgatcaatctatttgggaacttgtctgcaagatcattcctgcaaagctactgctccctaaatttctgtctactcacaataagccaattcatgacaatgtcactgaagaagagctcttcgcacttttaaccctagtagatagctacatcaaaaaacagggtgaacgtttatcctctgatgaatcgtctgcctgtgagcgtctcattacatttgttattgattgtgtaaatgcaagtgaggtaatccaaaaaagcgatttttatcagaagagtgggcatttaaagcttctaaaagtggaagctcttggttcgcaacagagcacaaaatatcgctccttaaacgaactcatagtgttaaaagaaaaataccagctgtttcttcgtggagggagcggaactttggtaaagggttggggaaagagctagttgcagtcgtgcctggcttggagctttgttttataagcaaggattttgaaattggtggcctatatgaagggcttaccgcttgttctgaagccgcgtgcctacgactgctttccacgtacccaaatcttggttcaaattcggcaagactagcgctcactaaagtattctctgccgagctctctacagatgaggagaaaagaggtttccggtatttgattcacggcagcaaagaagacgacttgagacaaacgctttggaagccaaacagggcaactaacccagtatggatgaaaatttggcgtatgtgtcagccagaagatttccctggatggtgtgagttagatgaagagttttctaatgctttgacaaaccagtacgaacattttattggcgttaaagagcagttctataaagacattatctctgaatacagaacaatactgcctgaatgcaattttgataactttgatgactgggaagtggagcaactgctcgcagatattggtagtcaaggagatgaaaggctatggaaagcgttgcctgtccataggacagctcataacactagagtcgcgattacgaccaaatgcctgatggaaggaagtgcaacagttccaagtgaatgggatgttcaccttattcaacattcagccattgctgaagtcgccgcttgccagcataaatgggtgaatcatggtctacctaaagagctgatcgagattgcgcttacccaatcaagtccagctcagtattccgcatttattttggaccagctctgcgctattcgtattgcgaatgaaggaattgagcatgagttggaaggcaagataaataataccaagtggctgcgattagcgtcaggaaccgaggtttcaccggaagctattttatctttctctgccaatgagctgcctgagtctgcaaagttctgcgagttaaaagagtcaaacatttacatgttctctcaactcgatggaaacatgtttgagcacgatcaagcacgtggtttcttgagagagtgggtcgcaaaaagtaacagctcagtttgctcgtgcattttggcagaagccgcgcaacatcaaagttatgtagttggtaatttttccaacatttctgctcaggtgctagaacagatttcatgcatcccgccattgatgcagctatctgcaggctggggcttactggttgagctctaccaaagccaatatctttcagtgaatgaaaacaagcaagtgatgctatgtaaggaaacagaaccacaatcattatggtgggcgctggagcgtattgctgatgatgatattcacggtcagtcaaaggaacttcggaaagcatttttagaagcgttgtgtaacaccgagggaggcgttgattatcttcctaaactgagatttcgcaatgagaacggaagttatgtatcgggcaacacactggtatcgaatgttgctcaggtagttgctgataacttaatttcgccacaagaatacgcagtcattgagagttattgcagtaaatctgctctcacgaatggtaatacgtcaaaaatcattgagttagcgggcgataatgcgccagtacttagtgattacttcgatgactgggaagggatggttccccctgatgccatagcgacatttatagcactgtttgctaaatctggtggcgtcgagaaattggttaacaattatctaagacagtcaacgctggagtcgataaagcaggggtatgaggaaaagtggaactccggaaagggacgtagaggcgaatttcacactatccgtatagctcgttatataaaagtgttgattttgaactggcaatttgtgcagaaaatgcggcgtacatgacgtcgattttcggcgaaagaattcaagttaaattacaaaaaacaccagattcattgcttgttcaccaagcgaacaagtccaagacgaaaaggatagagcttcgccgagttgatacaaagaatgtatcaaaagaccaacttctccgcatgcttgccaaagctgtagaaacgatttttactgatgtgtttggtgcagagtgtattcgatttgaaagtgaattttgaagaggtttggtgcttcagaacaggtagatattcagattacccgacagatagtcttggagaatgttgtcccccctacttgaaaggcttcaagtgcgagaagaaggactttgtgatttacgttcagattacaaacgtgaacagcgtgttttggcgagcagtgatccttctgtactacaagatcgctcacgccttaacagcgtccttacgaagattaaagagactcttgaaaataacgaaaaagtgcaatctttggtactcgaatctgtacgaaaagagatgagtaaacatttccaatactcgcctttcagcgtgccatttgagctgtttcaaaatgccgatgatgctttgtgtgaacttattgaaatgcagggcgactcaaccaatgtactgactcgatttgatgtggtttctggcagtgatgggactcttaacttctaccattgggggagagaggttaactactgtaaaagttcatatgtcgcaggcaaaaaccaatttgaccgcgacttagaaaagatggtgagtctcaacgtttcggataagtcagatggaaaaacaggcaagtttggactgggctttaaaagttcattgcttcttaccgacattccacgtttggtgagtggtgatatttgtgcagaaattcatgctggcgtattaccgagtgttcctagcaaaccagtgatgacggaacttaatcaaaatgtcgatgagtataaaattggaaatcgtaaaccgacattaatccagttgcctaaatgtgataagaagcgggcagatttgaagttggttttgggacgtttcaaaagtaacgctggcattctcacggtttttcacgacaaattcgagaaatcaatattgatgagcagcgatttgggtggtcgggacaggctctccataatatccctgaagtacttgtcggtgaagtgaaactgccaacaaatacttctgaagagtctaacgttatccttcgaagtaatagagtgcttattatcaataccgagtccggtcagttccttttgctttggattctaacggagttgtttctctttcgaatcgaaaaaacctaagtagcttttgggtgttaaacccgattgacgaagatctgaaattgggtttctgcatcaacgcgccatttgcggttgatattggtcgctctcagcttgctgtagataacggagacaatatcgatctttccagttcactcggcaaagcgttatcagctgtgttggtcaaaatgtttgcagcttcttcgaataattggaatgaatttgctgaagaggttggcctgggacaaagcagcacatttatcaagttttgggcgtcactttgggatgtaataacagcccattggccagcaaggcttggagagacgaactctaaagctgaactgattaaacaaatgttcacagtggaagatggtctgcttgcgttttaccagagatgtgcggctcttcctcgaaatcttggtgtaaaggaagattctcttgttcaacttaaaaacgttgatactggagcgaataaacctttgaccaaggcatttaataccttgggaaatcacccgatacttcaacggctatataaagaccaacaactcgtcgggcatgacacctttgagtttttgaagagtatcgattttagaccgaataatggtgcgttaactaagctcgaattgatcgatttgattggacaggactttcctcacaatgaagtaaaccacgacagagcaagtttctatggtcgcctatttggtaaaaactttgaaaagttaatgtcgaattttgaaatgacagtgactgagaaaaaggtgttggaagagcgtttttctgaattgaagtttctcaacaaaaccggtgtatacgtgactgcaagcaaactgattgttgaggggagccctgagagagacttgctatccaagtttgcaccagacagcgcgaagttaagtgaaaaatatgaccaagcatcaatggacttggttagcttcattcgtcgtgacgtaagctatgacattcattcatgggctaagcaaataagatctgaagaatctaacaggggaggaaagcaggaagggttgtgtagcttccttgttgaaggcggctatttagcatcatcgcttctcagaaaactacagacggatcaccccgcgtttcttacaaagggacgttttgatccgagcgtattaacagaaaaatgg

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		cgttggagttcttcaaaggcttcggctttcattagcatttggattgatacagaggaagataaagcaaggcacgtacgacaagcgcaaaaagagtttattccgaatgtgaccaatggtgagcagatcctcgaaaacatcacgaactggtggaatcaatgtcgtaatcaaagcttaattgattatgacaaacagctctatgctcaaccaatgccttggaaggcaatgacagaggacttcgagcttgaaacgttagaggttcgtaaaggttggttgaagttgttctatttagggagttgccaaacattaggtttcaataacgatgtagctaatcggaatgttgtttcttggttcgaggacaaggggtggtgggataaactagccgttgccaatggtcctagccctgaagtatggaaagaattaatggaagaatatcttcaaacagcacgcgttgatgagcgttatagagtttggattcaagttcttcctttgtatcgctttgctactaagctcaaggactatgtcgctctcttcatgaacgcttcctttattgataatcttgatgatttgttaaaaccaaatagttcaaacaagttatcaggctctggcatccaagtatctgagttaaaaggaacgctcggtattgggattaatttcattttacgagagttgcaaaggcaccaagttttggagcgtgagtattgtgaagatatccaaaagtacgcatttgttttgcctgctcgattacgaaagttactcaaaaaaatgggagcaggtttaagctttgacgcagagccagagaattcagagcgagcttacgactatttcgtttcggcattaaatagtgaaacccaccctcttcttaaggactttgacatcccatttagagtcttgttggctgataagcaagcgtttgaacgttgttttaattttgctctagatgagcagtttgaggaagtatatggataacattatacgcgttattcacccaaaattcggtgtcggtaccgtcgaattcgaaaaagctgagacatctcttgtccgatttgaacatggttttgaggagtgtttgaaaagtgagcttgaggcggtcgctgatcttaagtccgatcttgtttctggacagagtgtcgctgcctctgaacttgcgttaaaaacattagcgcactcactaaaaagtgttaatgaaaattggagtgtttttctaaatcgaacattaatttacttcctcatcagttatgggtatgccatcgagttctaaggcaatggccaacaaatcaactgattgctgatgatgttggtttaggtaaaacgatagaggcgggcttgattttatggccccttatcgagaggaaaagagtcaagcgtcttctgattttgacgccagcacctttggttgagcagtggcaccaaagaatgcttgatatgtttgatattcgtttgagtatgtatgcaccagaaaatgatacctcgcgcgtcaattactgggactcaaacaatatggttgtcgcttctctacctacgctaaggaacgacaagaatgggcgtttagagcggatgttaaatgctgagccgtgggatatgctcattgttgatgaggcgcaccatctaaattcaacggaagataagggtggaacgttaggctttcgctttatacagacgttgattgaaaatgataagtttgaatcgaagttatttttacagcgacgccgcatcgaggaaaagaacacggattcttctccttattgcagttgctgagaccggatttgttcaacgttaagcaaatggatgagcgagaaatgcgcccatttgtgaaagatgtgttgattcgaaacaataaacaatttgttacggatatgaatggtgagaggttatttaaacctctgtctgtgtcctcaagaacttacagttacagtgaacaagagcaacatttctatgacctcttaaccaagtttattgtatcgggtcaagcgtatgcatcctctttgaattcaagggatcaaagagcggttatgttggttcttaccgcaatgcagaagctcgcttctagttcaattgcagctatcgagagagctctaaaaggacggatagagaaacataaactaggtaagcaacgtcttcaggatattgaagttcaacaggctgctttattagaaaagcgtgaggagtcagaatcgcagtctgaaagcgagatatacagtgatgaattagcgcaattagaactggaatttattgaaacgacaacgcgggttcaattgatggatgatgagctccctagaattatggagttgttgtctgcttgtcagaaagttggctctgaaacaagaattttaacaatattagatatcctagaaacggagttcaaagatagaactgtcgtcttttttactgagtataaagctacgcaagcgctattaatgggtgctttgaataaaaagtatggtgaaggctgcgttacttttattaatggtgaaaatcgtcttctgaatgtagagaatggctcaggagtatgtgttgattatgtcaccgatagatacaatgccgcgaagcgttttaatgaaggcaaagtacgatttataatttctacagaggctggtggtgaagggattgatttacaacaaaattgtttttcaatgattcatgtcgacttgccttggaacccgatgcgacttcatcaacgtgtggggaggttgaatcgatatgggcaagtcaaaaacgtagaagtaatcactcttcgaaatcctgataccgtcgagtcaagaatctgggatttgctgaatacgaagatcgatttaatcatgcgttcggttggcggtgcgatggatgagccagaaaacctaatggagttgatattaggtatggcggatagcacattgtttaatgagttgtttacagaagcagccaatcgtaaaaactctgaatctctctctgcttggtttgaccataaaacaaaaacattcggtggcgagtctgtagtgcaaaaagtgaaagacttgattggtagagcagaaaaatttgactatcaagatcttgaggctgtaccgcgtttagatcttggagatttaaaaccgtttttactcagatgctttcatttaatcaaagacgttgtaagtatgatgaaaatggtggtttatcgtttttgacacctcacgcatggttggggcaatttggaaccagacgctcgtatgagaaattgcattttgaccgcaaagctaaacagcttgattcagaagctgacatcataggctttgggcatcccatgttttcaaaagcggttaatcaaggagagcaaatccctggaagttacgcgtttcttaacggtatagagaaagatcttgtagtgtttaaggttcaagatcaggttacgggaaccgatgcatcagtaaaagtgagtattgttggactggtgctcgatgataatggcgattgtgaattggtcaaggacgaagaccttatcgggtatttaaacgagtatcttaaaatttccaatgatgttgactctaaacgtacaccagaggatttagtgtctgttattcaaactgctaatgattatctaatggagaatgtgtcatcaattggcttaccatttaggctgcctaattctgaaccattaacggtattctacaaagcaagtaactaactattattctatagctgagcattacgaaaaagttcggtagtgattctggcttaatatttgggccgaagctaagaggtcgtt (SEQ ID NO: 39)
43	pLG045	gtcatagtcccttacggagataattcattgaaattaatatcttatacagcacatgtaaatagccgtggtgtatttttatccaatgaatcgttacaaaaataagatgcatgcccaccctgttctgtgtgaacgctacgaccagctacggatttataccaaaagtaggaattctatatgtcacgtattaccatcaacgttttatggttaaccgtaccaatagcgcggaagtgggcatgagcgaagtagcagatcaacagcaattggaaactcagccagcgggtgatgacctcctgcaaggtgtcaaacgcgttctcaggcatgccgttcaggcgtacggggatgggttaaaggtttatcaaagcctgcaaaatctcaacgaggtgattggcacggagtacggtaatcgggtcatttatgagttgattcaaaatgcgcatgatgcgcatacgtccgaagaacgtgggcggatagctgtcagcctggtgcttgaaaacctttcacggggaacgctctacatcgctaatgatgggcgagggtttcgccatcaggatgttgaagcggtcaaaaacctggcgatcagctccaaagagattggcgaaggtattggcaataaggggcttggatttcgcagtatcgaggcgctgacgcaatccgtgaggatctattctcgctcaaatacgaacggcaaggaccgatttgagggttactgtttccgtttcgcagatactgacgaaatcgcgcataatattcgcgatctcggtgttgatgacgcgatcagcaacgaagttgccaaaacgcttccccgctatcttgtgcctgttcctctagatgatcaaccggaggatgtccgcacttttgcccgcaacggtttctccaccgttatcgtggcaccgttagaaactgaagcggcagttacgcttgccagaacgcaggtgaaggagctgaccaatcgcgatgttccactgatgcttttcctcgatcgtattaccgaaatcagtatcgaaattttatccccggatgagaaagccgaaaagcgcaccatgcaacggcaggaaaaggcgctgggaagtattcctgacgcgcctgatgtcagtctctacgaagtcgatataggtcagcggaaacgcttttagtggccagaagcaatgtcgataaagcgcgcgtgcagcaagcggtgagcgatagcttattgactgcacctcagctaaagcgttggctgaactggcaagggataccggttgtttctgtcgccgttggcctgaacaatcaacagtaacttctggaagactctacaactttttgccaatgggcactgaggccgcttcaccgatttgcggctatatcgatgcaccatttttaccgatattgacaggcgtaacacgaacatgagtttgcagctgaaccggctgttaatggaagtggctgcggaaacctgtgccgctgctgctttgtccgtcgtatcccgtgagctggatataggtgcatctgcggtttttgatctgtttgcctggacggggaacatcgtcgcatgatgcaaacagcactggaacggaaagatacttcgctcagcaaagcccgcctgattccggtgatggctccgccaggaaaacagcaatggtcgagtcttgaagaagtcagtatctggccggaggtgaaatttgccatcctgaagccgaaagacgttgccagatacagtggcgcgcagttggtttctagcgaattgaatacgccgcgcatagtgcgtttgagggagataacaaaatttccctatatgtatcagtcattagatccttcggcgcagacactggtgaaatgggcagaagcctttgccctttcgctggtggaacggaaattctcccctgccagttggaccaaattctatgatgatttggtcaccttgtttgctgcggtaaaagtgaaactcaacacacttgagaactgcctgatcctgtatgaccgccagggcaaactccggcccgcaggcgggcataacagtaatgaacacaatggcgtttttgtacgtcggcatgtatccagaggcgacaaaaagaaagataagcgtaccgggattccgttgccgccagcgattgtttctcggcgctaccggtttctggatgaaaaaatcgtgcttagtgcggcgacgttcaatgcgtttaccgtcgccgacctgataagagagtacgatccgatcaaagccctgtcagggctgaatacggccctgagtaataaggcgacagtcagacagcgccaggatgcactattgtgggcatttgaggtctggcgcagcagtagtgtcgttgtcgatgtggagctgaaaa

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		aagccgatctccatattcccgtgcagtcgggttggtgtgcggcaagcaaggctatgttttcatcctcctggacgccaacagggaaggttgtggaaagctatttaaccggcgcgatggggatctcgcctgactgccgtctggcagcgggtttgttattgattgagctgcaagactggccgggcgtcgtgcaaaacagcaaaaccgactggattaaattcctccgcgtgcttggcgttgcagatggattacagccggttgaatctaaggtaagagcgcgagcatatggcgatagttggaatagctttttacgcaatggcgacgagcatgaggggtttgatagcgactggagggcagaagtaaagcgggcacatataagtttctaccatcctcagacggtctatacctcggaaggaaaaacatggcgattgcccgggcaacttgagcacgcaacattgccagacgatctgagggagctgttgtgtacgctgattttcgcctttctgaagtcgcagactacggagtttttacctttgaggtcggtcgttttgagcgacagaattcgcaaacagactcccgtacgctgccaacgccgcttggcacttttttacgcactaaagcttggcttgccagcactagctcactatctgaaggattgcattttagccgtccagatgcgtgctgggcttcgcgggagcggcgcaataaacctccgcgtttcctagaccatttgattgagcacaacgttgatattattgaagagagtcaactagcggagcgcttgttttctgcgaaaattggcctacgtgattggaatcataccgggacggcgttggatcgcattaaagaactggtctacattgttccgcagttgaacgctggcgataaggcggatttacagcgggaatatcaacgaagctggcgtgatatcctcgacagcgacgaagctcttcccgacggattggacctgattgtttttcgccgtgggcagcatgaagtgctgcgcggcaacagcgatctgcctcctgcggtgattgtcaccagtattgcacaaaaaattgaagcacaaatgcttgcttctgcaggctacgcaatactcggtattggcctggatgagaccgatacactcgtctcctgcctcggtgatacgggacgattttcaccccgtaagattaatgacggcggagtgcaactttacctcgatggtaagccgttttatcccgatgagagcgatccgttgcttatctccttcgacatgaactggttaccggaaatcctggttattggtctggcgttactcggggaaaacttagagcggggcgttcacgccaccaaggttgataagcagctgcgcgcaatcagggtacgccgttgtaagaccctctcttttgccgtgcagggcgatgatgccaccccaacggagtcgttcgtcagctattcctggccccatgaaacgatgccgacgctgattattgaagaggggctggtgtttaactggcagaccttagcgaagatttcccgcaacctctcacggctggtggataaccggttacgtttcattgaaaccttacttttgcgcctcgcagttggtcgcgataatggctcgttgagtaaaccggatgacgttaccctggcttgggagatgaattgcgatgttcaaacgatccgtgatcattacgcccgactgcgcacggacatcactcatgtgatagacatgctacttcctgtggtgacgtatctcaacggtattgagcttgctcaggttctcaagcgggaatatgccttatctaggtcagtatttgatgtgcgtagttggatttcatcacatctatctgatagtgatatacctgctgaaaagctgctggacgtgtgtgaaacagcaaccgatcgggttgaactccgtaaaatgctgtcgtttgattttcagcaatttaacctggctctggaagcgttaggggaaacaccgctgtccaatgaggatgctctgcgcagatttttacggcctttgtcggcagaggcgttcacatattatcgatcggttacgccgacactatctggcgaccttgataccggcggagatttgtcacaatacgttcagcataaatctttgggcttcatttccttcaactctgaatggattttgacacatgaaaccttggaaaaggagatggtggactcgcaggttgacacgcaacttttgagtgcgttaggaccggacaatggtgaagagctgtctgcacttaatacgttattagacgcgaatcgtaaaaatgtgcgcgaatttgccatgcaggctcagccgcgagtttccgcctggtgcagacaaaatgatgtcccggtgaatgctcactggcagtacaacgatcctcaggcgttttgccgacagctcgaaaataagggctttcttgatttccggctctttgagccggattcactaccggattactgcctgcgcgccgggctatggccaccaacgatgccgcccagcctagatcaggatgtgctgaatatcgacatgaggaaagtttcccaggaaaaagaacgcgctgagcaggcaaaacggcaacaggaacttgagcgtcgcagtatctttttttccgggcagtcgcttgatacagccagcccgctatttgccgatcaacttcgggaactggcgagtaccgatagtagttggcaggtgcgcagccagcacaagacgcaggccttgatggattttggcgtggtgacaatgcgtcaggcgagcggcggaggttgcggaaaaagaaccgggcgtgcgtatcgggagcctcgattgacacctgcacagcagcaagccatggggctggcgagcgagtggctggcttttcagtatctgcgcgatcgctttccggattatacggatgaaacttgctgggtatctggtaatcgggcttcgttttgcggg ggcgaggaaggagatgattcggccgggtatgatttcatagtgaagacgccgaaagtggaatggcttttcgaagtcaaatccaccctcgaagatggtcaggagtttgaactgactgccaatgaacttcgtgtggcaagtgcggcggctaaagacgcaagccgacgttaccgaatcctctacgtcccttatgtgctttcgccggatagatggtgcgttatcgaattaccaaacccgatgggcgataaaacacgcaatcacttcagcgttgtggggcatggatctttgcgtttgcgttttcagcggcaggagaactgacagcaaccctgctcagggaaacctgagcggggtttttaaatatggcctctatggataggggacactttctgcagtaaatggataataagaaagctaacgttgaagtctgattctgccattttccacgacagctaaatgctggatcttctttttaggatcccaacatacctagcagtaggacgtaagtatgcttgagttcatctcgatatccttgtttctgaatgacaggcattactatttcgtgggtgtgaaccgatgaaggggtgatgtcattggaaaataatgaggtagtagcaaggagaagttctgctcttatcatagtgaaaaagcggtttgggaacaaatcggaactgata (SEQ ID NO: 40)
44	pLG046	cactcaataccacacaattctcaactccgaaggacttcgtgaaacgtgagtaagcgtcaactcagctccgtctggtttacctcgtcaggctctgtagtttaggtgttgccatggcgtataaccctgccaacagaataacttaccttactccagtcaataccgccttcgctgtacgcttacgcttttcgctcaaactgtgtgaaaacgtttttgatcgcataaattaccaaaacagggctgaaaaccgcgctcatacgtaaaattcggctcaactaaccagtcgaccaatttcagattttgcgtagacgcgcgcacttcagttttagtcagggttttcacacagcctgcgctcatggctgctttaagctaaaacaaacagatagaaagaagttacgataccctgtgaattcttgcaggcagatatcaaggagggttcattggtagcgataaaaatgtatccggcaaaggatggggatgcttttcttattatttgcgatgaggaaaaaagtgcatttctgattgacggaggctacgcggaaacgttcaggcaacatatttgcctgacttacgtgagctgagttttaacggttaccggttacgtctggtcatggcaacacatattgattcagatcacattggtggtctcgtggacttctttcttgtaaatggacacgcagcagagcctgcagtgattactgttgaccgcgtatggcacaacagcctcagggcgatgacgagacccgaaaataatgcacaaaaagtggattcccgagaaatcactgactttttgagacggagatatcatgtcgaagccgataaagccaaaccgcatgaaatcagcgcgcgtcaggggagttcactggctgccagccttctggctggcgattatcattggaatgagggaaaagggtatcagtgtatctgcaccggtacctccattcccaacttgatgtgcgataacagtctaacaattctgagcccctctaaggagagaatttcagcgctctgcctgtggtggcgcagacaacttgcatcgctgggctttcgggacggtcctcctcgagtgaggcatttgatgatgctttcgaatttttttgtaaaagggaagcatctcaggttcctcttccgcatgtcatcaatgcaagaacaccgttgcttgagagggattatgcacgggatacctcgccaacaaatggcagttcgatagcgttcagtctggtgctcaataagaagagaatattgatgctaggagatgcctgggcggaagaagttgtgacatctctgggtgccagtggggcgtcccatcattttgatatcattaaaatctcacatcacggtagtattagaaacacaagcccgaatcttttaaagatcatagatgctcctgtgtacctgatctcaaccgacggaaaaaagcatgccagacaccctaacctggcggttctgaaagcgattgtggacagacctgcggcgtttacgcgaacgctctattttaactatgccaacagcgcatctgcttttatgaaaaattacctttctgcaagtggtgcacaattcagaatcattgaaggatcaacggattggataacactgtgagatatgctgctactgaaactgaaataaggaacgcaactgtactcattgaatgcgcgggttacactggttccggaaccctgatcgcagcagacaaggtccttacggctgcacattgtgtagtatcggatgatcctgagacaccaattacagtgacatttttggtgcggatgaagacgtctgtgtcaatgcgacaatttcagaaatagatacatcgtgcgatgcctgtctgctaacactttctgactctgtcgacattccgcctattacacttatgacacagccggagcgagagggaagccaatggaaagcctttggctatccggcatcacgcaatgggccatcacattatcttcatggcactataagtcagattttaccaaggcttttccatggcgttgatatggatttgtcggtcagtgccgattgtgttctggaagagtacagtggagtttctggtgccgccattctatcagaaaataaatgcattgcgatggtgcgcatcaggatggatggtggactaggtgcagtaagtcttgataagttaagcggtttgctgattcgaaacggcctcatcccagatgacattgcatccctgccagattcatcactgtcgggtgaagttgtcctgaaccgcacagaatttcgcgacaactttgaatcgttcgtcctggagcacaagggacgtgcagtgcttttggaaggtagtcccggctctggtaagactaccttctgccgccattatcagccccgtagtgagcaactcgcagtggcgggtgtctatgaatttacaccggaagacggtgctggtacgacattcaaaattcttcctgaggtatttgccgattggctgcataaccaggtttctatactgctttcaggtaggcctgctcgcagg

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		gaggaaacagaaaagatcaatctgacccaaaaggtgtctgaccttctacatactttctcagattactggaagcacaaaggaaaatatggcgtcattttcattgatgctgtgaatgaggcaagcgagtgcggggatgaggcagtatcgcgctttacagcattactgccggtgacacttccggagaacgtcaaacttgttttcaccgcaccatcattatcatcagctggtaaggctttccggcactggctcacacctcaggattgtatcagcctaacgcttttaagccatagggaggtgttacagctaacagctcgagagcttaaaacttccgccccttctttgtcactactcacacgagttagtgatatagctcagggccatccactttatctccgatacattcttgggtatctgaaagcgaatccggatcaggttaatctggagatattcccggttttcagtggcagcattgaaacctactacgaaaggctctggcaggggctggttaaggatgagagcgctgtaaatctgctcggtattctctcgcggatgcgctggggcattgatatttcatcactgatccctgttctaacaccgcaggaacagacggtgtttgttccaacccttgaccgtattcagcatctgcttcttaatgataaatcatcagcattgtgccaccaatcatttgcggcgtttatcaacagtaaaacggcggtaattaactcgctgctgcacggacgccttgccgacttctgccttaccagtggagagagttatggcctgattaatcgcgcttatcacctgctcctagcctctcacgacagacatcctgaagccgcattggtgtgcacgcaggaatgggctgacgcctgtatcgtcaaggggctcagccggatattctaattcacgatatccgtcagaccctgaagaacacgcttattcgtgccgatgcagtggcatcgattcgtctgttgctgcttttccaacgcatgaccttcagacaccatttttgtttctgcagtcagcttatcactcaggccttgccctggctgcacttggcagaccggatgaggcccttgagcagctcataccatctggaagcctcgttgttgatgcagttgatgcaattgtcagcgcacagactctcgcgcgtatgggaaacagtgaacacgcgctgaagctattggaaaaggtgaagtcagctgtcgaccaagaatttgaacgcaatcccgtcaatctatctgattttatcggcctttccctggcttgggtgagagctgagctgatggctggggtggttgatggccacggacgcacacgcgaggttgttgagtatttgtacggttgtgggcaagtcgttcgcgataattttgaacaatcagcgcatagtaaatcagcatatacacgcgctttttatcctcttcaggcagaaatggaagccgtgaacatagcctttaatgaccgctccgtatctttacggacggttaaagaaaagtttggtagcttaccggaaaatattcttgatctgatgctcagttcagttatgcgggcacatgacatcattctgcaacatcagttgccgatgccccagcatgctttgcaacccgtttggtacaatctggacagattacttcatactgatattccgtattcgaacgaaattcgttttaattcattaagtagccttatttttttcaatgcgccttctgctcttattatcaggatggcgggggtattttctttcgaagtagtacccgaaataacgttgctcaatgaagaaaatgagatagcagcagacagcattgacgttagtgaacagggacaactctggctggtgagcgcctaccttaatgaaacgcaaccctgtcccgatattaaacatccgagtcagggatgttctgaatggctcaagacattgactgaggctatttttggtacagcgggcaggcgcgccgggcagttattgacggcaacgatgagaaaaagaactgcttttagtcaaggtgcagaatgatattctccctgctctttcgtactcgctggaagagcgcatggcatggccgaattcatgggcaatgcctgaacagattatccccatgatttacgaagagttagtaaacatgttcggcgcatgctggcccgataagatatcagtgatcactgatttcattctggctcatacgcctcagcaatgtggactttattccgaggggtacaggcgtttactgaacagagttattcagactcttctaaatgagcatcggtttttggggcaatctgatacgacatttcaactacttgagacgttgcatgcgtttgtttctgcttttactgagaatcggcaggagctggttcctgaattactgaatattattccagcttatattagccttgatgctcctcagctggcacaggacacttacactgagcttttaggtgtgtcgatgggccctgactggtacaaagaagaccaatttgccctcatgacaactatgctgcgcgtgataccacagcatacagacacaaatactacactttcacaagttgcaggattccttgaacatgcttcgggtgaaatgacatttaggcgttatgttaggcaggaaaaatcacagtttattggcgaacttattcgtcgtgggaattatgcacacgggtttaactattatcgtcagcagtcctgcggatcccatgaggaaatgctcacccaacttagccacccagctgcagatagccctcatccattgaaaggcatgcggttcccgggggagcgctggatgaggaacatgctgtagaatgcattgtcagtgaactgcgaaacagagtcgactggcggcttcgctggggacttcttgaaatattcagctttggcagtattggtaatcttgcagtgccctttgctgaacttatcaatgaattttctgcagacactgaagaccttaatgaaatacccaaaaggttgcacaacattttacatggtgatgtgcctttctcagaacacagaaattttatcaaaaatttcacagagcaccttgcagacaaccataagccactctttgctgaatttatcagtttgctatccgaagacactagcgataacgacgttaagcctccccctctggtgatgctaaccagaagggtactgatacctcagatgatgtggcaatgcagccaggactttttgggaagcgttctgcgatcaatagggctgaagcctgcatggaaaatgcccgaaaagccgcagcacgcagaaacacagttcgtgcaagtgagttagccgttgaaagcctgcatataattcaggatggtgactggtcagtctggagaaagaacaaccatctggcggaacttacacggacgtacatattggacaactctgcggatgcaggttcggtcattcgtgcttatgcttcgcttgtagaaaaagaacgttatgccccggcatgggtaattgctagtcatctcatcgaaatagcagccagtaaattctctgatcaagaagcccaagctattaaccagatcgtacttgaacacaaccgccacatgcttgggaataccgaagcggatgctgcgcatttttcttttcttaatgaacctgatacctcagatgcaggtgaagaaacactctattttctgttttggctgctggaacacccactgaaattcagacgcgaacgggctctggaagtactgaagtggcttgcatcagacgatgataagattctgggccaatgcgtgacggaggcactcgtttcagacattgcctcacgagctgaagcactaatggcattgacagactgggtgtcagctagatctcctcagcgaatatgggactttatagttaaagagcgcagcctttttgaatggcttgaaggcactactgcactaagccaagtccatctcctggagcgagtaaccagcagagcgggatttgttttaagaaatgagattgccgcatttgagcgaccccgaaagcttttactgacatcagaagcctctggacaacggaatattccagaaaatttaccaacatgggtgcaatccttgtcgcagacccttgccgtgatggaaaagcagggaatagatatcccagctttgcttaccttactcgaaaaacgggttttacagcagagtggattggctgatatcacggtggcttttgagctggaaaagttacttgcgcgtggttttactgtgaatagaacaccaagtcaccatcgctgggagacgatggtgcgatttgcattaaaccagatcatacatgaggcggccgcacaggatgaactgcaaaacattgaacccttgctacgtgcctggaaccccgcgtcagaggagtgtgttgagccgtgggaggtttgtaaccgggcaaaacagattatctgcgctgttatggaaggtagacatcagcaagcttcgggcatagaggatggcttttttcttgcattatcttgatgaagtggaggtttcccgagaaggtcaaacgcatctggtggaaatctcagcggtgttaacgacagctcataatggtcatgagagccttagaccaggtgcagaaagcgaatttaatgcaacacagacacctgatatagagcggacgcttagtgtgcaccttacatgccagcgagtcaaaatgcagcctttgctttttgggggagctacgcctgccgcagtgtcgaaaaagtttatgcagatgactggaacgttgccttcagactttattcgcaggcaatggcgaagcgggcgttctcttagtaaaaacagatgggggaaccaataagcagaggaagtctgttactcatgaaaagaacaactaccctccctccaggactgggcttagcgtggtatgtcactgtcgatgggaagttgatgaatatattttcatatgccccgaggaggagataatgaaatacagttcaatggaaacgccaaaaacgcgagaggaatttgaggctcgctgttttcacctgctcaatgcgatcaagttaggacggtatcatggcattccgggtgaaggtaacaaagagcaggttccttttctccctaacggacgagttgatctggcaaacattgataccatgactcgcctctcgatgaactcgttatatgatttccactataacagggataattatccgcagtttgatctctctgaaaatgacgagaatgaagaggctacggattgagctggccgatagaataatgtgcttggatcttagaggggcttccaaagaattagaacgctaaggttgccaaagttgtgtacgaaaaatgattgatttggttgaacgctaaaaagaaagtgagtagcggtttgaagccaggctttcgagcttatataaacattctgc (SEQ ID NO: 41)
45	pLG047	caggaagaagcattctattgacgctactatgttattagtgggcgtttgcgacagaatcaatggatagaattcacgggcgatgtagcattttagacatctaagaagcactttagtcgataatctttcacctgttcgtctgtcaacatagatgcttgtgcgtggagtagtacgcatacggccgagggctattgaccatagtgcattgtttgcttaacgttagtgcgtaggaaagaaataatctgggaaaagaattgaaaaagatagaaaatattgcaacgtcgtgttaaaggcccgttttactggtacagggaaacaggcgctaggtgctggatgataatgacaggaaatgacgatgctgaatataaggatgtatcctgcccggaatggtgatgcgtttttgctttgtgcagatagagccacattgcttattgatggcgggtatagttcaacgtttaacaactatattgtcgacgatctacggaaactggcttcagaggggcaagcccttgatctggtgattaatacgcatattgatgccgatcatattggcggcatccttcgctttctatctattaacggcgcagcggcacgtcctgaaattatccagattaaacgcatctggcataacagtttacgcagtctgacggccccgcagactgagccggttgagcttaataatgaaattat

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		tttaaacacccttactcaacgcggttatttgaccccaatgaagaggggcagggcgccaaggctatcagtgcccggcagggcaata
		cgctcgcctctctcattcatgacgggcaatatgactggaatgaaggcgacggattacgccgtatctcagttgagtctatgcctgga
		atcaacttgcctggcgggcgcgttactgtactgacaccatcgaatacggcgctggatgcactactggtgttttggcaaaagagcct
		gaggcgctttggatttaagggtgaggtgggggctgatacgctggctgaagatgcctttgaatgcggtgtgtcacacctgcaggagg
		ccgtcgggaaaccaccttcgctaatttcagcaggtcgtcccaggcagcttgaagaagtttaccgacctgacacctctgtgacgaat
		gccagttccattgcgacgcttgttgaacttgatggttgtcgcattttaatgctggccgattcccctgcagaagacatcgttcatca
		gttgaaaattttgcaagctgagggctgttccctgctatttgatgcaatcaagatctcccatcatggcagttgcagtaatacaaatc
		ctgaactgctggggcttgttgatgcaccggtgtattttatttcatccgacggcagtcgacaccagcatccagatgtggaggtgttg
		acggccatcgttgacaggcctgccgcttttcccgcaccctttactttaactaccgaaccccgtcttcagactacttacaacatta
		tacgacgattactggggcaccttttaccgtagaagcaggcacgtcctgctggattgagattggaaaacgccaatgatgctggatgc
		ggaagtcaggcttgccacctgtaggattgcttgcgggaaagatacaggaaccggctggttgatatcacaggataaagtgctgacgg
		cgcgacactgcgttgagaatgccctttttaatcaagcgcccgtgtctctgacatttaggcaggcagacacacaggtggaactgaag
		gccacagtcctggatgaagatgaaaacacggacgtctgtttgctgttgcttgatgcaccgcaggatctgacccctgtacgattgag
		tgaaactcgcccgttgccggggagctccttttatgcctatggatggcctcagagtaaactgggcatcgggcatcgcgtggagggaa
		cgatcgcgcagatcctcgccgagccgctgctcggaatggatatagaaatagccatagagcagaatgcggtacttccccgctatgaa
		gggctatctggtgcggcacttatcaccgggggaactgtacggggatttgcgggtttccattgagaatacggtgggcgtcatttc
		agttgcagagatggcagcgtttctgcggcgtaacaacctgcttccggcacccgttacaccgacggagagttatgagaacaccagtg
		aggcgcagcgggttgaattccggcacagttttgagcgcgttattaccttaaaacgcgggggatatctatttctggagggcgcgcac
		ggtataggcaaatcgacgttttgtgcaaagtttacgcctaaagacccgacgattgagcattttgggacctatagctttaacacagg
		ccgtgacggcgtgaatgcagttcagcaggctcaacctgagaccttcgttaactggttaagtatgcaggtttccctattcctgacgc
		gggaacccgggcggcttatcaaaggggactactccgtactcatcaatgaagccggacaactgctgacgcgcctaggtgaagagtat
		gcccgccgcaacaagacaggggtgctcttcatcgatggacttgatgaggttgataagtacgatgaggccctgcttaatcggtttac
		agccctgttacccctgcagctcagtgaaggcttggtagtgatcttttctgccccgggctatacccgttattcagcacaactgggtg
		tcagggtatcgcctgcggactgctgcacactgccagctctgactcaggcatcagcgcgggaatactgcagacagtcgctcaaagaa
		gtaccatcgcaggggatgatcagggttatctgcgatcttgcgcaggggcatcctctgtatcttcgctatctgatcgatctggccaa
		tgcgggaaaagcagaggaagagcttgctcagttaccgctcattgacggacgtatccgaaattattatgaaatgctgtgggttagcc
		tgcaaaacaacccgctagtggttaatcttctggcgattatcgtgcgtttacgctggggaatttcacatgcgcagctcaccgaactg
		ctcagtcttgaagagctgagcgtcctagtcagcacacttgaacgcatcagccaccttctgatgacccctggtgagacaaccattta
		tcacgcctcatttgctgattttctggcagaaaaaactgtcctacgtgaagcagatattcagcagcggctgtctgcctactgtgaaa
		gtcaccctgacactaggtatggccttctgaatcttatgtatcacagcctgcgctgcgacccgacccggcagatgtgggcaatcagc
		cgctgcgatcagcactgggctgaccgctgtgttaccgaggggttaatccggcgttacttcttggcgatgttcgggaaacgctgaa
		tgccgcattggcaagcggcagtctgacggataccgtacgccttcttctgttatcccaccggctgagctttcgctacaacaccctttt
		ttgcgcaatctgctttactcacagccagggcattgatccggattggccatcctcaggaagcgttgcaacacgttattcgtttcggg
		cggctcagtctaccagtgacgcaagccctgcaggtggcgtttgacctgattcgtgcggataacgacagcgatgctcttgcgcttct
		cagtctggcagatgactgggtggaggagcagctggcagaggtaaaaaccggtctttcttatccggaattttacagctttatgata
		tgcgtatgaatatctactttctcaaagggctggccggagacaggcgtgcggaaggagatttaaagcaatttcagctttactggatg
		aacgtgattgagcaagtctgtgacgatgaggggacggtcaggggggcttcgcggtcagatgtgtgcctcgttctttgcaggcatgct
		gtttttccatggacgttatatttcgcttgcgaaactgagtgagaatttcacggggcccctgcaggaggtcacgcaatcgttcgtga
		taacgttcatgtattaccattttctctgtgaggagtttcaggtcagtattgatccggagctgctggaccagctctttaaagacctg
		acaacgctgagctgtctggaacatgaatctcctgtgtacgtagatccccggacacttgatgctatgatctcgtctggtgcccctgc
		gcaaatgataagaaattttcaggggatacatcagtaccactgcaaccggtacgtttcattggtgatgataatgtgtcagcgaatg
		atgtgtcgttcctggaggagatggctaaacataaaattcaggcatttgcgatccatcgtatgactgtccggcgcccgttgcgctg
		acagcaactggctggatcgtaggcatggaggaattgtgtaggatggtggcatggtgtgagggggcggcaggacgttttcatttaga
		gggagatgaagcagcccttgagtcggtgtggactgtcattgaaaagcaggtactgagcagcctgacatttccattatcagaccgtg
		tggcatggcatgatgcctatgctcttcctgaagctattgtaccacagctttatgaacggctggcactcctgatatcgtctgtttc
		ccttcccgactggacgcgcttttggcctttattgagcagcatttcccccgtcaatttgggctgtattcggaagggttccgagccac
		gttactaaagattctaacactcctgagccaggtggtggatgacggtggaattcagaaccgcctttatgatctggccttccgttggt
		atgagtttgtgctgggcaatctgcagaatcgccatgaacttgtgccagagttgttgcacctggtttcattatttgtccggctggat
		gcgggtgaaagtgcacggcaggcttaccagcaggtgctggcattctcaatgggccccgactggtataaagaggatcagtttggtct
		gatgataacagcgctcaagtcaatgagcgaggcggacgcgatccctcagcgtttgctcgcccgtattgcctgtctgctggatatgg
		ctggcggtgagatgacctttcagcgttacgtgcgatatgcgcgccgtgatttcactgcggcgttgtgccagcacggtaatttctcc
		caggcagccgcgtattttatgtgtcaaacatacggtacaacagctcagctttatgctgaagctacgcatggcgacatcgatcgtgt
		gtcattactgaaaggaacgcgtttccccgggggcgcactagatgaacaggatgtgatcctgaacattgtgcgtttcgctgtcccga
		tgtgtgactgggcgttatgctgggcattgcttgagacctaccattttggcgatgcgcgtcatcttgataattatgcagatgcctat
		gctcaaatgatgatcaacatgcaggactgtcaggatgcaatggcgatgatcgcacaacggctcacgcttattttgaagctgaact
		gatgcctgggaaccggcacctgtttatgaaatacctgcgaagcgcacttcctgaggctctcagggataaaactgattttctgaacg
		tttacctttcagataacaacagcgccccagcacagcagagcgagccatttgaagacgtcgcagaaacgcagcatgcaccgcctaat
		gtttttgcaagggcatcgcttgcgcttgatgaggctgaaagtcaattgcacagacgtaacacgtcacaggcgcagcacaaggcaat
		caatgcacttgagatgtttcagcaggagggatggtcggtatggagcgacttatcagaggagcatagacgtgcaggctccatactgc
		tgaaaagcacggattcggtgtcggaggttgtgacgctgagtagggcgttaatttctgcagagcagcatacggagagctggcgtatc
		gctgacaagctgattgaatggttgtctcctgcagcggatgagagtgtacaggctgagctggctgagcattcgctatcacacatgga
		gatactaaccggcatgcctgttgccgtcatcgaacggtatgatttcttaacaggaaagaggatcagcatccgtcttctgcgctta
		cccgtctgcttctgcatgctgttgatcatcctgtctggatgcgcagtgagaaagctgcggatatgttgctgtggctgctgcagcat
		catccccattacgtatccgacgttgggcctctggcattttcaatggtttcactgaaccatccggatgtgctgtgcgggatactcga
		taagctttctcaggatgatcctgggtctttatggactttgctgtcagcacatctggatgtggcagagacaaaaaaatcctgctgtc
		atgctggccggctcgccacattagggcgaattgcgagacgggctgcatccttggggaacgcgagtgctgctgaggcgctagcgtta
		ttgcatgacggggaagtacgccagcccttgcaggaaaaaatcgcacagcagagtccagcgtgtccaaaatgggctgagataattgc
		ttttcagtggcgacagttagcggatgccgggctggttgacggcagcctgtcagagagggcatttgctgtgctgtgtgaggcgtgtc
		atcccttcgggtgggaaacagtagaggctcttgaagaacttttggcgacgggcatgagcggaagcacggcctggaacggccgatgg
		gaggcaaaacttcgctttgccttacaggtagcacttatgtccgttctggacgatgcacagtgccttcaggctgaggctattttccg
		tatctgtaatcctgagccgactgacacattcagaattacgcatttttcatcgcctggtaagcaatggctcaaccagttgatgcagg
		ggaaggttaaattttcacctattgctgacagccagctctatctcgatttttacgagaggcggaatattaacggcgtactcgttctg

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		ttgaggctgacggcttatttctaccgtgacggggtagatgctccctgcttatccggacgttttcctgcaaccgctcttgccacatctgtgcgggcaggccaactggacacatgcgtgaatgttcaagcgacgcctgcatattttggcagtttcacgccagcaattccttctcaagggctaataacgctcactagggctctttcgcatcattttaaacgagctagttggcgaaaggggcgggatgttgagagtcaggggggcgcgcctctggaagaagggtgttatttatccattaaacggacgcgttcagactcccgccgggaataagggttgtatgggtttgtgaattcaacaacgaaccgattgcgcttatgaacgccgctggcgcactgaagattcactaggaagaatatgaatataccgttaacgcgaagtgaattcgagcaccgacttcatctgcttgagaatcattcaaaaacgggtcggctcatgctggcagaggggtatccggtgagagtttgcttaaagtcaggcgactgccaaacggccggattgattttctctccgtggatgaaactgcccgtcttcaggcgaatatgatggagtggatgaagtcgattcccctgccgaacataccgaacgatgagggcactccctaaacttaagtatcgagttaatcctagtagaaggggatgtgaaaagatacctttgaaaggtgcgaggtcaatggaacaactttcagagatttatctcttatctgaatgttcatcacggagctgcgttgtagtggccccgaaaaaactcactatagagaacggtctaggagaagactgtaaaagcatttgcttgcgttaattcg (SEQ ID NO: 42)
46	pLG048	gaaatttcgcgacagagatccttaacggtgcgtcgagcttcgacggaattcagaataatgatggtctggtgttcggtgaatcgtgctttgcgcatggcgatctcctatcagaacaaaaccagtatgccggatgatctctaaaagtgaatggaccgatatgcagggatgcttacagtgggtcttcgacctttataagcatagtaaagaatagaatatgccaatgtacgataatctgtgcactctattacctgcgcaaaaaagtacaccagaattgtttgtctggtttggcaaattgagatcattaggcggcatagcgaatgactttaaatgaaaagcccgattcatcaataaagattgttaaaacaaaaaccttgcccccagcagagggcgagcgccgggcaatgcgtggctatatgggccaatatgaaagagccggtgcagccatttatgctgaattagagcgtgggcaattggagtggataggcgtagcggaccgcagtgcgggtatcgttgatgatttagtacttggatttaatggccttatcgttgggcaccagttcaaaacgtcccgtttccctggtacatttacagtacagacactcttagtagggtctgatggtctgcttaagccattagtttgcgcctggcaaaatctttgtagtgctaacccaacgtctcaggtagaaattcgtttagttgtcaacgattatccatcagttaacgacgctcccggaatggaagctccagctcatagcgctgccttccttgatgagtttgaacattatcccaaacgcacgcttgaggaatggcgctacagtaactggggccgtttagtcgaaatattatttcaacattcctgcctaggtgacgatgatttcgagagatttttcatgcgttgcgcataattcatggttctgcagcagattttatacaattccataaactcagtgcagaacaagcgagactggcgtctgatatagcaaaaatattacctcgactggtctccgataaacgagatagggatcgatggtcctgtgaagaactattatatgaactagggtggaaagatcccaccaaaacacgccacttacatcgttttcccatcggtgctcacgtccaacgcaaccgcgatacggaactacaacttctccagacgatacgcaacacaatccagggctatgtggcattgattgggcctccaggttcggggaaatcgaccttgctacagacaaccctagctaccgagtataacactcgggtcgtgcgctatctggctttcataccgggcgctgcgcaaggtgtagggcgcggggaagctgatgatttcttcgaagacatttctgcccagttacgcagcagcgggctgcctggacttcgccttcgagacagcagccaatttgaaaggcgcgaacaattcggtgaactgctcaaacaagctggcgagcgttatcaacgtgatacagtaagaaccatcattattgttgatgggctggatcatatcccccgcgaagaactaccagcccattcgctgttaggggaattgccgctgcctgcagccatccctttgggcgtgacatttatacttggcacccagcgactggaactcaggcatctcaaacccgcagtacaggaacaggctgggcatccggatcgtctcgtaacaatgcatccacttgagagagtggcggtcgccaggatggcagacgttttaggtcttgattcaaccatttcgcgtgtaaaactttatgaacttagccgcggtcatccgctggcggccaattatctcattaaggcactgttatcggctgatgaacaggacatatcatgcatcctcgccggagggatggaatttaatggcgatattgaatcagtttacgcatctgcctggagagaaatcgcaaacgaccctgatgttatgcatgtactgggtttcattgcccgtgtcgaagctccgatgccgctgaaattgctggcaacaatcgtagatgctcaggcgatagagcgtaccttaaagaccgtccggcatttactcaaggaaacctcaaaggggtggactgtattccataacagcttccgtctatttgtgctctccaaaccaaagataacactgggcagtatagatgaaacctattcacaacatatttatcgtgaattagctaaactatctcgtcatgcaccagaacattcattacagtcctggctaacactgcgctatctcgcccggtcaggagagcgtgatgaacttctggcactcgcaactccagcatatttcgacaccagtttgcacatggacgttcctgttcagagattgatgcggacattcacttggctctgattgctgcgcgttccacgtatgatggtgtaattgccacacggttattactttgccgtgatgagatatccagacgaactcaagcactggagtatgccaatgaacttccgcgcgcgatgttaaaagttggcgatattgatgcggcgatctctttcgtccaggactttcccaatgcgggctatgaagttgttgaccttcttttggaacagggtgattttgaccgcgcgaaagaactgtttgagcaccttgagccattatctcaattgcataccccccagattcgagcactatggggattcgcataatctacaagaattcaaaaaatgggcaaaacgagttgttcacttccgcgacgctgagcaaattaagcaggcaatagactatttgaccgttgaggggtttaaacacgccacaagtgtatcaaccgatgaaaatatttcctctattcgcgaacagttaaagtggacagtggtcgaggcaattgttaactggcaatcagacgttaatattcaggatacctgcaatcagtatggcattcatgtgcaagagataccggttttgatgactcaggctggatttattgctagagacagaggaaataacaccttagcatcggaattatttaagactgccatggcattgtctgattttaatgatgtttctaatggggggcgaagatcgattgcattatttatgccacatcaggctgcaccgatctggcttcaaaattattcgaaaacctttttgcgcctgcaatttcgatgggagacaatgaattagaatcaacaaaagcactgacgcttgcagccatggaacatgcgcaactttgcgttttgctcggcaaatccttgcccgacgtagtcacctcaacacacgctatcttacgaccgctgcagacacatgcttcagaaacgggacgcttgttggggctgtccataataaatgcctcatgtattccttctggaaatattaaaatggtctgtcgcatggtgatgagatatgtaatgcaactcaatagctattctggaaacgatacctatcaggctcaattggcattgacagctacatcaccactgatttgtacattaattaaaatttctgcgctgtgtggtaaggttgaatattattcagtaataaatgaaattgataatgcaatgcctgctttaatattaaaaggcaatacactactccggcgtgaaatagcattggcaatgtatcaggctgacggtgaccgtgaaagggcggccgccagatttgagcctatggtaaacgagttggtagaaaatacacctagcgagcaactcgagactctgtcagttctggcaaacagctttgctgcaattggcgatgttgaccgggcactaaacttacttgcttcgatacatgaccactgtttaggctacgctctggcagcgcgtaaggaccctttatactctgtttggaaagacatattgattttggccaatgcggcagacccagaacaccgtgctcaacgaataggtcagttgatacgacaggttgatggtatgaaggaaaccgagggagcatctgccgcatatcgtttgacagaagtgttaatcaatgaagcaatgcgtatgaatgcgcacagtggttataccgtggcacagaaactcagcaactgggggctgattccatggccaaatcaggtaaatgaactggtaattggtatgctagatcgccgtcctgaaatggtgtttctctgtacacaaatttggtgcgggctatgccttccattctacattgaaccctattatcgtgaccctacacatgtaggcaattatattgacgttgctgcaaatgcagcggggccttcatcaattgccaaactggtatcaattctattaccggcaatccaggttcatagtcgagctcacgagcgactcacgctaataaatcgcctgagcaaggcggcattaagacacggttataccgataaccaacttgataatgccattactcgatggacttcagaggcccccgaagcccgccgctcctacacgccacaaacgtacgacgaagcttcaacccttgacgaacttcaacaggcatttgaatcaaatgattccgaacctgagtatcatgcgccttatcgtttttgtgagcttgcagagtccgccgcattagacaaggtggtgaaaatgtatgagtgctggcattgcctgcagtcggatgcacgttgtcgtttttttggttgcagagcggctagttaatgcgggggacacgacgttagccagaaaattagttgatgattacgataccagtagtgaccgggagatgtcatggagccaatggttaggaggaaatcgattccgtctcttccacgcgcgtaagctactcgatggagcagcaattcatcatgaagcatatgaagacttcatcagttcaattgtggctgggaaagagagcaccatgtcgttgctaacagatatggcagacattcttcctgtgatctgtgagtcgccagactggcccgccgtctggtctatcctggcagagcagatgtctttcactcgcgaacaccgtattggtgaacttttcgaatttggaaatgaaaatatgaccgacgaagagttacttgcggaattgctccatttttcattacgattgcctatcaccgaagctcgacgacacgcagagaaaactgcactaattctggcggtacattcaacaggagggcaaatcgtatttgagaacaccataacacgactcctgaacggcacccttgatgaaccattccaggcattgcaaattttgcttttgctaaaacagaaccactttgctgctaaatttggtgatttagtctctg

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		gccttacgaatcatcgtgatgtagctgttgctgaagctgcgtgcttgttagcacaatattggcagctacctgtatcgattgattttcatccgttgccgttgacctatcgattggcactcgacggagaccctgatcatgaaaatgctctgttagatcctgtgagtggggcaatgcgtattgaagtcgacttaggatggacacaaatgcttcgtcccgttgcacggagacttgcagagtttgctgattgtgacgaaatgaacatacgccagcgtgccgcaacgtttattcagcaatggggagggctggcagcctttggccctggagcaacaaaaaaaatcgaatctcagttacgcacactctcaatgcaaatcacctatcttaagccccatgcttacattggcatactggcacttcgtcatgtcgctggagagctgagcttggcaggcttgctctcgccaagggataaaccatcgctactggaacaaatggatgcagtacttccgccaactcctcgccctgaaatgcaaatccggccaactggcattaggcgaccgcttaaagtcaaggatgccccgtggagtgaagctgaagaaatgtggacaaatttggttgacgaggatgttaaaccctggataggtcgtgccgacgaattcgtaatagccgaggtttcacaattcaaaatgcatgatacccggcgtgctgaatatcaggtctatcgtattagcgcacctcaaattcatatttctgatgccaaattcatggcatggtatcaaagtttgcccgctgtcgtttggctgggaaaaatgatcccacttgacgaagacctcgcaccgacaatagtcaggcgtgtagtaagctccatcgggacaatgtcttcgccgggatatgccattgcattatgtcctaatatccagatgcatctgggatggcatgaatgctgcgagatgcctaatatttataccgaccagaactcaacaatcgtagcaagattagtgaactggcgagacgccgggccagtggatattgatgatgattatatatgggggaaggttgctatctgacgctttccaatgcaggcctgatacaagtcaagactctgttcggcgaattcaccgtgcgtaatttcgcaagcagggctgttcggcaattgcgacaaggcgaagcgcaaatgataaagacagctcagaatcagttcccgatactgtagcgagacgatttcacaacacggttcgattacctgacttctccaaccatggtctgaagaagtcagggagtgtagatcatgccggcattctgtttctgaatggcgcaggatttcgggtcagggtcaccacaacaggcttgtccttttct (SEQ ID NO: 43)
47	pLG049	acaatttttgccataagacgctttcctgaaactcttctcattctcagcaggaaagcgttctcttctcaatactctctggttatagagtattaaaaaataaggagttataatccttgtagcccaactgacataaggacgatgctcaatgtctgacagcctgcttgttcgcaccagtagagatggcgatcagtttcattatctttgggcggctcgccgcgcccttcgactactggaacctcagtcaactcttgttgccctgaccattgaaggggcatcaacgacggaaatgggctctcagccagtggttgaggatggggaggagctgattgatattgctgaatattacggcagtaacgagctcgcaacagcaacaactgttcgttatatgcagctaaagcattcaacaatgcactcagatactccatttccccctagtgggttacaaaaaaccatcgaaggttttgcaacccgttataaggcacttatacaaaaaataccggtagaaacgttacgcactaaactcgagttctggtttgtgacgaaccgtccagtcagtagcagcttcagtgaagcgatcaatgatgccgcgaaccaacacgttacacgccatccacatgatctggcgaaacttgagaaatttaccgggcttcaaggcgctgagttatcgatattctgccagcttttacatatagaaggtcagcaggacgatttatggagtcagcggaatatcctgctaagagaatcagcgggatatctccccgacctggatactgaagcccctctgaaattaaaagagctggttaacagaaaagcgttaaccgaaagcgccgcaaatccttccattaccagaatggatgtgttgcgtgctttgggggtggatgaaacagatctttttcctgcgccctgtcgtattgaaagaatagaaaattccgtctcaagaactcaagaggcgacgctggttcaacgtgttgttgaagcattcggcgcacctgtgatcatccatgccgatgccggtgtggggaaatcaatttttctctactcatatagaggagcatcttcccactggttctgttagcatcttatatgactgtttcggactgggtcagtaccgtaacgcgtcttcctaccgccaccaccatcgtacagcattggttcagatggctaatgaaatggcatctcgtggtctctgtcatccattgatcccaaatgctggtactggcatatcccagtatatgcgtgcgtttctgcatcgcctttctcagagcatttcaatactccgggcctctgagcccttggccgtattgtgtattattattgatgctgcggacaatgcacagatggcggcggaagaaatcggtgaaacgcgttcttttatcaaagatttaattagagaaaagcttcctgatggagtctgccttgttgcactttgccgaccttatagacgggaattacttgatccacctcctgaagcactcacattatccctacaaacttttaatcgcgatgagacagccgctcatcttcaccaaaaatttccagatgccagcgaaagtgatgttgacgagttccatcgtctaagctcttgcaacccccgggttcaggctctgtcattatcacaaaatcttccactgaacgacacattgagacttttggggccaaatcccaaaacggtagaagatactattggtgaagtgctggaaaaatccattgctcgcttacgtgatacagccggaatatctgaacgtgctcaaattgatacgatttgttccgcactggcaatattgcgtccattaattccattatctgtgctatctgccatttccggagtagctggttctgctattaaaagtttcgcacttgatctgggacgcccgttaatcgttagtggcgagactattcagttctttgatgaaccggccgaaacatggtttcagaggcgctttaggccatcggccgctgatctgcatcagtttattactaaactgagaccactaacaaaagatagttcctatgcagcatcagttttacctgcattgatgctggaaggaaaccagctttctgaactgatcgagctagcgatatcctcacaagctctgcctgaaaccagcgcggttgaacgcagggacatagaacttcaaagattacagtttgcgttaaaagcagccttacgcacaggtcgataccaggatgcggctaaactggcactgaaagctggtggagaatgcgcgggtgacaacaggcaaagagtcctgctgagggacaatatcgatctggcagcaaaatttgtgggaagcaacggcgttcaggaactggtttcccgtaacgcatttccagatactggctggcctggctccagaaatgcttattatgccgcaatactttccgaatatcctgaactctcaggagaggcccgcagtcgccttcgactcaccatggagtggttaacaaactggagtcaattaccagatgatgagcggagcaggcaaaatgttaccgatcaggacagagcggtaatgctcattgcctgcctgaatattcatggcgcggaagcggcagcaagggagctcagaaggtggcggcctcgaaaactatctttgacgctggaaaaattgttgccatgcagttactggcccacgcccgttatgatgaacttgatcagttggctattgcggctggaaacgatatcagcctggttatgggaattgtactggaagcaagaaaacttcaccgtccagtcgctgaacaagcaatcagaagaacctggcgcttgttaaaaagtcagcgagtcagcattaaagacagaaaccacgctaataaccagacaatagcagcaatcactggcatggttgaaatggcgcttatccaatctgtttgtactgaatcagaaagcatccagttgttggatcgttatttaccaaaggttcccccctatgctctgacttctgagtatagtaaagaaagagttgcttacgtccgggcatatgctctgcaggcaaacctgatgggctctcaattagcgcttagcgatttagcctccacagaggttaaaaaagaacttatggctgaaaaacgccacggcgaatctgatgacctgcgtcaactgaagcagtacagcggagtattaatcccttggtataatttatgggccaaagtaattcttggtaaaacaaggaaagcagacttagaaagtgagctaagtgatactcaaaaagaatcgacggctattaaaggtcattcttactctgagcattcattatcatcaaatgagatcgcaaatgtatggtttgatattctgatcgaagcaggtaatgtatcaaaagacgatgtggaaaacatcatcaaatggagtcagcataaagggaatagagtattcacaccaacgcttcaccgtttcagttctgtatgtgcagagatttcagggcttggagagctttcatatcacttcgcagaacttgccttatctttatggagggatgagcactctgatgctcagatcaaagctgacggctatatagacctttcccgttcactcatttcacttgatgaaccagaagctaaagaatactttaaccaagcgattgaagttacaaataagttaggcgatgaaaatttaagtcgatgggaagcgatacttgatcttgctgaatatgttgctggtaaaacgcaagtccctcctgaaatttcctataaactagcccgatgtgcggaactaaccagagaatatgttgatcgtgataaacattttgcatggagtgatactgttgagattttggctgagttatgtccatcttcagccctagcaataataagtcgttggcgtgaccgtacatttggcaatcatagaagcatactggcatggaccattgagcatcttgtaaagaaaaataaaattaatgcactcgatgcacttcctttaatcacatttgagaatgattggcataaatgcgacttgcttgattcagttttatcctcgtgtactgatgacaaagataagatcatggcattcgaagtggtttaccactatacaaaatttaacgtacaaaatatccaaaatcttaaaaagctggatgctatttctacatcattaggtattgaacacacagaactgaaagaaagaatttcaggtctacaacatactgagacggtttcaaaaaaatccagtctctcatcgaatgataatgagcaaggccatgaccaggaatgggagtccatttttaaagattgtgatttatcgtctattgatggtattagtgcagcatacgaaaaatttcgtaatgttcctgaattctattccaaagaaaccttcatcaagaaagcaataagccgagttaagacgggcaaagaatgtagtttcattactgccattggtgctatatttcactggggctttatgatttaaatatattcttgaatctatacccgacgaatggacatctcgtttaagcattaaaaccaccctggcaggtttaataaaagaatattgccaacgcttctgtatgcgaatcagaaaaagtcgcgtttacgagatttttcccttcagtctggccagcaggctttctggtataagtgaaaaagagattttcggtattaccctggaggccattgcagaatcgccagagcccgcaaactctgaccgtttatttagccttcctggccttcttgttagtaaactggagagtaatgaagcgttagatgtattatcttatgccttggatttattcgacgag

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		gtgctaaaagatgaggatggtgacggcccatggaacgagaaattatctccgccaactcatgtagaggattcacttgcaggctatatttgggcgcggctgggttctccggaggcggaaatgcgctggcaggcagcacatgcggttctggcactatgtcgaatgagtcgtacatgcgttatacaaggaattttccagcacgcaataaatgctaccactttacctttttgtgatcgcaatctgcccttttataccctccatgctcaattgtggttgatgatcgctgctgcaagggttgcgctggatgatggaaaatcgctgattcccaatattggttatttctaccattatgccactactgatcagccacatgtattaatccgtcattttgctgccagaactttacttgcactgcatgatagcgacctgatctctatcccagcacaagaagagaataaactccgaaatataaaccagtctacgactctccctgtgcttgataaggttgaagatcatagaggtgaagattcatatacttttggtatcgactttggcccttactggctaaaacctctgggacgttgtttcggtgtatctcaaaaacagttagaacctgaaatgcttcgcattattcgtgatgttcttggttttaaaggtagccgcaactgggatgaggatgagcgtaataaacgacgctattatcaagacagagataatcatcacagtcatggttcctatccacgggtcgatgactaccattttacttgtcataccatgcaatgtttatgaccgctgggcagttattagcgacaaaaccattagttggtagtgactacgacgatgtcgaggatgttttccaggactggttaagaagacatgatatttctcggaacgatcatcgctggctcgccgatcggagagatattcccccaaagagcgctccagttggcttaatagcagttctgacaatagggatgaatggctagcgtcaatctctgaaaatgtatttaacgaaacactatgtcccagccccggactattaacgctatggggacgttggtctgacgtttgttcagatcgaaaagaatctattattgtccattctgcgttagtatcgccggagcgatctttatcgctcctcagagcattacaaacaactaaaaatgtatatgactataaaatccctgatgctggagataatcttgaaatagatcacgcacactatcagctaaaaggatggattaaagatattgctgaatactgtggaattgatgagtttgatccctgggcaggtaatgtaaggtttccaatcccagaaccagcctcatttatcattgatgcgatgaaattaactactgataaagatcatcgggtatggtattcaccttctgatgttgaaccggcgatgatttccagtatctggggccatctatcaggtaaaaatgatgaggaaaaatcacatggttataggctatgtgcttcaatacacttcataaaatcagcattagaaacattcaacatggatctcatttagaggttgatgttgatcgctattcacggaacagcagatatgaacggaataatgaaaatgagctcgacaatatcccttcaagcactcgactcttcctcttccgacatgacggaaccatccacacgctatacggcaattatagaaatggggaaaaaactagttgatgagcttgagctaaatgactctgttgatacattaagcagatggatggctcatcatatcgcagagctcatttatgatgctgaacattgtacagacgacatcgtccgtacagctaaacaagcggagattagggactctatctggtcattctggtctaacagatacgaattgccaattggtagcagaccatttcaggagctcgaacctattctaagaaccttaaaaggtcttgatcctgaaaatgagcaaccgagatttttttcaccttaccgagatctaattaatgtagaaaaagaaaccagtgaggtccaaaaatggctaaccgccgctaaggatattgattcagcagcaaaaatactgattgattactgtttatcgttagcagcagaaaatgctatcgataaatcccaagaatgggtggaattagcacagaaagctggattgaacaaagatgttgatctgcttgaaattcgtatctttcagttacgaggtaccccagccaatacagacaatcccaataatgcacaacggagaatactggaaaaaaggcaaaaaaggcttgaagcttttctcttattgggctcccagttaaacgaacaactcaaatctcagcttgaagccttaccagcaattgaggatgagccaacggatgacgacgaagacttttgatatgacttgctttagcactggagacggctcacaagacggaccacataatagcctaacccaagacttttctactagtcctaatg (SEQ ID NO: 44)
48	pLG050	ttgtgcgtagcacttctccagtttttgttgaaacagataaagagactaaatcgatcattcgaacccaaaaatggccgatttgatgcagacaacgatttaagccatatctggtagcgcaatcgtcacctatgacaaaagttacatacttgtaatattctgaattcaatattcttcgtgaaattcattcaatgcttctttgagtagtgttttggcgttatgataatttcctaaatatcataaggttatcaggcggtgatgtatgaggcgatttgtctatggcgattaaaaacagcgcaatcatttatgcaggctatgattatcagacactccaaggtgtcaggctactggcggattggctcaatacaccaactaaatataaccgaatagcatttgaggctgatgcgaaacaagttgatgctccacaaggcattgatgatattgtctgcgaacgtcaggatggtaaaacagattttggcaagttaagtttacgccagataccgacaaagaagacaatcaactatcatgggaatggttactgaaacgtagtggtcatagtattcgagctcgttctatactgcaaaaaatagctgatgctgttgataaagtacctgcggaaagaaggggagatattactcttttgaccaataaaatacctaatcgtgagatagcaacttgcttgcgaaataacaaaatagattggaatcaggttccaattgctaagcagcaaagcattattcttcagttaggtacccaggaaagagcaaagcaattttttcgatatattacaaatatgtcatagtgatcaaagttatacgcgattaaatagtattgtcccagaactacttcgcaaacataccaacgaggaggggtatatcgcctgattgaacgagctaaacgttgggctatccagcgtaattcaccttcggatggtggatggatatgtcttgaacatattcgtgcagtgatttcaactaatagacctgaacctattccgcagacttttgtcttgccagataactatattgttcctgatgcagattttcacgacaaattcattgattcacttttaatcctactaatcgattagttgtcttaactggtgctccaggaaagggtaaaagtacttacatcagccatatttgtcagatattacaaactcgcgagtttccttatattcgccatcattattttcttgggttagatgatcgtacgacagatagattaagtcccagaatcgttgctgaagacttgatgtgtcaggtcaaagcatttgctcacaaatcgaaatgaaaaattatcatgcagagcacctacataaagtgctggctgaatgtgggcagatatataaagaagaaggtaaacgatttttcatcattattgatggtttggatcatgtctggcgtgataacggcaaagataaatctccactggatgagctattttgccaattgttaccgttgcctgataatgtaacattattggttggtactcaaccagtagatgatgagctattgccatcaagattgttacagaacagtccaagagaagaatggttgcacctaccaaatatgtcaggcgatgctattcgtaaatatctatcgggacaagttgaaagtggccgtatcgtattcaattttcatcaaagccagtatgaagaagttttatcacagtgtgctgagttgttgactactaaaactcaggatatcctcttcatgttatctactcatgtgaaaaattacatgttgaaggtaaagggttatcgcactgggaaatagaaaacctgcctcgctgcgaaggcggaaacattacaaattattataatgaattatggaaaatattaaattacgagcaacgcgatattcttcatctctgttgtgcttttccttttttatggcctgccacatcattttctgagatttttctgagaggactgaaactataccgaatgttaaggctgtaatccatttgctttatgagtccattgctggattaagaccgtttcatgaaagcttgattgtttttacccgtagcacaactgaacatgagaatagaataaaattattattgccagcgctaatttcatggctggagaaaagcgcacccaaaccgataaaaaattgttggtactggtcatgtcttgcttacaatggtgatccatatcctttaagaaatggcttaactagagactggatattggaacggttggctgaagggtatcgacaggatgagtttattcgattactcactcaggctgaaacttctgctttagccgaagggcattttagtgaggcctatcagcatcgttcacgcaagactcgactacttaatgctaggttgcaaatctgggatatgtcgacgttgggcgtttgcagtatgattaatgcttctgaagcattgcttaaacaatatcaatctacccagaatgtcagttcaccaaagatactggcaactttggctatcgctttatggtttcgtaatcatttcgatgaagcaaagcgcattacaagattggcgttacaacgctactcaaatgaatcatccgtatataccaataaaaatagcgatgagtcgcgtgctgacattcgtttattaatcaaagctgctgttttgactgagtgtttcgatgaaaaatggttggcaaccggttcagtacacaagtggagtgatagtaatattaatctgcttatcgaatgtgcggaatataaatcagatataggattactattttcattacatgatgtttttaagcaaactgtcataaaaaataaaatagtaaatgcgattgtcagagttgggattgttgaacaaatagatttagaatactggccacatttttctggtcttgactccgctctgctgcggttatacagtcatttatccactgcacatccatgttcacttataacagagcaaggtgaaagtgaaatcggtagatatcatgttcatccagaagtatcctacgatgaatggttctatgacagcctttttttatcgtcttaatgccagtggagattattgttggctaccggttagcacgggggaaggacaggaggaagtcagcagtcatttctccatttaaatgatttctcagatattattgctgaaagtatggctctaaatattcaacaaagcttcagcgattttttgttcacttattgctttggtatcagatcttaaagatcatcaaatgcaaatccaacagaagcgaatgtttttttaaaactgattgggtaagcattgctttaaatttacacttaatcatgcattgcaagccggttaatacggaagaaattgatattattcttaattctgagcatacagccctgtatcggctgcataaaactattcttaactttcatagtagagccttcgaatctgatgcaatagcaaactttctggtatttgaggatgggaggcagaaggaaaaactacaagagacaaatgaatatttggcgaataatcttgagttgtcagagattgcgcttcattatgatctcaatcaatcaatttttttgagcgagtcaagttatgttgggactatggtctgggatacggacatcataaagatatagctctgaatcaggtgctgactgcaataaaaactattg

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		caactgttgagcctaaatatgcattaacgcagcttgagcgtgtgagtccattggttcataatatttgtgacttcacagatggtgaccatactcaacattccgtaacggaattgtctgcgctatatgctcatctttctccccttactttaagtagtatctatgacagttatgttagcgagggtgagtggtatgatgcggataatgcattaacgcaatacttaaaacatgctgatctatcatcacctttcgttgagagtttatgccggacattactagatgatgggcaaattgaaataatacagaatcgtgctaaagacaatgccatattgactacgttttggccggaaatattaccacgaaaaatggattatagtagtagcgcaaaacgttcattaaggggactgaaaaatttgatccagcaaaaatcagccctgctgatgtaactaatttactcaatgttcggtcaagttatgaaaatattcctaagtggtatcattattggaaagaccaaggaaaagttacagaagtaattaacgtattgctgccaatcattaataatggcttgccagaatatagtgaatttcgttatatattatctgatttatttgaagatacattgcgtttgaaaggtaaaaaatatgcttttcccattttagtgcaggaacatattcagcgaaatggttggggtgaatgggggagtctgatgatcaaacatatgctcggttagataaagttatcagattgtatccggataaaattgatgactttctttacaagacgactcgacttcatcactataaaactaaagaagagaacttggtaattcccgggaataagctaacatatttattagtaaatgtaggccgagtggatgaggcgaaaagtctatgtgaagcgatgatttcggaggtagaggcagaaacccagaatcttccgttgtgcaaacctcaatggcaatgggagggagaattagataacgatatgatcgccgttaaattcatcattcgtcgtcttttttggcctgttcaatgtgtaaaacatcttgtcgctgatcaattgtctcatctcttagttaatggtcaatgtgctgaagaaattgaaaatttacttgtagttgagatgggaaatcgtcaactggagtcagaggtggtagatattttaactgttctctggttagctagtttgaaaggttataaggttcagaataatatatcttcctttatttatgctcgtagctttctttcagatgcattgctggaggctatcgttccaaatttaccaaacctcagtcgctatcaagtgctgtataaacatcctgatgatgatggtaatcactatggctttgaaaaaacacttggcaatgaacttccccatatattttgggatgaagtaaaaaggcttgaggagaaatctggagctccggctaaaatattaatgaaaaaagaatggaatgatatttgttataatcatgttcaacgatgggaaagggttgattatttcttcggttcagagcgtgatggttttactatgagtttttccacaaggaatacacgatttggtatatctgcatacttgagaaccattaaccggcttatcaacgaatttagaatgccaaagcattatgcagaacattattcgatttgtttaatgtcagccaacccattattttattccgtatctaatcaccgacctggttggttacctttatggcaatatggggagattaccacaaaggaaaatgtaaaaacatatgttgaggaatgcctgaatgcattcaaaaatgaacaggaaaattcaatattaggagcattgtcattacctgtacgcatcgatgaaaataattggttagatattacggctgttatggggatacaaacagaagaatatgcctcttttaagatacaacatgccgactgtggtcatagtgtagatagtttacttcaagcttatagaaatattaaattttcatttgcaaaatgggctgaataccaaaattgtgtaccactattgggaagtacacgcgaattactgagaatagcacggtgggatataatgtacgaatttcgtgggcttttctcattcggttgccaggaacaggttactgcctacccggctaaaaatcgtattaacttcgattatcagggtaaaaccatcggctatagtgacttctggcaagcaataccattatcaatttatcctaaggatatacgctcacctgttgctacttacactgcttatgataaggaccttgcctgtaactggaaaaatcatagcgtactgaaaaagcctaatatcatgttatgtgattgtaaggtactaaagagagaaaatagttacagtccttttgaaatatcagatattcgttttcactttgaatctgagccgttatagtaaggattattttgcgataattaatcaacggggagctggtcaaagtgcctgctcccatattgactaatatacaaatgtgtttgttaagacctttccaaaggtagggggaattatgaatttccgctcctcgctcatagccgcctgccagatttaaccccaccctaccacagggccccctcaagccaagccgccgccaatacaattttcccccacaccaaaacgcctccctccctagagcacgtactcacaacgccga (SEQ ID NO: 45)
49	pLG051	gggatttccaccacctcccaccgaccatctaagactttatgccactgtccctaggactgctatgtactaggagcggatgttaaactcagactcgtttcagctacattgcgttttgaataatattccatcataataactctttgaaaaatgtgatcttttcatttataacactgatgacttgcttatctcattgggatatcggaggagaatacttaactatgacaagcccgattattatgacactggctatattatatagattgatattaaaatgtaggattaggttcttgccaaggtgtcaagatttacagataggtttaaaaccatataaatatgttttacggtgagatacaatacatattgtaaggcataaacgcttggtaaaattttaattattggaagaagctaatcatggaacccatatcaattacagtggcaacttatgtagcaactaaacttattgatcaattcatctctcaagaaggatatggttgtattaagaaagcattattccccCaaaaaagatatgtggatagattatatcaactaattgaagagacggcaattgagtttgaagaaacatatccagtagaaagtggagcaataccattttatcattccgaaccattgtttgagatgttgaatgagcacatcttttttaaagagttccctgacaaagagatattattagacaagttcaaagaatatccaagtatcactcccccaactcaacaacaactcagcctttttatgagatgttatcattaaaaatcaataattgttcgaagttaaaaagctacatatcgaagaaacgtataaagaaaaaatattcgatattaatgaagagctcattcaagtcaaacttattttacggtctatagatgagaaactaacttttcacttaagtgatgattggttaaatgaaaaaaatagtcaagcaatagctgacttgggaggtcgatacacacccgaactcaacgtaaagctagaaatagcagagatatttgatggcctcggtagaactaatgatttttctaaaatatttattcgcatatagatagctttctggtcgctggaaagaaattacatagttgcgatgtaatttcctcagaattatttgaaataaaccagtccttaaaagaaatttctgatatatatcaggagattaatttttctaaattagatgaaatccctataaataaatttaataactatgtttctagctgccagacagctattggcggagcggtatcaatattgtgggaactccgagaaaagtcagagcaagtaggtgaaaccaagcattacagtgataagtattcatctactctgcgaatgcttcgggaatttgactatgcgtgcaatgaattacgtatattcattaattcaacaacagtgaagttggctaacaacccattcttacttctcgaaggaaaagcaggaattggtaagtctcatttactggctgatgtgattaaaaatcgaattgcttctgggtatccttcactactcatactagggcaacaacttacttcagatgaatctccatggtcacaaatcttcaagagattacagcttaaaatcacttctcgtgaattcctagaaaaactgaatttatatggcaaaaaaacaggaaaaagagtcttagttttattgatgctattaatgaaggtaatggaaataaattctggaatgacaatattaacagttttgtcgatgaaatcagatgctttgaatggcttggtctgataatgtcagtcagaacaacatatagaaatgtaacaatttcacatgagaatgttgtgcgaaataattttgaaattcatgaacatattggattccagaacgttgagttggaagcggttagtctattttatgattattacaatattgagaggccttcatctcctaaccttaatccagagtttaaaaatcctctatttcttaagttattgtgtgaaggcattaagaaaaatggtttaaccaaagtgcctgttggatttaatgggatttcaaatattttaactttttagttgaaggggtaaataaatcattagcatcgccaaaaaaatatgcattcgatcccagttttcctcttgttaaagatgctctcaatgaaatcataaaattcaaattagagattggtcgtaatagtatttcacttaaagatgctcactcagtggttcaatctgtagttaatgattatgttgctgataaaccttcctcagcgccttgattgacgaaggattattgactaaaggcatagtgagaaatgatgataattctactgaggaagtagtttatgtggcttttgaaaggtttgatgatcatttaactgttaatttttattaaatgatgttgaaaatatcgaaagtgaatttaagcctgatggtcgtctgaaaaaatattttcatgatgaatgtgatttttatataaaatcgggaatagtagaggcgttgtctattcaattgccagaaaggtatgaaaaagagctttatgaatttctgccggagttcagcaataatcttaaattactagaagcctttattgatagcttgatatggcgcgatattaaggctattgatttcgaaaaaattagacctttcatcaatgaacatgtttttaaatttaaagatagttttgatcatttcctcgaggcagtgatctctatttcaggtttagttggccatccctttaatgctaatttcttgcatgattggctaaaagattattctttggcaaatcgagattcgttttggactacagaacttaaatataaatatagtgaagactcagcatttaggcatctaatcgattgggcatgggccagaacagataaaagctttgtttcggatgagtcaatcgagctagttgcaactagtttatgctggtttttaacttctagtaaccgagaacttcgagattgctcaactaaggctttagtgagtttactcgagccaagaattcctgtattgagaaaaataattgataagttttatggtgtaaatgatccttacgtttgggaaagaatatttgcagttgcattaggctgtacattgcgaactgataatattaaagaactaaaatatttagccgaaactgtttaccaaaaggtattttgttctaagtatgtgtatccaaatatattacttagagattatgctagagagattattgaatttgctaatcatcttggattggaacttgaaagcattgaattatccaagactagaccaccctacaacagcatttggcctgacaagattccttcaaaagaggaactagagtccctttatgataaagaaccttatcgggaactctggagctctattatggaagatggtgacttttcacgatatactattggaacaaattataatcattctgattggtctggttgcaagtttaatgaaacccctgttgaccgtaagca

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		agttttaaaactttcaaatgtaaactaactgatcaacaaaaagacttgtatgatgccacagatcctttcatttatgatgataaat
		gcgaaggaattaaatttggtcgtgtggtcggtagaaaagcacaggaagaaataaaggcgagcaagaaattatttaagaattcattg
		tcatacgatctgttaagtgagtttgaaaatgaaatagagccatacctggatcataataataatctgctggaaactgataaacactt
		tgatcttcgactagctcaacaatttatattcaatcgtgttatagagcttggttgggatccggagaagcatggtaattttgaccaac
		aaataggaactggacgtggacgtagagaggcattccaagaacggattggtaaaaaataccaatggattgcttattatgaatacatg
		gcaaggctagccgataattttactcgttttgaaggttatggtgacgaacgaaaggaaaatccataccaagggccatgggagcctta
		cgtaagagatatagatcccactatcttacttaaagaaactggaacgaaaaaaataagcaataaagaaatgtggtggcttaatgatg
		aagtgtttgattggacttgctctaatgaagactgggttaaaagttctactactataactaattcatatgcttttattgaagttaaa
		gatgataatggtgatgaatggatagtattagaaagtcatccatcatggaaagaaccaaaaattattggaaacgatgattgggggca
		cccacgaaaagaggtttggtatcagatcagaagttatatcgttaaagttgaagaatttgaaaattttagatgttgggcaatagctc
		aagactttatgggcaggtggatgccggaatgtactgatagataccaattatttaatagggagtactattggtccgaagcatttaag
		tcttttaaatcagattattatggtggatctgactggacttcggtaacagaccgggagtctggagctaagatagctgatgttagtgt
		cacttcgattaattatttgtgggaagaggagttcgacaaatcaaaaatagaaactttgaatttttttgaagcctagtaacttaatct
		ttgaaaagatgggattaaaaagtggggaagtagagggtagcttcaatgatgaaaatggaactatggtttgctttgcagctgaagct
		gtatatgcttcaaagccgcatctacttgttaaaaaagaaccatttttaacaatgttaagggacaatggttttgaaatcgtttggac
		attattaggtgaaaagggcgttatagggggctcactcatatcaagtcatcattatggtcgacaggagtttagtggagcattttatt
		atgaagacagtcagctaacaggaagtcataaaactagctttacgagataaaaatgaatctcagagctgaatatataagtagtatta
		gaaaccgggttatacttaagaaatcaatcttaagtgtggcagtcgaatggtagctaatatgctagcggcgctaatgcctgtttgtt
		gctcataacaggcattcactttagttatggcagaaaagtatacatgctgggttgggaaagtgtgaaagaaaggaagattgctgcgc
		cgtttgtcgtcacgtttatcttcattggctatgca (SEQ ID NO: 46)
50	pLG052	aaatctctttcgcgtcaatagtggtaatatttttttatcattgtcctctttctactgacatactgattgtccgacagtggagccag
		tcgaaattgttgacagctagtcggggctcgtctggtctttctagcagtaagaaacgtattaatattggatcgccactagtttaaca
		gatacctcagaattatttatagactgacaccaccccggcagacgatcctgccctataggaagctaagtggaaacttatccagtaac
		agcttgtcgattttatcccagagggtgttcctcaggatgtatcgctgaaatcaaatccagcactaagaatgaggggtgagaaacca
		tttccttggtgggtctttgaccatttctgttgaactaatgtttttgggttatcaaggatacaaattcaaggcagtgtttcactaaa
		ccttacctcgcttcaataccaatacattttaatgggtataatatgtgactgcttttgccgcattattgacaggaacaaggactgg
		tgatgaatattgatttcagtttaattcgtagcgcccccaaaagccgtaacgatagctttgaagcactcgccgtacagttatttagg
		aaaacctgtcgagtaccgacaaattcaacatttattagtctgcgtggagatggtggagacggtggcgttgaggcatatttccgctc
		accggacggtgccgtattcggtgttcaggcaaaatactttttccagcttgcttccgcagagcttacacagattgatagttccctta
		aagctgcgctaagcaaccatcccacactaaccgaatactggatttatataccgtttgacctgaccgggcgtgttgctgcgggaaag
		cgaggaaaaagccaggcggaacgctttgaagaatggaaaagtaaagtcgaatcggaagcgtcagcgaaagggaagtcactttctat
		tgtcctttgtaccgctgctgttatctgcaatcaattacttgagatagacccttacggagggatgcgcaggtattggtttgatgaca
		cgttgctgacaacagctcaaattcaacaatgtctggaggacgccattgcttttgccgggccaagatatacttcaatgctggatgtg
		gtgacgaatgctcatgtcggcctggatttctttggtgggactggtgacttttgcgagtggtacgaaacatcattaacaccaatcgt
		tcgagagttccattcactgaatggatacggacgcaaatcgctggatatactcggcgaaacccgtgctacatctgccacggcattga
		ttgaagaaataattgcctactgtgagagcatgagagataacaatgtcacggccacatcggttacagatctttccgtcgctctgtca
		tccctattgacacttttcgctgatgcccgccatgctcaagaagataaattttatgaaaagcatggcaagcatagtgatacagaatc
		gttccgacagttccacgcagagtatatgtgtgcatttcctgccggagatatggatgcggcgagaaaatgggaagagcaggcgcagc
		aactgcaaaatttgctgacttctcaggtcattggtgccgcaacagcacattccttactgctggttgggccagcgggtatcggcaaa
		acccacgcgattgtcagcgcagcattgcgtcgactggaacatggtggtttttcactggtcgtctttggagacgactttggcaaagc
		agagccttgggaagtgctacgcagtaaaatagggctgggtgccgccatcgatcgttcgacattatttgaatgcatacaggcctgcg
		ccgaacatactggcttaccttttgtcatttatatcgatgcattgaacgaaagcccgcgagaagtgcgctggaaggacaagcttccc
		gaattgctcgctcaatgcaagtcttatccagacatcaaaatctgcgtttcaacccgagatacctatcgcaatcttgtggtcgattc
		acgctttccagggtttgctttcgaacacatcggtttttcaggacatcaattcgaagcggtacaagctttcgcagcctactatgagc
		tggatgcagagattacaccactttttcacccgaactcggtaatcctttatttttacacttggcctgtaaaacgctaaagggcgaa
		ggccgtgacagtctggatatttctttgccgggttttacctctctgtttcaaggacatctcaaacattgcgatgttttaattcgaga
		acgcctccactacgcaaaccctcgtaatctggtaagggctgcaatgatggcactcgcgaaaaccctgacacatgagttgccgcaga
		accgaacgtgggaaacctgttgcgaagcactgagcaaaatagtgggaactgagaccacacctgaatcctttttaaatgcattggca
		catgaaggcctcattatcctttctgttgtagatgaggataccttcctgatccgtctgggttatcaacgctacggtgacatactccg
		tgctatcagccttgtggaaactcttgattcggatacagtaaaactagcggagaaaattgcagcgttaacagaagaagatgctggat
		tgctggaagctcttgccgccgtgctgccagagaaaactgctcttgaaattactgctgaagaagtaggattaccatccgaacaagcc
		cataagctgttcatccagtcattggtttggcgctcccgacaaagtgtagtggaagaaattgatgaacacatccatgcagcactgca
		tacacctggattatgggagtcggtttatgaagcgctgttttcacttagtctggttcctgaccatcgtctaaacgcaactaactggc
		tggggccatttttacggcagtcatccttagctgaacgtgacacctacttgtcattagctgcgctgggatcatttgataataagact
		gctgtctattcactcatccatgcagcactatttgctgacataacccattggcctgctgaaagccggaggctggccagtctaacact
		tgcctggctcacttcgtgtgctgaccgccgaatcagggatttatcctcaaaagggctaagcagaatcctggcaaactacccggaga
		actgccaaacagtaatcagtgaatttgcatattgtgatgatgattacgtattagagcgtattagccttgctatctacagtgcatgc
		ttattgtcataccaacgcagaaatgcgtttatgccagcgctccctggtctattaagcattgcgtcagatagcaagaatattctgct
		ccgggatacggttcagctattagtaaacttgttgaaaacaggagaatttcccacagccgtaacaagccaattacagcattaccaga
		caaacgtatcattaccatcacgatggcctgtactggcggatgtcaaacccctcctagatctggaacatttaccatcaaacatggtg
		ctctggggagaatccatggccccggatttctggcgttatcaggtggaatcgaagatttccggctttgacttggagagcgccaatat
		cagccatgaaaacattgcctgttggttaatgcgagaagcacttaatttaggatatcccggttataaccactgcgcgctcaattatg
		atcgccatatcgggagtcagtatggctcgggacggggtagaaaagggtatgctgaccgactcggtaaaaaatattactggatcgcc
		ttacatcgactactgggcattctggccagtaatgttcccgcactggaagacccatattccgactacgaacctacaagtgatcttct
		atggtcagtcgacgtccgtaaagttgacctgaccgatgtacgcgatatcaccgcagaaggtgtctatccagtactgatggaggaaa
		caaattatgcattccctgaccacaattcagatatcaaaggttgggttaggaccgatgattttccaccttatgaagcttgtcttatt
		cgaactgacgaggaaggagagcagtgggtagcgctttcacatagctattgggatgacgataaagcgccgaatgaaaatagctggga
		ttccccgtacttgggagtgcgtgcttcctactcaagcgcactcataaatgaaagcatccagaactttaaacagaaaagatcacgcg
		atattttccaatataatcagggaagtagttgttatcgcggttatcttgctgaatatcctgacagcccggtatacaaacaacttctt
		aatagtgatgaagatagtgaagcgtttaattttacagaagtcagtttactgcgcggaaacgaatgggaatacgactactcatatac
		catgcccgagcgccaggataacctcattgcgccatgcctgggaattattcaaaaactcgaacttttatgggattgtcaaagcggtt

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		gggttgatcattctggcaaacttatcgccttccatcaaaaaggtgtaaaacaacgcggacttttcatccatcgttcggcattgaacgcctatctgtccataacaggtgaagagcttatacatcgccgttttgctaacagaggatattttgatttagctggtcgtaatagcacgcaaatagacctgaaaacttggatccagtaccgggcagacaaggcaccggtagttttacgagaagaggaactgccgtttaactgctgacaacgatacttattaagtaatcaactggctgccttggcatcgaatgccagaagagccatttcgcactaccaatttaagtagactgaaggaatacttggtacaagcaaacgcacgccatatcggatagaggggact (SEQ ID NO: 47)
51	pLG053	gcgcagctgacaaagattgaccgtgagcgctctgatggagaaagacgatagttgctgagtacgatatcgagggtacatttctctgtgtaggggtagttatttacaaaaaaataggagaataattaaatggtcaaaccaaactgggataactttaaagctaaatttagtgagaatcctcaaggtaattttgagtggttttgctacttgttgttctgtcaagaattcaaaatgcccgcaggtatatttagatataagaatcaatctggtatcgaaactaatccaataaccaaagataatgaaattatcggttggcaatctaaattctatgacacaaaattgtcggataacaaagctgatcttatagaaatgattgagaaaagcaaaaaggcttatccaggattaagtaaaatcattttctatactaatcaagagtgggggcaggggagaaagtcccatgaacctgaaggcgataagaacgctgataattatttggaaactgtcggaaatagtaacgatcccaaaataaaaattgaagttgatcagaaagcatatgagtcgggtatcgaaatagtatggagagttgctagtttttttgaatcaccgtttgtaatagttgagaatgaaaagattgctaaacatttcttctcccttaatgaaagcatctttgatttattagaagaaaagcgcaagcacacagaaaatgttttatatgaaattcaaaccaatatagagttcaaagacagaagtattgaaattgacagacgacattgcatagaacttctacatgagaatctagttcagaaaaaaattgtcatcgtcagcggagaaggtggggttggaaaaacagcagttatcaaaaaaatttatgaagcagaaaaacaatacactcctttctatgtctttaaggctagcgagtttaaaaggacagcattaatgagttattcggtgcgcatggcttagacgatttctctaatgctcatcaagacgaattacgtaaagtcatagtcgtagattctgctgaaaagcttttagaactgaccaatatcgatccttttaaagaattcctgactgttttaataaaggataaatggcaggttgttttcacaacccgtaacaattacttggcagatctgaactatgctttcatagatatttataagataactcctggaaacttagtaataaagaaccttgaacgcggcgagctaatagagttatctgataacaatggatttagccttcctcaagatgttcgattattagaactaatcaaaaatccattttatctaagtgaatatttgaggttctataccggtgaaagcatcgattatgtgagcttcaaagaaaagctatggaataagattatcgtcaaaaataaaccttctcgggagcagtgtttcttagcgactgcttttcagcgggctagtgagggccaatttttttgtctccccggcatgtgatactggaattttagatgagttagttaaagacggaattgtcggctatgaagctgctggttacttcattacacatgatatatacgaggaatgggcattagaaaagaaaatttctgtcgattatatccgtaaagcgaacaataacgagttcttcgaaaaaataggagaatcacttcctgttcgccgtagttttcggaattggatatctgaacgattgcttttagatgaccagtccataaagccttttatcgcagaaatagtctgtggagaaggaatatcaaatttttggaaagacgagttatgggtagctgtccttctttccgacaattcaagcatatttttttaattactttaaaagatatttacttagtagtgaccagaatctattaaaaagacttactttcttattgaggcttgcttgcaaggacgttgattacgatctgcttaaacagttaggtgtaagtaattcagatctgctttccattaaatatgttcttactaagcctaagggaactggttggcagagtgtgatccaatttatctatgaaaatttagatgaaatagggatcagaaatattaattttatacttcctgtgattcaggagtggaatcaaagaaacaaagtgggtgaaacgactcgattatctagtttgatagctctaaaatattatcaatggactatagatgaggatgtctatttatccggaagggataatgagaaaaatattctgcatacgattcttcatggggcggccatgattaaacctgaaatggaagaggttttagttaaggttcttaaaaataggtggaaagagcatggtaccccatatttcgaccttatgaccttaatccttactgacttagattcatatccggtttgggcatctctcccggaatatgttctacaattggcagatctgttctggtatcggccacttaaagaaacaggcgaacgttatcacagtatggatattgaagatgagttcggtctatttaggtctcatcacgactattatccagaaagtccatatcagactcctatatattggttactacaatcacagttcaaaaaaacaatagactttattcttgattttacgaacaagacaacgatatgttttgcccactcccatttgctaaaaacgaaattgaagaagtagatgtctttattgaagaaggaaagtttataaagcaatatatatgcaatcgtctgtggtgctcataccgaggaacacaggtctctacctacttactttcatcaattcatatggcattggaaaagtttttcttgagaattttaaaaatgcagactcgaaagtgttggaaagttggcttcttttcttgttaagaaataccaagtcagcttctatttctgcagtagttacgagtattgtacttgcattccctgagaagacattcaatgtagctaaagtactattccaaacaaaggacttcttccgttttgatatgaatcgaatggttctagacagaacacataaaagttcattaatctccctcagggatggctttggcggtacagattacagaaactctttgcacgaagaagatagaattaaagcttgcgatgatgtgcatagaaatacttatcttgaaaatcttgccttgcattatcaaatttcaggagtgaaaatgtaacggagaaagatgccattgaaaggcaacaagtgctctgggatattttcgacaaatactataatcagcttccagatgaagctcaagaaactgaagccgataagacgtggaggctctgcttggcaagaatggatcggcgaaagatgaaaataactaccaaggagaaagatgaagggattgagatatcattcaatcctgagattgaccctaaactaaagcaatatagtgaggaagcaataaagaaaaactccgagcatatgaagtatgtaacgctgaaactatgggcaagctataaaagagaaaaggatgaacgttataagaattatggaatgtatgaggacaatccgcaaattgctttacaagagaccaaagaaataataaaaaagcttaatgaggaaggggtgaagatttcagactattaaatggtaatataccagcagacgtttgttctgtattactgttagattattttaatcagttgaataatgaagagagagaatactgtaaagatattgttctagcgtattctaaacttccgttgaaggaaggctataattatcaggtacaagatggaacaacctcggcaatttcagccttacccgtgatttatcataattatccaatggaaagggagactataaaaacaatattactttgacactgtttaatgaccactctattggaatggcaggtgggcgctactcagtatttcctagtatggtgattcataaattatggctagactattttgatgatatgcagtccctattgtttggtttttttgattttaaagccaaaatatgtaatcctttcaagaaaaatcattcatgaaagttatcgtcaagtagactatgacattaaaaaaataaatattaataaggtgtttttaaataactataagcattgcatatcaaatgtcatcgataataaaatatctatagatgatttgggaagtatggataaagttgatctacatattttgaacacagctttccaattaattccagttgatactgttaatattgaacataagaaattggtttccttaattgttaaaagattttctacaagcctattgtcaagtgttcgagaagatagagttgattacgctcttcggcagtctttcttggaaagatttgcctactttacgcttcatgcgcccgtgagcgatattcccgattatataaaacctttttcttgatggtttcaacggttcagagcctatttcagagttatttaaaaaatttattctcgtcgaagatagattaaatacttacgccaaattttggaaggtttgggatttgtttttttgataaagtggttactttgtgcaaggatggagataggtattggtatgtagataaaattataaaaagttacctttttgctgaatctccatggaaagaaaactctaatggttggcacacatttaaagatagcaatagtcaattcttttgcgatgtatctaggactatgggccattgcccttcaactttatattctcttgccaaatctttgaataacattgccagttgctatcttaatcaaggtataacttggctttcagaaatattgtcggttaataaaaagctatgggaaaagaaattggaaaatgatactgtttattatttggaatgtttggttaggcggtatattaacaatgagcgtgagcgaattagacgaaccaaacagttgaaacaagaggtcttagtaatattggatttttttggtagagaaaggatcggttgttggttatatgtcacgggaaaatattctgtgatgtagttgaaaataataattttaatgagagcttttccaatttaggctccagggattggagcctttttattatcg (SEQ ID NO: 48)
52	pLG054	accttcttcgctaactgatggctaatgaggccgtaataaaacttaccttacctgtaaatacttttactactcattcagatcagaatgaagaggtttattttatttcattgaaaattaataaataaaaatattggcacggtatgtgcttatacagaatgccattttactaacaaggaatttaccgatgtcggaattaaaaaaatttcaggtacaaacagcacgtgcattgccggtgattgtgttggcggataccagtgggagtatgtcaacagatggcaagattgatgcacttaatctggggctcagggaaatgcttgatagttttaaacaagagagccgcctgcgcgctgaaattcaggtcagcgttattacgtttggtggtcaccaggctgaagttagcttgccattgacgcctgctcaccagttgcaaagtattacctccctggaggcaaatggcatgactccactgggtggcgcactatcgctggcctgcgagattattgaaaatccaacgcgaaaatttcagccgattatcgtgcttatctccgatggctaccctaacgacgactgggaagccccttttgctcgcctgattcacggtg

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		aacttactgccaaggcctcccgttttgccatggctatcggtgcagatgccgatgaatcaatgctcaacgaatttgcaaatgatcctgaggctcctctcttccacgcagaaaacgcgcgtgacattcgccgtttttcagagcggtaagcatgagcgtcagcgcacgaagccgttccgcaaccccgaatcagtctacaccgttgcagatcccgagtgctgatgatcaggactgggagttctgatgcgcctgtacgcttctggcacctcggtacgtggtcccgcacaccaacaggatgatgaacccaatcaggatgctgtagggatttacggtctgcgtggtggctggtgtattgccgttgctgacgggttgggtagccgatcaaaaagtcatttgggttcccgtaaggcagtcaatctgctgcggcagatcatgcgcggtgcggagatgctggtcgctgccgaagtgactccagcgttacgtgaagcttggctaaaccactttggtactgactatcacgattacgaaactacctgtttgtgggcctgtgtcgaggcgtcgggccatggcgtgatcggacaggtaggcgatggcctgctgctggtcagaagtgctggggtgttcaacgtaatgagcacaccacgacggggttacagcaatcacactgagactctggcacagcgtgcacatttagatagttgcagtgccagagtggcattaacccaacccggagatggcgtactgatgatgaccgacggtatcgctgatgaccttatcccggatcagctggagtcattctttaatgctatctaccaacggatacggcaatgcagcaagcgtcgtacacgtcgctggttaacacaggaacttaacggctggtcgactccaaatcatggtgacgacaagagcctcgctggaattttcaggatggactgaccacatgacatcaatagtaaaaacgcaaccaaaacgcgtggtgaaggataccaggggatcaagttacgagctgacagaggtaattaaccgtggtggacaaggcattgtttaccggacgacctatccgcaaaccctggtgaaaggttttactaatcaggacccacaggaacgccagcgctggcgcaaccatattacatggctgctcagccaggatcttagcgacctcaaacttgcacgtccattaatacttctggcggagcctcgctttggttacgtaatggagctgatggatggcctggttccattggatagcctgttgaacagctttataaacgcaggggaggagtctctggcggattatctgcgtcagggaggactccgtcggcggattcgtatcctttgccagctggcacgcacactcaatcagcttcacgcacgcggcatgttgtatggtgatctctcccccagcaatatttttgtttcagacgatccaagacacgcggagacctggcttatcgactgcgataacatcagcctgacagcccatcacaatctgactctgcataccgtggactatggtgctcccgaagtggtcaggggagaatcgttactgtccagcctgaccgatgtatggagcttcgccgtcattgcctggcaactgctgactcataaccatccgtttaaaggggaactggtcagtaatggtcctcctgagatggaagaagctgccatgcgcggtgaatacccgtggatcaatgacgcacaggatgacgcgaatcactgcttcgtcaatctgccaccggagctgattgcacatagtgcactgccaactctcttcgctcgctgctttgaacagggaaggtttgaacctcatgagcgtccgggtatggctgaatggcttgaggcgctgagtgctgtggatgagcgtctgtttacctgtgacagctgtgggggaagcacgctcctggcagaggaagcagaaagcgcgaacgatgccgtttgcttttactgtgacagtcccgccgaccgcctcctggtccggtttagtgaatatgtgactgagcaacaagacggctcgaatccagacaccaaaaccttgattgccacagggcgaaatgtatggctgcagccaggtcaccgtgttgagttaaagcgcctgttgccaagttttatctatgaccactggccatcagatcatctgcagattgattacaccgcccgcgggattgggatccatccgttgcttggcggagagctatacctacaacgcggtgaaactatcaaaccactgcgggggtttcagggactcaaaaacgagctgcgcggaacaggtggggagccttggcagatccatatcggcgatcctggccagtcgcatgtaatctggcagttcacgtggtgacaatatatgaaaattaacgaatttccactgatgtccaaagatattctgctgctggaaacggataaaggaaccaccgggttccggccaaagcaagctatcacctttcaggcgtatggtgagaattggctggcggtacaggggatcattgcgtaagtgtccagtgctcccctggtgatcacgaactctttagccgtctggtgatgagggatcaggttcgttggttgctgaccagtaaagcggaaaaacagttgcgggttcaatattgcacgcctgttgaagtcacaccaatgcagctcgagttgggaattgatgagcgaattgcggaagaccttttcgcgaaaaaacagatcaataacaacgatattgagcttgcctgccgctggtttgaagagacttttattgtccatagcgagtcagaaagtgactggttaacggttggccgttttagcaatcatgcagccaaaggtggttttcagctattgggaaacggctggcgtgcggatgttgagcgcaacccggaccacggctttcttatcagacgtattactggtcatttaagccatgatacaggcttctcgttgctggttggacacttcgccttccgggatatgtcagttgctgcggtgctgaatagtgcaacccagcaggcaatgctcgatgccgcactgcgagacagtgccagctaccttgagctctggaatctctacaacgataaagagtggcagagcgagttgaaaaaggccgaaacgctgggtgttctgcgctttgttgcgtgcgagggcaccgaagctggccgggaaaatgtctggcatctgactccccgaactcctgaagaatacagagaatttcgccagcgctggcgcgcgctcgatctgcccgcaggcactcaggttgacctgggcgctgaaactcccgactgggcagaagaactcagtaccgaagaggatacggtactgaaaacgccgcgcgggaagatcgagttcgctgatgaatatgtggtctttacttcagcctcgaatcgccgagacgtgcgccccgcaaagcctgaaggatggctctacctctcgttggcaggatatcgcacagtcggcaaacgtcgcctggcggcaaaacgtgccattgattccggtaaacgcatgccacagttgaagtggctgctggaaggggtcgttgttcctgctgctcggcgtcgcaacatccaggggatgacaccctacgcccgcgaaatctttaagggtggcaaaccaacgggcaaccaggaactggctgtgtttaccgctctgaacacacccgacattgctatcgtaattggcccgcccggaacagggaaaacccaggtgatcgctgcgctacagcgacgtctggcggaagaggcccaggaaaagaatattgctgctcaggttttaatcagcagttttcagcatgatgccgtcgataacgcgctggaccgcagtgacgttttcggtctgcctgcatcacgtgtgggcgggcgtcgtgcttcagtagaagacgagtcaccactggatccctggttgtctcgccacgccagtcatctgcaggagaaaattgctgaccagtatcaacgctacccggagttgaaaacaattgccgacctcacttcccggcttgccctgcagcgattggcaaacgacctgcctcaacaacgggcagaggctttttcgcatatttatcaggacgtcaattccctggcagagaaagggctggtcacggactcccggcttgagatacgtctgcaggactatattaagcatctgaaacaggatggtgttgctgaggtcagtacggtgatgaatgtagcagtattgcgccgcattcgcgcgttacggaccactcagactgctttctcagatgatggtgccgatcgtgcctgggatttgctgcgatggttgaagcggaatgttcctgacatcgacgctgagctgacctcggtattggaaatagctgccgatgccagagaagttcctgtggcactcgtcgagtgccagcaacagctgctggagcgttttctgcccgattatcgacctccggccctcaaaaataagatcgatgatgaaggactggctctactgaatgacctcgacaagcatctttccgacttgatgcatcggcgtaagcagggtgtggcatgggtgcttgaacaaatggccgatacgctggagatggaccgccgtgccgcacaggaggtggtggatgaatacgccatggtggtgggagcgacctgccagcaggccgccgggcaacagatggccagcctcaagtcggtttcaggagtcaagagcagtgacattgagttcgataccgtagtcgttgacgaggctgcacgcgccaaccctcttgacctgtttgtgcctatgtcgatggccacgcggagaattattctggtcggcgacgaccgccagcttccgcatatgctggaaccggatattgaaggccagttacaggaggagcatcagcttacggcactgcaactggctgcctttcgttcaagtctttttgagcgcatgaggctaaagctactggacctgcaaaagaaagataatttacagagggttgtgatgcttgataagcagttccgcatgcatccactgctgggagatttcatcagccagcagttttatgaaaaagaagggctggggagagtggaaccaggccgtagcgcagaggaatttgtctttgacgaaggtttcctgagagcgctggggccactggcgtcggcctatcgtgacaaggtctgccagtggatcgacctgcccgcttctgctgggctggcagaaaaatcaggaaccagccgtatccgcaccattgaagcggagcgtattgctcaagaggtggcacagttactgaaagccggaggagaaaccctctctgttggggtaattactttctatgccgcacaacgagaactgattatggaaaagttatccgaaatcaggctggaaggcgtgccactgatggaaaaacgtaacggaacctatgaaccgcatgaaaactttcgctgggtgcgcaagtaccgtgctgacggttcgttcagccaggaagagcggttacgagtaggttcggtggatgccttccagggtaaagagttcgatgttgtactgctatcctgcgtgcgcacctggcgtcagccgaggtcctcatctgccgccgatgatgcagctgccagggaacaaatgcttaatgaactgttcggtttcctgcgtctgcctaaccgcatgaacgtcgccatgagccgacaacgacagatgctgctttgcttcggcgatgcagcactggccaccgctcccgaagccctggaagccgcgccagcactggcagcatttcataccttatgcggaggcgttcatggcactcttcgctgaaacaggtatttatattcaatctgccccacggccgcagggtgaagcgcgcccgatactctggccagtcaggatacatagggtgctctacccggaaagctatcaggctcagatcaatgtcttccaacgcgcaattctcggattggtacgagcgcgcgtcgtacgtccgaccgaactggcagaactgaccggtctgcaccctaaacttattacgcttatcctggcacaaagcgtcagtaatggctggcttgagtccggtgaagataccctcacttcagcgggtcagcggttgctggatgatgaggatgacggtattggcaaacaaaaatcaggctatgtattgcaggatgctgtaagcggaaagttctggccgcgtctggtcagcacattgaagcaaatcgaaccggtcaatcctctggataaatatccgcaatttatactgaccaggaaaacaggagcgacactgcgacctttcctg

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		atgaatgccagccgatcgccactgccgcctctggaacgcaaagaactgaagcgtgcctggcgtgactatcgtgacgactatcgtgccagtcagcaactgggcgtcagccgtttgccgccacacattaacctgcacggtctgcagcagctagaggaaccaccgcagtgcgcacgaatactggtgtggatcaccactgatcgagagagtggacagctatggagtgccgcggacccatttgctctgcgcagtaacgcatggtggctggacctgccttcaatcgtggaaagtgactcccggttgcaaaagatactggaaccgctggttgtggtgccacgcgccgcagaacaaacctaccagcagtggcttgaggctatcgcgcacgaaactgattttaagatgatgagtcaatacccttgggccgaacgtttaccggatgtgaaacgttatttggtggcgctattggtacatagagggaggatcgagcagggtgataacggtcaaagtgagctggatgccgcactgaacgagtgccagaagctgctggaggttgttatgcagtggctgattcgtcgtcatccagccaacgcggaattattacccaagggccgcctggataaaattaatacggccaacttgctcaaggatatgaaaataccagcatttaccccatcagttattgatggcctatctggccagataatacgtcaggtgcgctacgcatgtagcaacccatccggctcattgaaggcactactttttgcagcggctgtcggtgcgaaccaggatccacagcacccatttggtcactggatgactcagcgttacaactgccaatgctgctgcaactggcggatcgtcgcaacaagagtagtcatggacagagtaaatatcttgataagccggtacaggaactcactcagcagatggttgaggaaagtatcagttatgcattgagttttaccgaacgttttaaggaatggatgtaatgtcaaaacgagcacaacagaagtatacctcacctattcccaagcagagaaatggctctgctgcggcatctgccatcaccacacttcagaggtctgcaatgacaaccgagtcgcagattattgccgcagcccatcacacagctcagagtgaaaagcttccaaaagatatcgattttgatgtgacatggctggaacgtatcagtcaacgtcttcagcaggaaggagatgatcaatttgtctcctggcttcagacatttactcttttctgccagaaactggcgcaaagggatgaagagacgcaagcagcagcacagcgtattcaacagctggagctgacgctggaggagcaaagcgaaaagttagaacaggaccgtgttgaacatgacattcaagctcgggaactggcggaaaagaaagccgggatcgtgagcaaagaacgagagctgaatgaacgtgagctcaacgccaaagcgggcttcagcgagcagaatgcagcatcgctgcgaaacctgacccagaggcagcagttactcgaccagcagcatcaggaggatattcaacagctcatcacacaaaagcaggggttaatgcgggaaatatcgcaggccattgtccagttgacccagttacaaatccagcaaagcgacgcggaggcacagcgcagcttgtcactggaccagcgcgaagaagacatcatcaggaaagaggaggatctgaagcgcgccagccgtcgtctggaacgagacgagcggtctgtagaggcggagagacaggcgctgaacgaatgtttggctgaagcaatgcaaacagaacgccttgagtttgaaaagaagctggatcagaaagagcgtcagttcgacaaagctcaggaacgggtgcaaaacctcagtgaacgcctcatggaatgggaggaacttgatcaggcgctcaatggccaatccgcttcgcaaatgctgaatgagctggataagttacgcgatgaaaaccgcgaacttaaaagtcagttcgcgcacactaacctagcagagctggagcgcgagaacaaatctctggccaacagcaaaagcgctcttaaaaatcagctggaaaatctgcttgcagagatggacaagctacaacgcgaggtggatcttcagcgagtggctgcgacccagcttgagacagtggcacgggagaagcggcttcttgagcagcagaaacatctgcttggtcaccagattgatgagattgaagctcgtattggcaagctgaccgatgccagcaaaacccagacgccgttccctgccatgtcacaaatggacgagaagaatgggctcaacgcaaaacgtgatcatcgagaggtcggtgacctgaaaaatttgccagtgagcttcagcagcgtattgctcaggcggaagagagcgtgcagctattctatccactggaaagtatccagctgctgcttggtggtctggcgatgagccaactgcacctgttccaagggatcagcgggaccggaaaaaccagcctcgccaaggcctttgcaaaagcgatgggggatttgtaccgatatttcggtgcaggctggctggcgtgaccgcgacgatcttctaggccactataatgccttcgagcggcgctattacgagaaagactgccttcaggcactctaccgtgctcaaacaccgtactggcaggacacctgtaatgtcattcttctcgatgagatgaatctttctcgaccggagcagtattttgctgagtttctctcggccctggagaagaacagccacgctgatcgaaaaattgcccttaccgaaacagctttactcaatgccccggaacggctcgttgaaggacgccatattctggtaccaggtaacctgtggtttattggcaccgccaaccatgatgaaaccacaaatgagctggccgacaaaacctacgatcgtgcccatgtgatgacactaccgaagcacgacactcgctttcctgtcagggagatggagaaaaccagctattcgtggcggtcactgcatgaagcctttgctaaagcaaaaacgcaacatgcggaaacggtcaggaacatgctggagcaactgtccggtcatgaatttactcacctgctggaaacagattttggcatcggctggggcaaccgttttgacaagcaggcgatggatttcatcccggtgacgatggcctccggggcagaagctgggcgcgcgctcgatcatctgctggcgacccgtattatgcgctcaggtaaggttaccgggcgctataatattggcttggaatcggtcacacgactcaaagaagaacttgaatttttctggattcaggtcggtctgcaaggcgatccggttgaatctatggcattgctggaggcagatatccgccgtctgtcaggtgcgcgctgatgtggcacgatcgtttaactggtaggcaacatgcacatcttccgcaacggattgatcacgggcgttactcaatcgaggcttcccctctgacgctaaatggacatacaccgaatttttcggattgctggtcagcgacggcggagcaaattgtcggctggacgatacgctgcataacttcattcagcctccgcccggccatgaagaggaaacccggctgctggaggaagccatcaccacgatcggtgccgcagttgatgatgacatcagtgtgctatcgccgctgatgccagcagctattgtcgataatcaaagcctttgctacctttcgaacgtgcactgctggaggtgatacaaaaaggacatttacagcatatatcacagcggccgcggctggatttacgttatgacgatgaggtggccgacgttgcccgcgtgcgtcgtctggcaaagggtgcactggtacatctggcgtcacactccgaatgctggcagcgtcagacactcggcggcgtggtacccaagcagatactggcacagtttagcgaagatgatttcaatatctacgagaatcgggtttatgcgcgattactggataagatcgaacgtcatttgtatcaccggctgcgcactttgagaagcctgcaatctactcttgcccaagcactggacttctatcaatctcaggaggtgaattaccgcctgcgcaatgctatttgtcagttgtgggggatgacttacgatgaggatgcgactgatggcgcatctcggcagctcaacgccacattggcgacgctggagcaaattttccgcatcatttccggtctgcgacaaagcggcctctatctgcgggtaagtcgtactgcgcaagtgacaggtggagttcatatgacgaatattttaagtcacgatcctcactatggtcatttgcctttactatgggcacagttggctgacggggctcagcccgaaaatttgcctcaacaacgcctcagagtgaaccagagcctggcagctgcgtatagcagctatgccgggttggtgttacgccatgcgttgcagccctggttacacggtaagagtgaaggaagctgggctggtcgcactctgcgacttcgccagcaaggcatggaatggctgctgagctgtgattccaatgacagtgccagtgaagagacgctgttgtctctggtgccatttctgaaccaccagcaggtagcggtagacctaccggaaaatcggtatatcgcctggccttgcgtggggcatttacagcaggcattacctgataaagagggctggattcggctttcacctttagatatgtactgtgtagagcgttttggcttactgatagataaaattcttagccgggaattattgcgaaactttgcccgtccggttatccgtattccccggtgcgtattaccacttgctacaaaactgtcttcactgacagttgatcaacagttaaatcagataacactgcatggggatctgactaaagctgagctggaacaattaacctctcatttaatcaacaacaatgctagcacacaggcagaggaaattacgctgcgataccgggaatggcgagcattgcaacagtgccctgtctgcgaccatacaaccgaactggtttatcaatatcccggtggatttaaaaccctctgtaaaaactgcaataccgctcgttatttcagccagcatgaaaatgcacactttttgaacaaaccagaacagtagaaagagaaagtaaaaccttcctggctcaggggcggagagtttttaactttcagttttagcagggtttttacgactcgctgcatttttaaagagttaagaataatgaaacttcagggcatcttttatatatcggtattacgcaaatcagtagtttcggttgcgcgttttgtatacataccggcaagtgtccaatcacagtgaatagccaaaatcgccgggagcacgttcggtcagcctgcggacatggtttttatcacgt (SEQ ID NO: 49)
53	pLG055	ggattcaccattatagtgacatgttcaagatgatgatatatctttgaaaagtgttctctttgcgaacggtatagaatttctagcgttacttttcataattacacttttagggttaggcaggcacaatctatgcgctgtcttagataactacatccatttttactggactaccaccaacaaaaatttagtggtgcaggagaaaacgtgaagtatcagatagtaggtggtgctggcctgcaccgcagcgaaaccaaaacagttgatatgatggttaagcagttaccagatagttggtttggctatgctggcttagttgttactgatagccaagggtcgatggaaatcgatatgctaattattactgctgaccgtctgctattagtcgagcttaaagagtggaatggtaacatcacatttgaaggggggaagtggctgcaaaatggtaagtcacgaggcaaaagtccctatcagatcaagcgtgagcatgcactgcgactaaaagatttgttgcaggaagagttatctcgtaagctgggttactttttgcatgttgaggctcatgtagtgctgtgtggcacagctggtcctgaaaacttgccattaagtgagaggcgctatgttcatacccgtgatgaattcttgactataggtaacccaaaaaattacgaaaagctggtgcaacacact

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		aactttttcatctttttgaaggggaaagcctcgaccaaattctgatgaggcattacctataattaagtccttctttgaaggacc
		aaaagtcaggcctttgccactaaaagaaagcggttatcttgcgaacgataagccattctttagtcaccctcacatggtctacaacg
		aattcagggctacccacaaagacaatagtcaacacagaggtctgctacggcagtggaactttgatgccttgggtgtagcaaacgca
		atgcaaacattgtgggctgagatagctctgcgtgagactcgagtcggtcgcctagttcgtcatggcagcgcaactatgcaggatta
		tatgttgcgtgctgtaagggaactatccgaggaggatataactgatgatgcccgtgagctgtatgagttacgccgtagttttagcc
		gattagatgagattctagatagcgaagctgacggatggagtaaatctgagcgtattgatcgcgttcgtgcattattagctccattc
		tcggaattacatagcttgggtatcagtcattgtgatattgacccgcacaatctatggtacgcaggggatcagaagagcattgtcgt
		tactggctttggcgcagcctcactggagggacataatagcctagagtcattgcgtccgacattgcaaagtgctccatatattttgc
		ccgaagatgcttttgaagaagcagttgagccctatcgcctagatgtattcatgttggctgtaattgcttatcgtatttgttttgca
		ggtgaatcattactgactcctggacagatgcctgaatggagagctccattaactgatccttttagcggtattctaaatagctggtt
		tgagcaagctcttaaccttgagccaagtaaacgctttccacgtgcggacataatgctcaatgagtttaatgcagctactaaggaac
		atagccaagaatttgatgaagctaaccagatttatcaagaattaaagcaaaacaaattctttcgcgaagggatgaacagcgttggt
		gtgttaattgagtttcctccacttcctgaacagttgtctatggtttactctgctcttgctgctattgctacgactggcagcatcag
		ttatcactgtgaacaaggtgggaaagctctgcaggtaaaattgtgggatggtgttattttgacccctcaacaacctggtgttaacc
		gccgtatccacgcttttaagcaacggatcgataagcttacgcatataaatctgccaactcctaaggtgcagtcctatggactatta
		ggacaaggcggcttgtatgtagtgagcgagtatgtggatggcctaccgtggtcacagtttattgctgagaacgtgttagtacaatc
		ccaacgttttacaattgcggaaaagttgatcaacaccattcatgcttttcatgaaaagcagttacctcatggagatctttgcccag
		agaaactgctggtacaagtcggggagcagacagtaattactctgattggattgcttgaattcagtgatgaattaactgcagataat
		cgctaccagccagagaatcccgaaagtactgatgcttttgggcgagattgctttgcagtatatcgtatggtggaggagctatttag
		tgaagatatgccagtactggtgcaggctgagctagaacgcgcaaaacaaaccgttgacggtatacctatcgcgctcgatcctttgc
		tgcagtcaattcgagcaccggaacaagctgagattaatcaagttgtggcgtctgagtcacaggataaggtaattcctgtttgctgg
		ggcacagatgattggccgcaagaagtgaagcttctagaacaaaatgatgggatctattatttcaatgtaactggtcatctaaccc
		acgctttgcgcatgaattgcgttgttacatcactggcctaggagagcggctattgatagacttagatcctgataatcgcactatta
		atagaatagtgtatgaaaaaggattatcgatcgaagaaagtataaaggctggtaaatattcccaggctaaaattaatactcaactt
		tcattacaacgtggctcacttaatcagcgtaatacttttattgaactactgtttaacctcgagccagtaattgatgccatcattga
		gcgagctaatcctaatcaagagatggatgaagatgacttcgatagtagtgagtcaagcccaattgagttatggcaggcattatctg
		atacagaagtagacctacgagatatagtcaacatcgactctactgactttcaggaatcaccgagtggttgcttactctacccatat
		actacggaatccggtgctgacctcagctttgaacttgatgataagatcattgtttatattaaagataagcgtgaatcagtgcaatt
		aggggaattgcagctaagtgagactacgccgagtctattggctattcgctttgattttgatgctgctcgtaagcgaattagtagcg
		gcagccagctacaattggaatcgatccgtgacaaatcatcaagagagttgcgtcaaagagcccttcaacgggtaattgaaaacaaa
		gcagagatccagcatctgccacagtattttgattaccaccagaaccctgcatgcagcaaatgcaaccgcggccatccgcggagac
		attacgcgcactttatgatcagcctggacaacgttttaatgaacagcagctaatggcatttcaacagttggtcgagtttggaccag
		ttggagttctgcagggaccacctggaacaggtaaaacaacatttatttcaaaatttattcactatctgtatcaacattgcggtgtg
		aataacattcttttggtcgggcaatcccatgcctctgttgataatgtagccatcaaggctcgagagctctgccatacgaaaggaat
		ggaactggatacagtacgtattggtaatgaacttatgattgatgagggtatgctaagtgttgcaactaaagctcttcagcgacaga
		ttcagcataaatttcaccgtgaatatgatctgcgagttagctccctaggaaagcgcctagggatggccccattattagtccaacag
		ttatgtcagttacatcgtacgctgaatcccttgatggtgacatatggccaatatagccgtgagctggataaagtagaacaaataaa
		gagtagtagtattagtcatcaagagcgactggctgaattattagaacaaagcaatcagcttaaactgcgaacacaagaaattatta
		actcaatattcgatgacagcttgctgaaaactcttgtctatgatgaaaccttgataagacagttggctgagcaagttgccatacaa
		tacaattataacaatccagagaaccttgaacgttttatgcagctattggaaatgagccaagagtggatggatgtattacgcggcgg
		cgaggctggatttgatcgatttatgttcaaaagtaagcgattggtttgtggaactcttgttggtgttgggaatcgtcgactagaac
		tagctgagtccagctttgattgggtaatagttgatgaggctggccgagcacaagctgctgaattgatggtagcgctgcaatcaggc
		aagcgggtgctgttggtaggggatcataaacaattgccaccattctatcatcaacagcatcttaagttagcctctaagaaattaga
		actcgggaaagggatcttttatgagtctgattttgaacgtgcttttaaagcaacaggcggcgtaacactcgatactcaatatcgaa
		tggtagaaccaattggcgagttagtatcggagtgcttttacgctcaagatatcggtaaactgcattcatcgaggaaagtctcgcca
		gattggtattccaagttaccaatcccttggaacaaaactgttacttggatcgatagttcgagccctaatgaagcaggtgcagaaga
		acataagggtaatggtcgttactataatcaacgagaagtccggctactgctagaggctttgcagtcattgtcgagtgatggctgca
		ttgcacagcttgagcaaactattaccacagaacagccatatcctattggtataatcacaatgtatcgtcagcaaaaagaggaaatt
		gacaatgctatcagtcgggctgaatgggctgcatcgttacgtggtttgatcaagatcgataccgttgattcatatcagggccagga
		aaacaagataattatcctcagtctggttcgcgataatcccaacaaactacaaggtttcctgcgcgacgcgccgcgaataaacgttg
		ctatttcgcgagctcaagaaaggttattgattctgggagcaaggcgtatgtggtcaaagaccaataatgattcagcacttggaaac
		gttcatgaatttattagtaaacaggttgcagtagatgaacccaactaccaaatcctgtgtggtcaaagtctgcttggagataacaa
		ctaatgtcagaaccacgtctgggtaatctgattaccgttttactacctgcgcgtagttacaagatcaactgcgctttgaccactga
		aaaactgatgcctggaattgaacagtttgcatgtcgcttgctgctgatttttgatcaactctatcccagcgagttacagaattact
		ttggtctaactgatcgtgagcgagaggtattgcttgatgggttgctggctaacagactgatcaacattaatcctgatgggcatatt
		gaggctagctcattcctacgtaagcatgcagctaataatggtgggaagccaagtttagttaaatatcaagaatgtacggaggaagt
		tgcattcgatctactaactctttcgatatgtaaaccgcaaccaaatcgtcgttttacttctggactgccagagctattgccgcggc
		atcagatcgggggagatgctgctgcggtaacagaggcttttagttcccagtttcggcaccatcttttgctcagccgcaacagcgag
		tatgagcgtcaacggactaaattatataagataatgggctgtagttcgcatgagatggtgcagctcccaatagagatagaggttag
		ctacggtgtttctgctgggagcattgagccgcagaaatttactcgttcctatgaatatttaggtaacacccggctgccgctttcaa
		acgagctggaagctcatatcgcagatttttgggagaacataaactagatgaattcggtatcgactgtgaagatttctgtaaacta
		gcaaatgataaagtgttgttacaatttgctaatggttataagttcaactattccggctggatagaggctcgtgaacaacgtaaaac
		tggctacggtacttcattgactaccggcatgttaggggctgtttatttgccgcacaattctaagctgttcattagtatgttgcata
		atgcattacgtgattatataggtaaaacagctccaaaagcgctgtggtatagcagtaaagtaccactgtggggagctaatggtagt
		caactttcgcgttttactcgcgctctaggcgatatacttggcaattatgccgatgataagattgctcgcatttcgcttttacactc
		aagtgcagatgaaggtgaaaaacgtcaagagcgtaagcggcacttaggtcgttttcctaccggtattggccttacttcagaggcta
		aatttgatcgtttggagatcctcttaattcctgatgtgattgctttggtgcaataccacggtcaacctaattctgatagtgcatta
		accctgccgattggttatataactgttgagccagagcgtttagaattacttaaaaaactaatgattaagcgaactgaaggggctgt
		tgcaaccattacttggtctgaatcaaaatttgaaaatttagcttcgctattacctgttgagtttctgattaaactgaataagaaaa
		gcggtgaagatgtggatgctgcaataaaaaaaatgcagatctataaccgtgctgaaaccgcacgggcaattttatcgctacgcaag

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		tagcatttatattgcaacgaataaatttttctaggttgctatgaactagctaaagggcaacaaatagataaacggcgttattcatgtcaaatgagataatgttaaattgatagggatttataccccgccggccattttgaatggtcggagttgttataaacgtta (SEQ ID NO: 50)
54	pLG056	cgtgatgaatgaagcggctaaatacattaatgataattataatttaattcattaaaatcagtaatatataaatataaaagttgtgaaatgtgatattcgtcaaagcatgtcaaaaagttttgactgttctttaggcatcattcgcaattgtctaacaacttgataggataggaacaatctcaaaaaggaaaatgacatatggcatacgaagctcaaatcagccgtactaatccagcagcatttcttttcgtcgtcgatcagtcaggttcaatgtccgacaaaatgtcttccggccgaagcaaggctgagtttgtcgccgatgctcttaatcgaactttaatgaacctaatcactcgctgcactaagtctgaaggcgtacgtgattatttcgaaattggtgttttgggttatggcggtcaaggggtttctaatggtttctctggttcactgggaggacaagtcctcaatccaatttctgctctcgaacagaatccagccagagtagaagatcgcaaacggaagatggatgatggagctggcggaatcatcgagacagcaattaagtttccagtatggttcgatcctattgctagtggcggcacgcctatgcgtgaagccctgaccagagccgccgaagagttggtgacttggtgtgatgcccatccggattgctatcctccgactatcctgcatgtgactgacggcgaatcaaacgacggtgacccggaagagattgccaatcatctacgacaaattcgcaccaatgacggtgaagttctgattcttaatatccatgtcagttctctcggaaatgatccaatcagattcccctcctcagacactggcttaccggatgcctacgctaaactgcttttccgtatgtccagccctcttccggaacatctggtgcgtttcgcgcaggaaaaaggtcatacggtcggtatagaatctcgtggattcatgttcaacgctgaggctgccgaactcgtcgatttcttcgacatcggaacccgcgcttctcagttgcgttgattcagcaatgaaactggagttcttagggacagttccgaaagatcctgaataccctaaggcgaatgaagataaatttgccttctccgaagatgggagaaggctggcgctatgtgatggcgcgagtgagtccttcaactcaaagttatgggccgatcttcttgctcgtaaatttactgcagatccgaaagtaaatcctgaatgggtagcatctgctttagcggaatattctgccacgcatgacttcccttctatgtcctggtcccagcaagcggcattcgaaagaggcagttttgcgacactaataggtgtagaggaatttgaagagcatcaggcggtagagattcttgctattggagatagcatcaccatgctggttgattgcgggaaactcatttgcgcatggcctttcgataatccagaaaaatttaatgagcggccaacactgcttgctacgctgtacgctcataacaatttcgtcggtggaagcactttctggacacggcatgggaaaacttttaccttgaaaaactcacccaacccaaactcctctgtatgacagatgcgctcggcgaatgggcactgaaacaagcgctggcagaggattctggtttatcgaattactttcgctgcaaactgaagaagagcttgcagagttagttctgagagagcgtgcagcaaaacgtatgcatatcgacgactcaacgctgcttgtactatcgttttaacgcggaaagtaaagatgccttacccatctcttgaacaatacaaccaagcgtttcagctacatagtaagctgctaatcgatcctgaattgaaatctggtaccgttgccacgacagggttgggtctcccccctagccatcagcggtggctttgcactgacctatacaatcaaatcaggcgctaagaaatacgccgttcgttgctttcatagagagtcaaaagccttagaacgccgttatgaggctatatccaggaagatttcaagccttcgctctccctactttctcgatttccagtttcagccccaaggggtcaaagtcgaaggaatatcatacctatcgtcaaaatggcatgggccaagggagagacgctaggagaattccttgaggtcaacaggcgttctgcacaagcaatagcgaaactatctgcatcgattgaatcacttgccgcctaccttgaaaaagaaaaaattgcacatggtgatttccagactggaaacctgatggtctccgacggaggtgcaaccgtccagttaatcgactatgacggcatgttcgttgatgagattaagacattaggaagctcggagttggggcatgtcaattttcagcatccccgtcgtaaagcaacgaatccgttcaatcacactctggatcgtttctcactaatttcactctggctggctcttaaagccttgcaaatcgatccgtccatttgggataaatcaaattcggaactggatgcaatcatttttcgagctaatgactttgtagaccccggttcatcttccatcttagggatgctatcgggaattcaacagctttccacccatgtaaagaattttgccgcagtctgcgcttcagcgatggaaaaaacgccttccctcggtgacttcattgcaagtaaaaacattcccatatcgctagcttcgatcagtatgaatggggatattccagtcagcaggctgaaacccggttatatcggtgcctacaccgtcctgtcagccttggattacagtgcttgccttcagcgagttggtgataaagttgaagttatcggaaagattattgacgtcaaactcaataagacccgaaatggcaaaccatatatctttgttaatttcggagattggcgcggtaatatctttaaaatatcaatatggagtgaaggcattagcgctttaccttcaaaacccgatgcctcatggatagggaaatggattagtgtaatcggccttatggaaccgccttacgttagcgggaaatacaaatattcacatatctcaattacagtaacgactatcggtcaaatgaccgttctttcagaaccagatgcccgctggcgtcttgctgggccaaacgaaagtcgacaaacattaacttctactagcagtaatcaggaagccttggagcgcattaagagtaagagcaccacttcaactcctatgcccatgaacactaacgccacaactgcaaatcaggcaatccttaacaagttacgggcttctacgcaaactgtagcggcagcaagagcgcaaactcagcatgtagtacctaataaatcatcaacgcattatgtggcaccgacgggaacatcagcttcgcagccagttcaaaatattccgagccctgctagtacctcaaagcagcaaacctctcaaaaaaatatagttacaaagattttgaaatggctttttggatgattggtacttgtaaagaacaagcgcaatttcagtggccgtatcacttgcgcttgaggtgcctgcgggtatgatcttgcgacatacaccactaaaacgaattcgtggcggcacttttagcctgcccctgtgttttcccgaggatttac (SEQ ID NO: 51)
55	pLG057	ggggcgaaaaggggaatgccggtcattgccggacgagtgcaccttaaaatgtgcggcagggggcgcccgcgggctgatccatttggcagaatggccgtgcatgcgacgatcgagcgcgggagacggctgaccctgatggacaaacgcgctttgagcgagcgggacatctgcactaagttcatcacgtcgtggcttgacagatgtttgccttgaccggtcgaatagccccattcggggccgtgtactttgcaaatgggccgaggtgcccgaaaaaccggtctggagccaggacaagaattacagtgcgcgaaccccaccggttactcacagcccgcttattggagttgatcgaaacccatcccgaaggattgcgactcgacgaggttcaggcgcgtacgcgtgttgaagggtgtcgcgcgggagtcgatgatctcgcagcagcgctactcgatctccagcaccaaggtcttgcacatataaacgcagcccggcgctggtttccgaagcgggcggcgagtgtacgaccatcctccgcagtcactggttcggatgacgtggcgggtgcagggctggtgctgcaggcgctaccggcgcgcatcactggcaacgatatggcggtagcaccagcacctgcattgagtgctaccggcacctcgctcaagccgacttggggcctgttacgcagcctgctgccgtattacgccgaggcgctagcccgcaatgaacgggcgttgctactcggaacgcctgagcgctacggcgagcagttcctgctcgtggcaccacgcggccgatggtggccagcagcagggttaggctacgggctagaactctcgcgtacgcatctgccggttgcttttctcaccgcgttagcccgacgcacgcgcgaaccgattcatgtagcctaccccatcgcgctggtgcggccccgcgacgccgcgcgcagcccctttctgttaccagtggcaactgtggcagcggactggaccctcgacgccgagaaactgcgcctgaatctgccggcccaaacgccggcgatcgaatggtcgtgggtgcgcggacagcgccagcgcggacgccagattcgcgagttgctcgatgcacttgatgtcaatgctgacgacgaagtctggcgggcaggctccttcgtcgactgggcgaccttcgtcgatcgtctcgctgcaaccacccctaccgaggtgcgcacaccgctcgatctcgctcagcccaacaatgagttggattgtggccaggcgggcggtatttacggggcgttggggctgttcctgtcgagcgaattgcagttcgcgcgcggggcggtgcgtgatctcaagtccatgacgcagtggtcagatgacgagctggccacaacggcgctggctgcgtgcttcagcgatgccatccacaaggcaccgaatccggtcatcgttccggtgctggagccgcttgtgcttggcgaggatcagcttgcggccgtgcgtgccgggctaaacgatcggctgaccgtggtaaccgggccgcccgggaccggcaagtcacaggtcgccgttgccctgatggctagcgcagcgcttgtcggtcgcagcgtcctgtttgccagccgcaatcatcaggcgatcgacgcagtcgtcgggcggctggccgaagtagttgaagaccggccgctggtaatccgtgccaatgcgcgcgaaagcgatgacagcttcgactttacccgtgcgatcgaagccatcctcgcgcggcccggtggtgagaggcccggcgaagggctggctggctcgatcgaagtgctgacgcggctcgatgcggcacggaccgctgcgatcgaacaggccgccactgctaaccaagcgatcaacgaactcgggcggctggaagcagcgatcggagatctgacggcagcccttggcatcgacgcagccgctccactaccgcgggatctgcccgctgccacacgacccttgcatagttggctagagcgcctgtttgcgccttgggtacggtaccggcgactacaacggctacggcgtctagcgctgggatggggccagcttggttttggcg

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		agtgcgacgaatcgacgctggagctacacgaacaacgtctactcgacctgcaggagctggctgcgctgcgggtcgagcgggatcaggcagaggcagccgtgcgtcaactccgttcaaccggcgatccgatcgcgctcggagagcggctgtgcgcttcatccaaattgcgtctgcaggggctcgccgaactgcttatcgagtgtgcgcctgaagatcgccgtgcgttgaccgcgttgcgcggcgatctggctctggcgcgcggtgatggcgccgccggtgctgcccgtgctcgggaactctggtcggctcagcgagccctgatcctcggccagatgccgctatgggccgtgtcaaacctcggcgcagccagccgcattccgctggtacccgggttgttcgattatgtggtgcttgacgaggcatcgcagtgtgatatcgcttcggctttgccgctgctggcccgggctcggcaggcgatcgtgattggtgatcccgcgcagcttacgcatatctcccaagtgcgccgggagtgggaagccgaaaccctgcgcaatgccggcttgatgaggcctggcatcggcagctatttgttctcgaccaacagtttgttccatcttgctgctgctgccgccggcgaccatcacctgctgcgcgatcacttccgctgccatgaagatattgccgactacattagtgccacattctacggcaatcgcctgcggccattgaccgacccgcgtagcctgcgggcaccagtcggacaggcagccggttttcactggacgaccgcgcccggtccgatccaaccagcccgcaccggctgctttgcaccagccgagatcgaagccatcgtgcacgaattgcattggttgctgggtgagggcggcttcactggaagcattggcgtagtcacatcgtttcgcgaacaggccaaccgtctacgcgaccgcatcgagcattgtttgagtgccgaggcgattgcaagcgcacgattggaggttcacaccgctcacggcttccagggcgatgcgcgcgatgtgattctactcagtttatgtatcggtccggatatgccggctggggcgcgagccttcctgcacgacacgggaaatctcgttaatgttgcggtgagccgtgcccgcgccgtttgccatatcttcggcaacctggagtatggagctcactgcggtatccggtatgtcgaggcactgctggcacggcgccatcgaacaggcgatgccactgccagtttcgaatcccccctgggaagaaaagctctggcgcgccttggctgagcgcggtatcgagacaacaccacaatacccgattgccggtcgccggcttgatctggcattgctgaccgacagtgtgcgtctcgatattgaggtcgatggcgaccgttttcatcgcgacctcgacggtcggcgcaaggtgggtgatctatggcgagatcatcaattgcaggcgctcggctggcgggtcgtgcgcttctgggtttacgaactgcgggagaacatggatggttgcgtcgaacgcatccttgtccacatccgaagcaccgattactgagcatcaccgttccccaccagcagcagccgtgccaccagcgaattggcggcgaatgcaactcgtgctcgggctggccggggctctggcgctggctagcctcgtcactgtattggtgggtgtaatcggcgacgccaccgaacgcgagagttggcgagtacggcgtagcgagcatcaggaggtgctgggcgcgctcagcaccgcacgtgcccagcttgatgaggaagtcgccaacctacgccgtaatcgtgctgcgctcgatgcagacctgaatcgtctccggaccagcgccgaagctgagcagggcggcgcagcacggctgcgtgaggaagtcgccgcactacgccaggagctcgccgccggccgcgccgagttggctgtggctacgcagcggcgcgacaccctgcaggctgcagtgaagacggccgatacgacgctggcggaactgaacgcgcgccgcgatgaggccgagcgtcagaccggtgaggcagcagaacgccggcgggtcgcggccgaagccgagcgggccgcgaaggcccagcagagcaaggccgaacaagcccgcgacagtgcggttgcacagcagaaggaggctgagcggcgcatcgagcagatccttcaggacctgaaaaccgccgaagaacgagtaggtggactgcgcacgcaagaggctcaactaaaagcggctacaactgcctccactgccgaacgtgaccggctggatgctgaagccaagcggctcggactggagcttgtcaagctcgatcagcagcgccagcagcttgagcgcgatacccgtactaccgccgaaactcgacggacggccgaggggctccagcagcagctcgaccaagcgaaccgggatctcggtaccgtccgcgaagccctgaagaccgcgcaggggcagctagccgaaacgcgcggccagcagacccaactcgccgacgaactggcccggctgcgcgcacagaaaaccggcctggatggcgtgatcaccgcggctgctaacgctcaagcggaacttgacaaactgcaggctcagcagaaacgggcggagcaagcagcagaaacgacgcgtctcgatgttcgtcagctcgaatctcggaaaacggcactggaagccgacatcatcaaattcaccgccagcggcaaggatttggaaaagttccgtgccgaactggctgataccaatgcagaactcgaacgtctgcgtcagcaattggttgaggcacggagccggcgcgagactatcgcgattgaagtggaacgcctaacgcaacagcgcggcgaactggagcgcaccatcggttcactaacgccgcgagcgcaggaggccgaagcgctacggatccggctccagcaagacaacggcactttgctcgccctgcgcgagcagattgaacgcttgcgcactgaacgtgacagcttgcagcagccggtcacatcttccatgcatgtccccggcgacaacgccgcggcacgctgatcaaggatcgcgctgatggacacgaacaccctggtctggcttgcatcgggtggcacgcttgccggcatcgtcagtgttatcaccgcattggtgtgcggcatgcactacggtgcggcgctacgccgcataccggctgcggcctttttggaagatatcgtcgcacgcgtcgcaactcgtcgcgaggaactcgaacggctggatgcccaattgggcgagcgccacaacggcctccagggcctgcggggcgaaacggagatgctgacggcccgccgggatgccttggcagcgcaactgcgcgaactgcaggaggacctggttgcactcgatgggcgccgggccgacatcgcttcggtgcgcgatgagttggcggaagcacggacgcaacttgccatgctcgtcagtgaactgaccgaacggcggacgcagcaggagcaactcgaacgcgcggccgaacgtgcccgtgcacaactgtccctgctcgaagaacgccggagcgagatcgaggcaatcgatacagccgagcgcgaagcacggatacggctcaccgaggcgcagacggaactgggcaccgtcgtccaggcgcgggaagcggcacggcgtgaagccgaggcggcagcgcgcgacagggagatgctggcaacgaacatcgaccggctcaccgatgagcgcaacgaactgcgcgctgacatcgccagtctccaagccgaacgcaatccgctgtcgactgaagttcagggcctgcgccggcacttggagcagttgcatcttcagcagcaggcactcgacggcgatcttcaacgcctgcaatccctacagccggtactggaagataaaatcagcggcctgcaacaggaagttgttacccggaccgctgaactcaaagaccttcaggccgaacgtgatccgctgtcgactgaagttcagggtctgcgccggcacttggagcagttgcaccttcagcggcagacactcgacggcgatcttcaacgcctgcaatccctacagccggtactggaagacaaaatcagcggcctgcaacaggaagttgttacccggaccgctgagctcaaagaccttcaggccgaacgtgatccgctggcagcggacattgatggcctgcgtcggcaactcgaaccgctgcgtacacagtgcgacgaagtcgaagcggaactcgcccgccgccgcgccgaactcgccgcgatcgagcaggagatccgtaccaaaggcggtggtagcgtcggcaacccggaagacgtgctcgccgatctcgaacaggcaccggcttgtctggtcggcgacggcggcaggggaccgttgatgccgaatccgcagcgcgacgacgacgaaacagcaatgctcggccgcgtgcggacacaccttgatcggctccgtctgcactttcccgagcgcactctttatgcttttcatactgcgctcaagacggcaacgattagtccgcttacagtgctggccggcatttccggtaccggcaagagtcagctgccgcgccgctatgccgaagcaatgggtatccatttcttgaaactgccggttcaaccacgttgggatagcccgcaggacatgctcggtttctacaattatttggagaagcgctacaaagcgaccgaatttgcacgggctctggtgcatttcgacacgtacaactggccgcttgcccggcctttcaaggatcggctactgttgatcctgcttgacgaactgaacctcgctcgcgtcgagtactacttcagcgagtttctgagccaactcgaaggccgtcccgcccgggcgatcgcgatcctgagcacatccgcagttcggaaatcgtgctcgatactggcggcgttggcggaccgccgccacgcatctatcccggccacaacctgctgttcgtcggcacgatgaacgaggatgagtcgacacagacactttccgacaaggtgctcgatcgcgccaacctgctgcgcttcccgcgccccgaaaaactggccggagaaacgctggcgagcggcggcgagccggcggaaggcttcctgccggcctctcgctggcatgcgtggcggcgcagttttggcacgctgccggcaacgctgcgcgaaccagtcgaacgttggatccacgatctcaatgagcatctagacgggctgcatcgaccgttcgcgcaccgtgtcaatcaggcgatgctcgcctacatcgccaactatccgggtgtcgccgagccgatggcgcaaaccagtcctctggatcaggcccgcattgcctttgccgatcaactcgaacagcgcattctgccgaagctacgaggcattgacctgggtgactctggagtcacccagcacctcgaccgcatccgtgcgttgatcgacaacgagttgcatgatgcaacactggctcgcgcctttcagcgcgccgcgcaagatgacggcagcggcaggccgttcgtgtggaaaggcgtacgccgtgaatcgatatgatcccgctggtgctggctatgccatggggactactggcacagactccgatcgccggccagccgacgcgccgaccgttacatgacggtgaaacggtcgaactcgatgggcggtacggtgccatggtggcgctacccgagcggaccgacctgcaactgggcagtcggcgctggccggtgcaggtggaaggtgccgcctttgcctggttcgagggatcctttcggttggtgtcgctgccgactgcagccttgaccagcgaacgtcagatccggttcgatcttctaacggcgggcgagtctgtgctgagtgtcgggctcgtgttgcgtaatcatctactgcgtccgcgcggagccggacgtgacgatccggccgccgatgcattgcacacctttgtgttgcaggttctcgaccgcatccgtgaggccgaaccgtccggtgccggagacgattgggatgatctcggcaccggttgggcgcggctgcgcaccgcctggcttgagcgcgatgcgcagatcgaagaagcgcgccgcgatctgatcgtcgaacatgctgaacaactcccggcccacatcaca
		gaaatcgctatccacccgcgtcgggtgctcaaacgcacccgcgagttgctgccgatcgatcgtatccaggaactcgacaccgcctg

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		tctcgaatggctgatccggcagcccggcgttaccgttgccgaaaaggccggtccgcgccagcgactgctcggcatcgcgcgcgaggagcatctcgatacgctcgaaaaccgggtgctgaaagatttcctgcgtctgagcgtcgaggctgccagcgtctggcagcgggagaaccggcgttttcacaacagtgagcgcgcccggctggtcgggcgttatctcgcgctgtgccgcatgcatcatcgcgaactgtgcgcggctggcatcggtgaccccatgcccccggtcgctccgaatttcgtgctgcaacaagattcccgctaccgcgtgatctggcgcgcgtaccgcgaactgttgagcgctgagcagcgtatggacgatctctggcgctggcagtgtcggttgtggagcgacttcgctcggcttgtcgtggtgatgggggtgcaagagttgtgcgacaagccgagtgcgctctcgcccctcttcgtgcgcagggaacaggcaagcggacgctggtcggacacgctcggcctgctcggtgtattcctgatcgacctgaacggcaggtcgtatgtggcggaagtctgtgatgcgagccagttgccccgaaacgacacgtcacgagcgaagctggcgtcctggcagtatgcactcggttgcacagcactcatccgcctcatcgatttgtggagtgggcattgtgcgagcctgtgtgtctgggccatgcatagcgctacagccgagacgcttccgttgaccgagttggtcgcttcagccgatgaagccctgagtacggccatcagacaggaaggtctgcgcaacggcgagcaacttcgggcacgtggactggtgatccgctcggcgccgccgggaaagaccgagtacgccacccaggctgggcaggtctacggactgacgctggccatcgggtcggaacatatccgcgaggcgcttggcgagtgcactttgatcctgcaggacagtctggagcgcctgtttgcatgagcggagtgcacggcattgatctcaatggtgtgctcgattgcgtggtgcgcctcgatcgggcaccgcgaccagcgccgacaccgccggtgatcgtctccggttcaccacagggcctgctgacgggagccgcggcactgcaatcgccctgcggccgacctggcatggaagccgaggaaggtatccgcctgccagtgctggccctgctgcacgcgctcagtggtgaggggcggcacgatacgcacgatacggccgtgctgctcggccgacacctgcgtagcctgttgtccgatgatacgcatgctgctgtcgtcgcagtgcctgacacacctggtttcgacgaacgagctcgcacccggctgctggatggcgcgctacgcgccgggctcgatctgcacctactatggcgcccggtcgcagcgttgcttggttggggcgaaacactgggaaacggcgaactccaagccctgcacggccggacggcctgcgtcgtgcagttgttgccggacggcatctcgattggcgatttcggcctcgaatgcgtggtgcagggtggccggccgacgttagtaccggtgcgccggcgcgacggcgaacgtcaattttactcgtggagcggtggtggactggttgcactgctcgcgcgcgaagctggaaccgacgaagccagtctgtgggtcggaccgtgggtatggaaggtcttgcttgggcagcctgcagaacgcgaggtgctggccgacccgcatgcaccgggtggttggcgactcgccagcggtccttccacactgtgcggcgccttagccgcggagttgcgcacaggcctgcgtatagcactcggagccgcgcgctcggcactgcgcaatgcagcggtcaUctgatcgaggggcctatcgccgatgcaccgcUtcggacgcaatgcagccaacactcgcgctacgccagatcgtggctgcggaactgaccgtggtgctcggcccgacggtgtccgcaagactcgtcgccatgccgctcgccgatgctctaattgccagaggggccgctatctgtgctgcgcgtcaagcggcgcggcagatcacgtattacgatttcctgccgatgctcgaaatcaatgtgctgcaggccggagagcatgcgttcgttgaactcatcggtcgcgaagagcgcatcgcggggggcatgagttacacgaatacgttggccgatcgcttcaccgttgccgcaagcacgcgctcgctcgagttctacctgctgaaagaggacgaagcaggcgctcgtcacagcgaaacggtgctgccggtaccgccggcagccgacgtggaaatcagcctgcacgtcacgcagacacccgctcaaggctacgcacgcgtggagatactctcggccgtccggggcgcgctcggtgaagcaccgatcctgctcgattggtcagcgatgacagagattgaaggctcgcgcgaggatattctgcgcgaactcgaattcgaggggctcggctatccggacatcgtaccgcaacgtgcacatcacctgctctgggattaccagcgcagtgacggcatgactatcgctgccgcgatgcgggccttcaattgtaagcctatcctaagttcaccgcgcaaccagtacaatcaattggttaaacaaacgcgcgcactcgtcggctgcgcagcaatctgtttttctgacaaagggcaccagttctgatcgtagtgcttacaccgccgtcgattcggatggccaattgccacctggaatcgcgccgacaatccaacaggaattcgaaaactttcgagtgcggctcgacacggattttgccgcaatcaccagcgtccgtaatcgacaagatatcgcaacccggcgtgaattggcgcgactgggcgcctgcttgtatgcagcgtgtcctaatgcaattgttcattacttccaacgcattgtcgcacgtagcgccgatgacctgacactggtgttgcatgccggcaaagtgctgagcaccgaaccagatcttgacagtcttttccattattgcgcgtctcgctacgatgaagccatccgcgctgtcaagagactgtcggtccacgtggtacgcgcggcaggcgatgctttggcttatcatgaaaaagctggaggcattcttgataaccgaagcgctgacaagttggctgaagctgcgctcctattgctaaaggaggaaatccaggcacataattacaaaatacgattccgtgccgccgcgcgactcggcctatttctgttacgccaccggcagcggcggcgcgatttcctgcatccgagtagcgctgacacggctaatcgtcggcgtgccaaagagttcgatgccctgttgatccaggctatcgcatcgaagcgccttaaccaagatctggaaaatgccttggaagaaatccgtgcacaaatccgatatcgcggtacaaatgcgatcgttgatatcgatcctgacgaagatggcgagattaacgagaacgaagtggagtagaggctgttgggcacccgctcgccatccctgtcgagcatcccggcttcgcgggcgcccatcccgtgcctttacggcgtgttcaacggccccggttcgccctgcgtatcgggctcctgctacgcccgtcgagacgcgctgcgcagactcgacgctcaaatggcttgacgccattctccctggctacc(SEQ ID NO: 52)
56	pLG058	tcgcgatcaaggggtgagcaggggataaacgcaaagacattgaagttgaggagaatttagttgccttacctgcgaaaaatctgagcgatcttgcattaaagattttctatctcaggccgatgctcataagagcatttcctgaatttcaccctttttttgctcgccatccctctgcgaataaggacaccgcgccagatatgtcactcatcacccatacattagaaaacctcacaaaagccttgcgtactgcgttgcgtgtctcaattgaatgcaatgagcgcagcgaaaatacccataaaattttaaacgtgttacgtcaggttgagctgacgctgatgctgcatcaacaacctatctatgccattgccggtacgcagggagcgggtaaaaccactctggcaaaaagcctgctgggcattgacgatagctggcttgaggcgaatccgggacggggcgagcagataccgttatttattgagcaacggcacgatgttcagggtgattatccgcaatttatttatgtctgtgctcaccacaaaaccggtgaaattttgacagccagccgcgcagtggcgatgagctgaaacagatgctgcgtgactggtcgcaaatggtgaatcaggagatagaagggggcaaaatcctctatccgaaattaatcattaataagtcagacagttttattgatgaagagatggtctgggcgctgttgcccggctacgagatcagcaacagccagaatcatcgctggcagggcatgatgcggcatgtcatggtcaacgccagaggcgtgttgctggtcactgacccgacgttaatggcaaatacgaaccagagcctgctggtgaacgatctgcgcagtgtgttcgccgatcgttctccggtgattgtcgtgaccaaaacagaaagcctgaacgatgcggagaaggccgaggtaaaagcgagcgctgccgcactttttcatgagacctcctcaccggtggtcgctgccggtgtcgataatcaagcgcagtggataggtgagctccgcactgcatttgctgagggtatccataatagcgccgcgtcagaagcggccgcgatcgaacgtttgatgactctggtcaatgacgatgttgcggatattattgataacctgaatctgctgtacgcggagcaggacagtggcgaggaacgtaccgtcgctattcttgaagcgttcgataaagcagccgagcgctatgaacagcaactgcgtaaagccatcaaacgagaaactgacgggcatcggcaaaaagccactgaatcttgccagcgccgttatcaggaagaagaagaagggccggtcaataatttaaaaggactcggtcgtcgtctgatgtttcaggggggcggagattgatcgtgaacgcaaaaatcgggtactggacgcctggcaaacccgctttgagcagcaatctctggccgatcacaatatggtcgcgctggaaacgctcaaccgtcgtgagttgaggcattacggtctttcacaggagacgctgtcaccccaacggttgacctcgcccgcggcgacaatgggatatttgtcggtggctgaggaggataatttttcctcgctggcccctttgcgccatctgctgggatcggctgcaacaagggatgcgccgccgcagttagaccagctttccacggtattaaaagtgctgcctgccatgacgatggaatatgcgcgcggttgggtggcgatcaaccaggcgatgcccgcagcgtcagagctaaccagcgagttgcggccacaacaaattctcgacgcgattttagcgcgcagagtagcatccacccggtgaaaaccgcgctgatggcgtttatcggtgccgacgccgcggacggcacgctggatggcgaagtgggcactccgcagaatgaagatagcggcgtatttacgcctgtcgcgatagcaggcaaagcgatgctggtcggtgcggcggtttatgcgttgtatcaggtggcgggcgtggtgagtgagagtgataaagctcaggcctggtatattgaacggatgatgaaggaactggcgcaatataatgaaaacgtcatcatcgagcgttatcaggacacgatgggcgatctgcgtcagctgattgaaatcaacctcaaccgtttatttggcgtgcaggatgtcctcacgcagaaaagctatctctggttagctattcagggactcacgacggtacaaaaggaagcccggcagtatgaagccagtatcaaacaatatctggcgtgatatttgccatgagcgttatcgatgggcggaaaatagctacatcaacctgctgcgtcaggttgatgccgagcggttaatccagcctcatgcagacatctcccgccagatatcggtcattgtctatggtccgacgcaggtgggaaa

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		aacctccctgattctgaccctgctgggcgtcagggatgactgttttaaagaacttaaccagctgctgcgtggtgggcaggcattaggtcacgcgtcaacggcgcgaacttaccgttaccggatatcacgggatgatgcctggtattttagccacaaagaccagggaacaaccgcctggtcggatagcggggcggcagatattttcgccagcctgcgtgcagaggttcaggcgggcaggcgctactttgacagtatcgacgtatttattccgcaacgtttcttccatcctcagcagcggcaaaatggtttgttaatccgcgacctgccgggtattcaggctgcggatgacaatgaaagggaatatgtgactcagcttgccagccagtttattcgttctgcggatgtgatcctgctgaccggcaaagcggattatttaggctttctgaaacccgaggagttgggtaatgacctactggctgactggttctggcagccacatcgctacaaaattgtattaacccggacttttagcaacagttccattcgggaaatgttgcgccgtgtttcccccgataaatcctggctgcaggcttatttgtttgagcaaatcaatacgctggaattgcaacttccggcggagatgcgtcaacacatttatccgctcgaatgcggtcactcctggcaaaccctgattgaggggggtgacgattatgctgactattgccaacggttgcgtgagcagatattaaccgacctgcgccatcatatgttgcaggcggtccatccactttctcgtttacgtacgggatacgccttacctgaattaattatccgccaccgggacaagttgcagcagcagtacacagcgctgcacagcacgctggacaaagaacaggaatattacctgcgtaaaaaagagcagctgtcgtctgtgcagactgaatattcccggcatctggcaaagagccagacacgactggacagattgcagcggctacgggaacggctgaataaaagacaggcgcgcaacgcgcatcaatccatcgctgtgccaccgatgggcacaagaacggtcagtgccttactgaaaatgattgctgaggcaagagaagagatggcgcttcatccggcgttaaagcaccttcctgcccatttcgctgcgcaacagattaaccaccatgccttcacggcgattgagcaaaagctgcatggctatcatgcggataattatctctttgccagcaactataagcatgactatcaggaaacgatcaacgcgatcaaacaacacctgaaactgatcaccacattagccgctaatttccagcgtagtgagctggagagacacatcaaggaacatcgtcgtcgccagcaacgtttacaacaccacaccacccggcgagacaaactcctgacggcagtgaccaataagcttacgcgcatcaatacgcagcaacaggaattaacgcacagccatatgcgtgacgaggatcattatcagcagctgattggcgagagccgtcgctttcaggaactgatcagagtggcgaaaaatgaacgagccaccctgattgaacaacacattaggcgtacggatattggtcaggctgagcgactggcctggctactcgctgcccgtgcgttaaagaaagactacgaatatgtcagagcattaggagagtagtgcatgtcagtggaacatgacccggttattgcgcaggataatgacgagcggatgctggatgaattggtgcaggaactgtttctgaccttgctgacgcgtgagctggcgcaacagaaagcggttatcgaaaccattaatgacaacgtctcgtatcaggctggtgagtcattaaaatcgttgaaacgggagatcaaactttccatcagcaccctgtcgaatgcgcaacagcaatatcaggaagagcaggccatcgccagggaggaatacgagaagcggctggagcagcagactcaaacatttgccagtgatgcggaaaaaatcaccaacagtcacagcagcagatggcagcacttcggcaaggtgagcagcagctggctgcacagttaacagatttgcagcaacagcatgccacacttcatcagcgctcaggtcagatgctgaatagcattaaatggctggtggtggggctggggggcgtcaacctgctgctgtttgcggctgtcatcatgatgtttttctcgggcatcgataatcatccgcgcatgcaggtttgtccggatatggtgcgcctggtgcaccatgacttttctctggcacggataaacggacgcacaggcagcgaatgacgcgccctgaataaactggcacaacttctgcattcatttcctcaggcttgtatacaaggccgcataccg (SEQ ID NO: 53)
57	pLG059	cgcatctgtaatgcaaacttattagacttaatccctataatgcaatataaatcatattgttaccttgtggctcctttatctgattgcacggatttatccctcgcgtacttattcagcatgatatagctgggtatcatgtgcctactcttaacctgaatgaaacttacaaacgttcgtggtatccacatgctaagtgaggctgagatagcaaaatttctcatatggttgctgcccctaagatcaacaacgcactgagcatgactctctggacaaggtgccacacaccaggcgcacgtctaaaaggaaatatacatcaaatacctgattgctaagttataccaagtggaaatcgggtatagtaggtcaaaacgaaagcgtgtcttaacactgcatattaacgatcaggaaggtcttagcatgtcaattaatatcaatacgttgcataatcttcgtcgcgcgttacttactgcgctggagctctcgattgagcacaatgaagaaacagaaaatgtcgatcacattactgatgttctgcggcaggtggagttgacagtacttttgcagcaagaatccatttacgccatcgcaggtatgcaaggggcaggtaaaacaaccttggcgaaagcgatccttggtattgatgatgaatggttagatgccaatccgggtcgtggcgaacaggtaccgctttttatcgaacaggtggatggcgatccctccgattttccacaagttgtctatcagtgcctaaaccttaaaacaggcgaaattgctccgcaaaagggcgagggtggggagcaacttcaaagtctgcttcgcgattggagcagtattcgtcgttatgaaaaagcgggctttaaactgctctaccctaaattgctgatcagtaaaaaaactcgttcatcaatgagcaagtgacttgggcgctgttgccgggctatgaggtagccacaagtaaaaactatctctggcaggatatgatgcgccacgtattggttaacgcccgtggtgtcatgttcgtgaccgatccctctctcttagccaatgacagcaaatccgcagtgctgcaagatttgcgagataacttcaaggaacgcggcccagtggtggtcatcagcaaaacagagatgctcggagaacatgaaatcaaacagctcaaaaccagtgccgctgaacgtgttttccccaatgttgggatgaaaaaagaggatatcgtagctactggttctggtaataacgacatctggattgatgcactacgtgacacagtcatcaataagctcaccagcagtgcggtatctgaagcaattgcactagataacttcatgggacttatccgcgaagacgtggccgaaataatcaataatctgaagatattggcggatacacagcagcatcacgaatccatagtggatgagatcctagacgttttcgatgaatcagcctccacccatgagcaaaaattacgtgaagcgatcaaaaaggagacccgtcagcactttactgatgcgcttaagtactgtgaaaaaagctataaagagaagaggtaggttttcaaaaaaacctcaaaattttcgcccgccgactgtcgtttcgcggcatagaagtggatgatgagcgcagtcaacgtattatagatgcttggaatagacagtacgaaaacatcagtattcacgaacataatttcgacgcactgacgtctgtgaatacccgggtgctgcgtgccaaggggctattgcctgtcgttgaaaatcagcaactattaccgggcagcgcagtcgggagaatggggtatctggttcaggataaacaagcagagtactcaataatggatcctgacctgatgacgggtttgtatacactgctcaaaaagccgggcggcgctcatcaagcaccgccgcctaaaaactcgctgcggcgctggagattatgcctgctttaatgctggaaaacgctcgtactaggttggcaatgcatcttgacccggcctgcacaacccaactggcagaggagatccagcctaaacaaattttgatgcgctcttttcgagcagagaacagtaccatcctattaaaacagccatgatggcgtttttgggtgctgatgcggcagatggaactgtagacggtaagagcacgccaaataccgaggggggattcgctccgctagcgctggtaggtaaagcggcattggtagcaagcgtggcttatggcatctatcaactaacaggagttattcgcgacagcgataaagcgcagatttattacattcgtcgtgtgatggaggaattgtcattccataacgaacagaccgttattggcaattataaggagatgattggcgaattgcgtgattatattgcgtataacctgaagcaaatatttggcgaaacggatgccctggcaaatcgaagcgccttgacgcttgccattaaaaatcttgttgccgcacaaaaggaagcaaaattgtatgaaactcacttccgaaaaatcctgggctgatctttgccaggagcgttatctgtgggcggaagagagttttgtcacgtttctacaaaaatttgacgcacagaggttgatccagtcggcagacaatgccaataggcaggtttcagtgatcctgtacggtccggcccaagtaggtaaaacctcattaatcctgaccctgctgggtattcgtgatgactgcttcaccgagctcaatactttgctacgcggcgagcaggggctgggcacaatgtccacggctcgcacctatcgctatcgcatggcgaaagatgacttctggtatttcagccatagggagtacggtgcaactcggtttagtgacaaggaggcgaaagtcattttttgcagattttcgtcaggctgtggagcagggcgagcgtgaattcgatagtgtggatgttttcctgccgcgccgtttttttgatccgaagttacagagcagtgcccagttgctgatccgtgatttacctggaactcactcaaccaacgccaacgagcagtattatgtcaacatgcttgccagccgatatcttgcttctgccgatgtggtactgctgaccggcaaggctgatgcgttggccttccttaagccggaagagttagacaatgctctgctgaacgactggcactggcaacgccaccgctacaagattgtactgacccgtgcttattcagatgccacactccagcgttttatcaaacaaaaacggtttgataaaaaagcaatgcggatatttttgcttcaacagattaataccatggatctgggcttgcctgaaagcatcagtgaactgatttaccccgtggagtgcggtcattcttggctggcaatcaatgccaaagatgacgagtttgcccgccagtgccgtgatttgcggcgagatgtattgcaagatttactcgactctctgcaccaggcatcgaacccattatcacgcttacgttcgggatacgcgctgccacatatcattaaacagcagatagctgtcgaaaaagagctttacgagacggaaaacgcattgctgcaaaaacagctctctcggctgggggaatatgttgatatgtacgagaaacgggtcagcagtaatagagataatcacctgaggttacaagtaaagctgcaagcactattacaaaaacgtgaggacgcgttgagtacagattttcgtgaacattcgaatgcgtttcaaataatttcgcaatcatctctcggttatcttaagtctcaaatttatgcatctcgtgaaacaaataccaaacgctggaacgatctg

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.		
Row No.	Vector	Locus
		ctggaaatctaccagcttccacttgaaagagtaccggagatgcccaatctagagcgggtcttaaaaagactaaacggctacttgtttgagacctattttcgagagaaaacacgtcagaatgatcagtatgagatagaagaggcaggctttaaagacgcaaactgcttaacgtatattttccacgaacgaatcaaggttaagtttggtgccgaagagcgcgccttgaacaataagatagccaaaaacgagcgggcagcgtgccgactggtgcgtatcgHgaacaattgtcgaaaaaaatggtgcacacgcagtcaagactcttccagatcaagcaggagttaggcgtatcgttaactctrtattttcagagatataaagagagtaaaaacttttcgaaagtcattgtttcggcgaaaaatactcgagcgcgtgaaatcgaatgcaacgctaaaaaaccgaatattacacgcagcgagcgtctcgcttgggtgctgatgtatagagcgttaaagaatgattttgactacgtaaagtccttagatgaggagagcactaaagttgaataaaaatcttgctgtcgcggaagtgtccagcgatgagcagttactggaccaactggtgcaggagctgtttttagagcatttgcgacgtgaactgggtgtgcagaagaagagtattgacgacagtaatgacaaactctttaatctcgaccgaaaatttgtcgctgaatttaaaaacgtgagcggattgcttgatacgatatccgacactcttggcgaacagactcgtgaactgaatgatgctaaagctgatgcccaaacacattatcgttctttgctgaatagtttggcacagaaccgaacggacaccgctgctctgcaagatatactccagcaactaagtagtaagcgtcataaggaacaaggcgagcaactgcaacggatccaggaacagttgtttcatcagagcgctgaactccaagcgcaatactccgtgttgacagaacagaatgcagtgttaaaccagcagcaggaggtccttcagaaacaacggttcactgctactctggccgaaatgcaagagcaaaacgtgacgctggcgtcacttacggaacagaataagtcgctgcatcgacagtttctcaccttagaagatgaacaacgtgcagattttcggacaaatagtcgctggggtaagcttgccgctggattctccatagcgaatacgcttatcctgataagcgtgaccgcactgtttatagttaagtactttctataaagaacccgcgtgcacaactcttcttcatataaaatatcttttccaacagatattgcattgaggatttcttttattgctgtttatgaaatggctaaatatcctccgacaaataagaacagtggcggatttttcatcctcgtctttttcagggag (SEQ ID NO: 54)
58	pLG060	atcagggcaaggaccgttgcccatatgtgactggttttggtgtcggctatgtggccaggctgcgtgaaagctactgatcgcttttttaatctaagtggtggatttatatgatcaatcattattgataaactcatgaagaaacctaatttatttaataaaattaaaaagtatacgattagatattgcgggtgtagatatgactcaccacattaaaggtcaaggcagacatcaggtgacgttgctctctgacgtgcttgatgattttgtcacagaagataaaaacacgttgaagagagaaaaatgaataccgcagaagactttaaccgcctctatgccgacgtttcacgcaatattcagcagacgctgactgatatcgctgcacttcatgttgaaaatgaagagggaaagcagcagctacaatcgatggtcactcagttgcaatccctgcaggatggctttaaccagaagctcacgtggctgcaaaagcatgccgaatgggacaaatttaccctggcattctttggcgaaaccaacgccggtaagagtacgataatcgaatcgctgcgcatcttgtttgacgaagaatcccgccgccagctgctgcaaaaaaaccacaacgacctggaaaaagccgagctggaattacaggaaatctcggaacgactgcgcagcgacttagggcggatctatagcgatgtagtggataaaatcaccgatatcagtttttccgctctgcgtctgatgcaaattctcgacaatgaaagcgccctgcgtcacaaacgggaagaggaagagagcaaggaacgcctgctggttgaaaagacggaaagccagtcgcgattgcaaattctgcaaaaacacaccagcgccaaaacacgattaaccctgtgcattgccgccgtcatctcttttgtcgcaggcgcaggcgcgagcgccgccgtggtgttcaatatgatggcggggcaataggatgagtaacgcactagatcttcaggctagtaccacgtcagtacgttcgcaacgaaagtcctcattgaatattcaggagctcctgaataaaacgctgcctcacctggttcagaccataatcaggaatgagagattaaaaaacaccctacttcaggttgatggtctcattatcggtaccggcgaggcggatttaccaaagggaatacccgctacgccttacatattgacgataagaccttccatctgctggacgtacccggcattgaaggcaatgagtcacgctatatcagccaggtgaaggaggctatcgccgaagcgcatatggtagtgtacgttaacggtaccaacaaaaagcctgaaaccgccaccgccgaaaagatcaaatcatacctcgaatacggtacgcaggtttatccgctggttaacgtgcgtggatatgccgacgcctatgaattcgaagaagatcgccacgatctgatgcagcaaggaggcgcaggagaagcgctgaagcaaaccgtcggggtactgcaaccggtgctgggctccgatgtgctgcttcccggtaactgcgttcaggggctgctggccttctgcgggctagcctatgacgatgcgacgcaaagcaccactatccacccctcgcgcgcgcacaacctcgccacgcaacagaaacgctatttccagcacttttcttctcgtcgggagatgcaggaatttagccagattgacgccattgcccgcgtcattcgcggtaaagtcgccacttttcgcgaagatattgttgaaagcaacaaaggcaaagtgcgagagtcactgggtcagtatctacaggtactaaacacgcaactcaccaatcatcgcgcatttctaaagaaaacagagccggaatttgacaaatgctgcgtcgcctttgctaacgccattgcagcctttgaacgccgaatcatcaataaccgccgtaaccgctggaacgactttttcaatgatctgatggaaaaaagcgacgacattgttgaagacgattttggtgataaagaggcgattgcccagcgtattagccagcagtttaaatcgcgtcgcgtcgaggtgaaaaaattaatgctccaggacactgaggagggcgttaaggccttacaggagcagatgattcaagcggtggctcgtttgttgcaagatattaagcacattgagttccagcagcatgtcgatttcgcccacggcggtgaattcgaatttggtcgcgagatcgcgctgggttatgaccttgggttaagggatttcggctcaatggcctttaaaatcggcagctacgccttaagcggcgccacagtcggtagcgccttcccggtgatcggtacggccattggtgccgtagcaggcgctttagtcggcgtcgtcatgaccgttgtcggtttctttaccagcaaagcgtcgaaagttcgcaaagcgcaggggaaagtgcgcgacaagctagaaagcgccagagataaagcgctggacggtattgatgatgaggtccgtaacctggttgcggctatcgagaatgaactgaaaagcagcctgctgcaaaaagtgaatgccatgcatacggcattgcagcagccgatcgccattttcgaacagcaaatcacgcaagtcacccatttaaaaaatcaactcgagaacatgccttatggaacaattcaaacagttcagtattgagaagcaggctgccattaactcgctgctacagctgcgcggcatgctggaaacgctgggcgaaatggagatcgatgtcaacgacgatctgcaaaaaatcgcgtcggccatcacagccgttgagtccgacgtgttgcgcattgccctgttggggctttttcggacggtaaaaccagcgttatcgccgcctggctcggcaaaatcatggaagatatgaatatctcgatggacgaatcttctgaccgtctgagcatctataagccggaaggattacccggagaatgtgagatcgtagataccccggggctgtttggtgataaagaacgagaaatagacggcaaacaggtgatgtatgaagatctcaccaaacgttttatttccgaagcgcatctgctttttttacgttgtcgatgccactaatccgcttaaagagagtcacagcgccatcgcaaaatgggtgctacgcgatctgaataagctgtcatcgaccatcttcatcatcaacaaaatggatgaagtgactgatttaaccgatcaggcgctgtttgcagaacaggcggccatcaaaaaagagaacctaaagggcaagctacagcgcgcggcaaacctgaatgcgctagagcttgaacagcttaatattgtttgcattgcttcaaatccaaacggtcgtggccttcccttctggttcaacaaacctgaacattacgaaagccgctcacgcatcaacgatctcaaaacagttgccgctgagattctgaaaaccaatgttcccgaagtgctgctggcgaaaactggcatggatgtggtgaaagatatcgtcacccagcgtatcaccagcgcccagctgcatctcagcaaactcagcacgttcgttgcgaaaaatgatgaagatacttcgcgttttacatgcgatatccagcaaagccgtaacgaggtcaaacgtctggctggcgaaatgtttgaagaacttagtttgctggaaaagcagctgatgagccagctacgcccgttggagctggatggcattcgcccctttatggacgacgaactgggctataacgatgagggcgtcggctttaaattacacctgcgtattaagcatattgtggatcgcttttttgcgcaatcctccgccgtcacgcagcgactgtcggacgatattactcgtcagcttaattccagcgagagcttcttaagcggagttggcgaaggggcatttaaatccctcggcggcgtgtttaaagggatttccaaaattagcccggagacgattaaaaccacgattttttgctgcacgcgataccattgggcaattaacgggctatgtctacacctttaaaccgtgggaagcgaccaaactggctggcggcatcgctaagtgggctggtccggccggggccgcatttaccatcggctctgatctatgggatgcctataaagcgcatgaacgtgagcgagagctggaagaggcgaaaaatgagttgacccggatgatcaaagatccgttcagcgatatctatagcgtcttgagttcagatgaaaagacgttcgctttctttgcccccagattcaagagatggaaaaagtcatttgcgatctgacagaaaaaagcgacaccattcggaagagccagcaaaagctaagcatactccagcagaagctcgagcagtttaaccgttcgagcgagcagcaagtgtcctgatacacaaacggcagcccgcaggccacgtttagttataaatcaaactaaacgtggccaggtgacatgcccccgttgattaacacacgttatcgtcgggtggaaaggacaacctcctacgtccgcttcacagcggacactcaggtttaacagtccagtacgtttagcttacggataaatcattttatgatgatgtggagaatgggggat (SEQ ID NO: 55)

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
59	pLG061	tattttgcgtagctagaacgcaatcaaatctagcagtccgctttgttcggagttcggacattatgagttggcaagtaaagtagcttgctaggaagccggatttgcacggtcggtataataagatgtaaccccttgccttcatttactcgaatgaacgtgcacattggataggaggaaaaggaatgcaattcattaccaacggccctgatattcctgatgagcttttgcaggcgcacgaggaagggcgcgttgtgttcttctgtggagcaggcatttcctaccctgctggtttacctggtttcaaagggttggtagaactaatttaccagaggaacggaacaacactttcagaaattgagcgtgaggttttcgagcgtgggcaatttgacggcacattagatttgctggaacggcgcttaccagggcagcgtatagccgtccgacgcgcgttggaaaaagcccttaagccaaagctccgtcgtaggggcgctattgatactcaggcggcgctgttacgtttagcccgtagccgcgagggtgcccttcgattggtcactaccaactttgaccgtctctttcatgtggcagctaaacgtacaggccaggcttttcaggcctatgtagcgccgatgctgccaattccaaaaaacagccgctgggatggacttgtatacctgcatgggctgttaccggaaaaggcggatgatactgccctgaatcgtctggttgttaccagcggtgactttggcttggcttatctcactgagcgttgggcagctcgctttgtgagtgagttatttcgtaactatgtggtctgcttcgttggctacagcatcaacgacccggtactgcgctacatgatggatgcgcttgcagcagatcggaggctcggtgaagtcacaccacaagtatgggcactgggggagtgtgagccggggcaggagcaccggaaagccatcgagtgggaggccaaaggggtcactcctatcctttacaccgtaccggcgggctccactgatcattcagtgctgcatcaaacgttgcacgcttgggcagatacttatcgagatggtatacagggcaaaaaggctatagtcgtcaaacatgctctggcccgcccgcaggacagcactcgtcaggacgatttcgttggtcggatgttgtgggccttgtcagataaatcaggtttaccagcaaaacgctttgcggaactcaatcctgcaccgccgctggattggttattgaaagctttctcggacgaacgatttaaatacagcgatctgccacgcttttgtgtatctccgcatgtcgaaattgacccgaaactccgattcagtctggttcagcgtcctgcgccctatgagctggccccgcagatgtcgctggtttctggatgtgtcagtgctagcaaatgggatgacgtaatgtcccatatagcccgttggctagttcgttatctgggcgaccctaggttgatcatatggattgctgaacgcggcggacaaatacacgaccgttggatgtttctgattgagagcgaactagatcgcttagcagcactgatgcgggagcgtaagacttctgagttagatgaaattctcttgcattcccccctggctattcctggtccacctatgtctactttatggcggcttctgcttagtggtcgtgtgaaatcgccattgcagaacctggatttgtatcgttggcaaaaccgcttaaagaatgaaggcttgacgactacattgcgcttggagttacgcgggttgctttctcccaaggttatgttgaggcggccgtttcgctatagtgaagacgattcgagcagcactgatgaacccttgcgaatcaagcaattggtggattgggagctggtgctgactgctgattacgtacgttcaaccctgttcgaccttgctgacgagtcatggaaatcgtccttgccatacctgttggaagattttcagcagttgttgcgtgatgcactggacttgttgcgggagttgggagagtccgacgatcgtcacgaccgctcgcattgggatttgccgtccatcactccgcactggcagaaccgggggttccgcgattgggtgagcctgattgaattacttcgggattcatggttagccgttcgagccaaagacagcgatcaggcctcgcgcattgctcagaattggtttgagttgccatatcccaccttcaaacgtctggcactgtttgccgcaagccaagacaactgcataccacctgagcggtgggttaattggttgttagaggacggttcatggtggttgtgggccacggatactcggcgagaggtattcagactgtttgttttgcagggacgacatctgacaggaattgcacaagagcgtctggaaactgctatcttggcagggcctccgcgcgagatgtacgaggataatttggaagcagacaggtggcattatttggtggctcattccgtctggttgtgtctagcgaagctcaggggagcgggccttgttttgggagagtctgcggctacacgtttgacggaaatatccacagcatacccaaaatggcaactggcaaccaacgagcgtgatgaattctctcactggatgagcggaaccggtgatccaggcttcgaggagagtatagatgtcgacattgcgccccgtaagtggcaggaattagtgcaatggctcgcaaagcctatgccagaaagactgcctttctatgaggacacttggagtgatgtttgccgtacgcgcttttttcacagtctgtatgcgttacgtaaactatcacaagatgatgtgtggcctgttggtcggtggcgtgaagctctgcagacttgggctgaaccagggatgattttgcgttcgtggcggtacgccgcaccgttggtgcttgacatgcctgacgcagtacttcaggagatttcccacgctgtcacttggtggatggaggaggcttcgaagaccatcctctgccacgaggagattctactggccctttgtcgtcgggttctgatgatagaaacaagcccagagtctagcaccattcgaaacggaattgagacctatgatcctgtttctacggcgatcaatcatcccattgggcatgtcacgcaatcactgatcaccctatggttcaaacagaacccgaatgacaatgatttgcttcctgttgaattgaaaacacttttcaccaaattgtgtaatgtacagatagagctattccgccatggtcgggtgttgctggggtcgcggctgatcgcattttttcgcgtagatcgaccttggaccgaacagtatctattgcccttgtttgcttggagtaatcccgtcgaagcaaaagctgtgtgggaaggcttcctctggtcgccacgcctgtatgaaccgttgctgatagctttcaagtcagattttttggagagcgccaatcactattctgatcttggcgagcaccggcagcaattcgctattttcctgacttatgcagctctgggccctaccgagggatataccgtggaggagttccgaacggcaattagtgctcttccacaagaaggtctggaggtagccgcgcaggcgttataccaggcacttgaaggtgcgggcgatcagcgcgaggagtattggaaaaatcgtgtccagccatttggcaacaggtttggccaaagtcccgcaacttggccaccccacgcatatccgaatcgttgactcgtatggtgattgctgcccgaggtgaatttccggcggctttggcagtggtgcaggactggctgcaaccgctcgaacaccttagctacgacgttcgccttttgctagaatcagatatttgcagccgatatcctgcggacgctctatccctgctgaatgccgtgattgccgaacaacactgggggcctcgagagttggggcaatgcttgcttcaaattgttcaagctgctccacaactggagcaagatgttcgttatcagcgattaaatgaatattctcgaaggcgcagcgtgtgaaagtgacaggcgttggacagtgcgaactgtggagcctaacaaggtaaagacactctaactgataatgctgcgccgctcgtgcaatgcaatacagtttttatctagcggtgaattatggtgttaaaagttagcccctgacacagggtgggtagttggctctgtgtcattgatgggtattagttctgatatgagctaataccca (SEQ ID NO: 56)
60	pLG062	gtaagacaagggttgagcaggctactaatcgttacacaggctaacaaaggcatattaagacgatttgtagcgctgtaaccttgaaaattatgtacaagcgccccgcattacgtcgttttaaaggccatcggattcaggcccgacgcggcttcacgcgattataaccgtgaaaaatcccccccgcatagaacctgaattatccccgccgccgcgcagaactgacagcgcttcagaaccgttaaccctctcagaaatcccgctttttactgtaaaaaaccatgcataaggtgcatggttttgcatgcgtttcaccgacactgaatcccccgccagcgccagcagtagcgtgccctgaggccgttaatgcacccgtattaaaagcgccctgttaagcgagcaggcggggcggggcgagcattgcgcgtcggtgttaccaattctatatggacattgagcaattcaaatataataaaggttgggtatatttcgtcctcaacgatgtcaaaaactgcaaaagcgtattataattcagatcattttcagaccacctattttaatcatgcatgcaaaatggaatatgtgatgacaaataaaaacaaaatcaaaccattattaaataatatatccgctcgcctttgggatggtcgtgcagctatattgataggagctgggttcagtcggaatgcaaagccattaacaagcaaggcaagaaagtttccaatgtggaacgacttaggtgacattttttatgaaagtgtttactgcaaaaaaaacgacaatagatattcaaatgtattgaagctaggagatgaagttcaggctgcatttggtagagcgacacttgataaattaatcatggatcatgttccagataaagaatatgaaccatccaaattacatgtttcccttctttccttgccgtggattgatgtttttacgactaattatgatacattacttgagcgagcaagtgttaatgtcgactccagaaaatatgacattgtccttaataaaaatgatttaatgaatgctgaaagaccaagaattataaaactgcatggtagcttcccatcagaaaggcccttcatagttacggaggaagattacagaaagtatcctttagaaaattctccttttgtgaataccgttcaacaatcattgattgagaatactctatgtctgataggattttcgggtgacgatcctaacttcttaaattggattggttggataagagataatcttggcacagaaaattcacccaaaatatacttgatcggtcttttttcatttaatgaagcacaacgtaagcttttagaaaaaagaaatatttccattgttgatttaagttttctaggtgattttggcaaggatcattatctagcacaccaacgctttatccaattcttatacgaatcaaaaaatcgagacaacctaatagagtggccaatagaaaccaattatgacagaattgtttttaatgatggcattgaattaaaaactgagaaaattaaaaagtgtatcttagaatgggctcagtcaagacaatcatacccgaactggcttattttgccggaatcaaacagaagtaatttatggcaaaacactatagattggttatctgttgctaattatgatgtcgcttgggatggttctgatgatcttgattttggatatgaaattacatggcgactaaataaagctttgctaccaattttcaatgatacatcagaattcttatttaagttgattgaaaaatatgagatcaattacgtttcggggataaataataaaatcattgactt

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		tgatgaaaaatactctcatataaccctcagtttaatgagattctgtcgacaagaaaaccttattgataaatggaagaatctaaacg atttattaattcaaaatcttgatcgattaacaccagaggtaaaatctgattattattatgaaaatatattattttcatacttcaat ttaaacttcgatgaagccagaaacaaactctccaactgggaaacgaataaactcctcccccatcatgaaataaaaagagcaggatt acttgccgaatttggaatgcttgatgaagcaatcaatcttcttgaagaaactttatctacgattcgaagaaacagtttgctttcat ctagaaacattgactattccagtgaatctcaagaagcatatggaatctatattttgcgaatgtttaaacggagtttgcgtttagat agcaaagatgacgattattcatctgagtataactcgcggttggctacattatcacaatatcgcagcgatcctgaaaacgaaataaa atacctagaaattaaactagagtcactaccaggtaccttcaagaataccaatgacacggatttcgatcttaacaaaagaacggtga ccacttatttaggaggaagcccaacagaagtgaggtcattagatgcttttagtttctttctactggcagaggaacttggcctccct ttccacataccaggaatgaacattttagtggaatagttgagaatgcagctcgacatatttatcaatactctccagagtgggctat tttttcaatatttagaacatttaacaaggataaggccaagagtctattcaatcgaaatagaatttcgtctcttgagcgaaaaaagg ttgaagatttatttgatggatactacaaaaaatatgagcaaattatcacaaaaaaaatagaagatagattaaacgataaacttgag atagaaatttctacgctatcaatcattcctgaaattctttcccggctagttacaaaagtatcatttaataaaaagaaagacattat tcacctttgcttaaactgtttaactcggataattttcatcaatacatggagactaaagatctattaaagcgcactatttccaatt tgagcgacttacaaaagatctcactaatagatattttcattgatttcccctccgcgcctcccaatacccaattacatatgggtcaa agatacaacttccttactccatttgaatgtctattaggggttacaataacccccccaaaagaaaactctaaaaaaatcgcatctgc aaaattaaaaaaagatataaacgatttaaaaagtgataatttagacttgaggaaagctgtatcacaaaagctcataacattatata acctagaaatgcttaacaaatctgacacgactaaacttataaaaaacctttggtcaaagcgtgataactttggattcccaataggc agtggttactataaatttttctttataaacaaccttaacccagataatgaaaatatagccgacaaattcatttctataattaaaac atacaaatttcctgtgcaagaaggaaaaagagttagtattacaggtgggttagatgagtattgtactgaactcaatggagcgctac accatataagtcttccagagaaaccctatctgaaataatttcaaaaatacatgactggtatgtcaaggatcgggcctggcttgaa aaaagagatgatttagccaaggagttcactcttagattcagaaatatcacaaatatcataacgacaattttagaacaccataagga caaattacatgctgaatctataaatgaaatatcaagcctactagataaaatgaaagaagacaagatacctgtaaactcagcagtaa caatgctttgtctgaaaaataaaagcacttacctcgagagaataaaagatatagagaatggactatatagctttaataaagatgat gttattgaagctatcaactcaacttatgtctttattagaaacaatgaatttccactaaccatcattcaagctatcagcgataaaat cgcatgggatagaaaccctcgccttcctgattgctacaatttaattgcatatataattaactcgtgtgaatttactcttccagatt atttaatagagaaaatccttcgagggctggcatatcaaataaacattgatgatagagattttgttgataacaatgaatatttgaat caccttgagaaaaaacttagtgcaacaaagctggctgcttctatgtttagaaaaaatgaaacactaggtattgaccaaccttctat cattcaagagtggaaaaacatgtgcaactctagaaatgagttcgatgaaattaggaatgaatggaacaacaatatataaataaagg aagaacacccaatttatattgggtgttctgttcacgaaaccctttaccataatcgaatggcaatataaattgagattgaaattta ttctcatctaattaatcagcccaccattg (SEQ ID NO: 57)
61	pLG063	actagctaagcaataagggcgatcggctctcccatagatcgaggccgaatgatgttagcaatgttcactcttggctggaatctgcc agaaatcgaggtcatatggtctgctttgagtgaggagcgcaaatggataaagccctcatgagttctttttcaatgacctaactttt gagaggcactgggttagatcatgtttcatgtttgcaatacaatatatatttaaacttaggtttataacttaaatgttagttcctga tctaaaccagattattaatcactcctagagtgaaatgagttaagccaagagttgataaaattaacagtttttttacaatatctgg atgtttgctagcgaacaggcatctaaaataactatgctgagctaaacttacaattcaaattgtaccgaggataaaatgcaagtaca acatcatactgaaccaaacttgaagaatgagattgtggctttatttaaggcttctcaattgatacctttttttggcagtggattta ctagagatattagagcaaaaaatggtaaagttcctgatgctattaaatttacggagttgattaggaatatagcggcagaaaaagaa gggttaacacaaacagaaatagatgaaattctaagaatcagccagcttaaaaaagcgtttggacttctaaatatggaggaatatat acccaaacgaaaatcgaaggcattattaggtaacatttttcagagtgtaaactctctgatcacgaaaagacaaaaataataaatt tagattggcctcatatttcacgtttaatattgacgatgctatagaaaacgttaataggaaatacaaaattctgcatccaaatcga gcagttcagagagaatttatatctgctaataagtgtctattcaaaattcatggcgatattactgaatttattaaatacgaagatca aaatctgatatttacttggcgtgaatatgcacacagtatagaagaaaataaatccatgctatcctttttatctgaggaagccaaaa actcagctttccttttcataggttgcagtcttgatggagagcttgatttaatgcatttatcaagaagcacaccatttaagaaatca atttatttgaagaaaggatatttaaatttagaagaaaaaatagctctttcggagtacggcatcgaaaaagtaattacctttgacac ttacgatcagatatatcaatggttaaataacacacttcagaatgttgagcgaaaatcccccacaagaagtttcgaactcgatgact ccaagttaatgaaagaagaggctataaatttattcgctaatggaggccctgtaactaaaatagtggataataaaagaatcctgcga aattctataactttttctcaacgagatgtctgtgatgatgcaattaaagcactacgtaatcatgactatatcctaattacaggtcg acgtttcagcggaaaatctgtacttttatttcaaattattgaggcaaaaaaagaatataatgcctcttattactcttcgactgaca cattcgatccttccattaaaaactcattgataaaattcgagaatcatatattcgttttcgactctaatttctttaatgcacaaagc attgatgaaattttaaccacaagggtgcatcctagtaacaaagttgttttatgctcgagttttggtgacgcagagttatatagatt caagttaaaggataaaaagatattacataccgaaattcagattaaaaataacttgattaatgaagaaggtaactatctcaatgata agctttcttttgaggggctaccactttataaatcttcagaaacgttgttgaattttgcttatcgatactatagcgagtataaaaat ttagactaagtggttctaatttatttaataagcaatttgatgaagattcaatgtttgttttgattttaattgcagcttttaataaa gccacatatggtcatatcaacagtcacaataaatatttgatattcagaattttatttcgcaaaatgatagattatttgaattgga gtcaactaacacagatccaagtggagttataatctgcaattcaccatcctggcttttaagagttatcagtgagtatattgataaga atcctgcatcttataaaacagtatctgatttaataatatctcttgcgtcaaaaggatttcttgcagcatcaaggaaccttataagc tttgataaactaaatgaacttgggaatggaaaaaatgtccataaatttatcaggggtatatataaggaaattgcacatacctatcg tgaagatatgcactactggttacaaagggctaagtcagaattaatatcggcacacacaattgatgacctcgtcgaaggaatgagtt atgcaagcaaagtaagactcgatagtgccgagtttaaaaatcaaacttattacagtgccacattagtattagcgcagttgtctgca agggctctatctataaataatgataaaatatatgcgctgagcttctttgaaagtagcctagaatccatccggaattataataataa ctcaaggcacataaacaaaatgatggataaaaatgatggtggctttagatatgcaatacaatatcttaaggataatccattaatag aactccttcctcgtaaggacgaagttaatgaattaattaacttctatgagagtcgtaagaaataatcatccttaaattaataaatg gcaagtaactcattcccttgtcatttattaaactcttaagagccttatcccgaaaagtattaatctgagctaataagattgttttt cagctatgtcattattttattgccaatatatttacacttaagcattgacaggtagcggatagttatttttggcttgtaaataagcc ttttaataatagaactgtaagacaatcgctctgatttttttgaaatttatctcaatgttaaattcttccgcttttggcacaaacggg ctagagcagacagatttaatgagataagggtatagatgaattctccatacccttgaacgattacttcccagttgatttgcttggtt tcagtcctggggtattaccgggtgtatccttattatcacgtctgcgttgatcgggttttcctgttgattttgcaattggttttgga ccaggtttaagccccataatcgtactccttagccatgtcagaggttattcctcagtgtggatataaggggagcggtaagaattatc aagcttggatgggcggtgaaaaatgactacttgactattatgtgagcaatgtcagcttttgacatttagaggccagcccattactg aagtaagccaaaaatgagtcgcgatgagccctcaacaatgagggccacctcggagattg (SEQ ID NO: 58)

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
62	pLG064	gacagcttccagggtatcgtggacgcgtcatgcaaagagatggggatgagggattttaatattctaccccttgtaccccatgccagtggtcgacctcataaatcattgattttaaaagcctcacttagggcgctcgctgccaccgatgccccacgatgcctgacgatcttcaacgactccccgcaaaagtccctatgcctcggaaaagccgccaaccccaacaacaccacctaacaacaagaaacaggacctcgtgccgagcttgttagcgcgactgactagccgtccgaaagcaaaaacaccgcgagccaaacaaggcaatttcttgccccccctaaggaaccacctgaggattgaacaccagcgcagcttactgtatataaaaacagttaaagtcctgttctcaggctgcatctggatcacacagccgccgttactcggaaacacggcggattagcgcgcacgctcaggccctccagccctaacggaatatgaatatccagaaaatcaaacacatatcagcctcacgcagcgcatagcgccctgccagaacacagcaggaagtcattgcgtttgcgttcctggcaatccatcattcacggttagggcccctataagacctgcagaagcagcgcgccatgggcagacccggcaaaagcccccaaacgggtgtggagaagctttatggagaaggaaatcccccacgaaggattcacaggctctagtaaagagccgctccagacgctccttccctttaatatcgatgaacccgggcaggagcccatgaaaatccaagatttccccccactccccgcctccgaacagccgttgatgtttgcagacttgtttgcaggctgtggtggcctgtccctcggtctctcactttcaggcatgaacggcgtgtttgccatcgaacgcgacaagatggctttctcgaccctatccgccaacttgcttgaagggcggaaggtgccggctccgcagttttcatggccctcatggctaggcaagaaagcctgggcaatcgacgaggttctcgaaaagcacccgattgagctcagtcagctaaagggcaagatccatgtcttggcaggaggaccaccctgccaaggtttcagctttgcaggaaaaaggaatgaatccgaccccgcaacaagctgttcgagaagtacgtcgaaatggtccaggccatccgaccatcggcccttgtcctggaaaatgtccctggaatgaaggtggcgcacgccacaaagaaatggaagcaactaggtatctcgatcaagccccagtcctactacgacaagctggtagagagtctggacaggatcggataccacgtccagggcaatatcgtcgactcctctcgcttcggggtacctcagaagcgcccacgcctgatagtaattgggctcagaaaggacctggcccagcacctcgaaggcggggtagcccgagcctttgtgctgctagaggaagcccggctcaagcagctacaagagttcgaccttcccgaggccatccatgccgaggatgccatctcggatatggagataggtcacgcgggaacgaggccctgcaatgaccctgactcccctaggaaattcgaagagattgcctataccggccctcgaacggcgttccaaaggctcatgcatcgaggctgtgatggcaccatcgatagcttgcgcctcgccaggcacaagccagagataaaggctaggttccaggcgatcatcgacgaccccaactgtgccaagggcgtacggatgaacgccgagatacgccaagcatatggactcaagaaacaccgcatctacccaatgcaggccagcgctccggctcccactatcacgacactgccggacgatgtcctccactacaaggagcccaggatactgaccgttcgggagtctgctcgactgcagtcattcccggactggttccagttccgaggaaaattcaccactggcggtagccaacggacgaaggagtgcccgcgctacacccaggtgggcaacgcggtaccaccttatttggcacgcgccgtcggcttggctatcaaggcaatgttggatgaggccgtgatgctcgccggccaacaggcagagcgagaacaagaagagaaaatgatagccatcgcttgaacacataggagtcgaggggaatggatagctcccaactggaaggggcgcaatacccggccgcgcttgtcgactgggccggccatcactcaggaggcgtaaaaaggctgctggataaaaatagcggccagcctaacaagcagctgctacggacgaacctttgtcccgtctccaggcctgggctaacaggcttcccaccgagacctcagctgtccccaggattgtcctgcttgtgggtggtcccgggaatgggaagacagaggcaatcgagtgcaccatccgctggctcgacgagagcctcggctgcgatggccggttggtcgaggaactctcgaaagccttccatccctcaaccggctccgcagtccccggctggccagggtagatgccggcagccttgccaagctagatagcagactgagcctcgacattgtccaggatgcctctgctaccgccgggcatgagggaagcaccgcccccgtccttcttatagaggagcttgccaggctactggatggacctccgacccaagcctatctctgctgtgtcaatcgtggtgtcctcgatgatgccctgatccacgcaatagacaacaatctggaacaagcacgaactcttctcgaggcggttacccgggctgtaagcctggcgtacaacgcgccttcatgctggcccctcgagggtttcccatccattgcagtctggccgatggatgccgagtcgctcttggtaaagccggacgacgagcccgtagcccctgccgagatactcctaggccaagccactgctcccgatatgtggccagcgaaaggggaatgcccagcaggcgacaaatgccctttctgcgccagccaggccatcctcgcgcgggatgagaacagggcatccttgctgaagatattgcgctggtatgagctcgccagtggcaagcgttggagtttccgggacctgttctccctcacctcgtacttgctagcaggccaccatcctgtagtccacgatccctcagggactccccaccagtccactccttgccaatgggctgcgaaccttgtcgacctcgaccaaaaggccctaacggcgaaaaggcatggcaagcagtcgctaactgccattttccacctgtcgacttcgagctaccaacatgcgctcttccatcgctgggacaaggacgcagctacctcgctccgccgcgacctcaaggatcttggcctcgagaaggaactcgagatggaggaagggcgaaccctaatggggcttgtctatttcctttcggagcgcaaaagccactatctcccagcgaccatcgcccctctgctggaggggctggtcgaaacgctagatccagccttcgcaagcccagacggagaagttgcagtcagcagtcgaaacacaatagtcctcggcgacttggatatgcgtttcagtcggtccctggccggaggtattgaattcgttcgtaagtaccaggtgctatcgccaaacgagctcgatttactccggcgcctatccgcatcagacgccatgctttcgttaccgagcatacggcgcaagaggccggtggccgccagccgagtccagcacgtcctccgtgatttcgcatgtcgcctagtacgcagaagcatatgcacccggacggccatcgtggcggacgctcccattctcgaggcattccagcaggtcgtcgaggacagcgacaagcaccatcacctcttcaaggtggtaaggcaagtaaaggaattgctgaacactgggaaggagttcgaggtgtcactaaccactacctttggccaaccactccccccctcgacaacgccaggcaacgctggtcgtcccgcagagcccggtccggatgtcccccagaacaacaagggacgccctcacccaccgatttgctatctccatgtcggccaagggcaatcagtccagccagtcccactgacctacgaccttttcaaagccgtgaaggaactggaaagagggctctcacctgcatcccttccacgcacagtcgttgcactgctggacacgactaaggcccggctttccggcccgattgtccgcgaccatgaactactcgatgatgcccggatccgcatcggcgcagatggcacggtggtcggccgctcgtggaatggttttgctgaaagccgggaggacgacgtatgagccttgcggatttcaagcagaccccgtggagcaaatcacatccgaactaccagaagtcggccctggcaatcagccctgcccctgagtatgcgagctcggaagtcctgcttgcctcgctctaccgaaccataggcttcgcaacagccagcgagggcggcgtgccgcaggccgggcgagatctagacaagcgtatccagaaactccgcgagaaacgccaatccccaccaacaggagcggtagtcggtgtagaggcttggaatactgtgcttcacgggatcctggagagcccgaagcttcccaaccagtcgtccaagcgtttcctccaggtaacgcccatcgtacccggggccgcactcttctccgggtctgcccgtctgagcagcaactcgtggcccgcaggcagcttgattcgccgcatggtctgcctgggatcgatggatggggagacggcgcaacgactttggcaacgcctcttcgctgcattgaacgtggacgacgaggacgatgtcttcgcacgctggcttgaccaagagacatcggcgtggaacccgggagcaagcaactgggcactctcgccaatacccgcggacgagatggtcacgttggagacggcagatttcctggggatcccctttctccccgcccggcgatttaccaaggacctacaggccatcatgcaggccaagggttcaatgacccgccggcagtggactagccttctcgaggcattgcttcgcctggcagccgcatcccacgtgacgtggctgtgcgacgtccacgccaggacttggagctgcctgtgggccgcactaacggatggcattgctccttccagtgaactggaagcaagacgggcgctgttcccggaagccccgcagtacatgacgtacgggggaaaagccctccaaggcatcaaggacaaggtgtctagctacctaaatgcccggctgggaatcaatgccctcctctggtctctggcgcagataggagctccctattctggcaacctctcctcgagcgccggaattgctgcactttgccagcatattcgtcagcacaaggccgagcttactcgcctaggcacgcttgagacgattgccgatgtgcgcgagcaagaagcccgtgcgcttctttgcaagaaaggcatcggctctaacctgctggagtttgcgcggcacgtccttgggcaacgccaggctgcagtcccattgctgagggggtacgaccagggatacatcctgaagaagaaaggcagcagcccgtccagcccatgggttgtctccctcggccccgtcgccgtgcttgccttggtccactgcgcccttgcaggaatgggcggtccccgctcggtccaccggcttggacagcacctagaggcttatggcatggccgtggacaagcatgacattggcaggaacgacctgggccaccagttgcgaatgctcggcctagtgctagatagccccgatgccgaaagtggcatgctgctactcccccgttccccataaaccaagccagccagggcccggaacatgaatagacttgcacactggcttgccgccactgtccacgagaaagtcaggggctcgacacaagggttcggaggtaccagcctagaatatcggcttatcttccgcggcccacccctcgagctactcgaaccggcctacgacgagctggcccgcaacggagggatccaggtgccaagcggggcagacggaggactggtgaccctgccggtactgctccagtatccagccggccagctgcagggacccaggccacgcatcggagcatccggtaagtgtgacaacgaccacttgcttgatatacgcaacgaccctgccaaccctagctttattgccctggtcccgccggg

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		actgcacaacaacctctcgatcgagtcaaccaccgacgaattcggattgggggcagccaccagcacggggcatgcatccttcgaacaatggtgggaggatggctttgtccagcaagcagtcaacgaggcgttgatcgctgccggcataacggacgcccagagggatgacgccaggggcctggtccgcgcaaccgcagcctcggtcgacgaggtggatccagacaagggaggtcatcgcgcggcctggcgcctactctcgcgcatctactcgatagcaaacgtgaatcaagggttgcctgcaggaacagcgctatcactggcatgtggtcttcccccaatgaaggagggaggaatttccgccaagactcagctttcggtcctgggaaaaatcgccgacgagcttgcggacggtttcaagactggcatcgagcgcctggcacaaggcgtccaacaaggggttgcgcaagcgctgcgcgaactgctttcccatctccactcgaattgcgacgtacctacggccttcgagcgtgccacagcggctttctacctgcccagtgccgatattgaactggcgcctcctccatcctggtggaccacgctcaccaccgagcagtggacggaactacttgccgacgagcctgacgaggtcgtcggcgagctaacgatccggtgtaccaatagtttgatccctatggggaaaggcttgccggccgtagtacgggacaaagtcgagctattgatttccacaagcgaagagagccaaccaaaggagctcctgttgacaggcggatcctacggcaaggttccgacgtcattgccagcgggccctaatgggactaccagccacattgacctatttccctcctcccacaaagcgccaatgagctacaaggtttccgcggacggctgcaagcctgcgagcgtccgggtcatctccctcgcgagctggaagcccggaatactcgttacctgcaggcttgcgacaaagctctcgccaccgaggaagccccgcaagaactcagctgcgatggactgggaaacatccctgtcgctgccgggctccggtcgttatgagctccagctccaccttgctccgggggcgagcattggaaaggtagaaggcttgccggacgatgccaccgaattcgaggagcagcgggagacaatcgaaccacggcaagttggggaatacgagtatctaatagaggtcgaggctgatggcaagtaccagctggacatcgcctttactgaagccggcgagcaagttccgaaggtctgccgggtatacctgacctgcgaagaggcaaaggaggaaggttgcaggagcgaattcgagcggctcatcaagctcaaccgacggcatctcgagaagttcgataccaaggctgttgtccatcttgaccggaacgcacgctcctccagcctgcagtcgtgggtgctggaggatcagaacgtatccaattccttcaggccactggtgatcgcggacgactatgcgtcccggtgggcccctcctgactgggacgccccgcacggccctgtactctcgaacgggcgtttccttcatgaccccgccccgaggccacgagcttccaacctcccaagggcttcatcgaggctcggcaggggatcgcccggtacatacgtggtagcgacgaccaatcggggctccttgagtcagcgccgcttggtgcctggctatccgaagaccctgggttccgctcccttgtcgaggactaccttggagcgttcatgtcttggctggacgccgacccgggtatcgcctgctggatcgacaccattgccgtctgctccctggagccggatggtcgtaccctgggaaggatcccagacgccatcatcctttccccctgcacccattgcgcctcgcatggcactgcttcgcccagaaagtactccgtgacgaggccgagggcgaagccccgtgcccggcagcaagcatcctcgatccggactgcgtccccgatctadgaccatctcgctgcaggcaccgggaggagtggatcaggtcgacttcctttccgtcgaatgcagctccgactactggtccgtgctttggaacggatcccggctgggacaaatacccgatcgcgctcgccgggccccgttcgacagtagcttcgggctggcagttggagggatatcgagcgggttcagccccgcccaggtctcacgagcactcgacgacgtcaccgacctcctggcagccaagcctatcgtcagcctggtagtgtccagcgcaggtggcaccacggatgcatgcaacgaagggttggccacctggtgcaccaagcgattcggcaacggggaccatgacaccccgcggcacggtgtcgggccaaggattgtggaggtattcgataccaggcaggctggccggcccgaccaggcgacgatcgccaacctctccgaggacacaggcaaccacgtccgctggtatgacaagcaaccaactgggtccaagccagacctgggcatcattgcccaactagattcggcccaacccgaatccaaggaggtcggaatgctttcgccgatgggaaccggcggactgatcaggcaccgcgtcaggcgccaactccaagcctccttcctaagtgaatcccggcagggcctgcagatgccaccctccggcgaaccgttcgcagataaggtttccgcatgcatgctcatgatggaaaggctcagggacggcaaggtcggcctgcagttctcccctaatgtccatgcagtgtccagcatgctcgaggaaaacagcgctgggttcgtcgctgtatcgtcgtcagcaatcgaccccgcctgcttcctcggaggctggatacaagggacgtatctatgggactacgacctcccctcgtactcgcatcgcgcaggcgacacaagcggctactacctgttatcacaggtcaagcaggctgatcgcgatgcgctacggcgagtcttgaagccccttccgggatgcgaggatctggacgatgatcaggtcgagcaaatcctcctcgaggttgcgcggaggggattcctacggtgcgaggcctctccggggacgatacgggggcgacggcgaccttggcctgttcctcgctgtccggctcctacaggatcagttccgtgtgacaggcaacaaggaaagcctgctgccggtgcttgccggatcaccggaggactcgacgatagcaataatcatccccgtcgaccccttccggggttacctttccgatcttgcccgctcccttggcaaggagcgcaaggatacctccctgtcgcgtcccgatctgctggtagtgggcgtgcgcgcatgcagcgacaagatccacctgcaccttacgcccatagaggtcaagtgcaggcaaggagtagtcttcggtgcaggcgaatcaaccgaggcactctcccaagccaaggccctgtcgtcattgcttcgtgccatcgaggaacgtgcaggtagttctctggcatggcgccttgccttccagcacctgttgctctcaatggttggctttggcctgcgagtctacagccagcatcaggcagtaggtgggcatgccggccgctgggctagctaccatgaacgtatcgctgcagccatactcagcccaaccccgccgatcagcatcgatgagaaggggcggctgatcgtggtggacgcgtcgctccagagcagcccgcatgatcgcgatggcgacaagtacacagagaccattgtcatttccagccgagatgccggtcgtatcatcgttgggaatgacgcacagtccttctatgatggcgtacgtgcaaaggtcgacgactggggctgctaccctgccaggcaagtgcggccggcaccccaatcgtgcagcccgacatcactcccccggacgatgtccagacgggcgaccccatagtagtcccagcagaagatatccccggggcatccaccagtctggtcgatcagacatctaccggcgtagcggaaccaggggcaagccctgcccccccaactgacgagccagggacagggatcattctctctgttggcaagactgtggatggtttcgagcctcgatcactatccctgaacatatccgacacccggctcaaccagttgaacattggtgtcgttggcgacctcgggacaggcaagacccagttcctcaaatcgttaatcctgcagatatccagggcccgcgaggccaaccgcggaatcacgccaaggttcctgatcttcgactacaagcgcgactacagcagccaggactttgtcgaggccacgggcgccaaggtggtgaaaccctatcgcctgcccctgaatctcttcgacaccacggggatgggggagtcctccgcaccatggctggacaggtttcgcttcttcgccgacgtactcgacaaggtgtattccggcatcggccccgtgcagcgggacaaacttaagggtgcagtccgcagcgcctacgaggtggctggtgggcaaggccgccagccaacgatctacgatatccatgccgagtaccgagagctgctcgcagggaagtcggactcgccgatggctatcatcgacgacctagtggacatggaggtcttcgcgcgctcaggggaaacgaagccgttcgacgagttcctggatggagtcgtggtgatatccctcgattccatggggcaggacgacaggagcaagaacctgctcgtcgccatcatgctgaatatgttctacgagaacatgctacgcacgccgaagcgccccttccttggcacgtccccacagctccgggccatcgactcgtacctattggtggacgaagcggacaacatcatgcgctatgagttcgacgtgctccgcaagttgctactgcagggccgcgagttcgggacgggcgtcatccttgcctcgcagtacctgcggcatttcaaggcaggggcaaccgactaccgggaaccattgctgacctggttcatccacaaggtacccaacgcaacacccgcggagcttggagtactcggcttcacctcggacctggcagagctatcagagcgagtgaagacccttcccaaccaccactgtctctacaagtcattcgacgtggctggagaggtcatacggggactgcctttcttcgaactcaccaaccaagcctgaccaacgcccggcctgcgaatacaggccgggcaaggaggctcctaatgacagacttcctttctcccgcagaacgctcggacaggatgtcacgtatccggggcaaggacacgcagcccgagctagcattacgcaaggtccttcaccggctcggactccgataccgattgcatggcgcggggctactaggcaagccagatctcgtgttcccgcgatacaggaccgtggtattcgtgcatgggtgcttctggcataggcacaagggatgcaatatcgccacgatccctaagagcaacacaccccttttggctggagaaattcgaaaagaatgtcgtacgtgacgcgcgagtagcaacagatttgcaggccttgggatggacggtacttgtcgtatgggagtgtgaactgacatctgccaaaaaagcccagaagactggcgaacgcctatatgaggttatccgtagtcgtagccacggaaagtatcggtaatcgactgaagcagccctgcggcctgtagtggtctactgatcccggacaccgatttaggcgaaaatcctcgccgtgagagaggtgtccg (SEQ ID NO: 59)
63	pLG065	cgaacggagcaggtagatccgcgctaactgacttgcccaatctggctgcattcgtccaacgctaggcggcttcgcaggaaaagcgaaacggagggagattctacgcgcacctttgtgcagacctgaggctccaccagacctgagagcccggcacgattgactgatcataggagtaaggccaagaagcgacttgatgcgcttgtaaggtaaattctcagcgaatcgaagtaatgacaccgaaacacgtgcggtcgacaaccgtgtaagattgctgataaaaagagcaggacgtcacaagaaatgaacttggaagtagtgccggcgagccggactttcatcgacct

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		cttctcgggatgcggaggtttgtcgctgggactttgccaggctggatggaaaggactcttcgccatcgagaaggccacggatgcgttcgagactttccgggagaacttccttggtgagaactcccgctttgcctttgattggcccagctggttggagcagcgcgcacactccatcgatgacgttttggcactgcgcggtctacatttgtcgaaaatgcggggtgaagtcgacctcatcgcgggtggtccgccatgtcaaggattctcgttcgcgggcaagcgaaacgcgaaggatccccgtaaccagctctcccagcggtacgtcgatttcgtcgagcgactccagccgaagtccctagttctggagaacgttcccggcatgaacgtcgcccataagtatgagcacgggaagagtcgcaagacttactacgaaaagcttctgcattcgctttcaatagccggctacgtggtgtcggggcgtgtcttggacgcggctgacttcggcgtcccgcagcgccgcactcgactaattgccgttgggattcggtcggatatcgcggataagcttgcatgcgcggctagctcgactcccgcagacgtgctcgagggcatcttcgatgcaatcaatcaggcaggcaagcgtcagctcgtccgatatggccagggcgcccatgtcacggttcgggacgcgatctctgatctcgcgattgggccggccgatcacgagaacaccgaagactacgtgggaagcgagcgatgtgcaggctacaggcaggtcaggtaccaggggccgaacacgccttaccagatcgccatggcttctggggtcaccccatccgaaatggacagcatgcgacttgcccgtcatcgtcctgatgtagaaaagcgcttcaaggcgatccttgaaacttgcccgcgaggggtcaacttgagcgccgagttgagggcgcagcatagaatgctgaagcataggacggtgccgatgcatcccgaaaagccggcgccaaccctgactaccctgccggatgacgtcctgcactaccgagacccgaggatcctgacggtccgggagtacgcccgaattcagtctttcccggactggttccgtttcaagggcaaatacaccacgggcggggcgtcccgtcgtcatgagtgcccgcggtacacgcaggttggcaatgcggtcccgccgctgctcgggcaggccattggctcaggattaatggcgtgcctctctttgagttcaacgcgagtgataagggccagtgcgcccagtctcgcgatggccgagaaaaaggcttttgccgtatagcaattagtcagctgcaagaatcgaacaggtggatagacgatgacgaaataccccgatggattgcttgattggtcgggcaatcgggctggaggagtcaagaaactcttctacggcggcagcggccgccccgtcgggaaggtgatagagactcctctactcacccgtctctgggaatggtcggatagcgtcgtccagttcgagccgggcattccgcgggcggtgttgctgttgggagggccgggaaacggcaagacagaggcaattgagcagacgcttcgccgaattgactcaaggcttgcgctgagcggagcgctcatcgacaagcttgcggctgtcttcgagtccaaggatggagtcccccaggacgccttgtggaggtggatcttggggcgctttcagggggggcgctcgagcgggacaatctcgattgtccaagacgcctcggaggggaatccgggctctcctgatcttccggcgcaattgctctgcaacgacctagcaggactcgtcgaagacaacgtgtcaaagcgcatctatttagcgtgcataaatcgcggcgtcctagatgatgccctgatacttgcgacggaaagaggtgacacagaaattggtgctttgctgaagcaaatcatccggtcggtgtcgatggcggcccatggcgtctcatgctggcctctgcagggatatccgggcatcgcagtctggccaatggatgtggagaccttggtcgcaggcgtccagggtcaaccttcacccgcggagcaggttcttcatattgcggccaatgccgaccattggcctgatttcggggcatgcgaagcgggtcagtattgcccgttttgcacaagtcgcaggctcctttccggcgagccccatgcgggatctctcgccaagctgctccgatggtatgagctggcgagcggaaagcgctggaacttcaggacctgttttcccttgtcgcccacctgttggctggaacccctagcaatgccgatgcgtccggttattcgccctgcaaatgggcggcaaaacaactgaatcccccggcggcgacccgcgcaaggccgatgtactccgaaagcgcggagtctttcggttgctggcttcccaataccaacacgcgctctttggcgactggccaatcgagcatgcgtcgggtctccgaagagacatcgccgacctagggcttggtgatttcccggcgcttgtggctatccagcagttcctggcgctggataagcggcgggagtcgacggcaaccctccgtgcccagctctccggcatgtcatccgtattggatccagcaaaggcaagccccaccttcgaggttagggtaagcgctaatactgttattcgttacgaagacttggataggcggttcagcctgtccatccaaggaggcagagagtacctccaagaatatcagtgcctctcggagatcgagatttcagcactcaaggtccttgaggaggccgacaataagttgtctgatcacttagtcaggcgatctcggccggcgacagcaattcgagtccaggcgcttctgagggccatcgcgtgcaggctggcaaggaggtcgattggcgtcaggtgttgtgtcacaaaggatgccgacgtcctcgaggagttccaccgcgtcaccaatggcgattcgtcggcgctgcagcaggcgatcaggcaggtcgaggcacttctcaacgtcaatcgccggttcgttgtttgtctcaacaacacctttggtgagccgctgcctcccccagagcggcgcgcgatgcttaccacggacattcagcgcgttaagccggtgcccgccttggagggtgttgagcggccgagatcgccgatgcccttcctgagggtcggcgcacaaggcaacgccaggcccatagccctgaccttcgatctcttcaaggcgacgaaatcccttaggcgtggcatggtcgcgtcgtcacttccgaggtcggtggtcgcgcttctcgatacgacccgagctggtcttgcgggagcgatcgtgcgagacgaagacgctctggaaggtgcggagatccggatcggaatcagggatgaggtcatagtgcggacctttggaagtttcgtcatccgccaggagggtgcttgatgtccatgcaggagtttctcgcttcaccatggaagaaagaagcctcgcaccgagccttcaacgaatcctctttggtatgaggtctgccccggagttcgcaactggcgaggtcgtcctgtcttcgctctaccgcgccgtcggctttgacggggtttccgaggagaaagtgccctcgcttggcaatgatttcaggaaggcgctggacaaggaacgcagaaagcagaacgcagctggtggtctgagcccagaagcctggcgcacggtcgtggatcgtgtcgtgcaaagtcctaaggttgcgcagcaatcctccaagcgattcctatcgctgtccccggtcgttcccgacgcggccatctactcgggcgccgcgcgccttggaggaaactcctggaacccggggcggctgatcaagcaaatggtcggaatcgggtcggagaccatggagggcgcggaaacgctttggggcgaactctacgatgctttgtccgtgacggaagcggatgatgtctgggcaagatggctccaaacagaatttagtcccaggcgcccagagcaaatagcgtgggccccaagaccgatggatcaaccagatttgcttccgcaatccgatagacggggagtttcctatcccgctcggcagttcgtggtggacctgcgaggaatcttggatgcgaagtccgccatgacgcggcggcagtggatcacactgctcgaggcgctacttcgaattggatcggtcagccatgtgctgtggctgtgcgacgtcaatgaccgcttgtggcgtgcgatgcgtgcggcgctcgagggcgaggcgagtggcgtgcccgccgatgccgccgccataagaaccgacattctggccgtcaggcggcggacgctctcgttcgggaatcccgctgtcccagcgattcgggacctggcctctcgatacctatccgcacgcctgggaatcaactgtgtcctttggacgctggacgaacttggcgtgggctcaagtcgactttgttcgtccgaagaaatccttgacttcatcaagagcgttcaggccaacgcaggggggctcaaggcccgtggcgtcatggatgccttccattccctgcaagacaaggaagtcaggaccattggctgtaagaaaggagtcggagcaaaccttctggaattcagccagtacacgcttggacagaggcagacgatggaccaggcactccgcgggtacgaccagagctatttcctcaggaagaacggggatgccaggaacgcgccatgggttctatctctagggcccgctgccgtacttgcgatggtccactcgtgcctacatgcggtggatggaccgcgatcgatacaaaggctttcatcccatctcgggagctacggcatcgagtttgatctccacggcgtcaacgatagcgtccttggaaagcaactccgaatgctcggactcgtactggatagcccggatgccgagagcggtatgctccttgtgcccccgttcgtagcctgaggaaggaggcaatgatgagcacgctagccaagggaattgcaagctgggtcgaaaaagccatggcgcgtgagatcgcgacgctggtggccgggaatatggagtgtcgcgcagtcttctgcggcccgccaaagcacatcctgaatcaagtatttgggcatcttatccacggtcgatcgctgatcgaagcgacaagggccgatggtcaggcggttcagtatcccgtgatccttcaggtcgaccgcctccctacagggtttcccatcggctccgccacacagtcgggatgccttcagttccatggactcgctgccgtcaggaacgacaggaatggtgttttcctagttcttgtcgagcccggtgctcaagcgagcgatacgcatgaatcaactcgaacttcgcttggactcgagccatcggtaaacgagggcggtgcctcgatcattgcctggtggtctgatccattcattcagtcgcttgttgattctgccctctcagaactctccggtcgcgacgccgcggctgccaaggatctactaaaggaggcgatgatcgccgccgacgcggcagatcagcacgaagtagcgagagttggagcctggcgcgtcatcgaacggttgtgggagctaaaagaacgcggcttgtctcttgaccaactcgttagcttggccgccggattcccgccctctagcgacggaagtattgaaccgagatccaagaccgccatcctttcagccatcgtggacaggatcgaagccgagaacttcggtggcttactgtcgtcccttctgcaaaaagccagggacgatatcgaaaaagaacacatcaccgcgtgcctctcgaatatgaggggcaggtgcgatgtggttactgcggttcggcgatgtgcgccatatgcgtacatgccttcggacgccatcgctggcgaagtctggtggaagtcgctcactgtcgagcgctgggaagagttgctcgatgatggcgctctacccgatgcgggcggcgacatcattattcagtgtgccaatccgatgatttcgcaccttaagggcatggttcccgtcgtcaagggatccgtgcaacttaggatcgaggttccagagaagtacgtgggcaggcggttggaggttatccgcgaggtcccgggtgcgaaggcggcgacgaaggtttggacagttgacgcggaacgcatgatccacgtcgaggacgacgagatcccccccacaagagtccgatgaagtactcggcaagcctcg

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		aaggatcagccggaaagaaggcgagcgttcgaattgtctcaatggatggctggctccctggggtggttgcctctgcgacgacggcgacaaaaggttccctcccgaaacgctcaaaagcagcgaagttagaggcgtcgctgtctctctccgggcaggggaggcactaccttgacatctacttaaggccgggcgtcgagctcgcgtcaatgctcgccaccggtagtgacgaggaaggaaatccagacccgtccatcacggcgccaatcggcatggtcgcggagggcgagttcggggtcgaaatcgaaatcgaaggggaatgcttcttcgacatcacgctcagggttccggaggttgcggatgatcaggtcatccggatcgaattgtcggcggagcaatcaagcccggaagagtgctcaagccacttcgaattgcagctccttaagaactctagcggtcggaagcccagcgcggtccacgttaatgctcagctaagaagtgcgcagcttcaaggttggatgctggagcaggggcgcgctggtcgctcctattatcccttcgttatggccgcggactatgccgccgactggcacaggcgggactggactggcgcagatgacacgatcttctcgaaggctagcttcctgtgcgatccccggccctcgccggaagaaatggcgccgccgcaggctttcatagatgccagagccgcactggccgccaggatcaggggtggtgacggaaatggcttggtcgaaggtgtgccgctcggtgagtggatggcaacggatcccgatttcgccggggaaatagacgtctacttgaaatcctacatgcactggcttgcgagcgatccagatggggcggtttggtgtgacgtagggttggtcgcgcggctcgagcctaacggacttaccttggtgcaagagccggatgcggtgatagttagcccgatgcatccggtaagacttgcttggcactgtgtggcccagcgagccatgttccttgccgcacgaaagagaccttgtccagccgccagcatcctcgatccggattgtgtgcccgatgcgatcactctcccactgagaaacgccatgggtggcaagaccaacgccacttttttctcggtcgaatgcagttcggactactggtcgattctttggaacgcggggcgcttggaagccctttcttcacatggggcgacagccccgcttgaccgggagtttggcctactcgtcggcggaatctccggtgggtttagtgtttcgcaggtgcacaaagcgctcgaggacatctgttcgatgctggtggcgaagccggtcgtcggcgtcctggtgtccagtaccgcgagccagaacaatgcgtgcaatgaaggtctgctttcctggggcaggaagtacttcggcggcggggatagggcggcaggcttggacgcctgggtcggggccagcgaggtcaggatctacgacgacagaccggaagatgcccggcctgatgatgcggagatttcaaatctggccgaggatacggcgaacgccgtgcactggtattccggcacggtggccggcgaggctcccgatctagcgatcatcgcccagcttgagacctccaatcccggtgcactcccaaccaaactaaattctccgttgggcttcggtgggctcgtgaggacccgaattcgggagccttccagcatggcggggggtcaactgctccgtgagtcgcgcatgtctggtcccgcggcgcccactggcgacgggctggccgacgctgtagcaagtgccatctcgtcgctcgagaacatctcggagcaacgccttggttacgtattcgcccctagcattcatgtgatcaagggggcgctggagagcgcggaatttgccgcagtttcctcttcgagcgttgacccggcctgctttctcggaagttggttggagggcacctatctttgggactacgagctcccgtcgtactcaggtcgtgccggagacagcaatggctactacttgttgtcacggatcaaggatctcgacctcgaaaccctgagaagcgtggtcaagaggttccccggttgcgaggagatgccggaagccgtgcttgctggaatagtcgaggaggtcgcacggcgtggtattccaaccgtcaggggcctcgccgcaggtgattctggcgcgacgggtgatttggggctactcgtggccacgaggctgcttcaggatagcttccgggcggccgaatcaggcgctggtctcctgacgccttggcgcagggagggagacatcgaagagcttgctctcgtcattccggtggatccattccagggctatcttgacgatctcgcgaaggcgctaaagcgccctacgctccaccgcccagacctattggtcgcgacggtgcgaatcagtgacctgggagttcaggtccgactgactcccatcgaggtcaagaaccggggtgctggagcggcgatgccgcaatccgatcgagaagccgcgcttgcccaggcacgctcgctggcatccctgctagatgcaatgctggcaacgtattctgaggatcaagagatggttctctggcggattgcgcaccagaacctcttgacctcgatgatcgggtacgcattccgtgtttacagccaacgtctggcagcccaaggcaagtcgggagactggtcgcgcctgcacgcacgagtcatggaagcaatcctgagctcccaggccgatgtgcgggtggattcgagaggccgcctgatcgtgatcgatggctctagccaaagtggtccgagggatacagatggagatggtttccacgagactatcgagctctcgcacaaggatgctgcgcttttcatccgtggcgagcacgatgcgctctgcacggccatgaagcagaagctaggtggctgggaaatgttccctgaagggagggatgccggactctccaatcaatcgccgcccgtggcccatgagactgcgcccttggtggatggcggcgttgaggtgccgtcccttcacgcgctccaagcaacggcggggcccgagggcagctcgctgccgtcttcgggagtcgaagccatgggcgcgtcgcagccggcctccccgggagccatcgacgtggatggcggcatggcccagtccgggctgatcattcgggtcggtgaaacgatcgatgggtttgagagccaaattcggcggctgaatcttggcaacacggccctgaaccaaatgaacatgggagtcgtcggcgatctggggaccggtaagacgcagctgctccagtctctggtttaccagatagccaaggggaaagatggaaatagaggtattgagccgagcgtcctcatcttcgactacaaaaaggattactcttcgaaggagttcgttgatgcggtagctgccagggtcattagccctcatcaccttcctctcaacttgttcgatgtttcaactgcatcgcagtccatcaatccaaagctcgagcgctacaagttcttctccgacgttctggacaagatctattcaggatcgggccgaagcagcgagaccgccttaagaactccgtcaaggacgcatatgtgcaagccgccgaagggcagtatccaacgatttacgacgtccatcgaaattacgtagaagcacttgatggaggcgcggactccctgtcgggaatcctaggcgacctcgtagacatggagctcttcacgccggatccaagtgtcgttgtttcgtcggccgaattcctgcgcggagtggtcgtgatatcgctaaatgaacttggttccgatgaccggaccaagaacatgctcgtggccatcatgctcaacgtcttctacgagcacatgctgcggatacagaagcggcctttccttggggagaaccgcaatatgcgtgttgtcgactccatgctgctcgttgacgaggccgacaacatcatgaagtatgaattcgacgtcctgcgtcgggtcctcctgcagggacgtgagtttggcgtcggggtgatcctcgcttcgcagtacttgagtcacttcaaggcaggtgcgacggactaccgggagcctttgctttcctggttcatacacaaggtcccgaacgttcgtccgcaggagctttcggcgcttggctttagtgatgcggtgggattgccgcaattggcggagcgtatccgtagccttggcgtccatgaatgtctctacaagactcatgacgtgcaaggtgagttcgtccgcggcgcgcccttctacagacggggtgagtgggccaaggaatgacttttcgtcgtgtcgatttatcgcctagttacgcttttggtcttaagttgcgttcctaagagaggtgggctgtgtccgacaatgcgtattacgtttatgcgctgaaagatccacggatggcgcccgcccagccgttctacataggtaaaggaaccgggacgcgctcccatgaccatcttgtaaggccagacgattcaaagaaggaagcaagatctccgagatcatggcctcagggcgtcaggtgctggtaacccggctcgtggacgggctcacagaagagcaagcgttgagaattgaggccgagcttattgccgcttttggcaccctcgatactgggggatgctcctgaattccgttctgccaagcgggttggtaaacaagagccgtagctcgctggttgtcccgtctggcgtaagggagaaggctcagattggtctggcccttctaaaggacgccgttctggagctggccaaggcgaatccgactggtatctcgaactccgatgctgcgagcatgctcggcctgcgtagcgactacggcggaggatcgaaggactatctgtcgtacagcctcctcgggctgctcatgcgggagggaaagctcgctcgggttgccggcactaagcggcacgttgctcaagtgagctagctgtggggttccggatcgggctggcccgctcggcgctgcgctacgaagctcgcttgcctgccaaggatgctgcggtcatcgaacgcatgaagcactacgccgcgctgtatccgcggttttgctatcgccggatccatatctatctggagcgcgagggcttccatctcggctgggaccggatgtt (SEQ ID NO: 60)
64	pLG066	gatggactggtactgtagattcaccgtggaccagcgaatctattatgtggtgagcagaacattaacacatcaatgtaacgccgtaatcattgagtctttgccggggacgcttgacatctccgaaagaattatatcgtgagtcttaaggggaatctcttgcttccggttatacatttaaccggatctagctataagactgttacatctattgggattaggtcaggacagatagcctgaaagcttttatagtgagggacttcagaaataccctagaaaaggaactgttatggtaggttcgcgctggtataaatttgattttcataaccatactccggcttcgcatgattacaaaattcctgacatcagccccagagagtggcttctggcttatatgaaacagcatgtcgattgtgttgtaatcagcgatcataacagcggagcctgggtcgacgtgttgaagggtgagctggagaatatgtcccgggacgccagcaccggcgacctgccggaatttcggccactgacactctttccgggggttgaactgacagcgaccggtaacgtacatattctggctgtgctgcacacgcacagtacaagtgccgatgtggaaaggcttctggcccagtgcaataataatagccccattccgagtgaagtccctaaccatcagctcgttcttcaactgggccccgccggcatcatcagtaatatccgccgtaatccgaaggctgtttgtattcttgcgcacattgatgcagccaaaggtgtcttaagtctgactaatcaggcagagctcaccgcagcctttcaggaaagtccccatgccgttgagattcgacaccgggtggaggatatcaccgacggaacccgccggcggctgattgataatttaccgtggctacggggctctgatgcgcaccatcctgaacaagccggcgtgcga

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		acctgctggctgaaaatgtcatcccctgattttgacggactcaggcatgcactgctcgatccggaaaactgtgtgctgtttgatcagctccctccggaggaacctgcgtcatatttgcgcagcctgaaattcagaacccgccactgccatcctgtgggtcaggattcggcctcggtggaattcagcccgttctataacgctgtaatcggctcaagaggcagcgggaagtccacgctcattgaaagcattcgtcttgcaatgcgcaaaacagaaggtctcactgcgacccaggggagtaagctggaccagttcattcggacggggatggaagcggattccttcatcgaatgtattttccacaaagaaggcacagatttccggctcagttggcgaccagacagtaagcatgaattacatatcttcagtgacggagaatggatgcctgacagtcactggtcggctgaccgttttccactctcgatttacagccagaaaatgctctatgagctggcttcggatactggtgcattcctgcgcgtctgtgatgagagcccggtggttaacaaacgggcctggaaagagcgctgggatcagctggaaagggaatatctgaatgaacaaatcacgttgcggggcctgcgtgccagacagggaagtgcggattcgctgcgggggaattatcggatgctgaacgtgccgtcagtcagctgcagtcaagcgcctattatccggtttgcagacagctggccctcgccagaaacgagctgtccgcagcaaccttacccctggagcactttgagcggcgtattgcagccattcaggctctggcagaagaaccgctgcagagatccgatatcccgccggaaccttccggtctgctgatggcatttatggcgcgcctgtcatctgtgcaacagcagtatgaccagcggctcaatactctcctggcagaatatgctgcagagctcgcgggtatcaggagagagcaatcttttattgccctccgaacagcagtgagtgaccaggaaacaaatgtagaaagtgaagctgtttccctgcgggccagagggcttaatcccgatgttctcaacgaactgatggcacgctgtgagtcactgaaaaatgagctgagaaattacgacggtcttgatggggcgatctctgcctctgttgcacggtctgagcagttgctggctgaaatgcgtgcccacagaatggcattgacagataaccggaaggcgtttctctcctccctgtcgctcagcgctctggaaatcaaaattcttcccctctgcgccccttatgaagatgttatatctggttaccagacggttaccggcatcagtaattttgccgaacgtatctacgataacgatgacgggagcggattactgagcgactttatcagtgaacgtccgttcagcccgttgcctgccgcaacagagaaaaaatacagggcgctggacgagctgaaagcgctgcatcacagcatccggctggataattcagaggctggggcggggcttcatggttctttccggaatcgtctcaggagtctgaatgaccagcagctggatgccctgcaatgctggtatcctgatgacggcatccacatacgttaccagacccccggggggcagatggaagacattgcctttgcttctccggggcaaaagggagcgagtatgctgcagttcctcttatcctatggcaccgatcctctactactggatcaaccggaggatgacctggactgcctgatgctgagcatgagcgtgatccctgccatcatgtcgaacaagaaacgccggcagctgattatcgtgtcgcactctgcccctatagtggttaacggcgatgcagaatatgttatcagtatgcagcacgatcgcacaggcctgtatccaggactctgcggtgcactgcaggaagctccgatgaaggcactgatatgccgtcaaatggagggggagaaaaagcgtttcgttcgcgctatgagcgtattcttagctgaagaacggaaccgtccttaaggcggccatgaccggagagtgggcctggcggctgaatgcctggataaaagacgcaaatgtcagactgatggcctctgcgtctttg (SEQ ID NO: 61)
65	pLG067	cctggtcctgccaattgctcccccagccatatgacataatccttttgaataatagggtttttatgcttgtactctagcccattcgcggtatcattttacgatctctcttccagttttatgcttaccgcctttgcctatcgtagaacaatgccgggaagcgttatcagcgattaagggcaaggaatgagaaaaagctggactatagaggaagattgtaagctgctaaccttggtgcgtcagctcttttccgcgctggtcagccataaccggctgaatgccacaatgccatttagccagcagctccacgatgcatttgactcacctgaccgcgatgccgcagcattgctttatcgcctcgaacaggcaaaaatcttgggatttgccagccgtcctggtggcgatcccactaaacaactgtttcgctgcctgataagcaatgatttggcgctatacgattacagcctcacctttcccaccctcagaaaagcattgcatccagataccgttgcggcagcactaaaccacttcacgattagcaatccacacgaaccactgtccaatactatcaatgaaatcgcgacagccttgcatcttgcccccatacaggtggaaaagattctgatcgacagcggccaaataaccatcaatagttaccgcaagtgtgagcgtgttggagagaaaaatatcaataataatctgcaagatctcatctctaggcaaattcctgacataacgctgattaaagagattaacgcctgtcgcgcccaagtctctcaactttaccacgtgcatgaacgtgatggcgctgaggtcatcttcagttccgacggcacggggttcggcaaaagctatggcgtgatccaagggtatgtcgaatatctggagcgcttcgccaaaacccaaaagtcagacgatctgtttcctgaaggtggctttaccaacctgctattcatgtcaccgcaaaaatcacaaatcgacctggacagcagtcagaaagagaaaattctggccgctagcggcgagttcatttgcgttctctcccgtaaggatgttgccgacctcgactttatggactgggcctctggtctgaaaaaccgcgaccgctatattcagtggtacgaaggggcgaaaggcagcaaatatatcggcggcgctatgcgttcgctcaattatcatgtcttacaaattgatcgctgtgaagagcagttaaaaaagctgacaacatacggttctcaggataccaactacgaaagagaaattctcgaagaacagctaaaaaactgccgtcacagtatccgcaatacgattgagtcagcctgtaaattactatttggaccagatagtgaaaaagcttccattaaagagtacattcgtcgcgggctccaggcgcggcaagagcgaatgcaaaacgcggagacagcacgaaaaccaggaaagcttgaacctaagataagcgtacacgaagtctatttcgagcttatcaaacaggtattgcctttcgaagtttgccagtaccgcccgtcagtgctattaatgaccacgaataagttcgacacatcaacttaccgactggcgcctcgtcagcgaggcgaaggtgtgcgttttgagtccgtaggtttcgacttgctgattggcggtaagctgactcccaaagatccacagattagcaccgttgcggcagccggtcataccgggcaggttacctatcttcgcgacgaacacttcagacgcaatccagattgtccttttcgccagaaaaatattcgttttacggtgatcattgatgaactacatgaagcctacactcgccttgaagaaacatgccatgtaaagctaatcacacaggaaaataacctggcgcacgttatttccgtcgcaggacgtattcacaacgcggtactcagcttagaacgccgaaacaagcccaaagaagcgcaaacgacctttgagcaagagatggtcaaattcatcactactctgcgcaatttactggcggaaaagtgcgaactatcccccggtacaaggctgggatcgatcctggagatgtttcgtgaccagttagggcatttgaagtcaacggcgacgccgccgaacgcatcatctcaatcacccgcaacgtattcagctttaaccccaaaatgtacgtcaatgaagaagggctgaaacgcattcgcatgcgcaacagcgaaggcgacataacgcgcaccgaactgtattacgaagtcgaaaatgatgccaatgacaccaaccccactctgcacgatctgttccagttggtctccgtcatcctcgccgcctgttctgaaatcaccaaccggcactttaagcgctgggtaaagaatggtggccaggacaactccagcagccagaatacgcctttgggccagtttgttgacgcagccaataacgtagccggcgtggtgcgacatatcttcgatcgcaccaccgataaaaacttgttgattgatcatttctacacttacctgcaacccaaaaccgtattcacgatgacgccgatagctgaactcaattacgtgaacaggggagccgagcgcacaattattctggcgttcgagatggatctggtacaagagttgcctgaagccatgctgctgcgtttattaaccggcacgcacaataaagtaattgggcttagcgccaccagcggttttagccacaccaaaaacggtaacttcaatcgtcacttcctggcgcactatagccgcgaccttggctaccgggtcgttgaacgcgaaaaggcagatatcgatacgcttaaggcattacgcgggttgagggccagtatccgcaacgtagacttcagggtgttcgatgataagcagttaaaattgaccgatatctaccaaaattgtgaaatctatcgcaggacgtatgacaactttttcgacgcgctgaagaaaccgctggaatacgacctgaaaaatacctataaacggcgtcagtgccagcgggaactggaagcgttactgcttgccgcctgggagggtaaaaacagcctgattctgtcactttcagggacgtttaagcgggcctttatcagcgcctggcgcacgcaccagacaacctggcgtcagcagtacggtatgcactcccggtgcgatgaaaaaacggataacggtaagaaacatgaccagatcctgacctttaccccattcaaagggcgtcacaccgtccatttggtctttttcgattcaccactggctaatgtcgaagatatcaggcaagaaacctatctccagaacagcaataccgtactggtatttatgagcagttataaaagtgcgggtaccggcctcaactactttgttaaataccatgacggcgatattaatgatatcaatgcaccacgtctggatgtcgattttgagcgcttagtgctcatcaactcctcgttttacagcgaagtaaaggacaacagcggcaacctcaatacattacctaactacgttaccgtgcttaaacactacgccgatgacgatattaccgtccacaagctggccgatttcaacgttaatttcgcccacggcgaaaactatcgcctgttaatggccgaacatgatatgagcttattcaaagtcgtcgtgcaggccgtagggcgagtcgagcgtcgcgacactctattgaaaacagaaatctttttaccccgcgatgtgttccgtaatgttgcatttcagttcgccgctcttagtgaagatagcggtaacgaggtggtatcagaaagtatgtctttgcttaaccaccgactcatggaggagtgcgaaaagctgagtcagggccagtcattcaataatgcggaacagcgactgacgtttgagcaagctatcgtcgcgaatggtcgccgcatcgatgaaattcacaaacgtgtccttaaaaccgactggattaataaggtacgcgctggcaatctcgattatctcgagatatgtaatttatt

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.		
Row No.	Vector	Locus
		ccgcgatcctgactcctttaccgatccccagcgctggctggcaaaactccaggctaatcccttgtataccgccaatcgacaaatgc aatctgttcacgacgctctgtttatcgatcgtcagcaagggaatcaaacgattttactttgccacaaacgcggcccggatggactt gcccacagagattattccgccctgtcggatttcgctggcggcgcaagagagtaccggccagagctcaccctctttccgcagtatag aaacgatgtcgattttaccccccggcaacctggtcggcgagttgattcgtgaatgtgacaacatccaggaaaaggcattcaaaaaat gggtacccaaccccaggctagttccgttgctcaaaggcaatgtcggtgaatatctcttcgataaagtgctaaaaagttatggtgtt accccactctccgaccagcaggtgtttgaacgccttgaaccgctggtctatgagtttttttgaccgctttattgaagtgggcgacga cctgctctgcatcgacgttaagcgctgggcgacacagttggacgatttgacgcgggcagaagaaacgcttgagaaaagcgacaaca agattcgccagatccgtaatatcgccagccaaaaggcggatactgaggggcagaaacagctccagacggcgctggcaggccgttat gaacgtattcgatttatctatctgaacgtcgcctacagccagaaccctaataatctgatgtggcaggataatgtggatcacacgat ccactacctcaacctgttgcaaactgactaccagtattatcagcccaaaaatcgagagagcggacgcgctcaggaaaactcgaaac tgcgcatgacattggatataaacccaatgttactaaccctgctgggtgtagaaaagttgccgactaaaggaaaagtatcatgatcc ctaatctgaatgagctgacggatactccgattgcccgtaccaatttgatcaagcttgaagaagatcagctgacaacaatccagcgt ctattggccccggtatctaatatctatacgatagactttatggttcagcactttactaaagagcgaaaagaaaaatccgctgatta ctatgcgcgaattcatcaggaggtaaaaacttgcgtgcggcagaagcttgggcttgaggccggacaggaagtaaaatatgagctgca ttgcttacccaattaccatcacgtctttttttttcctggcgcctgctgctgcaccgaacagcctagcgcatcggactttggcagaacg cattgaaacgctttgccagcgactcacagctgaaaattatgatttatctcgcctgattcagggattgttcagtctgcatttgaaaat ggtaatgctggaacaagccagcgagcgcttttcggtaccgccaacctacttcaactctacgttctatctcaacgctcgcctgagtca gcccgtcacgcagaaaagcggcactggagtgatggaggcattcgaactcgacatttatgcatcagaatataacgaactcgcctttac cctgcacaaacgaaaatttctggtcgaaccggaggatgaattgcatctctctctggacgatacctgcgtgtggtttaacatcgataa tcgtcggctcaaagcccggcgcaaactcgatgcccgggatagcaaactggactttttttcgtgagcgcagcggctatggtgaatgcca ggcctatacctataacgtggtcatgaatgccgcctgcgagcggctcagtgaactagagatcccgcatcagcctatcgcatttcaggc cacccacgaggtcaatcagttcgctaccgacctcgatcaacaactgactaatacgctgttggtggttaataacggcgtcgaatttag cgccacgcaagaagcttatttctttgacacattagccatccagttccccgggtatcaactctggcctctggcgtcgcttaaacattc tcagcaaaccggcttttctgagctgcctgccagtacatctattctggtactcaatgcagtagatgaagagcggagcaacagcatccg ccagcaagataatgaatctgttgagtacaatgatttctatgcggcctttgccgacgcccgaaaacaacccgaactcaattgggatac ttatacccagcttaaactagatcgtttgcaagggtggctaaatcagcaacctctgcccgtagtcttacagggtatgaatattgatca caagttgttggatgcgattgattttattaatgaacaattgacaagcaaccctactcaatacgaaatcgatcttacgaagcctcacag tcgtctcaagtcagcagttaccttacttaacagtaaggttcgccgaacaaaaaccgagctatggttcaaagagagcttactcaatca gcatcacatcccactaccagatttggcggacgggcactataccgcctatgcagtacgcaaaacgaaaagctatctccccctgcttgg atatgtcgaactaaaaatagaacacggccaacttagggtggttgataccgggatcgctgaaggtaaattagactatctgtctgttga tccccccctctctgggacgattaaagaaattattcgacaaaagcttctatctctacgaccacacagcagatgtcctgcttaccaccta caacagctcccgcgtaccgcgcctgattggcccggcgcaatttaatatcgtcgattcatacgcttatcaggaacaagaaaaaactct ggcagagcgtaaaggggataaatttaacgggtacgccatcacccgctctgcaaaaccggatcaaaacgtactgccctatctgatatc acctggccgctcgaaatacgactcgctgaccaaagcgcaaaagatgaagcatcaccatatttatctgcaaccgcatgagaatggtgt atttgttctggtaagcgatgcccagcctacaaatcctactattgcacggcctaacctggtggaaaatctgctgatatgggatgccca aggcaaagccgtagatgtatttagccacccgttaactggcgtttatctcaatagctttaccctggatatgctcaggagcggtgaaag cagcaagtgttcgatttttgccaagcttgcccggttgatggtagagaactagcggaaaatttagggcggtgtttttagaattcgtta tgtgtgaacctaactgatctccccccctgaaaacagtaccagtctaaactgaagtctccggtctttcttcctgctcacagagaggctt attaccatgaaaaagacccgttataccgaagaacagattgcgtttgcgctgaaacaggccgaaaccggcacccgcgtcggggaagtc tgcagaaagatgggtatttc (SEQ ID NO: 62)
66	pLG068	caactgaggcggatatggccggtgcgttcatgtcctgaattaattcgaaagacaaatcgcgttaccaagcgttgcgcgatttagca gcaaattgatagcttagccaccaacatttacacgttgtaggttgtcttggccgccattggtcttcagcaacctgcaacgctgatca gtcgctcagggaagatgaggtaccgcagatggacaagcacgcccccgagcacctgctccgccttctagcccaaggcgcctcgctgt gtggcaccgacagggccgaagcgtttaccgtgcttcaaagagcatccgcattgctctggcgcctggagcctagcgctccccccatg tcagcgatcaagcttgaaaatcagcttagcctacccttggaaaagtggttgccggatgcactgaggctagattattcgggcccact gctttactccaacatcgcgacgcagacctgcaacgaaatgctgcttgagctcgacgtcagtcagctctgggaagaagtccaggcaa gcgtaaatagggtcaagcaggcgtgtcggctgcgggcagagggagaaattcactaccgcaacttcaggcttttcctcattgagcac ggcgtgatctttccgtctgaagcccaagatgtcttcatcccgctcaacctctccctgaacgagttctacgaccccatccccctcca tctgtatcacaacggcttagtctacttgtgcccggaatgccggtggccaatgaatgcccagcggcacgaagtcagctgcgactcag cctggtgccaagacaaaaaaagcctttttgttcgtgaaggtacaagccttctcaaccgtgtgaacaacagcgtgctgcatggccag ccggtcgatggccgtctgatgctcaaacctgcgctgtggaaattcaccctgcagccaggacttatcgaaatcgccctggcgagtac gctggcgggaaaagggtttgatgtgagtctctggccggatgtggatcgaacagacctccgtatccagttaggccttattgagcagg acatcgatgccaaggtttgggtgtccccttacgagttggccaaacacatcgaatcgatcccctccagcaaaccacgttggatcgtg attcctgactatcagcgggagagcattccgtttctacgccagcgctgcaagtctggggtgagtgtatttacccaaagccagtgtgt gaaggaggccctgaaacatgctccccctttctgataccagcgtcatactgttcctcgccttagccgcgcgttacgtcggcaacgaa cccatggtggcggacgcagcagcgctctgcgcgggtcgcacacgaggctggagcacttggtacgtgctatccgagcctgaccagct acttatcgctgaaggcttgcgcctacgcccatcctccgtggcgcagcccaaacgcttcgtgatgaccgcagaggaaattatcaagg gcgaacgtagcccctttgagttagtcgactctggcaagctcagcagtgagctccacgagcaggattgctatcgcgtttcaccccac ctgaacgtcgatcagctcatcagggagcacctagatgcgttgagatatgggcgcccccatcggttcatgcacagattccagactc aggggatgtcgttctcaagcacatcacaggtgatcaggtcagggtgttcgtcgtcccacagagcgagcgaggggtgctcagtggcg cccaccagtacgttactgtcccaacctcccatgcagcccctgagacgaagtgggaacttgacgctctgaacgagctcgcggagtca ctcgatggtgcaaccggattgcacacgaatcatcgaagctcgttggccaacatttggggttcggatccgctacgcacagctgacgc aggtcatttttatcgtgtgaacgcgccgactggcaccggtaaaagtgtggctatggtcatgatgtcgatcgatgctgctcgcagag gacaccgggtggtgatcgcggtgccaacgttggttgagcttgagaacacggttcggattctcaagcaatccgctgcggtgacagcg cctgatatcacggttgccccccctgcactcagcaacacgcgtatacgagcgcggaaagcttcaatttcagcagggtcattctgcacc ggcctacgactatgcctgcttactcgatgcctatgcctcggatacgctgcaagttgaacctggaaaagaaccgtgctttaacgttc gggtatcgacacaggaagaaggtcgtgcagaacaatcaaagcggctgaatcactgccctttcctgttcaagtgcggacgaacaacg atgctgtcgcaagctctggaagcggacgtcgtggtgattaaccatcacgccctgttgtccggaacaacccgcattccattgtccga ctcagaccggtgtccaggcccacgcagcttcatagagctgctgctaagaacagcaccggtgtttcttgtcgacgaaatcgacggtc tactgaagtctgcgatcgacagcagcgtcatcgaattgaagctgggcaatcaaggtgacaacagcccgctgctccgtctattcaat acagtggccggtcgatccagcattcctgagattgatcgaagcagcatgtaccgcgtgaactgggcgcttacctactgcacgctgag

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		tgtcagccagctaatgaacctccagcaagaggaatatttcgagtggccaaagaaagaaaccacttggtcggacgcagacgacacgttcattaccgaaaagcttggtattgatcgtgagacgcttgagcacttgttcaacagcacgaaccgcataccgggctatctggaaaagctgagtcaccaccttgctcactggcaatcaaatgggggccagtacaagcttgaggccttggcaatcaatctgggccatctcgtcaaagagttgtccgacagcgacttgcttcctgcgcgtctcaaggagcacgatcaaatccgcctcaaggcgtcactcatcttgcgaggcacgttagaagcgatcgaaacgcacctgcgcaaccttcaggtcgagctacccagcttcgtgaacgccgaaataccttatgcctacgaggtcaaacggagtatcgcagggccggagccgctgagcccgactccgaatggccccttgcagcgagccgtatttggcttcaaacgtaaagacaccggagacaacgactcaactctgaacgttgtcgcaatgcgtggggatccgcacagcacactgctttcgctgccagatgtcagcgccttgggctatgccggtgtaaagcgattgtttatcggcttctcggcgactgcctacttccccggcgctagcgcttacgatcttcgtgctaaggatttcatcgacgttcccgatgtagctggccaggtgactttcgaaaatgtgcctcagacaaccgctatctctggcgctcagttctcgcagcgaaaattcctggtatcaaaattcgccaaagagatttggccgtggctacgcagccgacttgcaagcttggccaacgaccccgtcacgcagacgcgtgcccgcctgctgctggtcaccaatagcgatgcagacgctgaagttctggccatgaccctggccaggatgcagggcggtcctggtcagctggtaggctgggttcgtggacggcaaagcgactacaagccgtcctcgctagatgcacagcagatgcttgcatacgatgatctcgctgagttcaccaacggccgacacaaggacaaaactctgctggtcagcgccttgggcccaatggcgcgtggacacaacattgtgaacagcgacggattttcagccattggtgctgtggtgatctgtgtacgccctcttccatcgtcagatagccccaacaacaatctggcgcacatctgttacgaaaccagcaagtttgtagcgccatccagcagtccgggcgtattgatgatgcaggaacggaagcattccaatgcgctgctgcaaaagattcgtaccgcccgccccgcgttcagccagcagccggccaacatccgccactacacgatcatgaacatccttgtgagcctcacccaactgatcggtcgtggacgccggggcggcacacctgtgacttgctacttcgccgatgcggcatttctcgaaggtttgaagccgtggcctctgatgcttaacgagagcgttgaacagctcaagcaagacggcgattggaaccagtttgcccgtcatcatgccggcgttgcatcggcacttttgaaatacatcaatggatcagtgaaggacgcacgatgaaggttcttgaattacgcaccagcctctttgagttcgatccagcagctttgggacaaagctaccgcgtcgtggtaggcccgcattaccttgatgcctggcaagctcttcagggactggtaaggaaaccccatcctggcctaccgaccatagggcttgaggaaatgctcgccaccctctctggagggccggtcaaggtgaacctgtttccgcaaaaagaaggaggcgtctcggcgatccttttgctgaagcccctgcccgttgacaccatcaacgaagcgctccgcctttgggctatggacgtgatgcagttttacaaacaagaactgctcgaattcgaaggcaaactggtcgtcaccgacctggtacctatggacactgcccgcttggtcgcgtccggtgacgtatcgtcccttgcgtacacagtcattccttggttggtaggtcaagcgctgattgcgaagccaatgcaagcagcgaaacctcttaagctttatcaggctgccgacgggtgcgtgctcgcctgggacgacccagtcgtttcggaaagcgacgtacgctacgccagtgcgcttcacgccatcgagcctgcattggtgctgatctacggccaatccaagccctatctacagctgcgggtaaagctgactcaggtgatgccgaatctcaagggtcaaaagaagcatgcctgggtcaaaactggcgacctgattgtcaaagcaaaaatccggagcaagcccgacgggcatgggggctgggaaacatttacgaacatcccattgaaaagttgctgacctttatggggttccgtcgtttcctccaataatcgagggcgatatccctgtcgacagcgacgtgcgccctatctacgccattccaccctcgaaccccttgatcgcgtcaggcactggccccctgtttcttgaccaggcaggattccatctgcttgcttgtctaccaaggacaaagccgcttctggtcagaaaatctgtcgctgttctgcgcgaagaaaagaccaatgctacgggcgaggtgatcgacttgaacgtgatggtcttggcagctcacgcagacgtgatgctaaggcttcacggggcgagttcaaacttggccagggacagcaagttcttcaagaaagtcgccccaccacgtgtgacgctgtcacgtctggatgtgccagatgcgcagcgtatgctggagggggcagcatgacctgaacagcctcaacgaatggttattgaatcacgtggttccggcgagcagagtgctcgctcaaaacggcgccaaggtcatgattgttgagaccagtgcatcagcagcatcacgcgaaactggactcgatcccaagcacgtcatccgccgggtgctggcgaagcatggcatcgctacccaattcattatgcacgttgaccccgatgcacaggtgaagaagcgcaagcctaaggaagatgaccgtgatttcaaagcgatcaactcgatcatcgaagcgattcggttgagcggccagcaccctgcccctacacccaaggtcaagtcgatgccggccaacactacggtagtttcagtcctgctagatcgactccaggacaaaggctgggcgaaatttctacccgtgatcacgcgcaccacgctcggtggccacacccctgaaatcttctggtttgagtctggcgcagagtctgcaggcaaatggttcagctacagcgcgggactgactgcgatccatgccacggacacgctgctgacgcctgatcaattgaaaacactgatcacccaagcccttcttgattgcaaaatcaatccagctgactcgttgatcgtctgcctcgatgcagacctgagaactttttatgcaggcctaaaagacagtcctggtgaggggctaccaaccgtaccggacgatgcagcagtagtgcgaatccgtgcggaccatcaggtagcacagatcagtggtagccacaccttgtcgccgcaagcagcccactacattggcacgaaggtcggcgcgttccagtcctgtgagagtccctcagtgttttactttgtgtctccatccaagcagtttggcagcgttcgttcgcagcgtgacaacacccgttacgacgtacgggagagagatcttcgggatccttggcaacagctcggcgtcacggaaattgccatcatccagcctggagcctttgacggtgcagctgcggttgccgagcaagtggcgttgctctgtcgcaacccaccactgtgggatggtcatctgcgcctgcctggcccgatgcaccttggcaagcaagtagctgcagatcatccagttatggaagcgcggcgaaagacagaggctaatcgatcagccggttaaagccgcctggtaaccgttcattactagacacgtataagtcataacacccagcatttcacaaagagcgcga (SEQ ID NO: 63)
67	pLG069	atttgcctgagacttatttcccgtggcgcttagctagctaagagtgggcatcgtgagcaccattgatgatatgaaatgacggtatagcaatttaaccgtctggatttcaccagaaattagtgattcaataggaaattaaatacgttttatatttcaatgtgtatcaaaatcattcctgaaatttcctggtgctatatttgatgaaaacggataaacattctgttgattttaataaaattctgtctttcgatttagagcttacgcgtgatgaaaagttaaggcatatgggggccgtgctggcggaacgcacgttgagtttgaagataaatcaggatgaagcgattcatcaattggatgaaatggcaggcgatgcagatttaatcctcggtcataacatactggatcatgatttaccctggattgccaaacaacgcgtacgtgctcaaatattattagataaaccaatcattgatacccttttatttatcaccgctagcttttcccgcaaatccatacc atcggctgattaaagactataaactggtaagagatagcattaacgatccagtgaatgacgctaaattatcgcttcaggtattcaccgagcaaatatgtgcgctgcaagaaaagccgctggctcagttgcagctatatcagtatctttttgagcacggcgttgccagccatttcagtacacgtgggatggccagcatttttccgcactgacgggtcaggcgtccatatccgccgtagttttacctacgctagttaaatcggttgctcagaataaagcatgccctaaccagcttaatcgggttattggcgatgctcttaaacagcctttgcgcttactaccattggcttttgcctgtgcctggctccccgtatcgggagggaattctgttttaccgccctggatatggcgccgttttcccgtcaccgctgatatcatccgcgaactgcgtgagcaaaaatgccagtctgaaacttgccgctactgctgtgaaaaccatgatgctcgtcggcatttacagaaaattttcgagctgaacgattttcgtaaacttcctgatggctcgccgttacagcgcaatatcgttgagtacggattagctagtcgttcactgcttgggatattaccgactagcggagggaagtctttatgttatcaacttcctgcgattgtcaggaatctgcgaaatggttctttaaccattgttatttcgcctttacaagcgctgatgaaagatcaagtggataatttacgtcataaggcaggtattaaaggcgttgaggccatttcagggatgctaactttacctgagcgcggcgctattcttgagcaggtccgtaaggggatattgcgattctttacctctctcctgagcaattacgtaaccgcgcggtaaaacaagctatcaagcaacgtcagattagtggatgggtttttgatgaggctcactgtttatcaaagtggggccatgattttcgtcctgactatctgtattgtggcaaggttattgaatctttggcgcaggagcagtctgtgcagattcctccggtattttgctataccgcaacggcgaagttggatgtgattaatgatatttgtcggtattttgacaaaaaattatcgcacccattagctcgtttttcaggggagtagaaagaattaatcttcactatgaaatcattgcaagtaatggcttgagcaaaattagtcagattttgaatttgctcgataaattttttctaatgatgatgaaggtgcatgcattatctattgcgcgacccgccgttcggtagatgaaatcagcgatgtgttgacccaacagcaacctttaccggttgctcgttttttatgcccggcttgaaaatagtgaaaagaaagaaatccttgaagggtttattgctaaccgttatcgagttatttgtgctactaatgcctttggcatgggaatagacaaagaaaatgt

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		acgtttagtaatacatgcggagatccccggttctctggaaaattatctccaggaggcagggcgtgctgggcgggatacgctggacgcgcattgtgtgctattatttgatgagcaggacattgaaaaacagtttcgccttcaggctattagtgaagtaagctttaaagatatttatgcaatatttaagggaatcaaaaagaaagttaatgaaaataatgaagtcgttgccacaagtattgagctaattaatcatcctatggttaaaaccagtttctctatcgatgataacaatgcggatactaaagttaaaacggggatagcgtggctggaacgtgttggttatgtggagcgacttgataatataactcaggtttttcagggaaaagtggcctttccttctctggaagaagcgcaaagtaagatggcagcgctgcacttgaatcctgcggcgatggttctctggaatgctgttttacaggcgctattaaatgctaatgacgatgacggacttagtgccgacagcattgctgatgaggttgcccaatttcttccgcataaagaaaataatacgtcaggaattgaagcaaaagatgttatgcgcgtattgacacagatggctgatgttggcctggtcaccaggggaatgctgctgaccgtacgtatgcgccccaaagggaaagataatgcgaggatcacaactgagttaattcacaatattgaaatcgccatgttagggctgctgcgcgaagctcatcctgatattgaactggggatgccatggcctctccagattgcggttatgaatcaagagattattcagcaaggctatgatagaagtaataccacgttactacaaaatatattatttagctggtctcaggatgctcgagcaaacggtcataaagggcttattgattttcgttatggtacaaggaacagctaccagattattatgtatcgtgactgggcatatatcgaaagagccattttacaacgtcatcgtgtgacaagctccgtactgaattttatttatcaattggcattggatagtgatgaaagcagtatcaaaaaagtgatgctttctttctcactggaacaggttatcgattatttaagaaaagatgttgatattattccaatgatccaacagagacagggggggatgagcagcagtggctgatggctggtgcagaacgtgctctactttatcttcatgaacaacatgccattgtgctgcaaaatgggctggctgttttccggacagcgatgagcttgaaattgcaggctgaaaaatcgcaacggtatgtcaaagctgattatgaaccactggctctccattatcagcaaaagacgcttcagatccatgtgatgaatgaatacgccaggcttggtcttgaaaaacctaactatgcccaacggctcgtacaggattactttgctatggatgccgagtcatttgttccactttattttaaagggcggcgaaaaattctcgatctggcaaccagcgaaagctcatggaaacgcattgttgaaaatttgcataatcccgatcaggagcaaattgtgcaggcgagccttgaacaaaatacgttagttcttgccggaccaggctcagggaaaagtaaagttattatccatcgatgcgcctatcttttacgcgtgaagcaggtcgacccgcgtaaaatcctgttgctctgctataaccgtaacgcagcgatttccttaagacgcagattgaagtcgttgcttggtaaagatggcgccagcataatggtacaaaccttccacggattagcattgagccttacgggataccagattgagcggaaagataatgacgaaatcgattttgataacctgctctggaaagcaatagctttactcaaaggcgatgaaacgcagctcgggttagaagttgaagaacaacgtgaatacctcctcggcgggcttgagtatttactagtggatgaatatcaggatattgatgagccacagtatcagctgattgccgcgctggcaggtaaaaatgaaagtgaagatgatgctcgtcttaatctcatggcggtgggtgatgacgatcaatctatttatggtttccgtgatgccagcgtgcgatttattcgtttgtttgaaagcgattactccgcccgtactcatttttttaacgtggaattaccgctctacggccaatattattgcatgttcaaattatcttatcagtcataatcaggggagaatgaaatgcgagcatccgatcgtaatcgatcgcgctcgccagatgcttccgccaggcggagagtggagcgcacttgaaccttcggaaggcaaagttgttatccagcattgtaccggcgcggctcagcaggcggcagaagtcgtgcgccaaattcagtatattcaacggctgcagccggaatgccctcttgagaaaattgcggttattgcacgcaatgggctcgacaaaaaggagcttatttgggtccgttcagcccttgcggatgcaggtattccttgccgctttgcgctggagaaagattatggtttccccattcgccactgtcgggagatcgccaattatctgctatggctacgagaaagagcgctcgagtcgctgacgccagcagagctgtgtcagcaactaccggggcgagaccaggcgaaccgttggcacgatattatttatgaattaattgagcaatgggagctaagccagggaggcgagccattacctgccgcttattttgaacatttcatactggaatatttacatgcccagcacagccaggttcgctttggcctgggggttttgctgagcaccgtacatggcgtaaaaggtgaagagtttgagcatgtcattatattagatggaggttggcgtagttcgcactctctgcaacctgaaaataacgaagaagaacgaaggctcttttatgttggcatgacgcgagcgatatcccgacttgttattatgcatgatgatcgtgcgccaaatccctatatcgaacagttagatccagcggtcatcagccatactgctgcacaagccgttgcgcctgggatcttacgtcgtttctcgatcatcggattgcgccagctctatatcagttttgcaggtggacatccggctggtcatcccattcattcgttacttaccgatatgcaggttggggatagcgtccaactggtctctgtcgggaataccatcaaggtgaatgctaatcaatcggcaattgcgcagctttcaagtgccggaaagagccagtggcaattttctctttccgggatccgcaaaattgaagtgcttgccatgctacagcgcagcaaaacactaacagcagaggattatcaagttgcggtgaaagtggacaattggtatgtaccgatattattggttgaaacccgtgaagaagccgcttatgacaatattacttgaagcagaatac (SEQ ID NO: 64)
68	pLG070	tagctattgtgactatgctaaccatatgaatctattgtgtgattatgagtaatgactttttctaatatttgatttttaatgtagtaacttagctaattttaaaatttgtaaaaggatgtttatgtcgatttatcaaggtggtaacaagttaaatgaggatgattttcgttctcacgtttattccttgtgtcaattagataatgttggcgttctgttaggtgctggtgcttctgtcggttgtggtgggaaaacgatgaaagatgtatggaaatcgtttaagcaaaactaccctgagcttttgggagcacttattgataaatatcttctggtttcgcaaattgattctgataacaatttggtcaatgttgaacttttgatagatgaagcaactaaatttctttctgtagctaaaactagacgatgtgaagatgaagaggaggaattcaggaaaatattaagttcattatataaagaggttacgaaggctgcattattaacaggagaacagtttagagagaaaaatcagggtaaaaaagatgcgtttaaatatcacaaagagttaatttcaaaattaatttcaaatagacagcccggtcagtcggctccggcaatttttacaacaaattatgatttggccttagagtgggctgcagaagatttaggaatacagttgtttaatggtttttctgggctacatacacggcagttttatccccagaattttgatttggctttcagaaatgtaaatgcgaagggcgaagcaagattcggacattatcatgcgtatctctataaattacatggctcacttacgtggtatcaaaatgatagcttgactgttaacgaagttagtgcatctcaagcatatgatgaatatattaatgacataatcaataaagatgacttttatcgcggtcaacatttgatttatccagggcgaataaatatagccatacaatcggcttcgtttatggagagatgtttagacgttttggggagtttatttcgaaacctcaaacagcgttgttcataaatgggtttggtttcggtgattatcatataaatagaataatattaggcgcgttactgaatccatctttccatgttgttatatattatcctgaattgaaagaagcaattaccaaagtaagtaagggtggtggttcggaagctgagaaagctattgttactttaaaaaatatggctttcaatcaagtaactgtagttgggggaggaagcaaggcatattttaatagtttcgtagaacatctaccataccctgtgctctttccacgagataatattgttgatgagttggttgaagcaattgctaatctttctaaaggagaaggtaatgtccctttttaaacttactgaaatctcggctattggatacgttgtaggattagaaggggaaagaattaggataaacctgcatgaggggttgcaaggcagattagcatcgcatagaaagggggtgagctcagtaacgcaaccaggagatcttattgggttcgatgcaggtaatatattagttgtcgcaagagtgacagatatggcatttgttgaagcggataaagcgcataaggcaaatgtaggcacatctgatttagctgatatacctctaagacaaattatcgcctatgcaattggctttgtgaaaagggagttaaatggttatgtttttatatcagaagattggcgcttacctgcattgggttcttctgctgttcctttgacttcagatttttttgaacatcatttatagtattgataaagaagaactcccaaaagcggttgaattaggtgtggattctagaactaaaaccgttaagatatttgcaagtgttgataaattattgtcgcgacacttagccgttcttggtagtacaggatatggtaaatcaaatttcaatgctttgttaacgaggaaggtttctgaaaaataccctaactcaagaatagttatttttgacataaatggtgaatacgcgcaagcttttacaggtattccaaatgtaaagcacactattctaggggaatccccaaatgttgatagtttggaaaaaaagcagcaaaagggtgagctatatagtgaagagtattattgttataaaaagataccatatcaggcattaggttttgctgggttaattaaattattaagaccaagtgataaaacacaattgcccgcattaagaaatgcattaagtgcaattaatcggactcattttaaaagccgtaatatttacttggaaaaagatgatggtgaaactttctttttgtatgatgattgtcgtgacacaaatcaaagtaaattggctgagtggttggatttattaaggcgtagacgtcttaaaagaacgaatgtatggccaccgtttaaaagtttagcgactttggttgctgaatttggatgtgtagctgctgaccgttctaatggaagtaaacgtgacgcgtttggttttagtaacgtgttgccattggtaaaaatcatacaacaacttgcagaggatataagatttaaatctattgttaatttaaatggagggggtgagctagcagatggtggaacgcatt

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		gggataaagctatgagtgatgaagttgattacttctttggtaaggaaaaaggacaagaaaatgattggaatgttcatatagttaatatgaaaaatttggcacaagatcatgctccaatgttacttagtgcattgttggagatgtttgctgagatactatttagacgtgggcaggaacgttcgtatcctacggtacttttgttggaagaagcgcatcattacctgcgtgacccttatgctgaaattgactcacagattaaagcatatgaacgacttgctaaagaaggtaggaaattcaaatgctctttaattgtcagtactcagcgaccctcagagctttctcctactgttttggcaatgtgttcaaactggttttcgttacgtttgactaatgaaagagatttacaggctctcagatatgcaatggaaagcggtaatgaacaaatcttaaaacaaatatcaggtttaccaagaggtgatgctgttgcatttggttctgcatttaatttgcctgtaagaatttcaattaatcaagcaaggccagggccaaaatcttcagatgctgttttttctgaagaatgggctaattgtacagaattacgttgttaattacctgatgtacatggctagtgcaagttggtagcgcatgtctatatgcatttatttgcatgtgttttattgagtgagcgcacaagcttgatgacccgacaggtatgtatttagactgaa (SEQ ID NO: 65)
69	pLG071	gtgcgccttatgtgattacaacgaaaataaaaaccatcacaccccatttaatatcagggaaccggacataaccccatgagtgcaatagaaaatttcgacgcccatacgcccatgatgcagcagtattgaaaaatataacatatccaactgattgtattgaaaatttaaaatagccatataacaaaaggttacacataagctactttttggggtttcaggcaagaaactaaaaattattaacgccatcaaattattcacatcttaataattagcattgaaatttaatgtttttggttctttgtacatgtcaatggcttgtctttgtggcagaatcataaagctatgcaatcattgcattgttattaacacagcatatttttatatacttttaacaccttacctcaaaaaggataacaaagtggacagaagtgcggttgatacaattcgtgggtattgttatcaggttgataaaacgattattgagattttttcgttaccacaaatggatgactcgattgatatagagtgcattgaagatgttgatgtctacaacgatgggcatttaactgcgatacaatgcaaatattatgaaagtaccgattataaccactccgttatatcaaagcccataagattaatgttgtcacactttaaggacaataaagaaaaaggggctaattattatctttatgggcattataaatccggtcaagaaaagttaacactcccattaaaagttgactttttcaaatctaatttcctcacctacaccgaaaaaaaaatcaaacatgaataccatattgaaaatgggcttaccgaagaggatctacaagcctttttggatcggttagttataaatatcaatgcaaaatcatttgatgatcaaaaaaaagaaactatacaaataataaaaaaccatttccaatgtgaagattatgaggcagagcattatctttattctaatgctttcagaaaaacatatgatatctcttgtaataaaaaagatagaaggataaaaaaatctgattttgttgaaagtatcaacaaatcaaaagtcttatttaacatatggttttatcaatatgaaggaagaaaagaatatttaagaaaattaaaagaatctttcatacgcagaagtgtaaacacctcaccttatgctcgttttttcatcttagaatttcaagacaaaactgatataaaaacagttaaagactgtatatataaaatacaatcaaattggtctaatttatctaaaagaacagatcgaccatattctccttttttacttttttcatggcaccagcgatgccaatttatacgaattaaagaatcaattattcaatgaagatctaattttcactgatgggtacccttttaaaggaagtgtatttacccccaagatgttaatcgaaggtttttcaaataaagaaatccacttccaatttatcaacgacatagatgatttcaatgaaacactgaacagtattaatataagaaaagaagtttaccagttttatacggaaaactgccttgatatcccatcccaactaccccaggtaaacatacaagttaaagactttgccgacataaaggagatagtgtaatgagcaggaataatgatattaatgcagaagtagtatcggtatcgccaaataaattaaaaatttccgtagacgatcttgaagaatttaagatagcagaagaaaaattaggtgtaggatcttatttaagggtttcagataatcaagatgttgctcttctggcgatcatagataatttttctattgaagttaaagaaagccaaaagcagaaatacatgatagaagcaagtccaataggtcttgttaaaaatggaaaattctatcgcggtggagattcacttgcacttcctcctaaaaaagtggaaccagcgaaattagacgaaataatatccatatactcagatagtatagatataaatgaccgttttactttttcaagcttatcgcttaataccaaagtatccgtacctgtgaatgggaatagatttttcaataaacatatcgctatcgtaggttcaacgggttcaggtaaatcccacactgttgcaaaaatacttcaaaaagccgtagatgaaaagcaagaaggttataagggattaaacaattctcatataattatttttgatatacattctgaatatgaaaatgcattccctaattcaaatgtattaaatgtagatacattaacccttccatattggctattaaatggtgacgagttagaagagctttttcttgacacggaagcaaatgatcacaatcaaagaaatgtgttccgtcaggcaataacattaaataaaaagatacattttcaaggagatccagccacaaaggaaataataagctttcactcgccatattatttcgacattaatgaagtcatcaattatattaacaatagaaataatgaaagaaaaaataaagataatgaacatatttggtcagatgaggaaggaaatttcaagtttgacaatgaaaatgctcataggttattcaaagagaatgtaactcctgatggaagttcagccggtgctttaaatggaaaacttctcaattttgttgatcgattacaaagtaaaatatttgataagagattagattttattctgggtgaaggtagcaaatccgtaacatttaaagaaacattagaaactttaataagctatggaaaagataaatcaaacataacaatacttgatgtaagcggtgttccttttgaagtacttagcatatgtgtatcattgatatctcgattaatttttgaatttggctatcattcaaaaaaaataaaaagaaaatctaatgaaaaccaagatatcccaatattaattgtttacgaagaagcacataaatatgctcccaaaagtgatctgagcaaatacaggacatccaaagaagcaattgagaggattgcaaaagagggtagaaaatacggagtaacccttctccttgcaagtcagagaccttctgaaatttcagaaacaatattttctcagtgtaatacttttatctcaatgcgattaactaacccagacgatcaaaattatgttaagcgattactcccggatacagtaggtgatattacaaacctcctaccatcgctcaaagaaggtgaggccttaatcatgggggattcaatatcaataccttcgattgtaaaaatagaaaaatgtacaataccccatcgtcaattgacatcaaatatcttgatgaatggagaaaagaatgggtagattcggagtttgataagataattgaacaatggagtaaaagttaatttcagaagtggattcactcttgctcaagagtgaatccactaatatcatatcctaatgatatagtttaataaaatctattctggaatcattaggctgagag (SEQ ID NO: 66)
70	pLG072	ccattttttaaaataccctcttaaaggagggtattttaaaattatttgttttaataaaaattaaatattatattcattatcacaaccaataaaccgtttattttttacacttgcatactataaagacatgaaagatcccccttgtcaggactacgctaaagataataataacgtctattttcgtcatatataatatttgcttgttgcatttctaaaaaaaaagagtaaaatatcaaaatttaggagttacttttggacttatatgaaggcaattgacttatttgcgggggctggagggtttagtttatccgcccacaatacaggcgctatagatgttgttgctgctatagaattcgatagcgcggctgcaaacacctacagaaaaaatatgttagaaaggcttgagcataagaccgaacttttacaggaagatattttactcgtaggcccaaaaaagttaagaaaaaaaataaagctcaagaaaggcgagcttgatatgatacttggtggacctccgtgccaaggttttttccagtcatcgaattaatgatgctggtgttgatgatcctagaaataaattacttttaaggtatttcgatttgtttgtgaatttaaaccaaaagcttttttggtagaaaatgtctccggtttgttatggaagagacatgaagcccatttgaaacgctttaagtttttggcttccaaaaatggttatactttaattcattgcgatgtattaaatgctcgtgattatggtgttccgcaaaatcgcaaacgagttttcattgcaggtgtcagaaatgacattttaaaaaaaagaaataatattgagtttccacctcaagctactcatttcaaccctaattctaatgaagtaaaaaacaattcaaaaaatacgtggagaaccgcatcctctgtttttgagaagatgaatgataacttaattcaaagatatatatctgaatactttcttaaacatacttcttactcaattgatgaagcacaagagctacttgaaaacctagaatatcaagacgcacccataagcgaaaaagatccatgcaacatacatatgataccaactgagcgtatggaagagcgtttcagagccacaaaactcaatggcagtagaagcgatgcaggaaaagaatttgagctaaaatgtcattccaatggatacgcaggccataaagatgtttatggccgcataatgattcacctcccagccaatacaattacaactgggtgtaacaatccatctaagggaagattcattcatccatgggaaaatcacggcatcactttaaggcatgcggcaaggttgcaaacgttccctgatgactatattttttggggtaatgcgacagagcaagcaagacagattggtaatgcagttccccctatgttaggcacaatattaataaatgcattacttaacataattgcacccaatagataaggtgtaatgtatgaaaaatatcaaaattagaaacttaaatggaccaaaaaatcatttgatgattacttaccttataataatagaaggtgaaaaatggtaatttcagcagcttttcaaacaagagcaaggacaattgatcatctagggcgtgagcaaatagctgattgtccaaccgcaatttccgagctttggaaaaatgcatatgatgcttatgctcgtaatgtttctctaaatatatttgacggcaatacacctgtggcaactttagttgatgatgggcatggcatgtcgttagatgacattatcaataagtggcttacagtaggaaccgaatccaaggctacaaa

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		aaaagatattccatatgaagatagaaacggaatagatcatattcgagcaaagcaaggtcagaaaggcatcggtcgtctttcttgtg cggccttgggctcattaatgcttttagtttccaaaaagaaagatagccctcttgtagcttgcctgctcgattggcgtatatttgaa aacccatatttgatgcttaatgatataaagatacccattatggaatgcagtgataacaatgaattaatcactgttataccggaaat gtttgatgctttgatgggaaatctatggggtgatggtgatgatatattacgagataaccgtattgaacaagcttgggaaaattatt ctgaattagaaagaaatgaaaataattatattacaaaagaagctatcgagaatactgtaattaatgctttttttgaggaaaggcat tttcaatcttggcctgtgtggaataataaaaccactcacggcacagccatgtttatagctggaattcatgacgatttaatagctca gctatcaacagatgctggttcagaagctcaaggtgcagaggttcgggctaaagaacgctttcttcaaacattaaatagctttgtta atccatttaaaagagaaggcgaagaacagattactgatttcaatacaagtgttgtcgcatggaatggtaatctgcaacgatttatc atcgatgaagttagaaactttgatatttcaaactttgaccagctagaacatatagttgaaggaagtattgatgaaagtggattatt ttccgggaaagtgaaagccttcggagaatggtttgataatattacagtcaaacctaaatctgcatataagaccagaaaagatactc gctttggccctttctttttaagattaggcacatttgaagttataagaaaaaatagtacattatcagatgaacagcatgcaaccttc gaccgtatccgtgatcagtttggtggagtaatggtttttcgtgatgatttacgtgttatgccatacggacgtgaagataatgactt ttttgaaatcgaaaaaagacgttcaaaaaatgctggtttatatatgttcagtaatagggcatgttttggtggtgtatgtataacga aagaacataaccccaacctacgagataaagcaggtagagaaggtataattgacaataaagcatctaagttatttagagagatagtc gaaaacattttaatagaaattgcaaaaaggtttattggccgcgcatcaaatatacgagatgaaaagctagaggaaataaatgctaa acatgctgctttgaaagcagacgaagatagaaaaaaattattacgtaaagagcaaagaagaatcaaaacatcgattcaaagagatc gtatttctttagaacatttaagaaatgaattttatgaaatatcacagcttctaagcgacaagaataattttaaagaactagaggag ctattacagctcaaagaaaacatcgacgtattggatggtaccctaaaaaacctatctttaggttcagtaccaagaaatttagggag tatagagaaagactaccgtcagtatcgcgatttagagattgatgctaaaagtcttttaaagcagattaataactctgtatactcag cgcttgatcattttactgttaaagatgattattcaattgctgagaaagactttcgtagcaaagcagccatattacatgcgaaaata agaaaattttccaataaaggacgcaatatattaaaagaagagatgttgcgtttcgaaaagataacaaacaatacaaataaagcttt ccatgaaaaaacatctcaatatttatccgatctacaagaaaatagaacttcactcaaaaaaacacttgaaaatttagatcttgctt atcagattcaagacattgaaataggtcaaacctacgccccatatattaccgcattagaaagcttaagagaggaaattgatttagaa ggcctcgcgatctcttcagtcaacgaaaatacacggttgaagaaacaggtagagcaagtgaatgcactcgctcaacttggaataac tgtggagataattggtcatgaaatcgaaggtttcgatatgactattgagcgaggtataaatagactgtcatcaacaaacctcgatg aatatcagaaaaatgctttatcaagtattacccaagcacatcaatcattaagcgattcttggcgtttttttaagcccattaaaatta tcaggagataaggtaagagctttcttgagtggaaaagatattttttgattatgttaatcatttttcaacagtaaatttgaaaaaga ttcaattgaatttttcttgctctactaatttcctagatatttcattatatgatcaaccagccagaatttatcctgtgtttattaatt tagtaaacaactcacgatattgggttaaagaaactaaagaagagcgtcgaattattaggttagatgtacttgatggtttgatatat gttagtgataatgggccaggggttgatcctgatgacgtgtccgaacttttcactatatttttctccaagaaacaaagaggtggtcg cggggttggcctttatctctgcaaacaaaatttagcggtgagtggccatagtattttctacgaaacaagaacagagaaaaaaatac taaatggtgctaattttgtaattaatttcaaaggaattaaaaatgcttgataattctactttcgattacaaaccacatttaaaatc tgcttatattgatccgattagaactgtgacagtcatcgatgatgaatacccaactattgatgatttaatttcaccgaccaaagaca gtttttctcaagacaacatttctcgattaaaagatattattgatataagtcgaagtgaagaatataattggcttttagatgtctat aatggaaaagagaagaaaattcaagagggaaccgtatctaaccgtctttatcacagtgatctactaatcttggactatcatttaga tggagaggactctggatattgtaaaaaatctatagatattattaaaaatctatctgaaaatcgtcattttaatattgttgcagtgc atactaaaggttatgatggacaaaagggttcagttaatgaggtactaatcgatattattacttccttacaggaaagacccgctatt agtattttaaatgataaaatcaaatctagaatagatgatgctttagatgaatgggaaatcgaagatccaagtatcagggaagatct aattaattcagtttctacattagatttacttttcttgattaataaattcgggtcaaatttaagttcaggatgtttcgactacgaag ttcttgatgtttttcataatatatttgatcaaaaaccagacaatataaacatatccaaaatattgatttttaaatggatctcatca gaaaagttacatagatacgctgaccaatttaataataagacatcaaagttctttgattgggggacaaatgaaaaccacaattggat aaaaacagaagacttatttattactgtccttggtaaaaaagacacaccaatcagtgacataccgaatcaacttttggaggctttgt caaactctaaaccacatccgcacaaacttattttatcaaaactcagaagtgaaattgaaagtaatggtagctatgctgcaagtaat ataattaacaaaaaattcttacaggcggcgtggctaaaggaattacttcaaaaagaggatgaatatgctatcaaaacagctgcatg gcaagcagtaactaaattgtgggaagaattagcatacgaaataaaacagagtcttgatgattttacaattaatcttgtccgcgact taaagaaaattaactcacctttaaactatttcatagagaaatctacacttgatgctgaacttgaacaaattaaacatgcaaattgt ttcagttgttcaaaaaaaataactgctcatcatttggttacggggcatgttttggagttcaataataatcactggttgtgtctaac tcctatgtgtgaccttgttcctggtcagaaaaacggaaatagtttactccctgttacgctcgtgaaaatgtatgatgcgaaagttg ctttaaataatacacgtaaaaatatgcaaaacgagcttaaactacccaatttgccagaaatcaacgaagatgaatcaattagacaa atactaaattattccacacagaataatctattgttcgttcagtctgaacatgacgggaaaatacatattcttagtttcaccgttgg actcgatggcaaggcaaatcctaaagcaatggattgctatgtggaaaatcaaggtattttctctgaagataaaataatagcactaa aatatgccaagcccactgaaaatgaaatgaacataatatccgtagaagcaaaaatagttgctgaattacgctacgaatatgctttg aatttattaggtagactcggtgtatcaaaatctcgagtcggattagattttatcaactaaggtgcgttagcacgcacctagtctga caggtaccagttgtttatataggtatctgtcagactacatcctctttaggtttctctcgcccagataatttttttccatcaagtgac attttcattgatgtctaactctcagacattaaagtgtctaacttccttattaatgtcacaagcaacaattgaatttcaccgctttt gcgagcatgatcgcaataatatcagcccgttacccggttaattcctatgacatcactcgaaacactgcaatcggctatctctaacg tctctgtatggcgtcagggtgatgtatgcgcgccgcataaaccgttgctgctgctgtatgtgttgtcacagtacaaagcaggccac ccgcgcctgtttaactacggcctagagatccacgaaccactcactcgcctgctaaaagagtttggccccaagcgacgcactgacta tcccaatatgcctttctggcgactcagaactgacggcttctgggaaattgctaatgcggaaggctgcaaaccccgtagaggcaaca cccagccgacaaagaaagagctgattgataatcaggtagcgggggggttttgatgaaacagcttaccagcaactgcttgcacaccct gaagtaattgaccaactggcccagcagatcctgatggatcgtttccccgagagtattcagcggatcctcgccaaccaactgggtct ggattttatcgaccgttcaaagagccgcgatccgcgtttcagggatatcgtgcttcgggcttaccattcgcgatgtgctttctgcg gttacgatctacgactcgatggtgcgctggttggtattgaagccgcccatattcactggaaaacctatggcgggccgtgtgtggta aacaacggtctggcgctatgttcgctgcaccacgatgcttttgatatgggcgcattcgggctggatgaaaaccttaccatccgcat ctccggcggcgtcagccgtagcccggtggtggataacctgttctggcaacggaacggccagcagttacaccttcctcacgacaaat cgctgtggcccactgaacaatacgtcggctggcatcgtaaacagatcttcaaagcctgagaccgtgagcttcgcaggtatcatcga ttgcccaaactgctttatcccctacaacggataaattgcttttaacccctatagcggataaatccagcacaccagtgttggacttc agaataacgaatccaaactctagccctgagacaccaggctcttgattattattgataccgtattaatctgtacgaagtttgacccg c (SEQ ID NO: 67)
71	pLG073	gtaacaccgttgaacgtcggctgggtgttgttcataatccctttaaaggtctggggatggccatgacctcagggcggtagcgtga ccaaagttcatatccataccaattatttttatttaaaatatcaacttattcgagttgttttatttagttcaaagaaggtatcaaat

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		tgatagttatagatttttttgtggctgtggtggagccagtgaagggctacgtcaggctggctttgatatcgagcttggattagat attgaccaacaagcatcagaaacatttaaagctaatttccctgatgcaaaattcatccaagatgatattaggaaaatcgaacctca agatatctccgacatcattgatattaaagctaaacggcctttgttactgagtgcatgtgcaccatgtcaaccattttcgcaacaga ataaaaataaaactagtgacgactcaaggagaaatctactaaatgaaactcatcgttttattagagaacttcttcctgaatatatt atgcttgaaaatgttcctggaatgcaaaaaattgatgaagaaaaagaaggcccatttcaggagtttattaagctacttaaagagtt agagtataactatatatcttttatagccaatgctgagaactatgggattccccaaagaagaaaaagactcgtgctcttagctagtc gagtaggtaaagttaccctaccagagataacccatggtaaaaataaaatcccattcaaaactgtacgagattatatccaggacttc acaaagttatgttcaggagaaaccgaccccaaagatcctttacatagggctggaacactgagccctcttaacctaaaaagaattat gcacactccagaaggaggggatagaagaaattggccagaagagttagttaataaatgccataaaaattatgatggccacacagata cttatggaagaatgagttgggataagcctgcgcctacacttacgacgaaatgtaatagttactccaatggtcgttttgggcatcct gaccccactcaacatagagcaattagcataagagaagcatcaagattacaaacatttcctttaagctatgtttttaaaggttcgct gaattcaatggcaaagcaaatcggcaatgctgtaccttgcgaactcgctagactatttgggctacatctcatagaaaattgtacta ataaggattcatagatatatggctaaaataagaacaaaggctcgagctttggacatgcttggcagacaacaaattgcaggtatacc tactgccttgagtgagttatttaaaaatgctcatgatgcctatgctgataatgtcgaagttgatttttttaggaaagaaaatcttc ttatcttgagagatgatggattaggtatgacaaccgatgaatttgaagagaggtggttgactattggaacctccagcaaattaatc gacgatgatgcaattaataaaccagcagtggatagtaataaagcctttcgccctatcatgggagagaaaggaataggccgtttatc tatcgcagcaattggaccacaggtgctggttcttactagggccaaaagagacaatgagcttaagccattagttgctgcatttgtta attggagtttatttgctataccatcacttgatcttgatgatatagaaataccaattagaactattatcaacgacgaatgcttcact aaaaaaactcttgatgagatgattgagcaagcaagaaataatttagactctttatcacacaaaatatcaaaatcaaaagtatcaca aataaatacacaattatcatcttttgaatttgatcctattctatgggaaaaaaaattaggtgggctaagactatctggagatgggc atggaactcacttcataataatgcctaccgaagaaatattaatagatgacatttccacgagcgatagcaataaaacatcagagcag tcttctcgcttagaaaaagctttattaggttttacaaacacaatgtacagtgattcaaaccctcctattatagctcgttttagaga ctatctggaagatggtgagtgcattgacagaattagcgaatcaatttttttacaccgcaagaattcaatcttgcagatcaccaca ttgaaggatggttcaatgaatttggtcaattcagtggaactgtttctgtttatggtgaagagccaattcatcatgtcgtgacttgg aaaaataataatcaattaacccaatgcggtccatttaaaataaaattagcgtatattcatggtcggcttcgtgattcacgcttacc catggagttgtgggcccctctgaaggagaaaacagatagatatggtggtttatatatctatcgagatggattaagaattttgccct atggagattcagatacggattttctaaaaatagaaaagagaagaacgttatccgcttctgaatatttttctcatatcgacgtttg tttggagcaatagaattaacaaaagaaaacaatgcttcattagttgaaaaagctgggcgagaaggattcattgaaaataagccata taaacagtttaaagaaatgcttgaaaatttcttcatcgaaatcgcaagagatttctttaaggacgatggcgatatgtctgaattat ttgttgagacaaagcaacgtagaaatgaagaacatgatttgttatctaaaagatctaaacaaactaaagctaaaaaagatagatta aagaaagatctgtatgatttttttgataagttagataatgattactggaatattgaaataaataagctaatcaataaaaacgagga atatttctccagtacagaaataacagacaccaatatagattatgtatacaataaaattaaagaacaaaatgatgctatcattaaaa atctacgtaattctgtggatataaagaaaccctctggagttggattaacaaaagagttatctaatttatgggatagatatcaaata gaaagacaaaaaatactgttatcactaaatgagctaaaagataacgttgatagaaagcttatagaactggataataaaaataatga ttttctcaacttacggaagagacttgaagattctttgaatctacaacaaagttactatgaaaaagaactaacaaagttatataatg acgctaaaaatgctttgaaagatgtgcaatctaaagcaaataggttaatttctgataataagaaaaaacataagagtgaactaaaa aacatttcttatgaattccaatcaactaatctcaatggcaaagatactgcgtatatattggatgtaaaaagaaatctagaaagtaa aattgagaatacttcaaacgaagtgattaatgaaataagaaaactaaccgaccagattgcaataattagtgatagtaccacttctg aaaatttatcatcggctcaagtaactgaagcaatcgaaactgaacttgaacatttacgagaccaacaagcaaataacgcagagtta atactacttggcatggctctttctgtagtacatcatgaatttaatggtaatattagggcaattagaagtgcgctaagggaattaaa agcatgggctgacagaaatcctaagcttgatattatataccaaaaaatcagaactagttttgatcacttagatggttatttaaaaa cctttacaccattgacaagacgtttaagtcgctctaaaaccaatataactggaactgccattttagaatttatcagagatgtattc gatgatcgtcttgagaaagaaggaattgaattattcactacctcaaagtttgttaatcaagaaattgtaacttacacatcaaccat ttaccctgtctttataaatctaattgataacgcaatatactggcttgggaaaacaactggagaaaaaagacttatacttgatgcta ctgaaacaggatttgttattggtgatactggtcccggtgtttcaactagagatcgagatataatatttgatatgggatttacacga aaaacaggagggcgtggaatgggattattcatttccaaagagtgtttatctcgagatggatttactataagattggatgattacac tcctgaacagggtgctttctttattattgagccatcagaagaaacaagtgaatagcggatataaataaatgacaagctctactgat tttcataaactttctgaagactgcgttcgccgtttttttacattctgtagttgctgtagatgacaatatgtcttttggagctggtag tgatactttccctacagacgaagatattaatgctttagttgatcccgacgatgatcctacaccaataataacagcatcagcatccc caaggatagaatcaactaaatcaaaagcaaaggtaaaaaaccatccttttgattaccaagctctagcagaagctttcgccaaagat ggtattgcttgttgcggattattagctaagagttttaatgttgaagaaagagatataattacagcatcatcccacaaggcagatat aacaatacttgactgggatatgcaaagcgatagtgggcaatttgctattgaaataataaaatcgataatcgtttcagatataaatt ctggaggacgtttacgtcttctttctatttatactggtgaacatgttactgctgttataactaagttgaacaatgagttaaagaaa acataccgtagcgtaataaaaaatgatgatagtattttattgaagataactatgcactcgaacaatggtgtatagttgttattag taaagacgtttatgaaaaagatcttccaaatgtgttaataaaaaaattcactaaccttacagctgggttgctatccaacgccgcac tctcttgcatttctgaaataagagaaaaaacccatgggatattaacaaaatataataataaattagacactgcatatgtttcccac atcttaaatttaataaaatccaaggagtcaagggcatatgcttatgaaaatgctcatgattatgcagtagatttaatttctgaaga aataagatcaatattgcaaataagtgaaaacttaaagaaatctctaagcaaaaactccttatcccattggcctatttttcactatg caaaaaatggttgtaagaattttctattaactggaaaaaaacaaaaagacttatcagtagaacatctaaggaatatactctctgct gattctttagaagaaattcaacacgctattgaacacgcatctttaggtaaaaaggaatacttaagccaagatggtgaagaagataa aaagttaatgcaattatgctctctggaaatcacgcgcaggagtttaagatatcattctcatatagataatgtgtccttaaaacaag gaactttacttttagatgcatataattttgtctatctatgcatacaaccattatgtgatagcgtcagattgcatgaaaaagccgat tttttattcctcaggggaacactggacgataataattacaatttgttaatcgaagatgaatatggcggtttttataaaattaaaat gccggcaaaagcttctaatattatttcattttcatttggagtcgaaaatggaaacggtgtcatcatagggaaaaagaacaatctag ttaatactgactatatctcattcgttcctttactcgttgaaaaaatatctactccaaaagtattgaaatggatcggggaaataaaa acaacgtacgcgcaaaaaataacaactgatattgttgctaatctgtcaagaataggtttagatcaacatgagtggttacgaataaa atcaaaagatatataaatgattatatatgccgtcgttttataaaaactggcggcatgtatatctagttagtccatcatagaagtca agaaatttagtttgccctatatcttatagaaaatatatttatatgcttaaaaaacaccatctttataagatggcatttatgtgct ttgtttcgatcaattacaactg(SEQ ID NO: 68)
72	pLG074	gattattatccagcctttgcgcaggagagggcatgaactgctcactctgatagccgctcttgccatagttgagcttactccacaaa agtagacacattctgttcttacctagacgcctgctcaaaggcggccgggatgactatagcggtgatccagattgtacctgatccct

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		atacatgatttgtatcattgtcaagcttttgaacgatttaatctcttattggagttcatgatagccacttgaatttcgaaaataaggtactatatctagtaaagtcttagtcaatttttggtatatacagtggaagtggaaccatttcgtgtcctttgtttagatggcggtggaatgcgtggcgtgtatcaggcgacgtatctcaatacatttgcacagcgtctgcataactctggtgaaggagtcttagatccaggaaaggcatttgatttaattgtgggaaccagtacgggaggcatagttgcctgtgcgctagctgcgggggtctcacttgaaaaggttcttgcactttatcaagtgcatggcggaaaaatattccctcggcaacgattacgtgcactacctcgagtggggaagtatgtccgtggcctattttctggtcttgcgtctggcgaccaggctctgcgagcagtcctttctgattcattcggtaccgaaactatggggcaggtctatattcgtcgtggaattggtttagccatcactacagtggatctgaataggcatgctgccacagtttttaaaacccctcatatgagtcgtcttaatggacgtgacaacgatcgactattagtcgatgcctgtatggcgactagcgccgcccctatcctgagatcaatagctcgtctaactgaacctggcggtggagccactgttgattatgttgatggcggtctctgggcaaataatccgggggctgtcggcatgatagaagctcatgaaatccttcagcagagaggagagattgaacgtccgattcatttatttatgctcggtacgcttccattgcaaggaggtgaagaacttaagagcgcagataaattacatcgaggtgttttggggtggggagcagggattaaggccatcacagtaagtatgaattcacaggcagttgcgtacgactacttggctcggaaaatcgcagaattgcgaggatatggaagttttgcatatcgactcccagcacaatgcccatcaggagaactccagaaatatttggaaaatatggacgatgcacgtcctagggtgcttaatgcgcttgcccgacaagccgtctcagatgttgattacgcttgggctacggcagaatcagtaagtaaaatgggcgcgtttcgaactgcattggcaagttcgtccaattatagttgtcataaatccgaggaacaccatgaccattattgattgtaataaagagatgagagggtatcactcagaagaggtaaacctctcgaatgcagagcaggcagaaatgcgcggccgccgcgacaatggtcgaacaaggctccgaaacggattgacaaaggctggtcatcctttgccgaaggagttcagttctcaaggctcttatgcgatgcgaacaatggtccaggatgatgcatgtgactacgatattgatgatggcgcgtatttcgataaagaagaccttaagaactctgaaggcgattatcttagtgcgctagatgttcgtaagcgggttcggaaagcattgaaagacgaccgattggcatatgatgcggttgtcaaaaccaattgtgtgcgtcaaatgtatcccgatggatatcacattgatatccccatttatcgtacgacctgttctaaagatatttgggataatgacatcatagagtatgaattagcaagtggcgacgaatggaccaaatcagatgcacgtaaggtaacgagttggtacaacgatgcggttggtaatgaactgaaagcgggggaatctgataccagtcagatacgcaggatcaccaaacttactaagaaaatggctaggagccgtaatacctggaaaaaaaagacaaccagtggcatttgtatttcgaagttagttgtagacaatttcgttgcgcgctcaaatcgtgatgatgatgctttgcgtgatacctggaaggcaatcaaattgcagttagaagtcagtcaacgtattacccacccggtgtttacggacaaaaatcttgctgaggaaggagacgaatgcgttatttttttccgggaatgtttgggtgaggtgctggaaacattaaaggtgctcgacgagcatgactgcacaagtaagaaggctggcgacgcttgggatgaggtgtttaatacaacttattttagcgcccagtgtaccacggataacactacatctaaatcgctgctacggcctgcagttgcggccactgctagcctgtctttccctagttatcccgtacaacctaacaaatcatcggggtttgcctgatgaagtgggctatagacgatcccgtgcgtttcctgagggagaaggatgaactcacacatcttgaaaccgagacgggttggctaagcacggcttggcgtatatctgaagagggctcgatcaccgttgatatcgacatgtttatccatgggcgattgtttgctggggaaatgacatatccggacgcgtttccggattctccgccctacatacgtccgcgagataaatcagagcgatggactaaccatcaatatggcgtgggtggttcactgtgcttgcagtggcgggcagataactggcatagtaatgtgactggtgcagatatggtacgcagtgcgcacgagttgctgagtacagaacagcatcctgaattacctcattctgttccctctgcgcatcgcttgacggaggggcaaaaccttaatttcgtatttcgacgttatgtccctacctccgaagtcgaaaacatatttactatgctcccacttcagtctagaacccgaatatcatcttcaactgtgtataacgaagggtcggcggtaatgttcacagccagagtcgctgacgaacaggatgagcttcgaaatgttaccgatatccctcaagggctcatcgattttgttagtattttgtcgttgtcctatgagggctgggtctttagaagcgactactttagccagaggcaatccttagaatctgtagaagcattaatccagatattgatgatggccggttttaacaccgatgacattctggttaaggaaggggataagttcaaggctaggacgatcatattattaggcaaggaatggtcatcactgcgagtattcctgttagattctggggagcaaccagtgctgcgggagcatcgagttgttagatctccgaactcaaccttaagactttcggaagaatcacagaagttgagtaagatccgcgtaggaattgttggactgggatccgtaggtagcaaaattgcaatttcacttgctcgttcaggtgtcagacaattcttattagtcgatgacgactatctcacgcctggcaacttggtgcgtcatgagttggggtgggcccatgtgggagctcataaggcacgggccgtaagcaatactttagcgcttatagcggctggtgtgaaagtggatgtaaagactatgcgtcttgcggggcaggaatcggcggtgacagcagcggctgcactaaaggatctgtctaattgcgacttgttgatcgatgctacagctaatccagaagtttttttgctgttagctgcgactgcccagcgaaatggaataccgatgtgctgggggagatattcgcaggtggttacggaggcatgatcgctcgagcacgtcctaaacacgacccaaatccattagctgtgcgtgacgcttaccattcttatctctcaaccctccctgaagcaccatttaagaatatggctagctatgatgggagtgatgaacaaccacttatagcatacgacagcgatgtgggctttattactactgcactgacacggttggctgtggatactgctctatgcagagagccaagcgaatttccgtactctttgtacttgctgggtatgcgacgtgaatggattttcgaggagccatttgacacacggccagtcgaaataagtggagaaggctgggaacgcgacgaaaatgctgtgagagatgaagatagggtcgcagttgcaaaggcattggtaaatatgtttcaaggaaaacaaagtgctaacactgatcctacctcctaagcagcatgagttaatgatgactgcactccaaaatgctggtcaacgcgaagtcggcgggattcttatgggtgaacatgtcgggacaaatactttcatcgtccgggagataactatacatcgccgtggtacgtttgcttcctttgtacgacgtattgaggatgctattggtgggctccgtgtttttttaaaggaactggatacgattatgttcgcttcaattatatcggtgagtggcattctcacccttcatttgagccatacccaagcagaacagacgatctgtctatgttacagattgtaaaggatgaaaccgttggtgcaaattttgtggctttgttgataatcaagctcggacctgatggaaaaatggtttcaacagtccatacatatcttcccgatggttcgaagattctctcaactcttaagattcagccttaactcagaatgtcagattgtgaaattcatcttctagaggctaattgaagcatgctgattattttttgaggcggaagtatgttgcct (SEQ ID NO: 69)
73	pLG075	aactcacccgctctgaacgagccccttgaaacacaagacaccgtttttcccttaccataagggataggcaaacgactgtgtttatgactaccagcagagacaaaaccatcgaagtgctcggccacccatttgcgcctctaggttgctacgagactgcagaggatccatgtagcagattacctcggccatgaagctgctaacggaagcgaagccatagaccgtaggcgatacacgtacgtatggctttccggaagggcgatcctagtcaactgtctgatgtccgccaaatctttctcaatactggtcattcaccttttccttgaccggctgtcaggcccaacgtgcattcagatcgtcgcctaaatttgttgcatcacgtagagtctgccgcgtgctcgcccctatgccagactagtctgatgtggcggatgagataggtcacgacggtggtggctcggtagagtcggcatcgccgagtcaacgatggaacgtaaggggcgtgaatgcaaatcagccgtaagctcaacctttatgagatcgaggatctctaccagtcgcttggtacggattccaatctcaggcttcctatcagcatgagccacggcgggggttgggcgtggatgcttcgctggcccagttcatcgtcacctgggcacgtgcttgcgaaaaaaccgtccttcacctatatgccccgctggcgacgacgccatgacgcaaatcacgcagttggcgcagagtgcttctgggttcttcgcgctgatcatgtgcagtgaagtccacgctcagaatcatcaactgatcgatcggcgggaagcgcttctggcgatcaggccccttgtcgatgcgatgttcgcaggcgaccttcgtaacacctccaacatccgaggcgcccgtccaacggccatcaatctgttctgcgtgaacaacgcaaagcgtgagttcatcaagccgttttacttcgatcacgccgtgccgaaagtccagccgagatcttggttctcgactctcttggagacgtcatcgaagctgatgaatgctcgcagtggacaaggggcactgcttaggtcaggtctcccggcattgggcagcgtgctttgggagttgatctccaacgctgaccagcacgctgtcactgatgtaggcgggaacaagtacaagaaggcgctgcgtggcacctccatcaaactcaaccgaatgagtcgtcaggatgcgctgatgtattcagaccaagagccggagttggcgcgctttatcctgaagcatttcctgagagctgaggtactggacttcctggaagtctcggtcatcgacagcggtcctggactggcacggcggtggctgacggcgaaggaggggcggccagtagaaagcctggaggagctgagtcttgaggctgagcttgaggccacgctcgattgcttcaaaaagcacattacatccaagccgcagtctccgaactc

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		gggtatggggctgcataacgctgttcaagcactcaacaagctcaaggcgttcgtacgcgttcggacgggtcggctttcactgcatcaggcttttcagggaagtgatgagattatggagttcgatccgtcgattcgatacggtggccgtgtgttggccgctgtggaaggcactgtcttcaccatctgcattccggtgagctgacatgttcgatctcatggattttgaagtcgagttgcgtcagtcaggtaagccggttcatgtggtggttttcttcactggccctgatctcctcacagacacgcaagcggctcacgctctacagcaccaattgtcgggttacgtcatgcctgacctagtggtgtttctgatgcctggttacaccttggatgaattccgagcacaccaggcaaatgctacatcgcccctgatggcggagctaagccgtaaaggcccaggctcgcctcgcacctacgcgagtgcgttctatgacgtgaatggtgccattaccgagtacgtcaatatctctggccctgaggagcagttcgaggaactcatcaagcacaactctaacgctatcgcgaggactggcctgacccacctcgtcgaacgctccaacgtgctgaagaaggcgcctgcaggcttcttctactcaaagccctcttctcgggcttcgaactatttcattcgggcggaagacctgctctctgagaccttgcatgcccactacctggcgtttgcatgcctatctctcatcagtaaggcaacggaagatgggatggggacgcccgataccctgtatctggacacaatcgcattgctgcctctggcgctgtccatgcaggtgtacctcatgcgatttgagcagccgggctttgcgaatatccggtcattccattcgcacgaaggcctaatcaagggtgggcctttgcccaaggcagtttccgccctgtgtctcatttccgcatcgacccagtgcggcctcgcgcagcaatgggtgaaggtaaacagtgctccgccgacgcgcgtggccaccattctttcatttgagcgctcatcggactcctgctccgtcttgcacacactgaagcagcccgaagactttgaaatgttgggggagggtgaagcgagcgggattcgtctaattcggatccatggcgagcggttcgttgctgagcacagtgaaaccaagctgctgaacatcggcactgatcatgcgccgcccctgctgcaatccaagttctactcgttcatgggggccaacctgttcagctgcttcacccatgaccggccaggactgaggcctcggacagtgcatgtcgataaagataacctggtggctgccagcgatttcggtgaatggttcgacagggtactgcttgaggaagctgtcgcgtcgacccgttggatcatccacgatgacgacgctgccagtgcggccctggccgatcgagcgatcgcttacttagggatgtgtggcgtcaaggtcggtaacaaggtctccttcgatgacttcgatgccaacacgaattttgacgggtctgtcatcgtcattgccgctgctgccgaacgtggctcacgcctgcagagtgtgagccgacgcctgcgtaccgctcagcaatcgggtaccaggctttacattacgggggcactcttcgggcgcagctatcaactgatgaaggatctgcagagcaacctgacgcaacctgccaaggatcacagccggtatgttttcaagacgtacatggagatcccggcagcggagcttgcctgcacgagtcattgggccgaagagcagcggctgctcatctccttgcattcatttgcggaaactttctcgccagcgattacgcagcgcatggaagtatttgatcgcgcctctactgggggcttggtctgaacccattttggccgagcagtcacaccgggcagccgatgacacttagccgaggctttgcgtttgtcgacggtacgaaggatgtgagggcgcgacgtcaacggatatttacctaaccatcttgtggattctgcagaatgcccggtacagcggtaaggtgcagaacgccaagcggcttgagtccggtgagcttcagcaggtgctcctatcgccggatgtgttctcgcgcttcgacgatggcgttatccaggccgcattcttgcgcgcagcggtgccggcggagcttgactacagggctcatgaaacccacagcctggccatatcggacatcattcagcgcatcgccgcagggtacggacatgaacgtggtgaagccgccatggagtttgtcatggccttggctatcgggaagatacgactgcacaaggatgtcgataaccggctgcggagtaacttgatcaatatcttgacgccgcacgttcaggagatccgttatctgctggatccgaattacgaatcaccgttgtgatcaatttccgctaacccgttgcatgcgaggtatccagttaccggcaactcagctcatggctgagctgaaccctggttgctcttctagtttcgatggcttgccgattgccgggatcacccacctgcgtcggttctgcgacgaaggtctaagggcagggtggtggcacctggcttgctcattccgtttgacctcgccaccat (SEQ ID NO: 70)
74	pLG076	cgctcagtccggttggtggttttggttggtttggcgattgctcagatcgcacaatccgggctgagttccctttcagtgatctactattccgcgcagctatttagtggatataatcacgctttgaaaaaaaaacgggtcaattactcttcgccccacagcaacgaataaggagaaatttgtgagtaacgtcaacactttccttaaggaaaatttatcttcagtaagtaagaatgtttttgtggctcctggcatccctgaaaaaaaactgaataatgtcgctaaagcatttaatgttgtggataacttgaatactgtgctagccatttatgacaatacggtatttggtagcgcaaaagatggcatcgtttttaccggtgaaaaactggtcataaaagaagcttttgaaagtccttatgacttgttctacagcaatattgaagcagtagaatatatagaagatgtcacggtaaatgataaaggcaaggagaagcgaacagagtctgtttccctcaaactaaaaaatggcgaggtaaaacgaatcaaaggcttgatggagtgcaactataagaagttgagcgacattcttaagcataccatcagtgactttgatgagttcaaagaagaagatcagctcatcactcttgccgaaatgtcagaagctctcaaagtggcttatgtcaaaatcattgtgaacatggcgttctcagatgatggtcaggttgataaaaaagaatttgccgaaattctcttgttgatgacccgacttgagttaacgactgaatcccggtttacactgcgtagttatgtcggttcagaatccagtctgataccggttgaagaattaattgcgatcattgaccgggaatgtgtcccaagccataacaaatcaataaaagtctctcttgttaaagacctgattagcattttcatgagtgttaatgaaggtgaatataaaaaattcccgtttcttcagcaagtgcaacctttgctgggcgtaactgacgaagaaatagaactcgcagtaatggctattcagcaagattttaagatgttacgggaagatttttccgatgatgcgctgaaacgcagtatgaaagaacttacggcaaaagcaggtgcggtaggcgtgccactcgctgctgtctatctctctggctctgtcatcggtatgtccgcagcgggcatcacttctgggcttgcaacacttggacttggtggcgtgctgggtttttcaagtatggcaacaggtatcggtgttgcggtgttattaggtgtaggtgcctataaagggattcgtcatcttacgggtgccaatgaactggataaaaccaagcgccgggaactcatgcttaatgaagtcatcaagcagacacaatccacattgtccgcgctaattaatgatctaaattatatttctggaaagtttaacgacgccctggatgcgcataatcggcaaggagaaaaaattctaaaactccagaagatgatgaatgcattgaccggtgcagcagatgaattgaataagaaatctaataaaatgcaaaacagtgcactcaaacttaagtgccctgtttatcttgatgaggccaaactcagttcgctgacccgagagcccatcaaaaaacaattccatgatgttgttctttcattctacgaagaatatcttgttgaagagcaaaacgatgggaagagtgttgaagtgaaaaaacttaagatcaaagaaaacgcttccactcagcaattagagaaacttgccgcgatctttgaaggcatcggctatttcagagcgggggatgttattaaaggcaaactaactgggctattctcataatgaaaaaaccagatactcaggtatcggccttgctggtgcagaagcaccagcttgaacaaagcgagcatcaattgggtgaccttgatgctgctctagaagcgcttaacgctttgcaaactgataccgaagcttctttagatgaaatgattttggctatggatggtgttctggaacactcaggtatcacgtttgatgaggatatccacacaacggtttctagtgaattcagcgataccttgaatcctgtttgaccacgtcatcgtccagtatcagtaaactgtcgatgatagaaacaatagcgttcaccagcgatatggactgggaaacctattcccagtccatatcgcagtatgcccataaacacaatatcgatttaatagtcgatccgtttagcgccctgatgtctccaatccaaagaattgctctggaaaaacgtattcaggaagacttgaccttaaagactgcccgctgcgacaaatatgattacatgatcgctggcacctgtggcgttattggcggacttatcgatatttttctggtaggcgtacctggagcaggaaaactgacccagcttgcagataatgcagtggacggtgccgttgagaaattcgcttcagcctttggatggaagggcagttcagaagcaagcgattcgacaaaaagcgctatcggttttctggagagaaaattcaaaatcaattatgaccatcggcatggcggagatgttgacggtttgttcaggatgaacacgaagaatcaccatattaaaagtctcgcccactccccggacttagtcggtttatttttctcgatcctggatcaatttaccagtacggcacattttgtggcagacggaaaattggtttccgtagataccgagacttttgagcttaaagggaataacgttgtctctaaggtatttagtggtttcgtaaactggctgggccaccttttctctgatatggcaggttcttccggtgcagcagggagaggctccggtatccccattcctttcttttcattacttcagtttattaatgtgggtgaatttggccagcatcgccagtctttcgcaaccgtcgccgtccaggttttttgagaaagggtatgacttacggcatggattagcgatggcgatccccgtcatgattactgagttgcttgtgcgaatcacctggacggttaaacaacgttgctatcataagaaggactggggtgaatgtattccttcagcaaataaccctgaactcaggcgaatgttgcttgtggcgcatggaaccttgtgtctgatggatgtaggagatgcggcacttcgttcaggaggcgaaatgattcagttcctcctgagaacgaacctcatcggctggacgaggtttggaattctagcgattaaagaactccatgtctggtataaagcaggcggaattgatgccaatgctgtagatgaatatatggatcatgaacttcggcgaatgctaaaagcggggtagcgttacggctttgttgaataacattacgtttgggtgcttggctgtaaaaagctaggcaatggcgtatctgtcgacgcaatgcagaaaaggcaacttaattgcgaaacagaaatgttcgg

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		tgagttgcttgaccgtcctatggcagctaagtgccagaagtcgacgttgctaacatcagtatgtactcatcggcacagtccatgtc agagctattaactatagataaaaattcaataattaataaaataagaaccatctttctaggtggttcttattattaacaataaatat tacgatttcaacgagggttagaatg (SEQ ID NO: 71)
75	pLG077	cctggtcctgccaattgctcccccagccatatgacataatccttttgaataatagggtttttatgcttgtactctagcccattcgc ggtatcattttacgatctctcttccagttttatgcttaccgcctttgcctatcgtagaacaatgccgggaagcgttatcagcgatt aagggcaaggaatgggcttctggatatttgttattatgctggcggttatctggcttctgttttccaaaaagaaaaaatcgccgccc cccagagtaaacaacaaaatcatcaccaaaataaatcattcatctcgacagaaatctctcaataagccagataacagcatgacaaa tatgcattctcaggcctccgatgatgacgaactggcaacctttacttttgtgaacgggcagacggttgaatacagcaccagccgcc agccgtcacgagaaaacgccgcccgtagcaataccactccagcgcgatgggtcaaaccgggagaaagcatcaccattcaaaatgtc gtcattaatcacggttatttttatttcggcgggcggttaaaaacacattcatcaggagaatatggatatctttataacgatgactc cgacgcttcgctggttaatgacgcttttcccatcgagcctggttcacggcattattatgatgagtcactgggatactggcccagct ttgccacactctcccctcgctgccgtggcgcctatcttgactggctggcaagcgatcgcagcgatgcgagctgccccgttggctat gtttttatctatttttacggtctagaacgccgcgtactggccgatggcacacaagaagccatttctgacgatgaattcaaagcatt attcgaagagatatcgcgcctgagaaccgtatttcaggcaagcggttccttccggcattatgcaacgcagttgctggaaatgatga tcgttctccgaccgaagttgctttctatatataccgaaaacgaatatttctcatcgaggagttcattactgttcagattaaatcta gcgactgtggtcgataaaggacaacctatttgtgccgctctggcactggcatggatatactattttcctgattacaccctgcgcac gcctgcccgtcgatgtcatgctgaattttccgcattattcaaacagcgttatactcaaaaatacggtgacggtattgtcgtcaaac ccaataaaacacggttgtatttaagctataccccccgccagtggtacgcttcgggaacttcaggtaaaaaaacagatggatcttccc gatcccagcgttttaaaagccccagttcagaaattaatttctgttgcagaatcctgtatcaacgcgctggatgcctacagtcgcta tctcggtaaaaaagatgcctcaccaagtgatgtcgccgccatcatgctgcttcccgatgaaatactgaccgaagatgcagaacgtc tatttgctgaatttaaacactgggcagatgagaaaatccgtgaacattcaggactggcgacagtggctgatttctgggccagactg ggtatgcctgtaccggataagattaataagaaagaagccgagctgatgcaaaatttcgcccggcgagcaggctacggcattgcgcc ggatatgcgctatcaccttgtcagaccggatccagaaggtcatcttgttttatttcctgaagggcatgcggaattctacgtaccgt cggcggaatttacgtcagtctctgtggcgcttcggttgggtgccatgattgcacaaatggacaagcgcgtggatgttgctgaacag gccgcgctggagaaaacgattaatcataacgatgcgctgtcgccaacagaaaaacgttcgctgcacgcctacctcacctggcggct caatacgcctgcaaatcaggctggtctgaaaggtaaaattgagcaactcagcgataaagataaatccactattggcaacgtgatta tcagcgtcgcctgcgcagatggaaaaatcgatccggctgaaatcaaacaactggaaaaaatctacgccagcctcggtctggacagc agtgccgttaccagcgatatccaccgactgtcaaccgcagaaacaactccgacagctacgttacaaaccccatcagcgacgagcgg cgcgttttctcttgatgaacggatccttgcccgtcatgaatccgacacaacggacgtacgccagttactgaacaccatcttcaccg aagatgaacccgcagacgaatccccagcggagatcccgccacacgctggcgcaggtcttgatgaagcacatcatcaactttaccaa cgtttgcaggaaaaagaacgctgggcgcgaaacgaagtcgctgagctatgccagcagtttaatttgatgctaagcggcgcgattga agcaattaatgactggtctttcgaacaggttgacgccccggtgcttgatgatgacgatgatatttacgttgacctggaaattgcac aagaactcaaaggataatttatgtctggcattcgtattcgtctcaaagaaagagacgctattattcagtcactgaagtcaggtgtt acgcctaaaattggtattcagcacattcaggttggccgggtcaacgaaataaaagcgctgtatcaggatattgagcgtatcgctga tggcggcgcaggattccggctgattattggggaatatggctcaggtaagacattctttttaagcgttgtgcgctcaattgcgctag aaaaaaagctggtgacaatcagcgccgatttatccccggacaggcgcatccacgcgacgggtgggcaggcgcgtaacctctactcc gagctaatgaaaaatctatccacccgaaataagccggatggaaacgcattattaagcgtggttgagcgctttatcacggaagccag aaaagaagcagaaagtacaaatgtgtcagttccgacgattattcaccaaaagctcgccgccctgtctgatatggttggcggttacg atttcgccaaagtcattgaatgttactggcagggccacgagcaggataatgagacattgaaatcaaatgccatccgctggctaaga ggtgaatacaccacgaaaaccgacgcccgtaacgatctgggtgtgcgcaccattatttctgatgcctctttctacgattcgctaaa gctgatgagcctgtttgtccgtcaggccggatacgcgggtctgctggtgaatctggatgagatggtcaatctgtataagctcagta acactcaggcccgcgttgccaactatgaacagatactgcgtattctgaatgactgcctgcaaggacggctgaatatatcggtttt ttacttggcggtacgccagaattcctgttcgatccgcgcaaggggttgtacagctacgaagcgctccagtcccgactggcggaaaa tagcttcgctcagcgggctggtgtcattgattattcgtccccttccctgcacttagccagcctgacgccggaagaactctatattc tgttgaaaaaccttcgtcacgtttattccggcggcgatgcggataagtatctggttcctgatgatgctctgacggcatttttacgc cactgtagcaacactattggcgatgcctatttccgtacgccacgaaacacgattaaagccttcctggatatgctggccgtgctgga acaaaacccatccattcagtggtcacagttaatcgccggtgtcgcgatcgcggaagaaaaacccagtgatatggatgaaataacat cggcagaagatgccgatgaggacggtctggccgacttcagattatgatgaacgaataccagcggctggatccacggatacagaagt ggatataccggcagggatgggccgatctcagggaactgcaaaaaaaatccgtttcaccgatattagcgggcgatcgggatgttctg atcagcgccgcgactgccgcaggtaaaacagaagcgtttttcctgcccgcctgttctgccattgcggatattcagggcggctttgg cattttatacatcagcccgcttaaggccctgattaacgatcagtatcgaaggctggaaaacctcggtgatgcgttggagatgccgg tcacgccctggcatggtgatgttgcgcagagcaaaaagctgaaagcaaagaagaatcctgccggtattttgcttatcaccccggaa tcgctggaagcgatgctgatccgcaatgcgggatggttaaagcaggctttcgcgccactggcatatatcgccattgatgaattcca tgctttcatcggttctgagcggggtatgcagcttctctctctgttaaatcgagtcgatcacctgctgggaagaatcaacaatccag tcccccgagtcgcactcagcgcaacgctgggggaactggaacaggtgccgttatctctgcggccaaatcaacgtctgccctgtgac attattaccgacagtcagactcacgccacgctaaaagtacaggtgaaaggttatctggaaccgctgaccacctcgggccagcaatc tccaccgtcggcagagacgcaaatctgccatgatatctttcgcctctgtcgtggtgattcccatctggtgttcgctaatagtcgca aacggaccgaaagcattgccgccacgcttagcgatctcagtgaagcgagcatcgttcccaatgagttctttccccatcacggatct ctgtccagagatctgcgtgaaacgctggaacagaggcttcaacaaggcaacttacccaccaccgccatctgtacgatgacgttaga gcttggcatcgacatcggtaaagtcagctccgttgtgcaagttaccgcccccattccgtagccagcctgcgtcagcgaatgggac gctccggtcggcgcgactcgcctgccgtattgagaatgctgattgccgaacatgaactgacgccaacatcaggcattgtcgaccag ctcaggcttcagcttgttcagtcgctggccatgatccgcttacttatcggcaacaaatggtttgagccagctgatacccggcagat gcactattccaccctgttccatcagatcctggcgatcgtggcgcagtggggaggcgtgcgtgcggatcagatctggtcacagctat gcctgcaagggccatttcagaaagtccggatctatgacttcaaaacgttattgaaacatatgggggagcaccagtttctgacccag ctctcaagcggcgaactggttctgggcgtcgagggcgaacgtcaggtaaatcaatacaccttctacgccgtgttcagcacgccgga agagtttcgcattgtggcggggagcaaaacactgggctccattcccgttgattccccactgatgcctgatcaacacattattttcg gcggtcgacgctggaaggtaaccgatatcgatagtgataaaaaagttatttatgtcgaggcgacaaagggtgggcagccgccgtta tttggcggacaagggatgtccattcatgatgtcgtccgccaagaaatgctcactatttatcgggaaggcgactaccgcatcaccgt tggcaatcgcaaggccgattttgccgataccacggccaaaaacctgtttgatgaagggctgcactgttttcgcaacaataatctgg cttcggaatgttttattcagcagagacagcatgtctacattcttccctggctaggcgatcaaaccgtaaacacgttgtcggcatta cttatccaacgcggtttcaaggcgggctcatttgctggtgtggttgaagtagaaaaactacggtctcggaggttaaacaagcgtt

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		attcagcgcacttcaggaagggctaccttacgaatcccgtcttgccgaaagcatcgttgaaaagtgcctcgaaaaatatgatgagt atttacccgagacgttgctgacgcaggaatatggattacgtgcttttaatattgaacgcgtgacggagtggttgcaggggcattta tattaaggggaagaaga (SEQ ID NO: 72)
76	pLG078	cgtgattcagttcgccagactgcagcgttttccatgaatataactccatctggtttagaaagagttccaatctaacgatattggga ccagaatcacaggcggcagtggctttacgcttacaataactattctatcctgacaattttaagcctcgtttgttacgatgtaaccc tataactatgtggttcctcaacctttttttgcccaaaaaatgcccaatgaagtccaaagtggaaaacagatggttatccgttgatga gattgcagattacctcgcgattaagcgagacacggtatacaagtagatcgcaaagaaaggtatacctgcacacatgattggacgcc tttggaaatttaaaaaggatgaagtagatggctggatacgcgatggcaaagctggcgaaaacagtaatcaagaataaaaaagcaaa tttaggagcagtttaatgaaaaccgtacgtagtgcatgccagttgcaaccgaaggccttggaaatcaatgtcggcgaccagattga acagcttgatcaaatcatcaacgacaccaatggccaagagtactttaaaaagaccttcatcactgacggttttaaaactttgctct ccaagggtatggcacgcttagccggtaaatcaaacgatactgttttccacctgaagcaagctatgggtggtggtaaaacccacttg atggtcggctttggtttattagcaaaagatgctgcccttcgaaatagccacttaggatcaatgccataccaatcagattttggctc agccaaaatagcagcattcaatggacgcaataatcctcattcctatttctggggtgagatcgctcggcagctaggtcgagagggtg tattcaggagtactgggaatccggagccaaagctcccgatgaacaagcatggataaatatttttgatggtgaggaacccatccta atcttgttggatgaaatgccaccatacttccactactacagcacccaagtccttgggcaaggaactatagctgatgtagtgacacg ggctttttccaatatgttgaccgcagcgcagaagaaaaagaatgtatgtattgtagtttccgatcttgaggcagcttacgatacag gaggcaaactgattcagcgtgcattggatgatgctacgcaagaactcggacgcgccgaggtatccattacgccggtaaacctcgaa tccaatgaaatctacgagattctgcgtaaacgtttgtttttgtctctgccagacaaaaatgaggtctctgaaattgcgtcgatcta tgcatcaagacttgcggaagccgctaaagccaaaaccgtagagcgcagtgcagaagcattggcaaatgacatcgaatctacttacc cattccacccaagctttaaaagcatcgttgctttgttcaaagaaaacgaaaagttcaaacaaacccgtggtttgatggagttggtt tctagactgcttaaatcggtgtgggaaagcgatgaagaggtgtatttgatcggtgcccaacactttgatctttcgatacacgatgt tcgtgagaagctggctgaaatttcagaaatgcgcgatgttatcgcaagagatctttgggactccaccgacagcgctcatgctcaga tcattgacctcaataacggcaaccactatgcacaacaggttggtacgctattgctaacagccagcctctccaccgcagtgaactca gttaagggcttaaccgagagcgaaatgctggaatgtttgattgatcctaaccatcagggtagtgactaccgaaacgcattcactga acttgctaaatcagcttggtatttgcatcaaacacaagaagggcgcaattacttcagtcaccaagaaaatctcaccaaaaagcttc agggatatgccgacaaagcacctcaaaataaggttgatgaattaattcgtcaccgactagaggaaatgtatagaccagtcacgaaa gaagcatacgaaaaagtactaccactccctgaaatggatgaagcacaggccacactgaggagtggtcgtgccctgttaataatcag cccagatggcaaaacaccacctggtgtagtcggcaacttctttaagggcttggtaaacaaaaacaacattctggtattaacgggcg ataaatcctctattgccagtatagaaaaggctgcacgccatgtttatgctgttaccaaggcagacaacgaaattacagcatcacat ccgcagcgcaaagagttggatgagaagaaagcacagtatgagcaggacttccaaactacagtgctctctgtattcgataagctcct gttccccggtaacaatcgaggtgaagacgttttacggcctaaagcgctggatagcacctatccatccaacgaaccatacaacggtg aacgccaagtcgtgaagactctcacgtccgaccccatcaagctttacacccagattaacgaaaatttcgacgcactgagagcccga gcagagtcattgctgttcggtactttggatgaggcaagaaagacagatttgctcgataagatgaagcaaaaaacacagatgccttg gttgccaagccgtggcttcgatcaactcgctatcgaggcataccagcgaggtgtatgggaggatttaggcaatggctatattacga aaaagcccaagccaaaaaccactgaggtaatcatcagcgaggactcatcaccggatgatgccggcaccgttcgtcttaaaatcggc gtggctaatgcaggtaacagcccacgcattcattatgctgaagatgacgaagttaccgaaagcagcccagtacttagtgataacac gctagcaaccaaagcattgcgagtgcagtttttggcagtagaccctaccggtaaaaaccttactggaaacccaaccacctggaaaa atcgactgacattacgcaatcgctttgacgaagtggcgagaacagtcgaattgttcgttgcccccccgtggcacaatcaagtacacc ctagatggttcagaagcacgtaatggtgaaacctacaccgtgccaatccagctcgctgatcaggaagccactatctatgtctttgc tgaatgtgatggcttagaagagaagcgaaatttcacctttgcggcagcaggttctaaagaaataccgatcataaaagataagcccg ccactctggtcagcccctcacccaaacgtatggatagctcggcaaaaacctacgagggtttgaaaatcgccaaagagaaaggcatt gagttcgagcagattagcttaatggttggatctgcaccaaaggtgattcatatatcgctaggtgagatgaaaatcagcgccgaatt cattgaaaccgtattaacgcacttgcaaaccgtgttaagtccagaagcccctgtggtcatgaccttcaaaaaagcctacacacaga ctgggcatgatcttgagcaatttgttaagcagcttggcattgaaatcggtaatggcgaggtggaacaacgatgaataaaaccgttg attttggggcaccgtcagaattcggtatgcatcacttctatgtggagattcccgcagcgccccgtgacgctgttgtgatctatgaa gactatggctttgacggtgaagattctcgccgagaaacagtagagtgtcgcctgatattagccagagagctctggactaagatccg cgatgacgttcgccgtgactttaacgctcgcctaaagattaagaaacaaagctccggtacttggtctaccggtaaagtgaagcttg accgctttcttggacgtgagttgtgcgttcttggctgggcagcagaacatgcctcacccgatgaatgtctggttatttgccaaaag tggctggctttacgcccagaagaaagatggtggctttacagtaaaaccgcagctgaagcaggtcgtgatgatcaaacacaacgagg ctggcgtaaagcgctctattgcgcgctatcggatggagccaatatcaaattggaaaccaaaaagaagcccaagtctaaaaagctac aagttgaagatgagacccaggatctgtttgggtttatggaaaagggagagttttgatggccttgcaaccgtttgaatggagagaca aaccgtctcttattgagcacctgttcccggtacaaaaaatatctgccgagacctttaaagaacgaatggcaagccacggtcagttg ctggtgtcgttgggtgcttttggaaaggcagaaaacctctcatcttaaacaaagcgtgcattctgggctcattgttaccagcaac tgacaacccgcttgaagatttagaggtatttgagctgttaatgggcatcgactctgagtcaatgcaaaagagaattgaggcttcac taccagcatcaaaacaagaaacaatcggcgattacttggtattaccctatgccgaacaaatcaggattgctaagcgcccggaagaa attgatgaatctcttttcgtccatatttggaatcgggtcaacaatcatcttggtacttctgctcacacttttgcgcaactagttga ggaactaggtgttgcacggtttggccataggccaagagtggcagatgtattttctggttcgggtcaaattccgtttgaggctgctc gcttaggttgcgatgtctatgcctctgacttaaacccgatctcctgcatgcttacttggggcgctttgaacgttgttggtgcgagc gcgcaaaaaagagtagaaatagacaaagcccaacgggatatcgttaagaaagttcaaaaagagattgatgagcttgacattgagtc cgatggccgaggatggcgagcaaaggtattcctatactgcgttgaggtgacctgccctgaatccggttggcgtgtgcctttaattc caagtttgattatcagcaatagttttcgagttgttgctgagcttaagcccgttcctgctgagaggcgatatgatattagtatccgt gaagtatcgactgatgaggaactggagttctataaatcaggcaccatacaagatggcgaggtaattcactcgccagatggaaaaac tcagtatcgcgttaatatcaaaacaattcgcggtgactataaagaaggcaaggagaacctaaacaagctgcgaatgtgggagaaaa cagactttgctcctcgtcctgacgatatttttcaggatagattattttgcgttcaatggatgaaaaaaaaacctaaaggatcgcag tattactacgaatttcgtactgtaaccaatgacgacttaaaacgcgaaaaaaaggtaatagaacatgtcgcatccaaattagatga ctggcagaagcaaggtcttgttcctgatatggttattgaagcgggcgataaaacggatgagccaatcaggacgcgaggctggactc attggcaccatttattccatccaaggcagttgctatttttgagcttggtgaacaaatattcactcgcagaaggaaaatttaacttc ttgcagtgcatgaatcacttgtccaagctaactcgctggcgaccccaggccggtggtggtggcggttctgcggctacatttgataa tcaggcgctcaatactctgtacaactacccagttagagcaacaggatctatcgaaaatatcttggctgctcagcacaaccactgtg gaatcagcgagaatgtttcctttgtggttaattcacatccagcgccagagttagatgtggaaaacgacatttatattactgatccc ccatatggcgatgctgtcaagtatgaagaaatcacagagttctttattgcctggctgaggaaaatccgccgaaggaatttgccca

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		ctggacttgggatagtcgccgatctcttgcggtaaaaggagaagatgagggtttccgtacaggcatggttgctgcttatcgcaagatggcgcagaagatgccagacaatggtttacaggtgctaatgtttacccatcaaagtggcgctatctgggcagacatggctaatatcatttgggcgagcggccttcaagttactgccgcatggtacgtagttactgaaactgactctgcattacgtggtggttctaacgtaaaaggcaccatcatcctcattttacgcaagcgccatcaggcattagagaccttccgcgatgatttaggttgggaaatcgaagaagccgttaaagagcaagtcgaatcgttaatcggattggataagaaggttcgttcccaaggcgcggaaggcctctacaccgacgctgacctgcaaatggctggttacgcagccgcgttgaaagtactgacagcttattcccgtatcgacggtaaagacatggtgactgaagccgaggcaccacgccaaaaaggcaaaaaaactttgttgatgagttaattgatttcgccgtgcaaacggcagttcagtttttggtgccggttggcttcgagaaaagcgaatggcagaagcttcaagcggttgaacgcttctatctgaaaatggccgaaatggaacaccagggtgcaaaaaccttggataactatcagaacttcgccaaggcgttcaaggttcaccattttgatcaattgatgagtgatgcctcaaaggctaactctgctcggctaaagctttctaccgagttcagaagtaccatgatgtcaggtgatgccgaaatgactggcactcctctgcgagcccttctttatgccttatttgagatatcgaaagaagttgaagtagacgatgttcttttgcatctcatggaaaactgcccgaattacctgcccaataagcaactgcttgccaaaatggcggattacctggctgaaaagcgtgaaggtctaaaaggtaccaaaacgttcaaccctgagcaggaagcaagcagcgcgcgtgtccttgcggaagccattcgaaaccagaggttgtaatctatggcgattaagcgcttttcatcccgcacagaaagattagatacggaattcctcgctgaatcgttgaaaggggctgctaagtatttccggattgcgggttatttcaggagctccatctttgagcttgtaggcgaagagattgcaaagattccagaagttaagatcatctgtaattccgagcttgatctggctgacttccaggtagctactggccggaatacagcactcaaagagcgctggaatgaagtggatgtagaagctgaagcgctactgaaaaaggagcgctaccagatttggatcagctattacattcgggtaatgttgagattcgcgtagtccctagggagcggttattccttcacggcaaagcaggctcaattcattatgcagatggcagccgtaaatcttttattggctcagtgaatgaatctaaaagcgcattcgctcacaattatgagcttgtttggcaagacgatgatgaagaaagtgcggactgggtagaaagagaattttgggcactctggactgaaggcgtcccgctgcctgatgcgatcttagctgaaatccaccgtgtatctaatcgccgggaagtaaccgttgatgtattgaaaccagaggaagtcccagcggcggccatggcagaagcacctatctaccgtggaggggagcagttacagccctggcaacgctcgtttgtgactatgtttctggaacatagggagatctatggcaaggctcgcctactattggctgacgaggtgggtgttggtaaaacgctatcaatggcaaccagtgcattagtcagtgctttactagacgatggacctgttttgattctggcaccttctacactcacgattcagtggcaaattgagatgatggacaagctcggtgtgcctgctgcggtttggtcctcgcagaagaaagtttggctgggtgtagaggggcaaatactctcacctcgaggtgatgcctcctctatcaaaaaatgcccttatcgaattgccattatctctaccggactgattatgcatcagcgggagaagactgactttgttaaagaagctggaatgcttctgaagaatcgtttcggtaccgttattctggatgaggcgcataaagcccgtattcgtggaggattaggagatcaagcttcagaacctaataatctcatggccttcatgctgcagatcggcaggcgtacacggcatctggtactgggtactgcgacacctattcaaaccaacgtacgtgagttatgggatttattgggtattttgaactctggtgctgaatttgtactaggcgatgctctgtcgccatggcatgaccatgaacaagcgattccgttgataaccggccagactcaggtgacatctgaggctgaagtttggcattggttaagcaaccccctgccgccaagcaatgagcaccatactgttcagcaaattcgtgactacctgtccattgataataagtcctttggatattctcatcgtttcgaagatctcgactatatgattcagagtctttggctctccgaatgcatgacacctagcttctttaaagagaacaaccctatcctacgccatacagtgctgcgtaagcgtaaacagctggaagatgacggtctgttagagcgtgttggggtgaatacacatcccattaagcgcaacctagctcagtatcagtcgcggtttgtggggcttggcattccgaccaatacaccattccaggtcgcttacgaaaaagcggaagagttcagtaagttgcttcagtcacgcactcgagccgcaggcttcatgaaatctttgatgttgcaacggatctgctcaagtttcgcatcaggcttaaaaactgctcaaaagatgttgaaacatacggtttctgacgaagacgaggatctagttgaagatgttgagcacttactttcagaaatgactcctgcggaggtcgcttgtttaagagagattgaaacacaactgtcacgccccgaagccgttgactcaaaactgaacacagtgaaatggttcttaacggaattccgtaccgatggaaaaacttggctggaacacggctgtattattttcagccagtattacgacacggcggagtggatagcgaaagaactggccaagtccttaaaaggcgaagtggtagccgtttatgctggcgttggtaaaagcggcttattcaggggcgaacagtttaataacgttgaacgcgaattgattaaatccgcagtgaagacgcgcgagattctattagtggttgctacggatgccgcctgtgaaggcttaaacctgcaaaccttgggaacactcatcaatgtcgaccttccctggaacccatctcgtttagagcagcgcctcgggcgaatcaaacgttttggtcagacacgtaagtttgtggatatgctcaatcttgtgtacagcgaaacacaagacgagaaagtttataacgtgctgtcggaacgcttacgcgatacatacgacattttcggcagccttcccgatacgattgatgatgaatggatcgacaacgaggaagaactcaacactcgcatggatgaatacatgcatgaacgaaagaaagctcaagatgcgttctccgttaagtatcgcggtactctcgatcctgatgctcatctctgggaacgttgcgctacagtactgtcacgtagggacattgtaagtaagctcagcgaaccatggggaagctaattatgttgtgatgtggatgccccgctcagccaaggtcctgcacaactatgttggatgctctttttagagggctacatcatgaattcgatcaaagttattggtacaattctgagtaaatctgtctctcagggtatccatttcgagtg (SEQ ID NO: 73)
77	pLG079	gccagtcgcttgcaaagtattgagaattgatgtttatttgtgttttgaggtggtctttgaaaccaattttcgttgtcaggtcgagtattgggtgcagcagacgctattcaaacattccgtcccggttatccgaaggtttccggctcggtagaaggcctgaagcatgtctctggttttgaagacggttcgggcttttccgagaggtcggactaccgaagaattgcttgttctcgtcggtgcggctttctcaaatgacaagcggcttgcggctctcagcgaactggagacgctatttcgcgatggtttgatagtgaaaggcaaggacggtcgctggcgtgcaaaggcagatggtttcaaacccagacatgagagcgtgtcggcttcgagaggtggagggcctgagggcttcgttgatgtcattcacgctgccaatgcattcttctcctcggaaccgacggcggccgaactacctgatcaagaagacgaaagttcagatgctcccgatccgcaagcgctactgagatattggcgctcggccttgcgtgccgatccacgaggagccacgacccaggttctcgacaaacatggaatcgagtgggccttgatctctgggcgtggccctatcggtccagaagaagggcaaacgctgactgtttcaatcgaactcgacgcgattgatcctgcctttcgagaggctctggtgcgaagggaaggtcacgagaacgcgcttgcagtgggttggccgatggcggtcggacgacgtggcggagttcctgtctttcgacccgttggcatgttagcagcagcttgggatcgtaaggatgaccgtctaatcctgacgattgatgccgatgacgttttggtaaaccctgattgggtcaaaagtgccgctcgtgccagcggctggaagcgcgacgacctcgctgaccttttttcgtggacgatgggctggggctgcgggctcaggattttgtggagaaggtaaggattgccgttgccagtcagatacgtggtcgcgttgtcggcgagaatctcgccacacagctcgatgcctcggctcaagggattttgacagcgccgcgatcttcctaccgactgactcttctttcaccgcggggctgctcgtgacctggatgccattgcgacatggccgaaggaccgccttgagagaactgcgcttggcgcggtattcgggtttgaccttcaagacggcacggacaaggctgctgcaatcgacgcagttccgctgaacaaggaacagttgcgcgcggttcgatccgcatgccaagcgcctttgaccgtcgtgaccggtccgcccgggactggcaaaagccaagcgatcgtatctatggccgcgtcagtgctcgcagatggtggcagtgttctcgtcgcctccaagaaccatcaagcgcttgatgctgtggaggaccgtcttggctctcttgctccggacgtcccattcgccatccggacactgaacccgaatgacgaggcggatacgggcttcaaggacgccctcaaacaactcatcgacagcgaaaatgtgacgcgcaacgcatctgtcgacgaattcgcattaggcgagctcaaaagcgacgcgatcgcgagaagcgaagtggttagcgtgatcgataagatcacggaaacggaatgcgaaatttccgatattctggaccggattcaagtccgagaggatcgcgggcgccctgacaaccaagactctgaagacgtggatccgagacaaagtctcttactccgctttgtctcttggtttggatcgcttttcgccaagcgtcccccccaaagtagcgccagtgacagatcattcttcgtcccgccgcggaatgaacgtcaaagagcttcattgcgcgctggcagaaaaaagatatgaacgcgatgcgctcgggacacctgacgatccgatcgccttaggcgagaagatccgggaagcgaccgagaatcttctgcctcgcattctgtccgcccggacacatctcccagaggatgagaggcgcgaaatcgcagaactctacgatgactggacattcgacggggacgggg

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		acatcccccctactgatctttcgcgcgtcctcatttcgcatcggcctttgtggcttgcatcgatcttgggcacgcctcgacgcatac
		ctcttgatgacgggctgtttgacctcgtgatcttcgacgaggcgagccaatgcgacatcgcgacggccgttccgttgctggcgcgc
		gcgaagcgggccgtcgttgttggggatgatcgacaactgtcattcatccctcaactgggtcaggcgcaggatcgcaatctcatgca
		ggctcagggcctaccggtcgccagaatgggccgtttcgcccagagtcgccgttcgctattcgatttcgcatcgcgcgtgtctgttg
		ccgacaacaggattactctgaggcaccagtatcgttcagcaggccccatcgtcgattacatcagcgagaacttctacggaaaccag
		ttgcagacctcgtatgacccgaggcgactgaacgtgccagatggggtgcgccctggcctcgcatgggaacatgttcctgctcccgc
		ggtcccgcaaatgggcaacgtcaatccgtcggaagtaagcgcgattgttaggcacctgaaaaagctgatcgttgaagacaaataca
		ctggcagcatcggtgtcataacgccgtttcgcgctcaagtggccgctatcgagaacgcggtcgatgccgtcctggatgaaccgaag
		cgcattgcctgcgagctcaaggttggcacagttgacggttttcagggacaggagcgggatctcatcatgttctcgccttgcgtcgg
		tccacgcagcccgcagtctggcttgaccttctttcagcgagatacgcgccgtttgaacgttgcgatttcgcgggctcgggcggtcg
		cgatgatcttcggcgatcttgatttgcacgttcagggcaatcaaaagcgctggccaagctcgcttcgagggcgacggaagcgcgg
		acgaaacggggcgaaggtgtgttcgacagcgattgggaacgcaaagtctatcacgctctgaaggcccgaggtctggatccgcagcc
		gcagcacgaaatagctgggcggaggctggacttcgcgttgtttggagcgaatgatgtaaagctcgatctcgaggtcgacggacgca
		gatggcacgaaagcccagacggtcgtcgaaagacgtcagacctgtggcgcgatcatcaactgaagtccatgggatggcgggtgcgc
		cggttctgggtggacgaactttcaagggatatggagggttgtcttgaccgagtcgaacaagacctatcgtaagtcgagcaggaaca
		ccgcggttgcgttggggctgggtggcgccgccatccttgcctcgggctttctcgtcctgcaagtcaactcgctcgatcgccgatat
		ggtcgtatcgaggaaaatctgagctactacaccggggaactccaatccgcgcagcagcaactggcttttgctcgtgagcagtttcg
		cgaactttctgaccaaaagcaaagcttgtctcaggaagtcgcgagcgccgaacgcagccttcaaagcgcggctcagagagaggcgg
		atgcgcaggctagtgtcgaagcaagccaggccaaattgactgctgagcgggaccgtttggccgaagcccaaaaaacgattgcggat
		gcgcagcgaattgaacgtgaaactgctcaagctttgctgcgaagaaatggcctcgaaacagaggtggtcaaactgaaaggcgatgt
		gcaggcccttaaggagagccagcaagagttgtctgctggtgttgaccaaacgcaatcggctgtcgatcgcctcgaagagagaagag
		ctgaacttcaacgtgaagtggatagactcgcgcccgccgttgaagaccttcgtgcacaggagcggcttgtcgaacaactgcgaggt
		gacgaggatcgtctcgaacagagcctcgacgatttgaatgcgaacattgcaattgcacggactgaattggcgaccagcgcggaaaa
		ggtcgatgcggccgaggagaggctgcgtgcagggcaggaacaaatagcatccacagaagctcaacttgaaacactgaatttcgaag
		tcgatgacctcgagtcgagacagggcgaactgcaggcaagtgtctcgggagcagagacgcgtctttcttcattgcaaaatgaactg
		gagatcgcacagaacgcggtgacgcgagctgatgcgcagcgcgctgaaactacagaagcactcaacatcgctcaggaacagttttc
		gacgcgaagcgctcagctctctaccctccagtcgcagattgcatcggcagaggaagagcttgccgaacttgaagagagacgggcgg
		aattcagcagattgcaggctcaaatggaccagctgcaagcacgtcgaacgacactagaggaggttctccccgatcttgagaagcga
		gttcaagcagagcgggctaatttgggttctatcacgacagaagtggagacagagctcgggcgagttgctgtactcaaaggccaggg
		ttccagtctggaggccgacatcgagcgcctccaagagcgtcgcgacgaactcgggctggaaacgcagtccgccactgctgaggcgg
		aggccgcgcgcgcatcccttcaagctgagcttggtcaacttgcggaaaccgatgccctttcaagagcgcggactgccgatttgagg
		cgcttgagagaagctcttggagctgctgaaagagagctttccgaacttgaagagagacgggcggaattcagcagattgcaggctca
		aatagaccagctgcaagcacgtcgaacgacactagaggaggttctccccgaacttgagaagcgagttcaagcagagcgggctaatt
		tgggttctatcacgacagaagtggaaacagagctcgggcgagttgctgaactcaaaggccagggttccagtctggaagccgacatc
		gagcgcctccaagagcgtcgcgacgaactcgggctggaaacgcagtccgccactgctgaggcggaggccgcgcgcgcatcccttca
		agctgagcttggtcaacttgcggaaaccgatgccctttcaagagcgcggactgccgatttgaggcgcttgagagaagctcttgctg
		ctgccgatgatgagctttccgagacacgagcggaactgatggacggacagtctgtggaacaggaaccagtatcaaccattagtgaa
		ggcgctggcgcccgtgaaaacgctcagtctgacaactccgcgccatcgagcaccgacaattgaggtaaccgaaaatgcttacggac
		aatacaatacttgtgctggcgattgcgggtgtcctgatactgctcgccgtggttcaactttttctggccgcccgccacgaccgggc
		ggttacggcagcaggcccgatcgaagagcttgccgtctacgagaagcggctggaagaaaaacagcggctcatggacgatcttgaag
		ctgaagtggaaaaacgtcgggaggcaatggccgtcgttactgacctccgggctgaggtcgacggtctacggcgtcagaaggaggag
		ctccttacagaatgggagagtctccgtgaacgtcgcgacgaagttgcggcagttcgcaaggagactgaggacgccgttgtcgaacg
		ccagcaactcgaaacggagatcgccccgcttcgtgcggagtatctggagataaaggaaaggctggaaaaggcggaggagctcattg
		agcgcactgacgccttgagacgagagcacgacgaaatctccacacaggtcaaagatcttcgggacaagaagaggcaacttgaagag
		gccgaggaacgggtttctcgcctggaagagcgttccttcgaacttgagacatcgaatgctcggcttgagggacagaagtcttcgca
		tgaaagcgagttgtccgccttggaagcgcggatcgcctcggaacacggtgggttggcatctgcccaaaccgaacatgctcgcctcg
		atgcagaggttgcggctctgaaccaggaaacccgccgctccaggggcgaaatcgagacgctccaggacactcgaagcgcgcttgat
		gctcgattggcacacctcaaggccgagatagctcgccgagaaggtcgaaccgtcgacggggaaaccggcgaaacggatccgcttcg
		cgagctcaatgaaacaccaccggtcattacggagatgaggacctgggacaacgcgccccgcgagaacgaggcggatgccatcaaac
		gcgtcgaacgccgcctacgcgcaaagggtctcgactacccggctcgcacgcttcgcgcttttcacaccgccatgaaagtaaatgaa
		acaacgcagatggcggtccttgccggtatttccggaacgggcaagagccagctcccgcgtcaatacgcggccggtatgggcatcgg
		tttcttgcaagttccggtgcagccacgttgggatagtcctcaggatctgatgggatttacaactacatcgaaggcaagttccgac
		ccacagacatggcgcgtgcgctttgggcggtcgacgggcttaacaacgacgatgcggaacaggatcgcatgatgatgatcctgctg
		gacgagatgaacctcgcaagggtcgaatactatttctcggacttcctcagcaggctggaaagccgtccgcgtcccgatgacgtcga
		caatgaaaacgaacgcaaggacgctgtgatcgagcttgaaatcccgaacatggaacgcccccccaggatttttccgggctacaacc
		tcttgtttgcgggcactatgaacgaggacgaaagcacgcagtcgctatccgataaagttgtcgaccgtgcgaatatccttcgtttt
		tccgccccgaagaaaatcaaggacggacaggcagaaggaacggtcgagccgatttggccctttcgcaacagacatgggagagctg
		gggcggtcgagtgcgtctgtcgatggcggtcggcgtgtcaccaaccggattgaacaaatggttgatctgatgcgtgacttcaaac
		ggcctttcggtcatcggctcggacgcgcgatcatggcttacgcggcgaactatcctgaggttgaaggcggccgcggtgtcgacgac
		gctctcgcggatcaattagagatgcgccttctaccgaaactcaggggcgtggaaaccgacatggctggccctcagttctcgaggtt
		gatgacctttgtggaacgcgagctgggggacgacgccttggcccaagcaatcggtgagtcaatgtccctcgccgaggcaaccgggc
		agttcgtatggagtggagtcacgcgttgatgcggtttctggcccgtccctgggcggcgaaagcccttggagaggacgaagcctttg
		ggcccgaagactgtctgatcggtagctaccaggggggcgaacccaggcggctacgaatacgtgacgctcttgaggggaaacgtccga
		ggtagcgataccggaactgttctgtttccctatccaaagcgtgaggaagctgtcgggcccgcgcgtaagggcttcccggtgcgccc
		aaggtcggggcacgatcctgccactccggacgaagaagaaggcgcagaggcccttcgacacatgaacgaagttcttgcacgtatcc
		aagaactggaaggtgcgattgaagacccaagcgatacatgggggcgcctgagggatgcttggaagcgcgccgaaaatgaagccgaa
		cccaaaatggctgaaatcgtccggcaggcgcggggcatgcttccggtgcttcgcgatctggaaaaacgcatccgccgggttctacg
		taggcacagggagctaactccccttgatcgggtgcaggagatggatcggacctctatggtgtggctcagccgacagccagggcgaa
		gcatcgcggaacgtgcaggttcttcgcaacgaattcttgcgacggttcgccgtgagaatttcgatacgctcgagaaccgtgtcctg
		catgcctacacgcgtcttgccgcagatgttgcacgcgaatggacccgtgagcaccctcgtgcgaaggacagtgttcgctacaaaca
		ggttgaggcttttaggaaggcctgtcgagtattgtcgcgaacactcagtgacctcggtgtcatgatcgcgtcggccggcgtccagc
		caaactatgtgctcatgcaagatcgcagctatcgagaggttcatgagggatggctgaggcttctcttacgccgaaaaattgtagat

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.		
Row No.	Vector	Locus
		gatctttgggcttggcaggccgaaacttggacggatttctccgttctttcgatcattcttgccatcgacgaattggaagaggctgaacttgtcgctcagtcgccgatttcgtggagcggtgaggcaacaggcggacgctggttcaatcaggatcggccaatcgccgtcttttggctgcgcgacaccaaccgcattgttgaagtccaagcacgccctgagcgaccaggaaccatgttgagcgcggcacaagcgcacgtcgccctcagaatttccgatcccaaacgggctgaccttccgcgcaggatcgctgtctggacgccacatgccatgcgtagaattgatctcgaggatactgtgcgggggcagttcaactgcttcaccaaatccagcccctcgctcagacggaagttttgcggaatgggttgatcatgaccccagcacgtggtgtcgcagctgaagagagcgcaactcacggaagagcgatcgttacggcaatcgccataggcccagccggtgaagacctagcgaagggattccaggccgtgcgcgacttcattcgcagtgagctatacgaggtcgcaacatgatcgaccgaaaactatgcggcttcgatctcaacggatggagagatttcgttgcgaagaactggcgctccgtgccaggtgaagacgaggtcattggtccgaccgatatcgtcacaagtggccctctttcgtcgatcgtgcggatcggggaaagccgcctcgcaggttggatcggaggaccgcaggctgacattgctccgcacggtcgcggtggtggttggggtgatgtcgggtcagaacaaagacgcattcccgttcggtcactgctggaaatgcgtgatgacggggtcgaaaaactcgcccaggcacttgtgggatctgcgagcggttcggcaaacacagtcgtttcgatcgatgagggcccggatggcgatgaagccgtccaagagcaccttctcgaagcacttgcccgagggaagttccgaaatggctcattggtttggcgaccagttcttgccgccttgttcgccattcatcgcgatcaggtttcggaggggcagcttgtaggcgtcgtctcccatcagcgccaaggcttgtcagttcaaaagctgcgtattcgtagcgcaaggaatgtgctcgccccggagcgacgcgaggccgctgcccatataccgtgcgacgctggttacgagtccctattccgaggtgcccgcaacgccgctgtcggggcagagggtttttcggcgcgcacagctcatcgtgcgatcgcaagctcggtcggaaaagctggtttagggatggattgcaatcctgagatgctccgcatgcccaacggcgattgggagctcttggaccttaataaatttgacgcgtcggaagtggtgagtgtcccgagttccgagctcgatctggccgattgcgacgtcgttcttttcgagaccctttgtgaaggtcggctcaaaaaatgcctgagtgatgctatccaaagagcagctccagtcgaggtgctctctcttcccgcaacggctgttgcggaaggtgccttggaagcagcacgccgagccggggacggggaaccgatcttcttcgactttctaccacgattgtccaccatcgtgttcggatcggatggcgcaaagaatttcgatctcatacggaaagaagaaacgctcgaagcaggccggacctacagaagccctgaagcagcatctctcgcgataccggcagggcaggagagcgtctctgtctacctgaggaaagaggaagctccctggcctcgaaaggcaagggtgtcgcttggagctcctctgaagcatcaagctgccgtctcgctgtgggtcgaacagaaaccggccgccgggcgagcgcggatcctcatggaatcgccggacttggggcggaatttcgcggtggattgggatgaagcactggaagaggaacggccctggtctgagatcatcgagagcttggatacgcaagtgtcaattcccaaacgtctggttcttccctgcggcatggaggcatggcatgacagcgatcgatccgcaggtatgctaactttgctcgaatccgagcctaatcgcagccgcacggattgggcgacccttcggcaaaaactttcacagcgtccctttggcaaatactgcatctcaagtgacggcgacgtgcctccggagatcgcggcagaaaccctcgagcggtttgaaattctgaccagcaaagcgcttgaggttactgaaaagcgcctgaggggcgaaagcggctacggaacggaagacaatgaggctctcaaattcttgagttggcagttccgccgatgcccgcgcgatgtcgcgacgtggctgatggactgtattgaagcgtccgggcgcaaccatccgttcgtcaaacatcaagcaagttgggttctcgtatatcagggccttggccgcatcgtcggaaacgaagaggacgaagcgagagcaatgcggttgcttctgacttcgtccattgaggactgggtctggaaccgacaaagcgcggccatggcgttcatgctgtctcgttctgacagcgctccatcttacctggaacgagaagacgtagagaagctgaccaagaggactatcgcggacttccaacgtaatatcggcggccaatatacaatgtttaactacgcgcctttcttacttgcaggcctgataagatggcgtctcgttgatcctaaagctttggtgatcggggccgacccgttggcggatgacctcttggctatcattgagaaaacagagcacgacctgaaggcccgttgtgggtccaatatgaatttccaaaggcggcggtcgaagttcttgcctatcctccaagacctgaagtcagagctggcgggagaaggttcgaatcctgacctgttgttggatatctatggagcgagcggaacgtgaccatgagcgcgcaggtaccaagctgctggatcaagcctctgggacgagcagaagcccgtccgttgtataaaatcaccacagacacaagcaagagatatcgtgatactaaagcgctctggaggattcccatcagacctgatgaacatcgcaactgcatcgaaccagaaccatcttggcaactaggtgcggaccaaggactgaagcatgtgcctatcgctcaa (SEQ ID NO: 74)
78	pLG080	gggctgtttggttgaattaaaaatacgaactaaaaccaacaagagtcggaaaaaacttcaaaatgctgcttatggataatagtcatcttaaaaatgtacggaaaaagagactaaaatcagaaaaacatctgttatacattgacttaaagtcatcatctccgctatgagtcctcaatccaagttgacaaatgtttagccaggagttcccgtgaacgagcatctctctcatatggatgtacataccttgtttgaagaaatggacgagcaggctgatggaataacgtttaaatactcatttgatgacatagcaaagagcaacgcattggttgtcactgagtttgtcaatttgagcgtgacagcacggtagctttactcgccagccttcttactctcccggcacaccaatctcagtgtttgcgctttgagcttctgacgagccttgcactaattcactgcaaaggtcagcagatagcaaatatcgatgacgtgaaacgctggtatgtcactattggggagtcgagtagtatcgttggagaagatcctgctgaggacgtcttcgtcgcccttgttgataataaaaaaggtgattaccgtgtgctagaggggtttgggaggcggcaggttttatacacaattaatggtcgaaattgtatccgacatgccggatacgcaccgctatcgctcgctgaaacttgctatacaggcaattctccgtctctcagatgtcatttgtgctcgctctggcctttatcgttttcaggaaggcgcagacgaattccctgactctcttgacaccgctggtcttgatgagaaaacgctctgttcaagggtaacgttgtccgagcgttctcttcgagctgaggggatcaaacttgctgacttagcacctttcattcttgaaccttctcatataagtatgcttggaaatcaggtccctggggagggaatgcttgaacaacggccattgctccgcacacgcgatggtattgtggttgtacttcctaccgccatgaccattgcacttcgccaggcagtgataacatttgcaaagcgcacagaagaattgagcgagctagacaaagcgttagctaacgtctacagccttactttctccgagatgccggtcttcggtaatggaggaaggttaagaagactgacatgggagaagtacaaaatgagccgaacaacgatggtaacctccatcgtggatgctggtcatttgatggtacttcagttcgttttgccttccatacagcaatatgccgataccggtttcaacaacttgctacagctagatgaagagaccacgcaatttctagataactctgttgaacaaattacagttgacctcgccaaacaacccggctttcagcgtggcatcgtcgtgcgcattgcatgtgggtgggggcgggttttatgggggtccctccccaactgccagatggttggggatttgaatggatgtctggtgcggactttgtccggttcggggcattacccgatatgtcaccaattgccttctggcgtgtgcaagacgcagtcgaaacgatcaggcaagctggtgttcgattaatcaatatgagcggaactctcaatcttcttgggtggatacgtgccaatgatggccatatggttcctcatgaccagttaccagatgaccgtatcacaccggaacacccgctaatgttaatgattcccacgaatttactccgtggtatacgaatagcggcagacacaggatatgaccggcatcgcattagtgacaacaatggtaaatggcatcgagtgatgaggccttcggcagaagatttctttcccaccgagcgtcagagcaagtgctacgcatcaattgatgatcttgaagcgcaacggctgacctgtgtatatgaggggcagggtaatctttgggtaacgctcgaagctccagaaatggaagattggatgctcctcgttgagcttgccaaaatggttcgaacatggattgggcggattggcgaggcactggaggtcttgagtgagcaaccaataaaaaaatcattaaaggtgtatctgcattttgatggtaacgacaatatcggcagatttgatggtgagaattttctgatgatatgaatacattttggcgacttgaacgaatccatgagcatggggcgattcgtgtggttcttcaagatgggtatcttgcaggttttcgtctaccggataaccgtgcagaacgagctctggtgcgcgcactcggtacggcgtttgccacacttcttcggatgaaagagccagtagacaaaggggtcactgttgagcagatagcggtgcccaatgacagagcgcgcagcttccacataatgcaggcttatgacttcaaccaatatttaggccgttcactaactaaacgtcttttagctattgaagatatcgactcagccgcagcccgaattgagctagcatggcgtgctgtttcgacagatgcaccatcacgatatcagggtaaaaaggaagttggaaagctccttaatgatgtggttgatgtgctgatccaagacttactaagcgaactttcaagatttgaccgtaaacagacagtaatgcgattacttgaaaacgttgtaaaggcacgttgtgaagaggcgcactggcgtagtactgcagcagcggtccttggcttgcatgcaggagaagagggtgtcgaagagacgatagctcaagaaatgagccgttatgcgggcgcagcgttaacttcccggctaatcattgaacttgccatctgtgtgtgcccgacaagcggtggaattgaaccttctgatatggcactcagtaaacttcttgcacgggcatcactgctttttcgcataggtgg

TABLE 3B -continued

Sequences of loci of row numbers 1-78 of Table 3A.

Row No.	Vector	Locus
		tatgtcagatgccgtacgtttcggtgctttgcctgctgatattcgcatctccccttaggtgatctcctctttcgcgatgaactcggcaaaatggtgcttgaaccaatgctttcaaaagttactaacgaacggtttgaggaacaagcggcacaattcgagcaacactatgtgaaaactgccggaggggatgatgagaatagcaaacaagatagtgttgcggctgaaaccaccgaggaccaaaccgatattttccttgcattctggaaagcagaaatgggcttcactctcgaggatggaatgcgatttatccagttccttgagtccatcggaatagagcaagaatcagcaatcttcgagatgcgaagaagccaattagcggatgctgctaaatcggctgggctcgcagatgaaactattgatgcgttcctcaaccagtttatccttagcgcgcgtccgaaatgggatgtagtgcccgatggatttgacctttctgatatatatccctggaggtttggccgacgcctttcagttgctgtacgtcccttgttacagattgaagagagtcacgatccactaattgttatcgcaccaggactcttgaatctgtcccttaaatacgttttcgatggcgcatacactgggcaatttaagcgtgacttctttcgcacagagggtatgagagacacttggttaggtggagcgcgggaaggacacacattcgaaaaaactttggagagagaacttcgtgaaataggctggacagttcgacgtggcataggctttcctgaaattcttcgcaggaatctaccaggtgatccggggatattgatcttcttgcctggcgctcagaccgcaatcaagttctcgttatcgaatgtaaggacctctcacttgctcgtaattactcagaagttgcctcgcaactatctgaatatcaaggtgatgacataaagggcaaaccagataaactcaagaaacaccttaaacgcgtattactagccaaagaaaacatcgataatttgccaagttcacttcgatagcgaatcccgagattgtatcgtggctcgttttcagtggagcatctcccattgcctatgctcaatccaagattgaggctttggcaggaactaatgttggccgcccaagtgatcttctgaacttttgatagatatgctgtgcgataagacgccctggcaactaagttaatcgttcctactactgatagttttaaatcaagg (SEQ ID NO: 75)

Variants and Mutations

[0113] One or more components of the systems herein may comprise one or more mutations compared to corresponding wildtype counterparts. In some embodiments, the one or more mutations may be in the catalytic domain of an enzyme of a system herein. The mutation(s) may alter (e.g., increase) the activity of the enzyme.

Polynucleotides and Vectors

[0114] The present disclosure further includes polynucleotides comprising coding sequences of one or more components of the systems. In some embodiments, the present disclosure comprise vectors. The vectors may comprise the polynucleotides with coding sequences of one or more components of the systems. In one aspect, the present disclosure provides cells comprising one or more of the polynucleotides and/or vectors herein.

[0115] A vector refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. A vector may be a replicon, such as a plasmid, phage, or cosmid, into which another DNA segment may be inserted so as to bring about the replication of the inserted segment. Generally, a vector is capable of replication when associated with the proper control elements. Examples of vectors include nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g. circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. A vector may be a plasmid, e.g., a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques.

[0116] Certain vectors may be capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors." Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. A vector may be a recombinant expression vector that comprises a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory elements, which may be selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. As used herein, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence (e.g. in an in vitro transcription/translation system or in a host cell when the vector is introduced into the host cell).

[0117] A vector may be a viral vector, wherein virally-derived DNA or RNA sequences are present in the vector for packaging into a virus. Viral vectors also include polynucleotides carried by a virus for transfection into a host cell. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g. bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome.

[0118] In some embodiments, the polynucleotide herein may be a part of a vector or a pair of vectors that is/are introduced into cells for inducing diversification (e.g., site-specific mutagenesis) of the variable region and/or support replication of the molecules. Non-limiting examples of vectors include plasmids and virus based vectors, including vectors for phage display that may be used to express a diversified variable region sequence. Other non-limiting embodiments are vectors containing variable sequences that have been subjected to the methods of the instant invention and then removed from an operably linked template region, including by preventing the expression of template regions, so as to produce without further diversification quantities of the variable region-encoded protein for uses including as a diagnostic, prognostic, or therapeutic product.

Regulatory Sequences

[0119] The vectors or polynucleotides may further comprise one or more regulatory sequences. In some cases, the regulatory sequences may direct the expression of the nucleic acids in specific types. The term "operably linked" as used herein refers to linkage of a regulatory sequence to from a DNA sequence such that the regulatory sequence regulates the mediates transcription of the DNA sequence. Regulatory sequences include transcription control sequences, e.g., sequences which control the initiation, elongation and termination of transcription. In some cases, regulatory sequences include those control transcriptions. Examples of such regulatory sequences include promoters, enhancers, operators, repressor, transcription terminator sequences.

[0120] The variable region (or the gene overlapping or including the variable region sequence), the template region, and the coding sequence for reverse transcriptase may be operably linked to the same regulatory sequence (e.g., promoter). Alternatively or additionally, the variable region (or the gene overlapping or including the variable region sequence), the template region, and the coding sequence for reverse transcriptase may be operably linked to different regulatory sequences. In some cases, the variable region (or the gene overlapping or including the variable region sequence) and the template region are operably linked to the same regulatory sequence; and the encoding sequence for reverse transcriptase is operably linked to a different regulatory sequence. In some cases, the template region and the coding sequence for reverse transcriptase are operably linked to the same regulatory sequence; and the variable region (or the gene overlapping or including the variable region sequence) is operably linked to a different regulatory sequence.

Promoters

[0121] In some examples, the regulatory sequences are promoters. The promoter may be suitable for expressing the component(s) in the systems, e.g., the variable region, the template region, and/or the reverse transcriptase in desired cells. A promoter refers to a nucleic acid sequence that directs the transcription of a operably linked sequence into mRNA. The promoter or promoter region may provide a recognition site for RNA polymerase and the other factors necessary for proper initiation of transcription when a sequence operably linked to a promoter is controlled or driven by the promoter. A promoter may include at least the Core promoter, e.g., a sequence for initiating transcription. The promoter may further at least the Proximal promoter, e.g., a proximal sequence upstream of the gene that tends to contain primary regulatory elements. The promoter may also include the Distal promoter, e.g., the distal sequence upstream of the gene that may contain additional regulatory elements. In some cases, the promoter may be a heterologous promoter, e.g., promoting expression of nucleic acids or proteins in cells that do not normally make the nucleic acids or proteins.

[0122] The promoters may be from about 50 to about 2000 base pairs (bp), from about 100 bp to about 1000 bp, from about 50 bp to about 150 bp, from about 100 bp to about 200 bp, from about 150 bp to about 250 bp, from about 200 bp to about 300 bp, from about 250 bp to about 350 bp, from about 300 bp to about 400 bp, from about 350 bp to about 450 bp, from about 400 bp to about 500 bp, from about 450 bp to about 550 bp, from about 500 bp to about 600 bp, from about 550 bp to about 650 bp, from about 600 bp to about 700 bp, from about 650 bp to about 750 bp, from about 700 bp to about 800 bp, from about 750 bp to about 850 bp, from about 800 bp to about 900 bp, from about 850 bp to about 950 bp, from about 900 bp to about 1000 bp, from about 950 bp to about 1050 bp, from about 1000 bp to about 1100 bp in length.

[0123] The promoters may include sequences that bind to regulatory proteins. In some examples, the regulatory sequences may be sequences that bind to transcription activators. In certain examples, the regulatory sequences may be sequences that bind to transcription repressors.

[0124] In some cases, the promoter may be a constitutive promoter, e.g., U6 and H1 promoters, retroviral Rous sarcoma virus (RSV) LTR promoter, cytomegalovirus (CMV) promoter, SV40 promoter, dihydrofolate reductase promoter, β-actin promoter, phosphoglycerol kinase (PGK) promoter, ubiquitin C, U5 snRNA, U7 snRNA, tRNA promoters or EF1α promoter. In certain cases, the promoter may be a tissue-specific promoter may direct expression primarily in a desired tissue of interest, such as muscle, neuron, bone, skin, blood, specific organs (e.g. liver, pancreas), or particular cell types (e.g. lymphocytes). Examples of tissue-specific promoters include Ick, myogenin, or thy1 promoters. In some embodiments, the promoter may direct expression in a temporal-dependent manner, such as in a cell-cycle dependent or developmental stage-dependent manner, which may or may not also be tissue or cell-type specific.

[0125] In some cases, the promoters may be inducible promoters. The term “inducible promoter”, as used herein, refers to a promoter that, in the absence of an inducer (such as a chemical and/or biological agent), does not direct expression, or directs low levels of expression of an operably linked gene (including cDNA), and, in response to an inducer, its ability to direct expression is enhanced. Examples of inducible promoters include, promoters that respond to heavy metals, to thermal shocks, to hormones, promoters that respond to chemical agents, such as glucose, lactose, galactose or antibiotic (e.g., tetracycline or doxycycline). Examples of inducible promoters also include Drug-inducible promoters, for example tetracycline/doxycycline inducible promoters, tamoxifen-inducible promoters, as well as promoters that depend on a recombination event in order to be active, for example the cre-mediated recombination of loxP sites. Examples of inducible promoters further include physically-inducible promoters, e.g., particular a temperature-inducible promoter or a light-inducible promoter.

[0126] The promoters may be suitable for expressing the component(s) in the systems in desired types of cells. In some cases, the promoters are for expressing the component (s) in prokaryotic cells. Examples of such promoters include filamentous haemagglutinin promoter (fhaP), lac promoter, tac promoter, trc promoter, phoA promoter, lacUV5 promoter, and the araBAD promoter. In some cases, the promoters are for expressing the component(s) in eukaryotic cells. Examples of such promoters include the cytomegalovirus (CMV) promoter, human elongation factor-1E promoter, human ubiquitin C (UbC) promoter, and SV40 early promoter. In some examples, the promoters are for expressing the component(s) in yeasts. Examples of such promoters include Gal 11 promoter and Gal 1 promoter. In some cases, the promoters may be used for expressing the components in a cell-free system. In such cases, the promoters may be selected based upon the source of the cellular transcription components, such as RNA polymerase, that are used.

Codon Optimization

[0127] In some embodiments, at least one or more regions of the polynucleotide molecule may be codon optimized for expression in a eukaryotic cell. In certain embodiments, the polynucleotide molecules that encode one or more components of the systems as described in any of the embodiments herein are optimized for expression in a mammalian cell or a plant cell.

[0128] An example of a codon optimized sequence is in this instance a sequence optimized for expression in a eukaryote, e.g., humans (i.e. being optimized for expression in humans), or for another eukaryote, animal or mammal as herein discussed. It will be appreciated that other examples are possible and codon optimization for a host species other than human, or for codon optimization for specific organs is known. In some embodiments, an enzyme coding sequence encoding a component in the system is codon optimized for expression in particular cells, such as eukaryotic cells. The eukaryotic cells may be those of or derived from a particular organism, such as a plant or a mammal, including but not limited to human, or non-human eukaryote or animal or mammal as herein discussed, e.g., mouse, rat, rabbit, dog, livestock, or non-human mammal or primate. In some embodiments, processes for modifying the germ line genetic identity of human beings and/or processes for modifying the genetic identity of animals which are likely to cause them suffering without any substantial medical benefit to man or animal, and also animals resulting from such processes, may be excluded. In general, codon optimization refers to a process of modifying a nucleic acid sequence for enhanced expression in the host cells of interest by replacing at least one codon (e.g., about or more than about 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more codons) of the native sequence with codons that are more frequently or most frequently used in the genes of that host cell while maintaining the native amino acid sequence.

[0129] Various species exhibit particular bias for certain codons of a particular amino acid. Codon bias (differences in codon usage between organisms) often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, among other things, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, genes can be tailored for optimal gene expression in a given organism based on codon optimization. Codon usage tables are readily available, for example, at the "Codon Usage Database" available at www.kazusa.orjp/codon/ and these tables can be adapted in a number of ways. See Nakamura, Y., et al. "Codon usage tabulated from the international DNA sequence databases: status for the year 2000" Nucl. Acids Res. 28:292 (2000). Computer algorithms for codon optimizing a particular sequence for expression in a particular host cell are also available, such as Gene Forge (Aptagen; Jacobus, Pa.), are also available. In some embodiments, one or more codons (e.g., 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more, or all codons) in a sequence encoding a component in the system corresponds to the most frequently used codon for a particular amino acid.

Nuclear Localization Signals

[0130] In some embodiments, the systems and compositions herein further comprises one or more nuclear localization signals (NLSs) capable of driving the accumulation of the components, to a desired amount in the nucleus of a cell.

[0131] In certain embodiments, at least one nuclear localization signal (NLS) is attached to the nucleic acid sequences encoding the components in the systems. In some embodiments, one or more C-terminal or N-terminal NLSs are attached (and hence nucleic acid molecule(s) coding for the components in the systems can include coding for NLS(s) so that the expressed product has the NLS(s) attached or connected). In a preferred embodiment a C-terminal NLS is attached for optimal expression and nuclear targeting in eukaryotic cells, e.g., human cells.

[0132] Non-limiting examples of NLSs include an NLS sequence derived from: the NLS of the SV40 virus large T-antigen; the NLS from nucleoplasmin (e.g., the nucleoplasmin bipartite NLS; the c-myc NLS; the hRNPA1 M9 NLS; the sequence of the IBB domain from importin-alpha; the NLSs of the myoma T protein; the NLS of human p53; the NLS of mouse c-abl IV; the NLSs of the influenza virus NS1; the NLS of the Hepatitis virus delta antigen; the NLS of the mouse Mx1 protein; the NLS of the human poly(ADP-ribose) polymerase; and the NLS of the steroid hormone receptors (human) glucocorticoid. Examples of such NLSs include those described in paragraph [00131] in Zhang et al. WO2014093595A1.

[0133] In some embodiments, a NLS is a heterologous NLS. For example, the NLS is not naturally present in the molecule it attached to.

[0134] In general, strength of nuclear localization activity may derive from the number of NLSs in the nucleic acid-targeting effector protein, the particular NLS(s) used, or a combination of these factors. Detection of accumulation in the nucleus may be performed by any suitable technique. For example, a detectable marker may be fused to the nucleic acid-targeting protein, such that location within a cell may be visualized, such as in combination with a means for detecting the location of the nucleus (e.g., a stain specific for the nucleus such as DAPI).

[0135] In some embodiments, a vector described herein (e.g., those comprising polynucleotides encoding the components in the systems comprise one or more nuclear localization sequences (NLSs), such as about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs. More particularly, vector comprises one or more NLSs not naturally present in the the components in the systems. Most particularly, the NLS may be present in the vector 5' and/or 3' of the the components in the systems. In some embodiments, the the components in the systems comprises about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs at or near the amino-terminus, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs at or near the carboxy-terminus, or a combination of these (e.g., zero or at least one or more NLS at the amino-terminus and zero or at one or more NLS at the carboxy terminus). When more than one NLS is present, each may be selected independently of the others, such that a single NLS may be present in more than one copy and/or in combination with one or more other NLSs present in one or more copies. In some embodiments, an NLS is considered near the N- or C-terminus when the nearest amino acid of the NLS is within about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, or more amino acids along the polypeptide chain from the N- or C-terminus.

[0136] In certain embodiments, other localization tags may be fused to the Cas and/or transposase(s), such as without limitation for localizing to particular sites in a cell, such as organelles, such mitochondria, plastids, chloroplast, vesicles, golgi, (nuclear or cellular) membranes, ribosomes, nucleoluse, ER, cytoskeleton, vacuoles, centrosome, nucleosome, granules, centrioles, etc.

Fusion Proteins and Linkers

[0137] The components, e.g., proteins, domains, and nucleic acids, in the systems (from the same or different systems) may be associated (e.g., fused). The fusion may be via a linker. The term "linker" as used in reference to a fusion protein refers to a molecule which joins the proteins to form a fusion protein. Generally, such molecules have no specific biological activity other than to join or to preserve some minimum distance or other spatial relationship between the proteins. However, in certain embodiments, the linker may be selected to influence some property of the linker and/or the fusion protein such as the folding, net charge, or hydrophobicity of the linker. In some embodiments, components in different systems may be associated (e.g., fused). In some embodiments, the two or more different systems herein may be associated (e.g., fused). For example, two or more of the ATPase(s), deaminase(s), and reverse transcriptase(s) may be associated (e.g., fused) together.

[0138] Suitable linkers for use in the methods of the present invention are well known to those of skill in the art and include, but are not limited to, straight or branched-chain carbon linkers, heterocyclic carbon linkers, or peptide linkers. However, as used herein the linker may also be a covalent bond (carbon-carbon bond or carbon-heteroatom bond). In particular embodiments, the linker is used to separate the Cas protein and the ligase by a distance sufficient to ensure that each protein retains its required functional property. Preferred peptide linker sequences adopt a flexible extended conformation and do not exhibit a propensity for developing an ordered secondary structure. In certain embodiments, the linker can be a chemical moiety which can be monomeric, dimeric, multimeric or polymeric. Preferably, the linker comprises amino acids. Typical amino acids in flexible linkers include Gly, Asn and Ser. Accordingly, in particular embodiments, the linker comprises a combination of one or more of Gly, Asn and Ser amino acids. Other near neutral amino acids, such as Thr and Ala, also may be used in the linker sequence. Exemplary linkers are disclosed in Maratea et al. (1985), Gene 40: 39-46; Murphy et al. (1986) Proc. Nat'l. Acad. Sci. USA 83: 8258-62; U.S. Pat. Nos. 4,935,233; and 4,751,180. For example, GlySer linkers GGS, GGGS (SEQ ID NO: 76) or GSG can be used. GGS, GSG, GGGS (SEQ ID NO: 76) or GGGGS (SEQ ID NO: 77) linkers can be used in repeats of 3 (such as (GGS)$_3$ (SEQ ID NO: 78), (GGGGS)$_3$ (SEQ ID NO: 79)) or 5, 6, 7, 9 or even 12 or more, to provide suitable lengths. In some cases, the linker may be (GGGGS)$_{3-15}$, For example, in some cases, the linker may be (GGGGS)$_{3-11}$, e.g., GGGGS (SEQ ID NO: 77), (GGGGS)$_2$ (SEQ ID NO: 80), (GGGGS)$_3$ (SEQ ID NO: 79), (GGGGS)$_4$ (SEQ ID NO: 81), (GGGGS)$_5$ (SEQ ID NO: 82), (GGGGS)$_6$ (SEQ ID NO: 83), (GGGGS)$_7$ (SEQ ID NO: 84), (GGGGS)$_8$ (SEQ ID NO: 85), (GGGGS)$_9$ (SEQ ID NO: 86), (GGGGS)$_{10}$ (SEQ ID NO: 87), or (GGGGS)$_{11}$ (SEQ ID NO: 88).

[0139] In particular embodiments, linkers such as (GGGGS)$_3$ (SEQ ID NO: 79) are preferably used herein. (GGGGS)$_6$ (SEQ ID NO: 83), (GGGGS)$_9$ (SEQ ID NO: 86) or (GGGGS)$_{12}$ (SEQ ID NO: 89) may preferably be used as alternatives. Other preferred alternatives are (GGGGS)$_1$ (SEQ ID NO: 77), (GGGGS)$_2$ (SEQ ID NO: 80), (GGGGS)$_4$ (SEQ ID NO: 81), (GGGGS)$_5$ (SEQ ID NO: 82), (GGGGS)$_7$ (SEQ ID NO: 84), (GGGGS)$_8$ (SEQ ID NO: 85), (GGGGS)$_{10}$ (SEQ ID NO: 87), or (GGGGS)$_{11}$ (SEQ ID NO: 88). In yet a further embodiment, LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR (SEQ ID NO: 90) is used as a linker. In yet an additional embodiment, the linker is an XTEN linker. In particular embodiments, the CRISPR-cas protein is a Cas protein and is linked to the ligase or its catalytic domain by means of an LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR (SEQ ID NO: 90) linker. In further particular embodiments, the Cas protein is linked C-terminally to the N-terminus of a ligase or its catalytic domain by means of an LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR (SEQ ID NO: 90) linker. In addition, N- and C-terminal NLSs can also function as linker (e.g., PKKKRKVEASSPKKRKVEAS (SEQ ID NO: 91)).

[0140] Examples of linkers are shown in the Table 4 below.

TABLE 4

GGS	GGTGGTAGT (SEQ ID NO: 92)
GGSx3 (9)	GGTGGTAGTGGAGGGAGCGGCGGTTCA (SEQ ID NO: 93)
GGSx7 (21)	ggtggaggaggctctggtggaggcggtagcggaggcggagggtcgGGTGGTAGTGGAGGGAGCGGCGGTTCA (SEQ ID NO: 94)
XTEN	TCGGGATCTGAGACGCCTGGGACCTCGGAATCGGCTACGCCCGAAAGT (SEQ ID NO: 95)
Z-EGFR_ Short	Gtggataacaaatttaacaaagaaatgtgggcggcgtgggaagaaattcgtaacctgccgaacctgaacggctggcagatgaccgcgtttattgcgagcctggtggatgatccgagccagagcgcgaacctgctggcggaagcgaaaaaactgaacgatgcgcaggcgccgaaaaccggcggtggttctggt (SEQ ID NO: 96)
GSAT	Ggtggttctgccggtggctccggttctggctccagcggtggcagctctggtgcgtccggcacgggtactgcgggtggcactggcagcggttccggtactggctctggc (SEQ ID NO: 97)

Adaptor Proteins

[0141] The adaptor proteins may include orthogonal RNA-binding protein/aptamer combinations that exist within the diversity of bacteriophage coat proteins. A list of such coat proteins includes, but is not limited to: Qβ, F2, GA, fr, JP501, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ϕCb5, ϕCb8r, ϕCb12r, ϕCb23r, 7s and PRR1.

Heterologous Components

[0142] In some embodiments, when a system or composition herein comprises multiple components, the components may be heterologous, i.e., they do not naturally occur together in the same cell or an organism. In some examples, the system comprises an ATPase and an adenosine deaminase that are heterologous. In certain examples, the system comprises two or more heterologous reverse transcriptases.

Cas Proteins and Variants

[0143] In some embodiments, the systems may further comprise a Cas protein or a variant thereof, and one or more guide molecules. One or more components described herein in the systems may be associated (e.g., fused) with a Cas protein or a variant thereof (a catalytically inactive). The Cas protein and guide molecule(s) may guide the components such as ATPase, deaminase, reverse transcriptase etc. to target a desired target sequence.

[0144] The Cas proteins, variants thereof, and guide molecules may be those in a CRISPR-Cas or CRISPR system, refers collectively to transcripts and other elements involved in the expression of or directing the activity of CRISPR-associated ("Cas") genes, including sequences encoding a Cas gene, a tracr (trans-activating CRISPR) sequence (e.g. tracrRNA or an active partial tracrRNA), a tracr-mate sequence (encompassing a "direct repeat" and a tracrRNA-processed partial direct repeat in the context of an endogenous CRISPR system), a guide sequence (also referred to as a "spacer" in the context of an endogenous CRISPR system), or "RNA(s)" as that term is herein used (e.g., RNA(s) to guide Cas, such as Cas9, e.g. CRISPR RNA and transactivating (tracr) RNA or a single guide RNA (sgRNA) (chimeric RNA)) or other sequences and transcripts from a CRISPR locus. In general, a CRISPR system is characterized by elements that promote the formation of a CRISPR complex at the site of a target sequence (also referred to as a protospacer in the context of an endogenous CRISPR system). See, e.g, Shmakov et al. (2015) "Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems", Molecular Cell, DOI: dx.doi.org/10.1016/j.molcel.2015.10.008.

Class 1 Systems

[0145] The Cas proteins may be Cas proteins in class 1 CRISPR systems. In certain example embodiments, the Class 1 system may be Type I, Type III or Type IV Cas proteins as described in Makarova et al. "Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants" Nature Reviews Microbiology, 18:67-81 (February 2020), incorporated in its entirety herein by reference, and particularly as described in FIG. **1**, p. 326. The Class 1 systems typically use a multi-protein effector complex, which can, in some embodiments, include ancillary proteins, such as one or more proteins in a complex referred to as a CRISPR-associated complex for antiviral defense (Cascade), one or more adaptation proteins (e.g. Cas1, Cas2, RNA nuclease), and/or one or more accessory proteins (e.g. Cas 4, DNA nuclease), CRISPR associated Rossman fold (CARF) domain containing proteins, and/or RNA transcriptase. Although Class 1 systems have limited sequence similarity, Class 1 system proteins can be identified by their similar architectures, including one or more Repeat Associated Mysterious Protein (RAMP) family subunits, e.g. Cas 5, Cas6, Cas7. RAMP proteins are characterized by having one or more RNA recognition motif domains. Large subunits (for example cas8 or cas10) and small subunits (for example, cas11) are also typical of Class 1 systems. See, e.g., FIGS. **1** and **2**. Koonin E V, Makarova K S. 2019 Origins and evolution of CRISPR-Cas systems. Phil. Trans. R. Soc. B 374: 20180087, DOI: 10.1098/rstb.2018.0087. In one aspect, Class 1 systems are characterized by the signature protein Cas3. The Cascade in particular Class1 proteins can comprise a dedicated complex of multiple Cas proteins that binds pre-crRNA and recruits an additional Cas protein, for example Cas6 or Cas5, which is the nuclease directly responsible for processing pre-crRNA. In one aspect, the Type I CRISPR protein comprises an effector complex comprises one or more Cas5 subunits and two or more Cas^7 subunits. Class 1 subtypes include Type I-A, I-B, I-C, I-U, I-D, I-E, and I-F, Type IV-A and IV-B, and Type III-A, III-D, III-C, and III-B. Class 1 systems also include CRISPR-Cas variants, including Type I-A, I-B, I-E, I-F and I-U variants, which can include variants carried by transposons and plasmids, including versions of subtype I-F encoded by a large family of Tn7-like transposon and smaller groups of Tn7-like transposons that encode similarly degraded subtype I-B systems. Peters et al., PNAS 114 (35) (2017); DOI: 10.1073/pnas.1709035114; see also, Makarova et al, the CRISPR Journal, v. 1, n5, FIG. **5**.

Class 2 Systems

[0146] The Cas proteins may be Cas proteins in class 2 CRISPR-Cas systems. Class 2 systems are distinguished from Class 1 systems in that they have a single, large, multi-domain effector protein. In certain example embodiments, the Class 2 system can be a Type II, Type V, or Type VI system, which are described in Makarova et al. "Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants" Nature Reviews Microbiology, 18:67-81 (February 2020), incorporated herein by reference. Each type of Class 2 system is further divided into subtypes. See Markova et al. 2020, particularly at Figure. 2. Class 2, Type II systems can be divided into 4 subtypes: II-A, II-B, II-C1, and II-C2. Class 2, Type V systems can be divided into 17 subtypes: V-A, V-B1, V-B2, V-C, V-D, V-E, V-F1, V-F1(V-U3), V-F2, V-F3, V-G, V-H, V-I, V-K (V-U5), V-U1, V-U2, and V-U4. Class 2, Type IV systems can be divided into 5 subtypes: VI-A, VI-B1, VI-B2, VI-C, and VI-D.

[0147] The distinguishing feature of these types is that their effector complexes consist of a single, large, multi-domain protein. Type V systems differ from Type II effectors (e.g., Cas9), which contain two nuclear domains that are each responsible for the cleavage of one strand of the target DNA, with the HNH nuclease inserted inside the Ruv-C like nuclease domain sequence. The Type V systems (e.g., Cas12) only contain a RuvC-like nuclease domain that cleaves both strands. Type VI (Cas13) are unrelated to the effectors of Type II and V systems and contain two HEPN domains and target RNA. Cas13 proteins also display collateral activity that is triggered by target recognition. Some Type V systems have also been found to possess this collateral activity with two single-stranded DNA in in vitro contexts.

[0148] In some embodiments, the Class 2 system is a Type II system. In some embodiments, the Type II CRISPR-Cas system is a II-A CRISPR-Cas system. In some embodiments, the Type II CRISPR-Cas system is a II-B CRISPR-Cas system. In some embodiments, the Type II CRISPR-Cas system is a II-C1 CRISPR-Cas system. In some embodiments, the Type II CRISPR-Cas system is a II-C2 CRISPR-Cas system. In some embodiments, the Type II system is a Cas9 system. In some embodiments, the Type II system includes a Cas9.

[0149] In some embodiments, the Class 2 system is a Type V system. In some embodiments, the Type V CRISPR-Cas system is a V-A CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-B1 CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-B2 CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-C CRISPR-

Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-D CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-E CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-F1 CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-F1 (V-U3) CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-F2 CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-F3 CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-G CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-H CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-I CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-K (V-U5) CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-U1 CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-U2 CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system is a V-U4 CRISPR-Cas system. In some embodiments, the Type V CRISPR-Cas system includes a Cas12a (Cpf1), Cas12b (C2c1), Cas12c (C2c3), Cas12d (CasY), Cas12e (CasX), and/or Cas14.

[0150] In some embodiments the Class 2 system is a Type VI system. In some embodiments, the Type VI CRISPR-Cas system is a VI-A CRISPR-Cas system. In some embodiments, the Type VI CRISPR-Cas system is a VI-B1 CRISPR-Cas system. In some embodiments, the Type VI CRISPR-Cas system is a VI-B2 CRISPR-Cas system. In some embodiments, the Type VI CRISPR-Cas system is a VI-C CRISPR-Cas system. In some embodiments, the Type VI CRISPR-Cas system is a VI-D CRISPR-Cas system. In some embodiments, the Type VI CRISPR-Cas system includes a Cas13a (C2c2), Cas13b (Group 29/30), Cas13c, and/or Cas13d.

Specialized Cas-Based Systems

[0151] In some embodiments, the system is a Cas-based system that is capable of performing a specialized function or activity. For example, the Cas protein may be fused, operably coupled to, or otherwise associated with one or more functionals domains. In certain example embodiments, the Cas protein may be a catalytically dead Cas protein ("dCas") and/or have nickase activity. A nickase is a Cas protein that cuts only one strand of a double stranded target. In such embodiments, the dCas or nickase provide a sequence specific targeting functionality that delivers the functional domain to or proximate a target sequence. Example functional domains that may be fused to, operably coupled to, or otherwise associated with a Cas protein can be or include, but are not limited to a nuclear localization signal (NLS) domain, a nuclear export signal (NES) domain, a translational activation domain, a transcriptional activation domain (e.g. VP64, p65, MyoD1, HSF1, RTA, and SETT/9), a translation initiation domain, a transcriptional repression domain (e.g., a KRAB domain, NuE domain, NcoR domain, and a SID domain such as a SID4X domain), a nuclease domain (e.g., FokI), a histone modification domain (e.g., a histone acetyltransferase), a light inducible/controllable domain, a chemically inducible/controllable domain, a transposase domain, a homologous recombination machinery domain, a recombinase domain, an integrase domain, and combinations thereof. Methods for generating catalytically dead Cas9 or a nickase Cas9 (WO 2014/204725, Ran et al. Cell. 2013 Sep. 12; 154(6):1380-1389), Cas12 (Liu et al. Nature Communications, 8, 2095 (2017), and Cas13 (International Patent Publication Nos. WO 2019/005884 and WO2019/060746) are known in the art and incorporated herein by reference.

[0152] In some embodiments, the functional domains can have one or more of the following activities: methylase activity, demethylase activity, translation activation activity, translation initiation activity, translation repression activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, nuclease activity, single-strand RNA cleavage activity, double-strand RNA cleavage activity, single-strand DNA cleavage activity, double-strand DNA cleavage activity, molecular switch activity, chemical inducibility, light inducibility, and nucleic acid binding activity. In some embodiments, the one or more functional domains may comprise epitope tags or reporters. Non-limiting examples of epitope tags include histidine (His) tags, V5 tags, FLAG tags, influenza hemagglutinin (HA) tags, Myc tags, VSV-G tags, and thioredoxin (Trx) tags. Examples of reporters include, but are not limited to, glutathione-S-transferase (GST), horseradish peroxidase (HRP), chloramphenicol acetyltransferase (CAT) beta-galactosidase, beta-glucuronidase, luciferase, green fluorescent protein (GFP), HcRed, DsRed, cyan fluorescent protein (CFP), yellow fluorescent protein (YFP), and auto-fluorescent proteins including blue fluorescent protein (BFP).

[0153] The one or more functional domain(s) may be positioned at, near, and/or in proximity to a terminus of the effector protein (e.g., a Cas protein). In embodiments having two or more functional domains, each of the two can be positioned at or near or in proximity to a terminus of the effector protein (e.g., a Cas protein). In some embodiments, such as those where the functional domain is operably coupled to the effector protein, the one or more functional domains can be tethered or linked via a suitable linker (including, but not limited to, GlySer linkers) to the effector protein (e.g., a Cas protein). When there is more than one functional domain, the functional domains can be same or different. In some embodiments, all the functional domains are the same. In some embodiments, all of the functional domains are different from each other. In some embodiments, at least two of the functional domains are different from each other. In some embodiments, at least two of the functional domains are the same as each other.

[0154] Other suitable functional domains can be found, for example, in International Patent Publication No. WO 2019/018423.

Split CRISPR-Cas Systems

[0155] In some embodiments, the CRISPR-Cas system is a split CRISPR-Cas system. See e.g., Zetche et al., 2015. Nat. Biotechnol. 33(2): 139-142 and International Patent Publication WO 2019/018423, the compositions and techniques of which can be used in and/or adapted for use with the present invention. Split CRISPR-Cas proteins are set forth herein and in documents incorporated herein by reference in further detail herein. In certain embodiments, each part of a split CRISPR protein are attached to a member of a specific binding pair, and when bound with each other, the members of the specific binding pair maintain the parts of the CRISPR protein in proximity. In certain embodiments, each part of a split CRISPR protein is associated with an inducible binding pair. An inducible binding pair is one which is capable of being switched "on" or "off" by a protein or small molecule that binds to both members of the inducible binding pair. In some embodiments, CRISPR proteins may preferably split between domains, leaving domains intact. In particular embodiments, said Cas split domains (e.g., RuvC and HNH domains in the case of Cas9) can be simultaneously or sequentially introduced into the cell such that said split Cas domain(s) process the target nucleic acid sequence in the algae cell. The reduced size of the split Cas compared to the wild type Cas allows other methods of delivery of the systems to the cells, such as the use of cell penetrating peptides as described herein.

Guide Molecules

[0156] The guide molecules (i.e., a molecule comprising a guide sequence) refer to polynucleotides capable of guiding Cas to a target genomic locus and are used interchangeably as in foregoing cited documents such as International Patent Publication No. WO 2014/093622 (PCT/US2013/074667). In general, a guide molecule may be any polynucleotide sequence having sufficient complementarity with a target polynucleotide sequence to hybridize with the target sequence and direct sequence-specific binding of a CRISPR complex to the target sequence. The guide molecule can be a polynucleotide.

[0157] The ability of a guide sequence (within a nucleic acid-targeting guide RNA) to direct sequence-specific binding of a nucleic acid-targeting complex to a target nucleic acid sequence may be assessed by any suitable assay. For example, the components of a nucleic acid-targeting CRISPR system sufficient to form a nucleic acid-targeting complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target nucleic acid sequence, such as by transfection with vectors encoding the components of the nucleic acid-targeting complex, followed by an assessment of preferential targeting (e.g., cleavage) within the target nucleic acid sequence, such as by Surveyor assay (Qui et al. 2004. BioTechniques. 36(4)702-707). Similarly, cleavage of a target nucleic acid sequence may be evaluated in a test tube by providing the target nucleic acid sequence, components of a nucleic acid-targeting complex, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible and will occur to those skilled in the art.

[0158] In some embodiments, the guide molecule is an RNA. The guide molecule(s) (also referred to interchangeably herein as guide polynucleotide and guide sequence) that are included in the CRISPR-Cas or Cas based system can be any polynucleotide sequence having sufficient complementarity with a target nucleic acid sequence to hybridize with the target nucleic acid sequence and direct sequence-specific binding of a nucleic acid-targeting complex to the target nucleic acid sequence. In some embodiments, the degree of complementarity, when optimally aligned using a suitable alignment algorithm, can be about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting examples of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g., the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; available at www.novocraft.com), ELAND (Illumina, San Diego, Calif.), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net).

[0159] A guide sequence, and hence a nucleic acid-targeting guide, may be selected to target any target nucleic acid sequence. The target sequence may be DNA. The target sequence may be any RNA sequence. In some embodiments, the target sequence may be a sequence within an RNA molecule selected from the group consisting of messenger RNA (mRNA), pre-mRNA, ribosomal RNA (rRNA), transfer RNA (tRNA), micro-RNA (miRNA), small interfering RNA (siRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), double stranded RNA (dsRNA), non-coding RNA (ncRNA), long non-coding RNA (lncRNA), and small cytoplasmatic RNA (scRNA). In some preferred embodiments, the target sequence may be a sequence within an RNA molecule selected from the group consisting of mRNA, pre-mRNA, and rRNA. In some preferred embodiments, the target sequence may be a sequence within an RNA molecule selected from the group consisting of ncRNA, and lncRNA. In some more preferred embodiments, the target sequence may be a sequence within an mRNA molecule or a pre-mRNA molecule.

[0160] In some embodiments, a nucleic acid-targeting guide is selected to reduce the degree secondary structure within the nucleic acid-targeting guide. In some embodiments, about or less than about 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1%, or fewer of the nucleotides of the nucleic acid-targeting guide participate in self-complementary base pairing when optimally folded. Optimal folding may be determined by any suitable polynucleotide folding algorithm. Some programs are based on calculating the minimal Gibbs free energy. An example of one such algorithm is mFold, as described by Zuker and Stiegler (Nucleic Acids Res. 9 (1981), 133-148). Another example folding algorithm is the online webserver RNAfold, developed at Institute for Theoretical Chemistry at the University of Vienna, using the centroid structure prediction algorithm (see e.g., A. R. Gruber et al., 2008, Cell 106(1): 23-24; and PA Carr and GM Church, 2009, Nature Biotechnology 27(12): 1151-62).

[0161] In certain embodiments, a guide RNA or crRNA may comprise, consist essentially of, or consist of a direct repeat (DR) sequence and a guide sequence or spacer sequence. In certain embodiments, the guide RNA or crRNA may comprise, consist essentially of, or consist of a direct repeat sequence fused or linked to a guide sequence or spacer sequence. In certain embodiments, the direct repeat sequence may be located upstream (i.e., 5') from the guide sequence or spacer sequence. In other embodiments, the direct repeat sequence may be located downstream (i.e., 3') from the guide sequence or spacer sequence.

[0162] In certain embodiments, the crRNA comprises a stem loop, e.g., a single stem loop. In certain embodiments, the direct repeat sequence forms a stem loop, e.g., a single stem loop.

[0163] In certain embodiments, the spacer length of the guide RNA is from 15 to 35 nt. In certain embodiments, the spacer length of the guide RNA is at least 15 nucleotides. In certain embodiments, the spacer length is from 15 to 17 nt, e.g., 15, 16, or 17 nt, from 17 to 20 nt, e.g., 17, 18, 19, or

20 nt, from 20 to 24 nt, e.g., 20, 21, 22, 23, or 24 nt, from 23 to 25 nt, e.g., 23, 24, or 25 nt, from 24 to 27 nt, e.g., 24, 25, 26, or 27 nt, from 27 to 30 nt, e.g., 27, 28, 29, or 30 nt, from 30 to 35 nt, e.g., 30, 31, 32, 33, 34, or 35 nt, or 35 nt or longer.

[0164] The "tracrRNA" sequence or analogous terms includes any polynucleotide sequence that has sufficient complementarity with a crRNA sequence to hybridize. In some embodiments, the degree of complementarity between the tracrRNA sequence and crRNA sequence along the length of the shorter of the two when optimally aligned is about or more than about 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, or higher. In some embodiments, the tracr sequence is about or more than about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, or more nucleotides in length. In some embodiments, the tracr sequence and crRNA sequence are contained within a single transcript, such that hybridization between the two produces a transcript having a secondary structure, such as a hairpin.

[0165] In general, degree of complementarity is with reference to the optimal alignment of the sca sequence and tracr sequence, along the length of the shorter of the two sequences. Optimal alignment may be determined by any suitable alignment algorithm and may further account for secondary structures, such as self-complementarity within either the sca sequence or tracr sequence. In some embodiments, the degree of complementarity between the tracr sequence and sca sequence along the length of the shorter of the two when optimally aligned is about or more than about 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, or higher.

[0166] In some embodiments, the degree of complementarity between a guide sequence and its corresponding target sequence can be about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or 100%; a guide or RNA or sgRNA can be about or more than about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, or more nucleotides in length; or guide or RNA or sgRNA can be less than about 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, or fewer nucleotides in length; and tracr RNA can be 30 or 50 nucleotides in length. In some embodiments, the degree of complementarity between a guide sequence and its corresponding target sequence is greater than 94.5% or 95% or 95.5% or 96% or 96.5% or 97% or 97.5% or 98% or 98.5% or 99% or 99.5% or 99.9%, or 100%. Off target is less than 100% or 99.9% or 99.5% or 99% or 99% or 98.5% or 98% or 97.5% or 97% or 96.5% or 96% or 95.5% or 95% or 94.5% or 94% or 93% or 92% or 91% or 90% or 89% or 88% or 87% or 86% or 85% or 84% or 83% or 82% or 81% or 80% complementarity between the sequence and the guide, with it being advantageous that off target is 100% or 99.9% or 99.5% or 99% or 99% or 98.5% or 98% or 97.5% or 97% or 96.5% or 96% or 95.5% or 95% or 94.5% complementarity between the sequence and the guide.

[0167] In some embodiments according to the invention, the guide RNA (capable of guiding Cas to a target locus) may comprise (1) a guide sequence capable of hybridizing to a genomic target locus in the eukaryotic cell; (2) a tracr sequence; and (3) a tracr mate sequence. All (1) to (3) may reside in a single RNA, i.e., an sgRNA (arranged in a 5' to 3' orientation), or the tracr RNA may be a different RNA than the RNA containing the guide and tracr sequence. The tracr hybridizes to the tracr mate sequence and directs the CRISPR/Cas complex to the target sequence. Where the tracr RNA is on a different RNA than the RNA containing the guide and tracr sequence, the length of each RNA may be optimized to be shortened from their respective native lengths, and each may be independently chemically modified to protect from degradation by cellular RNase or otherwise increase stability.

[0168] Many modifications to guide sequences are known in the art and are further contemplated within the context of this invention. Various modifications may be used to increase the specificity of binding to the target sequence and/or increase the activity of the Cas protein and/or reduce off-target effects. Example guide sequence modifications are described in International Patent Application No. PCT US2019/045582, specifically paragraphs [0178]-[0333]. which is incorporated herein by reference.

Methods of Identifying Defense Systems

[0169] The present disclosure further provides methods of identifying defense systems. In some embodiments, the methods are based on the facts that genes of defense systems often form clusters in the genome. Thus, candidate defense system genes may be those co-locate with known defense system genes in the genomes of multiple cells of a species or strain. Accordingly, novel defense system be identified by recording or identifying candidate genes located close to known defense systems and identifying homologs of the candidate genes in multiple genomes of the species or cells. The candidate genes that have a significant number of homologs close to known defense system genes may be selected as a putative novel defense system genes. The selected putative defense system genes may be further validated by experiments, e.g., by testing their effects on phage resistance.

[0170] In some examples, the methods of identifying a defense system in a microorganism may comprise identifying genes of known defense systems in a plurality of genomes of the microorganism; recording candidate genes located within 50 kb from the identified genes of known defense systems on the genomes; identifying homologs of each candidate gene on the genomes; and selecting candidate genes wherein at least 10% of homologs of the candidate genes are within 5000 nucleotides and/or 5 genes from one or more known defense systems on the genomes. FIGS. **4** and **8** show flow charts of exemplary methods of identifying novel defense systems.

[0171] In some cases, the recorded candidate genes may be located less than 50 kb, less than 40 kb, less than 30 kb, less than 20 kb, less than 10 kb, less than 8 kb, less than 6 kb, less than 4 kb, less than 2 kb, less than 1000 bp, less than 800 bp, less than 600 bp, less than 400 bp, or less than 200 bp from the identified genes of known defense systems on the genomes. In some cases, the recorded candidate genes may be located less than 20, less than 18, less than 16, less than 14, less than 12, less than 10, less than 8, less than 6, less than 4, or less than 2 open reading frames from the identified genes of known defense systems on the genomes.

[0172] The methods of identifying defense systems may comprise obtaining sequence data of multiple genomes. The multiple genomes may be those from different microorganism cells of the same species or strain. The sequence data used may be from at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least

20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 200, at least 400, at least 600, at least 800, at least 1000, at least 2000, at least 4000, at least 8000, at least 10,000, at least 20,000, at least 40,000, at least 60,000, at least 80,000, at least 100,000, at least 120,000, at least 140,000, at least 160,000, at least 180,000, or at least 200,000 genomes.

[0173] The methods of identifying defense systems may comprise identifying known defense system genes in multiple genomes. The known defense systems or their genes may be identified using sequence alignments and comparing with known sequences, motifs or domains in a protein or nucleic acid domain database. The domains within the gene members of each system may be analyzed bioinformatically using the tools HHpred (Soding J, Biegert A, Lupas A N. (2005) The HHpred interactive server for protein homology detection and structure prediction, nucleic Acids Res. 33: W244-W248; Alva V, Nam S-Z, Soding J, Lupas A N, I. S, S. C, et al. (2016) The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis, nucleic Acids Res. Oxford University Press; 44: W410-W415), Phyre2 (Kelley L A, Mezulis S, Yates C M, Wass M N, Sternberg M J E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. Nature Research; 10: 845-858), PSI-BLAST (Altschul S F, Madden T L, Schaffer A A, Zhang J, Zhang Z, Miller W, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, nucleic Acids Res. 25: 3389-402).

[0174] In some examples, the database may be PFAM. The term "pfam" may encompass a large collection of protein domains and protein families maintained by the pfam consortium and available at several sponsored world wide web sites, including for example: pfam.sanger.ac.uk/(Welcome Trust, Sanger Institute); pfam.sbc.su.se/ (Stockholm Bioinformatics Center); pfam(dot)janelia(dot)org/(Janelia Farm, Howard Hughes Medical Institute); pfam(dot)jouy(dot)inra (dot)fr/(Institut national de la Recherche Agronomique); and pfam.ccbb.re.kr/. pfam domains and families are identified using multiple sequence alignments and hidden Markov models (HMMs) (see e.g. R. D. Finnet et al. nucleic Acids Research Database (2010) Issue 38: D211-222). By accessing the pfam database, for example, using any of the above-reference websites, protein sequences can be queried against the hidden Markov models (HMMMs) using HMMER homology search software (e.g., HMMER3, hmmer(dot)j anelia(dot)org/).

[0175] In some examples, the database may be NCBI's Conserved Domain Database (CDD) (Marchler-Bauer A, Lu S, Anderson J B, Chitsaz F, Derbyshire M K, DeWeese-Scott C, et al. (2011) CDD: a Conserved Domain Database for the functional annotation of proteins, nucleic Acids Res. 39: D225-D229).

[0176] In some examples, the database may be COG. The term "COG (clusters of orthologous groups)" may encompass a large collection of protein families classified according to their homologous relationships available at e.g. the NCBI COG website (www(dot)ncbi(dot)nlm(dot)nih(dot) gov/COG). Each COG comprises a group of proteins found to be orthologous across at least three lineages and likely corresponds to an ancient conserved domain [see e.g. Tatusov et al. Science 1997 Oct. 24; 278(5338):631-7; and Tatusov et al. nucleic Acids Res. 2000 Jan. 1; 28(1): 33-36].

[0177] The methods may further comprise filter false positives among the identified known defense genes.

[0178] The methods may further comprise, after the false positives of the known defense genes are filtered, identifying known defense systems. A defense system may comprise one or more defense proteins or nucleic acids involved in defense function. Examples of the known defense systems used in the methods include mobilome, a CRISPR system, Type I RM and McrBC system, BREX-associated system, Zorya system, Wadjet system, Druantia-associated system, Hachiman system, Lamassu system, Thoeris-like system, Gabija system, Septu system, pAgo system, Shedu system, Kiwa system, DUF499-DUF1156 system, and Toxin/antitoxin system.

[0179] The methods may further comprise recording (e.g., tabulating) candidate genes, which are genes within certain distance of a known defense system gene. The candidate genes may be on the 5' side or the 3' side of the defense system gene. For examples, the candidate genes may be within 50 kb, 40 kb, 30 kb, 20 kb, 18 kb, 16 kb, 14 kb, 12 kb, 10 kb, 9 kb, 8 kb, 7 kb, 6 kb, 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 900 bp, 800 bp, 700 bp, 600 bp, 500 bp, 400 bp, 300 bp, 200 bp, or 100 bp from the known defense system. In some examples, the candidate genes are within 10 kb of a defense system. In some cases, each of the candidate gene is called a seed.

[0180] The methods may further comprise, for each of the candidate gene, identifying homologs in the genomes. A homolog of the candidate gene may be a gene that share at least 50%, 60%, 70%, 80%, 90%, 95%, 99%, or 100% sequence identity with the candidate gene. In some examples, the homologs share at least 70% of sequence identity with the candidate genes.

[0181] In some cases, the homologs may have an E-value of 10^{-3} or lower, 10^{-4} or lower, 10^{-5} or lower, 10^{-6} or lower, 10^{-7} or lower, or 10^{-8} or lower. The Expect value or E-value refers to a parameter that describes the number of hits one can "expect" to see by chance when searching a database of a particular size. Essentially, the E-value describes the random background noise. For example, an E value of 1 assigned to a hit can be interpreted as meaning that in a database of the current size one might expect to see 1 match with a similar score simply by chance. The lower the E-value, or the closer it is to zero, the more "significant" the match (e.g., homology, identity) is.

[0182] The methods may further comprise selecting putative defense system genes from the candidate genes. The selected putative defense system genes may have at least a portion of the homologs in proximity to the known defense system genes. For example, a selected putative defense system genes may have at least 5%, at least 10%, at least 11%, at least 12%, at least 13%, at least 14%, at least 15%, at least 16%, at least 17%, at least 18%, at least 19%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, or at least 50% of its homologs. In some examples, a selected putative defense system genes may have at least 15% of the its homologs in proximity to the known defense system.

[0183] In some embodiments, the selection of putative defense system genes comprises selecting putative cassettes comprising multiple candidate genes. Each of the candidate genes in the putative cassette may have at least 5%, at least 10%, at least 11%, at least 12%, at least 13%, at least 14%, at least 15%, at least 16%, at least 17%, at least 18%, at least

19%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, or at least 50% of its homologs. In some examples, each of the candidate genes in the putative cassette may have at least 15% of the its homologs in proximity to the known defense system.

[0184] When a candidate gene or its homolog is in proximity to a known defense gene, the candidate gene or its homolog may be within 1000 nt, 900 nt, 800 nt, 700 nt, 600 nt, 500 nt, 400 nt, 300 nt, 200 nt, 100 nt, 80 nt, 60 nt, 40 nt, 20 nt, 10 nt, 5 nt, 4 nt, 3 nt, 2 nt, or 1 nt from the known defense gene.

Validation of Identified Defense Systems

[0185] In some embodiments, the methods further comprise validating the selected putative defense systems and genes. The validation may be performed by introducing the putative defense system in host cells, infected the cells with virus (e.g., phages), and test phage infection efficiencies. Host cells introduced with a functional defense system may significantly suppress the phage infection efficiency. Examples of methods of validation include those described in Doron S. et al., Science. 2018 Mar. 2; 359(6379), Systematic discovery of antiphage defense systems in the microbial pangenome.

Methods of Use

[0186] The defense systems herein may be introduced to host cells to manipulate the cells' function and activity. In some examples, the defense systems may be introduced to bacteria to manipulate their resistance to phage infection. In some embodiments, the defense systems may be introduced to eukaryotic cells to manipulate the function, structure, level, and/or expression of proteins or nucleic acids.

Protection of Bacteria

[0187] In some embodiments, the defense systems may be introduced to bacteria or other host cells to increase the cells' resistance to an infection. In some cases, the defense systems may be used to protect bacterial fermentation from phage infection and contamination, which is a main cause of slow fermentation or complete starter failure. The lack of bacteria which survive adequately can result in milk products which do not have a desirable taste.

[0188] In some embodiments, the defense systems may be introduced to bacteria useful in the manufacture of dairy and fermentation processing such as, but not limited to, milk-derived products, such as cheeses, yogurt, fermented milk products, sour milks, and buttermilk. In some embodiments, the bacteria are useful as a part of the starter culture in the manufacture of dairy and fermentation processing. In some embodiments, the starter culture is a food grade starter culture. Examples of such bacteria include lactic acid bacteria, which encompass Gram positive, microaerophillic or anaerobic bacteria which ferment sugar with the production of acids including lactic acid as the predominantly produced acid, acetic acid, formic acid and propionic acid. Examples of the bacteria include *Lactococcus* species, *Streptococcus* species, *Lactobacillus* species, *Leuconostoc* species, *Oenococcus* species, *Pediococcus* species, *Bifidobacterium* species, and *Propionibacterium* species. In some embodiments, bacteria protected in a method of protecting bacteria from phage infection comprises bacteria selected from a *Lactococcus* species, a *Streptococcus* species, a *Lactobacillus* species, a *Leuconostoc* species, a *Oenococcus* species, a *Pediococcus* species, a *Bifidobacterium*, and a *Propionibacterium* species. In some embodiments, a method of protecting bacteria from phage infection comprises protecting a *Lactococcus* species of bacteria. In some embodiments a method of protecting bacteria from phage infection comprises protecting a *Streptococcus* species of bacteria. In some embodiments a method of protecting bacteria from phage infection comprises protecting a *Lactobacillus* species of bacteria. In some embodiments, a method of protecting bacteria from phage infection comprises protecting a *Leuconostoc* species of bacteria. In some embodiments, a method of protecting bacteria from phage infection comprises protecting a *Oenococcus* species of bacteria. In some embodiments, a method of protecting bacteria from phage infection comprises protecting a *Pediococcus* species of bacteria. In some embodiments, a method of protecting bacteria from phage infection comprises protecting a *Bifidobacterium* of bacteria. In some embodiments, a method of protecting bacteria from phage infection comprises protecting a *Propionibacterium* species of bacteria.

Enhancing Bacteria Susceptibility to Infection

[0189] In some embodiments, the defense systems may be introduced to bacteria or other host cells to decrease the cells' resistance to an infection. In some examples, the defense system may be engineered to reduce or eliminate its defense function. In certain examples, one or more modulating agents that manipulate the function or level of the defense systems may be introduced to the host cells.

[0190] In some examples, the present disclosure provides methods of treating bacterial infection in a subject in need thereof, the method comprising administering to the subject a therapeutically effective amount of the anti-Defense System agent, thereby treating the bacterial infection in the subject. In some embodiments, there is provided the agent, for use in the treatment of bacterial infection in a subject in need thereof. In some examples, the present disclosure provides methods of generating cells as reagents that can be easily infected by phages. Such cells may be used as research tools in biotechnology.

Engineered Cells

[0191] The present disclosure provides engineered cells comprising the systems and/or polynucleotides herein. In some cases, the cells may be where the plasmids and/or vesicles are produced. For example, the cells may be host cells, such as bacterial cells. In some examples, the cells may be eukaryotic cells, in which the systems are used for manipulating the function and other activities of the cells.

[0192] The cell may be a prokaryotic cell. The prokaryotic cell may be a bacterial cell. The prokaryotic cell may be an archaea cell. Examples of bacterial cells include those from the genus *Escherichia, Bacillus, Lactobacillus, Rhodococcus, Rodhobacter, Synechococcus, Synechoystis, Pseudomonas, Psedoaltermonas, Stenotrophamonas*, and *Streptomyces*. Examples of bacterial cells include *Escherichia coli* cells, *Caulobacter crescentus* cells, *Rodhobacter sphaeroides* cells, *Psedoaltermonas haloplanktis* cells. Suitable strains of bacterial include, but are not limited to BL21 (DE3), DL21(DE3)-pLysS, BL21 Star-pLysS, BL21-SI, BL21-AI, Tuner, Tuner pLysS, Origami, Origami B pLysS, Rosetta, Rosetta pLysS, Rosetta-gami-pLysS, BL21 Codon-

Plus, AD494, BL2trxB, HMS174, NovaBlue(DE3), BLR, C41(DE3), C43(DE3), Lemo21(DE3), Shuffle T7, Arctic-Express and ArticExpress (DE3).

[0193] The cell can be a eukaryotic cell. The eukaryotic cells may be those of or derived from a particular organism, such as a plant or a mammal, including human, or non-human eukaryote or animal or mammal as herein discussed, e.g., mouse, rat, rabbit, dog, livestock, or non-human mammal or primate. In some aspects the engineered cell can be a cell line. Examples of cell lines include C8161, CCRF-CEM, MOLT, mIMCD-3, NHDF, HeLa-S3, Huhl, Huh4, Huh7, HUVEC, HASMC, HEKn, HEKa, MiaPaCell, Pancl, PC-3, TF1, CTLL-2, C1R, Rath, CV1, RPTE, A10, T24, J82, A375, ARH-77, Calul, SW480, SW620, SKOV3, SK-UT, CaCo2, P388D1, SEM-K2, WEHI-231, HB56, TIB55, Jurkat, J45.01, LRMB, Bc1-1, BC-3, IC21, DLD2, Raw264.7, NRK, NRK-52E, MRCS, MEF, Hep G2, HeLa B, HeLa T4, COS, COS-1, COS-6, COS-M6A, BS-C-1 monkey kidney epithelial, BALB/3T3 mouse embryo fibroblast, 3T3 Swiss, 3T3-L1, 132-d5 human fetal fibroblasts; 10.1 mouse fibroblasts, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A172, A20, A253, A431, A-549, ALC, B16, B35, BCP-1 cells, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO-IR, CHO-K1, CHO-K2, CHO-T, CHO Dhfr−/−, COR-L23, COR-L23/CPR, COR-L23/5010, COR-L23/R23, COS-7, COV-434, CML T1, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepalc1c7, HL-60, HMEC, HT-29, Jurkat, JY cells, K562 cells, Ku812, KCL22, KG1, KYO1, LNCap, Ma-Mel 1-48, MC-38, MCF-7, MCF-10A, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCK II, MDCK II, MOR/0.2R, MONO-MAC 6, MTD-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, OPCN/OPCT cell lines, Peer, PNT-1A/PNT 2, RenCa, RIN-5F, RMA/RMAS, Saos-2 cells, Sf-9, SkBr3, T2, T-47D, T84, THP1 cell line, U373, U87, U937, VCaP, Vero cells, WM39, WT-49, X63, YAC-1, YAR, and transgenic varieties thereof. Cell lines are available from a variety of sources known to those with skill in the art (see, e.g., the American Type Culture Collection (ATCC) (Manassas, Va.)).

[0194] Further, the cell may be a fungus cell. As used herein, a "fungal cell" refers to any type of eukaryotic cell within the kingdom of fungi. Phyla within the kingdom of fungi include *Ascomycota, Basidiomycota, Blastocladiomycota, Chytridiomycota, Glomeromycota, Microsporidia*, and *Neocallimastigomycota*. Fungal cells may include yeasts, molds, and filamentous fungi. In some embodiments, the fungal cell is a yeast cell.

[0195] As used herein, the term "yeast cell" refers to any fungal cell within the phyla Ascomycota and Basidiomycota. Yeast cells may include budding yeast cells, fission yeast cells, and mold cells. Without being limited to these organisms, many types of yeast used in laboratory and industrial settings are part of the phylum Ascomycota. In some embodiments, the yeast cell is an *S. cerervisiae, Kluyveromyces marxianus*, or *Issatchenkia orientalis* cell. Other yeast cells may include without limitation *Candida* spp. (e.g., *Candida albicans*), *Yarrowia* spp. (e.g., *Yarrowia hpolytica*), *Pichia* spp. (e.g., *Pichia pastoris*), *Kluyveromyces* spp. (e.g., *Kluyveromyces lactis* and *Kluyveromyces marxianus*), *Neurospora* spp. (e.g., *Neurospora crassa*), *Fusarium* spp. (e.g., *Fusarium oxysporum*), and *Issatchenkia* spp. (e.g., Issatchenkia *orientalis*, a.k.a. *Pichia kudriavzevii* and *Candida acidothermophilum*). In some embodiments, the fungal cell is a filamentous fungal cell. As used herein, the term "filamentous fungal cell" refers to any type of fungal cell that grows in filaments, i.e., hyphae or mycelia. Examples of filamentous fungal cells may include without limitation *Aspergillus* spp. (e.g., *Aspergillus niger*), *Trichoderma* spp. (e.g., *Trichoderma reesei*), *Rhizopus* spp. (e.g., *Rhizopus oryzae*), and *Mortierella* spp. (e.g., *Mortierella isabellina*).

[0196] In some embodiments, the fungal cell is an industrial strain. As used herein, "industrial strain" refers to any strain of fungal cell used in or isolated from an industrial process, e.g., production of a product on a commercial or industrial scale. Industrial strain may refer to a fungal species that is typically used in an industrial process, or it may refer to an isolate of a fungal species that may be also used for non-industrial purposes (e.g., laboratory research). Examples of industrial processes may include fermentation (e.g., in production of food or beverage products), distillation, biofuel production, production of a compound, and production of a polypeptide. Examples of industrial strains can include, without limitation, JAY270 and ATCC4124.

[0197] In some embodiments, the fungal cell is a polyploid cell. As used herein, a "polyploid" cell may refer to any cell whose genome is present in more than one copy. A polyploid cell may refer to a type of cell that is naturally found in a polyploid state, or it may refer to a cell that has been induced to exist in a polyploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). A polyploid cell may refer to a cell whose entire genome is polyploid, or it may refer to a cell that is polyploid in a particular genomic locus of interest.

[0198] In some embodiments, the fungal cell is a diploid cell. As used herein, a "diploid" cell may refer to any cell whose genome is present in two copies. A diploid cell may refer to a type of cell that is naturally found in a diploid state, or it may refer to a cell that has been induced to exist in a diploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). For example, the *S. cerevisiae* strain S228C may be maintained in a haploid or diploid state. A diploid cell may refer to a cell whose entire genome is diploid, or it may refer to a cell that is diploid in a particular genomic locus of interest. In some embodiments, the fungal cell is a haploid cell. As used herein, a "haploid" cell may refer to any cell whose genome is present in one copy. A haploid cell may refer to a type of cell that is naturally found in a haploid state, or it may refer to a cell that has been induced to exist in a haploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). For example, the *S. cerevisiae* strain S228C may be maintained in a haploid or diploid state. A haploid cell may refer to a cell whose entire genome is haploid, or it may refer to a cell that is haploid in a particular genomic locus of interest.

[0199] In some aspects, the cell is a cell obtained from a subject. In some embodiments, the subject is a healthy or non-diseased subject.

[0200] In some embodiments, a cell transfected with one or more vectors described herein is used to establish a new cell line comprising one or more vector-derived sequences. The cells can be used to produce the engineered systems. In some embodiments, the engineered systems are produced, harvested, and delivered to a subject in need thereof. In some embodiments, the engineered cells are delivered to a subject. Other uses for the engineered cells are described elsewhere herein.

[0201] In some aspects, the present disclosure also provides tissues, organs, or subjects (e.g., animals, plants, etc.) comprising one or more cells described above.

Engineered Animals

[0202] The present disclosure further provides engineered organisms that comprise the systems, polynucleotides, and/or vectors. The engineered organism, in some embodiments, can be an animal; for example, a mammal. In aspects, the organism is a non-human mammal. In an aspect, the invention provides a non-human eukaryotic organism; e.g., a multicellular eukaryotic organism, comprising a eukaryotic engineered cell according to any of the described embodiments. In other aspects, the invention provides a eukaryotic organism, preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. The engineered organism in some embodiments of these aspects may be an animal, for example, a mammal. In some embodiments, the engineered organism can be an arthropod such as an insect. In some embodiments, the engineered organism can be a farm or other production animals, including but not limited to pigs, goats, cattle, chickens, and sheep.

[0203] Various methods of generating transgenic animals that contain exogenous genetic material can be generated by various methods that will be appreciated by those of ordinary skill in the art. Such techniques include, but are not limited to, polynucleotide or virus microinjection into a pronucleus in a developing embryo, cell cytoplasm, or into the vasculature or blastoderm of a developing embryo (for example, in chickens); embryonic stem cell or other stem cell (e.g. pluripotent, multipotent, or induced pluripotent stem cell) manipulation (e.g. introduction of transgene or modification via gene editing); techniques utilizing a cre-lox approach, viral vectors, nuclear transfer, primoridial germ cell manipulation, spermatogonial manipulation. Many variations of these basic techniques have been done and are included within the scope of this disclosure. Exemplary methods for generating various transgenic animals can be found, for example, in any of the following, which are incorporated by reference as if expressed in their entirety: "Transgenic Animal Science: Principles and Methods" (1991) Charles River Laboratory; Hammer R. E, Pursel V. G, et al: Production of transgenic rabbits, sheep and pigs by microinjection. Nature 1985; 315(6021):680-683; Jaenisch R: Germ line integration and Mendelian transmission of the exogenous Moloney leukemia virus. Proc Natl Acad Sci. 1976; 73:1260-1264; Brackett B G, Boranska W, Sawicki W, Koprowski: Uptake of heterologous genome by mammalian spermatozoa and its transfer to ova through fertilization. Proc Natl Acad Sci.1971; 68:353-357; Gordon J. W, Scangos G. A, Plotkin D. J, Barbosa J. A, Ruddle F. H: Genetic transformation of mouse embryos by microinjection of purified DNA. Proc Natl Acad Sci.1980; 77:179-184; Lavitrano M, Camaioni A, Fazio V. M, Dolci S, Farace M. G, Spadafora C: Sperm cells as vectors for introducing foreign DNA into eggs: genetic transformation of mice. Cell 1989; 57(5):717-723; Chang K, Qian J, et al: Effective generation of transgenic pigs and mice by linker based sperm-mediated gene transfer. BMC Biotechnol. 2002; 2(1):5; Perry A. C, Wakayama T, Kishikawa H, Kasai T, Okabe M, Toyoda Y, Yanagimachi R: Mammalian transgenesis by intracytoplasmic sperm injection. Science 1999; 284 (5417):1180-1183; Clark J, Whitelaw B: A future for transgenic livestock. Rev. Genet. 2003; 4(10):825-833; Bowen R. A: Efficient production of transgenic cattle by retroviral infection of early embryos. Reprod. Dev. 1995; 40(3):386-390; Shim H, Gutierrez-Adan A, Chen L. R, BonDurant R. H, Behboodi E, Anderson G. B: Isolation of pluripotent stem cells from cultured porcine primordial germ cells. Reprod. 1997; 57(5): 1089-1095; Maclean, N: Animals with Novel Genes. Cambridge University Press. Cambridge, UK, 1995; Ebert, K. M, and Schindler J. E. S: Transgenic farm animals: Progress report. Theriogenology 1993; 39: 121-135; Gossler et al: Transgenesis by means of blastocyst-derived embryonic stem cell line, Proceedings of National Academic Science 1986; 83:9065-9069; Makoto Nagano, Clayton J. Brinster, et al: Transgenic mice produced by retroviral transduction of male germ-line stem cells. PNAS2001; 98(23):13090-13095; Alexander Baguisi et al: Production of goats by somatic cell nuclear transfer. Nature Biotechnology 1999; 17:456; Esponda P: Transfection of gametes. A method to generate transgenic animals. J. Morphol. 2005; 23(3):281-284; Andreas Sched, Zonia Larin, et al: A method for the generation of YAC transgenic mice by pronuclear microinjection. Nucleic Acids Research1993; 21(20):4783-4787; Ralph L. Brinster. Germline Stem Cell Transplantation and Transgenesis. Reproductive Biology Journal 2002; 296: 2174; Hofmann A, Zakhartchenko V, et al: Generation of transgenic cattle by lentiviral gene transfer into oocytes. Reprod. 2004; 71(2):405-409; Sang H. M: Transgenics, chickens and therapeutic proteins. Vox Sanguinis. 2004; 87(2):S164-5166; Meade H. M, Echelard Y, et al: Expression of recombinant proteins in the milk of transgenic animals. In Gene expression systems: using nature for the art of expression. Academic Press, San Diego. 1999; 399-427; Rudolph N. S: Biopharmaceutical production in transgenic livestock. Trends Biotechnol. 1999; 17(9):367-374; Kuroiwa Y, Kasinathan P, et al: Cloned transchromosomic calves producing human immunoglobulin. Nature Biotechnol. 2002; 20(9):889-894; Swabson M. E, Martin M. J, et al: Production of functional human hemoglobin in transgenic swine. Biotechnology 1992; 10(5):557-559, Niemann H: Transgenic pigs expressing plant genes. natl Acad. Sci.2004; 101(19):7211-7212.

Engineered Plants and Algae

[0204] The engineered organism, in some embodiments, can be a plant and algae that comprise the systems, polynucleotides, and/or vectors. In general, the term "plant" relates to any various photosynthetic, eukaryotic, unicellular or multicellular organism of the kingdom Plantae characteristically growing by cell division, containing chloroplasts, and having cell walls comprised of cellulose. The term plant encompasses monocotyledonous and dicotyledonous plants. In some embodiments, the engineered plant is a dicotyledonous plant belonging to the orders *Magniolales, Illiciales, Laurales, Piperales, Aristochiales, Nymphaeales, Ranunculales, Papeverales, Sarraceniaceae, Trochodendrales, Hamamelidales, Eucomiales, Leitneriales, Myricales, Fagales, Casuarinales, Caryophyllales, Batales, Polygonales, Plumbaginales, Dilleniales, Theales, Malvales, Urticales, Lecythidales, Violales, Salicales, Capparales,* *Ericales, Diapensales, Ebenales, Primulales, Rosales, Fabales, Podostemales, Haloragales, Myrtales, Cornales, Proteales, San tales, Rafflesiales, Celastrales, Euphorbiales, Rhamnales, Sapindales, Juglandales, Geraniales, Polygalales, Umbellales, Gentianales, Polemoniales, Lamiales, Plantaginales, Scrophulariales, Campanulales, Rubiales, Dipsacales*, and *Asterales*. In some embodiments, the plant is a monocotyledonous plant such as one belonging to an order of the group of: *Alismatales, Hydrocharitales, Najadales, Triuridales, Commelinales, Eriocaulales, Restionales, Poales, Juncales, Cyperales, Typhales, Bromeliales, Zingiberales, Arecales, Cyclanthales, Pandanales, Arales, Lilliales*, and *Orchid ales*, or with plants belonging to *Gymnospermae*, e.g. those belonging to the orders *Pinales, Ginkgoales, Cycadales, Araucariales, Cupressales* and *Gnetales*. In some embodiments, the engineered plant can be a plant of a species included in the non-limitative list of dicot, monocot or gymnosperm genera hereunder: *Atropa, Alseodaphne, Anacardium, Arachis, Beilschmiedia, Brassica, Carthamus, Cocculus, Croton, Cucumis, Citrus, Citrullus, Capsicum, Catharanthus, Cocos, Coffea, Cucurbita, Daucus, Duguetia, Eschscholzia, Ficus, Fragaria, Glaucium, Glycine, Gossypium, Helianthus, Hevea, Hyoscyamus, Lactuca, Landolphia, Linum, Litsea, Lycopersicon, Lupinus, Manihot, Majorana, Malus, Medicago, Nicotiana, Olea, Parthenium, Papaver, Persea, Phaseolus, Pistacia, Pisum, Pyrus, Prunus, Raphanus, Ricinus, Senecio, Sinomenium, Stephania, Sinapis, Solanum, Theobroma, Trifolium, Trigonella, Vicia, Vinca*, Vilis, and *Vigna*; and the genera *Allium, Andropogon, Aragrostis, Asparagus, Avena, Cynodon, Elaeis, Festuca, Festulolium, Heterocallis, Hordeum, Lemna, Lolium, Mus*a, *Oryza, Panicum, Pannesetum, Phleum, Poa, Secale, Sorghum, Triticum, Zea, Abies, Cunninghamia, Ephedra, Picea, Pinus*, and *Pseudotsuga*.

[0205] Specifically, the engineered plants are intended to include without limitation angiosperm and gymnosperm plants such as acacia, alfalfa, amaranth, apple, apricot, artichoke, ash tree, asparagus, avocado, banana, barley, beans, beet, birch, beech, blackberry, blueberry, broccoli, Brussel's sprouts, cabbage, canola, cantaloupe, carrot, cassava, cauliflower, cedar, a cereal, celery, chestnut, cherry, Chinese cabbage, citrus, clementine, clover, coffee, corn, cotton, cowpea, cucumber, cypress, eggplant, elm, endive, *eucalyptus*, fennel, figs, fir, geranium, grape, grapefruit, groundnuts, ground cherry, gum hemlock, hickory, kale, kiwifruit, kohlrabi, larch, lettuce, leek, lemon, lime, locust, pine, maidenhair, maize, mango, maple, melon, millet, mushroom, mustard, nuts, oak, oats, oil palm, okra, onion, orange, an ornamental plant or flower or tree, *papaya*, palm, parsley, parsnip, pea, peach, peanut, pear, peat, pepper, persimmon, pigeon pea, pine, pineapple, plantain, plum, pomegranate, potato, pumpkin, radicchio, radish, rapeseed, raspberry, rice, rye, sorghum, safflower, sallow, soybean, spinach, spruce, squash, strawberry, sugar beet, sugarcane, sunflower, sweet potato, sweet corn, tangerine, tea, tobacco, tomato, trees, triticale, turf grasses, turnips, vine, walnut, watercress, watermelon, wheat, yams, yew, and zucchini.

[0206] The term plant also encompasses Algae, which are mainly photoautotrophs unified primarily by their lack of roots, leaves and other organs that characterize higher plants. Thus, in some embodiments, the modified organism is an algae. "Algae" and "algae cells," include but are not limited to, algae or cells thereof selected from several eukaryotic phyla, including the Rhodophyta (red algae), Chlorophyta (green algae), Phaeophyta (brown algae), Bacillariophyta (diatoms), Eustigmatophyta and dinoflagellates as well as the prokaryotic phylum Cyanobacteria (blue-green algae). The term "algae" includes for example algae selected from Amphora, *Anabaena, Anikstrodesmis, Botryococcus, Chaetoceros, Chlamydomonas, Chlorella, Chlorococcum, Cyclotella, Cylindrotheca, Dunaliella, Emiliana, Euglena, Hematococcus, Isochrysis, Monochrysis, Monoraphidium, Nannochloris, Nannnochloropsis, Navicula, Nephrochloris, Nephroselmis, Nitzschia, Nodularia, Nostoc, Oochromonas, Oocystis, Oscillartoria, Pavlova, Phaeodactylum, Playtmonas, Pleurochrysis, Porhyra, Pseudoanabaena, Pyramimonas, Stichococcus, Synechococcus, Synechocystis, Tetraselmis, Thalassiosira*, and *Trichodesmium*.

[0207] As noted above, part of the plant may be engineered to include and/or express one or more components of the engineered system described herein. As used herein, "plant tissue" refers to part of the plant and includes cells. The term "plant cell" as used herein refers to individual units of a living plant, either in an intact whole plant or in an isolated form grown in in vitro tissue cultures, on media or agar, in suspension in a growth media or buffer or as a part of higher organized unites, such as, for example, plant tissue, a plant organ, or a whole plant.

[0208] As used herein, "protoplast" refers to a plant cell that has had its protective cell wall completely or partially removed using, for example, mechanical or enzymatic means resulting in an intact biochemical competent unit of living plant that can reform their cell wall, proliferate and regenerate grow into a whole plant under proper growing conditions.

Therapeutic and Diagnostic Applications

[0209] In another aspect, the present disclosure provides methods for treating diseases or conditions in a subject with the systems described herein. In some embodiments, the methods comprise administering one or more components of the systems, the polynucleotides, the vectors the cells, or any combination thereof, to a subject (e.g., a subject in need thereof). The systems may comprise or may cause production of therapeutic and/or diagnostic agents, such as the genetic modulating agents. in certain examples, the methods may comprise administering one or more cells comprising the vesicles or plasmids into a subject.

[0210] The diseases may be genetic diseases. Genetic diseases that can be treated are discussed in greater detail elsewhere herein. Other diseases include but are not limited to any of the following: cancer, Acubetivacter infections, actinomycosis, African sleeping sickness, AIDS/HIV, ameobiasis, Anaplasmosis, Angiostrongyliasis, Anisakiasis, Anthrax, Acranobacterium haemolyticum infection, Argentine hemorrhagic fever, Ascariasis, Aspergillosis, Astrovirus infection, Babesiosis, Bacterial meningitis, Bacterial pneumonia, Bacterial vaginosis, Bacteroides infection, balantidiasis, Bartonellosis, Baylisascaris infection, BK virus infection, Black Piedra, Blastocytosis, Blastomycosis, Bolivian hemorrhagic fever, Botulism, Brazilian hemorrhagic fever, brucellosis, Bubonic plague, Burkholderia infection, buruli ulcer, calicivirus invention, campylobacteriosis, Candidiasis, Capillariasis, Carrion's disease, Cat-scratch disease, cellulitis, Chagas Disease, Chancroid, Chickenpox, Chikungunya, Chlamydia, Chlamydia pneumoniae, Cholera, Chromoblastomycosis, Chytridiomycosis, Clonochiasis, Clostridium difficile colitis, Coccidioidomycosis, Colorado tick fever, rhinovirus/coronavirus infection (common cold), Cretzfeldt-Jakob disease, Crimean-congo hemorrhagic fever, Cryptococcosis, Cryptosporidiosis, Cutaneous larva migrans (CLM), cyclosporiasis, cysticercosis, cytomegalovirus infection, Dengue fever, Desmodesmus infection, Dientamoebiasis, Diphtheria, Diphylobothriasis, Dracunculiasis, Ebola, Echinococcosis, Ehrlichiosis, Enterobiasis, Enterococcus infection, Enterovirus infection, Epidemic typhus, Erthemia Infectisoum, Exanthem subitum, Fasciolasis, Fasciolopsiasis, fatal familial insomnia, filarisis, Clostridum perfingens infection, Fusobacterium infection, Gas gangrene (clostridial myonecrosis), Geotrichosis, Gerstmann-Straussler-Scheinker syndrome, Giardasis, Glanders, Gnathostomiasis, Gonorrhea, Granuloma inguinales, Group A streptococcal infection, Group B streptococcal infection, Haemophilus influenzae infection, Hand, foot, and mouth disease, hanta virus pulmonary syndrome, heartland virus disease, Helicobacter pylori infection, hemorrhagi fever with renal syndrome, Hendra virus infection, Hepatitis (all groups A, B, C, D, E), herpes simplex, histoplasmosis, hookworm infection, human bocavirus infection, human ewingii ehrlichiosis, Human granulocytic anaplasmosis, human metapneumovirus infection, human monocytic ehrlichiosis, human papilloma virus, Hymenolepiasis, Epstein-Barr infection, mononucleosis, influenza, isoporisis, Kawasaki disease, Kingell kingae infection, Kuru, Lasas fever, Legionellosis (Legionnaire's disease and Potomac Fever), Leishmaniasis, Leprosy, Leptospirosis, Listeriosis, Lyme disease, lymphatic filariasis, lymphocytic choriomeningitis, Malaria, Marburg hemorrhagic fever, measles, Middle East respiratory syndrome, Melioidosis, meningitis, Meningococcal disease, Metagonimiasis, Microsporidosis, Molluscum contagiosum, Monkeypox, Mumps, Murine typhus, Mycoplasma pneumonia, Mycoplasma genitalium infection, Mycetoma, Myiasis, Conjunctivitis, Nipah virus infection, Norovirus, Variant Creutzfeldt-Jakob disease, Nocardosis, Onchocerciasis, Opisthorchiasis, Paracoccidioidomycosis, Paragonimiasis, Pasteurellosis, Pediculosis capitis, Pediculosis corporis, Pediculosis pubis, pelvic inflammatory disease, pertussis, plague, pneumococcal infection, pneumocystis pneumonia, pneumonia, poliomyelitis, prevotella infection, primary amoebic meningoencephalitis, progressive multifocal leukoencephalopathy, Psittacosis, Qfever, rabies, relapsing fever, respiratory syncytial virus infection, rhinovirus infection, rickettsial infection, Rickettsia pox, Rift Valley Fever, Rocky Mountain Spotted Fever, Rotavirus infection, Rubella, Salmonellosis, SARS, Scabies, Scarlet fever, Schistosomiasis, Sepsis, Shigellosis, Shingles, Smallpox, Sporotrichosis, Staphylococcal infection (including MRSA), strongyloidiasis, subacute sclerosing panencephalitis, Syphilis, Taeniasis, tetanus, Trichophyton species infection, Tocariasis, Toxoplasmosis, Trachoma, Trichinosis, Trichuriasis, Tuberculosis, Tularemia, Typhoid Fever, Typhus Fever, Ureaplasma urealyticum infection, Valley fever, Venezuelan equine encephalitis, Venezuelan hemorrhagic fever, Vibrio species infection, Viral pneumonia, West Nile Fever, White Piedra, Yersinia pseudotuberculosis, Yersiniosis, Yellow fever, Zeaspora, Zika fever, Zygomycosis and combinations thereof.

[0211] Other diseases and disorders that can be treated using embodiments of the present invention include endocrine diseases (e.g. Type I and Type II diabetes, gestational diabetes, hypoglycemia. Glucagonoma, Goiter, Hyperthyroidism, hypothyroidism, thyroiditis, thyroid cancer, thyroid hormone resistance, parathyroid gland disorders, Osteoporosis, osteitis deformans, rickets, ostomalacia, hypopituitarism, pituitary tumors, etc.), skin conditions of infections and non-infection origin, eye diseases of infectious or non-infectious origin, gastrointestinal disorders of infectious or non-infectious origin, cardiovascular diseases of infectious or non-infectious origin, brain and neuron diseases of infectious or non-infectious origin, nervous system diseases of infectious or non-infectious origin, muscle diseases of infectious or non-infectious origin, bone diseases of infectious or non-infectious origin, reproductive system diseases of infectious or non-infectious origin, renal system diseases of infectious or non-infectious origin, blood diseases of infectious or non-infectious origin, lymphatic system diseases of infectious or non-infectious origin, immune system diseases of infectious or non-infectious origin, mental-illness of infectious or non-infectious origin and the like.

[0212] In some embodiments, the disease may be neuronal diseases. The systems herein may be delivered to neuronal cells or related cells for treating such diseases. Examples of diseases and cells include those described in Bergen J M et al., Nonviral Approaches for Neuronal Delivery of Nucleic Acids, Pharm Res. 2008 May; 25(5): 983-998.

Pharmaceutical Compositions

[0213] The systems, polynucleotides, vectors, and cells herein may be formulated as pharmaceutical compositions. A pharmaceutical composition may comprise an excipient, such as a pharmaceutically acceptable carrier, that is conventional in the art and that is suitable for administration to cells or to a subject.

[0214] In certain embodiments, the methods of the disclosure include administering to a subject in need thereof an effective amount (e.g., therapeutically effective amount or prophylactically effective amount) of the treatments provided herein. Such treatment may be supplemented with other known treatments, such as surgery on the subject. In certain embodiments, the surgery is strictureplasty, resection (e.g., bowel resection, colon resection), colectomy, surgery for abscesses and fistulas, proctocolectomy, restorative proctocolectomy, vaginal surgery, cataract surgery, or a combination thereof.

[0215] The term "pharmaceutically acceptable" as used throughout this specification is consistent with the art and means compatible with the other ingredients of a pharmaceutical composition and not deleterious to the recipient thereof. As used herein, "carrier" or "excipient" includes any and all solvents, diluents, buffers (such as, e.g., neutral buffered saline or phosphate buffered saline), solubilisers, colloids, dispersion media, vehicles, fillers, chelating agents (such as, e.g., EDTA or glutathione), amino acids (such as, e.g., glycine), proteins, disintegrants, binders, lubricants, wetting agents, emulsifiers, sweeteners, colorants, flavourings, aromatisers, thickeners, agents for achieving a depot effect, coatings, antifungal agents, preservatives, stabilisers, antioxidants, tonicity controlling agents, absorption delaying agents, and the like. The use of such media and agents for pharmaceutical active components is well known in the art. Such materials should be non-toxic and should not interfere with the activity of the cells or active components.

[0216] The precise nature of the carrier or excipient or other material will depend on the route of administration. For example, the composition may be in the form of a parenterally acceptable aqueous solution, which is pyrogen-free and has suitable pH, isotonicity and stability. For general principles in medicinal formulation, the reader is referred to Cell Therapy: Stem Cell Transplantation, Gene Therapy, and Cellular Immunotherapy, by G. Morstyn & W. Sheridan eds., Cambridge University Press, 1996; and Hematopoietic Stem Cell Therapy, E. D. Ball, J. Lister & P. Law, Churchill Livingstone, 2000.

[0217] The pharmaceutical compositions can be applied parenterally, rectally, orally or topically. For example, the pharmaceutical composition may be used for intravenous, intramuscular, subcutaneous, peritoneal, peridural, rectal, nasal, pulmonary, mucosal, or oral application. In a preferred embodiment, the pharmaceutical composition according to the invention is intended to be used as an infuse. The skilled person will understand that compositions which are to be administered orally or topically will usually not comprise cells, although it may be envisioned for oral compositions to also comprise cells, for example when gastro-intestinal tract indications are treated. Each of the cells or active components (e.g., modulants, immunomodulants, antigens) as discussed herein may be administered by the same route or may be administered by a different route. By means of example, and without limitation, cells may be administered parenterally and other active components may be administered orally. In some cases, the composition or pharmaceutical composition may by intramuscular injection. In some cases, the composition or pharmaceutical composition may by intravascular injection.

[0218] Liquid pharmaceutical compositions may generally include a liquid carrier such as water or a pharmaceutically acceptable aqueous solution. For example, physiological saline solution, tissue or cell culture media, dextrose or other saccharide solution or glycols such as ethylene glycol, propylene glycol or polyethylene glycol may be included.

[0219] The composition may include one or more cell protective molecules, cell regenerative molecules, growth factors, anti-apoptotic factors or factors that regulate gene expression in the cells. Such substances may render the cells independent of their environment.

[0220] Such pharmaceutical compositions may contain further components ensuring the viability of the cells therein. For example, the compositions may comprise a suitable buffer system (e.g., phosphate or carbonate buffer system) to achieve desirable pH, more usually near neutral pH, and may comprise sufficient salt to ensure isoosmotic conditions for the cells to prevent osmotic stress. For example, suitable solution for these purposes may be phosphate-buffered saline (PBS), sodium chloride solution, Ringer's Injection or Lactated Ringer's Injection, as known in the art. Further, the composition may comprise a carrier protein, e.g., albumin (e.g., bovine or human albumin), which may increase the viability of the cells.

[0221] Further suitably pharmaceutically acceptable carriers or additives are well known to those skilled in the art and for instance may be selected from proteins such as collagen or gelatine, carbohydrates such as starch, polysaccharides, sugars (dextrose, glucose and sucrose), cellulose derivatives like sodium or calcium carboxymethylcellulose, hydroxypropyl cellulose or hydroxypropylmethyl cellulose, pregelatinized starches, pectin agar, carrageenan, clays, hydrophilic gums (acacia gum, guar gum, arabic gum and xanthan gum), alginic acid, alginates, hyaluronic acid, polyglycolic and polylactic acid, dextran, pectins, synthetic polymers such as water-soluble acrylic polymer or polyvinylpyrrolidone, proteoglycans, calcium phosphate and the like.

[0222] If desired, cell preparation can be administered on a support, scaffold, matrix or material to provide improved tissue regeneration. For example, the material can be a granular ceramic, or a biopolymer such as gelatine, collagen, or fibrinogen. Porous matrices can be synthesized according to standard techniques (e.g., Mikos et al., Biomaterials 14: 323, 1993; Mikos et al., Polymer 35:1068, 1994; Cook et al., J. Biomed. Mater. Res. 35:513, 1997). Such support, scaffold, matrix or material may be biodegradable or non-biodegradable. Hence, the cells may be transferred to and/or cultured on suitable substrate, such as porous or non-porous substrate, to provide for implants.

[0223] The pharmaceutical compositions may comprise one or more pharmaceutically acceptable salts. The term "pharmaceutically acceptable salts" refers to salts prepared from pharmaceutically acceptable non-toxic bases or acids including inorganic or organic bases and inorganic or organic acids. Salts derived from inorganic bases include aluminum, ammonium, calcium, copper, ferric, ferrous, lithium, magnesium, manganic salts, manganous, potassium, sodium, zinc, and the like. Particularly preferred are the ammonium, calcium, magnesium, potassium, and sodium salts. Salts derived from pharmaceutically acceptable organic non-toxic bases include salts of primary, secondary, and tertiary amines, substituted amines including naturally occurring substituted amines, cyclic amines, and basic ion exchange resins, such as arginine, betaine, caffeine, choline, N,N'-dibenzylethylenediamine, diethylamine, 2-diethylaminoethanol, 2-dimethylaminoethanol, ethanolamine, ethylenediamine, N-ethyl-morpholine, N-ethylpiperidine, glucamine, glucosamine, histidine, hydrabamine, isopropylamine, lysine, methylglucamine, morpholine, piperazine, piperidine, polyamine resins, procaine, purines, theobromine, triethylamine, trimethylamine, tripropylamine, tromethamine, and the like. The term "pharmaceutically acceptable salt" further includes all acceptable salts such as acetate, lactobionate, benzenesulfonate, laurate, benzoate, malate, bicarbonate, maleate, bisulfate, mandelate, bitartrate, mesylate, borate, methylbromide, bromide, methylnitrate, calcium edetate, methyl sulfate, camsylate, mucate, carbonate, napsylate, chloride, nitrate, clavulanate, N-methylglucamine, citrate, ammonium salt, dihydrochloride, oleate, edetate, oxalate, edisylate, pamoate (embonate), estolate, palmitate, esylate, pantothenate, fumarate, phosphate/diphosphate, gluceptate, polygalacturonate, gluconate, salicylate, glutamate, stearate, glycollylarsanilate, sulfate, hexylresorcinate, subacetate, hydrabamine, succinate, hydrobromide, tannate, hydrochloride, tartrate, hydroxynaphthoate, teoclate, iodide, tosylate, isothionate, triethiodide, lactate, panoate, valerate, and the like which can be used as a dosage form for modifying the solubility or hydrolysis characteristics or can be used in sustained release or prodrug formulations. It will be understood that, as used herein, references to specific agents (e.g., neuromedin U receptor agonists or antagonists), also include the pharmaceutically acceptable salts thereof.

[0224] Methods of administrating the pharmacological compositions, including agents, cells, agonists, antagonists, antibodies or fragments thereof, to an individual include, but are not limited to, intradermal, intrathecal, intramuscular, intraperitoneal, intravenous, subcutaneous, intranasal, epidural, by inhalation, and oral routes. The compositions can be administered by any convenient route, for example by infusion or bolus injection, by absorption through epithelial or mucocutaneous linings (for example, oral mucosa, rectal and intestinal mucosa, and the like), ocular, and the like and can be administered together with other biologically-active agents. Administration can be systemic or local. In addition, it may be advantageous to administer the composition into the central nervous system by any suitable route, including intraventricular and intrathecal injection. Pulmonary administration may also be employed by use of an inhaler or nebulizer, and formulation with an aerosolizing agent. It may also be desirable to administer the agent locally to the area in need of treatment; this may be achieved by, for example, and not by way of limitation, local infusion during surgery, topical application, by injection, by means of a catheter, by means of a suppository, or by means of an implant.

[0225] Therapy or treatment according to the invention may be performed alone or in conjunction with another therapy, and may be provided at home, the doctor's office, a clinic, a hospital's outpatient department, or a hospital. Treatment generally begins at a hospital so that the doctor can observe the therapy's effects closely and make any adjustments that are needed. The duration of the therapy depends on the age and condition of the patient, the stage of the cancer, and how the patient responds to the treatment. Additionally, a person having a greater risk of developing an inflammatory response (e.g., a person who is genetically predisposed or predisposed to allergies or a person having a disease characterized by episodes of inflammation) may receive prophylactic treatment to inhibit or delay symptoms of the disease.

Vaccines

[0226] The systems, vesicles, plasmids, and cells may be used as vaccines. In some examples, the vesicles may comprise molecules capable of eliciting T cell and B cell immune responses. In some examples, the vesicles may not replicate once delivered in a target cell.

Bioproduction

[0227] The engineered system molecules, vectors, engineered cells, and/or engineered systems can be used for bioproduction of various molecules including engineered systems. In some embodiments, the engineered cells can be used in an in vivo (e.g. a modified animal or plant), in vitro, or ex vivo cell system to produce engineered systems. As previously mentioned, the engineered system molecules, vectors, engineered cells, and/or engineered systems can be used to make a modified animal that can produce engineered systems. In some embodiments, the animal can be engineered to produce engineered systems in one or more bodily fluids or product (e.g. an egg as in the case of modified avians). As previously mentioned, the engineered system molecules, vectors, engineered cells, and/or engineered systems can be used to make a modified plant that can produce engineered systems. In some embodiments, the plant can be engineered to produce engineered systems in one or more parts of the plant. In some embodiments, production can be in a harvestable portion of the plant.

[0228] In some embodiments, the objective can be to make and/or harvest a particular molecule from a producer cell. This can be useful for generating and harvesting molecules that are otherwise difficult to generate and/or harvest outside of a cell or via other processes and techniques. In some embodiments, the molecule is one that is naturally produced by the producer cell (which can be an engineered cell). In some embodiments, the producer cell can be engineered to increase production of one or more endogenous molecules. In some embodiments, the producer cell is engineered to produce an exogenous molecule. In some embodiments, endogenous and/or exogenous molecules produced can be packaged into engineered systems, which can be subsequently harvested from the producer cell. The molecules can then be further harvested from the engineered systems. Methods of purifying engineered systems are described elsewhere herein and will be appreciated by those of ordinary skill in the art. Similarly, methods of harvesting the molecules from the engineered systems will be appreciated by those of ordinary skill in the art.

[0229] In some cases, endogenous producer cell molecules or exogenous molecules of interest are normally secreted by the producer cell. Packaging these into engineered systems prior to secretion followed by subsequent purification of the engineered systems carrying the packaged endogenous molecule can be an alternative to obtaining conditioned media to obtain these normally secreted endogenous molecules.

[0230] The systems (e.g., the systems comprising ATPase(s) and adenosine deaminase(s) described herein) may be used to modify polynucleotides in vitro, in cells, and in vivo. Examples of applications, e.g., in plants, fungi, animals, therapeutic and diagnostic applications, include those described in International Patent Publication Nos. WO 2019/071048 (e.g. paragraphs [0528]-[0837]), WO 2019/084063 (e.g., paragraphs [0676]-[0892]), which are incorporated by reference herein in their entireties.

Delivery

[0231] The one or more components of the systems herein may be introduced to cells for expression. Examples of methods of introducing the components into cell include lipofection, nucleofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid: nucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA. Lipofection is described in e.g., U.S. Pat. Nos. 5,049,386, 4,946,787; and 4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, WO 91/17424; WO 91/16024. Delivery can be to cells (e.g. in vitro or ex vivo administration) or target tissues (e.g. in vivo administration). Physical methods of introducing polynucleotides may also be used. Examples of such methods include injection of a solution containing the polynucleotides, bombardment by particles covered by the polynucleotides, soaking a cell, tissue sample or organism in a solution of the polynucleotides, or electroporation of cell membranes in the presence of the polynucleotides. Examples of delivery methods and vehicles include viruses, nanoparticles, exosomes, nanoclews, liposomes, lipids (e.g., LNPs), supercharged proteins, cell permeabilizing peptides, and implantable devices. The nucleic acids, proteins and other molecules, as well as cells described herein may be delivered to cells, tissues, organs, or subjects using methods described in paragraphs [00117] to [00278] of Feng Zhang et al., (WO2016106236A1), which is incorporated by reference herein in its entirety.

EXAMPLES

Example 1—Identification of Bacterial Defense Systems

[0232] Bacterial defense systems were identified using method outlined in FIG. **5**, FIGS. **6**A-**6**B show the examples of the identified bacterial defense systems, their domain structures, and their effects on phage growth. Selected identified bacterial defense systems and mutated forms were tested for their effects on phage growth (FIG. **7**).

Example 2—Diverse Enzymatic Functions Mediate Antiviral Immunity in Prokaryotes

[0233] Bacteria and archaea possess multiple defense systems to protect against attacking viruses and other foreign genetic elements through a variety of mechanisms, including sequence-specific endonucleases and toxin-antitoxin systems. Here, using a systematic approach to identify defense-associated genes in bacterial and archaeal genomes, Applicants identified a diverse set of putative defense gene cassettes that remain functionally uncharacterized. Applicants heterologously reconstituted 50 of these cassettes in *Escherichia coli*, demonstrating that 29 of them mediated defense against specific bacteriophages. These new defense systems include retrons; a widespread family of reverse transcriptases with unusual domain associations; and STAND ATPases, which are homologs of essential eukaryotic apoptosis effectors but whose role in prokaryotes has remained enigmatic. In addition, Applicants demonstrated that a two-gene system containing a divergent adenosine deaminase mediates RNA editing upon exposure to phage, representing a novel mechanism of defense. The discovery of these novel defense systems highlighted the immense untapped diversity of molecular functions employed by microbes in their wars against viruses and provides clues to the evolutionary origins of microbial immune mechanisms.

[0234] Bacterial and archaeal viruses are the most abundant, and possibly the most diverse, biological entities on earth (Cobián Güemes et al., 2016; Suttle, 2013). To defend against the incessant and varied virus attacks, prokaryotes have evolved multiple, diverse antivirus defense systems. These include the adaptive immune systems CRISPR-Cas, which provide immunity by memorizing past infection events (Hille et al., 2018), and a variety of innate immune systems, such as restriction-modification (RM)-based systems, including DNA phosphorothioation, DPD, DISARM (Ofir et al., 2018), and BREX (Goldfarb et al., 2015; Gordeeva et al., 2019), which target specific, pre-defined sequences within the phage DNA; abortive infection (Abi) systems, which induce altruistic cell dormancy or death upon phage infection; and additional systems with mechanisms that have not yet been investigated (Doron et al. 2018). Antivirus defense systems range in complexity from a single small protein (e.g., certain types of Abi systems) to large cassettes of eight or more proteins acting in concert (e.g., type I and type III CRISPR-Cas systems).

[0235] The arms race between microbes and viruses is a powerful evolutionary force that sculpts the host genomes. A distinctive outcome of this process is the modularity of defense systems, whereby components of one system are often recruited by other systems. For example, restriction-modification enzymes have been found in association with a number of additional proteins, leading to expanded defense systems, such as DISARM (Ofir et al., 2018). Toxin-antitoxin systems are particularly prone to swapping, resulting in nearly every possible combination of toxin and antitoxin (Makarova et al., 2013). Another key feature of the evolution of microbial anti-parasite defense is the persistent exchange of components between defense systems and mobile genetic elements (Koonin et al., 2019). In particular, nucleases encoded by both transposons and toxin-antitoxin modules apparently have been recruited for roles in CRISPR-Cas systems, and conversely, components of CRISPR-Cas systems have been recruited by mobile genetic elements for antidefense and other functions, such as RNA-guided transpositions (Faure et al., 2019; Klompe et al., 2019; Strecker et al., 2019). The extensive modularity and baroque evolutionary patterns of defense systems yield extraordinary diversity and highlight the potential for discovery of additional systems with novel mechanisms.

Domain-Independent Identification of Uncharacterized Defense Systems

[0236] A distinctive property of anti-phage defense genes is their tendency to cluster together within defense 'islands' in bacterial and archaeal genomes (Makarova et al., 2013; Makarova et al., 2011). As a consequence, an uncharacterized gene whose homologs consistently occur next to, for instance, restriction-modification genes has an increased probability of being a new defense gene (Shmakov et al., 2019; Shmakov et al., 2018). A recent analysis (Doron et al., 2018) identified and validated 10 new defense systems, based on the requirement that each (putative) system contain at least one annotated protein domain that is enriched within defense islands.

[0237] To test whether additional unknown systems existed which either lack annotated domains, or only contain domains that are typically non-defense but have been co-opted in specific instances to perform defensive functions, Applicants developed an expanded computational approach in which putative novel systems were identified independent of domain annotations (FIG. **8**A). Applicants analyzed all 174,080 bacterial and archaeal genomes available in Genbank as of November 2018, encoding a total of 620 million proteins. To identify candidate novel defense systems, Applicants first compiled a list of all proteins within 10 kb or 10 open reading frames of known defense systems (see Methods). This list ($n=6\times10^5$ after redundancy reduction) was a mix of novel defense genes with many non-defense genes. For each entry in the list ('seed'), Applicants identified all homologs within the original set of genomes with an alignment coverage of at least 70% and an E-value of 10^{-5} or lower. Each detected homolog was then assessed for its proximity to a known defense system. For each seed, if the fraction of homologs within 5 kb of 5 genes of a known defense system ('defense association score') (Shmakov et al., 2019) was sufficiency high, the seed was retained for further analysis (see Methods). For each retained seed, the gene neighborhoods of 30 representative homologs were examined to identify conserved operons that contain the seed gene and putatively constitute a minimal intact defense system.

[0238] To determine an appropriate cutoff for the defense association score, Applicants performed the same analysis for a selected set of seeds from known systems. From this analysis, a value of 0.15 was chosen because >90% of the known seeds had a score higher than this value (FIG. **8**B). Applying this threshold to the novel seeds resulted in a final list of 1.5×10^4 defense gene candidates (10.5% of all seeds; minimum 50 identified homologs) (FIG. **8**C). This analysis suggested that uncharacterized defense systems substantially outnumbered the currently known ones. Furthermore, the defense-enriched seeds included a diversity of identified enzymatic activities, including those that had not been previously implicated in antivirus immunity.

Candidate Defense Systems Exhibited Antivirus Activity in a Heterologous System

[0239] Applicants selected 50 candidate defense systems to test experimentally by heterologous reconstitution in *E. coli*. Candidate systems were prioritized for testing based on the following criteria: presence of identified molecular functions not previously implicated in defense; broad phylogenetic distribution; and for multi-gene systems, conservation of component genes. For each system, 1-4 homologs were selected and cloned from the source organism into the low-copy vector pACYC and transformed into *E. coli* (FIG. **9**A). BREX type I (Goldfarb et al., 2015; Gordeeva et al., 2019), Druantia type I (Doron et al., 2018), and the abortive infection reverse transcriptase RT-Abi-P2 (Odegrip et al., 2006) were included as positive controls. Each system was then challenged with a diverse panel of coliphages with dsDNA, ssDNA, or ssRNA genomes, and phage sensitivity was compared to that observed with an empty vector control.

[0240] Applicants observed anti-phage activity in at least one homolog for 29 out of the 50 tested candidates (58%). The most active representative in each of these 29 systems was further tested with an expanded panel of phages in two *E. coli* strains (FIG. **9**B). All 29 systems were active against at least one dsDNA phage; three were active against ssDNA phages (M13 or φX174); and none were active against ssRNA phages (MS2 and Q(3). Phage specificity was typically narrow and varied widely across systems. In addition, the abundance of these systems within sequenced genomes spans two orders of magnitude, ranging from ~0.1% to ~10% of the genomes (FIG. **9**B and FIG. **14**).

RADAR Contained a Divergent Adenosine Deaminase that Edits RNA in Response to Phage Infection

[0241] One of the validated systems was a two-gene cassette consisting of a KAP-family ATPase (~900 residues) and a divergent adenosine deaminase (~900 residues); this system was active against dsDNA phages T2, T3, T4, and T5. Applicants focused on this system for further investigation because deaminase activity had not previously been implicated in anti-phage defense. These systems appear in diverse defense contexts, adjacent to CRISPR, BREX, RM, Zorya, and Wadjet, and form three distinct subtypes (FIG. **10**A). In some cases, this system had the ATPase and deaminase only, but some variants also included a small membrane protein, either a SLATT domain (Burroughs et al., 2015) or the type VI-B CRISPR ancillary gene csx27 (Makarova et al., 2019). Mutations in either the ATPase Walker B motif or in the putative Zn2+-binding H×H motif of the deaminase abolished defense activity (FIG. **10**B).

[0242] Applicants further tested whether it acted on nucleic acids. Indeed, whole-transcriptome deep sequencing showed an enrichment of A to G substitutions in sequencing reads at specific sites in the presence of phage, whereas C, G, or U bases were not affected (FIG. **10**C), consistent with base editing of adenosine to inosine. Editing occurred when both the defense system and the phage were present. In this experiment, expression of the defense system without the phage resulted in a near-baseline level of editing, and no editing was detected in the absence of the system. The editing sites were distributed throughout the *E. coli* transcriptome as well as the phage transcriptome (FIG. **10**D). RNA secondary structure analysis indicated a characteristic stem-loop structure at strong editing sites; specific adenosines in loops were edited with up to ~90% frequency, whereas adenosines within the stem were not edited within the limit of detection.

[0243] Based on these results, Applicants named this system phage restriction by an adenosine deaminase acting on RNA (RADAR). Growth kinetics at varying phage multiplicity of infection (MOI) revealed a threshold MOI above which RADAR-expressing cells had a lower OD600 compared to the empty vector control, suggestive of RADAR-mediated growth arrest (FIG. **10**E). Collectively, these results are consistent with an abortive infection mechanism that is activated by phage.

A Widespread Family of RT-Containing Defense Systems

[0244] The defense systems identified by the pipeline herein included a diverse family of reverse transcriptases (RTs). Although RTs are typical components of diverse mobile retroelements as well as retro-transcribing viruses, some RTs encoded in bacterial genomes show no evidence of mobility (Zimmerly and Wu, 2015). Two of these RTs have been previously shown to play a role in anti-phage defense, namely RT-Cas1, which mediated acquisition of CRISPR spacers from RNA via reverse transcription (Silas et al., 2016), and RT-Abi, a set of abortive infection genes that catalyzed untemplated dNTP polymerization in vitro (Emond et al., 1997; Odegrip et al., 2006; Wang et al., 2011).

[0245] Recent computational analyses have revealed a vast diversity of bacterial RTs, including 16 'unknown groups' (UGs) that either remained functionally uncharacterized, or were identified to perform metabolic roles (Kojima and Kanehisa, 2008; Simon and Zimmerly, 2008; Toro and Nisa-Martinez, 2014; Zimmerly and Wu, 2015). Many of these RTs were independently identified by the computational pipeline herein, suggesting that they might represent a widespread family of uncharacterized defense genes. Applicants found that at least 7 of these RT groups (UG1, UG2, UG3, UG8, UG9, UG15, and UG16) provided robust protection against dsDNA phages (FIG. **9**B), and mutations in the (Y/F)×DD (SEQ ID NOS: 1-2) active site of the RTs abolished activity (FIG. **11**A-**11**C). Many of these RTs contained an uncharacterized C-terminal domain, and some were fused to or associated with required enzymatic domains that had not been previously implicated in anti-phage defense, including a nitrilase-family C—N hydrolase and a family A DNA polymerase (FIGS. **11**A, B and FIG. **15**).

Retrons Mediated Anti-Phage Defense

[0246] Applicants also identified defense functions for a group of retrons, a distinct class of RTs that produce extrachromosomal satellite DNA (multi-copy single-stranded DNA, msDNA) by reverse transcribing a segment of the 5' region of its own mRNA (Lampson et al., 2005). Retron cDNA is covalently linked to an internal guanosine of the RNA via a 2'-5' phosphodiester bond. Retrons had been harnessed for bacterial genome engineering (Farzadfard and Lu, 2014), but their native biological function had remained unknown. Applicants found that the original *E. coli* retrons Ec67 (Lampson et al., 1989) and Ec86 (Lim and Maas, 1989), as well as the Ec78 retron (Lima and Lim, 1997) and a novel TIR domain-associated retron, mediated defense against dsDNA phages. In addition, the absence of additional domains typical for group II introns in the UG2 group, together with the presence of a large upstream region that formed a identified highly structured RNA, suggested that UG2 was yet another retron-like element. Mutations in the (Y/F)×DD (SEQ ID NOS: 1-2) active site of the RT, as well as a G to A substitution at the branching guanosine, abolished activity, indicating that the defense function depends on msDNA synthesis. Notably, these retrons were associated with other domains, including TOPRIM (topoisomerase-primase) (Aravind et al., 1998) and TIR (Tol/interleukin 1 receptor) domains, that were required for activity (FIG. **11**C). The TOPRIM domain can possess nuclease activity (Aravind et al., 1998) whereas the TIR domain can be a NAD+ hydrolase that is involved in programmed cell death pathways in animals and plants (Horsefield et al., 2019).

Additional Molecular Functions

[0247] Applicants identified other defense systems with diverse molecular functions, including a three-gene cassette containing a von Willebrand factor A (vWA) domain protein, a PP2C-like serine/threonine protein phosphatase, and a serine/threonine protein kinase provided strong protection against T7-like phages (T3, T7, and φV-1). In this experiment, all three genes were required for activity (FIG. **12**). This system, termed the TerY-phosphorylation triad (TerY-P), was previously analyzed computationally in the context of Ter-dependent stress response systems (Anantharaman et al., 2012) and can operate as a phosphorylation switch that couples the activities of the kinase and the phosphatase.

[0248] Four systems contained an N-terminal SIR2 (sirtuin) deacetylase domain (FIG. **12**), which was present in the Thoeris system (Doron et al., 2018) and had also been detected in the same neighborhoods with prokaryotic Argonaute proteins (Makarova et al., 2009), but had not been functionally characterized in prokaryotes. Additionally, a large 1300 residue P-loop ATPase containing two transmembrane helices inserted into the ATPase domain, similarly to the KAP family ATPases (Aravind et al., 2004), protected against both dsDNA and ssDNA phages.

[0249] Applicants also demonstrated defense function for several identified NTPases of the STAND (signal transduction ATPases with numerous associated domains) superfamily (FIG. **12**). This expansive superfamily consists of multidomain proteins that include eukaryotic ATPases and GTPases involved in programmed cell death and various forms of signal transduction (Danot et al., 2009; Leipe et al., 2004). Typically, STAND NTPases contain a C-terminal helical sensor that, upon target recognition, induces oligomerization via ATP or GTP hydrolysis, leading to activation of the N-terminal effector domain. The functions of prokaryotic STAND NTPases remain poorly characterized. Those few for which experimental data are available contain a helix-turn-helix domain and have been shown to regulate transcription (Danot et al., 2009). Several identified STAND NTPases were active against dsDNA phages (FIG. **9**B); these proteins contained different putative effector domains, including DUF4297 (a putative PD(D/E)×K-family nuclease that is also present in the Lamassu defense system (Doron et al., 2018)), an Mrr-like nuclease, SIR2, a trypsin-like serine protease, and an uncharacterized helical domain.

[0250] The findings described here substantially expanded the space of protein domains, molecular functions, and their interactions that are employed by bacteria in anti-phage defense. Some of these functions, in particular RNA editing, had not been previously implicated in defense mechanisms. The high success rate of the identification of defense systems based solely on the evolutionary conservation of the proximity to previously identified defense genes validated the defense island concept (Makarova et al., 2013; Makarova et al., 2011) and demonstrated its growing utility at the time of rapid expansion of sequence databases.

[0251] Despite similarities in domain architectures among some of the identified defense systems, their phage specificities differed substantially. The molecular basis of such narrow specificity remained to be uncovered, but these observations emphasized the importance of multiple defense systems for the survival of prokaryotes in the incessant arms race with viruses. Furthermore, these results were compatible with the concept of distributed microbial immunity, according to which defense systems encoded in different genomes collectively protect microbial communities from the diverse viromes they confront. The remarkable variability of the discovered defense systems implied that their sensor and effector components were involved in diverse molecular interactions. Several of the identified defense systems incorporated molecular functions from typically non-defense sources, highlighting the versatility of activities that were recruited for antiviral defense. The notable cases in point include the RNA deaminase activity of the RADAR system, as well as reverse transcriptases of different families, in particular retrons. The demonstration of the defense functions for multiple RTs that were generally associated with mobile genetic elements was consistent with the 'guns for hire' paradigm whereby enzymes are shuttled between MGE and defense systems during microbial evolution (Koonin et al., 2019).

[0252] The discovered defense systems can be characterized mechanistically, e.g., by mutating the catalytic residues. Applicants showed here that the respective enzymatic components were functionally important. Many of these systems can function via an abortive infection mechanism, e.g., by causing growth arrest or programmed cell death in the infected hosts as demonstrated here for the RADAR system. In particular, this can be the mode of action of STAND NTPases, homologs of essential eukaryotic programmed cell death effectors, whose role in prokaryotes has long remained enigmatic (Koonin and Aravind, 2002; Leipe et al., 2004). In addition, the membrane-associated ATPase can function analogously to the STAND NTPases to which they are distantly related (Aravind et al., 2004).

[0253] Many of the identified defense systems contained enzymatic activities as well as identified sensor components that had not been previously detected in defense contexts, suggesting the possibility of reengineering for novel biotechnology applications. Further experimental characterization of these systems, as well as others Applicants identified computationally, can be expected to greatly expand the repertoire of such functions.

Methods

[0254] Detection of known antivirus defense systems. All bacterial and archaeal genomes (n=174,080) were downloaded from Genbank (ftp://ftp.ncbi.nih.gov/genomes/genbank/) in November 2018. For genomes where gene annotations were incomplete or missing, genes were identified using Prodigal (Hyatt et al., 2010). Known defense-related protein domains were annotated using RPSBLAST version 2.8.1 from a set of position-specific scoring matrices curated from the NCBI Conserved Domain Database (CDD) (Doron et al., 2018; Makarova et al., 2011; Marchler-Bauer et al., 2017; Punta et al., 2012). To reduce the false positive rate, a multi-gene system containing a ubiquitous protein domain was required to include two or more of its component genes in close proximity. For example, the type I restriction-modification endonuclease hsdR was called as a defense gene only if the corresponding methylase (hsdM) or specificity protein (hsdS) was also encoded in the vicinity. Toxin-antitoxin systems were excluded from the set of known defense systems due to their overall low enrichment within defense islands.

[0255] Candidate novel defense genes. All translated protein-coding sequences within either 10 kb or 10 genes of known defense systems (whichever was greater), including the components of the known defense systems themselves, were compiled into a preliminary list ($n=8.7\times10^6$). Highly similar sequences (at least 98% sequence identity and coverage) were discarded using the linclust option in MMseqs2 (Steinegger and Riding, 2017, 2018) with parameters—min-seq-id 0.98-c 0.98, resulting in a reduced list of 2.5×10^6 sequences. A second round of redundancy elimination was then applied to this reduced list, using the default cluster option in MMSeqs2, yielding a final list of 6.0×10^5 candidate sequences.

[0256] Scoring candidate genes for defense enrichment. For each of the 6.0×10^5 candidate genes, a 'defense enrichment score' was computed as (number of homologs in proximity to one or more known defense systems)/(total number of homologs). A gene was considered to be located in proximity to a known defense system if it occurred no more than 5 kb or 5 genes away from the locus encoding that system. Candidate sequences with a defense enrichment score of 0.15 or higher were retained for subsequent analysis, with the exception of mobilome components (such as transposons), toxin-antitoxin, or abortive infection components, which were discarded. This cut-off was chosen because more than 90% of the known defense genes scored higher than this value. To identify homologs of the candidate proteins, all 6.2×10^8 proteins in the original set of Genbank genomes were tabulated, and highly similar proteins (at least 98% sequence identity and coverage) were removed using linclust, resulting in a reduced list of 1.3×10^8 proteins. Each seed sequence was then searched against this non-redundant protein sequence database using MiMseqs2. To qualify as homologs, matches were required to have a minimum coverage of 70% and a maximum E value of 10^{-5} (parameters—coy-mode 0-c 0.7-e 0.00001).

[0257] From genes to defense systems. For each defense-enriched candidate protein, the gene neighborhoods of 30 homologs in proximity to known defense genes were randomly selected and examined on a case by case basis, in order to determine whether the candidate was a stand-alone defense gene system or a member of a conserved multi-gene cassette. Protein domains were identified using HHpred, and the resulting identification were used to infer the involvement of the respective proteins in the activity of the respective identified defense system (Zimmermann et al., 2018).

[0258] Abundance estimation of defense systems. To estimate the abundance of each validated defense system within the microbial pangenome, Applicants downloaded n=205214 genomes available in Genbank as of August 2019. For each defense system, initial protein sequence seeds of the signature genes were taken from experimentally validated loci. Initial seeds were aligned and converted into HMM profiles. Applicants then used a constrained 2 iteration HMM profile search to generate highly specific HMM profiles and retrieve related systems as follows. Each ORF of size 150aa or greater with one or more hits was searched against all HMM profiles using HMMER3.1 and assigned to the profile that had the highest scoring match. For each system, ORFs with profile hits with less than 500 bp of intergenic distance on the same strand were grouped into candidate loci. For multi-protein systems, a putative locus was considered a hit if every signature gene profile for the system had a match in the locus with a bitscore of at least 25. For single gene systems, a locus was considered a hit if the protein had a match to the system's single signature gene profile with a bit score of at least 50 and an alignment coverage of at least 70%. Signature proteins from the identified systems were separately clustered at 50% identity using MMseqs2 and subsequently aligned using MAFFT. The alignments were used to create a new set of signature gene profiles as input to the next iteration. For BREX and Type I RM, Applicants used preexisting pfam profiles for the signature genes in place of iterative HMM profile searching. The final abundance was calculated as the number of system hits divided by the number of genomes (n).

[0259] Bacteria and phage strains. Phages T2, T3, T4, T5, T7, P1, λ, φV-1, M13, φX174, MS2, and Qβ, as well as host *E. coli* strains K-12 (ATCC25404) and C (ATCC13706), were obtained from the American Type Culture Collection (ATCC). The genome of phage φV-1, originally isolated from a measles vaccine (Milstien et al., 1977; Petricciani et al., 1973), was sequenced and found to be 92% similar to enterobacteria phage 285P, a T7-like phage (Xu et al., 2014).

[0260] Cloning. To facilitate experimental validation using coliphages, the source organism of each candidate defense system was chosen to be as similar as possible to *E. coli*, in particular, from other strains of *E. coli* whenever possible. Candidate defense systems were cloned into a variant of the low-copy plasmid pACYC184 containing 7 synonymous mutations in the chloramphenicol resistance gene to remove restriction sites. When possible, genomic DNA from source organisms was obtained from ATCC, NCTC, or DSMZ, and the genes of interest were amplified with Q5 (New England Biolabs) or Phusion Flash (Thermo Scientific) polymerase, using primers with 5' ends homologous to the ends of the plasmid backbone. Plasmids were assembled using the NEBuilder HiFi DNA Assembly mix (New England Biolabs). When the source organism was not readily available from public culture collections, genes were chemically synthesized (GenScript) with optional human codon optimization of the open reading frames. When possible, the native promoter was retained. For some source organisms outside of Enterobacteriaceae, or when the candidate system was operonized with other upstream genes, the system was placed under a bla or lac promoter.

[0261] Sequence verification of plasmids. The full sequences of all plasmids were verified by high-throughput sequencing. To prepare sequencing libraries, 25-50 ng of each plasmid was mixed with purified Tn5 transposome loaded with Illumina adapters and incubated at 55° C. for 10 min in the presence of 5 mM MgCl2 and 10 mM TAPS buffer (Picelli et al., 2014). The quantity of Tn5 was titrated to generate an average fragment size of ~100-400 bp. Tagmentation reactions were subsequently treated with 0.5 volumes of 0.1% sodium dodecyl sulfate for 5 min at room temperature and amplified with KAPA HiFi HotStart polymerase using primers containing 8 nt i7 and i5 index barcodes. Barcoded amplicons were sequenced on a MiSeq (Illumina) with at least 150 cycles for the forward read. Reads were aligned to the reference plasmid sequence by the Geneious read mapper, and error-free plasmids were retained for subsequent experiments.

[0262] Competent cell production. *E. coli* strains K-12 and C were cultured in ZymoBroth with 25 μg/mL chloramphenicol and made competent using Mix & Go buffers (Zymo) according to the manufacturer's recommended protocol.

[0263] Phage plaque assays. *E. coli* host strains were grown to saturation at 37° C. in Luria Broth (LB). To 10 mL top agar (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 7 g/L agar) was added chloramphenicol (final concentration 25 μg/mL) and 526 μL *E. coli* culture, and the mixture was poured on 10 cm LB-agar plates containing 25 μg/mL chloramphenicol. For phages T2, T4, T5, P1, λ, M13, MS2, and Qβ, dilutions of phage in phosphate buffered saline were spotted on the plates, and plaque counts were recorded after overnight incubation at 37° C. If individual plaques were too small to be counted, the most concentrated dilution at which no plaque formation was visible was recorded as having a single plaque. For phages T3, T7, φV-1, and φX174, a total of 3 μL of phage containing 5×106 virions was spotted, and the area of the plaque was measured after incubation at 37° C. for 68 hr.

[0264] Phage cultivation. Phages T2, T3, T4, T7, φV-1, M13, φX174, MS2, and Qβ were propagated in liquid culture. The host *E. coli* strain for each phage was grown to an OD600 of 0.2-0.4 at 37° C. in LB and infected with a slab of top agar containing phage plaque from a previous lysis. Cultures were grown overnight at 37° C. with 250 rpm agitation. Phages T5, P1, and λ, were propagated by the double agar overlay method; after overnight incubation at 37° C., plaques were scraped in LB. For both liquid culture and double agar overlay, phage samples were centrifuged to pellet cellular debris, and the supernatant was filtered through with a 0.22 μm sterile filter.

[0265] Whole transcriptome sequencing. *E. coli* ATCC25404, containing either an empty vector or the candidate defense system, was grown to log phase in LB and diluted to an OD600 of 0.2. The culture was then split into two tubes, one of which was infected with phage T2 at an estimated MOI of 2. Both subcultures were incubated at 37° C. for 1 hr with 250 rpm agitation. RNA was extracted using TRIzol Reagent (Thermo Fisher Scientific) and treated with DNAse I, followed by a RiboMinus ribosomal RNA depletion kit (Thermo). Sequencing libraries were prepared using NEB Ultra II directional RNAseq library prep kit (New England Biolabs) and paired-end sequenced (2×75 cycles) with a NextSeq (Illumina). Adapter sequences were trimmed from sequencing reads using CutAdapt (with parameters—trim-n-q 20-m 20-a AGATCGGAAGAGC-A AGATCGGAAGAGC (SEQ ID NO: 472)), and trimmed reads were aligned to the *E. coli* MG1655 reference genome using the Geneious read mapper.

[0266] RNA secondary structure. Minimum free energy RNA secondary structures were generated using the Turner (2004) energy parameters at 37° C. (Turner and Mathews, 2010).

[0267] *E. coli* growth kinetics. Cells were grown to log phase in LB and diluted to an OD600 of 0.2. Cultures were infected with phage T2 at varying MOI at grown at 37° C., and the OD600 was measured every 2 min for a total duration of 4 hr on a Synergy Neo2 plate reader (BioTek).

[0268] Anantharaman, V., Iyer, L. M., and Aravind, L. (2012). Ter-dependent stress response systems: novel pathways related to metal sensing, production of a nucleoside-like metabolite, and DNA-processing. Mol Biosyst 8, 3142-3165.

[0269] Aravind, L., Iyer, L. M., Leipe, D. D., and Koonin, E. V. (2004). A novel family of P-loop NTPases with an unusual phyletic distribution and transmembrane segments inserted within the NTPase domain. Genome Biol 5, R30.

[0270] Aravind, L., Leipe, D. D., and Koonin, E. V. (1998). Toprim—a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. Nucleic Acids Res 26, 4205-4213.

[0271] Burroughs, A. M., Zhang, D., Schäffer, D. E., Iyer, L. M., and Aravind, L. (2015). Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. Nucleic Acids Res 43, 10633-10654.

[0272] Cobián Güemes, A. G., Youle, M., Cantú, V. A., Felts, B., Nulton, J., and Rohwer, F. (2016). Viruses as Winners in the Game of Life. Annu Rev Virol 3, 197-214.

[0273] Danot, O., Marquenet, E., Vidal-Ingigliardi, D., and Richet, E. (2009). Wheel of Life, Wheel of Death: A Mechanistic Insight into Signaling by STAND Proteins. Structure 17, 172-182.

[0274] Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., Amitai, G., and Sorek, R. (2018). Systematic discovery of antiphage defense systems in the microbial pangenome. Science 359.

[0275] Emond, E., Holler, B. J., Boucher, I., Vandenbergh, P. A., Vedamuthu, E. R., Kondo, J. K., and Moineau, S. (1997). Phenotypic and genetic characterization of the bacteriophage abortive infection mechanism AbiK from *Lactococcus lactis*. Appl Environ Microbiol 63, 1274-1283.

[0276] Farzadfard, F., and Lu, T. K. (2014). Synthetic biology. Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. Science 346, 1256272.

[0277] Faure, G., Shmakov, S. A., Yan, W. X., Cheng, D. R., Scott, D. A., Peters, J. E., Makarova, K. S., and Koonin, E. V. (2019). CRISPR-Cas in mobile genetic elements: counter-defence and beyond. Nat Rev Microbiol 17, 513-525.

[0278] Goldfarb, T., Sberro, H., Weinstock, E., Cohen, O., Doron, S., Charpak-Amikam, Y., Afik, S., Ofir, G., and

Sorek, R. (2015). BREX is a novel phage resistance system widespread in microbial genomes. EMBO J 34, 169-183.

[0279] Gordeeva, J., Morozova, N., Sierro, N., Isaev, A., Sinkunas, T., Tsvetkova, K., Matlashov, M., Truncaite, L., Morgan, R. D., Ivanov, N. V., et al. (2019). BREX system of *Escherichia coli* distinguishes self from non-self by methylation of a specific DNA site. Nucleic Acids Res 47, 253-265.

[0280] Hille, F., Richter, H., Wong, S. P., Bratovič, M., Ressel, S., and Charpentier, E. (2018). The Biology of CRISPR-Cas: Backward and Forward. Cell 172, 1239-1259.

[0281] Horsefield, S., Burdett, H., Zhang, X., Manik, M. K., Shi, Y., Chen, J., Qi, T., Gilley, J., Lai, J. S., Rank, M. X., et al. (2019). NAD. Science 365, 793-799.

[0282] Hyatt, D., Chen, G. L., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11, 119.

[0283] Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S., and Sternberg, S. H. (2019). Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration. Nature 571, 219-225.

[0284] Kojima, K. K., and Kanehisa, M. (2008). Systematic survey for novel types of prokaryotic retroelements based on gene neighborhood and protein architecture. Mol Biol Evol 25, 1395-1404.

[0285] Koonin, E. V., and Aravind, L. (2002). Origin and evolution of eukaryotic apoptosis: the bacterial connection. Cell Death Differ 9, 394-404.

[0286] Koonin, E. V., Makarova, K. S., Wolf, Y. I., and Krupovic, M. (2019). Evolutionary entanglement of mobile genetic elements and host defence systems: guns for hire. Nat Rev Genet.

[0287] Lampson, B. C., Inouye, M., and Inouye, S. (2005). Retrons, msDNA, and the bacterial genome. Cytogenet Genome Res 110, 491-499.

[0288] Lampson, B. C., Sun, J., Hsu, M. Y., Vallejo-Ramirez, J., Inouye, S., and Inouye, M. (1989). Reverse transcriptase in a clinical strain of *Escherichia coli*: production of branched RNA-linked msDNA. Science 243, 1033-1038.

[0289] Leipe, D. D., Koonin, E. V., and Aravind, L. (2004). STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. J Mol Biol 343, 1-28.

[0290] Lim, D., and Maas, W. K. (1989). Reverse transcriptase-dependent synthesis of a covalently linked, branched DNA-RNA compound in *E. coli* B. Cell 56, 891-904.

[0291] Lima, T. M., and Lim, D. (1997). A novel retron that produces RNA-less msDNA in *Escherichia coli* using reverse transcriptase. Plasmid 38, 25-33.

[0292] Makarova, K. S., Gao, L., Zhang, F., and Koonin, E. V. (2019). Unexpected connections between type VI-B CRISPR-Cas systems, bacterial natural competence, ubiquitin signaling network and DNA modification through a distinct family of membrane proteins. FEMS Microbiol Lett 366.

[0293] Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2013). Comparative genomics of defense systems in archaea and bacteria. Nucleic Acids Res 41, 4360-4377.

[0294] Makarova, K. S., Wolf, Y. I., Snir, S., and Koonin, E. V. (2011). Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. J Bacteriol 193, 6039-6056.

[0295] Makarova, K. S., Wolf, Y. I., van der Oost, J., and Koonin, E. V. (2009). Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. Biol Direct 4, 29.

[0296] Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., et al. (2017). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res 45, D200-D203.

[0297] Milstien, J. B., Walker, J. R., and Petricciani, J. C. (1977). Bacteriophages in live virus vaccines: lack of evidence for effects on the genome of rhesus monkeys. Science 197, 469-470.

[0298] Odegrip, R., Nilsson, A. S., and Haggard-Ljungquist, E. (2006). Identification of a gene encoding a functional reverse transcriptase within a highly variable locus in the P2-like coliphages. J Bacteriol 188, 1643-1647.

[0299] Ofir, G., Melamed, S., Sberro, H., Mukamel, Z., Silverman, S., Yaakov, G., Doron, S., and Sorek, R. (2018). DISARM is a widespread bacterial defence system with broad anti-phage activities. Nat Microbiol 3, 90-98.

[0300] Petricciani, J. C., Chu, F. C., Johnson, J. B., and Meyer, H. M. (1973). Bacteriophages in live virus vaccines. Proc Soc Exp Biol Med 144, 789-792.

[0301] Picelli, S., Björklund, A. K., Reinius, B., Sagasser, S., Winberg, G., and Sandberg, R. (2014). Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res 24, 2033-2040.

[0302] Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., et al. (2012). The Pfam protein families database. Nucleic Acids Res 40, D290-301.

[0303] Shmakov, S. A., Faure, G., Makarova, K. S., Wolf, Y. I., Severinov, K. V., and Koonin, E. V. (2019). Systematic prediction of functionally linked genes in bacterial and archaeal genomes. Nat Protoc 14, 3013-3031.

[0304] Shmakov, S. A., Makarova, K. S., Wolf, Y. I., Severinov, K. V., and Koonin, E. V. (2018). Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. Proc Natl Acad Sci USA 115, E5307-E5316.

[0305] Silas, S., Mohr, G., Sidote, D. J., Markham, L. M., Sanchez-Amat, A., Bhaya, D., Lambowitz, A. M., and Fire, A. Z. (2016). Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. Science 351, aad4234.

[0306] Simon, D. M., and Zimmerly, S. (2008). A diversity of uncharacterized reverse transcriptases in bacteria. Nucleic Acids Res 36, 7219-7229.

[0307] Steinegger, M., and Soding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol 35, 1026-1028.

[0308] Steinegger, M., and Soding, J. (2018). Clustering huge protein sequence sets in linear time. Nat Commun 9, 2542.

[0309] Strecker, J., Ladha, A., Gardner, Z., Schmid-Burgk, J. L., Makarova, K. S., Koonin, E. V., and Zhang, F. (2019). RNA-guided DNA insertion with CRISPR-associated transposases. Science 365, 48-53.

[0310] Suttle, C. A. (2013). Viruses: unlocking the greatest biodiversity on Earth. Genome 56, 542-544.

[0311] Toro, N., and Nisa-Martinez, R. (2014). Comprehensive phylogenetic analysis of bacterial reverse transcriptases. PLoS One 9, e114083.

[0312] Turner, D. H., and Mathews, D. H. (2010). NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. Nucleic Acids Res 38, D280-282.

[0313] Wang, C., Villion, M., Semper, C., Coros, C., Moineau, S., and Zimmerly, S. (2011). A reverse transcriptase-related protein mediates phage resistance and polymerizes untemplated DNA in vitro. Nucleic Acids Res 39, 7620-7629.

[0314] Xu, B., Ma, X., Xiong, H., and Li, Y. (2014). Complete genome sequence of 285P, a novel T7-like polyvalent *E. coli* bacteriophage. Virus Genes 48, 528-533.

[0315] Zimmerly, S., and Wu, L. (2015). An Unexplored Diversity of Reverse Transcriptases in Bacteria. Microbiol Spectr 3, MDNA3-0058-2014.

[0316] Zimmermann, L., Stephens, A., Nam, S. Z., Rau, D., Kithler, J., Lozajic, M., Gabler, F., Soding, J., Lupas, A. N., and Alva, V. (2018). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. J Mol Biol 430, 2237-2243.

TABLE 5

Source organism strains of validated defense systems.

#	System	Genes	Organism	Strain	Promoter
	BREX type I	6	*E. coli*	DSM5212	Native
	Druantia type I	5	*E. coli*	DSM5212	Native
	RT-Abi-P2	1	*E. coli*	ECOR30	Native
1	RT_retron-TIR	1	*Shigella dysenteriae*	NCTC2966	Native
2	RT_retron-TOPRIM (Ec67)	1	*E. coli*	NCTC8623	Native
3	Nuc_deoxy + RT_retron (Ec86)	2	*E. coli*	BL21	Native
4	RT_UG2	1	*Salmonella enterica*	NCTC8273	Native
5	RT_UG15	1	*E. coli*	21-C8-A	Native
6	RT_UG16	1	*E. coli*	KTE25	Native
7	RT_UG1-nitrilase	2	*Klebsiella pneumoniae*	NCTC9143	Native
8	RT_UG3 + RT_UG8	2	*E. coli*	ECOR12	Native
9	ATPase_AAA + Ada	2	*Citrobacter rodentium*	ATCC51459	Native
10	ATPase_KAP_TM	1	*E. coli*	ECOR25	Native
11	ATPase_KAP + QueC + DNase_TatD	4	*E. coli*	NCTC9009	Native
12	DUF4011-Helicase_SF1_Dna2-Nuclease_Vsr-DUF3320	1	*E. coli*	ATCC43886	Native
13	ATPase_GHKL + Helicase_SF2_HepA	2	*Vibrio harveyi*	ATCC43516	bla
14	MBL + Protease_S1-ATPase_STAND	3	*Erwinia piriflorinigrans*	CFBP5888	bla
15	DUF4297-ATPase_STAND	2	*Salmonella enterica*	NCTC13175	Native
16	ATPase_STAND	1	*E. coli*	NCTC9087	Native
17	Nuclease_Mrr-ATPase_STAND	1	*E. coli*	NCTC11132	Native
18	SIR2-ATPase_STAND	1	*E. coli*	NCTC13384	Native
19	SIR2-DUF4020	1	*E. coli*	NCTC9112	Native
20	SIR2	1	Cronobacter sakazakii	NCTC8155	Native
21	SIR2 + Helicase_HerA	2	*E. coli*	NCTC11129	Native
22	Nuclease_DUF4297 + Helicase_HerA	2	*E. coli*	NCTC11131	Native
23	vWA + phosphatase_PP2C + STK-IB	3	*E. coli*	NCTC9094	Native
24	Phosphoesterase_PHP-ATPase_SMC	1	*E. coli*	NCTC8620	Native
25	Nuclease_DUF1887	1	*Salmonella enterica*	NCTC6026	Native
26	ATPase_AAA + Protease_S8	2	*E. coli*	ECOR52	Native
27	ATPase_DUF499 + DUF3780 + Methylase_DUF1156 + Nuclease_PLD-Helicase_HepA	4	*E. coli*	ECOR58	Native
28	RT_IG9 + DNA Po1A	2	*Pseodomonas brassicacearum*	Wood1	lac Native
29	RT_retron _ ATPase_AAA + HNH (Ec78)	3	*E. coli*	ECONIH5	Native

TABLE 6

PCR primers used to amplify genomic DNA source organisms containing validated defense systems.

#	Primer	Sequence
BREX type I	Fwd	gctaacttacattaattgcgttgcgcaACAGCACCACGTTCATCTTCC (SEQ ID NO: 98)
	Rev	ccaaggggttatgctagttattgcgGTTCATTAAAATAGTTACTACGTTAATTCACACCC (SEQ ID NO: 99)
Druantia type I	Fwd	gctaacttacattaattgcgttgcgcaGGTGAACGTTTGGTTGATAGGG (SEQ ID NO: 100)
	Rev	ccaaggggttatgctagttattgcgCTCAATGGGCATAATTTTACATTGTGC (SEQ ID NO: 101)
RT-Abi-P2	Fwd	gctaacttacattaattgcgttgcgcaACATCCCGTCATCATGCCATC (SEQ ID NO: 102)
	Rev	ccaaggggttatgctagttattgcgCTCCTCGGAATAGAATGTTATGTTCG (SEQ ID NO: 103)
1		Synthesized
2	Fwd	gctaacttacattaattgcgttgcgcaCGCGCTATCACGTAAAATAGGC (SEQ ID NO: 104)
	Rev	ccaaggggttatgctagttattgcgCGAAAAATCAGCCTTAGCGTTCATAAC (SEQ ID NO: 105)
3	Fwd	gctaacttacattaattgcgttgcgcaGCTCATGTTATGCATGTGCATG (SEQ ID NO: 106)
	Rev	ccaaggggttatgctagttattgcgATTAGGTCTTCGCTTTATTTAAAGGGTTC (SEQ ID NO: 107)
4		Synthesized
5		Synthesized
6		Synthesized
7	Fwd	gagctaacttacattaattgcgttgcgcaGTCCTTAAACACGACAAAACCTGTG (SEQ ID NO: 108)
	Rev	cccaaggggttatgctagttattgcgCGCAATGTAACACCCACCC (SEQ ID NO: 109)
8	Fwd	gctaacttacattaattgcgttgcgcaTCTCAACTTCCCCAAATGTCCG (SEQ ID NO: 110)
	Rev	cccaaggggttatgctagttattgcgTTAGCAAAATACGCCCACGAAGTC (SEQ ID NO: 111)
9	Fwd	gctaacttacattaattgcgttgcgcaGAGGATTTATGCACAAAATCCTGATGC (SEQ ID NO: 112)
	Rev	ccaaggggttatgctagttattgcgGATTTAATCTGTTGTTCCGAACGG (SEQ ID NO: 113)
10	Fwd	gctaacttacattaattgcgttgcgcaACCGTGCTGGCATGTTTTTAC (SEQ ID NO: 114)
	Rev	ccaaggggttatgctagttattgcgAGGAAGATCCGTGACCAGGAG (SEQ ID NO: 115)
11	Fwd	gctaacttacattaattgcgttgcgcaGAAATTATTTGGAATGGATGATGGCG (SEQ ID NO: 116)
	Rev	ccaaggggttatgctagttattgcgACTTCTACCTCCCTTTAGAAAAGTTAATG (SEQ ID NO: 117)
12	Fwd	gctaacttacattaattgcgttgcgcaCGGATTGAATCTGTTTATGAAATTTGGCTG (SEQ ID NO: 118)
	Rev	ccaaggggttatgctagttattgcgCCGACAGTTGTCACTGTTCTTATTACC (SEQ ID NO: 119)
13	Fwd	ccctgataaatgcttcaataatattgaaaaaggaagagtATGGCGGGTGCTTCAATAGAC (SEQ ID NO: 120)
	Rev	cccaaggggttatgctagttattgcgTTAGTTACTTGCTTTGTAGAATACCGTTAATGG (SEQ ID NO: 121)
14	Rev	cccaaggggdatgctagttattgcgTCAATCCGTAGCCTCTTCATTCTCG (SEQ ID NO: 122)
	Fwd	ataaatgcttcaataatattgaaaaaggaagagtATGGTAGCGATAAAAATGTATCCGGC (SEQ ID NO: 123)

TABLE 6-continued

PCR primers used to amplify genomic DNA source organisms containing validated defense systems.

#	Primer	Sequence
15	Fwd	gctaacttacattaattgcgttgcgcaACAATTTTTTGCCATAAGACGCTTTC (SEQ ID NO: 124)
	Rev	ccaaggggdatgctagdattgcgCATTAGGACTAGTAGAAAAGTCTTGGG (SEQ ID NO: 125)
16	Fwd	gctaacttacattaattgcgdgcgcaGGGATTTCCACCACCTCCC (SEQ ID NO: 126)
	Rev	ccaaggggdatgctagdattgcgTGCATAGCCAATGAAGATAAACGTG (SEQ ID NO: 127)
17	Fwd	gctaacttacattaattgcgdgcgcaGCGCAGCTGACAAAGATTGAC (SEQ ID NO: 128)
	Rev	ccaaggggdatgctagdattgcgCGATAATAAAAAGGCTCCAATCCCTG (SEQ ID NO: 129)
18	Fwd	gctaacttacattaattgcgdgcgcaACTAGCTAAGCAATAAGGGCG (SEQ ID NO: 130)
	Rev	ccaaggggdatgctagttattgcgCAATCTCCGAGGTGGCCC (SEQ ID NO: 131)
19	Fwd	gctaacttacattaattgcgdgcgcaTATTTTGCGTAGCTAGAACGCAATC (SEQ ID NO: 132)
	Rev	ccaaggggdatgctagdattgcgTGGGTATTAGCTCATATCAGAACTAATACCC (SEQ ID NO: 133)
20	Fwd	gctaacttacattaattgcgdgcgcaGTAAGACAAGGGTTGAGCAGGC (SEQ ID NO: 134)
	Rev	ccaaggggdatgctagdattgcgCAATGGTGGGCTGATTAATTAGATGAG (SEQ ID NO: 135)
21	Fwd	gctaacttacattaattgcgdgcgcaTAGCTATTGTGACTATGCTAACCATATG (SEQ ID NO: 136)
	Rev	ccaaggggdatgctagdattgcgTTCAGTCTAAATACATACCTGTCGGG (SEQ ID NO: 137)
22	Fwd	gctaacttacattaattgcgdgcgcaGTGCGCCTTATGTGATTACAACG (SEQ ID NO: 138)
	Rev	ccaaggggdatgctagdattgcgCTCTCAGCCTAATGATTCCAGAATAG (SEQ ID NO: 139)
23	Fwd	gctaacttacattaattgcgdgcgcaCGTGATGAATGAAGCGGCTAAATAC (SEQ ID NO: 140)
	Rev	ccaaggggdatgctagdattgcgGTAAATCCTCGGGAAAACACAGG (SEQ ID NO: 141)
24	Fwd	gctaacttacattaattgcgdgcgcaGATGGACTGGTACTGTAGATTCACC (SEQ ID NO: 142)
	Rev	ccaaggggdatgctagdattgcgCAAAGACGCAGAGGCCATCAG (SEQ ID NO: 143)
25	Fwd	gctaacttacattaattgcgdgcgcaGGGCTGTTTGGTTGAATTAAAAATACG (SEQ ID NO: 144)
	Rev	ccaaggggdatgctagdattgcgCCTTGATTTAAAACTATCAGTAGTAGGAACG (SEQ ID NO: 145)
26	Fwd	gctaacttacattaattgcgdgcgcaATAGAACGATGAAGGATGGAAGCTAC (SEQ ID NO: 146)
	Rev	ccaaggggdatgctagttattgcgTTGTATTTTGTTGTGTATGGGCGG (SEQ ID NO: 147)
27	Fwd	gctaacttacattaattgcgdgcgcaCGTGATTCAGTTCGCCAGAC (SEQ ID NO: 148)
	Rev	ccaaggggdatgctagdattgcgCACTCGAAATGGATACCCTGAG (SEQ ID NO: 149)
28		Synthesized
29		Synthesized

TABLE 7

Predicted protein domains within validated defense systems. Transmembrane helices were identified using TMHMM, and all other domains were identified using HHpred.

ID	Gene	Domain	Representative HHpred Hit	Probability	Start	End	Residues
BREX type I	A	DUF1819	PF08849.11	100	6	189	201
	B	DUF1788	PF08747.11	100	65	187	200
	C	ATPase	PF07693.14	96.66	43	348	1213
	C	DUF499	PF04465.12	99.88	247	846	1213
	D	Methyltransferase	PF02384.16	99.7	210	622	1201
	E	PglZ	PF08665.12	99.12	474	650	865
	F	Lon protease	PF13337.6	100	30	484	694
	F	Lon protease	PF05362.13	99.9	486	693	694
Druantia type I	A	DUF4338	PF14236.6	99.92	45	339	404
	B	CoiA	PF06054.11	99.77	1	182	548
	C	Macoilin	PF09726.9	96.72	167	323	627
	E	Helicase	PF00270.29	98.45	99	388	1836
	E	Helicase	5V9X_A	97.55	1071	1208	1836
	E	DUF1998	PF09369.10	98.92	1626	1710	1836
RT-Abi-P2	A	RT	PF00078.27	99.09	68	291	515
1	A	RT	PF00078.27	99.43	105	309	542
	A	TIR	PF13676.6	97.91	411	536	542
2	A	RT	PF00078.27	99.45	48	262	586
	A	TOPRIM	cd01026	96.88	367	465	586
3	A	Nuc_deoxy	PF15891.5	96.04	29	128	307
	B	RT	PF00078.27	99.52	53	248	320
4	A	RT	PF00078.27	99.63	54	328	425
5	A	RT	PF00078.27	99.12	67	296	540
6	A	RT	PF00078.27	99.14	59	263	494
7	A	RT	PF00078.27	99.06	80	382	1232
	A	Nitrilase	PF00795.22	98.89	953	1216	1232
	B	Transmembrane	—	—	4	26	144
8	A	RT	PF00078.27	99.39	53	251	398
	B	RT	PF00078.27	98.96	63	323	667
9	A	ATPase	PF07693.14	99.6	33	364	851
	B	Adenosine deaminase	PF00962.22	99.52	166	831	856
10	A	ATPase	PF07693.14	97.62	39	390	1273
	A	Transmembrane	—	—	160	177	1273
	A	Transmembrane	—	—	199	218	1273
11	A	ATPase	PF07693.14	99.8	15	385	643
	C	QueC	PF06508.13	99.67	150	369	457
	D	TatD DNase	PF01026.21	99.94	13	254	263
12	A	DUF4011	PF13195.6	99.81	33	308	1911
	A	ATPase	PF13086.6	97.93	427	552	1911
	A	Helicase	PF01443.18	97.82	1379	1636	1911
	A	Endonuclease	PF18741.1	98.7	1683	1780	1911
13	A	GHKL ATPase	5V44_A	99.46	1	241	2511
	A	GHKL ATPase	5V44_A	99.03	1544	1756	2511
	B	Helicase	6BOG_B	100	1	873	893
14	A	MBL-fold hydrolase	PF00753.27	98.79	8	324	386
	B	Protease	PF02122.15	98.23	2	187	1935
	B	ATPase	PF14516.6	99.36	204	535	1935
15	A	DUF4297	PF14130.6	98.41	8	223	2092
	A	ATPase	PF14516.6	99.44	250	597	2092
16	A	ATPase	PF14516.6	98.93	316	643	1484
17	A	Mrr	PF13156.6	97.05	17	162	1587
	A	ATPase	PF14516.6	99.07	204	476	1587
18	A	SIR2	cd00296	99.26	22	244	769
	A	ATPase	PF14516.6	97.6	312	464	769
19	A	SIR2	cd00296	99.44	21	253	1275
	A	DUF4020	PF13212.6	98.39	1114	1268	1275
20	A	SIR2	cd00296	99.47	21	240	1207
21	A	SIR2	cd00296	99.59	26	338	415
	B	HerA helicase	4D2I_B	100	10	608	610
22	A	DUF4297	PF14130.6	99.05	1	191	394
	B	HerA helicase	4D2I_B	100	7	568	571
23	A	VWA	PF00092.28	98.93	14	203	277
	B	Phosphatase	PF00481.21	99.74	5	232	239
	C	Kinase	PF00069.25	100	34	296	561
	C	ssDNA-binding	PF01336.25	96.18	344	435	561
24	A	PHP	cd07436	99.36	4	238	891
	A	ATPase	PF13166.6	99.74	266	836	891
25	A	DUF1887	PF09002.11	92.5	1105	1272	1272
26	A	ATPase	PF13654.6	97.36	5	349	384
	B	Protease	PF00082.22	99.87	264	561	754

TABLE 7-continued

Predicted protein domains within validated defense systems. Transmembrane helices were identified using TMHMM, and all other domains were identified using HHpred.

ID	Gene	Domain	Representative HHpred Hit	Probability	Start	End	Residues
27	A	ATPase	PF07693.14	96.47	49	312	1022
	A	DUF499	PF04465.12	100	79	745	1022
	B	DUF3780	PF12635.7	100	1	187	195
	C	DUF1156	PF06634.12	99	18	81	945
	C	Methyltransferase	PF01555.18	96.08	150	202	945
	C	Methyltransferase	PF01555.18	97.76	548	682	945
	D	PLD	cd09179	99.17	4	177	907
	D	Helicase	6BOG_B	100	218	865	907
28	A	RT	PF00078.27	99.35	136	351	613
	B	DNA PolA	2KFZ_A	100	31	515	515
29	A	RT	PF00078.27	99.37	34	241	311
	B	ATPase	PF13175.6	99.8	64	432	550
	C	HNH	PF01844.23	97.57	43	85	216

TABLE 8

Amino acid sequences of validated defense systems.

#	Gene	Sequence
BREX type I	A	MIKNDKAWIGLLGGPLMSRESRVIAELLLTDPDEQTWQEQIVGHNILQASSPNTAKRYAATIRLRLNTLDKSAWTLIAEGSERERQQLLFVALMLHSPVVKDFLAEVVNDLRRQFKEKLPGNSWNEFVNSQVRHLPVLASYSDSSIAKMGNNLVKALAEAGYVDTPRRRNLQAVYLLPETQAVLQRLGQQDLISILEGKR (SEQ ID NO: 150)
	B	MIDPVLEYRLSQIQSRINEDRFLKNNGSGNEIGFWIFDYPAQCELQVREHLKYLLRHLEKDHKFACLNVFQIIIDMLNERGLFERVCQQEVKVGTETLKKQLAGPLNQKKIADFIAKKVDLAAQDFVILTGMGNAWPLVRGHELMSALQDVMGFTPLLMFYPGTYSGYNLSPLTDTGSQNYYRAFRLVPDTGPAATLNPQ* (SEQ ID NO: 151)
	C	MNIEQIFEKPLKRNINGVVKAEQTDDASAYIELDEYVITRELENHRHFFESYVPATGEPRIRMENKIGVWVSGFFGSGKSHFIKILSYLLSNRKVTHNGTERNAYSFFEDKIKDALFLADINKAVHYPTEVILFNIDSRANVDDKEDAILKVFLKVFNERIGYCADFPHIAHLERELDKRGQYETFKAAFADINGSRWEDERDAYYFISDDMAQALSQATQQSLESSRQWVEQLDKNFPLDINNFCQWVKEWLDDNGKNILFMVDEVGQFIGKNTQMMLKLQTITENLGVICGGRAWVIVTSQADINAAIGGMSSRDGQDFSKIQGRFSTRLQLSSSNTSEVIQKRLLVKTDEAKAALAKVWQEKADILRNQLAFDTTTTTALRPFTSEEEFVDNYPFVPWHYQILQKVFESIRTKGAAGKQLAMGERSQLEAFQTAAQQISAQGLDSLVPFWRFYAAIESFLEPAVSRTITQACQNGILDEFDGNLLKTLFLIRYVETLKSTLDNLVTLSIDRIDADKVELRRRVEKSLNTLERLMLIARVEDKYVFLTNEEKEIENEIRNVDVDFSAINKKLASIIFDDILKSRKYRYPANKQDFDISRFLNGHPLDGAVLNDLVVKILTPKDPTYSFYNSDATCRPYTSEGDGCILIRLPEEGRTWSDIDLVVQTEKFLKDNAGQRPEQATLLSEKARENSNREKLLRVQLESLLAEADVWAIGERLPKKSSTPSNIVDEACRYVIENTFGKLKMLRPFNGDISREIHALLTVENDTELDLGNLEESNPDAMREVETWISMNIEYNKPVYLRDILNHFARRPYGWPEDEVKLLVARLACKGKFSFSQQNNNVERKQAWELFNNSRRHSELRLHKVRRHDEAQVRKAAQTMADIAQQPFNEREEPALVEHIRQVFEEWKQELNVFRAKAEGGNNPGKNEIESGLRLLNAILNEKEDFALIEKVSSLKDELLDFSEDREDLVDFYRKQFATWQKLGAALNGSFKSNRSALEKDAAAVKALGELESIWQMPEPYKHLNRITPLIEQVQVNHQLVEQHRQHALERIDARIEESRQRLLEAHATSELQNSVLLPMQKARKRAEVSQSIPEILAEQQETKALQMDADKKINLWIDELRKKQEAQLRAANEAKRAADSEQTYVVVEKTVIQPVPKKTHLVNVASEMRNATGGEVLETTEQVEKALDTLRTTLLAVIKAGDRIRLQ* (SEQ ID NO: 152)
	D	MNTNNIKKYAPQARNDFRDAVQIKLTTLGIAADKKGNLQIAEAETIGETVRYGQFDYPLSTLPRRERLVKRAREQGFEVLVEHCAYTWFNRLCAIRYMELHGYLEHGFRMLSHPETPTAFEVLDHVPEVAEALLPENKAQLVEMKLSGNQDEALYRELLLGQCHALHHAMPFLFEAVDDEAELLLPDNLTRTDSILRGLVDDIPEEDWEQVEVIGWLYQFYISEKKDAVIGKVVKSEDIPAATQLFTPNWIVQYLVQNSVGRQWLQTYPDSPLKDKMEYYIEPAEQTPEVQAQLAAITPASIEPESIKVLDPACGSDHILIEAYNVLKNIYEEGYRGRDIPQLILENNIFGLDIDDRAAQLSGFALLMMARQDDRRIFTRDVRLNIVSLQESLHLDIAKLWQQLNFHQQVQTGSMGDMFAENNALTQTDSAEYQLLMRTLKRFVNAKTLGSLIQVPQEEEAELKVFLDALYREQEGDFQQKTAAKAFIPFIQQAWILAQRYDAVVANPPYMGGNYMETELKNFVSSYYPQGKADLYSSFMVRLLLQLKDNRTLSLMTPFTWMNLSSFEELRKIILTNFSIQSLVQPEYHSFFESAYVPICAFSISNTPLSWNAKFFDLSDFYGEKNQAPNFQYAIKNDNKCHWKYNRITTDFLTPGYIIAYSLPDSALSCFKTSKKLHDVCNLKQGLITGDNERYLRFSHESIYNSFSLNEKRKKTKWFPYQKGGAYRKWYGNNDYVVDWENDGYSIKNFYNDKGKLRSRPQNIQFYCKEGLTWTSLTISSLSMRYVPNGYIFDAKGPMCPKSSLDIWNILGYANSKVIDIFLKQLAPTMDYSQGPVGNVPFKFNDGDLNEIIKELVNIHKRDWDENETSFEFKRDMLVHFSRDINTIKGSFTLRQGENKKAINRTKFLEEMNNSFFINCFNLTDILSPEIELNKITLTHATIEIDIQKIISYAIGCQMGRYSLDREGLVYAHEGNNGFADLVAEGAYKSFPADSDGILPLMDEEWFDDDVTSRVKEFIRTVWGEEYLRENLDFIAEVLKPKKGESALEITIRRYLSTQFWKDHLKMYKKRPIYWLFSSGKEKAFECLVYLHRYNDATLSRMRTEYVVPLLARYQANIDRLNDQLDEASGGESTRLKRERDSLIKKFSELRSYDDRLRHYADMRISIDLDDGVKVNYGKFGDLLADVKAITGNAPEVI* (SEQ ID NO: 153)

TABLE 8-continued

Amino acid sequences of validated defense systems.

#	Gene	Sequence
	E	MQNQDFIAGLKAKFAEHRIVFWHDPDKRFIEELEQLKLESVTLINMTHESQLAVKKRIEIDEPEQQFLLWFPHDAPPHEQDWLLDIRLYSSEFHADFAAITLNTLGIPQLGLREHIQRRKAFFSTKRTQALKNLATEQEDEASLDKKMIAVIAGAKTAKTEDILFNLITYQYVNQQIEDDSELENTQAMLKRHGLDSVLWEMLNHEMGYQAEEPSLENLLLKLFCTDLSAQADPQQRAWLEKNVLLTPSGRASALAFMVTWRADRRYKEAYDYCAQQMQAALHPEDHYRLSSPYDLHECETTLSIEQTIIHALVTQLLEESTTLDREAFKKLLSERQSKYWCQTQPEYYAIYDALRQAERLLNLRNRHIDGFHYQDSATFWKAYCEELFRFDQAYRLFNEYALLVHSKGAMILKSLDDYIEALYSNWYLAELSRNWNEVLEAENEMQAWQIPGVPRQQNFFNEVVKPQFQNPQIKRVFVIISDALRYEVAEELGNQINTEKRFTAELRSQLGVLPSYTQLGMAALLPHEQLCYQPGNGDIVYADGLSTSGIPNRDTILKNYKGMAIKSKDLLELKNQEGRDLIRDYEVVYIWHNTIDATGDTASTEDKTFEACRTAVAELKDLVTKVINRLHGTRIFVTADHGFLFQQQALSVQDKTTLQIKPENTIKNHKRFIIGHQLPADDFCWKGKVADTAGVSDNSEFLIPKGQIRFFSGGARFVHGGTMLQEVCVPVLQIKALQKTAAEKQPQRRPVDIVAYHPMIKLVNNIDKVSLLQTHPVGELYERPRILNIYIVDNANNVVSGKERISFDSDNNTMEKRVREVTLKLIGANFNRRNEYWLILEDAQTETGYQKYPVIIDLAFQDDFF* (SEQ ID NO: 154)
	F	MQTHHDLPVSGVSAGEIASEGYDLDALLNQHFAGRVVRKDLTKQLKEGANVPVYVLEYLLGMYCASDDDDVVEVQGLQNVKRILADNYVRPDEAEKVKSLIRERGSYKIIDKVSVKLNQKKDVYEAQLSNLGIKDALVPSQMVKDNEKLLTGGIWCMITVNYFFEEGQKTSPFSLMTLKPIQMPNMDMEEVFDARKHFNRDQWIDVLLRSVGMEPANIEQRTKWHLITRMIPFVENNYNVCELGPRGTGKSHVYKECSPNSLLVSGGQTTVANLFYNMASRQIGLVGMWDVVAFDEVAGITFKDKDGVQIMKDYMASGSFSRGRDSIEGKASMVFVGNINQSVETLVKTSHLLAPFPTAMIDTAFFDRFHAYIPGWEIPKMRPEFFTNRYGLITDYLAEYMREMRKRSFSDAIDKFFKLGNNLNQRDVIAVRRTVSGLLKLMHPDGAYSKEDVRVCLTYAMEVRRRVKEQLKKLGGLEFFDVNFSYIDNETLEEFFVSVPEQGGSELIPAGMPKPGVVHLVTQAESGMTGLYRFETQMTAGNGKHSVSGLGSNTSAKEAIRVGFDYFKGNLNRVSAAAKFSDHEYHLHVVELHNTGPSTATSLAALIALCSILLAKPVQEMQMVVLGSMTLGGVINPVQDLAASLQLAFDSGAKRVLLPMSSAMDIPTVPAELFTKFQVSFYSDPVDAVYKALGVN* (SEQ ID NO: 155)
Druantia type I	A	MHKYPSIIVNINLREAKLKKKVREHLQSLGFTRSDSGALQAPGNTKDVIRALHSSQRAERIFANQKFITLRAAKLIKFFASGNEVIPDKISPVLERVKSGTWQGDLFRLAALTWSVPVSSGFGRRLRYLVWDESNGKLIGLIAIGDPVFNLAVRDNLIGWDTHARSSRLVNLMDAYVLGALPPYNALLGGKLIACLLRSRDLYDDFAKVYGDTVGVISQKKKQARLLAITTTSSMGRSSVYNRLKLDGIQYLKSIGYTGGWGHFHIPDSLFIELRDYLRDMDHAYADHYMFGNGPNWRLRTTKAALNALGFRDNLMKHGIQREVFISQLAENATSILQTGKGEPDLTSLLSAKEIAECAMARWMVPRSIRNPEYRLWKARDLFDFISNDSLNFPPFDEIAKTVV* (SEQ ID NO: 156)
	B	MNYAIDKFTGTLELAARATKYAQYVCPVCKKGVNLRKGKVIPPYFAHLPGHGTSDCENFVPGNSIIVETIKTISKRYMDLRLLIPVGSNSREWSLELVLPTCNLCRAKITLDVGGRSQTLDMRSMVKSRQIGAELSVKSYRIVSYSGEPDPKFVTEVERECPGLPSEGAAVFTALGRGASKGFPRAQELRCTETFAFLWRHPVAPDFPDELEIKSLASKQGWNLALVTIPEVPSVESISWLKSFTYLPVVPARTSITAIWPFLNQKTSINHVECVYSDTILLSTNMAPTSSENVGPTMYAQGSSLLLSAVGVETSPAFFILNPGENDFVGVSGSIEQDVNLFFSFYKKNVSVPRKYPSIDLVFTKRNKEKTIVSLHQRRCIEVMMEARMFGHKLEYMSMPSGVEGVARIQRQTESNVIKLVSNDDIAAHDKSMRLLSPVALSQLSDCLANLTCHVEIDFLGLGKIFLPGSSMLSLDDGKFIELSPNLRSRILSFILQMGHTLHGFSLNNDFLLVEKLVDLQPEPHLLPHYRALVKEVKTNGFECNRFR* (SEQ ID NO: 157)
	C	MSYQYSQEAKERISKLGQSEIVNFINEISPTLRRKAFGCLPKVPGFRAGHPTEIKEKQKRLIGYMFQSHPSSEERKAWKSFSLFWQFWAEEKIDKSFSMIDNLGLKENSGSIFIRELAKNFPKVARENIERLFIFSGFADDPDVINAFNLFPPAVVLARDIVIDTLPRILDELEARISLIADNVEKKNNHIKELELKIDAFSEQFDNYFNNEKSSLKIINELQSLINSETKQSDIANKAIDELYHFNEKNKQLILSLQEKLDFNALAMNDISEHEKLIKSMANDISEFKNALTILCDNKIKNNELDYVNELKKLTERIDTLEINTSQASEVSVTNRFTKFHEIAHYENYEYLSSSEDISNRISLNLQAVGLTKNSAEKLARLTLATFVSGQIIQFSGSLADIIADAIAIAGAPRYHIWRVPVGIISDMDAFDFIETIAESSRCLLLKGANLSAFEIYGAAIRDIVVQRQIHPTNYDHLALIATWKQGPATFPDGGMLAELGPVIDTDTLKMRGLSATLPQLKPGCLAKDKWTNIDGLHLDSVDDYVDELRALLDEAGFDGGTLWKRMIHIFYTSLIRIPNGNYIYDLYSVLSFYTLTWAKIKGGPVQKIEDIANRELKNYSAKISS*(SEQ ID NO: 158)
	D	MEWRAVSRDKALDMLSTALNCRFDDEGLRISAVSECLRSVLYQYSISETEEARQTVTSLRLTSAVRRKLVPLWPDIADIDNAIHPGIMSILNSLAELGDMIKLEGGNWLTAPPHAVRIDNKMAVFFGGEPSCTFSTGVVAKSAGRVRLVEEKVCTGSVEIWDANEWIGAPAEGNEEWSSRLLSGTISGFIDAPGNMSETTAYVRGKWLHLSELSFNKKQIYLCRMSVDNHFSYYLGEIEAGRLCRMNSLESSDDVRRLRFFLDTKCNCPLKVRIKISNGLARLRLTRRLPRRETKVLLLGWRESGFENEHSGITHHVFPEEILPIVRSAFEGLGIIWINEFTRRNEI* (SEQ ID NO: 159)
	E	MINKNKVTERSGIHDTVKSLSENLRKYIEAQYHIRDEGLIAERRALLQQNETIAQAPYIEATPIYEPGAPYSELPIPEAASNVLTQLSELGIGLYQRPYKHQSQALESFLGENASDLVIATGTGSGKTESFLMPIIGKLAIESSERPKSASLPGCRAILLYPMNALVNDQLARIRRLFGDSEASKILRSGRCAPVRFGAYTGRTPYPGRRSSRRDELFIKPLFDEYNKLANNAPVRAELNRIGRWPSKDLDAFYGQSASQAKTVYSGKKTGKQFVLNNWGERLITQPEDRELMTRHEIQNRCPELLITNYSMLEYMLMRPEIRNIFEQTKEWLKADEMNELILVLDEAHMYRGAGGAEVALLIRRLCARLDIPRERMRCILTSASLGSIEDGERFAQDLTGLSPTSSRKFRIIEGTRESRPESQIVTSKEANALAEFDLNSFQCVAEDLESAYAAIESLAERMGWQKPMIKDHSTLRNWLFDNLTGFGPIEITLIEIVSGKAVKLNILSENLFPDSPQQIAERATDALLALGCYAQRADGRVLIPTRMHLFYRGLPGLYACIDPDCNQRLGNHSGPTILGRLYTKPLDQCKCASKGRVYELFTHRDCGAAFIRGYVSSEMDFWHQPNGPLSEDEDIDLVPIDILVEETPHVHSDYQDRWLHIATGRLSKQCQDEDSGYRKVF

TABLE 8-continued

Amino acid sequences of validated defense systems.

#	Gene	Sequence
		IPDRVKSGSEITFDECPVCMRKTRSAQNEPSKIMDHVTKGEAPFTTLWTQISHQPASRPIDGKHPNGGKKVLIFSDGRQKAARLARDIPRDIELDLFRQSIALACSKLKDINREPKPTSVLYLAFLSVLSEHDLLIFDGEDSRKVVMARDEFYRDYNSDLAQAFDDSFSPQESPSRYKIALLKLLCSNYYSLSGTTVGFVEPSQLKSKKMWEDVQSKKLNIESKDVHALAVAWIDTLLTEFAFDESIDSTLRIKAAGFYKPTWGSQGRFGKALRKTLIQYPAMGELYVEVLEEIFRTHLTLGKDGVYFLAPNALRLKIDLLHVWKQCNDCTALMPFALEHSTCLACGSNSVKTVEPSESSYINARKGFWRSPVEEVLVSNSRLLNLSVEEHTAQLSHRDRASVHATTELYELRFQDVLINDNDKPIDVLSCTTTMEVGVDIGSLVAVALRNVPPQRENYQQRAGRAGRRGASVSTVVTYSQNGPHDSYYFLNPERIVAGSPRTPEVKVNNPKTARRHVHSFLVQTFFHELMEQGIYNPAEKTAILEKALGTTRDFFHGAKDTGLNLDSFNNWVKNRILSTNGDLRTSVAAWLPPVLETGGLSASDWFAKVAEEFLNTLHGLAEIVPQTAVLVDEENEDDEQTSGGMKFAQEELLEFLFYHGLLPSYAFPTSLCSFLVEKIVKNIRGSFEVRTVQQPQQSISQALSEYAPGRLIVIDRKTYRSGGVFSNALKGELNRARKLFNNPKKFIHCDKCSFVRDPHNNQNSENTCPICGGILKVEIMIQPEWGPENAKELNEDDREQEITYVTAAQYPQPVDPEDFKFNNGGAHIVFTHAIDQKLVTVNRGKNEGESSGFSVCCECGAASVYDSYSPAKGAHERPYKYIATKETPRLCSGEYKRVFLGHDFRTDLLLLRITVGSPLVTDTSNAIVLRMYEDALYTIAEALRLAASRHKQLDLDPAEFGSGFRILPTIEEDTQALDLFLYDTLSGGAGYAEVAAANLDDILTATLALLESCECDTSCTDCLNHFHNQHIQSRLDRKLGASLLRYALYGMVPRCASPDIQVEKLSQLRASLELDGFQCIIKGTQEAPMIVSLNDRSIAVGSYPGLIDRPDFQHDVYKSKHTNAHIAFNEYLLRSNLPQSHQNIRKMLR* (SEQ ID NO: 160)
RT-Abi-P2	A	MKKVYELTSEEALSYFLRHDSYTTLELPAYINFTTLLNDINSSIHNKKIKIEPTAKELMGKDINYEVLVSKDGLYSWRRITLINPLYYVYFCRKITAPATWEIITEKFKSFESNDLFTCSSIPVRKDNSSNIAASVMNWWEDFEQKSLALALEYEFWSTDISNFYPSIYTHSFEWVFISKEEAKKKKSKNNPGGLIDSHIQMMMNNQTNGIPLGSTLMDTFAELILGQIDIELRKKTNELKIINYKWRYRDDYRIFSNSKDDLDIISKCLVNVLGDFGLDLNSKKTELYEDIILHSLKQAKKDYIKEKRHKSLQKMLYSIYLFSLKHPNSKTTVRYLNDFLRNLFKRKTIKDNGQQVDAMLGIISSIMAKNPTTYPVGTAIFSKLLSFLYGDDTQKKLTKLEQLHKKLDKQPNTEMLDIWFQRTQAKINLEWNKSYKSALCVRINDELTKEKTFSVNNLWNIDWIQGKETSPNKAKILSLLRKTKIVDTDKFDKMDDNITPEEVNLFFKEHSN* (SEQ ID NO: 161)
1	A	MSLHDKLLMHNFALANKKSPDFISELPQIEPKPYSNGHKIKWINHTLTSTEVTPPDNLIKICILIESGEIAITSVSDIANLLGVPAGQLLYILYRKKDNYRTFEIEKKNGKKRVINAPCGGLSILQTRLKPVLEYFYRPKKSAHGFDCGKSIITNAGMHIKKNFWNIDLENYFESISFARVYGIFKSKPFNFAHPAATVLAQLCTHNGKLPQGACTSPILAMASASLDKQLTQFAGRKKISYSRYADDITFSFNQRNIDIIKKNDDGSYSLSETIDNIISKNGFKINYDKFRVQTRNTRQSVTGLWNDKVNINRRYIRITRSMIHRWTDDKLKYALLFATEKGYQAKDNNHAIQIFRNHIYGRLSFIKMVRGKDYPGYLKLMSYMSHNDPLKTQEGLRAMKETENFDVFICHASEDKKDIAIPIYDELTKLKISAFIDHVEIKWGDSLIDKINAALVKSKYVIAILSANSVNKEWPQKELRAVLASEISSGDVKLLTLLKKEDEEVVNLSLPLLSDKFYMVYDNNPEVVANNIKSLLQR* (SEQ ID NO: 162)
2	A	MTKTSKLDALRAATSREDLAKILDVKLVFLTNVLYRIGSDNQYTQFTIPKKGKGVRTISAPTDRLKDIQRRICDLLSDCRDEIFAIRKISNNYSFGFERGKSIILNAYKHRGKQIILNIDLKDFFESFNFGRVRGYFLSNQDFLLNPVVATTLAKAACYNGTLPQGSPCSPIISNLICNIMDMRLAKLAKKYGCTYSRYADDITISTNKNTFPLEMATVQPEGVVLGKVLVKEIENSGFEINDSKTRLTYKTSRQEVTGLTVNRIVNIDRCYYKKTRALAHALYRTGEYKVPDENGVLVSGGLDKLEGMFGFIDQVDKFNNIKKKLNKQPDRYVLTNATLHGFKLKLNAREKAYSKFIYYKFFHGNTCPTIITEGKTDRIYLKAALHSLETSYPELFREKTDSKKKEINLNIFKSNEKTKYFLDLSGGTADLKKFVERYKNNYASYYGSVPKQPVIMVLDNDTGPSDLLNFLRNKVKSCPDDVTEMRKMKYIHWYNLYIVLTPLSPSGEQTSMEDLFPKDILDIKIDGKKFNKNNDGDSKTEYGKHIFSMRVVVDKKRKIDFKAFCCIFDAIKDIKEHYKLMLNS* (SEQ ID NO: 163)
3	A	MNKKFTDEQQQQLIGHLTKKGFYRGANIKITIFLCGGDVANHQSWRHQLSQFLAKFSDVDIFYPEDLFDDLLAGQGQHSLLSLENILAEAVDVIILFPESPGSFTELGAFSNNENLRRKLICIQDAKFKSKRSFINYGPVRLLRKFNSKSVLRCSSNELKEMCDSSIDVARKLRLYKKLMASIKKVRKENKVSKDIGNILYAERFLLPCIYLLDSVNYRTLCELAFKAIKQDDVLSKIIVRSVVSRLINERKILQMTDGYQVTALGASYVRSVFDRKTLDRLRLEIMNFENRRKSTFNYDKIPYAHP* (SEQ ID NO: 164)
	B	MKSAEYLNTFRLRNLGLPVMNNLHDMSKATRISVETLRLLTYTADFRYRIYTVEKKGPEKRMRTIYQPSRELKALQGWVLRNILDKLSSSPFSIGFEKHQSILNNATPHIGANFILNIDLEDFFPSLTANKVFGVFHSLGYNRLISSVLTKICCYKNLLPQGAPSSPKLANLICSKLDYRIQGYAGSRGLIYTRYADDLTLSAQSMKKVVKARDFLFSHPSEGLVINSKKTCISGPRSQRKVTGLVISQEKVGIGREKYKEIRAKIHHIFCGKSSEIEHVRGWLSFILSVDSKSHRRLITYISKLEKKYGKNPLNKAKT* (SEQ ID NO: 165)
4	A	MNNDDYPWFRKRGYLHFDEPVSLKKAVKYVSSPEKIIKHSFLPFLSFEVKSFKIKKDKSTKQLSKTEKLRPIAYSSHLDSHIYAFYAEYLTGHYELLIQENNLHENILAFRSLNKSNIEFAKRAFDTITEMGECSAVALDLSGFFDNLDHQILKHQWCKVIGTEALPQDHFAIYKSITRYSKVDKNRAYEILGISKNNPKYNRRKICTPVDFRNKIRKNGLITVNNSQKGIPQGSPTSALLSNIYMLDFDTEMRDYAQERGGHYYRYCDDMLFIVPTKYNKTLAGDVAQRIKHLKVELNTKKTEIRDFIYKDSTLVANMPLQYLGFIFDGSNILLRSSSLARYSERMKRGVRLAKATMDSKNRIRENKGEALKALFKKKLYARYSHIGRRNFLTYGYRAAKIMNSKAIKRQLKPLQKRLENEILK* (SEQ ID NO: 166)

TABLE 8-continued

Amino acid sequences of validated defense systems.

#	Gene	Sequence
5	A	MVIFDEKRHLYEALLRHNYFPNQKGSISEIPPCFSSRTFTPEIAELISSDTSGRRSLQGYDCVEYYATRYNNFPRTLSIIHPKAYSKLAKHIHDNWEEIRFIKENENSMIKPDMHADGRIIIMNYEDAETKTIRELNDGFGRRFKVNADISGCFTNIYSHSTPWAVIGVNNAKIALNTKVKNQDKHWSDKLDYFQRQAKRNETHGVPIGPATSSIVCEIILSAVDKRLRDDGFLFRRYEDDYTCYCKTHDDAKEFLHLLGMELSKYKLSLNLHKTKITNLPGTLNDNWVSLLNVNSPTKKRFTDQDLNKLSSSEVINFLDYAVQLNTQVGGGSILKYAISLVINNLDEYTITQVYDYLLNLSWHYPMLIPYLGVLIEHVYLDDGDEYKNKFNEILSMCAENKCSDGMAWTLYFCIKNNIDIDDDVIEKIICFGDCLSLCLLDSSDIYEEKINNFVSDIIKLDYEYDIDRYWLLFYQRFFKDKAPSPYNDKCFDTMKGYGVDFMPDENYKTKAESYCHVVNNPFLEDGDEIVSFNDYMAIA* (SEQ ID NO: 167)
6	A	MTSTIDFYESDFSATLYPLKTNQILLKHHSQEMSEYIYQKVINPAYPTDSFLSQQKVFSTKPKGHLRRTVKLDPVAEYFTYDVTYRNRKIFRPEVSESRKSFGYIFRNGSRIPIHVSYNEYKQSLKKYSELYSHSIHFDIASYFNSLYHHDIIHWFSSKEGVSPADVEALGQFFREINSGRSIDFMPQGIYPAKMIGNEFLKFVDLHGRLKSAQIVRFMDDFTIFDNDIETLNNDFIRIQQLLGQVSLNINPSKTTFDNVMGDVNETLTQIKSSLKEIITEYEHIPTASGVEWETNIEIIKHLDDEQVNKLIDLLKDEKIEESDADLILGFLRTHNDSLLSQMPMLLGRFPNLIKHIYTICSGITDKSGLVKILLSYLNTNNNFLEYQLFWIGAIVEDYLLGVGEYGSVLHKLYELSGDFKIARAKVLEIPEQGFGFKEIRNEYLRTGQSDWLSWSSAIGTRNLKSAERNYILDYFSKGSPINYLVASCVKKL* (SEQ ID NO: 168)
7	A	MKLLDKKYYNLEPKYEYLKDSFILGLAWKKTDSFVRTHNWYADILELDKCAFDISDEVTNWSNEISKNALSKSDIELIPAPKGASWFINQGKWTTNKDNRKIRPLANISIRDQSFATAVTMCLADAIETRQKDCSLSNLGYAEHVKNKVVSYGNRLVCDWDNERARFRWGGSEYYRKFSSDYRSFLQRPIYIGRETVNKVSGIDDVYIISLDLKNFFGSIKINLLLEKIKKISADHYAAKFINDNEFWTLANRILSWDWPEESLSLLESLDKEKNVGLPQGLASAGALANAYLIEFDESLISKLRTKIEDSQIILHDYCRYVDDIRLVISGEALESNKIKESIHALVQGILDETLAQNPSDNEPYLKINDSKTYILELSDIDNGSGLTNRINEIQHEVGASSIPERNGLDNNIPALQQLLLTEQDNFSEDVDSLFPGFKNDKSIKVESVRRFSAHRLEKSLAKKSKLISPEERKQFDNETSLIAKKLLKAWLKDPSIMVIFRKAIAINPNLDAYSTILEIIFSRIQRNRDKRDKYIMLYLLSDIFRSVIDVYRNLESEYVDDYQKLMGEVTLFAQKILSCKSFIPNYAYQQALFYLAVINKPFIASNKASFDLARLQCVLIKQHLEPLNSSDGYLFEVSAQISKDYRANAAFLLSHTNSNKVVDLIIEKFAFRGGEFWNAIWKEIVRMQDKDRINEFRWAISKYESKPNSSEHYLSSVISFKENPFRYEHALLKLGVALVELFDDTEKNVWQPDGKQYSPHEIKVKLEGNSTSWGELWRPNFSISCSIDKKGEPGKDPRYISPEWLANYPQTQNDEQKIYWVCSVLRSAALGNVDYTQRNDLKLDKAKYDGIHSQFYKRRMGMLHTPESIVGSYGTITDWFASFLQHGLQWPGFSSSYISQEDILSITNIIEFKNCLLERLGYLNKQICISSNVPTLPTVVNRPELASNHFRIVTVQQLFPKDTNFHPSDVTLANPDVRWKHREHLAEICKLTEQTLNAKLKTESREHTSTADLIVFSELAVHPEDEDIVRALAFRTKAIIFSGFVFCEQDGRIVNKARWIIPDSSESGTQWRVRDQGKHHMTSDEVALGIQGYRPSQHIISIEGFIPEGPFKLTGAICYDATDIKLAADLRDLTDMFVIAAYNKDVDTFDNMASALQWHMYQHIVITNTGEYGGSTMQAPYKEKYHKLISHAHGTGQIAISTADIDLAAFRRKLQTYKKTKTQPAGYNRKH* (SEQ ID NO: 169)
	B	MDTLVKLATIISPLISAGVAIWAILVAKKTISESKEIAKKTIADTAYQAYLQLAMENPQFSKGYSADCRQERDPMYDQYVWYVARMIFCFEKIIEVEVNLKDSSWANTLEKHLKFHSEHFKKTNVVEEALYIPPILDLIRCAAN* (SEQ ID NO: 170)
8	A	MLNQSFSVSNLIKLLKKTDPKRYKTGRNSAEYKKYIADKVNGSIETYSFGSISNSRTNNKNVYIFKDFMDVLVARKINDNIKRVYSVKQNNRHDIIKKVNTVLSEPVNYYIYRLDIKSFYESIDKNIVFQRINNNPIISHNTKKFINGLFKHNAFSANNGLPRGMGLSATLSEIFMEEFDAELARLPEVFYASRYVDDIIVFSFYKIPDYKNYFSRILPNGLHLNERKCSEYTIEDTSTKHSEIEFLGYSFIIHHGLKNQRRHVVIRISEEKIKKIKRRIALAVKDYSMNSDAELLKKRIKYLTGNTLVNSNSNKTDALYSGIYYNYQHLTDKTQLKELDIFKNRMLFSSKGEVGRKILAAGHNLLTAPKKYSFLAGFEKRLLSSFKREDIIKINKVW* (SEQ ID NO: 171)
	B	MKIKISKSDYKRVLLTDILPYEVPILFSNEGFYKLISENKVLPGTFSEGLKLDSYTIPYSYKIKKGLASSRSLGIIHPSTQLRICDFYDKYEHLMVHMCTKSPFSLRYPSKIGSYYYEKDFLKSRINLKDGLVQFHNHGFDSQETSSSSHFSYKKYPFIYKFYESYEFHRLERKFRKLLKLDIAKCFSHIYTHSVSWAVKSKEFSKVNRTYNSFEGCLDKLFQDANYGETNGIIIGPEFSRIFAEIILQRVDLNVESHLNLEPGIVKDKSYATRRYVDDYFIFADDDETFKLTEFVLANELEKYKLYLNESKKEFIERPFVTGATMAKNDIAEIIEDLYGSLIHTEKLDELTAMVNLNPDVKIQPENMNDLFPLKGVWNKKLHADKFIKRIKIAVRKNNTTFDLVSSYLLSAIKSKFFKVIRLLRMFDLSGKEDITYKFFSIFNEVIFFIYAMDFRVRQTYIISQVILEINSFANKQASDISEVIKKNTFDELLMCMKSMGNIHERPVELSNLLICMKGLGEQYKLNPDEFKDLLGISENECFYDLEYFSTCSMLHYIGDDVLYLKMKEDIVLAIQSLISGRNDIKKDTETFMLFLDMMTCPYLTVKHKRIIYRTYVEANTGQKRFTNAVIDSEIDSLKNNVIFFNWSGDADLEHVLYKKELRTAYE* (SEQ ID NO: 172)
9	A	MTSEIVLNLDFPEYKDDFCTDSIDEQDNELWQQQANKKLLSFLEVMGEEARRYKENNSRSTHPHYKTLSSYHHAIFISGARGAGKTVFMRNARFSWQKHYNKDLKRPKLYFIDVTDPTLLNIDDRFSEVIIASIYATVEKRMKQPDIAQNIKDNFINSLKTLSGALGKSKDYDEYRGIDRIQKYRSGIHLEKYFHQFLISSVELLDCDALVLPIDDVDMKIDNAFGVLDDIRCLLSCPLVLPLVSGDNDLYRFIAKSKFEELLNRKANSNYAKEGSEIAERLSEAYITKVFPSHVKIPLQPIDELLPYLYIHSNEDENKQHTSYSEFIKLVQQKFYFLCNGQERSTNWPQPRSAREVTQLIRSLPPSTLSKEDDSGTDLWQRFAVWAEERRDGLALTNVESYLFIKNAKAVEDLNLSNLIAFNPLLQKGKYPWAEKDFYKQQSQRRKELNAPETNSGILNTVFSEQRKDFILRSMPALELIMEPMYVTKTVAEKND

TABLE 8-continued

Amino acid sequences of validated defense systems.

#	Gene	Sequence
		NSALIAIYTHSDYYSQQQNRRCHIFFGRAFEIMFWSVLAKTENLPQEFYEKDKFKSLFGNIFKKVPFYSIFSMNPTKVVDEENDDGSEPDFSQKLDDSINELVEDIYIWATSNKLRAFKNKNLIPLMTCVFNKVFSMNVLRKNVQDRVKFRDEHLSDLAKRFEYMFINAEFTFIREGVVVNTNVATGAAPARVRNLSEFNRYDKTLSRNMSGILSVKEDNGLTIVKESEGDIADLLFEIWHSPLFKLTTRTCYPIGKINSQNTAQENLSSDFNSFFENGINFELIKQYYWQTSNHDNIRTADVREWATSRLNEAIILFSWMKESKSIKAKIDGQSYEGRLFRGLQQALEGYEEV* (SEQ ID NO: 173)
	B	MFNQDPYWLIPTLCLASDRIFYAQLRDHLGQKSSGERKKEKNGYILVQAAQDYQFYFGGRIRKEDVQNNALMWQIETGNENCLSMLDSLSAYFLTWRGNCFEVRRERLEPWLMICSVIDPAWIIAYAYQQLIKQNVVCDSELISLLTEHQCPFAFPKGRGDISFADNHVHLNGHGYSSISMLNFIDGNYKVKKGIKWPYRQEYTLFESGLLDKNDLPRWLSAYSSCLLKNVYNSFQQGKRSEVDFTCLKDAVETVLADEDKYYFLEVASLYDVVTLQQRVLYEAAQQKYHSHQRWLLYTCGIMLGTESEDYANALANLIRISNILRNYMVVSAVGLGQFIDFFGFNYRRITKPADTNNRVHYDSSAGISREYRVSPDFVLGSGVMPDIYARQLFDFYCTQARKGVPEQGHIVVHFTRSFPDKKSTYDKLLTECRERLRSQCDYFGRFLTSLTLQSIEYKNLSTDEDRSIDIRKLVRGYDVAGNENELQIEVFAPVLRVLRAAKFKGEGVNFKRLQRPFITVHAGEDYCHILSGLRAMDEAVEFCMLGEGDRIGHGLALGVDIKLWANRQKRAYLTVGQHLDNLVWAYHQAVLLSQHIVEHIPVMHELRDKIHYWSHQLYSETYTPDLLFKAWLLRRNWPDYKSIISDPANINEWVPDQHILVSTDETTAKARKIWERYLNSGLAENDVFNRIISVNCAPDTAQNFSMTFNENEDILSKGELLLYEAIQDFLIEKYSRLGLVIEACPTSNIYIGRLEKYHEHPLFRWNPPDSQWIKPGGKFNRFGLRTGPLSVCINTDDSALMPTTIENEHRLMRDCAIHFYGIGTWMADLWINSIRIKGIEIFKGNHLSQDLDNLI* (SEQ ID NO: 174)
10	A	MIMSTPWLTPIVADSDHAEANAVSYEALTPTELDSDKAGCYISALNYAYEHPDIRNIAVTGPYGAGKSSVLKTWCKAHNGTLRVLTVSLADFDMQRHVDESNGDSSSDEGTKNTGSVEKSEYSELQQILYKNKKHELPCSRIDRISDVTAGQILRSASFLTGTILLSGAALFFLAPDYVTTKLSLPGAFARYLLECPFGVRVSGAVASVMGSLCLLLNQLHRIGIFDRKVSLDKVDLLKGAVTTRASSPSLLNVYIDEIVYFFDSTKYDWIFEDLDRFNNGRIFVKLREINQIINNCLSDRKPVKFIYAVRDGIFNSAESRTKFFDFVMPVIPVMDNQNAYEHFVKKFKEEEINNNLSECISRIATFIPNMRVMHNITNEFRLYQNLVNSRENLAKLLAMIAYKNLCAEDYHGIDSKKGVLYHFIQSYLDHEIQNELLHSANNELEDMAQSLVAITNEKLANRENLREELLMPYLSKNYSGALVFYTEGRQISLDDLIQDEDEFLMLLDKENIQVVTPYNRQNFLMINQRDTEKLKQQYEKRCHLIETKSVDNITRVKNNISSLESLRTEILSGTVADIAEKMTNEGFVAWIKKKEDTGVLTIQSEHEQIDFIFFLLSSGYLSTDYMSYRSIFIPGGLSETDNLFLKDVMSGKGPEKTFSFHLDNVNNIVERLKKLGVLQRDNAQHPAVIRWLIDNDPDTLKNNIMALLSQTGSQRVVSLLMLMQNDFTTYVRLRYLEIFMSDEHILNRLLAHLCASEERTPEQKFFVQEIAAHLLCLTEKSNIWQSVEINKRIGELIDSSPILITAVPKGYGDAFFEVLKDNTLSVSYIPGDVGDEKCSVIRKTAGAGLFKYSVSNLKNVYLCLTQDKNEERMSFSLYPFHCLESLAISELTEILWTNIEDFILSVFIESEEIDRIPELLNSSEVSMTVVEQIIAKMDFCINNLDDIINRSECADNNASGRNIYSMLLQHDRIFPSFDNIEHLLHDTSINTSGELVQWVNEKHFEFEPSDIVINDTGIFNNFISELICSPVISEEALLKVLSNLNVVIIDVPENIPLRNAELLCSEKKLAPTVNVFTVLFNALSENVDDINRMNTLLGNLIAQRPEIITQEPEDIFYIEGDFDEELASELFRHKLIGMNIKVAALRWLRDNKPGILDKSYLLSLDILAELSPWMGDDDLRLTLLKRCLVAGDAGKDALCVVLNSFADESYHGLLPHDRFRKIPHSVDLWEVAELISNLGFIQPPKMGSGRDEHKIVTTPVRYVRDVEFYD* (SEQ ID NO: 175)
11	A	MFLNDQETSTDLLYYTAIASTVVRLVDETSDAPITIGVHGDWGAGKSSVLKMLEAACEKKDKTHCIWFNGWTFEGFEDAKTVIIETIVEDLVASRPMSTKVAEAAKKVLRRIDWLKMAKKAGGLAFTAFTGIPTFDQIKGMYELASDFLSAPQDKLSAADFKAFAEKAGGFIKEADTDSNTLPKHIHAFREEFRALLDAAEEKLVVIVDDLDRCLPKTAIETLEAIRLFLFVEKTAFVIGADEAMIEYAVKDHFPDLPQSTGPVSYARNYLEKLIQVPFRIPALGTAETRIYTTLLLAENALGSEDDNFKALLNKAREEMKRPWISRGLDREAVMAALNGKIPEWENALLFSLHVTPMLSSGTHGNPRQIKRFLNSMMLRQAIADERGFGSDIKRPVLAKIMLAERFYPSVYGKLVQLVSNHPEGKPEALAEFEALVRGGKTAPKSRADSKENSSESEDVQNWLKIDWAIGWAKAEPALSGEDLRPYVFVTRDKHSTLSNLVVSSHLIPIMEKLLGPKIGMVKIKGDLEKLSPPDADELFEMLSDKLFQEDSFNRKPRGFDGLEYLVETQPHLQRRLIDFARRIPVKKAGGWLATRIAQSLVDPTLIEEYTKLIQEWASDENLSLSKSAKATLQLSGYQH* (SEQ ID NO: 176)
	B	MGTSKAYGGPVHGLIPDFVENPSPPTLPPVDPADDSTLDTPLIPPDSSGSGPLSTPKANFTRYSRSGSRSSLGKAVAGYVRNGVGGAGRASRRMGASRAAAGGLLGLISDYQQGGATQALERFNLGNLAGQSASTALLSLVEFLCPPGGSVDEGVARQAMLETIADMSDVGEENFDELTPDQLKEWIGFVVHSIEGRLMADIGKNGIKLPDDIDAIVSIQEDLHDFVDGATRTQLREELRNLTGLSGDAIDRKVEEIYTVAFELLAREGERLE* (SEQ ID NO: 177)
	C	MSHHTLVARLGTDDNSDLQLSRQSTHLTEINFLKENGKLDFGLGQALNGLSDLGLTPMDVSVDLALLAATVTAADTRISRGHNAQDLWTREIALYIPVASPTLWNSQTGLLSRMLNFLTGDRWTIHFRSRPVIEFIGLIQRSSKERSVNPTSVCLFSGGLDSFIGAIDLLSNGGTPLLISFIYWDTTTSVYQQKCAQLLSERYGQSFSHVRARVGFEKTTIEGEDGENTLRGRSFMFFSLATMAADALGGPVTINVPENGLISLNVPLDPLRVGALSTRTTHPFYMARFNELLGNLGISAHLENPYAYKTKGEMAIHCHDHAFLRQHAADTMSCSSPQSTRWNPALNEQQSTHCGRCVPCLIRRASLFTAFGTDDTIYRIPDLRSRVLDSSKPEGEHVRAFQFALARLARSPSRAKFDIHKPGPLSDYPDCLAEYEGVYLRGMKEVERLLSGVITRPLT* (SEQ ID NO: 178)
	D	MKLAGQKPAPQWVDFHCHLDLYPNHSALIRECDISRVATLAVTTTPKAWMRNRELTSDSPYVRVALGLHPQLIAEREHEIALLEHYLPSARYVGEIGLDASPRFYRSFEAQERIFSRILNACFEQGDKELSIHSVRAAAKVLGHLENTRLTENCKAVLHWFTGSISEARRAVELGCYFSINEEMLRSPKHRKLVSFLPFERILTETDGPFVFHEEKAIHPRDVQRTVHEIAQIHHVSDTDAAMRILYNLRSLVTNSSHSENSS* (SEQ ID NO: 179)

TABLE 8-continued

Amino acid sequences of validated defense systems.

#	Gene	Sequence
12	A	MSTVDTSTAEELNQGGSDFILTSLEAMRKKLLDLTSRNRLLNFPITQKGSSLRIVDELPEQL YETLCSEIPMEFAPVPDPTRAQLLEHGYLKVGPDGKDIQLRAHPSAKDWAHVLGIRTDFDLP DSHKTVVSDSDRELLEKAHQFELQYAQGQNGKLTGIRSEYVNQGIALSALKEACCLAGYEGL EDFERQAKAGNEISISSSNPSHDDNRIQALLYPNELEACLRAIYGKAQTALEESGANILYLA LGFLEWYESDSSEKARYAPLFTIPVRCERGKLDPKDGLYKFQLYYTGEDILPNLSLKEKLQA DFGLALPLFNEEETPESYFASVKKVVEQHKPKWSVKRYGALSLLNFGKMMMYLDLDPARWPC DKRNILSHEVIRRFFTSQSCGQENSGLPGGFGQHEYCIDSYPDIHDKVPLIDDADSSQHSAL IDAIRGQNLVIEGPPGSGKSQTITNLIAAALLNGKKVLFVAEKMAALEVVKRRLDRAGLGQF CLELHSHKTHKRKVLDDINARLVSQATMPTMEEIDAQILRYEDLKQQLNEYAALINNQWAQT GKTIHQILSGATRYRHKLDIDATALHIENLSGKQLDKVTQLRLRDQIVEFSRIYKEVREQVG ANAEIYEHPWSGVNNTQIQLFDSARTVDLLQTWQTSIIDFQHSYQEYVDKWALEGESLNTLQ YIEQLVEDQSNLPVLCGSEHFPALSELDSPDAIARVRHYLDRFELLQGHYVALSQVIEPQKL RLLEQGQSCDFPREELEKYGAAEDFTLRDLVRWLESIQSIHDELSSIYAQLNDFKNALPDGI ASYIDDSQAGLLFCSELLSILGALPTELIRVRDPLFDDDDIDAVLRDLMCQIETLRPLRDGL STLYQLDQLPSQEMLAHAVAVIQQGGLFAWFKSDWRSAKALLMAQSRKPDTKFAELKRCSAD LLKYSELLQRFEQSDFGNQLGNAFRGLDTDCEQLMLLRDWYKKVRACYGIGFGKRVAIGSGL FNLDGEIIKGVHLIEKSQISSRLMTLVKRVEHEAKLLPRISSLLEEHASWLGEQGVLMQSYR QVRNTLIALQGWFINPDISLEQMTHSSEILQNINDLQISLENDSLQLGAFLQLTPACGAYKN KQLTLDTINDTLNFAEQLVDKINCVSLATQIRHLASGSDYDLLCRDGGEIVSKWNEQIKNAE LYALETKLERSQWLKSTDGSLNTLIERNERAIQQPRWLNGWVNFIRCYEQMHENGLQRIWSA VLAGSLPIEKVELGLALAIHDQLAREVIHIHPELMRVSGSQRNALQKSFKEYDKKLIELQRQ RIAAKIACRNIPEGNSGGKKSEYTELALIKNELGKKTRHIPIRQLVNRACNALVAIKPCFMM GPMSAAHYLEPGRMEFDLVVMDEASQVKPEDALGVIARGKQLVVVGDPKQLPPTSFFDRSAD GEDDDDAAALSDTDSILDAALPLFPMRRLRWHYRSRHEKLIAYSNRHFYNSDLVTFPSPNAE SPEYGIKFTYVSKGRFSNQHNIEEAQAVAEAVLHHAHHRPGESLGVVAMSSKQRDQIERAID ELRRNRPEFNDAIDGLHAMEEPLFVKNLENVQGDERDVIFISFTYGPSEHGGKVYQRFGPIN SDVGWRRLNVLFTRSKKRMHVFSSMRSEDVLTSETSKLGVISLKGFLQFAESGKLDSLTTHT GRAPDSDFEVAVMEALNHAGFECEPQVGVAGFFIDLAVKDPGCPGRYLMGIECDGAAYHSAK SARDRDRLRQEVLERLGWRISRIWSTDWFSNPDEVLSPIIRKLHELKTLAPDVVVPSYEYVE TIESSAEVASDSIDSLMPNLGLKEQLKYFATHVIEVELPNVDADRRLLRPAMLEALLEHQPL SRSEFVERIPHYLRQATDVYEAQRFLDRVLALIDGAEAEANDAAFESELA* (SEQ ID NO: 180)
13	A	MAGASIDAIGVINQIKDNLTDRYEDGFPVLKEIIQNADDAGANELTIGWSKGFCNAENELLN APALFFINDAPLAEEHRDAILSIAQSSKATSKASVGKFGLGMKSLFHMGEAFFFMSDQWRIE HWASDVFNPWDKYRDAWNEFGENDKCQIATKLKGFLSTDKPWFVVWVPLRTKALAKAHNNYI IINNFSGDEKLPSFFNQAHLSEKTSEILPQLKNLKDIGFFCESDKGVFDEVTSIQLHEDSSR SSFCGEPRLNNGDSFAVFSGKIYSNSNEERCALDYAGCERVIFDERLNQLKDENMGWPKSYQ FDKKANLPVEALDKAEQHASVTFSRFKTKGQAYLKANWAVFLPLSQTKELVAVPIEGEYDYN LYLHGYFFVDAGRKGLHGHDNLGFSTSLEHVKNDEKKLREVWNIILASEGTFNLVLPALNEF CQKLRLPHQIKTVLTKALYDLLIERYRKEVSKSANWIINIDDKGAAWSLLDKNAQCLPIPRP ENSDYSRIWSTLPGLSKLLDKKSLYEATGNEFLTEQNQRDSWNITLLEEALGSGVVNAFYRS INIEYLLQFLQLAKEQCTTEDFDNLIIPQFREVLSTHKLAELSLNKALNTQVFELVSAPKTV VLPIDKDDQSIWELVCKIIPAKLLLPKFLSTHNKPIHDNVTEEELFALLTLVDSYIKKQGER LSSDESSACERLITFVIDCVNASEYIQKSDFYQKSGHLKLLKVEALGSQQSTKYRSLNELIV LKEKYQLFLRGGERNFGKGLGKELVAVVPGLELCFISKDFEIGGLYEGLTACSEAACLRLLS TYPNLGSNSARLALTKVFSAELSTDEEKRGFRYLIHGSKEDDLRQTLWKPNRATNPVWMKIW RMCQPEDFPGWCELDEEFSNALTNQYEHFIGVKEQFYKDIISEYRTILPECNFDNFDDWEVE QLLADIGSQGDERLWKALPVHRTAHNTRVAITTKCLMEGSATVPSEWDVHLIQHSAIAEVAA CQHKWVNHGLPKELIEIALTQSSPAQYSAFILDQLCAIRIANEGIEHELEGKINNTKWLRLA SGTEVSPEAILSFSANELPESAKFCELKESNIYMFSQLDGNMFEHDQARGFLREWVAKSNSS VCSCILAEAAQHQSYVVGNFSNISAQVLEQISCIPPLMQLSAGWGLLVELYQSQYLSVNENK QVMLCKETEPQSLWWALERIADDDIFIGQSKELRKAFLEALCNTEGGVDYLPKLRFRNENGS YVSGNTLVSNVAQVVADNLISPQEYAVIESYCSKSALTNGNTSKIIELAGDNAPVLSDYFDD WEGMVPPDAIATFIALFAKSGGVEKLVNNYLRQSTLESIKQGYEEKWNSGKGRRGEFSHYPY SSLYKSVDFELAICAENAAYMTSIFGERIQVKLQKTPDSLLVHQANKSKTKRIELRRVDTKN VSKDQLLRMLAKAVETIFTDVFGAECIRFESEFLKRFGASEQVDIQITRQIVLENVVPLLER LQVREEGLCDLRSDYKREQRVLASSDPSVLQDRSRLNSVLTKIKETLENNEKVQSLVLESVR KEMSKHFQYSPFSVPFELFQNADDALCELIEMQGDSTNVLTRFDVVSGSDGTLNFYHWGREV NYCKSSYVAGKNQFDRDLEKMVSLNVSDKSDGKTGKFGLGFKSSLLLTDIPRLVSGDICAEI HAGVLPSVPSKPVMTELNQNVDEYKIGNRKPTLIQLPKCDKKRADLKLVLGRFKSNAGILTV FSRQIREINIDEQRFGWSGQALHNIPEVLVGEVKLPTNTSEESNVILRSNRVLIINTESGQF LFALDSNGVVSLSNRKNLSSFWVLNPIDEDLKLGFCINAPFAVDIGRSQLAVDNGDNIDLSS SLGKALSAVLVKMFAASSNNWNEFAEEVGLGQSSTFIKFWASLWDVITAHWPARLGETNSKA ELKQMFTVEDGLLAFYQRCAALPRNLGVKEDSLVQLKNVDTGANKPLTKAFNTLGNHPILQR LYKDQQLVGHDTFEFLKSIDFRPNNGALTKLELIDLIGQDFPHNEVNHDRASFYGRLFGKNF EKLMSNFEMTVTEKKVLEERFSELKFLNKTGVYVTASKLIVEGSPERDLLSKFAPDSAKLSE KYDQASMDLVSFIRRDVSYDIHSWAKQIRSEESNRGGKQEGLCSFLVEGGYLASSLLRKLQT DHPAFLTKGRFDPSVLTEKWRWSSSKASAFISIWIDTEEDKARFIVRQAQKEFIPNVTNGEQ ILENITNWWNQCRNQSLIDYDKQLYAQPMPWKAMTEDFELETLEVKKGWLKLFYLGSCQTLG FNNDVANRNVVSWFEDKGWWDKLAVANGPSPEVWKELMEEYLQTARVDERYRVWIQVLPLYR FATKLKDYVALFMNASFIDNLDDLLKPNSSNKLSGSGIQVSELKGTLGIGINFILRELQRHQ

TABLE 8-continued

Amino acid sequences of validated defense systems.

#	Gene	Sequence
		VLEREYCEDIQKYAFVLPARLRKLLKKMGAGLSFDAEPENSERAYDYFVSALNSETHPLLKDFDIPFRVLLADKQAFERCFNFALDEQFEEVYG* (SEQ ID NO: 181)
	B	MDNIIRVIHPKFGVGTVEFEKAETSLVRFEHGFEECLKSELEAVADLKSDLVSGQSVAASELALKTLAHSLKSVNENWSVFSKSNINLLPHQLWVCHRVLRQWPTNQLIADDVGLGKTIEAGLILWPLIERKRVKRLLILTPAPLVEQWHQRMLDMFDIRLSMYAPENDTSRVNYWDSNNMVVASLPTLRNDKNGRLERMLNAEPWDMLIVDEAHHLNSTEDKGGTLGFRFIQTLIENDKFESKLFFTATPHRGKEHGFFSLLQLLRPDLFNVKQMDEREMRPFVKDVLIRNNKQFVTDMNGERLFKPLSVSSRTYSYSEQEQFIFYDLLTKFIVSGQAYASSLNSRDQRAVMLVLTAMQKLASSSIAAIERALKGRIEKHKLGKQRLQDIEVQQAALLEKREESESQSESEIYSDELAQLELEFIETTTRVQLMDDELPRIMELLSACQKVGSETRILTILDILETEFKDRTVVFFTEYKATQALLMGALNKKYGEGCVTFINGENRLLNVENGSGVCVDYVTDRYNAAKRFNEGKVRFIISTEAGGEGIDLQQNCFSMIHVDLPWNPMRLHQRVGRLNRYGQVKNVEVITLRNPDTVESRIWDLLNTKIDLIMRSVGGAMDEPENLMELILGMADSTLFNELFTEAANRKNSESLSAWFDHKTKTFGGESWQKVKDLIGRAEKFDYQDLEAVPRLDLGDLKPFFTQMLSFNQRRCKYDENGGLSFLTPHAWLGQFGTRRSYEKLHFDRKAKQLDSEADIIGFGHPMFSKAVNQGEQIPGSYAFLNGIEKDLVVFKVQDQVTGTDASVKVSIVGLVLDDNGDCELVKDEDLIGYLNEYLKISNDVDSKRTPEDLVSVIQTANDYLMENVSSIGLPFRLPNSEPLTVFYKASN* (SEQ ID NO: 182)
14	A	MVAIKMYPAKDGDAFLIICDEEKSAFLIDGGYAETFRQHILPDLRELSFNGYRLRLVMATHIDSDHIGGLVDFFLVNGHAAEPAVITVDRVWHNSLRAMTRPENNAQKVDSREITDFLRRRYHVEADKAKPHEISARQGSSLAASLLAGDYHWNEGKGYQCICTGTSIPNLMCDNSLTILSPSKERISALCLWWRRQLASLGFSGRSSSSEAFDDAFEFFCKREASQVPLPHVINARTPLLERDYARDTSPTNGSSIAFSLVLNKKRILMLGDAWAEEWTSLGASGASHHFDIIKISHHGSIRNTSPNLLKIIDAPVYLISTDGKKHARHPNLAVLKAIVDRPAAFTRTLYFNYANSASAFMKNYLSASGAQFRIIEGSTDWITL* (SEQ ID NO: 183)
	B	MRYAATETEIRNATVLIECAGYTGSGTLIAADKVLTAAHCVVSDDPETPITVTFFGADEDVCVNATISEIDTSCDACLLTLSDSVDIPPITLMTQPEREGSQWKAFGYPASRNGPSHYLHGTISQILPRLFHGVDMDLSVSADCVLEEYSGVSGAAILSENKCIAMVRIRMDGGLGAVSLDKLSGLLIRNGLIPDDIASLPDSSLSGEVVLNRTEFRDNFESFVLEHKGRAVLLEGSPGSGKTTFCRHYQPRSEQLAVAGVYEFTPEDGAGTTFKILPEVFADWLHNQVSILLSGRPARREETEKINLTQKVSDLLHTFSDYWKHKGKYGVIFIDAVNEASECGDEAVSRFTALLPVTLPENVKLVFTAPSLSSAGKAFRHWLTPQDCISLTLLSHREVLQLTARELKTSAPSLSLLTRVSDIAQGHPLYLRYILGYLKANPDQVNLEIFPVFSGSIETYYERLWQGLVKDESAVNLLGILSRMRWGIDISSLIPVLTPQEQTVFVPTLDRIQHLLLNDKSSALCHQSFAAFINSKTAVINSLLHGRLADFCLTSGESYGLINRAYHLLLASHDRHPEAALVCTQEWADACIVKGAQPDELIHDIRQTLKNTLIRADAVASIRLLLLFQRMTFRHHFLFLQSAYHSGLALAALGRPDEALEQLIPSGSLVVDAVDAIVSAQTLARMGNSEHALKLLEKVKSAVDQEFERNPVNLSDFIGLSLAWVRAELMAGVVDGHGRTREVVEYLYGCGQVVRDNFEQSAHSKSAYTRAFYPLQAEMEAVNIAFNDRSVSLRTVKEKFGSLPENILDLMLSSVMRAHDIILQHQLPMPQHALQPVWYNLDRLLHTDIPYSNEIRFNSLSSLIFFNAPSALIIRMAGSFEVVPEITLLNEENEIAADSIDVSEQGQLWLVSAYLNETQPCPDIKHPSQGCSEWLKTLTEAIFWYSGQARRAVIDGNDEKKELLLVKVQNDILPALSYSLEERMAWPNSWAMPEQIIPMIYEELVNMFGACWPDKISVTTDFILAHTPQQCGLYSEGYTIRLLNRVIQTLLNEHRFLGQSDTTFQLLETLFIAFVSAFTENRQELVPELLNIIPAYISLDAPQLAQDTYTELLGVSMGPDWYKEDQFALMTTMLRVIPQHTDTNTTLSQVAGFLEHASGEMTFRRYVRQEKSQFIGELIRRGNYAHGFNYYRQQSCGSHEEMLTQLSHPAADSPHPLKGMRFPGGALDEEHAVECIVSELRNRVDWRLRWGLLEIFSFGSIGNLAVPFAELINEFSADTEDLNEIPKRLHNILHGDVPFSEHRNFIKNFTEHLADNHKPLFAEFISLLSEDTSDNDVKPPPSGDANQKGTDTSDDVAMQPGLFGKRSAINRAEACMENARKAAARRNTVRASELAVESLHIIQDGDWSVWRKNNHLAELTRTYILDNSADAGSVIRAYASLVEKERYAPAWVIASHLTEIAASKFSDQEAQAINQIVLEHNRHMLGNTEADAAHFSFLNEPDTSDAGEETLYFLFWLLEHPLKFRRERALEVLKWLASDDDKILGQCVTEALVSDIASRAEALMALTDWVSARSPQRIWDFIVKERSLFEWLEGTTALSQVHLLERVTSRAGFVLRNEIAAFERPRKLLLTSEASGQRNIPENLPTWVQSLSQTLAVMEKQGIDIPALLTLLEKRVLQQSGLADITVAFELEBCLLARGFTVNRTPSHHRWETMVRFALNQIIHEAAAQDELQNIEPLLRAWNPASEECVEPWEVCNRAKQIICAVMEGRHQQASGIEDGFFLHYLDEVEVSREGQTHLVEISAVLTTAHNGHESLRPGAESEFNATQTPDERTLSVHLTCQRVKMQPLLFGGATPAAVSKKFMQMTGTLPSDFIRRQWRSGRSLSKNRWGEPISRGSLLLMKRTTTLPPGLGLAWYVTVDGKLMNIFSYAPRRR* (SEQ ID NO: 184)
	C	MKYSMETPKTREEFEARCFHLLNAIKLGRYHGIPGEGNKEQVPFLPNGRVDLANIDTMTRLSMNSLYDFHYNRDNYPQFDLSENDENEEATD* (SEQ ID NO: 185)
15	A	MSDSLLVRTSRDGDQFHYLWAARRALRLLEPQSTLVALTIEGASTTEMGSQPWEDGEELTDIAEYYGSNELATATTVRYMQLKHSTMHSDTPFPPSGLQKTIEGFATRYKALIQKIPVETLRTKLEFWFVTNRPVSSSFSEAINDAANQHVTRHPHDLAKLEKFTGLQGAELSIFCQLLHIEGQQDDLWSQRNILLRESAGYLPDLDTEAPLKLKELVNRKALTESAANPSITRMDVLRALGVDETDLFPAPCRIERIENSVSRTQEATLVQRVVEAFGAPVIIHADAGVGKSIFSTHIEEHLPTGSVSILYDCFGLGQYRNASSYRHHHRTALVQMANEMASRGLCHPLIPNAGTGISQYMRAFLHRLSQSISILRASEPLAVLCIIIDAADNAQMAAEEIGETRSFIKDLIREKLPDGVCLVALCRPYRRELLDPPPEALTLSLQTFNRDETAAHLHQKFPDASESDVDEFHRLSSCNPRVQALSLSQNLPLNDTLRLLGPNPKTVEDTIGEVLEKSIARLRDTAGISERAQIDTICSALAILRPLIPLSVLSAISGVAGSAIKSFALDLGRPLIVSGETIQFFDEPAETWFQRRFRPSAADLHQFITKLRPLTKDSSYAASVLPALMLEGNQLSELIELAISSQALPETSAVERRDIELQRLQFALKAALRTGRYQDAAKLALKAGGECAGDNRQRVLLRDNIDLAAKFVGSNGVQELVSRNAFPDTGWPGSRNAYYAAILSEYPELSGEARSRLRLTMEWLTNWSQLPDDERSRQNVTDQDRAVMLIACLNIHGAEAAAREL

TABLE 8-continued

Amino acid sequences of validated defense systems.

#	Gene	Sequence
		RRWRPRKLSFDAGKIVAMQLLAHARYDELDQLAIAAGNDISLVMGIVLEARKLHRPVAEQAIRRTWRLLKSQRVSIKDRNHANNQTIAAITGMVEMALIQSVCTESESIQLLDRYLPKVPPYALTSEYSKERVAYVRAYALQANLMGSQLALSDLASTEVKKELMAEKRHGESDDLRQLKQYSGVLIPWYNLWAKVILGKTRKADLESELSDTQKESTAIKGHSYSEHSLSSNEIANVWFDILEAGNVSKDDVENIIKWSQHKGNRVFTPTLHRFSSVCAEISGLGELSYHFAELALSLWRDEHSDAQIKADGYIDLSRSLISLDEPEAKEYFNQAIEVTNKLGDENLSRWEAILDLAEYVAGKTQVPPETSYKLARCAELTREYVDRDKHFAWSDTVEILAELCPSSALAIISRWRDRTFGNHRSILAWTIEHLVKKNKINALDALPLITFENDWHKCDLLDSVLSSCTDDKDKIMAFEVVYHYTKFNVQNIQNLKKLDAISTSLGIEHTELKERISGLQHTETVSKKSSLSSNDNEQGHDQEWESIFKDCDLSSIDGISAAYEKFRNVPEFYSKETFIKKAISRVKTGKECSFITAIGAIFHWGLYDFKYILESIPDEWTSRLSIKTTLAGLIKEYCQRFCMRIRKSRVYEIFPFSLASRLSGISEKEIFGITLEAIAESPEPANSDRLFSLPGLLVSKLESNEALDVLSYALDLFDEVLKDEDGDGPWNEKLSPPTHVEDSLAGYIWARLGSPEAEMRWQAAHAVLALCRMSRTCVIQGIFQHAINATTLPFCDRNLPFYTLHAQLWLMIAAARVALDDGKSLIPNIGYFYHYATTDQPHVLIRHFAARTLLALHDSDLISIPAQEENKLRNINQSTTLPVLDKVEDHRGEDSYTFGIDFGPYWLKPLGRCFGVSQKQLEPEMLRIIRDVLGFKGSRNWDEDERNKRRYYQDRDNHHSHGSYPRVDDYHFYLSYHAMFMTAGQLLATKPLVGSDYDDVEDVFQDWLRRHDISRNDHRWLADRRDIPPKERSSWLNSSSDNRDEWLASISENVFNETLCPSPGLLTLWGRWSDVCSDRKESIIVHSALVSPERSLSLLRALQTTKNVYDYKIPDAGDNLEIDHAHYQLKGWIKDIAEYCGEDEFDPWAGNVRFPIPEPASFIIDAMKLTTDKDHRWVTSPSDVEPAMISSIWGHLSGKNDEEKSHGYRLCASIHFIKSALETFNMDLILEVDVDRYSRNSRYERNNENELDNIPSSTRLFLFRHDGTIHTLYGNYRNGEKTS* (SEQ ID NO: 186)
	B	MAHHIAELIYDAEHCTDDIVRTAKQAEIRDSIWSFWSNRYELPIGSRPFQELEPILRTLKGLDPENEQPRFFSPYRDLINVEKETSEVQKWLTAAKDIDSAAKILIDYCLSLAAENAIDKSQEWVELAQKAGLNKDVDLLEIRIFQLRGTPANTDNPNNAQRRILEKRQKRLEAFLLLGSQLNEQLKSQLEALPAIEDEPTDDDEDF* (SEQ ID NO: 187)
16	A	MEPISITVATYVATKLIDQFISQEGYGCIKKALFPQKRYVDRLYQLIEETAIEFEETYPVESGAIPFYHSEPLFEMLNEHIFFKEFPDKEILLDKFKEYPSITPPTQQQLSLFYEMLSLKINNCSKLKKLHIEETYKEKIFDINEELIQVKLILRSIDEKLTFHLSDDWLNEKNSQAIADLGGRYTPELNVKLEIAEIFDGLGRTNDFSKIFYSHIDSFLVAGKKLHSCDVISSELFEINQSLKEISDIYQEINFSKLDEIPINKFNNYVSSCQTAIGGAVSILWELREKSEQVGETKHYSDKYSSTLRMLREFDYACNELRIFINSTTVKLANNPFLLLEGKAGIGKSHLLADVIKNRIASGYPSLLILGQQLTSDESPWSQIFKRLQLKITSREFLEKLNLYGKKTGKRVLVFIDAINEGNGNKFWNDNINSFVDEIRCFEWLGLIMSVRTTYRNVTISHENVVRNNFEIHEHIGFQNVELEAVSLFYDYYNIERPSSPNLNPEFKNPLFLKLLCEGIKKNGLTKVPVGFNGISNIFNFLVEGVNKSLASPKKYAFDPSFPLVKDALNEIIKFKLEIGRNSISLKDAHSVVQSVVNDYVADKTFLSALIDEGLLTKGIVRNDDNSTEEVVYVAFERFDDHLTVNFLLNDVENIESEFKPDGRLKKYFHDECDFYIKSGIVEALSIQLPERYEKELYEFLPEFSNNLKLLEAFIDSLIWRDIKAIDFEKIRPFINEHVFKFKDSFDHFLEAVISISGLVGHPFNANFLHDWLKDYSLANRDSFWTTELKYKYSEDSAFRHLIDWAWARTDKSFVSDESIELVATSLCWFLTSSNRELRDCSTKALVSLLEPRIPVLRKIIDKFYGVNDPYVWERIFAVALGCTLRTDNIKELKYLAETVYQKVFCSKYVYPNILLRDYAREIIEFANHLGLELESIELSKTRPPYNSIWPDKIPSKEELESLYDKEPYRELWSSIMEDGDFSRYTIGTNYNHSDWSGCKFNETPVDRKQWKTFKCKLTDQQKDLYDATDPFIYDDKCEGIKFGRVVGRKAQEEIKASKKLFKNSLSYDLLSEFENEIEPYLDHNNNLLETDKHFDLRLAQQFIFNRVIELGWDPEKHGNFDQQIGTGRGRREAFQERIGKKYQWIAYYEYMARLADNFTRFEGYGDERKENPYQGPWEPYVRDIDPTILLKETGTKPGSNKEMWWLNDEVFDWTCSNEDWVKSSTTITNSYAFIEVKDDNGDEWIVLESHPSWKEPKIIGNDDWGHPRKEVWYQIRSYIVKVEEFENFRCWAIAQDFMGRWMPECTDRYQLFNREYYWSEAFKSFKSDYYGGSDWTSVTDRESGAKIADVSVTSINYLWEEEFDKSKIETLNFLKPSNLIFEKMGLKSGEVEGSFNDENGTMVCFAAEAVYASKPHLLVKKEPFLTMLRDNGFEIVWTLLGEKGVIGGSLISSHHYGRQEFSGAFYYEDSQLTGSHKTSFTR* (SEQ ID NO: 188)
17	A	MVKPNWDNFKAKFSENPQGNFEWFCYLLFCQEFKMPAGIFRYKNQSGIETNPITKDNEIIGWQSKFYDTKLSDNKADLIEMIEKSKKAYPGLSKIIFYTNQEWGQGRKSHEPEGDKNADNYLETVGNSNDPKIKIEVDQKAYESGIEIVWRVASFFESPFVIVENEKIAKHFFSLNESIFDLLEEKRKHTENVLYEIQTNIEFKDRSIEIDRRHCIELLHENLVQKKIVIVSGEGGVGKTAVIKKIYEAEKQYTPFYVFKASEFKKDSINELFGAHGLDDFSNAHQDELRKVIVVDSAEKLLELTNIDPFKEFLTVLIKDKWQVVFTTRNNYLADLNYAFIDIYKITPGNLVIKNLERGELIELSDNNGFSLPQDVRLLELIKNPFYLSEYLRFYTGESIDYVSFKEKLWNKIIVKNKPSREQCFLATAFQRASEGQFFVSPACDTGILDELVKDGIVGYEAAGYFITHDIYEEWALEKKISVDYIRKANNNEFFEKIGESLPVRRSFRNWISERLLLDDQSIKPFIAEIVCGEGISNFWKDELWVAVLLSDNSSIFFNYFKRYLLSSDQNLLKRLTFLLRLACKDVDYDLLKQLGVSNSDLLSIKYVLTKPKGTGWQSVIQFIYENLDEIGIRNINFILPVIQEWNQRNKVGETTRLSSLIALKYYQWTIDEDVYLSGRDNEKNILHTILHGAAMIKPEMEEVLVKVLKNRWKEHGTPYFDLMTLILTDLDSYPVWASLPEYVLQLADLFWYRPLKETGERYFISMDIEDEFGLFRSHHDYYPESPYQTPIYWLLQSQFKKTIDFILDFTNKTTICFAHSHFAKNEEEVDVFIEEGKFIKQYICNRLWCSYRGTQVSTYLLSSIHMALEKFFLENFKNADSKVLESWLLFLLRNTKSASISAVVTSIVLAFPEKTFNVAKVLFQTKDFFRFDMNRMVLDRTHKSSLISLRDGFGGTDYRNSLHEEDRIKACDDVHRNTYLENLALHYQIFRSENVTEKDAIERQQVLWDIFDKYYNQLPDEAQETEADKTWRLCLARMDRRKMKITTKEKDEGIEISFNPEIDPKLKQYSEEAIKKNSEHMKYVTLKLWASYKREKDERYKNYGMYEDNPQIALQETKEIIKKLNEEGGEDFRLLNGNIPADVCSVLLLDYFNQLNNEEREYCKDIVLAYSKLPLKEGYNYQVQDGTTSAISALPVIYHNYPMERETIKTILLLTLFNDHSIGMAGGRYSVFPSMVIHKLWLDYFDDMQSLLFGFLILKPKYVILSRKIIHESYRQVDYDIKKININKVFLNNYKHCISNV

TABLE 8-continued

Amino acid sequences of validated defense systems.

#	Gene	Sequence
		IDNKISIDDLGSMDKVLHILNTAFQLIPVDTVNIEHKKLVSLIVKRFSTSLLSSVREDRVDYALRQSFLERFAYFILHAPVSDIPDYIKPFLDGFNGSEPISELFKKFILVEDRLNTYAKFWKVWDLFFDKVVTLCKDGDRYWYVDKIIKSYLFAESPWKENSNGWHTFKDSNSQFFCDVSRTMGHCPSTLYSLAKSLNNIASCYLQGITWLSEILSVNKKLWEKKLENDTVYYLECLVRRYINNERERIRRTKQLKQEVLVILDFLVEKGSVVGYMSRENIL* (SEQ ID NO: 189)
18	A	MQVQHHTEPNLKNEIVALFKASQLIPFFGSGFTRDIRAKNGKVPDAIKFTELIRNLAAEKEGLTQTEIDEILRISQLKKAFGLLNMEEYTPKRKSKALLGNIFSECKLSDHEKTKIINLDWPHIFTFNIDDAIENVNRKYKELHPNRAVQREFISANKCLFKIHGDITEFIKYEDQNLIFTWREYAHSIEENKSMLSFLSEEAKNSAFLFIGCSLDGELDLMHLSRSTPFKKSIYLKKGYLNLEEKIALSEYGIEKVITFDTYDQIYQWLNNTLQNVERKSPTRSFELDDSKLMKEEAINLFANGGPVTKIVDNKRILRNSITFSQRDVCDDAIKALRNHDYILITGRRFSGKSVLLFQIIEAKKEYNASYYSSTDTFDPSIKNSLIKFENHIFVFDSNFFNAQSIDEILTTRVHPSNKVVLCSSFGDAELYRFKLKDKKILHTEIQIKNNLINEEGNYLNDKLSFEGLPLYKSSETLLNFAYRYYSEYKNRLSGSNLFNKQFDEDSMFVLILIAAFNKATYGHINSHNKYFDIQNFISQNDRLFELESTNTDPSGVIICNSPSWLLRVISEYIDKNPASYKTVSDLIISLASKGFLAASRNLISFDKLNELGNGKNVHKFIRGIYKEIAHTYREDMHYWLQRAKSELISAHTIDDLVEGMSYASKVRLDSAEFKNQTYYSATLVLAQLSARALSINNDKIYALSFFESSLESIRNYNNNSRHINKMMDKNDGGFRYAIQYLKDNPLIELLPRKDEVNELINFYESRKK* (SEQ ID NO: 190)
19	A	MQFITNGPDIPDELLQAHEEGRVVFFCGAGISYPAGLPGFKGLVELIYQRNGTTLSEIEREVFERGQFDGTLDLLERRLPGQRIAVRRALEKALKPKLRRRGAIDTQAALLRLARSREGALRLVTTNFDRLFHVAAKRTGQAFQAYVAPMLPIPKNSRWDGLVYLHGLLPEKADDTALNRLVVTSGDFGLAYLTERWAARFVSELFRNYVVCFVGYSINDPVLRYMMDALAADRRLGEVTPQVWALGECEPGQEHRKAIEWEAKGVTPILYTVPAGSTDHSVLHQTLHAWADTYRDGIQGKKAIWKHALARPQDSTRQDDFVGRMLWALSDKSGLPAKRFAELNPAPPLDWLLKAFSDERFKYSDLPRFCVSPHVEIDPKLRFSLVQRPAPYELAPQMSLVSGCVSASKWDDVMSHIARWLVRYLGDPRLIIWIAERGGQIHDRWMFLIESELDRLAALMRERKTSELDEILLHSPLAIPGPPMSTLWRLLLSGRVKSPLQNLDLYRWQNRLKNEGLTTTLRLELRGLLSPKVMLRRPFRYSEDDSSSTDEPLRIKQLVDWELVLTADYVRSTLFDLADESWKSSLPYLLEDFQQLLRDALDLLRELGESDDRHDRSHWDLPSITPHWQNRGFRDWVSLIELLRDSWLAVRAKDSDQASRIAQNWFELPYPTFKRLALFAASQDNCIPPERWVNWLLEDGSWWLWATDTRREVFRLFVLQGRHLTGIAQERLETAILAGPPREMYEDNLEADRWHYLVAHSVWLCLAKLRGAGLVLGESAATRLTEISTAYPKWQLATNERDEFSFIWMSGTGDPGFEESIDVDIAPRKWQELVQWLAKPMPERLPFYEDTWSDVCRTRFFHSLYALRKLSQDDVWPVGRWREALQTWAEPGMILRSWRYAAPLVLDMPDAVLQEISHAVTWWMEEASKTILCHEETLLALCRRVLMIETSPESSTIRNGIETYDPVSTAINHPIGHVTQSLITLWFKQNPNDNDLLPVELKTLFTKLCNVQIELFRHGRVLLGSRLIAFFRVDRPWTEQYLLPLFAWSNPVEAKAVWEGFLWSPRLYEPLLIAFKSDFLESANHYSDLGEHRQQFAIFLTYAALGPTEGYTVEEFRTAISALPQEGLEVAAQALYQALEGAGDQREEYWKNRVQPFWQQVWPKSRNLATPRISESLTRMVIAARGEFPAALAVVQDWLQPLEHLSYDVRLLLESDICSRYPADALSLLNAVTAEQHWGPRELGQCLLQIVQAAPQLEQDVRYQRLNEYSRRRSV* (SEQ ID NO: 191)
20	A	MTNKNKIKPLLNNISARLWDGRAAILIGAGFSRNAKPLTSKARKFPMWNDLGDIFYESVYCKKNDNRYSNVLKLGDEVQAAFGRATLDKLIMDHVPDKEYEPSKLHVSLLSLPWIDVFTTNYDTLLERASVNVDSRKYDIVLNKNDLMNAERPRIIKLHGSFPSERPFIVTEEDYRKYPLENSPFVNTVQQSLIENTLCLIGFSGDDPNFLNWIGWIRDNLGTENSPKIYLIGLFSFNEAQRKLLEKRNISIVDLSFLGDFGKDHYLAHQRFIQFLYESKNRDNLIEWPIETNYDRIVFNDGIELKTEKIKKCILEWAQSRQSYPNWLILPESNRSNLWQNTIDWLSVANYDVAWDGSDDLDFGYEITWRLNKALLPIFNDTSEFLFKLIEKYEINYVSGINNKIIDFDEKYSHITLSLMRFCRQENLIDKWKNLNDLLIQNLDRLTPEVKSDYYYENILFSYFNLNFDEARNKLSNWETNKLLPHHEIKRAGLLAEFGMLDEAINLLEETLSTIRRNSLLSSRNIDYSSESQEAYGIYILRMFKRSLRLDSKDDDYSSEYNSRLATLSQYRSDPENEIKYLEIKLESLPGTFKNTNDTDFDLNKRTVTTYLGGSPTEVRSLDAFSFFLLAEELGLPFHIPGMNIFSGIVENAARHIYQYSPEWAIFSIFRTFNKDKAKSLFNRNRISSLERKKVEDLFDGYYKKYEQIITKKIEDRLNDKLEIEISTLSIIPEILSRLVTKVSFNKKKDIIHLLLKLFNSDNFHQYMETKDLLKRTTSNLSDLQKISLIDEFIDFPSAPPNTQLHMGQRYNFLTPFECLLGVTITPPKENSKKIASAKLKKDINDLKSDNLDLRKAVSQKLITLYNLEMLNKSDTTKLIKNLWSKRDNFGFPIGSGYYKFFFINNLNPDNENIADKFISIIKTYKFPVQEGKRVSITGGLDEYCTELNGALHHISLPEKTLSEIISKIHDWYVKDRAWLEKRDDLAKEFTLRFRNITNIITTILEHHKDKLHAESINEISSLLDKMKEDKIPVNSAVTMLCLKNKSTYLERKDIENGLYSFNKDDVIEAINSTYVFIRNNEFPLTIIQAISDKIAWDRNPRLPDCYNLIAYIINSCEFTLPDYLIEKILRGLAYQINIDDRDFVDNNEYLNHLEKKLSATKLAASMFRKNETLGIDQPSIIQEWKNMCNSRNEFDEIRNEWNNNI* (SEQ ID NO: 192)
21	A	MSIYQGGNKLNEDDFRSHVYSLCQLDNVGVLLGAGASVGCGGKTMKDVWKSFKQNYPELLGALIDKYLLVSQIDSDNNLVNVELLIDEATKFLSVAKTRRCEDEEEEFRKILSSLYKEVTKAALLTGEQFREKNQGKKDAFKYHKELISKLISNRQPGQSAPAIFTTNYDLALEWAAEDLGIQLFNGFSGLHTRQFYPQNFDLAFRNVNAKGEARFGHYHAYLYKLHGSLTWYQNDSLTVNEVSASQYDEYINDIINKDDFYRGQHLIYPGANKYSHTIGFVYGEMFRRFGEFISKPQTALFINGFGFGDYHINRIILGALLNPSFHVVIYYPELKEAITKVSKGGGSEAEKAIVTLKNMAFNQVTVVGGGSKAYFNSFVEHLPYPVLFPRDNIVDELVEAIANLSKGEGNVPF* (SEQ ID NO: 193)
	B	MSLFKLTEISAIGYWGLEGERIRINLHEGLQGRLASHRKGVSSVTQPGDLIGFDAGNILVVARVTDMAFVEADKAHKANVGTSDLADIPLRQIIAYAIGFVKRELNGYVFISEDWRLPALGSSAVPLTSDFLNIIYSIDKEELPKAVELGVDSRTKTVKIFASVDKLLSRHLAVLGSTGYGKSNFN

TABLE 8-continued

Amino acid sequences of validated defense systems.

#	Gene	Sequence
		ALLTRKVSEKYPNSRIVIFDINGEYAQAFTGIPNVKHTILGESPNVDSLEKKQQKGELYSEEYYCYKKIPYQALGFAGLKLLRPSDKTQLPALRNALSAINRTHFKSRNIYLEKDDGETFLLYDDCRDTNQSKLAEWLDLLRRRRLKRTNVWPPFKSLATLVAEFGCVAADRSNGSKRDAFGFSNVLPLVKIIQQLAEDIRFKSIVNLNGGGELADGGTHWDKAMSDEVDYFFGKEKGQENDWNVHIVNMKNLAQDHAPMLLSALLEMFAEILFRRGQERSYPTVLLLEEAHHYLRDPYAEIDSQIKAYERLAKEGRKFKCSLIVSTQRPSELSPTVLAMCSNWFSLRLTNERDLQALRYAMESGNEQILKQISGLPRGDAVAFGSAFNLPVRISINQARPGPKSSDAVFSEEWANCTELRC* (SEQ ID NO: 194)
22	A	MDRSAVDTIRGYCYQVDKTIIEIFSLPQMDDSIDIECIEDVDVYNDGHLTAIQCKYYESTDYNHSVISKPIRLMLSHFKDNKEKGANYYLYGHYKSGQEKLTLPLKVDFFKSNFLTYTEKKIKHEYHIENGLTEEDLQAFLDRLVININAKSFDDQKKETIQIIKNHFQCEDYEAEHYLYSNAFRKTYDISCNKKDRRIKKSDFVESINKSKVLFNIWFYQYEGRKEYLRKLKESFIRRSVNTSPYARFFILEFQDKTDIKTVKDCIYKIQSNWSNLSKRTDRPYSPFLLFFIGTSDANLYELKNQLFNEDLIFTDGYPFKGSVFTPKMLIEGFSNKEIHFQFINDIDDFNETLNSINIRKEVYQFYTENCLDIPSQLPQVNIQVKDFADIKEIV* (SEQ ID NO: 195)
	B	MSRNNDINAEVVSVSPNKLKISVDDLEEFKIAEEKLGVGSYLRVSDNQDVALLAIIDNFSIEVKESQKQKYMIEASPIGLVKNGKFYRGGDSLALPPKKVEPAKLDEIISIYSDSIDINDRFTFSSLSLNTKVSVPVNGNRFFNKHIAIVGSTGSGKSHTVAKILQKAVDEKQEGYKGLNNSHIIIFDIHSEYENAFPNSNVLNVDTLTLPYWLLNGDELEELFLDTEANDHNQRNVFRQAITLNKKIHFQGDPATKEIISFHSPYYFDINEVINYINNRNNERKNKDNEHIWSDEEGNFKFDNENAHRLFKENVTPDGSSAGALNGKLLNFVDRLQSKIFDKRLDFILGEGSKSVTFKETLETLISYGKDKSNITILDVSGVPFEVLSICVSLISRLIFEFGYHSKKIKRKSNENQDIPILIVYEEAHKYAPKSDLSKYRTSKEAIERIAKEGRKYGVTLLLASQRPSEISETIFSQCNTFISMRLTNPDDQNYVKRLLPDTVGDITNLLPSLKEGEALIMGDSISIPSIVKIEKCTIPPSSIDIKYLDEWRKEWVDSEFDKIIEQWSKS* (SEQ ID NO: 196)
23	A	MAYEAQISRTNPAAFLFVVDQSGSMSDKMSSGRSKAEFVADALNRTLMNLITRCTKSEGVRDYFEIGVLGYGGQGVSNGFSGSLGGQVLNPISALEQNPARVEDRKRKMDDGAGGIIETAIKFPVWFDPIASGGTPMREALTRAAEELVTWCDAHPDCYPPTILHVTDGESNDGDPEEIANHLRQIRTNDGEVLILNIHVSSLGNDPIRFPSSDTGLPDAYAKLLFRMSSPLPEHLVRFAQEKGHTVGIESRGFMFNAEAAELVDFFDIGTRASQLR* (SEQ ID NO: 197)
	B	MKLEFLGTVPKDPEYPKANEDKFAFSEDGRRLALCDGASESFNSKLWADLLARKFTADPKVNPEWVASALAEYSATHDFRSMSWSQQAAFERGSFATLIGVEEFEEHQAVEILAIGDSITMLVDCGKLICAWPFDNPEKFNERPTLLATLYAHNNFVGGSTFWTRHGKTFYLEKLTQPKLLCMTDALGEWALKQALAEDSGFIELLSLQTEEELAELVLRERAAKRMHIDDSTLLVLSF* (SEQ ID NO: 198)
	C	MPYPSLEQYNQAFQLHSKLLIDPELKSGTVATTGLGLPLAISGGFALTYTIKSGAKKYAVRCFHRESKALERRYEAISRKISSLRSPYFLDFQFQPQGVKVEGISYPIVKMAWAKGETLGEFLEVNRRSAQAIAKLSASIESLAAYLEKEKIAHGDFQTGNLMVSDGGATVQLIDYDGMFVDEIKTLGSSELGHVNFQHPRRKATNPFNHTLDRFSLISLWLALKALQIDPSIWDKSNSELDAIEFRANDFVDPGSSSILGMLSGIQQLSTHVKNFAAVCASAMEKTPSLGDFIASKNIPISLASISMNGDIPVSRLKPGYIGAYTVLSALDYSACLQRVGDKVEVIGKIIDVKLNKTRNGKPYIFVNFGDWRGNIFKISIWSEGISALPSKPDASWIGKWISVIGLMEPPYVSGKYKYSHISITVTTIGQMTVLSEPDARWRLAGPNESRQTLTSTSSNQEALERIKSKSTTSTPMPMNTNATTANQAILNKLRASTQTVAAARAQTQHWPNKSSTHYVAPTGTSASQPVQNIPSPASTSKQQTSQKNIVTKILKWLFG* (SEQ ID NO: 199)
24	A	MVGSRWYKFDFHNHTPASHDYKIPDISPREWLLAYMKQHVDCVVTSDHNSGAWVDVLKGELENMSRDASTGDLPEFRPLTLFPGVELTATGNVHILAVLHTHSTSADVERLLAQCNNNSPIPSEVPNHQLVLQLGPAGIISNIRRNPKAVCILAHIDAAKGVLSLTNQAELTAAFQESPHAVEIRHRVEDITDGTRRRLIDNLPWLRGSDAHHPEQAGVRTCWLKMSSPDFDGLRHALLDPENCVLFDQLPPEEPASYLRSLKFRTRHCHPVGQDSASVEFSPFYNAVIGSRGSGKSTLIESIRLAMRKTEGLTATQGSKLDQFIRTGMEADSFIECIFHKEGTDFRLSWRPDSKHELHIFSDGEWMPDSHWSADRFPLSIYSQKMLYELASDTGAFLRVCDESPVVNKRAWKERWDQLEREYLNEQITLRGLRARQGSADSLRGELSDAERAVSQLQSSAYYPVCRQLALARNELSAATLPLEHFERRIAAIQALAEEPLQRSDIPPEPSGLLMAFMARLSSVQQQYDQRLNTLLAEYAAELAGIRREQSFIALRTAVSDQETNVESEAVSLRARGLNPDVLNELMARCESLKNELRNYDGLDGAISASVARSEQLLAEMRAHRMALTDNRKAFLSSLSLSALEIKILPLCAPYEDVISGYQTVTGISNFAERIYDNSDGSGLLSDFISERPFSPLPAATEKKYRALDELKALHHSIRLDNSEAGAGLHGSFRNRLRSLNDQQLDALQCWYPDDGIHIRYQTPGGQMEDIAFASPGQKGASMLQFLLSYGTDPLLLDQPEDDLDCLMLSMSVIPAIMSNKKRRQLIIVSHSAPIVVNGDAEYVISMQHDRTGLYPGLCGALQEAPMKALICRQMEGGEKAFRSRYERILS* (SEQ ID NO: 200)
25	A	MNEHLSHMDVHTLFEEMDEQADGITFKYSFDDIAKSNALVVTEFVNFERDSTVALLASLLTLPAHQSQCLRFELLTSLALIHCKGQQIANIDDVKRWYVTTGESSSIVGEDPAEDWVALVDNKKGDYRVLEGVWEAAGFYTQLMVEIVSDMPDTHRYRSLKLAIQAILRLSDVICARSGLYRFQEGADEFPDSLDTAGLDEKTLCSRVTLSERSLRAEGIKLADLAPFILEPSHISMLGNQVPGEGMLEQRPLLRTRDGIVVVLPTAMTIALRQAVITFAKRTEELSELDKALANVYSLTFSEMPVFGNGGRLRRLTWEKYKMSRTTMVTSIVDAGHLMVLQFVLPSIQQYADTGFNNLLQLDEETTQFLDNSVEQITVDLAKQPGFQRGIVVRIACGWGAGFMGVPPQLPDGWGFEWMSGADFVRFGALPDMSPIAFWRVQDAVETIRQAGVRLINMSGTLNLLGWIRANDGHMVPHDQLPDDRITPEHPLMLMIPTNLLRGIRIAADTGYDRHRISDNNGKWHRVMRPSAEDFFPTERQSKCYASIDDLEAQRLTC

TABLE 8-continued

Amino acid sequences of validated defense systems.

#	Gene	Sequence
		VYEGQGNLWVTLEAPEMEDWMLLVELAKMVRTWIGRIGEALEVLSEQPIKKSLKVYLHFDGNDNIGRFDGENSDDMNTFWRLERIHEHGAIRVVLQDGYLAGFRLPDNRAERALVRALGTAFATLLRMKEPVDKGVTVEQIAVPNDRARSFHIMQAYDFNQYLGRSLTKRLLAEDIDSAAARELAWRAVSTDAPSRYQGKKEVGKLLNDWDVLIQDLLSELSRFDRKQTVMRLLENVVKARCEEAHWRSTAAAVLGLHAGEEGVEETIAQEMSRYAGAALTSRLIIELAICVCPTSGGIEPSDMALSKLLARASLLFRIGGMSDAVRFGALPADIRISPLGDLLFRDELGKMVLEPMLSKVTNERFEEQAAQFEQHYVKTAGGDDENSKQDSVAAETTEDQTDIFLAFWKAEMGFTLEDGMRFIQFLESIGEQESAEEMRRSQLADAAKSAGLADETIDAFLNQFILSARPKWDVVPDGFDLSDIYPWRFGRRLSVAVRPLLQEESHDPLIVIAPGLLNLSLKYVFDGAYTGQFKRDFFRTEGMRDTWLGGAREGHTFEKTLERELRETGWTVRRGIGFPEERRNLPGDPGDIDLLAWRSDRNQVLVECKDLSLARNYSEVASQLSEYQGDDIKGKPDKLKKHLKRVLLAKENIDNFAKFTSIANPEIVSWLVFSGASPIAYAQSKEALAGTNVGRPSDLLNF* (SEQ ID NO: 201)
26	A	MDYLSEVLKIIEGATKANASMASNYAGLLADKLEQKGEVKQARMIRERLLRAPQALAGAQRAGGGISLGSLPVDIDSRLNTVDVSYPKLDSSEIFLPAAISTRVEEFITNVQRYDEFVKADAALPSRMLVYGKPGTGKTMLSKYIATRLDFPLLTVRCDTLISSLLGQTSKNLRQVFDYVMQRPSVLFLDEFDALAGARGNERDIGELQRVVTSLLQNMDAASEDTVIIASTNHEQLLDPAIWRRFSFRIPMPLPDIHQRELIWKNRLKNMICSDLDLSDLSRKSEGLSGAIIEQVSLDARRDAVIEGASVINHHKLYRRLYLAQSLMEGVNLSTYEDEIRWLRSKDKKLFSIRVLANLYKLTSRVISNILKESGAYEQKGYTV* (SEQ ID NO: 202)
	B	MSRRGTQFSNAKVTNPMLRIPFSSSDLGAIVNAGGGAKVLVDVTAEYRQGLVRNLTTSKHYLESKLSEYPGSLGTLVFKLRDQGIAKTHRPNKIAQEAGLQNAGHAKIDEMLVAAHAGCFDVLESVILHRNIKAILANLSAERIEPWDENRKVPGGTDGLFESSNILVRLFEYTGEDATYNNYENVISILEQHGVKYDEIRQKCGLPLLRIMDLSPNDRYILDILIDYPGIRTLIPEPKYSAFPVSVSDSVGIETNSFPVPSEELPIVAVFDTGVSPIAATITPWVVSRETYVIPPDTSYEHGTMVSSLISGAHFLNDNHPWIPDTKSKIHDVCALDENGSYISDLILRLADAVNKRPDIKVWNLSLGGGPCNEQTFSDFAMELDRLSDKFGILFVVAAGNYVDEPIRTWPNPDPLGGADLISSPGESVRALTVGSVSHMEANDALSEIGTPTPYTRRGPGPVFTPKPDIIHAGGGVHRPWNVGASSLKVVGPDNRLCSNFGTSFAAPIVASLAAHTWQRIATNTDFNVSPSLIKALLIHSAQLSSPDYSPSERRYLGAGIPNEVIETLYDSDDRFTLIFQTFLVPGVRWRKDNYPIPSALIQNGKFKGEIVITAAYAPPLNPNAGSEYVRANVELSFGLIENNTIKGKVPMEGENGQSGYERAQIEHGGKWSPVKIHRKAFNKGITSGNWALQAKTTLRANEPALMEPLPVTIVVTLKSLDGNTQVYADGVRALNANNWAHYPLPARVPVSV* (SEQ ID NO: 203)
27	A	MKTVRSACQLQPKALEINVGDQIEQLDQIINDTNGQEYFKKTFITDGFKTLLSKGMARLAGKSNDTVFHLKQAMGGGKTHLMVGFGLLAKDAALRNSHLGSMPYQSDFGSAKIAAFNGRNNPHSYFWGEIARQLGREGVFREYWESGAKAPDEQAWINIFDGEEPILILLDEMPPYFHYYSTQVLGQGTIADVVTRAFSNMLTAAQKKKNVCIVVSDLEAAYDTGGKLIQRALDDATQELGRAEVSITPVNLESNEIYEILRKRLFLSLPDKNEVSEIASIYASRLAEAAKAKTVERSAEALANDIESTYPFHPSFKSIVALFKENEKFKQTRGLMELVSRLLKSVWESDEEVYLIGAQHFDLSIHDVREKLAEISEMRDVIARDLWDSTDSAHAQIIDLNNGNHYAQQVGTLLLTASLSTAVNSVKGLTESEMLECLIDPNHQGSDYRNAFTELAKSAWYLHQTQEGRNYFSHQENLTKKLQGYADKAPQNKVDELIRHRLEEMYRPVTKEAYEKVLPLPEMDEAQATLRSGRALLIISPDGKTPPGVVGNFFKGLVNKNNILVLTGDKSSIASIEKAARHVYAVTKADNEITASHPQRKELDEKKAQYEQDFQTTVLSVFDKLLFPGNNRGEDVLRPKALDSTYPSNEPYNGERQVVKTLTSDPIKLYTQINENFDALRARAESLLFGTLDEARKTDLLDKMKQKTQMPWLPSRGFDQLAIEAYQRGVWEDLGNGYITKKPKPKTTEVIISEDSSPDDAGTVRLKIGVANAGNSPRIHYAEDDEVTESSPVLSDNTLATKALRVQFLAVDPTGKNLTGNPTTWKNRLTLRNRFDEVARTVELFVAPRGTIKYTLDGSEARNGETYTVPIQLADQEATIYVFAECDGLEEKRNFTFAAAGSKEIPIIKDKPATLVSPSPKRMDSSAKTYEGLKIAKEKGIEFEQISLMVGSAPKVIHISLGEMKISAEFIETVLTHLQTVLSPEAPVVMTFKKAYTQTGHDLEQFVKQLGIEIGNGEVEQR* (SEQ ID NO: 204)
	B	MNKTVDFGAPSEFGMHHFYVEIPAAPRDAVVIYEDYGFDGEDSRRETVECRLILARELWTKIRDDVRRDFNARLKIKKQSSGTWSTGKVKLDRFLGRELCVLGWAAEHASPDECLVICQKWLALRPEERWWLYSKTAAEAGRDDQTQRGWRKALYCALSDGANIKLETKKKPKSKKLQVEDETQDLFGFMEKGEF* (SEQ ID NO: 205)
	C	MALQPFEWRDKPSLIEHLFPVQKISAETFKERMASHGQLLVSLGAFWKGRKPLHNKACILGSLLPATDNPLEDLEVFELLMGIDSESMQKRIEASLPASKQETIGDYLVLPYAEQIRIAKRPEEIDESLFVHIWNRVNNHLGTSAHTFAQLVEELGVARFGHRPRVADVFSGSGQIPFEAARLGCDVYASDLNPISCMLTWGALNVVGASAQKRVEIDKAQRDIVKKVQKEIDELDIESDGRGWRAKVFLYCVEVTCPESGWRVPLIPSLIISNSFRVVAELKPVPAERRYDISIREVSTDEELEFYKSGTIQDGEVIHSPDGKTQYRVNDCTIRGDYKEGKENLNKLRMWEKTDFAPRPDDIFQDRLFCVQWMKKKPKGSQYYYEFRTVTNDDLKREKKVIEHVASKLDDWQKQGLVPDMVIEAGDKTDEPIRTRGWTHWHHLFHPRQLLFLSLVNKYSLAEGKFNFLQCMNFILSKLTRWRPQAGGGGGSAATFDNQALNTLYNYPVRATGSIENILAAQHNHCGISENVSFVVNSHPAPELDVENDIYITDPPYGDAVKYEEITEFFIAWLRKNPPKEFAHWTWDSRRSLAVKGEDEGFRTGMVAAYRKMAQKMPDNGLQVLMFTHQSGAIWADMANIIWASGLQVTAAWYVVTETDSALRGGSNVKGTIILILRKRHQALETFRDDLGWEIEEAVKEQVESLIGLDKKVRSQGAEGLYTDADLQMAGYAAALKVLTAYSRIDGKDMVTEAEAPRQKGKKTFVDELIDFAVQTAVQFLVPVGFEKSEWQKLQAVERFYLKMAEMEHQGAKTLDNYQNFAKAFKVHHFDQLMSDASKANSARLKLSTEFRSTMMSGDAEMTGTPLRALLYALFEISKEVEVDDVLLHLMENCPNYLPNKQLLAKMADYLAEKREGLKGTKTFNPEQEASSARVLAEAIRNQRL* (SEQ ID NO: 206)
	D	MAIKRFSSRTERLDTEFLAESLKGAAKYFRIAGYFRSSIFELVGEEIAKIPEVKIICNSELDLADFQVATGRNTALKERWNEVDVEAEALLKKERYQILDQLLHSGNVEIRWPRERLFLFIGKA

TABLE 8-continued

Amino acid sequences of validated defense systems.

#	Gene	Sequence
		GSIHYADGSRKSFIGSVNESKSAFAHNYELVWQDDDEESADWVEREFWALWTEGVPLPDAILAEIHRVSNRREVTVDVLKPEEVPAAAMAEAPIYRGGEQLQPWQRSFVTMFLEHREIYGKARLLLADEVGVGKTLSMATSALVSALLDDGPVLILAPSTLTIQWQIEMMDKLGVPAAVWSSQKKVWLGVEGQILSPRGDASSIKKCPYRIAIISTGLIMHQREKTDFVKEAGMLLKNRFGTVILDEAHKARIRGGLGDQASEPNNLMAFMLQIGRRTRHLVLGTATPIQTNVRELWDLLGILNSGAEFVLGDALSPWHDHEQAIPLITGQTQVTSEAEVWHWLSNPLPPSNEHHTVQQIRDYLSIDNKSFGYSHRFEDLDYMIQSLWLSECMTPSFFKENNPILRHTVLRKRKQLEDDGLLERVGVNTHPIKRNLAQYQSRFVGLGIPTNTPFQVAYEKAEEFSKLLQSRTRAAGFMKSLMLQRICSSFASGLKTAQKMLKHTVSDEDEDLVEDVEHLLSEMTPAEVACLREIETQLSRPEAVDSKLNTVKWFLTEFRTDGKTWLEHGCIIFSQYYDTAEWTAKELAKSLKGEVVAVYAGVGKSGLFRGEQFNNVERELIKSAVKTREILLVVATDAACEGLNLQTLGTLINVDLPWNPSRLEQRLGRIKRFGQTRKFVDMLNLWSETQDEKVYNVLSERLRDTYDIFGSLPDTIDDEWIDNEEELNTRMDEYMHERKKAQDAFSVKYRGTLDPDAHLWERCATVLSRRDIVSKLSEPWGS* (SEQ ID NO: 207)
28	A	MSEQFVSEAAGTPHLAEQDDGLKNLKLLEESFNTDKLNSSEQKKLQELRSILSPLLKKGGVLADLFQDGKDVLAFPIDVDSVLQHLNQDMRDDWFTDTLQHKDLLSNKQSLHEVLHELLNEGNGQYIGSFRSVYNIPKKGLGIRYSLETDFYDRFIYQAICTFLIQFYDPLLSHRVLSHRFNKDRKSEKYIFKSRIDLWQTFEGVTRTALSNNQSLLATDLINCYENITIETIRTAFERSIEHINTSGPNKVLIRNAVQTLCNLLSRWGYSERHGLPQNRDASSFIANWLNDIDHEMVRLGYDYYRYVDDIRYICPNTRVAKKALTELINQLRKVGMNINSGKTKILTQDSTANEVDEFFPTSDDRSLTIDNMWRSRSRRVIARSAKYIFQELKECIEEKQTQSRQFRFAVNRLIKLTDAGIFDIHATIATDLKALLISSLEDHAASTDQYCRLLGILDLNEHELNDIYNHLSDHERSVHSWQNFHLWLLLANRKYKSTNLITLATARIESDILQPEIAAIFIYLKCVGEAQVLIDNISKFESAWPYYHQRNFLLACSDFDHNQLKPLISKLGPKLKWTGSRAKPYFTNGMPLVERDKIAMLDLYDEITPYD* (SEQ ID NO: 208)
	B	MTESKKALLFIADYTDQGQDRIFLWSDGTLGEVTISDLVDQKHELVCHDLWLIAPSLYRATNKLPSNITDIEELRILTSGKKKERESRDKKDISQLLSSFVSEETIARYKEIFNRKIPLDEAVLSSIGEALLKCSEWKSDANTAGEWERFITERPVNDYLIRSTSEGISISEEKLRYHKNKIEFEFYMALKSFSSDYDMPLEVPSDQAVIEYLEPKGFDFTGLDVDYILNFVPMQSHFAEDLIRLRKIQNSRRVLAAIPLSQSRIYPIVDSFGSITSRIYFKDPSLQNLAKHHRDILIPDTNKQLSYIDYDQFEAGVMAALSGDEKLLELYNSSDVYEIAAKEIFDDKSKRKQAKRLFLSYAYGMKRQHILAAAQGFGADRQNAKKFFEQFKTFEAWKVLVHEEFHRTGRIGTALGNYMHRERKGELTSKEKRSAISQIVQGTASLIFKKALLCLSSISEVKLKLPMHDAVLLEHPADYDMDRVINIFSEIMSEHFQNKIQGKASLSQFHEDL* (SEQ ID NO: 209)
29	A	MSVIRGLAAVLRQSDSDISAFLVTAPRKYKVYKIPKRTTGFRVIAQPAKGLKDIQRAFVQLYSLPVHDASMAYMKGKGIRDNAAAHAGNQYLLKADLEDFFNSITPAIFWRCIEMSSAQTPQFEPQDKLFIEKILFWQPIKRRKTKLILSVGAPSSPVISNFCMYEFDNRIHAACKKVEITYTRYADDLTFSSNIPDVLKAVPSTLEVLLKDLFGSALRLNHSKTVFSSKAHNRHVTGITINNEETLSLGRDRKRFIKHLINQYKYGLLDNEDKAYLIGLLAFASHIEPSFITRMNEKYSLELMERLRGQR* (SEQ ID NO: 210)
	B	MTKQYERKAKGGNLLSAFELYQRNSDKAPGLGEMLVGEWFEMCRDYIQDGHVDESGIFRPDNAFYLRRLTLKDFRRFSLLEIKLEEDLTVIIGNNGKGKTSILYAIAKTLSWFVANILKEGGSGQRLSEMTDIKNDAEDRYSDVSSTFFFGKGLKSVPIRLSRSALGTAERRDSEVKPAKDLADIWRVINEVNTINLPTFALYNVERSQPFNRNIKDNTGRREERFDAYSQTLGGAGRFDHFVEWYIYLHKRTVSDISSSIKELEQQVNDLQRTVDGGMVSVKSLLEQMKFKLSEAIERNDAAVSSRVLTESVQKSIVEKAICSVVPSISNIWVEMITGSDLVKVTNDGHDVTIDQLSDGQRVFLSLVADLARRMVMLNPLLENPLEGRGIVLIDETELHLHPKWQQEVILNLRSAFPNIQFITTTHSPIVLSTIEKRCIREFEPNDDGDQSFLDSPDMQTKGSENAQILEQVMNVHSTPPGIAESHWLGNFELLLLDNSGELDNHSQVLYDQIKAHFGIDSIELKKADSLIRINKMKNKLNKIRAEKGK* (SEQ ID NO: 211)
	C	MRELARLERPEILDQYIAGQNDWMEIDQSAVWPKLTEMQGGFCAYCECRLNRCHIEHFRPRGKFPALTFIWNNLFGSCGDSRKSGGWSRCGIYKDNGAGAYNADDLIKPDEENPDDYLLFLTTGEVVPAIGLTGRALKKAQETIRVFNLNGDIKLFGSRRTAVQAIMPNVEYLYTLLEEFDEDDWNEMLRDELEKIESDEYKTALKHAWTFNQEFA* (SEQ ID NO: 212)

[0317] Sequence of vector backbone. Inserts were cloned between the HindIII and EcoRI restriction sites (underlined).

(SEQ ID NO: 213)

CCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATC

GGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTT

GCGCAAGCTTCTGCAGAATTCGCAATAACTAGCATAACCCCTTGGGGCCT

CTAAACGGGTCTTGAGGGGTTTTTTGCTGAAACCTCAGGCATTTGAGAAG

CACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCAATA

-continued

GACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGA

ATTTGCTTTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGG

CGTAGCACCAGGCGTTTAAGGGCACCAATAACTGCCTTAAAAAAATTACG

CCCCGCCCTGCCACTCATCGCAATACTGTTGTAATTCATTTAACATTCTG

CCGACATGGAAGCCATCACAGACGGCATGATGAACCTGAATCGCCAGCGG

CATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATAGTGAAAACGG

-continued

```
GGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAA

CTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTT

AGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATA

TGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAA

AACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATC

CCATATCACCAGCTCACCGTCTTTCATCGCCATACGGAACTCTGGATGAG

CATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGC

TTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGT

CTGGTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTT

TACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTTC

TCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATAC

GCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGT

GCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTA

TCAACAGGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCAC

AGGTATTTATTCGGCGCAAAGTGCGTCGGGTGATGCTGCCAACTTACTGA

TTTAGTGTATGATGGTGTTTTTGAGGTGCTCCAGTGGCTTCTGTTTCTAT

CAGCTGTCCCTCCTGTTCAGCTACTGACGGGGTGGTGCGTAACGGCAAAA

GCACCGCCGGACATCAGCGCTAGCGGAGTGTATACTGGCTTACTATGTTG

GCACTGATGAGGGTGTCAGTGAAGTGCTTCATGTGGCAGGAGAAAAAAGG

CTGCACCGGTGCGTCAGCAGAATATGTGATACAGGATATATTCCGCTTCC

TCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTGCGGCGAGCGGAAAT

GGCTTACGAACGGGGCGGAGATTTCCTGGAAGATGCCAGGAAGATACTTA

ACAGGGAAGTGAGAGGGCCGCGGCAAAGCCGTTTTTCCATAGGCTCCGCC

CCCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGCGAAAC

CCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGCGGCTCCCTCGT

GCGCTCTCCTGTTCCTGCCTTTCGGTTTACCGGTGTCATTCCGCTGTTAT

GGCCGCGTTTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGT

TCGCTCCAAGCTGGACTGTATGCACGAACCCCCCGTTCAGTCCGACCGCT

GCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGAAAGACATGCA

AAAGCACCACTGGCAGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCT

TGAAGTCATGCGCCGGTTAAGGCTAAACTGAAAGGACAAGTTTTGGTGAC

TGCGCTCCTCCAAGCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCAGAG

AACCTTCGAAAAACCGCCCTGCAAGGCGGTTTTTTCGTTTTCAGAGCAAG

AGATTACGCGCAGACCAAAACGATCTCAAGAAGATCATCTTATTAATCAG

ATAAAATATTTCTAGATTTCAGTGCAATTTATCTCTTCAAATGTAGCACC

TGAAGTCAGCCCCATACGATATAAGTTGTAATTCTCATGTTAGTCATGC
```

Example 3—Diverse Enzymatic Activities Mediate Antiviral Immunity in Prokaryotes

[0318] Bacteria and archaea are frequently attacked by viruses and other mobile genetic elements and rely on dedicated antiviral defense systems, such as restriction endonucleases and CRISPR, to survive. The enormous diversity of viruses suggests that more types of defense systems exist than are currently known. By systematic defense gene prediction and heterologous reconstitution, here Applicants discovered 29 widespread antiviral gene cassettes, collectively present in 32% of all sequenced bacterial and archaeal genomes, which mediate protection against specific bacteriophages. These systems incorporate enzymatic activities not previously implicated in antiviral defense, including RNA editing and retron msDNA synthesis. In addition, Applicants found a diverse set of other defense genes. These results highlight an immense array of molecular functions that microbes employ against viruses.

[0319] Domain-independent identification of uncharacterized defense systems

[0320] Many antiviral defense genes in bacterial and archaeal genomes show a distinctive tendency to cluster together within defense 'islands' (7, 10). As a consequence, an uncharacterized gene whose homologs consistently occur next to, for instance, restriction-modification genes has an increased likelihood of being involved in defense (11, 12).

[0321] Applicants found that additional, unknown defense systems exist which either lack annotated domains, or only contain domains that are not typically associated with defense but have been co-opted in specific instances to perform defense functions. Applicants developed an expanded computational approach in which novel defense systems were identified independent of domain annotations (FIG. **16**A). Applicants analyzed all bacterial and archaeal genomes available in Genbank as of November 2018, collectively encoding 620 million proteins. To identify candidate novel defense genes, Applicants first compiled a list of all genes within 10 kb or 10 open reading frames away from known defense systems (see Methods). This initial list ($n=8.7\times10^6$) which evidently contained both novel defense genes and non-defense ones, was clustered to yield 6×10^5 representative sequences ("seeds"). To distinguish between defense and non-defense seeds, Applicants identified all homologs of each seed present in Genbank and analyzed their gene neighborhoods. The seed was predicted to be a defense gene if these neighborhoods resembled those of known defense gene—in particular, if a high percentage of homologs were located in proximity to known defense genes ("defense score") and displayed context diversity (FIGS. **16**B, **21**A-**21**D, and Methods). All clustering and homolog detection steps were performed based on amino acid sequences, without invoking existing domain annotations and thus allowing the identification of novel types of defense genes.

[0322] After all filtering and curation steps, Applicants identified a total of 7,472 seeds that represented candidate defense genes, along with 4,555 seeds for known defense genes under the same analysis parameters (FIG. **16**C). These seeds were analyzed with additional, more sensitive analysis of their domain content. Of the uncharacterized genes, 1,687 (23%) had either no annotated domains or contained only domains of unknown function (DUFs), and an additional 2,756 (37%) contained only domains that are different from the characteristic domains of known defense genes. These results suggested the existence of a diverse set of defense genes with mechanisms that remain to be investigated.

[0323] Candidate defense systems exhibit antiviral activity in a heterologous system

[0324] To characterize the functional diversity among the predicted defense genes, Applicants selected 48 candidate systems to test experimentally for defense activity. Candidate systems were prioritized based on the presence of predicted molecular functions not previously implicated in defense; broad phylogenetic distribution; the presence of at least one protein larger than 300 amino acids (to increase the likelihood of the presence of enzymes); and, for multi-gene systems, conservation of the component genes. Because wild-type bacterial strains are likely to harbor multiple active defense systems, thereby maintaining phage resistance even if one of the systems were knocked out (13), Applicants elected to assay activity by heterologous reconstitution. For each system, 1-4 homologs were selected, cloned from the source organism into the low-copy vector pACYC and transformed into *Escherichia coli* (FIG. **17**A), comprising a total of 395 kb of exogenous DNA (see tables 9-16 for sequence, accession, and source organism information). Three previously identified defense systems, BREX type I (13, 14), Druantia type I (4), and the abortive infection reverse transcriptase RT-Abi-P2 (15) were included as positive controls. Each system was then challenged with a diverse panel of coliphages with dsDNA, ssDNA, or ssRNA genomes, and phage sensitivity of the bacteria was compared to that observed with the empty vector control (FIGS. **17**B-**17**C).

[0325] Applicants observed anti-phage activity for 29 of the 48 tested candidates (60%) (FIG. **22**). Systems from source organisms outside the Enterobacteriaceae family, which had *Escherichia* and closely-related genera including *Salmonella* and *Klebsiella*, had little to no activity, suggesting the importance of host compatibility. The most active representative in each of these 29 systems (representing 4% of the uncharacterized defense seeds) was further tested with an expanded panel of phages in two *E. coli* strains (FIGS. **17**D and **23**). All 29 systems were active against at least one dsDNA phage, and four were active against ssDNA phages (M13 or φX174). Phage specificity was typically narrow and varied widely across systems. The abundance of these defense systems among the sequenced bacterial and archaeal genomes spans two orders of magnitude, ranging from ~0.1% to ~10% of the genomes (FIG. **17**D). Overall, 32% of all sequenced bacterial and archaeal genomes contain at least one of these novel defense systems, which are broadly distributed across bacterial and archaeal phyla (FIG. **24**).

[0326] RADAR with a divergent adenosine deaminase that edits RNA in response to phage infection

[0327] Applicants identified a two-gene cassette consisting of an ATPase (~900 residues) and a divergent adenosine deaminase (~900 residues) that was active against dsDNA phages T2, T3, T4, and T5. Because deaminase activity had not been previously implicated in antiviral defense, Applicants focused on this system for further investigation. The system appeared in diverse defense contexts and forms three subtypes (FIGS. **18**A and **25**A). In most cases, it had the ATPase and deaminase only, but some variants also included a small membrane protein, either a SLATT domain (16) or the type VI-B CRISPR ancillary protein Csx27 (17). Mutations in the ATPase Walker B motif or in the putative divalent metal cation-binding H×H motif of the deaminase abolished defense activity, whereas the SLATT domain membrane protein was required for resistance against phage T5 but not against phage T2 (FIG. **18**B).

[0328] Given the large size of the deaminase compared to typical metabolic adenosine deaminases and its sequence divergence due to large insertions within the deaminase domain (FIG. **25**B), Applicants found that it acted on nucleic acids rather than on free nucleosides or nucleotides. Applicants performed whole-transcriptome sequencing and found an enrichment of A to G substitutions in sequencing reads at specific sites in the presence of phage, whereas C, G, or U bases were not affected (FIGS. **18**C and **26**A), consistent with RNA editing of adenosine to inosine. Furthermore, the overall expression of phage genes, including early genes, was reduced by ~100-fold even at a multiplicity of infection (MOI) of 2 (FIG. **18**D). Since most of the cells in the culture were expected to be infected, this suggested that defense activity occurs early in the infection cycle, which was not evident from efficiency of plating (EOP) alone.

[0329] RNA editing occurred only when both the defense system and the phage were present; expression of the defense system without the phage resulted in a near-baseline level of editing, and no editing was detected in the absence of the system. Mutations in the ATPase or deaminase active sites abolished editing, and no DNA editing was detected (FIG. **26**B). Editing sites were broadly distributed throughout the *E. coli* transcriptome (FIGS. **18**E, **26**A, **27**, and table 17), and editing could also be induced by co-expressing specific phage proteins with the system (FIGS. **28**A-**28**F and table 18). RNA secondary structure predictions indicated a characteristic stem-loop structure at strong editing sites; specific adenosines in loops were edited with up to ~90% frequency, whereas adenosines within the stem were not edited within the limit of detection (FIGS. **18**E and **27**). Finally, some of the editing sites were deleterious to the host cell, resulting in nonsynonymous mutations such as at the UAA stop codon of the transfer messenger RNA (tmRNA) (FIG. **28**B), which rescues ribosomes stalled during translation (18).

[0330] Based on these results, Applicants named this system phage restriction by an adenosine deaminase acting on RNA (RADAR). Growth kinetics at varying phage multiplicity of infection (MOI) revealed a threshold MOI above which RADAR-expressing cells had a lower OD600 compared to the empty vector control, suggestive of RADAR-mediated growth arrest (FIG. **18**F). Together with the abundance and broad distribution of editing sites in the host transcriptome (FIGS. **26**A-**26**B, **27**), these results are consistent with an editing-dependent abortive infection mechanism that is activated by phage.

[0331] A widespread family of defense systems containing reverse transcriptases

[0332] Applicants discovered that a family of uncharacterized reverse transcriptases (RTs) are active defense systems. Although most RTs in prokaryotes are components of mobile retroelements, distinct clades of RTs that lack the hallmarks of mobility also exist, including 16 'unknown groups' (UGs) (19-22). Applicants independently identified many of these uncharacterized RTs via the pipeline, suggesting that they might be defense genes (FIG. **19**A). Indeed, six of these candidates (UG1, UG2, UG3, UG8, UG15, and UG16) provided robust protection against dsDNA phages. In all cases, mutations in the RT active site ((Y/F)×DD (SEQ ID NOS: 1-2) to (Y/F)×AA) abolished activity (FIGS. **19**B and **29**A-**29**B). Applicants named these genes defense-associated RTs (DRTs).

[0333] Each of these RT systems displayed a distinct pattern of phage resistance (FIG. **17**D). Moreover, while UG2 (drt2), UG15 (drt4), and UG16 (drt5) act as individual genes, the UG3 (drt3a) and UG8 (drt3b) RTs were components of the same defense system (DRT type 3), with both RTs required for defense activity. Like RADAR, some subtypes of the UG1 (DRT type 1) and DRT type 3 systems were also associated with small membrane proteins (FIG. **19**A). Moreover, DRT type 1 encompassed a much larger protein (~1200 residues) than the other five RTs and also contains a C-terminal nitrilase domain. Mutation of the catalytic cysteine of the nitrilase (C1119A) abolished the activity (FIG. **19**B). Nitrilases typically function in processes unrelated to defense, such as nucleotide metabolism and small molecule biosynthesis (23). Thus, DRT type 1, which is divergent from typical nitrilases and forms a distinct clade in the phylogenetic tree of the nitrilase family (FIGS. **30**A-**30**C), exemplifies a non-defense domain that was apparently co-opted for a defense function.

[0334] To further characterize these RTs, Applicants performed whole transcriptome sequencing of RT-expressing *E. coli* during phage infection. These experiments revealed substantial differences in phage gene expression across the different RTs (FIG. **19**C). For instance, DRT type 1 strongly suppressed the expression of phage late genes, such as capsid proteins, whereas early and middle genes were not substantially affected, suggesting that it is active prior to the late stage of infection but does not prevent the injection of phage DNA into the host cell. In contrast, DRT type 3 did not strongly suppress expression of any of the phage genes, despite growing at a rate similar to DRT type 1 during phage infection (FIG. **31**A). Transcriptome sequencing also identified a highly expressed, structured non-coding RNA at the 3' end of the DRT type 3 system that is required for activity (FIGS. **19**B, **19**D-**19**E).

[0335] Retrons Mediate Anti-Phage Defense

[0336] Applicants also found that retrons, a distinct class of RTs that produce extrachromosomal satellite DNA (multicopy single-stranded DNA, msDNA), are active anti-phage defense systems. The retron msDNA is produced from the 5' UTR of its own mRNA and is covalently linked to an internal guanosine of the RNA via a 2'-5' phosphodiester bond (24). First identified over 30 years ago, retrons have been harnessed for bacterial genome engineering (25), but their native biological function has remained unknown. Applicants found that the original *E. coli* retrons Ec67 (26) and Ec86 (27), as well as a homolog of the Ec78 retron (28) and a novel TIR (Toll/interleukin 1 receptor) domain-associated retron, mediate defense against dsDNA phages. Of note, the Ec86 retron is natively present in the widely-used laboratory *E. coli* strain BL21. Mutations in the (Y/F)×DD (SEQ ID NOS: 1-2) active site motif of the RT, as well as at the branching guanosine, abolished activity, indicating that the defense function depends on msDNA synthesis (FIGS. **19**B and **29**C). Furthermore, perturbations to the msDNA also abolished activity (FIG. **31**), suggesting that its structure, and not simply formation, is essential for the defense activity. Indeed, a single nucleotide mismatch in the msDNA hairpin reduced activity by 100-1000 fold, but introducing a second mutation on the complementary strand to restore the structure of the msDNA also restored wild-type activity (FIG. **31**). Notably, these retrons are associated with other domains, including TOPRIM (topoisomerase-primase) (29), TIR (30), a nucleoside deoxyribosyltransferase-like enzyme, and the Septu defense system (4), all of which play a role for activity (FIG. **19**B).

[0337] Additional Molecular Functions of Defense Systems

[0338] Applicants investigated several additional systems with diverse components (FIGS. **20**, **32**A-**32**B). These include a three-gene system containing a von Willebrand factor A (vWA) metal ion binding protein, a PP2C-like serine/threonine protein phosphatase, and a serine/threonine protein kinase that provided strong protection against T7-like phages (T3, T7, and φV-1). This system, dubbed TerY-phosphorylation triad (TerY-P), has been previously analyzed computationally in the context of tellurite resistance-associated stress response and might operate as a phosphorylation switch that couples the activities of the kinase and the phosphatase (31).

[0339] Additional systems include proteins containing a SIR2 (sirtuin) deacetylase domain that is also present in the recently-discovered Thoeris system (4) and has also been detected in the same neighborhoods with prokaryotic Argonaute proteins (32); ApeA, a predicted HEPN-family abortive infection protein (33) and a putative ancestor of the type VI CRISPR effector Cas13; a ~1300 residue P-loop ATPase containing an unusual insertion of two transmembrane helices into the ATPase domain, similar to the KAP ATPases (34); and a four-gene cassette containing a 7-cyano-7-deazaguanine synthase-like protein (QueC), suggestive of small molecule biosynthesis. All of these components are essential for defense activity (FIG. **20**).

[0340] Finally, Applicants also demonstrated defense functions for several predicted NTPases of the STAND (signal transduction ATPases with numerous associated domains) superfamily (FIG. **20**). This expansive superfamily comprise multidomain proteins that include eukaryotic ATPases and GTPases involved in programmed cell death and various forms of signal transduction (35, 36). Typically, STAND NTPases contain a C-terminal helical sensor domain that, upon target recognition, induces oligomerization via ATP or GTP hydrolysis, leading to activation of the N-terminal effector domain. The role of the STAND NTPases in prokaryotes has long remained enigmatic (35, 37); the few for which experimental data are available contain a helix-turn-helix domain and have been shown to regulate transcription (36). Several STAND NTPases were active against dsDNA phages (FIG. **17**D); these proteins contained different putative effector domains, including DUF4297 (a putative PD-(D/E)×K-family nuclease), an Mrr-like nuclease, SIR2, a trypsin-like serine protease, and an uncharacterized helical domain. Applicants named these systems antiviral ATPases/NTPases of the STAND superfamily (AVAST). As homologs of essential eukaryotic programmed cell death effectors, AVAST systems are likely to function via an abortive infection mechanism, i.e. by causing growth arrest or programmed cell death in infected hosts.

[0341] These findings substantially expanded the space of protein domains, molecular functions, and interactions that are employed by bacteria and archaea in antiviral defense. Some of these functions, including RNA editing, have not been previously implicated in defense mechanisms. The high success rate of defense system prediction based on the evolutionary conservation of their proximity to previously identified defense genes supported the defense island concept (4, 7, 10) and demonstrated its growing utility at the time of rapid expansion of sequence databases. Furthermore, the computational approach implemented in this work provided for a substantial expansion of the range of the identified putative defense systems. Many of these previously unknown defense systems contain enzymatic activities as well as predicted sensor components that potentially could be engineered for novel biotechnology applications.

[0342] Despite similarities in domain architectures among some of the identified defense systems, their phage specificities differ significantly, emphasizing the importance of multiple defense mechanisms for the survival of prokaryotes in the arms race against viruses. These observations are compatible with the concept of distributed microbial immunity, according to which defense systems encoded in different genomes collectively protect microbial communities from the diverse viromes they confront (38). Additionally, several of the identified defense systems incorporate molecular functions from typically non-defense sources, highlighting the versatility of activities that are recruited for antiviral defense. These include the RADAR deaminase, nitrilases, and reverse transcriptases of different families, including retrons. The demonstration of defense functions for multiple RTs, which are generally associated with mobile genetic elements, is consistent with the 'guns for hire' paradigm whereby enzymes are shuttled between MGEs and defense systems during microbial evolution (8). Finally, most of these defense systems do not appear to be substantially enriched within prophages, suggesting that they are dedicated host defense genes, rather than virus superinfection exclusion modules (FIGS. **33**A-**33**C and Methods).

[0343] The overall patchy pattern of phage specificity observed for the different defense systems was unexpected. In some cases, the same system exhibited widely varying levels of protection against similar phages; for instance, DRT type 3 offered full protection against phage T2 but no protection against phage T4, which is ~98% identical to T2.

[0344] The range of domains contained within these systems indicates that they employ diverse biochemical activities. The identification of these defense systems, as well as others Applicants have predicted computationally, provides a foundation for mechanistic investigation.

[0345] The results described here have broad implications for understanding antiviral resistance and host-virus dynamics in natural populations of microbes, as well as for technological applications such as the development of antibacterial therapeutics, DNA and RNA editing, molecular detection, and targeted cell destruction.

[0346] 1. C. A. Suttle, Viruses: unlocking the greatest biodiversity on Earth. Genome 56, 542-544 (2013).

[0347] 2. A. G. Cobián Güemes et al., Viruses as Winners in the Game of Life. Annu Rev Virol 3, 197-214 (2016).

[0348] 3. F. Hille et al., The Biology of CRISPR-Cas: Backward and Forward. Cell 172, 1239-1259 (2018).

[0349] 4. S. Doron et al., Systematic discovery of antiphage defense systems in the microbial pangenome. Science 359, (2018).

[0350] 5. J. E. Samson, A. H. Magadan, M. Sabri, S. Moineau, Revenge of the phages: defeating bacterial defences. Nat Rev Microbiol 11, 675-687 (2013).

[0351] 6. J. Bondy-Denomy, A. Pawluk, K. L. Maxwell, A. R. Davidson, Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. Nature 493, 429-432 (2013).

[0352] 7. K. S. Makarova, Y. I. Wolf, E. V. Koonin, Comparative genomics of defense systems in archaea and bacteria. Nucleic Acids Res 41, 4360-4377 (2013).

[0353] 8. E. V. Koonin, K. S. Makarova, Y. I. Wolf, M. Krupovic, Evolutionary entanglement of mobile genetic elements and host defence systems: guns for hire. Nat Rev Genet, (2019).

[0354] 9. G. Faure et al., CRISPR-Cas in mobile genetic elements: counter-defence and beyond. Nat Rev Microbiol 17, 513-525 (2019).

[0355] 10. K. S. Makarova, Y. I. Wolf, S. Snir, E. V. Koonin, Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. J Bacteriol 193, 6039-6056 (2011).

[0356] 11. S. A. Shmakov, K. S. Makarova, Y. I. Wolf, K. V. Severinov, E. V. Koonin, Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. Proc Natl Acad Sci USA 115, E5307-E5316 (2018).

[0357] 12. S. A. Shmakov et al., Systematic prediction of functionally linked genes in bacterial and archaeal genomes. Nat Protoc 14, 3013-3031 (2019).

[0358] 13. J. Gordeeva et al., BREX system of *Escherichia coli* distinguishes self from non-self by methylation of a specific DNA site. Nucleic Acids Res 47, 253-265 (2019).

[0359] 14. T. Goldfarb et al., BREX is a novel phage resistance system widespread in microbial genomes. EMBO J 34, 169-183 (2015).

[0360] 15. R. Odegrip, A. S. Nilsson, E. Haggard-Ljungquist, Identification of a gene encoding a functional reverse transcriptase within a highly variable locus in the P2-like coliphages. J Bacteriol 188, 1643-1647 (2006).

[0361] 16. A. M. Burroughs, D. Zhang, D. E. Schïffer, L. M. Iyer, L. Aravind, Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. Nucleic Acids Res 43, 10633-10654 (2015).

[0362] 17. K. S. Makarova, L. Gao, F. Zhang, E. V. Koonin, Unexpected connections between type VI-B CRISPR-Cas systems, bacterial natural competence, ubiquitin signaling network and DNA modification through a distinct family of membrane proteins. FEMS Microbiol Lett 366, (2019).

[0363] 18. C. D. Rae, Y. Gordiyenko, V. Ramakrishnan, How a circularized tmRNA moves through the ribosome. Science 363, 740-744 (2019).

[0364] 19. S. Zimmerly, L. Wu, An Unexplored Diversity of Reverse Transcriptases in Bacteria. Microbiol Spectr 3, MDNA3-0058-2014 (2015).

[0365] 20. N. Toro, R. Nisa-Martinez, Comprehensive phylogenetic analysis of bacterial reverse transcriptases. PLoS One 9, e114083 (2014).

[0366] 21. K. K. Kojima, M. Kanehisa, Systematic survey for novel types of prokaryotic retroelements based on gene neighborhood and protein architecture. Mol Biol Evol 25, 1395-1404 (2008).

[0367] 22. D. M. Simon, S. Zimmerly, A diversity of uncharacterized reverse transcriptases in bacteria. Nucleic Acids Res 36, 7219-7229 (2008).

[0368] 23. H. C. Pace, C. Brenner, The nitrilase superfamily: classification, structure and function. Genome Biol 2, REVIEWS0001 (2001).

[0369] 24. A. J. Simon, A. D. Ellington, I. J. Finkelstein, Retrons and their applications in genome engineering. Nucleic Acids Res 47, 11007-11019 (2019).

[0370] 25. F. Farzadfard, T. K. Lu, Synthetic biology. Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. Science 346, 1256272 (2014).

[0371] 26. B. C. Lampson et al., Reverse transcriptase in a clinical strain of *Escherichia coli*: production of branched RNA-linked msDNA. Science 243, 1033-1038 (1989).

[0372] 27. D. Lim, W. K. Maas, Reverse transcriptase-dependent synthesis of a covalently linked, branched DNA-RNA compound in *E. coli* B. Cell 56, 891-904 (1989).

[0373] 28. T. M. Lima, D. Lim, A novel retron that produces RNA-less msDNA in *Escherichia coli* using reverse transcriptase. Plasmid 38, 25-33 (1997).

[0374] 29. L. Aravind, D. D. Leipe, E. V. Koonin, Toprim—a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. Nucleic Acids Res 26, 4205-4213 (1998).

[0375] 30. S. Horsefield et al., NAD. Science 365, 793-799 (2019).

[0376] 31. V. Anantharaman, L. M. Iyer, L. Aravind, Ter-dependent stress response systems: novel pathways related to metal sensing, production of a nucleoside-like metabolite, and DNA-processing. Mol Biosyst 8, 3142-3165 (2012).

[0377] 32. K. S. Makarova, Y. I. Wolf, J. van der Oost, E. V. Koonin, Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. Biol Direct 4, 29 (2009).

[0378] 33. V. Anantharaman, K. S. Makarova, A. M. Burroughs, E. V. Koonin, L. Aravind, Comprehensive analysis of the HEPN superfamily: identification of novel roles in intra-genomic conflicts, defense, pathogenesis and RNA processing. Biol Direct 8, 15 (2013).

[0379] 34. L. Aravind, L. M. Iyer, D. D. Leipe, E. V. Koonin, A novel family of P-loop NTPases with an unusual phyletic distribution and transmembrane segments inserted within the NTPase domain. Genome Biol 5, R30 (2004).

[0380] 35. D. D. Leipe, E. V. Koonin, L. Aravind, STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. J Mol Biol 343, 1-28 (2004).

[0381] 36. O. Danot, E. Marquenet, D. Vidal-Ingigliardi, E. Richet, Wheel of Life, Wheel of Death: A Mechanistic Insight into Signaling by STAND Proteins. Structure 17, 172-182 (2009).

[0382] 37. E. V. Koonin, L. Aravind, Origin and evolution of eukaryotic apoptosis: the bacterial connection. Cell Death Differ 9, 394-404 (2002).

[0383] 38. A. Bernheim, R. Sorek, The pan-immune system of bacteria: antiviral defence as a community resource. Nat Rev Microbiol 18, 113-119 (2020).

[0384] 39. D. Hyatt et al., Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11, 119 (2010).

[0385] 40. M. Punta et al., The Pfam protein families database. Nucleic Acids Res 40, D290-301 (2012).

[0386] 41. A. Marchler-Bauer et al., CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res 45, D200-D203 (2017).

[0387] 42. M. Steinegger, J. Soding, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol 35, 1026-1028 (2017).

[0388] 43. M. Steinegger, J. Soding, Clustering huge protein sequence sets in linear time. Nat Commun 9, 2542 (2018).

[0389] 44. R. J. Roberts, T. Vincze, J. Posfai, D. Macelis, REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res 43, D298-299 (2015).

[0390] 45. D. Cohen et al., Cyclic GMP-AMP signalling protects bacteria against viral infection. Nature, (2019).

[0391] 46. G. Ofir et al., DISARM is a widespread bacterial defence system with broad anti-phage activities. Nat Microbiol 3, 90-98 (2018).

[0392] 47. K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30, 3059-3066 (2002).

[0393] 48. L. Zimmermann et al., A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. J Mol Biol 430, 2237-2243 (2018).

[0394] 49. J. C. Petricciani, F. C. Chu, J. B. Johnson, H. M. Meyer, Bacteriophages in live virus vaccines. Proc Soc Exp Biol Med 144, 789-792 (1973).

[0395] 50. J. B. Milstien, J. R. Walker, J. C. Petricciani, Bacteriophages in live virus vaccines: lack of evidence for effects on the genome of rhesus monkeys. Science 197, 469-470 (1977).

[0396] 51. B. Xu, X. Ma, H. Xiong, Y. Li, Complete genome sequence of 285P, a novel T7-like polyvalent *E. coli* bacteriophage. Virus Genes 48, 528-533 (2014).

[0397] 52. S. Picelli et al., Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res 24, 2033-2040 (2014).

[0398] 53. E. S. Miller et al., Bacteriophage T4 genome. Microbiol Mol Biol Rev 67, 86-156 (2003).

[0399] 54. D. H. Turner, D. H. Mathews, NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. Nucleic Acids Res 38, D280-282 (2010).

[0400] 55. Y. Zhou, Y. Liang, K. H. Lynch, J. J. Dennis, D. S. Wishart, PHAST: a fast phage search tool. Nucleic Acids Res 39, W347-352 (2011).

[0401] 56. D. Arndt et al., PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res 44, W16-21 (2016).

[0402] 57. J. Strecker et al., RNA-guided DNA insertion with CRISPR-associated transposases. Science 365, 48-53 (2019).

[0403] 58. S. E. Klompe, P. L. H. Vo, T. S. Halpin-Healy, S. H. Sternberg, Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration. Nature 571, 219-225 (2019).

[0404] 59. E. V. Koonin, K. S. Makarova, Y. I. Wolf, Evolutionary Genomics of Defense Systems in Archaea and Bacteria. Annu Rev Microbiol 71, 233-261 (2017).

[0405] 60. S. Yamamoto, K. Kiyokawa, K. Tanaka, K. Moriguchi, K. Suzuki, Novel toxin-antitoxin system com-

posed of serine protease and AAA-ATPase homologues determines the high level of stability and incompatibility of the tumor-inducing plasmid pTiC58. J Bacteriol 191, 4656-4666 (2009).

[0406] Materials and Methods

[0407] Detection of known defense systems. All bacterial and archaeal genomes (n=174,080) were downloaded from Genbank (NCBI) in November 2018. For genomes where gene annotations were incomplete or missing, genes were predicted using Prodigal (39). Known defense-related protein domains were annotated using RPSBLAST version 2.8.1 and the set of position-specific scoring matrices curated from the NCBI Conserved Domain Database (CDD) (4, 10, 40, 41). To reduce the false positive rate, a multi-gene system containing a ubiquitous protein domain was required to include two or more of its component genes in close proximity. For example, the type I restriction-modification endonuclease hsdR was called as a defense gene only if the corresponding methylase (hsdM) or specificity protein (hsdS) was also encoded in the vicinity. Genes were predicted for known defense systems including HsdRMS, McrBC, BREX, Druantia, Zorya, Wadjet, Thoeris, Hachiman, Lamassu, Gabjia, Septu, Shedu, Kiwa, pAgo, and other RM systems. Toxin-antitoxin systems were excluded from the set of known systems due to their overall low enrichment within defense islands (FIGS. **21**A-**21**D).

[0408] Candidate novel defense genes. All translated protein-coding sequences within either 10 kb or 10 ORFs of known defense systems (whichever was greater), including the components of the known defense systems themselves, were compiled into a preliminary list (8.7×10^{6} genes), which was expected to consist of both defense and non-defense genes. Highly similar sequences (at least 98% sequence identity and coverage) were discarded using the linclust option in MMseqs2 (42, 43) with parameters—min-seq-id 0.98-c 0.98, resulting in a reduced list of 2.5×10^{6} sequences. These sequences were then further clustered using the cascaded clustering option in MMSeqs2, yielding a final list of 6.0×10^{5} representatives ("seeds").

[0409] Scoring candidate genes for defense enrichment. For each of the 6.0×10^{5} seeds, a "defense enrichment score" was computed as (number of homologs in proximity to one or more known defense systems)/(total number of homologs). A gene was considered to be located in proximity to a known defense system if it occurred no more than 5 kb or 5 ORFs away from the locus encoding that system. CRISPR-Cas systems were omitted from the defense score calculation due to their low defense island association (10). Candidate sequences with a defense enrichment score of 0.1 or higher were retained for subsequent analysis, with the exception of predicted mobilome components (such as transposons), which were discarded. This cut-off was chosen because more than 90% of the known defense genes scored higher than this value, whereas most mobilome, toxin-antitoxin, and other non-defense genes scored lower (FIGS. **16**B, **21**A-**21**D). To identify homologs of the candidate proteins, all 6.2×10^{8} proteins in Genbank were tabulated, and highly similar proteins (at least 98% sequence identity and coverage) were removed, resulting in a reduced list of 1.3×10^{8} proteins. Each seed sequence was then searched against this non-redundant protein sequence database using MMseqs2. To qualify as evidence of homology, the resulting alignments were required to have a minimum coverage of 70% and a maximum E value of 10^{-5} (parameters—coy-mode 0-c 0.7-e 0.00001). The set of identified homologs was further clustered at 90% sequence identity to perform stringent redundancy reduction. In order to accurately compute defense association frequencies, seeds with fewer than 50 homologs after redundancy reduction were discarded.

[0410] Filtering defense-enriched genes based on context diversity. To select for genes that are likely to encode components of independent defense modules, defense-enriched seeds were further required to have sufficient context diversity. For each seed, the number of homologs within 5 kb or 5 ORFs of different defense system categories was counted, and the seed was retained if the entropy of this list, defined as $\Sigma-p_i \ln p_i$, where p_i is the normalized frequency of category i, was at least 0.9. This value corresponds to halfway between 2 and 3 non-zero entries in the case of a uniformly distributed frequency vector. Seeds were further filtered based on the proportion of homologs next to predicted toxin-antitoxin/Abi, mobilome, and CRISPR-Cas genes (FIGS. **21**A-**21**D).

[0411] Refining the classification of putative defense genes. A total of 12,027 seeds passing filter was identified, consisting of both known and putative defense genes. To determine whether each gene was putative or known, the original classification was refined as follows. A list was compiled of the amino acid sequences of reported homologs of known systems, including 288,776 restriction-modification proteins from REBASE (44); 517 proteins for BREX (14); and 27,775 proteins for other recently-identified systems (4, 45, 46). This list was supplemented with additional curated homologs and, following redundancy reduction, searched against the putative defense seeds using MMseqs2. Seeds that matched one or more of these known defense genes (at least 70-80% coverage with a maximum E value of 10^{-5}) were labeled as known. A subset of labels were adjusted by an additional round of manual curation, resulting in a classification of 4,555 known and 7,472 putative defense genes.

[0412] Domain analysis of predicted defense genes. The 7,472 putative defense seeds were further analyzed with additional, more sensitive methods to assess their domain content. For each seed gene, a multiple sequence alignment (MSA) of its homologs was created using MAFFT (47). If the number of homologs was 1,000 or fewer, all homologs were included in the alignment; otherwise, 1,000 homologs were randomly selected for inclusion. MSAs were searched against the Pfam 32.0 database using HHpred (48), and domain predictions with at least 80% probability were retained. Of these 7,472 genes, 3,029 (41%) contained at least one pfam domain that has been reported to be defense-associated (4, 10, 45). Although some of these 3,029 proteins could be distant homologs of known defense proteins, many were included in this category because they contained ubiquitous pfam domains that are also employed by some known defense systems (in particular, AAA-family ATPases, helix-turn-helix (HTH) motifs, and (P)D–(D/E)×K-family nucleases); these are predicted to be uncharacterized defense genes. The remaining 59% either had no domain hits or contained only domains that were not in the set of defense-associated pfams.

[0413] From genes to defense systems. For each selected candidate defense protein, the gene neighborhoods of 30 homologs in proximity to known defense genes were randomly chosen and examined to identify conserved (predicted) operons that contained the seed and could be expected to constitute a minimal, intact defense system. Protein domains were predicted using HHpred, and the resulting prediction was used to infer the potential involvement of the respective proteins in the activity of the respective predicted defense system.

[0414] Estimation of defense system abundance. To estimate the abundance of each validated defense system in microbial genomes, Applicants downloaded n=205,214 genomes available in Genbank as of August 2019. For each defense system, initial protein sequence seeds encoded by the corresponding signature genes were taken from experimentally validated loci. Initial seeds were aligned and converted into HMM profiles. Applicants then used a constrained 2 iteration HMM profile search to generate highly specific HMM profiles and retrieve related systems as follows. Each ORF of size 150aa or greater, with one or more hits, was searched against all MINI profiles using HMMER3.1 and assigned to the profile that had the highest scoring match. For each system, ORFs with profile hits with less than 500 bp of intergenic distance on the same strand were grouped into candidate loci. For multi-protein systems, a putative locus was considered a hit if every signature gene profile for the system had a match in the locus with a bit score of at least 25. For single gene systems, a locus was considered a hit if the protein had a match to the system's single signature gene profile with a bit score of at least 50 and an alignment coverage of at least 70%. Signature proteins from the identified systems were separately clustered at 50% identity using MMseqs2 and subsequently aligned using MAFFT. The alignments were used to create a new set of signature gene profiles as input to the next iteration. For BREX and Type I RM, Applicants used preexisting pfam profiles for the signature genes in place of iterative MINI profile searching. The final abundance was calculated as the number of hits for the given system divided by the number of genomes (n).

[0415] Bacteria and phage strains. Phages T2, T3, T4, T5, T7, P1, λ, φV-1, M13, φX174, MS2, and Qβ, as well as host *E. coli* strains K-12 (ATCC25404) and C (ATCC13706), were obtained from the American Type Culture Collection (ATCC). The genome of phage φV-1, originally isolated from a measles vaccine (49, 50), was sequenced and found to be 92% similar to enterobacteria phage 285P, a T7-like phage (51).

[0416] Cloning. To facilitate experimental validation using coliphages, the source organism of each candidate defense system was chosen to be as phylogenetically similar as possible to *E. coli*, in particular, from other strains of *E. coli* whenever possible. Candidate defense systems were cloned into the low-copy plasmid pACYC184. When possible, genomic DNA from source organisms was obtained from ATCC, NCTC, or DSMZ, and the genes of interest were amplified with Q5 (New England Biolabs) or Phusion Flash (Thermo Scientific) polymerase, using primers with 5' ends homologous to the ends of the plasmid backbone. Plasmids were assembled using the NEBuilder HiFi DNA Assembly mix (New England Biolabs). When the source organism was not readily available from public culture collections, genes were chemically synthesized (GenScript). When possible, the native promoter was retained. For source organisms outside of Enterobacteriaceae, or when the candidate system was operonized with other upstream genes, the system was placed under a bla or lac promoter.

[0417] Sequence verification of plasmids. The full sequences of all plasmids were verified by high-throughput sequencing. To prepare sequencing libraries, 25-50 ng of each plasmid was mixed with purified Tn5 transposome loaded with Illumina adapters and incubated at 55° C. for 10 min in the presence of 5 mM MgCl2 and 10 mM TAPS buffer (52). The quantity of Tn5 was titrated to generate an average fragment size of ~100-400 bp. Tagmentation reactions were subsequently treated with 0.5 volumes of 0.1% sodium dodecyl sulfate for 5 min at room temperature and amplified with KAPA HiFi HotStart polymerase using primers containing 8 nt i7 and i5 index barcodes. Barcoded amplicons were sequenced on a MiSeq (Illumina) with at least 150 cycles for the forward read. Reads were aligned to the reference plasmid sequence by the Geneious read mapper, and error-free plasmids were retained for subsequent experiments.

[0418] Competent cell production. *E. coli* strains K-12 and C were cultured in ZymoBroth with 25 μg/mL chloramphenicol and made competent using Mix & Go buffers (Zymo) according to the manufacturer's recommended protocol.

[0419] Phage plaque assays. *E. coli* host strains were grown to saturation at 37° C. in Luria Broth (LB). To 10 mL top agar (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 7 g/L agar) was added chloramphenicol (final concentration 25 μg/mL) and 526 μL *E. coli* culture, and the mixture was poured on 10 cm LB-agar plates containing 25 μg/mL chloramphenicol. For phages T2, T4, T5, P1, λ, M13, MS2, and Qβ, dilutions of phage in phosphate buffered saline were spotted on the plates, and plaque counts were recorded after overnight incubation at 37° C. If individual plaques were too small to be counted, the most concentrated dilution at which no plaque formation was visible was recorded as having a single plaque. For phages T3, T7, φV-1, and φX174, a total of 3 of phage containing 5×10^{6} virions was spotted, and the area of the zone of lysis was measured after incubation at 37° C. for 68 hr. A total of 2-4 technical replicates was collected for each infection condition. Initial screening of defense system candidates was performed in *E. coli* K-12 (ATCC25404), excluding phage φX174 due to its inability to infect *E. coli* K-12; systems with observed defense activity were further tested as described above.

[0420] Phage cultivation. Phages T2, T3, T4, T7, φV-1, M13, φX174, MS2, and Qβ were propagated in liquid culture. The host *E. coli* strain for each phage was grown to an OD600 of 0.2 -0.4 at 37° C. in LB and infected with a slab of top agar containing phage plaque from a previous lysis. Cultures were grown overnight at 37° C. with 250 rpm agitation. Phages T5, P1, and λ were propagated by the double agar overlay method; after overnight incubation at 37° C., plaques were scraped in LB. For both liquid culture and double agar overlay, phage samples were centrifuged to pellet cellular debris, and the supernatant was filtered through with a 0.22 μm sterile filter.

[0421] Phage genome sequencing. DNA from phage φV-1 was isolated using QuickExtract DNA extraction solution (Epicentre) following the manufacturer's recommended protocol. After tagmentation and PCR amplification steps described earlier for plasmid sequence verification, the library was sequenced on a MiSeq with 200 cycles for the forward read and 110 cycles for the reverse read. Trimmed reads were assembled into contigs with SPAdes 3.13.0 using the—careful option, and contigs were subsequently scaffolded into a full genome using the genome sequence of enterobacteria phage 285P (51) as a reference.

[0422] Whole transcriptome sequencing. *E. coli* ATCC25404, containing either an empty vector or the candidate defense system, was grown to log phase in LB and diluted to an OD600 of 0.2. The culture was then split into two tubes, one of which was infected with phage T2 at an estimated MOI of 2. Both subcultures were incubated at 37° C. for 1 hr with 250 rpm agitation. RNA was extracted using TRIzol Reagent (Thermo Fisher Scientific) and treated with DNAse I, followed by a RiboMinus ribosomal RNA depletion kit (Thermo). Sequencing libraries were prepared using NEB Ultra II directional RNAseq library prep kit (New England Biolabs) and paired-end sequenced (2×75 cycles) with a NextSeq (Illumina). Adapter sequences were trimmed from sequencing reads using CutAdapt (with parameters—trim-n-q 20-m 20-a AGATCGGAAGAGC-A AGATCGGAAGAGC (SEQ ID NO: 472)), and trimmed reads were aligned to the *E. coli* MG1655 reference genome using the Geneious read mapper.

[0423] Phage fragmentation. Phage fragments were amplified from the genome of phage T2 by PCR, cloned into an ampicillin-resistant plasmid after an IPTG-inducible T7 promoter, and sequenced verified as previously described. Each fragment was then transformed into NovaBlue(DE3) *E. coli* expressing the *Citrobacter rodentium* RADAR system. Independent colonies for each fragments were grown to saturation at 37° C. in LB with 25 μg/mL chloramphenicol and 100 μg/mL ampicillin. Cultures were then diluted 1 to 5 in the same media, and IPTG was added to a final concentration of 0.5 mM. After 4 h growth at 37° C., cells were pelleted by centrifugation, and total RNA was extracted by a Direct-zol RNA purification kit (Zymo). The *E. coli* tmRNA was subsequently amplified by RT-PCR (QuantBio) and sequenced with a MiSeq (Illumina).

[0424] *E. coli* growth kinetics. Cells were grown to log phase in LB and diluted to an OD600 of 0.2. Cultures were infected with phage T2 at varying MOI at grown at 37° C., and the OD600 was measured every 2 min for a total duration of 4 hr on a Synergy Neo2 plate reader (BioTek).

[0425] Classification of phage genes. Phage T2 genes were classified as putative early, middle, or late genes based on the closest promoter on the same strand, as annotated based on the genome of phage T4 (53). Genes that could not be unambiguously classified were labeled as unknown.

[0426] RNA secondary structure prediction. Minimum free energy RNA secondary structures were predicted using the Turner (2004) energy parameters at 37° C. (54).

[0427] Prophage analysis. Prophage and phage DNA sequences were downloaded from PHASTER (55, 56). All clusters (seed gene plus identified homologs) with hits matching the experimentally validated systems, as well as one cluster matching the rexA gene of phage lambda as a positive control, were searched against the PHASTER database with tblastn for near identical matches (≥95% identity). For each cluster, phage association frequency was calculated as the number of proteins in the cluster with unique matches to the PHASTER database divided by the total number of unique proteins in the cluster (number of proteins after clustering at 90% sequence identity). The cutoff for frequent phage association of a system was defined as half of the frequency for rexA. Applicants note that PHASTER does not predict all instances of prophages and prophage remnants, and Applicants have also considered an alternative approach of identifying prophage association based on proximity to integrases, which may allow a greater number of prophages to be identified. However, a challenge with the latter approach is that defense islands often appear to derive from mobile genetic elements other than prophages and contain many integrases that originate from non-phage sources (e.g., CRISPR-associated transposases (57, 58)), leading to a high rate of false positives. The use of PHASTER provided the advantage of substantially reducing the false positives that would otherwise be expected for an approach based on integrase association.

[0428] Computational analysis of the RT (UG1) nitrilase domain. Homologs of the RT (UG1) defense gene were identified with a PSIBLAST search seeded on the experimentally validated sequence (WP_115196278.1), and highly similar homologs (≥90% identity) were removed. An MSA of the nitrilase domain was then created using MAFFT, and a custom position-specific scoring matrix (PSSM) was derived from this alignment. Bacterial and archaeal proteins in Genbank (redundancy-reduced at 98% sequence identity and coverage) were then searched against this profile with RPSBLAST, and the E-values of proteins with a match covering a minimum of 20% of the length of the profile were recorded. Known nitrilase enzymes were identified using a separate RPSBLAST search against the same set of Genbank proteins using 36 PSSMs from the CDD database (E-value≤10^{-6}; minimum 40% profile coverage): cd07197, cd07564, cd07565, cd07566, cd07567, cd07568, cd07569, cd07570, cd07571, cd07572, cd07573, cd07574, cd07575, cd07576, cd07577, cd07578, cd07579, cd07580, cd07581, cd07582, cd07583, cd07584, cd07585, cd07586, cd07587, COG0388, pfam00795, PLN02504, PLN02747, PLN02798, PRK10438, PRK13286, PRK13825, TIGR00546, TIGR03381, and TIGR04048.

[0429] Establishing an abi response. Abortive infection (abi) systems, which are based on altruistic cell suicide or dormancy (59), typically induce non-specific or deleterious biochemical activity targeting the host cell that also interferes with the phage reproduction cycle. Abi responses can be characterized through traditional assays such as efficiency of the center of infection (ECOI), adsorption, host survival, and one-step growth curve measurements. However, because the events of phage DNA injection and expression of toxic early genes are likely to be deleterious to an infected cell even if the production of progeny phages is ultimately suppressed, these assays may not be informative in terms of distinguishing between abi vs. non-abi mechanisms. An alternative approach to establishing the existence of an abi response is to identify the biochemical activity of the defense system, which Applicants have focused on for the RADAR system.

[0430] Gene knockouts vs. heterologous reconstitution. To further assess the feasibility of performing knockout experiments in the source bacterial strains for each defense system, Applicants performed analyses which suggested that different defense systems with overlapping phage specificities often co-occur. For instance, *E. coli* strain DSM5212 contains both BREX type I and Druantia type I (FIG. **2**D), both of which were included as positive controls; if BREX were to be knocked out in this strain, the presence of Druantia would likely ensure that its phage resistance profile across the 12 phages in Applicants' assay would remain unchanged. Similarly, the SIR2+HerA system from *E. coli* strain NCTC11129 primarily confers resistance to phage lambda (FIG. **2**D); the source strain NCTC11129 additionally contains BREX type I, which also confers resistance against phage lambda. Collectively, these observations suggested that the knockout of a single defense system may not be sufficient to make its corresponding source strain phage-sensitive, motivating the use of heterologous reconstitution as the primary assay for defense activity.

TABLE 9

List of validated defense systems and their domain architectures.

#	WT	Mutants	Type	Name	Domain Architecture*
1	FIG. 17D	FIG. 19B	Retron	Retron-TIR	RT_etron-TIR
2	FIG. 17D	FIG. 19B	Retron	Ec67	RT_retron-TOPRIM
3	FIG. 17D	FIG. 19B	Retron	Ec86	Nuc_deoxy + RT_retron
4	FIG. 17D	FIG. 29C	Retron	Ec78	RT_retron + ATPase_AAA + HNH
5	FIG. 17D	FIG. 19B	RT	DRT type 1	RT_UG1-nitrilase
6	FIG. 17D	FIG. 29A	RT	DRT type 2	RT_UG2
7	FIG. 17D	FIG. 19B	RT	DRT type 3	RT_UG3 + RT_UG8
8	FIG. 17D	FIG. 29B	RT	DRT type 4	RT_UG15
9	FIG. 17D	FIG. 19B	RT	DRT type 5	RT_UG16
10.A	FIG. 17D	FIG. 18B	RNA	RADAR	ATPase_AAA + ADA
10.B	FIG. 18B	FIG. 18B	RNA	RADAR	ATPase_AAA + ADA
11	FIG. 17D	FIG. 20	RNA	apeA	RNase_ApeA
12	FIG. 17D	FIG. 20	STAND	AVAST type 1	MBL + Protease_S1-ATPase_STAND
13	FIG. 17D	FIG. 20	STAND	AVAST type 2	ATPase_STAND
14	FIG. 17D	FIG. 20	STAND	AVAST type 3	Nuclease_DUF4297-ATPase_STAND
15	FIG. 17D	FIG. 20	STAND	AVAST type 4	Nuclease_Mrr-ATPase_STAND
16	FIG. 17D	FIG. 20	STAND	AVAST type 5	SIR2-ATPase_STAND
17	FIG. 17D	FIG. 20	Other	dsr1	SIR2-DUF4020
18	FIG. 17D	FIG. 20	Other	dsr2	SIR2
19	FIG. 17D	FIG. 20	Other	SIR2 + HerA	SIR2 + Helicase_HerA
20	FIG. 17D	FIG. 20	Other	DUF4297 + HerA	Nuclease_DUF4297 + Helicase_HerA
21	FIG. 17D	FIG. 20	Other	tmn	ATPase_AAA_TM
22	FIG. 17D	FIG. 20	Other	qatABCD	ATPase_AAA + QueC + DNase_TatD
23	FIG. 17D	FIG. 20	Other	hhe	HEPN_DUF4011-Helicase_SF1_Dna2-Nuclease_Vsr-DUF3320
24	FIG. 17D	—	Other	mzaABCDE	Ankyrin-sigma + ATPase_MutL + ATPase_AAA-Z1 + Nuclease_DUF4420 + AIPR
25	FIG. 17D	FIG. 20	Other	TerY-P	vWA + phosphatase_PP2C + STK-OB
26	FIG. 17D	FIG. 20	Other	upx	Nuclease_DUF1887
27	FIG. 17D	FIG. 20	Other	ppl	Phosphoesterase_PHP-ATPase_SMC
28	FIG. 17D	FIG. 20	Other	ietAS**	ATPase_AAA + Protease_S8
29	FIG. 17D	FIG. 20	Other	Restriction-like system	ATPase_DUF499 + DUF3780 + Methylase_DUF1156 + Nuclease_PLD-Helicase_HepA

*Dashes (-) indicated domain fusions and (+) represents separate proteins.

**ietAS is also a previously-described plasmid stabilization toxin-antitoxin system (60).

TABLE 10

Source organism strains of validated defense systems and controls.

#	Source Organism	Strain	Promoter	Codon	Genes	bp
BREX type I	*Escherichia coli*	DSM5212	Native	Native	6	13703
Druantia type I	*Escherichia coli*	DSM5212	Native	Native	5	11823
RT-Abi-P2	*Escherichia coli*	ECOR30	Native	Native	1	1921
1	*Shigella dysenteriae*	NCTC2966	Native	Native	1	2064
2	*Escherichia coli*	NCTC8623	Native	Native	1	2038
3	*Escherichia coli*	BL21	Native	Native	2	2188
4	*Escherichia coli*	ECONIH5	Native	Native	3	3551
5	*Klebsiella pneumoniae*	NCTC9143	Native	Native	2	4451
6	*Salmonella enterica*	NCTC8273	Native	Native	1	1780
7	*Escherichia coli*	ECOR12	Native	Native	2	4995
8	*Escherichia coli*	21-C8-A	Native	Human	1	1838
9	*Escherichia coli*	KTE25	Native	Native	1	1608
10.A	*Citrobacter rodentium*	DBS100	Native	Native	2	5526
10.B	*Pluralibacter gergoviae*	ATCC33028	Native	Native	3	6689
11	*Escherichia coli*	NCTC8008	Native	Native	1	1981
12	*Erwinia piriflorinigrans*	CFBP5888	bla	Native	3	7246
13	*Escherichia coli*	NCTC9087	Native	Native	1	5109
14	*Salmonella enterica*	NCTC13175	Native	Native	2	7175

TABLE 10-continued

Source organism strains of validated defense systems and controls.

#	Source Organism	Strain	Promoter	Codon	Genes	bp
15	*Escherichia coli*	NCTC11132	Native	Native	1	4964
16	*Escherichia coli*	NCTC13384	Native	Native	1	3411
17	*Escherichia coli*	NCTC9112	Native	Native	1	4212
18	*Cronobacter sakazakii*	NCTC8155	Native	Native	1	4329
19	*Escherichia coli*	NCTC11129	Native	Native	2	3308
20	*Escherichia coli*	NCTC11131	Native	Native	2	3419
21	*Escherichia coli*	ECOR25	Native	Native	1	4415
22	*Escherichia coli*	NCTC9009	Native	Native	4	5408
23	*Escherichia coli*	ATCC43886	Native	Native	1	5958
24	*Salmonella enterica*	NCTC5773	Native	Native	5	9416
25	*Citrobacter gillenii*	NCTC9094	Native	Native	3	3605
26	*Salmonella enterica*	NCTC6026	Native	Native	1	4100
27	*Escherichia coli*	NCTC8620	Native	Native	1	3066
28	*Escherichia coli*	ECOR52	Native	Native	2	3676
29	*Escherichia coli*	ECOR58	Native	Native	4	9809

TABLE 11

PCR primers used to amplify validated defense systems and controls.

#	dfd	Sequence
BREX type I	Fwd	gctaacttacattaattgcgttgcgcaACAGCACCACGTTCATCTTCC (SEQ ID NO: 14)
	Rev	ccaaggggttatgctagttattgcgGTTCATTAAAATAGTTACTACGTTAATTCACACCC (SEQ ID NO: 215)
Druantia type I	Fwd	gctaacttacattaattgcgttgcgcaGGTGAACGTTTGGTTGATAGGG (SEQ ID NO: 216)
	Rev	ccaaggggttatgctagttattgcgCTCAATGGGCATAATTTTACATTGTGC (SEQ ID NO: 217)
RT-Abi-P2	Fwd	gctaacttacattaattgcgttgcgcaACATCCCGTCATCATGCCATC (SEQ ID NO: 218)
	Rev	ccaaggggttatgctagttattgcgCTCCTCGGAATAGAATGTTATGTTCG (SEQ ID NO: 219)
1		Locus synthesized
2	Fwd	gctaacttacattaattgcgttgcgcaCGCGCTATCACGTAAAATAGGC (SEQ ID NO: 220)
	Rev	ccaaggggttatgctagttattgcgCGAAAAATCAGCCTTAGCGTTCATAAC (SEQ ID NO: 221)
3	Fwd	gctaacttacattaattgcgttgcgcaGCTCATGTTATGCATGTGCATG (SEQ ID NO: 222)
	Rev	ccaaggggttatgctagttattgcgATTAGGTCTTCGCTTTATTTAAAGGGTTC (SEQ ID NO: 223)
4		Locus synthesized
5	Fwd	gagctaacttacattaattgcgttgcgcaGTCCTTAAACACGACAAAACCTGTG (SEQ ID NO: 224)
	Rev	cccaaggggttatgctagttattgcgCGCAATGTAACACCCACCC (SEQ ID NO: 225)
6		Locus synthesized
7	Fwd	gctaacttacattaattgcgttgcgcaTCTCAACTTCCCCAAATGTCCG (SEQ ID NO: 226)
	Rev	cccaaggggttatgctagttattgcgTTAGCAAAATACGCCCACGAAGTC (SEQ ID NO: 227)
8		Locus synthesized
9		Locus synthesized
10.A	Fwd	gctaacttacattaattgcgttgcgcaGAGGATTTATGCACAAAATCCTGATGC (SEQ ID NO: 228)
	Rev	ccaaggggttatgctagttattgcgGATTTAATCTGTTGTTCCGAACGG (SEQ ID NO: 229)

TABLE 11-continued

PCR primers used to amplify validated defense systems and controls.

#	dfd	Sequence
10.B	Fwd Rev	gctaacttacattaattgcgttgcgcaTGTGGTTAGTTATCACAGCACTAACC (SEQ ID NO: 230) ccaaggggttatgctagttattgcgGTGTATAAGAATCCGAGACCGAAC (SEQ ID NO: 231)
11		Locus synthesized
12	Fwd Rev	ataaatgctcaataatattgaaaaaggaagagtATGGTAGCGATAAAAATGTATCCGGC (SEQ ID NO: 232) cccaaggggttatgctagttattgcgTCAATCCGTAGCCTCTTCATTCTCG (SEQ ID NO: 233)
13	Fwd Rev	gctaacttacattaattgcgttgcgcaGGGATTTCCACCACCTCCC (SEQ ID NO: 234) ccaaggggttatgctagttattgcgTGCATAGCAATGAAGATAAACGTG (SEQ ID NO: 235)
14	Fwd Rev	gctaacttacattaattgcgttgcgcaACAATTTTTTGCCATAAGACGCTTTC (SEQ ID NO: 236) ccaaggggttatgctagttattgcgCATTAGGACTAGTAGAAAAGTCTTGGG (SEQ ID NO: 237)
15	Fwd Rev	gctaacttacattaattgcgttgcgcaGCGCAGCTGACAAAGATTGAC (SEQ ID NO: 238) ccaaggggttatgctagttattgcgCGATAATAAAAAGGCTCCAATCCCTG (SEQ ID NO: 239)
16	Fwd Rev	gctaacttacattaattgcgttgcgcaACTAGCTAAGCAATAAGGGCG (SEQ ID NO: 240) ccaaggggttatgctagttattgcgCAATCTCCGAGGTGGCCC (SEQ ID NO: 241)
17	Fwd Rev	gctaacttacattaattgcgttgcgcaTATTTTGCGTAGCTAGAACGCAATC (SEQ ID NO: 242) ccaaggggttatgctagttattgcgTGGGTATTAGCTCATATCAGAACTAATACCC (SEQ ID NO: 243)
18	Fwd Rev	gctaacttacattaattgcgttgcgcaGTAAGACAAGGGTTGAGCAGGC (SEQ ID NO: 244) ccaaggggttatgctagttattgcgCAATGGTGGGCTGATTAATTAGATGAG (SEQ ID NO: 245)
19	Fwd Rev	gctaacttacattaattgcgttgcgcaTAGCTATTGTGACTATGCTAACCATATG (SEQ ID NO: 246) ccaaggggttatgctagttattgcgTTCAGTCTAAATACATACCTGTCGGG (SEQ ID NO: 247)
20	Fwd Rev	gctaacttacattaattgcgttgcgcaGTGCGCCTTATGTGATTACAACG (SEQ ID NO: 248) ccaaggggttatgctagttattgcgCTCTCAGCCTAATGATTCCAGAATAG (SEQ ID NO: 249)
21	Fwd Rev	gctaacttacattaattgcgttgcgcaACCGTGCTGGCATGTTTTTAC (SEQ ID NO: 250) ccaaggggttatgctagttattgcgAGGAAGATCCGTGACCAGGAG (SEQ ID NO: 251)
22	Fwd Rev	gctaacttacattaattgcgttgcgcaGAAATTATTTGGAATGGATGATGGCG (SEQ ID NO: 252) ccaaggggttatgctagttattgcgACTTCTACCTCCCTTTAGAAAAGTTAATG (SEQ ID NO: 253)
23	Fwd Rev	gctaacttacattaattgcgttgcgcaCGGATTGAATCTGTTTATGAAATTTGGCTG (SEQ ID NO: 254) ccaaggggttatgctagttattgcgCCGACAGTTGTCACTGTTCTTATTACC (SEQ ID NO: 255)
24	Fwd Rev	tgagctaacttacattaattgcgttgcgcaATGATGAAGATCACCTAAAATGATAGGTTG (SEQ ID NO: 256) cccaaggggttatgctagttattgcgCAGCTGTTAATTGTATATTGATGCGATGC (SEQ ID NO: 257)

TABLE 11-continued

PCR primers used to amplify validated defense systems and controls.

#	dfd	Sequence
25	Fwd	gctaacttacattaattgcgttgcgcaCGTGATGAATGAAGCGGCTAAATAC (SEQ ID NO: 258)
	Rev	ccaaggggttatgctagttattgcgGTAAATCCTCGGGAAAACACAGG (SEQ ID NO: 259)
26	Fwd	gctaacttacattaattgcgttgcgcaGGGCTGTTTGGTTGAATTAAAAATACG (SEQ ID NO: 260)
	Rev	ccaaggggttatgctagttattgcgCCTTGATTTAAAACTATCAGTAGTAGGAACG (SEQ ID NO: 261)
27	Fwd	gctaacttacattaattgcgttgcgcaGATGGACTGGTACTGTAGATTCACC (SEQ ID NO: 262)
	Rev	ccaaggggttatgctagttattgcgCAAAGACGCAGAGGCCATCAG (SEQ ID NO: 263)
28	Fwd	gctaacttacattaattgcgttgcgcaATAGAACGATGAAGGATGGAAGCTAC (SEQ ID NO: 264)
	Rev	ccaaggggttatgctagttattgcgTTGTATTTTGTTGTGTATGGGCGG (SEQ ID NO: 265)
29	Fwd	gctaacttacattaattgcgttgcgcaCGTGATTCAGTTCGCCAGAC (SEQ ID NO: 266)
	Rev	ccaaggggttatgctagttattgcgCACTCGAAATGGATACCCTGAG (SEQ ID NO: 267)

TABLE 12

Protein accession numbers of defense system components (proposed gene names underlined).

#	Gene	Name	Protein Accession
BREX type I	A	brxA	WP_085962535.1*
	B	brxB	WP_000566901.1
	C	brxC	WP_001019648.1
	D	pglX	WP_021524842.1
	E	pglZ	WP_001180895.1
	F	brxL	WP_001193074.1
Druantia type I	A	druA	WP_000549798.1
	B	druB	WP_001315973.1
	C	druC	WP_021520530.1
	D	druD	WP_000455180.1
	E	druE	WP_000608843.1
RT-Abi-P2	A	—	WP_047657908.1
1	A	—	WP_005025120.1*
2	A	Ec67	WP_000169432.1
3	A	—	WP_001034589.1
	B	Ec86	WP_001320043.1
4	A	Ec78	WP_001549208.1
	B	ptuA	WP_001549209.1
	C	ptuB	WP_001549210.1
5	A	drt1a	WP_115196278.1
	B	drt1b	WP_040189938.1
6	A	drt2	WP_012737279.1
7	A	drt3a	WP_087902017.1
	B	drt3b	WP_062891751.1
8	A	drt4	GCK53192.1
9	A	drt5	WP_001524904.1
10.A	A	rdrA	WP_012906049.1
	B	rdrB	WP_012906048.1
10.B	A	rdrA	WP_155731552.1
	B	rdrB	WP_064360593.1
	C	rdrD	WP_064360592.1
11	A	apeA	WP_000706972.1
12	A	avs1a	WP_023654314.1
	B	avs1b	WP_084007836.1*
12	C	avs1c	WP_023654316.1
13	A	avs2	WP_063118745.1
14	A	avs3a	WP_126523998.1
	B	avs3b	WP_126523997.1*

TABLE 12-continued

Protein accession numbers of defense system components (proposed gene names underlined).

#	Gene	Name	Protein Accession
15	A	avs4	WP_044068927.1
16	A	avs5	WP_001515187.1
17	A	dsr1	WP_029488749.1
18	A	dsr2	WP_015387030.1*
19	A	—	WP_021577683.1
	B	herA	WP_021577682.1
20	A	—	WP_016239654.1
	B	herA	WP_016239655.1
21	A	tmn	WP_001683567.1
22	A	qatA	STG85056.1
	B	qatB	STG85057.1
	C	qatC	STG85058.1
	D	qatD	STG85059.1
23	A	hhe	WP_032200272.1
24	A	mzaA	VEA06816.1*
	B	mzaB	VEA06814.1
	C	mzaC	VEA06812.1
	D	mzaD	VEA06810.1
	E	mzaE	VEA06808.1
25	A	terY	WP_115257868.1
	B	—	WP_115257869.1
	C	—	WP_115257870.1
26	A	upx	WP_060647174.1
27	A	ppl	STM52149.1
28	A	ietA	WP_000385105.1
	B	ietS	WP_001551050.1
29	A	—	WP_000860009.1
	B	—	WP_001044652.1
	C	—	WP_001207938.1
	D	—	WP_000985714.1

*Probable error in annotated protein start position corrected.

TABLE 13

Predicted protein domains within validated defense systems and controls. Transmembrane helices were predicted using TMHMM, and all other domains were predicted using HHpred.

ID	Gene	Residues	Domain	Representative HHpred Hit	Probability	Start	End
BREX type I	A	201	DUF1819	PF08849.11	100	6	189
	B	200	DUF1788	PF08747.11	100	65	187
	C	1213	ATPase	PF07693.14	96.66	43	348
			DUF499	PF04465.12	99.88	247	846
	D	1201	Methyltransferase	PF02384.16	99.7	210	622
	E	865	PglZ	PF08665.12	99.12	474	650
	F	694	Lon protease	PF13337.6	100	30	484
			Lon protease	PF05362.13	99.9	486	693
Druantia type I	A	404	DUF4338	PF14236.6	99.92	45	339
	B	548	CoiA	PF06054.11	99.77	1	182
	C	627	Macoilin	PF09726.9	96.72	167	323
	D	347	(none)	—	—	—	—
	E	1836	Helicase	PF00270.29	98.45	99	388
			Helicase	5V9X_A	97.55	1071	1208
			DUF1998	PF09369.10	98.92	1626	1710
RT-Abi-P2	A	515	RT	PF00078.27	99.09	68	291
1	A	542	RT	PF00078.27	99.43	105	309
			TIR	PF13676.6	97.91	411	536
2	A	586	RT	PF00078.27	99.45	48	262
			TOPRIM	cd01026	96.88	367	465
3	A	307	Nuc_deoxy	PF15891.5	96.04	29	128
	B	320	RT	PF00078.27	99.52	53	248
4	A	311	RT	PF00078.27	99.37	34	241
	B	550	ATPase	PF13175.6	99.8	64	432
	C	216	HNH	PF01844.23	97.57	43	85
5	A	1232	RT	PF00078.27	99.06	80	382
			Nitrilase	PF00795.22	98.89	953	1216
	B	144	Transmembrane	—	—	4	26
6	A	425	RT	PF00078.27	99.63	54	328
7	A	398	RT	PF00078.27	99.39	53	251
	B	667	RT	PF00078.27	98.96	63	323
8	A	540	RT	PF00078.27	99.12	67	296
9	A	494	RT	PF00078.27	99.14	59	263
10.A	A	851	ATPase	PF07693.14	99.6	33	364
	B	856	Adenosine deaminase	PF00962.22	99.52	166	831
10.B	A	907	ATPase	PF07693.14	99.48	29	349
	B	914	Adenosine deaminase	PF00962.22	97.63	789	901
	C	245	SLATT	PF18183.1	96.01	120	241
			Transmembrane	—	—	44	63
			Transmembrane	—	—	78	100
			Transmembrane	—	—	127	146
			Transmembrane	—	—	151	168
11	A	601	HEPN	PF18739.1	86.57	507	532
12	A	386	MBL-fold hydrolase	PF00753.27	98.79	8	324
	B	1935	Protease	PF02122.15	98.23	2	187
			ATPase	PF14516.6	99.36	204	535
	C	93	(none)	—	—	—	—
13	A	1484	ATPase	PF14516.6	98.93	316	643
14	A	2092	DUF4297	PF14130.6	98.41	8	223
			ATPase	PF14516.6	99.44	250	597
	B	207	(none)	—	—	—	—
15	A	1587	Mrr	PF13156.6	97.05	17	162
			ATPase	PF14516.6	99.07	204	476
16	A	769	SIR2	cd00296	99.26	22	244
			ATPase	PF14516.6	97.6	312	464
17	A	1275	SIR2	cd00296	99.44	21	253
			DUF4020	PF13212.6	98.39	1114	1268
18	A	1207	SIR2	cd00296	99.47	21	240
19	A	415	SIR2	cd00296	99.59	26	338
	B	610	HerA helicase	4D2I_B	100	10	608
20	A	394	DUF4297	PF14130.6	99.05	1	191
	B	571	HerA helicase	4D2I_B	100	7	568
21	A	1273	ATPase	PF07693.14	97.62	39	390
			Transmembrane	—	—	160	177
			Transmembrane	—	—	199	218
22	A	643	ATPase	PF07693.14	99.8	15	385
	B	274	(none)	—	—	—	—
	C	457	QueC	PF06508.13	99.67	150	369
	D	263	TatD DNase	PF01026.21	99.94	13	254

TABLE 13-continued

Predicted protein domains within validated defense systems and controls. Transmembrane helices were predicted using TMHMM, and all other domains were predicted using HHpred.

ID	Gene	Residues	Domain	Representative HHpred Hit	Probability	Start	End
23	A	1911	DUF4011	PF13195.6	99.81	33	308
			ATPase	PF13086.6	97.93	427	552
			Helicase	PF01443.18	97.82	1379	1636
			Endonuclease	PF18741.1	98.7	1683	1780
			DUF3320	PF11784.8	98.1	1841	1885
24	A	679	Ankyrin repeat	COG0666	99.52	10	188
			Sigma	COG1191	99.81	411	657
	B	500	MutL	COG0323	99.81	1	352
	C	952	ATPase	PF13872.6	97.51	117	349
			Z1	PF10593.9	100	437	672
	D	342	DUF4420	PF14390.6	100	9	317
	E	601	AIPR	PF10592.9	100	245	562
25	A	277	vWA	PF00092.28	98.93	14	203
	B	239	Phosphatase	PF00481.21	99.74	5	232
	C	561	Kinase	PF00069.25	100	34	296
			ssDNA-binding	PF01336.25	96.18	344	435
26	A	1272	DUF1887	PF09002.11	92.5	1105	1272
27	A	891	PHP	cd07436	99.36	4	238
			ATPase	PF13166.6	99.74	266	836
28	A	384	ATPase	PF13654.6	97.36	5	349
	B	754	Protease	PF00082.22	99.87	264	561
29	A	1022	ATPase	PF07693.14	96.47	49	312
			DUF499	PF04465.12	100	79	745
	B	195	DUF3780	PF12635.7	100	1	187
	C	945	DUF1156	PF06634.12	99	18	81
			Methyltransferase	PF01555.18	96.08	150	202
			Methyltransferase	PF01555.18	97.76	548	682
	D	907	PLD	cd09179	99.17	4	177
			Helicase	6BOG_B	100	218	865

TABLE 14

Sequence of vector backbone. Inserts were cloned between the HindIII and EcoRI restriction sites (underlined).

CCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATC

GGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTT

GCGC**AAGCTT**CTGCA**GAATTC**GCAATAACTAGCATAACCCCTTGGGGCCT

CTAAACGGGTCTTGAGGGGTTTTTTGCTGAAACCTCAGGCATTTGAGAAG

CACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCAATA

GACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGA

ATTTGCTTTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGG

CGTAGCACCAGGCGTTTAAGGGCACCAATAACTGCCTTAAAAAAATTACG

CCCCGCCCTGCCACTCATCGCAATACTGTTGTAATTCATTTAACATTCTG

CCGACATGGAAGCCATCACAGACGGCATGATGAACCTGAATCGCCAGCGG

CATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATAGTGAAAACGG

GGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAA

CTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTT

AGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATA

TGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAA

AACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATC

TABLE 14-continued

Sequence of vector backbone. Inserts were cloned between the HindIII and EcoRI restriction sites (underlined).

CCATATCACCAGCTCACCGTCTTTCATCGCCATACGGAACTCTGGATGAG

CATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGC

TTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGT

CTGGTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTT

TACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTTC

TCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATAC

GCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGT

GCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTA

TCAACAGGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCAC

AGGTATTTATTCGGCGCAAAGTGCGTCGGGTGATGCTGCCAACTTACTGA

TTTAGTGTATGATGGTGTTTTTGAGGTGCTCCAGTGGCTTCTGTTTCTAT

CAGCTGTCCCTCCTGTTCAGCTACTGACGGGGTGGTGCGTAACGGCAAAA

GCACCGCCGGACATCAGCGCTAGCGGAGTGTATACTGGCTTACTATGTTG

GCACTGATGAGGGTGTCAGTGAAGTGCTTCATGTGGCAGGAGAAAAAAGG

CTGCACCGGTGCGTCAGCAGAATATGTGATACAGGATATATTCCGCTTCC

TCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTGCGGCGAGCGGAAAT

TABLE 14-continued

Sequence of vector backbone. Inserts were cloned between the HindIII and EcoRI restriction sites (underlined).
GGCTTACGAACGGGGCGGAGATTTCCTGGAAGATGCCAGGAAGATACTTA
ACAGGGAAGTGAGAGGGCCGCGGCAAAGCCGTTTTTCCATAGGCTCCGCC
CCCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGCGAAAC
CCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGCGGCTCCCTCGT
GCGCTCTCCTGTTCCTGCCTTTCGGTTTACCGGTGTCATTCCGCTGTTAT
GGCCGCGTTTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGT
TCGCTCCAAGCTGGACTGTATGCACGAACCCCCCGTTCAGTCCGACCGCT
GCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGAAAGACATGCA
AAAGCACCACTGGCAGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCT
TGAAGTCATGCGCCGGTTAAGGCTAAACTGAAAGGACAAGTTTTGGTGAC
TGCGCTCCTCCAAGCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCAGAG
AACCTTCGAAAAACCGCCCTGCAAGGCGGTTTTTTCGTTTTCAGAGCAAG
AGATTACGCGCAGACCAAAACGATCTCAAGAAGATCATCTTATTAATCAG
ATAAAATATTTCTAGATTTCAGTGCAATTTATCTCTTCAAATGTAGCACC
TGAAGTCAGCCCCATACGATATAAGTTGTAATTCTCATGTTAGTCATGC
(SEQ ID NO: 268)

TABLE 15-A

Sequences of validated defense systems (sequences shown in Tables 15-B and C)

Row No.	#	Name	Description	Source Organism	Strain	bp	Gene	Gene Name	Accession	Residues
1	Control	BREX type I	—	*Escherichia coli*	DSM5212	13703	A	brxA	WP_085962535.1*	201
2							B	brxB	WP_000566901.1	200
3							C	brxC	WP_001019648.1	1213
4							D	pglX	WP_021524842.1	1201
5							E	pglZ	WP_001180895.1	865
6							F	brxL	WP_001193074.1	694
7	Control	Druantia type I	—	*Escherichia coli*	DSM5212	11823	A	druA	WP_000549798.1	404
8							B	druB	WP_001315973.1	548
9							C	druC	WP_021520530.1	627
10							D	druD	WP_000455180.1	347
11							E	druE	WP_000608843.1	1836
12	Control	RT-Abi-P2	—	*Escherichia coli*	ECOR30	1921	A		WP_047657908.1	515
13	1	—	Retron-TIR	*Shigella dysenteriae*	NCTC2966	2064	A		WP_005025120.1*	542
14	2	Ec67	Retron-TOPRIM	*Escherichia coli*	NCTC8623	2038	A	Ec67	WP_000169432.1	586
15	3	Ec86	Nuc_deoxy + retron	*Escherichia coli*	BL21	2188	A		WP_001034589.1	307
16							B	Ec86	WP_001320043.1	320
17	4	Ec78	Retron + ATPase + HNH	*Escherichia coli*	ECONIH5	3551	A	Ec78	WP_001549208.1	311
18							B	ptuA	WP_001549209.1	550
19							C	ptuB	WP_001549210.1	216
20	5	DRT type 1	RT-nitrilase (UG1)	*Klebsiella pneumoniae*	NCTC9143	4451	A	drt1a	WP_115196278.1	1232
21							B	drt1b	WP_040189938.1	144
22	6	DRT type 2	RT (UG2)	*Salmonella enterica*	NCTC8273	1780	A	drt2	WP_012737279.1	425
23	7	DRT type 3	RT (UG3) + RT (UG8)	*Escherichia coli*	ECOR12	4995	A	drt3a	WP_087902017.1	398
24							B	drt3b	WP_062891751.1	667
25	8	DRT type 4	RT (UG15)	*Escherichia coli*	21-C8-A	1838	A	drt4	GCK53192.1	540
26	9	DRT type 5	RT (UG16)	*Escherichia coli*	KTE25	1608	A	drt5	WP_001524904.1	494
27	10.A	RADAR	ATPase + deaminase	*Citrobacter rodentium*	DBS100	5526	A	rdrA	WP_012906049.1	851
28							B	rdrB	WP_012906048.1	856
29	10.B	RADAR	ATPase + deaminase	*Pluralibacter gergoviae*	ATCC33028	6689	A	rdrA	WP_155731552.1	907
30							B	rdrB	WP_064360593.1	914
31							C	rdrD	WP_064360592.1	245
32	11	apeA	ApeA (HEPN)	*Escherichia coli*	NCTC8008	1981	A	apeA	WP_000706972.1	601
33	12	AVAST type 1	MBL + protease-STAND	*Erwinia piriflorinigrans*	CFBP5888	7246	A	avs1a	WP_023654314.1	386
34							B	avs1b	WP_084007836.1*	1935
35							C	avs1c	WP_023654316.1	93
36	13	AVAST type 2	STAND	*Escherichia coli*	NCTC9087	5109	A	avs2	WP_063118745.1	1484
37	14	AVAST type 3	DUF4297-STAND	*Salmonella enterica*	NCTC13175	7175	A	avs3a	WP_126523998.1	2092
38							B	avs3b	WP_126523997.1*	207
39	15	AVAST type 4	Mrr-STAND	*Escherichia coli*	NCTC11132	4964	A	avs4	WP_044068927.1	1587
40	16	AVAST type 5	SIR2-STAND	*Escherichia coli*	NCTC13384	3411	A	avs5	WP_001515187.1	769

TABLE 15-A-continued

Sequences of validated defense systems (sequences shown in Tables 15-B and C)

Row No.	#	Name	Description	Source Organism	Strain	bp	Gene	Gene Name	Accession	Residues
41	17	dsr1	SIR2-DUF4020	*Escherichia coli*	NCTC9112	4212	A	dsr1	WP_029488749.1	1275
42	18	dsr2	SIR2	*Cronobacter sakazakii*	NCTC8155	4329	A	dsr2	WP_015387030.1*	1207
43	19	—	SIR2 + HerA	*Escherichia coli*	NCTC11129	3308	A		WP_021577683.1	415
44							B	herA	WP_021577682.1	610
45	20	—	DUF4297 + HerA	*Escherichia coli*	NCTC11131	3419	A		WP_016239654.1	394
46							B	herA	WP_016239655.1	571
47	21	tmn	Transmembrane ATPase	*Escherichia coli*	ECOR25	4415	A	tmn	WP_001683567.1	1273
48	22	qatABCD	ATPase + QueC + TatD	*Escherichia coli*	NCTC9009	5408	A	qatA	STG85056.1	643
49							B	qatB	STG85057.1	274
50							C	qatC	STG85058.1	457
51							D	qatD	STG85059.1	263
52	23	hhe	DUF4011-helicase-Vsr	*Escherichia coli*	ATCC43886	5958	A	hhe	WP_032200272.1	1911
53	24	mzaABCDE	MutL + Z1 + DUF + AIPR	*Salmonella enterica*	NCTC5773	9416	A	mzaA	VEA06816.1*	679
54							B	mzaB	VEA06814.1	500
55							C	mzaC	VEA06812.1	952
56							D	mzaD	VEA06810.1	342
57							E	mzaE	VEA06808.1	601
58	25	TerY-P	vWA + PP2C + STK-OB	*Citrobacter gillenii*	NCTC9094	3605	A	terY	WP_115257868.1	277
59							B		WP_115257869.1	239
60							C		WP_115257870.1	561
61	26	upx	DUF1887	*Salmonella enterica*	NCTC6026	4100	A	upx	WP_060647174.1	1272
62	27	ppl	Phosphoesterase-ATPase	*Escherichia coli*	NCTC8620	3066	A	ppl	STM52149.1	891
63	28	ietAS	ATPase + protease	*Escherichia coli*	ECOR52	3676	A	ietA	WP_000385105.1	384
64							B	ietS	WP_001551050.1	754
65	29	—	Restriction-like system	*Escherichia coli*	ECOR58	9809	A		WP_000860009.1	1022
66							B		WP_001044652.1	195
67							C		WP_001207938.1	945
68							D		WP_000985714.1	907

*Probable error in annotated protein start position corrected.

TABLE 15-B

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.		Cloned Sequence
1	Control	acagcaccacgttcatcttcctttttaactgattttacagagactttaatacagttaaaattta
2		tttcctgagctgtaatcgattaagttgatgcatttaatgggaatgatatagggtcatttccagtct
3		cacttatagaaatggctaaagcatgactctcgccaaaaccgtttatgtgttgtacataacgcgatc
4		atccctctcacaaattgccttttctcatggcatctcgcccggtcccccattacaatcacttttgt
5		tttttgcgagctgcattccagtcttcagagggttttcgatgattaaaaatgacaaggcatggata
6		ggagacttgctgggcggaccgctcatgagcagggaaagccgcgtcattgccgaactgttgctaacc
		gatcccgatgaacagacatggcaagagcaaattgttggccacaacattttacaagcctcttctcct
		aacaccgcaaaacgttacgcggcaacaatcaggcttcgcctgaacacgctggataaaagcgcgtgg
		acattgattgccgaaggtagtgaacgggaacgccaacaacttctgtttgtggctctgatgctacat
		tcgccggtagttaaggattttctggctgaagtggtgaacgatctgcgcaggcagttcaaggaaaag
		ttgcctggcaatagctggaacgaatttgtgaatagccaggttcgcctacatccggtactcgccagc
		tactcagattcatctattgcaaaaatgggaaacaatctggtgaaggcgcttgctgaagcgggttat
		gtggatacgccccgcagacgtaacctgcaggcagtttaccttttaccggaaactcaggcagtgtta
		cagcgcctgggacaacaggacttgatatctattctggagggaaaacggtgatagatcccgttcttg
		aatatcgcctgtctcaaatccagagtcgcattaacgaagatcgcttcctcaaaaataacggctccg
		gaaatgaaattggtttttggatctttgattatcccgcgcagtgcgaactgcaggtacgggagcatt
		tgaaatatctgctccggcatctggaaaaggaccataaatttgcctgtctgaatgtcttccaaatca
		tcatcgatatgctcaatgaacgcggccttttcgagcgcgtctgccagcaggaagtcaaagtgggta
		ctgagacgctgaaaaagcagcttgctggtccgttaaatcagaaaaagatcgctgattttatagcga
		aaaaagtcgatctggctgcccaggattttgtcattcttaccggcatgggcaacgcctggccattag
		tacgcggtcatgaactgatgagtgccttgcaggatgtcatggggttcaccccactgctgatgtttt
		atcctggcacctacagcgggtacaacctttccccgctcacagacaccggttcacaaaattattatc
		gcgctttcagactggtaccagatacgggacccgcagcaacattgaatcctcaatgaagagcataac
		aatgaatattgaacagatttttgaaaaacctctaaaacgaaatataaacggggtagtcaaagcaga
		gcaaaccgatgatgccagcgcgtacatcgagttagatgaatatgtcatcacccgcgaactggaaaa
		ccatcttcgccatttcttcgaatcctatgttcctgccactggcccggaacggatccgtatggaaaa
		caagatcggcgtatgggtttcaggcttcttcggttcaggtaaatcgcactttattaagattctttc
		ttatcttttatctaaccgcaaagttacacataacggtacggaacgtaatgcttactccttctttga

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.	Cloned Sequence
	agataaaatcaaagatgcattattccttgccgatattaacaaagcggtgcattacccgactgaagt cattctgttcaatattgattcgcgtgccaacgtagatgacaaagaagatgccattcttaaagtctt cctgaaagttttcaacgaacgcattggatactgcgctgattttccgcatattgcccatcttgagcg cgagctggataaacgcggtcagtatgaaacctttaaagccgcgtttgccgatatcaatggctcgcg ctgggaagacgagcgcgacgcttactacttcatcagcgatgacatggcacaagcattaagccaggc cacgcagcagagtcttgaatcctcccgccaatgggtggaacaactcgacaaaaacttcccgctgga tatcaataattttgccagtgggtaaaagagtggctggatgacaatggtaagaacatcctctttat ggtggatgaagtcggtcagttcattggcaaaaatacgcaaatgatgctgaagctgcagactattac tgaaaaccttggggtaatttgcggtggccgcgcatgggttatcgtgacttcgcaggccgatatcaa cgcggcaatcggtggtatgagcagtcgcgacggacaggacttctccaagatccaggggcgcttctc tacacgcctgcaactttccagctctaacacatcagaagttatccagaaacgtttgttggtaaagac tgacgaagcaaaagcggcactggcaaaagtgtggcaagagaaagccgatatcctgcgtaaccagct ggcttttgacactacaacaactactgcactacgtccttttaccagcgaagaagagttcgttgacaa ctacccgtttgtcccgtggcactatcagattctgcaaaaagtgtttgaatctattcggacgaaagg tgcagcgggtaaacaattggccatgggtgagcgttctcagctggaggcattccagacggcggcgca gcaaatctcagcgcaagggctggattctctggtgcctttctggcgcttctatgccgccattgagag cttcctggaacctgccgttagccgcaccatcactcaggcttgccagaatggcattcttgatgagtt cgatggcaacctgcttaaaacgctgttcctgatccgctatgtggaaacgctgaaaagcaccctgga taacctggtcacattgtctatcgataggatcgatgccgataaagttgagttgcgccgccgggtcga aaaaagtctcaacacgcttgaacgcctgatgctcattgcgcgcgttgaagataaatatgtgttcct gaccaacgaagagaaagagatcgaaaacgagatccgtaacgttgatgtcgatttctctgcgatcaa caaaaaactggcatcgatcatctttgatgacattctgaaaagccgtaaatatcgttatccggctaa caagcaagactttgatatcagccgcttcctgaacgggcatccattagacggcgcagtgcttaacga tctggtggtgaagatcctgacccctaaagatccgacttattcgttctataacagcgatgcgacctg tcgcccttatacgtcagaaggcgacggctgtattttgattcgtctgcccgaagagggccgtacctg gagcgatattgatttagtcgtccagactgaaaagttcctcaaagataacgccgggcaacgtccgga acaggcaaccctgctctcagaaaaagcgcgtgaaaacagcaaccgggaaaaattactccgtgttca gttggaatcactacttgcagaagcagacgtctgggcgattggcgaacgcttaccgaaaaaatcctc cacgccatcgaacattgtcgatgaagcctgccgttacgtgattgaaaacaccttcggcaagctgaa gatgctgcggccttttaacggtgacatctcccgtgaaattcatgcattactgacggttgagaacga caccgaactggatctcggtaacctcgaagagtccaaccccgacgccatgcgcgaggtagaaacctg gatcagcatgaatatcgaatacaataaacctgtgtatttacgcgatattctgaaccattttgcgcg tcgcccttatggctggcccgaagacgaagtgaaactgctagtagcccgtctggcctgcaaaggtaa attcagcttcagccagcaaaacaacaacgtcgagcgaaaacaggcgtgggagttatttaataacag ccgccgccatagcgaattgcgtctgcataaagttcgccgtcatgatgaagcgcaggtgcgtaaagc cgcgcaaaccatggctgacatcgctcagcagccgtttaacgaacgggaagagccggcgctggttga acatattcgtcaggtatttgaagagtggaagcaagagctgaacgtattccgcgccaaggcagaggg cggaaacaatccggggaaaaacgagattgaatccggtctgcgcctgcttaatgccattcttaatga gaaagaagattttgccctgatcgaaaaagtctcatcgctgaaagatgaacttctggatttcagcga agaccgtgaagatttggtcgacttctaccgtaagcaattcgccacctggcaaaaactgggtgctgc gctgaatggcagctttaaatctaaccgcagcgcgctggaaaaagacgccgcagcggttaaagcgct gggcgagctggaaagcatctggcaaatgccggaaccttataagcatctcaatcgcatcacgccgtt gattgaacaggtccagaacgtcaaccatcagttagtcgaacagcatcgccagcacgccctcgaacg cattgacgcccgcattgaggaaagccgtcaacgcttgctggaagcgcacgccacgtcggagctgca aaacagcgttctgctgccgatgcaaaaagccagaaaacgcgctgaagtcagccagtcgattccgga aattttggcggaacagcaagagacaaaagcgctgcaaatggatgcagataaaaagattaacctgtg gatcgacgagctgcgtaaaaagcaagaagcacaactccgggcagcaaatgaagctaaacgcgctgc cgactcagaacagacttatgttgtggtggaaaaaaccgttatccaaccggtaccgaaaaaaacgca tctggtgaatgtcgccagtgagatgcgtaatgccaccggtggtgaagttctggaaacgaccgaaca ggtggaaaaggcgctcgacacgttacgcacaacgctgctggccgtcattaaagcaggcgatcgcat tcgccttcagtaactcccatttcagggcagcactctgctgccctttgcaggattttctatgaatac caataacattaaaaaatatgccccacaggcccgtaacgacttccgcgatgcggtgatccagaagct aacgacgcttgggatcgctgcagataaaaaggcaatttgcagattgccgaggccgaaaccattgg cgagaccgtgcgttacggtcagtttgattacccgttatcgacccttccccgccgcgaacggctggt aaaacgcgcccgtgagcagggttttgaggtgctggttgagcactgcgcctacacctggtttaaccg cttatgtgcaattcgctatatggagctacacggttatcttgagcacggcttccgtatgttgtccca cccggagacgccgaccgcgtttgaggtgctggatcatgtgccggaagtggcagaagccctgctgcc ggaaaataaggcgcagctggttgaaatgaagcMccggtaatcaggacgaagccctgtaccgcgaac tgctgctggggcagtgccacgccctgcaccacgcgatgccgttcctgtttgaagcggtagatgacg aagcggaactgctgttgccggataacctgacccgtaccgactctattctgcgtgggctggttgatg atattccggaagaagactgggagcaggtagaggttatcggctggctgtatcagttctatatttcgg aaaagaaagatgccgtgattggcaaagtggtgaagagcgaagatattcctgccgccacccagctgt ttacgccaaactggattgtgcagtatctggtacaaaactccgttggccgccagtggttgcagacct acccggactcgccgctgaaagacaaaatggagtactacatcgagcctgcggaacaaacgccggaag tgcaggcgcagctggcggcgattaccccagccagcattgaacccgaaagtattaaagtgctcgacc cagcctgcggctccggtcatattttgattgaagcctataatgtgctgaaaaatatctacgaagagc gtggttatcgcgggcgtgatattccacaactgattctggaaaataatatttttggtcttgatatcg acgaccgcgcggcacagctttccggctttgcattattaatgatggcgcgtcaggatgaccgcagaa tatttacccgcgatgtacgtctgaatattgtctctttgcaggaaagcctgcatctggatatcgcca aactctggcagcaactgaatttccaccagcaggtacaaaccggcagtatgggggatatgtttgctg aaaataacgcgttaacccaaactgacagcgcagaatatcagctgctgatgcgcacgctgaaacgct ttgtgaatgcaaaaacgctgggctcactgattcaggtgccgcaggaagaagaagcggaactgaagg

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.	Cloned Sequence
	tattcctggacgcgttgtatcgcctggaacaggaaggcgatttccagcagaagacggcggcaaaag cgtttattccgtttattcagcaggcgtggattttagcgcagcgatatgatgcggtagtggcgaatc cgccgtatatggggggtaattatatggagacagaacttaagaatttcgtctcttcttactaccctc aaggaaaggcggatctttattcttcatttatggtcagattacttttacaattaaaagataatcgca ctttaagcctaatgacccccttacttggatgaatttatcatcatttgaagagctccgaaaaatta tacttacaaatttcagcattcagtcattagtacagcctgaatatcattcatttttgagtcagctt atgtcccaatttgtgcttttagcatttcaaataccccattaagctggaatgcaaatttttttgatt tatcagatttttatggagaaaaaaatcaagctccaaattttcagtatgcaattaaaaatgacaata aatgtcattggaaatataacagaatcaccacggactttctatgtactcccggatatatcattgctt actctctgcctgattctgcgttatcttgcttcaaaacatccaaaaaacttcatgatgtttgcaatc taaaacaaggattaattactggtgataatgaaagatacctaagattctggcatgaaatcagctata actctttcagtctcaatgaaaaaagaaaaaaaacaaaatggttcccatatcaaaaaggtggtgcat accgtaaatggtatggtaataatgattatgttgttgactgggagaatgatggttattccattaaaa acttttataatgacaaaggtaaattacgctcacgccctcaaaacatacaattttattgtaaagagg gtttaacatggacaagtttaactatttcgtcactatcgatgagatatgtaccaaatggatatattt ttgatgcaaaaggacctatgtgttttccgaaatcctctttggatatctggaatattcttggctatg cgaatagcaaagtaatagatatatttctcaaacaattagcgcccaccatggattattctcaagggc ctgttggaaatgtcccattcaaatttaacgatggtgatttgaacgagataataaaagaactcgtaa acattcacaaacgtgactgggatgaaaatgaaacatcttttgagtttaagagagatatgttggttc atttttcaagagatattaacactattaagggtagttttacactaaggcaaggggaaaataaaaaag cgattaacagaacaaaatttttagaagaaatgaataactctttctttataaattgctttaatctaa ctgatattttatctccagaaattgaactaaacaaaatcacgttaacgcatgcaactattgaaattg atattcaaaaaataatttcatatgcaataggctgccaaatgggacgttactcccttgatcgcgaag gtctggtatacgctcatgaaggcaataatggcttcgccgatcttgtcgccgaaggtgcttataaaa gcttcccggctgatagtgacggcattctgccgctaatggatgaagagtggtttgacgatgacgtca cctctcgcgtcaaggagtttatccgcaccgtttggggcgaagaatatttgcgcgaaaacctcgatt ttatagccgaagttctcaagcccaaaaaaggcgaatctgcgctggagaccattcgtcgctatcttt ccacccagttctggaaagatcatctgaaaatgtataaaaagcgtccaatctactggctattcagct ccggtaaagagaaagcgtttgagtgcttggtgtatctgcatcgctataacgatgccacgctgtcga gaatgcgtaccgaatatgtggtgccgctgctggcgcgttatcaggccaatattgatcgcctgaacg atcaacttgatgaggcttctggcggtgaatccacacgtctgaaacgcgaacgcgacagcctgatca aaaaattcagcgaactgcgcagctatgacgatcgcctgcgtcactatgctgatatgagaatcagta ttgatctcgacgatggcgttaaggttaactacggcaagtttggcgatctgctggcagatgtcaaag ccatcaccggcaatgccccagaggtgatctaaaccagacggcacgttctcctgttgccgggttctg cccggtggcaaataccaccgggaaacgcgccgctgctgacatttctccacctcacttcatgataaa atgcgccaccgtgtcaaaatctcctttcgcgttttggcgctttcttattcatcgtaacaacatgg gattgtgaacttgcaaaatcaggactttattgctggccttaaagctaaatttgccgaacatcgcat cgttttctggcacgatcccgataaacgttttattgaggaactggaacagctcaagcttgaaagcgt cacgctaatcaacatgacccacgagtcacagctggcggtaaaaaaacgcatcgagattgatgagcc agaacagcagttcctgctgtggttcccccatgatgcgccgcctcatgaacaagactggctgctgga tatccgcctttacagcagcgaattccatgccgattttgccgccatcaccctgaacacgctgggcat tccccagcttggcctgcgcgagcatattcagcgacgcaaggccttcttcagcactaaacgcacgca ggcgctgaaaaatctggcgacagaacaggaagatgaagcctcgctggataagaaaatgattgcggt gatcgctggcgcaaagaccgcgaaaaccgaagacattttgttcaacctgattacccagtacgttaa ccaacaaatagaagacgacagcgaactggaaaacacgcaggcgatgctgaaacgccacggtctgga ctcggtattgtgggaaatgctcaaccacgaaatgggctaccaggcagaggagccatcgctggaaaa cctgctcctgaaactgttttgtaccgatctctctgcccaggccgacccacagcagcgcgcctggct ggaaaaaaatgtcctgctgacgccatccggcagagcatctgccctggcatttatggtgacctggcg tgccgatcgtcgctataaagaggcttatgactactgcgctcagcaaatgcaggccgccctgcaccc ggaagatcattaccgactcagctcgccgtatgatttgcacgaatgcgaaaccaccctcagcatcga acaaaccattattcatgcgctggtaacacagctgctggaagagagcaccacgctcgatcgggaagc ctttaaaaaactgctctctgagcgccagagcaaatactggtgtcagacacaaccagagtattacgc catctatgacgcattgcgccaggctgagcggttgctgaacctgcgcaatcgccacatcgatggttt ccactaccaggacagcgccaccttctggaaagcctactgcgaagaactgttccgcttcgaccaggc ttatcgcctgtttaatgaatatgccttgctggttcacagcaaaggagcgatgatcctcaagagcct ggatgattatatcgaggcgctctacagcaactggtatctggcagagttaagccgtaactggaacga agtgctggaagcggaaaatcgtatgcaggcgtggcaaatccctggcgtgccgcgtcagcagaactt cttcaatgaggtggtgaagccacagttccaaaatccgcaaatcaaacgcgtgttcgtgataatttc cgatgccctgcgttatgaagtggcggaggagctggggaatcaaatcaataccgagaaacgctttac cgcagaactgcgctcgcagctcggcgtgctccccagctacacccaactgggaatggcggcattgct gccccatgaacaactttgctatcaacccggtaacggcgacatcgtttatgctgatgggctgtcgac ctcgggtattcctaaccgcgataccattctgaagaactataagggaatggcgataaaatcgaagga ccttctggagttaaaaaatcaggaagggcgagaccttattcgcgattacgaagtggtgtatatctg gcataacacgattgatgccactggcgacacggcatccacggaagataaaaccttcgaagcgtgccg cacggcggtggctgaactgaaagatttagtcaccaaggtgatcaaccgcctccacggcacacgcat tttgttacggcggatcacggtttcctgttccagcaacaggcgctttcggttcaggataaaaccac tctgcaaattaagccggaaaacaccatcaagaaccacaaacgctttattatcggccatcagcttcc cgccgatgattttgctggaaagggaaagtggcggataccgcaggcgtgagcgacaacagcgagtt cctgattccgaaagggatccagcgcttccatttctctggcggcgcgcgcttcgttcatggcggcac catgttgcaggaggtttgcgttccggtattgcagataaaagccctgcaaaaaaccgccgcagaaaa acagccacagcgccgcccggtggatattgtcgcttaccatccgatgattaagctagtgaacaatat cgataaagtgagcctgttgcagacgcatccggtgggcgaactttatgaaccgcgtatcctgaacat

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.		Cloned Sequence
		ttacattgtcgacaacgccaacaatgtggtctcgggcaaagagcgcatcagctttgacagtgataacaacaccatggaaaaacgcgtacgcgaagttacgctgaagctgattggcgctaacttcaaccgtcgcaatgagtactggttgatactggaagacgcacaaacggaaacggggtatcagaagtacccggtcattatcgatctggcgttccaggatgatttcttctaagtgaggcgatatgcaaacccatcatgatttacctgtttcaggcgtatccgcaggggaaattgcctccgagggttacgatctggacgccctgctgaaccagcattttgctggtcgcgtggtgcgtaaagatctcaccaagcaactcaaggaaggggcaaacgtcccggtgtatgtgcttgagtatctgctcggcatgtactgcgcctctgacgatgacgatgtggtcgagcaagggttgcaaaacgttaagcgtattctggctgataactatgtgcgcccggatgaagcggagaaagtgaagtcgctgatccgcgagcgtggttcgtacaaaatcatcgataaagtgtcggtgaaactgaaccagaaaaaagacgtttacgaagcccagctttctaacctcggcatcaaagacgcgctggtgccctcgcagatggttaaagacaacgagaagctactgacgggcggtatctggtgcatgattaccgtcaactatttctttgaagaagggcagaagacctcacccttctcattgatgacgctcaagcctatccagatgccgaatatggatatggaagaggtgttcgatgcgcgtaaacactttaaccgtgaccagtggatcgatgtgctgctgcgctcggtgggtatggagcccgccaatattgagcaacgcaccaaatggcaccttatcacccgtatgatcccgttcgtggagaacaactataacgtttgcgagctggggccgcgtggcaccggtaaaagccatgtgtataaagagtgttctcctaactccctgttagtttccggcgggcaaacgaccgttgccaacttgttctacaacatggccagtcgccagatcggcctggttggcatgtgggatgtggtagcgttcgacgaagtcgcggggatcactttcaaagataaagacggcgtgcaaatcatgaaagattacatggcgtcaggatctttttctcgcggcagagattcgattgaaggtaaagcgtcgatggttttcgtcggcaacatcaatcaaagcgtagagactctcgttaaaaccagccatttgctggcaccatttccgactgcgatgattgatacagcatttttcgaccgctttcatgcctatattcccggttgggaaatccccaaaatgcgcccggaattctttaccaaccgttacgggctgattacggattatctcgctgaatatatgcgcgaaatgcgcaaacgcagtttctctgatgcgattgataaattctttaagctgggtaacaacctcaaccagcgtgacgttattgccgttcgacgtaccgtgtcggggttgttaaaactcatgcatcccgatggcgcgtacagcaaagaagatgtgcgagtctgcctgacctatgcgatggaagttcgccgccgcgtgaaagagcaacttaaaaaactgggcggtctggagttcttcgatgtgaactttagctacatcgacaacgaaacgctggaagagtttttgtgagcgtaccggaacagggcggcagcgaacttattcctgccggaatgccaaagccgggtgttgtgcatctggtcactcaggcagaaagcggcatgaccgggctgtatcgttttgaaacacagatgactgccggtaatggtaagcatagtgtatcgggtctgggttcaaatacctccgcgaaagaagctatccgcgtcggtttcgattacttcaaaggcaatttgaatcgggtaagcgcggccgcgaaattctccgatcatgaatatcaccttcatgtcgttgaactgcataatactggcccaagcaccgcaaccagtcttgctgcgcttatcgctttatgttcgatattgctggcaaaaccggtgcaggaacagatggtggtgttgggcagtatgacgcttggtggggtaattaacccggtgcaggatcttgccgccagtttgcagctcgccttcgacagcggtgcaaaacgggttctgttgccgatgtcctcggctatggatattccaacggttccggcagagttatttaccaagtttcaggtgagtttttactcagacccggttgatgctgtttataaggcgctgggtgtgaattaacgtagtaactattttaatgaac(SEQ ID NO: 269)
7 8 9 10 11	Control	ggtgaacgtttggttgatagggtagtaaaactagtaatcatcctataattagctatattcgtggttattagattgaaaacagataacattaacaaaatctataaatcgatttgaatgatttttttcatcaatactgttgtaagctcctgctatcaaaagttttgcacacaatctataagctcccagaattgcttgtataaatgctatcattggcgctgtcccgatcgagggagcaaggaggggactctcttgtgccatgcgattaatcactggggctctaagtgaaatttagtgggactaaatactaattggaacgtgagataaaaatgcacaaatatccctctataatagttaatatcaaccttcgagaagccaaactgaaaaagaaggtacgtgagcatttacaatccttgggttttacaagatctgattctggagcgctccaggccccgggaaataccaaagatgtaatacgggctcttcatagttctcaacgagctgagcggatatttgcaaaccaaaagttcataacgctaagagcggcaaagcttattaaatttttcgcatccggcaatgaggtcattccggataagatttcaccggtacttgaacgtgtaaagtcaggaacctggcaaggagatctctttaggttagcagcattaacttggtccgtacctgtttcaagcggatttggaaggcgtctccggtatcttgtatgggatgaaagcaacggaaaattgatagggctgatcgcaattggtgaccctgtgttcaaccttgcagtccgagataatttgattgggtgggatactcatgccagaagttcccggcttgttaatttgatggatgcatacgtcctcggtgctcttccccttataatgccctgctgggaggaaaattaattgcatgtctgcttcgtagccgcgatctttatgatgactttgcaaaggtctatggtgataccgttggagtaatatctcaaaaaaagaaacaagcacgtcttttggctattacaacaacatcgtctatggggcgctcatcggtatataaccgtttaaagctggatggaattcaatatttaaaatcgattggatatacaggcggttgggggcattttcatatacctgatagcttgttcattgaattacgtgattacttacgtgatatggatcacgcttatgcagatcattatatgtttggtaatgggcctaactggcgtttacgtacaactaaggcagctttaaatgcactaggatttagagataatttgatgaagcatggaattcaacgtgaagtgtttatcagtcagctagcagaaaatgcaactagtattctgcaaacaggcaaaggtgaaccagatctaacctctttgctttctgctaaagagatagctgagtgtgcgatggcacgatggatggttccacgatcaattcgcaatccagaatatcggctttggaaagcaagagatctatttgattttattagtaatgactcgctaaactttcccccgtttgacgagatagcgaaaacagttgtctaatcttaactgaaggggagtaagtgaattacgctattgataagttcaccgggacactgatattagcagctcgagcaacgaaatatgctcaatatgtttgcccagtttgtaaaaaaggtgttaacctccgtaaagggaaggttataccccccatattttgctcatttgcccggacatggtacgtcagactgtgaaaattttgttcccggaaattctatcattgtcgaaactattaaaactatttcaaagcgatatatggatttgcgcttattgattcctgtcggaagtaatagtcgagagtggtcattagaattagtgttgccaacctgtaatttatgtagagcaaagataacgttagatgtaggaggcagaagccaaacgcttgatatgaggagtatggtaaagagtcgccagattggtgctgaattatcagtaaaatcttaccgtattgtttcatatagtggtgaaccagatccaaaatttgtaacagaagttgaaagagaatgcccaggtttaccttctgagggagcagcagttttcactgctttagggcgtggggcatcgaagggatttccacgagcacaagagttaagatgtactgaaacatttgcctttctttggcgacaccctgttgctccagattttcctgatgaattagaaataaaaagtttagctagtaaacagggatggaatttagctcttgttacaattcctgaagtcccttctgtggagagtatttcatggctaaaatcttttacataccttcctgttg

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.	Cloned Sequence
	ttcctgccagaacatctattacagcaatttggccgttcctaaatcaaaaaacaagtattaatcatg
	tcgaatgtgtttattctgacacaatattgttgtcaacaaatatggcaccaacatcatcagaaaatg
	ttggaccaactatgtacgcacaaggttcctctttattactttcagcggttggtgttgaaacatcac
	ctgctttcttcattctaaatcctggagaaaatgactttgtgggcgtttctggctcaattgagcagg
	acgtaaacttatttttttctttctataaaaaaaacgtttctgtacccagaaaatatccctcaatag
	atttggtttttactaagaggaataaagaaaagaccatcgtttccttacatcaaagaagatgcattg
	aagttatgatggaagcacgaatgtttggccataaattagaatacatgtctatgccttctggtgttg
	aaggagtggcaagaattcaaagacaaactgaaagtaatgttattaagttagtttctaatgatgaca
	ttgcagctcatgataagagcatgcggttactatctcctgttgcgttatctcaattatctgattgct
	tagcaaacttaacatgtcatgtagaaatagatttttaggtcttggtaaaatatttttacctggtt
	cttctatgctatcattagatgacgggaaatttattgaattatctcctaatcttcgctcacggatat
	taagttttatacttcaaatggggcacaccctccatggttttagtttaaataatgatttttattag
	ttgagaaattagtggatttgcagccggaaccacacttattaccgcattatagagcattggtaaaag
	aagttaagaccaatggatttgaatgtaaccgctttagataaggtgccttcgaatgagttaccaata
	tagccaagaggcaaaggaacggatctctaagttgggacaatccgaaattgttaactttatcaatga
	gatttctccaactttacgacgtaaagcttttggttgtttaccaaaagtaccgggattcagggcagg
	acatcccactgaaattaaagaaaaacagaaaagattgattgggtatatgttccagtcacatccttc
	ctctgaggagagaaaagcatggaaaagtttttctctttttttggcagttttgggctgaagagaaaat
	tgacaaatcatttagtatgattgataatttaggattaaaagaaaactctggctctatttttattag
	agagcttgctaaaaactttcctaaagttgctagagagaatatcgagcgcctgtttatctttagtgg
	gtttgctgatgatccagacgttataaatgcatttaacctttttcctcctgcagttgttcttgcccg
	cgatatcgtgattgatactcttccaattcgtttagatgagcttgaagcacgtattagtttaattgc
	cgataatgttgagaaaaaaaataatcatattaaagaacttgagttaaaaatagatgcttttccga
	acagtttgataattactttaataatgaaaagagcagtttaaaaataattaatgaactacaatcttt
	gataaactcagagactaaacaatctgatattgctaataaagctattgacgagctttatcattttaa
	tgaaaaaaacaaacagctaatattatctcttcaagaaaaattagattttaatgctctggctatgaa
	tgatatttctgagcatgaaaaattgataaaaagtatggctaatgacatttcagaatttaaaaatgc
	attaacgatcttgtgtgataataaaataaagaataacgagttagattatgtcaatgaattaaaaaa
	actcactgaacgaatagatacacttgaaataaacacatctcaagctagcgaagtgagtgtcaccaa
	tagatttacaaaattccatgaaatagcgcactatgaaaattatgagtatattcatcctccgaagac
	atatctaatagaatttctttaaatttacaggctgttggattgacaaaaaattcagcagaaaaattg
	gctagattgacattagctaccttcgtttctggacaaatcattcaattcagtggctctttggcagat
	attatcgcggatgcaattgccattgctattggtgcaccacgttatcacatatggagagttccagtt
	ggtattatttctgacatggatgcttttgattttatagagactatagctgaatcatctcgctgtctc
	cttttgaaaggggccaatctttcagcatttgagatttatggagcggcaattagagatatagttgtt
	caacggcaaatacatccaacaaattatgaccatctggcattgatagctacctggaaacaaggccca
	gctacattccctgatggaggaatgttggccgagttgggacctgttattgatactgatacattaaaa
	atgcgtggtttatcagctactttaccccaattgaaaccaggttgtcttgccaaggataaatggaca
	aatattgatggactacatcttgatagtgttgatgattatgtagatgaattaagagcattactggac
	gaagctggatttgatggggaactttgtggaagagaatgattcatatttctatacttcactcata
	aggatccctaatggaaattatatttatgatctttattctgtcttgtctttttatactcttacatgg
	gcaaaaattaaaggtggccccgtccaaaagatagaagatattgccaatcgtgaattaaaaaattat
	agtgcaaaaatatcttcttgaggaggtggttaatggagtggagagcagtatcacgagacaaagcac
	tggatatgttatcaactgcattaaattgtcgatttgatgatgaagggttgagaatttcagcagttt
	cagaatgcttaaggagcgtattatatcaatattctatatctgaaacagaagaagctaggcaaactg
	taacctcgcttcgactcactagtgcagtaaggcgaaaattggtacctttatggccagacattgctg
	atattgataatgctatacatccgggcattatgtctatattgaacagcttggctgaattgggtgaca
	tgattaagttagaaggtggtaattggctaacagctcccccacatgcagtacgaattgacaataaga
	tggctgtttttttggtggagagccttcctgtacattttcaacgggcgtggtagctaaatctgctg
	gaagagttcgcttggttgaagaaaaagtgtgtactggaagtgttgaaatctgggatgcaaatgagt
	ggattggtgccccagcagaaggcaatgaagaatggtcatccagactactatctggaactatttccg
	gctttatcgatgcacctggcaatatgagtgaaacgactgcatatgtgcggggaaaatggctccatt
	tgtcagaactttcttttaataaaaagcaaatctacttatgcagaatgtccgttgataatcactttt
	cctattatttaggagaaattgaagctggacgcttatgtagaatgaattcgttagaatcgtctgatg
	atgtcagaagattacgtttttttctcgatacaaaagataattgtccgctaaaggtccgtatcaaaa
	tatctaatgggctagcaagattaagattaaccagaagattaccaagacgagaaacgaaggtactcc
	tgctaggctggagagaatcaggttttgaaaatgaacattcaggaataacacaccatgtattccccg
	aggaaatattacccatagtgcgtagcgcttttgaagggcttggtattatttggattaacgaattca
	cgcgacggaatgaaatatgattaataaaaataaagtaactgaacgttcaggtatacatgataccgt
	gaaaagccttagtgaaaatctgagaaaatacattgaggcacaatatcatatccgggatgaagggtt
	aattgctgagcgacgagcgcttttacagcaaaatgaaactattgctcaagctccttatatagaagc
	aaccccaatttatgaacctggtgcgccatacagtgaattgcctattcccgaagcagcaagtaatgt
	gctaactcaactatcagaacttggaattggcctctatcaacgcccctataaacaccaatcacaggc
	acttgagtcatttcttggcgaaaacgcttctgatctggtcattgcaacaggtacaggctccggtaa
	gactgaaagctttctaatgccaattattggaaaattggcgattgaatcttccgagagacctaaatc
	tgcatcccttccaggttgtagagcaattttattatatccaatgaatgcattagttaacgatcaact
	tgctcgtatcagacgtctttttggtgattctgaagcctctaaaatactgagatctggaagatgtgc
	ccctgtacgctttggcgcttatacgggaagaacgccttaccctggtcgtcgtagctctagacgaga
	cgagctttttatcaaacccctttcgatgagttttacaataaactcgcaaataacgcccccgtacg
	tgcggaactgaaccgcattggtcgctggccaagtaaagatcttgatgctttttatgggcaaagcgc
	atctcaggctaaaacctacgtctcaggcaaaaaaacgggtaagcaatttgttttgaacaattgggg
	ggagaggctaattacccagcctgaggatcgtgagctaatgacccggcatgaaatacagaatcgctg

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.	Cloned Sequence
	tccagaattactgataacgaactactccatgcttgagtatatgctgatgcgacctatcgagcgtaatatttttgagcagactaaggaatggctcaaagctgatgagatgaatgagcttatcttagtgcttgatgaagcgcatatgtatagaggagcagggggagcagaggtagcccttttaatacgtcgcctctgtgctcggttggatattccccgggaacgtatgcgctgcatccttaccagtgctagtctagggtccattgaggatggagaacgttttgcccaagacttaactggcttatcaccaacctcttcgaggaaatttcgaattattgagggtacaagggaatcgcgtcctgagtcacaaattgttaccagtaaagaagctaatgcactggctgaattcgacctaaattcatttcagtgcgtagctgaggatcttgaatctgcatatgcagcaatagagtctcttgccgaacgaatgggctggcaaaagccgatgataaaagatcatagtacactacgtaattggttatttgataatttgactggttttggtcctattgaaacgcttattgaaatagtttcaggtaaagcggttaagctaaatatcttgagtgaaaacctttttccagactctccacagcaaatcgcagagcgagcaacagatgcattactcgcattgggttgctatgctcagagggcatccgatggcagagtgcttattccaactcgcatgcatcttttttatcggggattaccaggtctttatgcctgtatagatcccgattgtaatcaacgtttgggtaaccatagcgggccaactatacttggccgcctttatacgaaaccactggatcaatgtaaatgcgcttcaaaagggcgagtctacgaattatttacccaccgtgactgcggtgcggcttttattcgtggatacgttagttccgaaatggactttgtatggcaccagccgaacggaccattatcagaagatgaggatatcgatcttgttcccatagatatattggtcgaggaaacacctcatgtacatagtgattaccaggacagatggctacatatagcaacaggacgcctttctaaacagtgtcaagatgaggattctggttatcgtaaagtctttatacctgaccgagttaagtctggatcagaaattacatttgatgaatgccctgtttgtatgcgtaagacaagaagtgctcagaatgaaccgtctaaaattatggatcatgttacaaaaggggaagcaccttttacaacgttagtacgtacacagatatctcaccagccagcgagtcgtcctattgatggtaaacatcccaatgggggaaaaaaagtacttattttttctgatggccgacaaaaagcagctcggcttgcacgtgatattcctagagatattgagcttgatttgtttcggcaatccattgctctcgcctgttctaaactgaaagatatcaatcgggaacccaaaccaacatcagtactttaccttgctttcctatcagtcctttctgaacatgacttgcttatttttgatggggaagattcacgaaaagttgtaatggcccgtgatgaattttatcgtgattataatagcgatctggctcaagcttttgatgatagcttcagcccccaagagtcaccgtcacgatataaaatagcgttgcttaaacttttatgtagcaattactattctctttccggaacaacagttggttttgttgaaccatcgcagcttaaatcaaaaaaaatgtgggaagatgtgcagtccaagaagctcaatattgagagcaaggatgttcatgctttagctgttgcttggattgataccttactcactgaatttgcttttgatgaatctattgattcgacactacgaatcaaagcagctggattctacaaacccacttggggtagtcaaggacggtttggaaaagctcttaggaaaaccctgatacagtatcctgctatgggggagctttatgtggaagttttggaggagatttttcgtactcatctgacattaggaaaagatggtgtctactttcttgctccaaatgcactacgtctgaaaatagatctcttgcatgtctggaaacaatgtaatgactgcacggcactaatgccatttgctttagaacattctacttgccttgcttgtggtagtaacagtgtcaaaacagtcgagccgtcggaaagcagctatattaatgcacgaaaaggattctggcgttcgccggtagaagaagttttggtttcaaattcgcggcttctaaaccttagcgttgaagagcatactgctcaactctcacatagagatagggccagcgttcatgccactacagaactctacgaactgagattccaagatgttcttattaatgataacgacaagcccattgatgtacttagttgtacgacgacgatggaagtgggggttgatattggatctctggttgctgttgctttaagaaacgtccctccgcaacgagaaaattatcagcaacgtgctgggcgagcaggccgccgtggcgcatctgtttcaacggtggttacatattctcaaaatggccctcatgatagttattatttccttaatcctgaacgcattgttgcaggttctcctcgtacacctgaagtgaaagtaaataatcccaaaatagccagaagacacgttcattcttttttagttcagaccttttttcacgagttaatggaacaaggaatttataatcccgcagagaaaactgccatacttgagaaagcacttggtactacacgagatttttttcatggagcaaaagatactggcctaaatctcgatagctttaataattgggttaaaaaccgtattctatctactaatggtgatttgagaacaagtgttgcagcatggcttcctcctgttcttgaaactggagggctttctgccagtgactggtttgctaaggtagcagaggaatttttaaatacactccatgggctggctgaaattgttccacaaactgccgttcttgttgatgaggaaaatgaagatgatgagcagacttctggtggaatgaaatttgcacaagaagaattacttgagttcctgttttaccatggtttattaccaagttatgcatttcctacaagcctctgtagtttcttggtagaaaaaattgtaaagaatattagaggttcttttgaggtgcgaacagtacaacagcctcagcaatcaatttctcaggctctgagtgaatatgccccgggacgtttgattgttattgataggaaaacctatcgctctggtggtgttttttctaatgcattgaaaggcgaactaaaccgggcaagaaagctttcaataatcccaaaaagtttattcattgcgataagtgctcttttgtccgcgatcctcataataatcagaatagcgaaaatacttgtccgatctgtggtggcattctaaaagtagaaataatgattcagcccgaagtctttggacctgaaaatgccaaggaacttaatgaggacgacagagagcaagaaatcacctatgtaacagcggcacaatatccacaacctgttgatcctgaagattttaagttcaataatggaggtgctcatattgtttttactcacgcaatagatcagaaactggtgacggtgaaccgagggaaaaatgagggggagtccagtggtttttcagtatgttgcgaatgtggtgcggcctccgtttatgattcctactcaccggcaaagggggcacatgaaagaccgtataaatatatagcaactaaggaaacgcctcgcttatgctctggcgagtataaacgcgtttttctcggacatgatttccgtactgatttgcttttattacgaataaccgttgggtctccgcttgtaactgatacttcaaatgctatcgttttacggatgtatgaagatgcattatatacaatagcggaagcactaaggcttgcagctagtcgccataaacaactggatcttgatcctgctgagtttggctctggtttcagaattttacccactatagaggaagatactcaggcattggatctcttcctttatgatactttatccggcggtgcgggttatgcggaagtagcagcagcgaatctagatgacattcttactgcaacactcgcattgttagaaagctgtgagtgcgatacctcctgtacagattgtctcaatcatttccacaaccagcatatacaaagccgtctcgataggaaactaggtgcatctttacttcgttatgcactatacggaatggttcctcgttgtgcttcacctgatattcaggtagaaaaattgtctcaattgagggcaagtctggaattggatggttttcaatgcataattaagggaactcaggaggcacctatgattgtgagtttgaatgaccgttctattgcagtgggaagttatcctggtcttattgatcgacccgactttcaacacgacgtatataagtcaaagcatactaatgctcatatagcctttaatgaatatcttcttcgttcaaatctgccacaatcgcatcaaaatattagaaaaatgttgcgctgatagcagcagtattgagtgccctaaagccctgtagggcactcaaggttttcagtgcgtgagcgggctttaactgaagccataaa

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.		Cloned Sequence
		tgtacgtatgggagaaaatgtgaccatttaactcgccagcaactattgcacaatgtaaaattatgcccattgag (SEQ ID NO: 270)
12	Control	acatcccgtcatcatgccatcacgacgcgctgagacgctgaaaaaataaaatcagcaccaccgtcagcgcgcagtgctttccccgcctcgcccgcccgcttcatgagacggttttaatgcagttgcattatgtcccgctcctcagtgctgcgctccatcctgattacaaaaaccgttatcaaaaacacatgcaaatagacgcagtcaaatgcgctaccgcctctcgcaataccttcaatttcatgataaaaaacatcatccctaacaagagcattatcctcatgaaaaaagtatatgaactaaccagtgaagaagcactgtcatatttcttcgccatgactcctacacaacattagaattaccggcttatattaatttcaccacattattaaatgatattaattcatctatccataacaaaaaaattaaaattgaaccaaccgccaaggagctgatgggtaaagatatcaattatgaggtgcttgtcagtaaagatggtctatatagctggcgtaggataacacttatcaatcccctttattatgtctacttctgtagaaaaatcacagcaccagcaacctgggaaatcataacagaaaaattcaaatcttttgaatcaaacgacctttttacatgttcaagcatccccgtcagaaaagacaactcgtcaaacattgctgcgtctgtaatgaattggtgggaagattttgaacaaaaaagccttgcccttgctcttgaatacgaattcatgttcagcactgacatctcaaacttctacccatcaatatatactcatagttttgaatgggtattcatatcaaaagaagaggcaaagaagaaaaaaagcaaaaataacccaggggggattaattgacagccacattcaaatgatgatgaacaaccagacaaatggtattccactcggcagcacattgatggatacatttgctgagcttatcttgggtcaaatcgatatagaattaagaaaaaaaactaacgaactcaaaataataaactacaaggtagtacgctaccgtgatgattaccggatcttctctaatagcaaagatgatttagacataatatcaaaatgtttagtcaatgtattgggcgattttggtttagatctaaactcaaaaaaaactgaactatatgaagacatcatacttcattcgttgaaacaagctaaaaaagactacatcaaagaaaaaagacataagtcactccagaaaatgctctattcaatatatttattttcacttaaacatccaaactcgaaaacaaccgttagatatctaaatgattttcttaggaatttatttaagcgaaagacaattaaagataacggccaacaggttgatgctatgcttggtattatttcaagcatcatggcaaaaaaccctacaacgtacccagtaggaacggcaattttctcaaaactcctcagttttctttatggtgatgacacccaaaaaaaattaacaaagctagaacaactccataaaaaactggataaacaacccaatacagaaatgcttgacatatggtttcagcgaactcaagcaaaaataaacctagagtggaataaatcttataagtcagctctatgcgtccgtataaatgatgaactcacaaaagagaaaacattttctgtaaataatttatggaatattgactggatccaaggaaaagaaacaagccccaataaagccaaaatattatccttgctaagaaaaacaaaaatcgttgacacagataaatttgataaaatggatgacaatataacacctgaagaagttaatctattctttaaagagcacagcaattaatatcccaaagccatgttagtaacataacatggctttttaaatcactcattatcagttatcaagaacgaacataacattctattccgaggag (SEQ ID NO: 271)
13	1	agttaatgactattgtgagcgagaaacgcgctactactatatatagacagacaagatgcacttactgaataaatactcataacggagaaaccagctgtatagtgaacaatagatttccagtagcatatttttacttcacttttagttattaatatgataatcataaactacggctctgccttaaatttgtgaggttgtttcgcctcgaaggaactaatgttaggacatacgccaccgttcagtcgatggtaacgcttcttaactagtggtccgctaagtgatgcgcaaagtgattgggcagagccgaaacgtttacaatccgataggagttggttttgtcgctacatgataaattattaatgcataacttcgcattagccaataaaaaaagccctgacttcatatctgaacttcctcaaattgaacctaaaccatacagcaatggacataaaattaaatggataaaccacacacttactagcactgaagttactccccctgataacctgattaaaatatgcatattgattgagtcaggggaaattgctataacatcagtaagtgatattgccaatttacttggagttcctgctggccaattactttatatactatatcgtaaaaaagataattatcgtacttttgaaatagaaaagaagaatggtaaaaaaagagtcattaatgctccttgtggcggtctatcgatactccaaacgagactaaagcccgttcttgaatatttctacaggccaaagaaatctgctcatggttttataaaggaaagagcatcattactaatgctgggatgcatattaaaaaaaatttgtcgtaaacattgatctagaaaactatttcgaatcaataagttttgctagggtttatggaatatttaaaagtaaacctttttaattttgctcatcctgcagctactgttttagctcagttatgtactcacaatggaaaattacctcaaggtgcgtgtacatcgccaatattagcaaatattgcatcagcttctctagacaaacagctcacccaatttgcaggaagaaaaaaatatcttattctaggtatgctgacgacataactttttctttcaatcagagaaatattgatataatcaaaaaaaacgacgacggaagttatagtcttagtgaaactatagacaatattatttcaaaaaatggctttaaaataaattatgataaatttagagttcaaaccagaaatacaagacaaagtgttactggcttagtggttaatgataaagttaacattaacagaagatatataagaattacacgttcaatgattcatagatggacagatgataagctaaagtatgcacttctctttgctacagaaaaaggatatcaggcaaaggataataaccacgcaattcaaattttccgaaatcatatttatggaaggcttagctttataaaaaatggttagagggaaagactatccaggatatttaaaactgatgtcatacatgagtcataacgatccattaaaaacccaagaaggattgcgagcaatgaaagaaacagaaaactttgatgtttttatatgccatgcaagcgaagacaaaaaagacattgcaattccaatatatgacgagttaactaaacttaaaatttcagccttcatagatcatgttgagataaaatggggcgactccttaattgataaaataaatgcagcactagttaaatcaaaatatgtcatcgctattttatctgctaattcagtcaataaggaatggcctcaaaaagaattaagagcagttttagccagcgaaatatcgagtggcgacgtaaaacttttgaccttattaaaaaaagaagacgaggaggtcgtaaacctatcattacctttacttagtgataagttttatatggtctatgataataatcctgaagtagtcgccaacaatattaaatcactcttacaacgataattctctcacaaaagaaaatgtgcagattgatgcgtattaagtattaatctgcacatacaaaaaaaataataaaataatacatttttcataacttgtagg(SEQ ID NO: 272)t
14	2	cgcgctatcacgtaaaataggcaaaatacttctggaaaacagaaagttgaagtgatatgttcataaacacgcatgtaggcagatttgttggttgtgaatcgcaaccagtggccttaatggcaggaggaatcgcctccctaaaatccttgattcagagctatacggcaggtgtgctgtgcgaaggagtgcctgcatgcgtttctccttggcctttttttcctctgggatgaagaagaaatgacaaaaacatctaaacttgacgcacttagggctgctacttcacgtgaagacttggctaaaattttagatgttaagttggtatttttaactaacgttctatatagaatcggctcggataatcaatacactcaatttacaataccgaagaaaggaaaag

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.		Cloned Sequence
		gggtaaggactatttctgcacctacagaccggttgaaggacatccaacgaagaatatgtgacttac
		tttctgattgtagagatgagatctttgctataaggaaaattagtaacaactattcctttggttttg
		agaggggaaaatcaataatcctaaatgcttataagcatagaggcaaacaaataatattaaatatag
		atcttaaggatttttttgaaagctttaatttcggacgagttagaggatattttctttccaatcagg
		atttttattaaatcctgtggtggcaacgacacttgcaaaagctgcatgctataatggaaccctcc
		cccagggaagtccatgttctcctattatctcaaatctaatttgcaatattatggatatgagattag
		ctaaactggctaaaaaatatggatgtacttatagcagatatgctgatgatataacaatttctacaa
		ataaaaatacatttccgttagaaatggctactgtgcaacctgaaggggttgttttgggaaaagttt
		tggtaaaagaaatagaaaactctggattcgaaataaatgattcaaagactaggcttacgtataaga
		catcaaggcaagaagtaacgggacttacagttaacagaatcgttaatattgatagatgttattata
		aaaaaactcgggcgttggcacatgctttgtatcgtacaggtgaatataaagtgccagatgaaaatg
		gtgttttagtttcaggaggtctggataaacttgaggggatgtttggttttattgatcaagttgata
		agtttaacaatataaagaaaaaactgaacaagcaacctgatagatatgtattgactaatgcgactt
		tgcatggttttaaattaaagttgaatgcgcgagaaaaagcatatagtaaatttatttactataaat
		tttttcatggcaacacctgtcctacgataattacagaagggaagactgatcggatatatttgaagg
		ctgctttgcattctttggagacatcatatcctgagttgtttagagaaaaaacagatagtaaaaaga
		aagaaataaatcttaatatatttaaatctaatgaaaagaccaaatatttttttagatctttctgggg
		gaactgcagatctgaaaaaatttgtagagcgttataaaaataattatgcttcttattatggttctg
		ttccaaaacagccagtgattatggttcttgataatgatacaggtccaagcgatttacttaattttc
		tgcgcaataaagttaaaagctgcccagacgatgtaactgaaatgagaaagatgaaatatattcatg
		ttttctataatttatatatagttctcacaccattgagtccttccggcgaacaaacttcaatggagg
		atcttttccctaaagatattttagatatcaagattgatggtaagaaattcaacaaaaataatgatg
		gagactcaaaaacggaatatgggaagcatatttttttccatgagggttgttagagataaaaagcgga
		aaatagattttaaggcattttgttgtatttttgatgctataaaagatataaaggaacattataaat
		taatgttaaatagctaatgaacagccctaacgttatgaacgctaaggctgatttttcg
		(SEQ ID NO: 273)
15	3	gctcatgttatgcatgtgcatgaaaaccactgcataaagcgggcaggcgtggcggggatacgagcg
16		cgcgccatgtggtatggagattggatctattcataacttgatgtataaagtagaaaaaaaagcggg
		gagattatgaataaaaaatttaccgatgagcagcaacaacagcttataggacatctcacaaagaaa
		ggcttctatcgaggagctaatattaaaataaccatttttctatgtggtggtgacgttgctaatcat
		caatcttggcgtcatcaattatcacaatttttagcaaagttcagtgatgttgatatattttatcca
		gaagatctatttgatgatcttttggctggtcaagggcagcatagccttttaagtttagaaaatatt
		ctggctgaagctgtcgatgtaataattttatttcctgaaagtccggggtctttcacagagcttggt
		gcgttctctaataatgaaaacttaaggagaaagttgatttgcattcaagatgcaaaatttaaatca
		aaacgtagctttattaactatggtcctgttcgcctgttgcgtaagtttaattcaaaatctgttttg
		cgttgtagttcaaatgaactaaaagaaatgtgtgattcatctattgatgttgccagaaaattacga
		ttatataaaaaattaatggcatctattaagaaggttaggaaagaaaataaagtatcaaaagatatt
		ggaaatatattatacgcagagcggtttctattgccttgtatctatttactggatagtgtcaactac
		cgcacactgtgtgaactagcttttaaagcgataaagcaagatgatgttttatctaaaattattgtt
		agatccgttgtttctcgtctaataaatgaacgaaaaatacttcaaatgactgatggttatcaggtc
		actgctttgggggctagctatgttaggagcgtctttgatagaaagacacttgaccgattgcggctt
		gagattatgaattttgaaaaccgtagaaaatcaacatttaactatgataagattccgtatgcgcac
		ccttagcgagaggtttatcattaaggtcaacctctggatgttgtttcggcatcctgcattgaatct
		gagttactgtctgttttccttgttggaacggagagcatcgcctgatgctctccgagccaaccagga
		aacccgtttttttctgacgtaagggtgcgcaactttcatgaaatccgctgaatatttgaacactttt
		agattgagaaatctcggcctacctgtcatgaacaatttgcatgacatgtctaaggcgactcgcata
		tctgttgaaacacttcggttgttaatctatacagctgattttcgctataggatctacactgtagaa
		aagaaaggcccagagaagagaatgagaaccatttaccaaccttctcgagaacttaaagccttacaa
		ggatgggttctacgtaacattttagataaactgtcgtcatctcctttttctattggatttgaaaag
		caccaatctatttgaataatgctaccccgcatattggggcaaactttatactgaatattgatttg
		gaggatttttcccaagtttaactgctaacaaagtttttggagtgttccattctcttggttataat
		cgactaatatcttcagttttgacaaaaatatgttgttataaaaatctgctaccacaaggtgctcca
		tcatcacctaaattagctaatctaatatgttctaaacttgattatcgtattcagggttatgcaggt
		agtcggggcttgatatatacgagatatgccgatgatctcaccttatctgcacagtctatgaaaaag
		gttgttaaagcacgtgatttttttattttctataatcccaagtgaaggattggttattaactcaaaa
		aaaacttgtattagtgggcctcgtagtcagaggaaagttacaggtttagttatttcacaagagaaa
		gttgggataggtagagaaaaatataaagaaattagagcaaagatacatcatatattttgcggtaag
		tcttctgagatagaacacgttaggggatggttgtcatttattttaagtgtggattcaaaaagccat
		aggagattaataacttatattagcaaattagaaaaaaaatatggaaagaaccctttaaataaagcg
		aagacctaat (SEQ ID NO: 274)
17	4	acgtgtcttgatttaagttgacttcaagactataaagtctcaagtaacagtcggttagcttccttc
18		atgggttggtcatgccgggttgttaagtatggctgtttgcgataagctttaaatactctttagcgt
19		tggacggttacgtctagtcgggtgattagccagactctaacttattgaacgtattaagggttgcga
		aagtgtcgcaacccgagatcgttcctctctcgggttgcgacactttcgcttcctcaagtaaagagt
		gaagcccggcgcaaatgcgccgggccattttcaggtactgttatgtctgttattcgtggattagct
		gcggttttacgtcaaagtgactccgatatcagcgcctttcttgtaaccgccccgagaaagtacaaa
		gtttacaaaatccctaagcgtacgacgggatttagagtcattgcccagcctgccaaagggctaaaa
		gatatccaacgagcctttgttcagctctatagcctccctgttcatgatgcttcaatggcctatatg
		aaagggaagggaattcgtgataatgctgcagcacatgctggcaaccagtatctcctaaaggcggat
		ctggaggatttttttaactcaattacaccggcaatttttttggcgttgcattgaaatgtcatctgcg

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.	Cloned Sequence
	caaacacctcaatttgaacctcaggataagcttttattgaaaagatccttttctggcaaccgata aagcgtcgcaaaaccaaattgatattgagtgttggtgcgccttcttcaccagtcatatccaatttc tgtatgtatgagttcgataatcgaattcatgcggcttgcaagaaggtggagataacatacacacgc tatgcagatgatctcacgttctcgtctaatatccctgatgtactgaaagcagttccttcaacgctt gaggtcttactgaaggatttatttggaagcgcgctcagacttaatcacagcaaaacggtttttca tcaaaagcacataaccggcatgtgactggtataacaataaataatgaagagacactttcactcggg cgcgatagaaaaagatttatcaaacatctgattaaccagtataagtatggactccttgataatgag gataaagcttatctgatcgggctgttagcatttgccagccatatcgagcctagtttcatcacacgg atgaacgaaaaatactcattagaactcatggaacgcctgagaggacagagatgaccaagcaatatg aaagaaaagcaaagggtggaaatttactgtcagcattcgaactttaccaacgtaatagtgataaag cgcctggtctgggtgaaatgttagtgggtgagtggttcgaaatgtgcagggattacattcaggatg gacatgttgatgagtcaggaatatttcgtccagataatgcgttctatcttcgccgcctgacgttaa aggattttcgccgtttctctcttctggaaattaaactcgaagaagatctgacagtcattattggca acaatggtaaagggaagacaagtatcttatatgcgattgcaaaaacgctgagttggttcgtcgcga acatcctgaaggaaggtggtagtggacaaaggttaagcgaaatgactgacataaaaaatgacgctg aagacaggtattcagatgtcagtagcactttcttctttggcaaaggacttaagagtgtgccgatca gattgtcacgctcagcccttggtacagccgaaaggcgggacagcgaggttaagcctgccaaggatt tagctgatatatggcgagtcatcaatgaggtgaatacgatcaacttgccgacgttcgctctttaca acgttgagcgatcgcaaccgtttaaccgcaacataaaagataataccggacgcagagaagagcgct ttgatgcctatagtcaaacgctcggtggcgcaggacgtttcgatcatttcgttgagtggtacattt acctccataagcgtactgtatcagatatctcaagttctattaaagaacttgaacaacaggttaatg acttacagcgtaccgttgatggcggtatggtttcggtaaaatcacttctggaacagatgaagttta agcttagtgaagctatagaaagaaatgatgctgcggtttcctcgagagtgttaactgagtctgttc aaaaaagtattgttgagaaagcaatctgctcggttgtccctagtatcagcaatatatgggttgaaa tgataacgggttctgatttagtcaaagttacaaatgatgggcatgatgttactattgaccaattat ctgacgggcagcgtgtatttctgtcgttggtggccgatcttgcgcgaagaatggttatgctgaatc ccctgctggaaaatccattagagggacgtggcattgttttaattgatgaaatagaacttcaccttc atcctaagtggcagcaggaagttatcctgaacctgcgcagtgcattccctaacattcaatttatta ttacaacacacagtcccattgttctttctacaattgagaaacgctgtattcgtgagtttgagccca acgatgatggcgaccaatcattccttgattctcccgatatgcaaacaaagggaagtgagaatgctc aaattcttgagcaggtaatgaacgtacattctacaccgcctggtattgctgaatctcattggttag gtaattttgaactattgcttttagataattctggagaacttgataaccactctcaagtgctttacg accaaatcaaggcgcactttggcatcgatagtattgagttgaagaaagcagatagccttattcgca ttaataagatgaagaataaactgaacaagataagggccgagaaggggaaatagtaatgagagagtt agcccggctggagagaccggagattcttgaccagtatatagccggtcaaaatgactggatggagat tgatcagtctgcggtatggccgaaattaactgaaatgcagggcggattttgtgcctattgcgagtg ccggttgaacagatgtcatattgagcatttcaggccaaggggaaagtttcctgctctgacgtttat ctggaataacctgtttggttcttgtggcgattcaagaaaaagtggcgggtggtcacgttgcggtat atataaggacaatggtgctggcgcctacaatgctgatgatcttataaaacctgatgaagaaaatcc tgacgactacctgctatttctcactactggagaggttgtaccggctatcggactcacggggagagc gcttaaaaaagcgcaggaaactatccgtgtttttaacctgaacggtgacataaagttgtttggcag tcgcagaactgcagtgcaagcaatcatgcctaatgtcgaatatttgtatactctactcgaagagtt tgacgaagatgactggaatgaaatgcttagagatgagctcgaaaagatagaatctgatgaatacaa aacggccctaaaacatgcatggactttcaaccaagagttcgcataatcctaaa (SEQ ID NO: 275)
20 21	gtccttaaacacgacaaaacctgtgatacttaccatggattcctctatgaaggaaaggtagtatag ccattttgggtgatacatacagtgaatgtcattgctgtagttgaagtgagtaagagcgcttaagat taagttgagagaaaatgaaactacttgataaaaagtattacaacctcgagcccaaatatgagtacc ttaaggactcatttattttaggactggcatggaaaaaaacagatagttttgtaagaactcacaatt ggtatgcagatattttagagctggacaagtgtgcgtttgatattagtgatgaagtcactaattggt caaacgagatctcaaagaacgctctttccaaaagtgatattgaattgataccggctccaaaaggag caagctggttcattaatcaaggtaaatggactaccaataaagataatagaaagataaggcctttgg ctaacatatctattagggatcagtcttttgctacagcagtaacaatgtgccttgctgatgctatag aaacaagacagaaagactgttcgttgagcaatcttggctatgctgagcatgtaaagaacaaggttg ttagttacggaaataggcttgtctgcgattgggacaatgaaagggcaagatttcgttggggaggaa gtgaatattataggaagttctcttccgattatcgaagctttctacaaagacctatctatataggca gggaaacagtaaataaagttagcggaattgatgatgtatatatcatcagtttagatctgaaaaatt ttttcggttctataaaaataaaccttctgttagaaaaaatcaaaaaaatatccgctgatcattatg cagctaaattcataaatgataatgaattttggactttggcgaatcggattttaagttgggattggc ctgaagaatctttatctttacttgagagtttggatataaaagaaaaaaatgttggtcttccccagg gattagcttctgctggtgctctggcgaatgcatatctcattgagtttgatgaatctttaatttcta agcttcgtactaagatagaagacagccaaataatactgcatgattattgtcgatatgtcgatgata ttagattagtgatttcaggagaagcactagaaagtaataagattaaggaatctattcatgcattag ttcagggcattcttgatgagacattggctcaaaatccgtcagataatgaaccatatttaaaaatta acgatagcaagacttatattcttgagctttcagacattgacaacggaagtgggcttacaaatcgaa tcaatgaaattcagcatgaagtaggagcttcgagtatcccagagcgtaacggactcgataataata tcccggcacttcaacaattattactgaccgaacaggataattttccgaggatgttgatagtttat ttcccgggtttaaaaatgataagtcgataaaggtagaatctgtacgtagattttctgcccataggc tggaaaaaagtttggctaaaaaaagcaagctaatttcacctgaggagaggaaacaatttgataatg aaacctcactgattgcaaaaaaattattaaaagcttggctaaaagatccatcaattatggttatct tccgcaaagcgatagctatcaatcctaatctagatgcttatagcaccattcttgaaattattttt

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.		Cloned Sequence
		caagaatacaacgcaatcgtgataaacgagataaatatataatgctgtatcttctttctgatatat ttcgtagcgtcattgatgtctatcgaaacctagaatcagaatacgtcgacgattatcaaaaattga tgggtgaagttacattgtttgcccaaaaaatactttcctgcaaatcttttattccaaattacgcat atcagcaagcattattttatctcgcagtgatcaataaaccatttatagctagtaataaagcttctt ttgatcttgcaaggcttcaatgcgtcttaattaaacagcatttagaaccgttgaatagtagtgatg gatacctatttgaggtatctgctcaaatcagtaaagactaccgagcaaatgccgcttttctacttt ctcatacaaatagtaacaaagtagtagacttaattatcgaaaaatttgctttccgaggaggtgaat tctggaatgcaatttggaaagaaattgttaggatgcaagataaagataggattaacgaatttagat gggccatatcaaaatatgagtcaaagccaaatagttcggagcactatctttcatcagtgatcagtt tcaaggaaaacccatttagatatgaacatgcgcttctcaagctaggtgtagcattagttgaactct ttgatgatacagagaaaaacgtatggcaacctgatggtaagcagtattctccacatgaaataaaag taaaattagaaggtaactcaacctcatggggtgaattatggcgtccaaattttagtatttcatgct cgatagataagaaaggtgaacctggtaaagacccacgctatataagccctgagtggttggcaaatt atccacagactcaaaatgatgaacaaaaaatctattgggtttgcagtgtgctaagaagtgctgctt taggcaatgtagattatactcaaagaaatgatttaaaacttgataaagctaagtatgatggtatcc attctcagttttacaagcgacgtatgggaatgttacatacaccagagtcaattgttggttcatatg gaactataacagattggtttgcaagttttcttcagcatggattgcaatggccaggttttcttctt cgtatataagccaagaagatatattgtcaattactaatattattgagtttaaaaactgtttattgg aacggctaggctacttaaataagcagatatgtatttcatcgaatgttccaaccttaccgactgttg tcaacaggcctgaattagcatctaaccattttagaattgttacggttcagcagttatttcctaagg atactaatttccatccttctgacgtgactttggctaatcccgatgtgcgctggaagcacagagagc accttgcggaaatctgtaagctaacggagcaaactttaaatgcaaaacttaaaactgagtctaggg aacatacaagcacagctgatctaatcgtttttttctgagttagcagttcacccagaagatgaagata tagttagagcactggcatttagaaccaaagccatcatttttttccggctttgtcttctgtgaacaag atggccgaatagttaacaaagctcgttggattattccagactcttcagagtctgggacccaatggc gtgtccgtgatcaggggaaacatcatatgaccagtgatgaagtggctcttggcattcaaggatata gaccatcccaacatattatttcaattgagggtcaccctgagggaccatttaaattaactggtgcga tttgctacgatgcaacagatataaagcttgcggcagatctgagagatttgactgacatgtttgtca ttgcagcatacaataaagatgtagacacatttgataatatggcttcagcactacaatggcatatgt atcagcatattgttattacgaatacgggagaatatggaggctcaactatgcaagccccgtacaaag agaaatatcataaattgatttctcatgctcatgggactggtcaaatagcaattagtactgctgata tagatttagcagcattcaggcggaagctacaaatatataaaaagaccaaaacccagcctgctggat acaatagaaaacattaaggatttttatggatactttagttaagttagctacaattatttctccatt aattagtgctggagtagctatttgggcaattttggttgctaaaaaaaccatcagtgaaagcaaaga aattgccaagaaaaccatcgctgatacggcctaccaagcatatttgcaattagccatggagaaccc acaattttcgaaaggctacagcgcagattgtagacaggagcgagaccctatgtatgatcaatatgt ttggtacgtggctaggatgatattctgctttgagaaaatcatcgaggttgaagtaaacttaaaaga tagttcttgggcaaatacgttggaaaaacatttgaagtttcattctgaacattttaagaaaacgaa tgttgtcgaagaggctctctatattccccctattttggatctcataagatgtgcagctaactaata acttatcccaataggattatattccacacgataagcccactggaaaatgtaacatcccaagatagt ttttgggattgtttcccagtgggcggaaagtatcatgatagttgtcacccccggtggagctgcaaa gatttttatggggtgggtgttacattgcg (SEQ ID NO: 276)
22	6	acacgatataaaaccatctcattgcttgctgggttaactgagttgctgaatttttttctagaattt cgcaaaatttaataggtaaaccttgtttttttaaatttacgatgatataaaaataatgccctaaac aaaggtttaggggtattgtacaggttgtcaagcctcccacaggtcttggtgaaaccaatcactgtg acgacggtaagcaacacttggatgatattcataattgactccacgctactgattacattatacagc atatctaacatttgcggcgaggttcacaatttgtatttaggtactgattgtggatgagaaggttgg agaaagaccacttggttaagccggaggatgtgtcctagaattgtcgctattctgtcatcctccggt tttgctaatttcattcagggaatataatgaataatgatgattacccatggttcagaaaacgtggtt atttgcatttcgatgaacctgtttcattaaaaaaagcggttaaatatgtttcctctccagaaaaaa taataaaacattcttttctgccatttttaagctttgaagtaaaatcgtttaaaatcaaaaaagaca aatcaacaaaacaattaagtaaaactgaaaaattaagacctattgcctattcctcacatttggata gtcatatttatgcattttacgcagaatatcttactggacattatgaattattgatccaagaaaaca atttacacgagaacatccttgccttcagatctttaaataaaagcaatatagaatttgccaagagag catttgatacaattactgaaatgggtgagtgtagcgctgttgcattagatctttctggtttttttg acaatttagatcatcaaattttgaaacaccagtggtgcaaagttattgggactgaagcgttgccgc aagaccattttgccatatacaaaagtataacaagatattctaaagttgataaaaatagagcgtatg agattttaggtatatcaaagaataaccccaagtataatagacgcaagatctgcacccctgttgatt ttagaaataagattagaaaaaatggtcttattatagttaataattcccaaaaaggtataccccaag gctcgccaattagtgctctactttcaaatatatatatgcttgactttgatattgaaatgagagatt acgcgcaggaacgtggtggccattattatcgctattgtgatgatatgctattcattgtaccaacta agtataataaaactctagcaggtgatgtagcccagcggattaagcatcttaaggtagaactcaata ctaagaaaactgagattcgagattttatatacaaagacagtaccttagtggcaaatatgcctttac agtatcttgggtttatttttgatgggagtaatatattattacgttcatcttctctcgcaagatatt cggaacgaatgaaaagaggtgtccgcttagcaaaagctacaatggacagcaagaataggattagag aaaataaaggtgaagctttaaaagctttatttaagaaaaaattatatgccagatattcacatattg gaagaaggaatttttttgacttatggttatcgcgccgcgaagatcatgaattcgaaagctataaaaa gacagttaaaaccattgcagaaaagattggaaaatgaaatactaaaataaatatttgctggcccga atcatacagggccacaatacagttgaaaacaagctataataaacaacatctaatttttatatac (SEQ ID NO: 277)

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.		Cloned Sequence
23	7	tctcaacttccccaaatgtccgtattcatccataaataccctgatttataacaattttaccgtttt
24		ttagtccatcatcgtccgcagccatccagtagaatccgataaagaatgtgtataggattgtgtata
		tgttcctgttcggtcatggattcctatacacatgcctttaaacgatatgcagattcgccgcgctaa
		gcctgaagataaaccctatacgcttggggatgggcaaggcttgtcattgcttatagaacctaatgg
		aagcaagagctggcggttccgctatcgctatgccggtaaacccaagatgatctcgcttggcgttta
		cccaacgatcactcttgccgatgctcgttcccgtcgtgatgaagctcgaaaacttgtggcagaagg
		aaagaaccctagtgatgttcgaaaagagcaaaagctggctctgcaagcagagtcagagaacgcctt
		cgaaaagatagccagagagtggcatcaacttaaatctgctaaatggtcggcaggatatgcatcaga
		catcatggaagcgtttaagaacgacattttccttatgtgggaacaaggcctgtgagtgagattaa
		accgctagagctgctgaacgtactgcgtaaaattgagaaacgtggtgcgttggagaaaatgcggaa
		agtgcgtcagcgttgctctgaagtgtttcgctacgcaattgcaacgggtagagcggagtacaatcc
		tgcggcagatctttccagcgctctcgaagtgcaccaatccaatcatttcccgttcctaaaagctga
		tgagatacccgaatttctgcgtgccttagagagttacaccgggagtaagcttgtccagatagcaac
		gaaattactgatgattacgggcgtgagaaccatcgaattacgcgcggcattatggcaagaatttga
		tctggataacgctatttgggaaattcctgctgaaaggatgaaaatgcgcaggccgcatcttgtgcc
		attgtcgacccaagcgttagatttactccatgaactcaagataatgacagggaactatcgttatgt
		ttttccaggacggaacgatccgaacaaaccgatgagcgaagctagcataaatcaagttatcaagcg
		tatcggttacgaaggccgactcactggtcacgggttcagacatatgttatcaacaattttgcatga
		agaaggttttcaatcagcatttattgaagtccaattagctcatgttgatagaaataatataagagg
		aacttataatcatgccatataccttatggaaaggcagaagatgatgcaatggtacagtgattatct
		tcgcaaaaaaaaggggttataatatgttaaaccagtcattttccgtttcgaacttaattaagcttt
		taaaaaaaaccgatccaaaaagatacaaaattggtaggaattcagctgaatataaaaaatatatag
		ctgataaagttaatggctcaattgaaacatactcatttggttcgatctcaaattcaagaattaaca
		acaaaaatgtgtatatatttaaagattttatggatgtacttgtcgccaggaaaataaatgataaca
		ttaagcgtgtgtatagtgttaaacaaaacaacagacatgacatcataaaaaaagtaaatacagtgt
		taagtgagcctgtaaattattatatttacaggctggatattaagagtttttatgaatcaatagata
		aaaatatcgttttccaaagaattaataataacccgattatttctcataatactaaaaaatttatca
		atggtctttttaaacataacgctttctctgcaaataacggacttccccgtggtatgggattaagtg
		cgactttatcagaaatatttatggaggaatttgatgctgagttggcgaggctgcctgaagtatttt
		atgcttcaagatatgtggatgatatcatagtttttttcattctataaaataccagattataaaaatt
		atttttcaaggattttaccaaatggattacatttaaatgaaagaaagtgcagtgagtataccatag
		aggacacttcaactaaacattctgaaattgagtttttgggatattcatttattatacaccatggat
		taaaaaatcagcgtcgtcatgttgtgatcagaatttcggaggagaaaataaagaaaataaaaagaa
		ggattgcacttgcggtaaaagattactcaaataattctgatgcagaactcttgaagaaaagaataa
		agtatttaactggtaatatattagtaaactccaatagtaataaaactgatgctttatatagtggaa
		tttattacaattatcaacatttaactgataaaacacagctcaaggaacttgatatatttaagaata
		ggatgctattttcttcaaagggcgaggtggggagaaaaattttagcagcaggtcacaacttattaa
		ctgcgcctaaaaaatactcatttttggctggtttgaaaaacggctactgtcttcttttaaacggg
		aagatattattaaaataaataaggtttggtgattcatgaaaattaaaatatcgaagagtgattata
		aaagagtacttctcacggatattttaccatatgaagtccctatcctttttttctaacgaaggtttct
		ataagttaatttctgaaaataaagttttacccggaacattttcagaaggccttaagctggattctt
		ataccatcccttactcctataaaataaaaaaggggctggcgagttctcgaagccttggcattatac
		atccttcaacgcagttaagaatctgtgattttatgataagtatgaacatttgatggttcatatgt
		gtacaaaaagtccgttttcgctacgttatcctagcaaaatagggagctattattacgaaaaggact
		tcttaaaaagtagaataaatctaaaagatggtcttgtacaatttcataatcatggctttgattccc
		aagaaacttcctcatcttcccatttttcatataagaaatatcctttcatctataagttttatgagt
		catatgaatttcatagattggaaaggaagtttaggaaacttttaaagcttgatattgctaagtgtt
		ttagtcatatatatacacacagcgtttcatgggctgtaaaatctaaagaattctctaaggttaata
		gaacttataacagctttgaaggttgtttggataagctttttcaagatgccaattatggtgaaacaa
		atggcataataattgggcctgaattttcaaggatatttgcggagattatattacagcgcgttgact
		tgaatgttgagtctcatttgaatcttgagccaggcatagttaaagataagagctatgctataagac
		gttacgttgatgattattttatatttgcggatgatgatgaaacatttaagctaatagaatttgtac
		tggcaaatgaactcgaaaaatataagctttatttgaatgaatctaaaaaggaatttatcgagaggc
		cattcgtgactggagctacgatggctaaaaatgatattgcagaaatcattgaggatttatatggat
		cgttaatccatactgagaagttggatgagttaacagctatggttaatttaaatccagacgtcaaaa
		ttcagcctgaaaatatgaatgacctttttccattgaaaggtgtgtggaataaaaagctacacgcgg
		acaaatttataaaacgaatcaaaattgcggttagaaaaaacaataccacatttgatcttgttagct
		catacttattaagtgcgattaagagtaagtttttcaaagtaattaggctgttgaggatgttcgatc
		tgtcaggaaaagaagatataacttataaattcttctcaatattcaatgaggtgatttttttttattt
		atgctatggattttcgagtccgacagacatacataattagccaagttattttggaaataaattcat
		ttgctaataagcaagcttcagacattagtgaagttataaaaaagaatacttttgatgagcttctta
		tgtgcatgaaaagcatgggtaatattcatgagaggccagtggagttatctaacttacttatatgta
		tgaaaggtttgggggagcagtataaactcaatccagatgaatttaaggatttgttgggtattagtg
		agaatgagtgtttttacgatttagaatatttttctatatgcagcatgttacactatataggcgatg
		atgttctctatctaaaaatgaaagaagatattgtccttgctatacagagtttgataagtggtcgga
		acgatataaaaaaagacactgaaacatttatgctattccttgatatgatgacgtgcccatatctta
		cagttaagcataagagaataatttatagaacatatgtcgaagcaaatacaggtcaaaaaagattta
		cgaatgcagtaattgattctgaaattgattctttaaaaaataatgtaatctttttttaactggtctg
		gagatgctgatcttgagcacgttctttataaaaaagagttgcgaacagcatatgaatagtagtatt
		ttaatttcgttaaagggttgcgatgcctaaggtttcgacctgaagcagataccggaagatcggctt
		ttgaatgttcatccgaaagatattcgcgatacgttttgaggatggaccgatttagacacactattg
		cctttttagctaaacaggccgcgaaagcggcctttttaatgaatcagatttcccctcaccgatctca

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.		Cloned Sequence
		atacttcccctcagcgtgcgcagccccgcccgcctgcccgcttcgcttaacagactggttttcatgcaccccttaaatcgtctcagaagccaccacacaagggctttcgcgtcaaaaatggcgcatgagactcatgcgttttcatgcgccatagatatgcactcatacgctctcaggccagctagggaaaaagcgtaaaaaatcccggtactggaccgagacttcgtgggcgtattttgctaa (SEQ ID NO: 278)
25	8	agcatcggagcaaagtaactcaataccgaacaataaatatgagcccttcgtgaaaccgggtaaggtcaaactcataaaccaacaaaaggggaaaagtgggatatgtgaggcgtgtatgattttttatttattgggcttcgttaaaaatggtgatttaatagccctttaaatttatcactttttaactaactccgagggtttatggtcatttttgacgagaagcgacacctgtacgaggcactgctgcggcataactacttccctaatcagaaaggctctatttccgaaatccccccttgtttcagctcccggacctttacaccagagatcgccgagctgatctctagcgatacctccggccggagatctctgcagggctacgactgcgtggagtactatgccaccaggtataacaatttcccacgcacactgagcatcatccaccccaaggcctactccaagctggccaagcacatccacgacaattgggaggagatcaggtttatcaaggagaacgagaacagcatgatcaagcccgatatgcacgccgacggcaggatcatcatcatgaattacgaggatgccgagaccaagacaatcagggagctgaacgacggattcggcaggcgctttaaggtgaacgccgatatcagcggctgtttcaccaatatctattctcacagcatcccttgggccgtgatcggcgtgaacaatgccaagatcgccctgaacacaaaggtgaagaatcaggacaagcactggtctgataagctggactactttcagcggcaggccaagagaaacgagacccacggagtgcctatcggaccagccacatcctctatcgtgtgcgagatcatcctgagcgccgtggataagaggctgcgcgacgatggcttcctgtttcggagatacatcgacgattacacctgctattgtaagacacacgacgatgccaaggagttcctgcacctgctgggcatggagctgagcaagtataagctgtccctgaacctgcacaagaccaagatcacaaatctgcctggcaccctgaacgacaattgggtgtctctgctgaacgtgaatagcccaaccaagaagcggttcacagatcaggacctgaacaagctgagctcctctgaagtgatcaacttcctggattacgccgtgcagctgaacacacaagtgggcggcggctccatcctgaagtacgccatcagcctggtcatcaacaatctggatgagtataccatcacacaggtgtacgactatctgctgaatctgtcctggcactaccccatgctgatcccttatctgggcgtgctgatcgagcacgtgtacctggacgatggcgacgagtataagaacaagttcaatgagatcctgtctatgtgcgccgagaacaagtgcagcgatggcatggcctggaccctgtacttctgtatcaagaacaatatcgacatcgacgatgacgtgatcgagaagatcatctgctttggcgattgtctgtccctgtgcctgctggatagctccgacatctatgaggagaagatcaacaatttcgtgtctgatatcatcaagctggactacgagtatgatatcgaccggtactggctgctgttttatcagagattctttaaggacaaggcccccaagcccctacaacgataagtgtttcgacatcatgaagggctatggcgtggacttcatgcctgacgagaattacaagacaaaggccgagtcctattgccacgtggtgaacaacccctttctggaagacggagacgagattgtgagtttcaacgactacatggctatcgcatgacttttaggcctcatt (SEQ ID NO: 279)
26	9	aagtgaacggatgtatattgagtgcaatgtgattaactatctgttgttacaatatttagataggtgataaaatatgacatctaccattgatttttatgaatctgatttctcagccacattatacccattaaaaccaatcaaatattactcaagcatcactcacaagagatgtcagaatatatttatcagaaggtcattaatcctgcatatccaacagatagttttctgtctcagcaaaaagtcttttcgactaaacctaaaggtcatttgagacgaactgtaaaattagatccagtagctgagtattttatttatgatgttatctatcgaaacaggaagatatttaggccagaagtaagcgagtcgagaaaaagctttggatatatttttaggaacggtagcaggatacctatccacgtttcctataatgaatataaacaaagcttaaaaaaatattctgagctatattctcacagtatacattttgacatagcatcttattttaatagtttatatcaccatgatataatccactggtttagctcaaaagaaggagttagccctgcggatgttgaagctctcggacagtttttcgcgaaattaactcaggacgaagtatcgattttatgccccaaggaatttatccggcaaaaatgatcggtaatgagtttctaaaattcgttgatttacatggtcgcctaaaatctgctcaaatagtaagatttatggatgactttactatttttgacaatgacattgaaacactaaataatgatttcatcagaatacagcagttattagggcaagtatccttaaatataaatccgtcaaaaaccacatttgacaatgtgatgggagatgtgaatgaaaccttaactcagatcaagtcatcacttaaagaaatcattacggaatatgaacatatacctacagcctcaggggtagaggtagtcgagactaatattgaaatcataaagcaccttgatgatgaacaagttaacaaattaatagacttgctaaaagatgaaaaaatagaagagtctgatgccgatttaattcttggtttttgagaactcataatgatagtttactttctcagatgccaatgctattaggcagattcccaaatttaataaaacatatttatacgatctgttcaggtattaccgataaatcaggattagtaaaaatattgctcagctatttaaatactaataataactttttagaatatcaattgttttggattggagcaatagttgaagactatctattaggtgtaggtgagtatggctccgttttacacaagttatatgagttatctggtgattttaaaattgccagagcaaaagtattagagataccggaacagggttttggtttcaaagaaataaggaatgaataccttagaaccggacaatcagattggttatcatggtcttcggctatcggtacgagaaatcttaaatcagcagagagaaactatattcttgattatttctcaaaaggctcaccaataaattatcttgttgcatcttgcgtcaagaaactttaatttaaaagccaccttcttgaaaggtggctttaaaaaatacctttagttcc (SEQ ID NO: 280)
27 28	10.A	gaggatttatgcacaaaatcctgatgcgaaatgttttcaaaaattgtcaggttaacgttcctgcagatctttgcgttacatgtcatttctggatcctttcccgacaggttaggttgtgattgatatgatgcccatctctcattttagtgatcgttatccctttataaacaggagtttatatgttatctatatgcaatagacttaaatcgatatacgtgcgcagcttacgattcacctctctacttactatttaaggaaaagagtgaggggagaattgattttcattaagatattatgagagaattatgactagtgaaatagtgttaaatcttgatttcccagaatataaggatgatttttgtactgatagcattgatgagcaagataatgagttgtggcagcaacaggccaataaaaagctactttcgtttctcgaggtgatgggggaggaagcaagacgatataaagaaaataattcccgtagtacgcatccacattataagacattgagtagttatcaccatgcaatctttatcagtggcgcgcgggggggcggggaaaactgttttcatgagaaatgccagatttagctggcaaaaacattataataaagatctaaaacgccctaagctatattttattgatgtgattgacccgacgctattgaatattgatgaccgtttttctgaagtcattatcgcttcaatatatgctacggtagaaaagc

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.	Cloned Sequence
	ggatgaagcaacctgatattgcgcagaatatcaaagataattttattaattcgcttaagacgttgt ccggtgcattaggtaaatcaaaagattatgatgaatataggggcattgatcgtattcaaaaatatc gttctggaatccaccttgaaaaatatttccatcagttcttgatttcaagcgttgagttactggatt gcgatgcgctggttttgccgattgatgatgttgatatgaaaatagataacgcttttggtgttctgg acgatattcgctgcctgttgtcatgtccattagttctaccattagttagtggggataatgatcttt atcggttcattgccaaaagtaaatttgaggaattattaaatcgtaaagcaaactctaattatgcta aagaaggcagcgagatagcagaaagattatcagaagcatatattactaaagtattccccagccatg tgaagatacccctccaaccgatagatgagttgttgccatatctttatatacattctaatgaagatg aaaataaacaacatacaagctattctgaatttatcaaacttgtacaacaaaaattctactttcttt gtaatgggcaagaacgaagcacaaattggccgcagccgagaagcgcacgtgaagttacgcaactaa tccgttctttacctccgtctactcttagtaaggaagatgattcgggaactgatttatggcaacgct tcgctgtctgggcggaagaacgtcgcgatggattagcattaaccaatgttgaatcttatctgttta ttaagaatgcgaaagcagtagaagatttaaatctgtcaaatcttattgcttttaatcctttactgc aaaaaggaaaatatccctgggcagaaaaggatttttataaacagcagtcccaacgtcggaaagagc tcaatgcccccgaaacaaattcaggtatccttaataccgtattttccgaacaaaggaaagatttta ttttaagaagtatgcctgcgctggaactcattatggagcctatgtatgtcactaagacggtagcag aaaaaaatgataattctgcgcttatagcgatctatacccattctgattattacagccagcagcaga acagacgatgtcatatattttttggcagagcttttgaaataatgttctggtcagtattagcgaaaa ctgaaaatcttccacaagaattttatgaaaaagataagtttaaatctttatttggtaatattttca aaaaagtaccattctactcaatattttcaatgaaccctacaaaggttgttgatgaagaaaatgacg atggcagtgaacctgatttttcgcaaaaactggacgatagcattaatgaactggtggaagatatat atatctgggcaaccagtaataaattgcgagccttcaaaaataaaaatttaatacccttaatgacgt gcgtttttaataaggtattttcacagatcaatgtactgagaaaaaacgtgcaggacagagttaaat ttagagatgaacatttgtcagatctggctaagcgatttgagtatatgtttattaatgctatcttta ctttcatcagagaaggggtagttgtcaataccaatgtggcaacaggcgcagctcctgccagagtac gtaatttatcagagtttaataggtatgataaaacattatccaggaatatgtccgggattttatccg tgaaagaggataatggcttaacgatagtcaaagagagtgagggcgatatcgcagatctgttatttg aaatttggcatagcccattatttaaattaacaaccaggacatgttacccaataggtaaaataaatt cgcaaaatacggcccaggaaaatttatcatcagattttaattcattttttgaaaatggtatcaact tcgaattgataaaacaatattattggcaaacttcaaatcatgataatatcaggacagcagacgtta gggaatgggcaacttcacgtcttaatgaagcaatcatccttttttcatggatgaaagaaagcaagt ctattaaagcgaaaattgacggacagagctacgagggtcggctctttcgcgggcttcagcaggcgc tggaaggttatgaggaggtctgagtatgtttaatcaggatccttattggctcattcctaccctttg tctggcatcagaccgaattttttatgcacaattgcgagaccacttaggccagaaaagtagcggtga acgcaaaaaagaaaaaaatggatatatactggtacaggcggcacaagactatcaattctattttgg cggccgtattcggaaagaggatgtgcaaaataatgccttaatgtggcagatagaaactggtaatga aaattgcttatcgatgcttgatagtttgtcagcatatttcctcacatggcgcggcaattgttttga ggtcaggcgtgagcgacttgaaccctggctgatgatctgttccgtgatagatcccgcatggattat tgcctatgcataccaacaattgattaaacaaaatgttgtatgtgatagtgagcttatttctttgct gacagaacatcaatgtccatttgcctttccaaaaggcagaggggacatttcctttgctgataatca tgtccatcttaatggtcatggttatagttcaatttcaatgctgaactttatagatggaaattataa ggttaaaaaagggataaaatggccctatcggcaggaatacaccctctttgaaagtggtcttctgga taaaaatgatcttccccgctggctgtccgcttatagctcttgcttacttaaaaatgtatataattc atttcaacaaggaaaaagatccgaggtagatttcacatgtctgaaggatgcggtcgaaacggtgct tgcggatgaggataaatattattttttagaggtagcttcgctatatgatgttgtcaccttgcagca aagagtgctttatgaagccgcccagcagaaatatcactcacatcaacgttggttactgtatacttg cggaataatgttaggtacagaatctgaagattatgcgaatgcgctggctaacctgatccgaatcag caatattctaagaaactatatggttgtatctgcggttggattgggacaatttattgattttttcgg cttcaactatcgtcgaataacaaagccagctgatacaaacaaccgagttcattatgattcttctgc tggtatttccagagaatatcgtgtctctcctgattttgtactgggtagcggcgtaatgcctgatat atatgccaggcaacttttcgatttttattgtacccaagcacgcaagggcgtacccgaacaaggaca tattgttgttcattttacacgttcctttcctgacaaaaaatcaacatatgataaattgctaaccga gtgtcgcgaacggttacgttctcagtgtgattattttggccgttttttaacatcgcttactttgca gtcgatagaatataaaaatttatctactgatgaagatcgaagcatagacattagaaaattagttcg tggctatgatgttgctggaaatgaaaacgagctacaaatagaggtatttgccccggttctccgggt actgcgtgctgctaaatttaaaggggagggggtgaactttaaaaggctacagcgcccttttattac tgtacatgctggtgaggattattgtcatatactcagtggccttcgggctatggatgaagccgttga attttgtatgttaggagaaggcgatcgtatagggcatggattagctctgggagtagatataaaact atgggcgaatcgccaaaagcgagcatacctgacggttggacaacatcttgataatttggtttgggc atatcatcaggcagtattactttctcaacatattgtcgagcatataccagtaatgcatgaattaag ggataagatccattattggtctcatcaattatatagtgaaacttatacgccagatttactctttaa agcatggctgctccgccgtaactggccggattataagtcaatcatatctgatccagcaaatatcaa tgaatgggtgcctgaccaacatattttagtcagtacagatgagactacagctaaggccagaaaaat ttgggaacgttatttaaatagcggtctggcagaaaatgatgtttttaacagaataatttcagtaaa ttgtgcgcccgatacagcgcaaaatttttcaatgacctttaatgaaaatgaagatattttatccaa aggggaattattattgtatgaagctatccaggatttcttaatcgaaaaatatagtaggttgggttt agtcatagaagcttgtccaacctcaaatatttatattggcagactggagaaatatcatgagcaccc attattccgttggaatcctcctgactcccaatggattaaacctggtgggaaatttaatcgctttgg attgcgcacaggacctttatctgtctgtataaatacagatgacagtgcattgatgccaaccacaat tgaaaacgaacatcgcttaatgagagactgcgccatacatttttatggtattggaacatggatggc ggatttatggataaactcaatacgcataaaaggtattgaaatattcaaaggtaatcatttaagtca

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.		Cloned Sequence
		ggatttagataatttaatctaaatgtaaacaagaaatccacgcaaatgcgtggattttaagtcaac ttattattctctgaaacggtttaaccgttcggaacaacagattaaatc (SEQ ID NO: 281)
29 30 31	10.B	tgtggttagttatcacagcactaacctattttcgagcttttttgattgaccaataccatttcttta attatgaataatgatgcgtcaaccgatggcgaacgggccaaatccactcttctacaactgcccatt gtcacggtgtggaataattaaaaatttagatttttgagattattctcattaccatcttgattta tttggttttgcatcaaaattcatagttcacaagcttttctcactccaaaaacaactgtaaagggat tattgtgaacacgatatacataccattagacagcggagagtctgcggttcttaaggatccagatac cttacttccccgaaatatttacgaacagcttactcgatttattgaaaaggctgttaatgaagtacc gaagcctcacgaagcgcttaatgaaacccgtagccataaggctatatcgattgacggcgcaagggg gacaggaaaaacgtcggtgctagtgaatttgaacgactatctgcagagtaatgctcagcaactggc ggggaaaattcatatccttgatcctatcgatccgactctacttgaagatggtgagtcgctgttctt gcatattattgttgctgccgtgcttcatgataaagagatcaaaactgcccaaagcagagacctcga taagtccagagtgtttacccagaagcttgagaacttggcacacggactggagtccgttgatttgca acagaatcaacgtggaatggataaaattcgctccttatatggcagcaagcatctggcaaattgcgt tgaagagtttttaaaatctgcgttggagttgatcggaaagaaattattgatactaccgattgatga tgtggacacttcactaaaccgggcatttgaaaatctggaaatattgcgtcgttatcttacctctcc gtatgttttgccggtagtgagcggcgatcgccgtttatatgatgaggtctgctggcgagattttca tggaaggttgaataaggattcagcatataatcgcaagaacacatatgatattgctagagatttggc aattgagtatcagcgtaaaattctgccgctaccgcgcagactgagtatgcccgatgtaagtgatta ctggcagcaagatggtatcgaagttacgctagataaaaatggcattcctctgcgtaattttatggc atggttgaaaatatttattactggccccgtgaatggccttgagggtagtgatttacctctaccgat accttcaatacgtgctttaacccagttcatcaaccattgcagggatttaattcgtgagcttcctga accattcagaaagaaagtcagtacgctggccttacgtcgtatgtggcaaatgcctgatgttcctct tgatgttcttgaaagttttgctgaaaaacatcgggaattgagtaaagaagctaagcgtgaatatgg ggaggcttacaagctattttatgatggactaaagaattttactgcttgggatagtaaggcttatct agaagatgataaacaatctgcatggctcgataggttgtgtgagtattttcgttttgaacctaaggc tggggctgtgtttttaacgcttcaggcaaaacagttctgggtctcatgggcgcagggtgacaatcg taatcaatcgattcttgcgactccgcttttcaacccttattgcataattttcgtgaatacgatgt ctttgaaaggtatgatgatctttctgattgggaatctcagttaagaacaaggttaccggagagttg gttgactgccattaaagggcaaaaaacgcttttaccctatcctgtagcagaagcgggaattaatac cagtttaaagtggaggtattgggaagaattagagaactatgggtttgatcctgctttggaaagcaa ggcaaatttcctttttgtccacgttgatgcagaggaatttttatacaaactctaaacagtcagtcgt gataaatattggtagagtttttgaaataattattgctagtcttgtttcggatttagagttggccga cttgcagagaattagacaacgttctccatttttactctgctagcgcgcttgcacctaccaaaacgtt agatttggaagaggattttacgaaaaagaatacaagatttatgaataacagaagtgaaactgacag agacatttctgatgatattcttgttgatgtgccggataaaaatgaggacgcatggaaaaaaatttg tgatgaaataaaccattggagaaagacacacaatgtggctagtacaaacttatcaccttggctggt ttataaggtctttaataaaacatatagtcaggttgctaataatgtgtttgttcccagtggaatgca aaatgttgatgcggctctaaatgtttttggtagggtttttttatgcagtttggtcagcatttggtag ttttgaaaaaggcgaattgttcggactatccgatgtggttgctacaactaatattatttcggcaaa aaatttttataatcatgataacttccgagtgaatgttggaccgtttacgcctgagcaaaaccaaaa ttctgacagcgatcgtgaggcatatcagcatcgcaaaatgtatggtgaaaaaaccagagcggtaag ttatgtattagcaactcatccgctgaaaaaatggatcgacgaggtattacgcactgagtttaaaca aaaacagaatgctcagattcagaccgagagaaaaatgccgattcaggctgagaaaattatagatat cagcccggcaagagagtttatcacaagaaaactttcattaaattcacactcccggttggttaaaac acgtataataaaacagcttaagatgttatatccaaactacgataaggctaaggacttcattgatga agttacaaaccacttccctcagaatgatcccgcaattaatacgcttcagaaagcatttgcagaact ttaccccgatggtgacaaataatgttaactcggtctctaagtgaacatgctgcagggtgtttttttc actgatgagcgtctgtcacaacgctttctagatatccttttatcgccacccaaggattttgaaacg tggtcatcattgcaggaggaatctttcaagctgctcgttaagagcatcgatagccgatatccacgc acttaccggttaaccgacgtacgccagcttgtggggaacatatgtgacaacgggttactgacgagt ccgacactaccttggctcgatgtcattgcggatcagttactgttgcggaatggcgacttactctat taccgcgaaaataaggttcaagactacgtgcgaatagctgcggaactcgaccctgcccttctagtg ggatggcgtcttggcgactggcttttgcaaagcccaccgccgcgattgacggacataacccgtgtg gtgatggcgcagaatccgtttttttgctccacctgctaatgcaggtaaaccttttgccgaggggcac gtacatctcgggggagtgacggctggagatactattttggatggctatctttttgaagagattgaa ctacccaaaagcaaagatatgttgttgtgggcgcacaaagagcatgatgagttaacaccgttgata aatcgagcaaagtctttgcttacagttctactttctgccccccctcaaacggtttctgagcaaact caaaatggttttgatcagcgtaaaactgtatctgagaagtacaaggcattacagaacccaatggat agcatccatcgtctcccagactggttattgcttgctaaaaagaatcgcggaactgaaagcgtcagc cccggctggttttaaaccaactggcgcatgcctccgaaaaaaaacatccctcgcgctggctgtgg ctgcagctatacctttgccactcttatcagcttaaagacactcatccactggagcgcacggcaata ctctgtttttggcttacggtaaatgcgctacggcgtcacattattatggacggacaggggcttgcg tgttttaccgagcgttattttaatggtgctttacgtgcgggtaagaaagctgacagtagcaatatg cgctacctgtttgccggtaaagacgatgtggccgaagtgaaagcatccccaaaggctttcgatcat gagatggtcactggattttcctcgacattgctgaaaaccctcggcattccagctgtttttccaccg tatatttttggtgagcatgagattaagccagatgaacgcgtgctgcgctatattggagcactggag cgctggcagttttgtgggcacttttctcgctctaaaactgcaagtcgcggcaagcgagcaaaggct gatttgcaggctaactggacagaagcggagcgattgttacagaaactgtacagtcataatggctgg aatcatcccgtcttcttagggggtaaacgtaacccacattttcattttcagccgtcgaactggttt cgggggcttgatgttgcaggggatgaaaacgtactaaaaattgcaggctttgccccgatgctgcgc tggctacgaagtggattatatcccgtaccagaagggcttcgcgccagtatgagttttcatttcagt

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.		Cloned Sequence
		attcatgccggggaggattacgcacatccggcgtcaggattgcgtcatattgatgaaacggttcgcttctgcgaaatgcgggagggagaccggctaggacatgctctggctctcggaattgaacctgcgctctgggcgaaacggcatggtgaaatgatactacctctggatgaacatttagataatcttgtctggcagtggcactatgctacgcttttatcggcttcattgcctctcgctcaggcggtattaccgctgcttgagcgtagaattgcacgctttattgcacggtgcgaatggtgcaaaaagagacctccgcaaatagataacagtgtggtggggaaacaggcctgtagtgatgataaacctctggaaaatattacacctgatacgctctaccgggcctggctactgcggcgtaattgttcatatcgactccagcaactccacggcggttcccctttgacctcgcaagagaaatgtgcgctgccggattgggccacgctcagcgataaaggcaatgtggcggcgcagctttatcagcaaagacactcgagtctccttgacgatatgccgccgcaactggtagttgtgcgtgtagcggacgaatggggaactcaggagcttattggcttgggaaatcctggtaaactgcgtcagcaggctcttgacggtaaagatatcctccaagacattgatacgccggtagagctgcaatttatgcatgctttacaggactatttgctagatcactatgatcgtaaagggttaattatagaaaccaacccaacatcaaacgtatatatcgcgcgattcaaaaagcacgtagagcatcctatttttcgttggaatcctccggatgaagaactgttgaaaccaggcgctgaatttaatcgttatggattgcgccgtgggccagtcagggttctggtcaatactgacgatccagggattatgcctacgacattacggacggaatttttactactgcgagaggctgcgattgagcgtggtgtcagccgaacgatggcagaatattggctggaaaggctgcgcctgtacgggctggaacagtttcagcgtaatcatttaaatgtatttgaagttattgaatagaggattttatcgtgagtggtacattcccttacttgcaatatacggatgtcaatgggctacaacctaagctcaaagaagagttgaaaaatttacggagaaaagagtatttgtcctactggcctcgttttctgatacgtagaatttcgctttatgctcttccattcctcatgttcttcactttttttcttttgtctgagtctgacgaagaaagttggggcagaggaagtgactaatattcttggaaccgtgagtatatccttcagtagttgcctgctgctggggattattatttctggtgtcgtgttactcttgcagtggacgtgcttcaactgtaaatacagtccgcaggatacgaatggagttgttggggctcgtaagttaaattataaattacttgctcatgttgtatttgttattgcatgcgtgcttttatttgtttttatttattgcaccaataataaagtgttttatggttttatcgtgtttcttggtttgacattattaccattggtaattgaccgtaccttgggggtgactcgtcaaaatgaacgtcacaaactctatatcagaaggttagagcgcctcgatgaattgaatattctccgggagaaaatgaatattaaattcgaagaatcccatttcatcgagtatatgaagcttgttgatgaagctgatcacggaaaaaaccaggatacagtaagcgatacatcctattttatgacgttgatagaaaataagctaaaagtgtaatcggttttaatatgatgctgtataaaaaactacgcaattgcgtggttttttgtcggactatgagggcaaggttgccctaaaacagaggttaaacgttgggatgtgatttattgcacatcatgccgtgcccatccagtagaatccggttcgaaatgtgtataggattgtgtatatgtttctgttcggtctcggattcttatacac (SEQ ID NO: 282)
32	11	ttttagaaatattgtgtaaaacttcttactctttactggtcatccctcagtcgtggaaaaaacacactgttccatataggttttatttgtgatataatgaacaagttcttatttaagaaacctataaacattaagcgacggaaatatatcatgaaaatagtcagcaataccgtttgggatggacttaaactgcctgattatagggctcgtttttttatagaagtttggaaggagattttgtacgtcaacactccttcattttatcaatctaaaatgattaatacgatgtcaggtgccgaggagttagtcgaagccattgatgattacatacaagatgataagagtaaaaaaagcttattatcaatgatagaagattacaaaggtaatttaaaaaaagactctatagcaaaagacacttttaaaaacttgcatgcaacgctgttaaaaaaaattgagactgttcctgacccaatatctagtaattatattttagaattaaaaacaattgttaaattagtattatccaaagaaagtgactattatcacgaacttaaaaagcagctaaaatcatctattttgtctaacgctgatttgaataaaaaagcccgtttaatggactccatttatcaattaactaaaagctttattggctatctcctgtggaaggggtattcaccaacttatttatataatagaatggagtatcttacgagaattaaaaattatggcagtagagacttttccgctcaatttaatagttgccttgataaattaactattaggattcatgattatacagtttattttcttattacccctttgtctaaatatctgattgaattgaataatatccttgatgttagctttatcaatcgagaaggtattattaatgaaaaaaactacaataaaatttcacaaggggttgaatcttcggtattagccaaaattgttgttaatacaacagactacgtttccgcggcgtggcaggcaaatgaaaaactggataaagtcatagattatttagaaatagagaagccagaatataatattagatattctcctgtatgtcttacagagttttcaaatggtagattcacacaccgtcagactataaacataggcagattgaaacaattcattacaagtaaaaattacagcattcttgaaaatatacctaatgagtccaaggtactcttacgagagtctataaaactagacagatatgatgtactgacaagatctttaaggtatttaagagttgcaaaagaatcaacttcacttgagcaaaaattgctgggcgtatggatagctcttgaatgtattttcgagagcacatcaggtaatatcatttctggaataactaaccatatccctacgttctatagcactcaaagtctagaaattagaattagatattctaaagatttattagaagcccgattgaagcctatttcagatagccttttagagattacagccaatcagaaatctaaatttcgagacctttctttaaaagaatactttgacatagtgaaaatcgaaaaaaacaggaataaaatttcgatgagttagtttccaaggggggatgagtttgccgtttttcgactaataaaaatatttgaatcattcggaacgtcaaagaaaataaatgatagatttaatgatactaaaaaggatgttgagtctcagctttatagaatttacaaggtaagaaataaaataacccatagagcatactacggaaatattaggccccaattagtggatcatctttatagctatttactaagtgcatatagcacactaatttatagtttaagatataatgcaataaataaatttgaaccacaagatatgtttaatgcatatattatctcgtgcgagagtttaatattcaatgttgaagaagaaaaaaaacttgaaaatataactatggatgaaataattttatcatagtgaatgttttctaggtgtcgtattc (SEQ ID NO: 283)
33 34 35	12	atggtagcgataaaaatgtatccggcaaaggatggggatgcttttcttattatttgcgatgaggaaaaaagtgcatttctgattgacggaggctacgcggaaacgttcaggcaacatattttgcctgacttacgtgagctgagttttaacggttaccggttacgtctggtcatggcaacacatattgattcagatcacattggtggtctcgtggacttctttcttgtaaatggacacgcagcagagcctgcagtgattactgttgaccgcgtatggcacaacagcctcagggcgatgacgagacccgaaaataatgcacaaaaagtggattcccgagaaatcactgacttttgagacggagatatcatgtcgaagccgataaagccaaaccgcatgaaatcagcgcgcgtcaggggagttcactggctgccagccttctggctggcgattatcattggaat

TABLE 15-B-continued

Sequences of validated defense systems (Cloned sequences corresponding to row No. 1-68 in Table 15-A)	
Row No.	Cloned Sequence
	gagggaaaagggtatcagtgtatctgcaccggtacctccattcccaacttgatgtgcgataacagt
	ctaacaattctgagcccctctaaggagagaatttcagcgctctgcctgtggtggcgcagacaactt
	gcatcgctgggcttttcgggacggtcctcctcgagtgaggcatttgatgatgctttcgaatttttt
	tgtaaaagggaagcatctcaggttcctcttccgcatgtcatcaatgcaagaacaccgttgcttgag
	agggattatgcacgggatacctcgccaacaaatggcagttcgatagcgttcagtctggtgctcaat
	aagaagagaatattgatgctaggagatgcctgggcggaagaagttgtgacatctctgggtgccagt
	ggggcgtcccatcattttgatatcattaaaatctcacatcacggtagtattagaaacacaagcccg
	aatcttttaaagatcatagatgctcctgtgtacctgatctcaaccgacggaaaaaagcatgccaga
	caccctaacctggcggttctgaaagcgattgtggacagacctgcggcgtttacgcgaacgctctat
	tttaactatgccaacagcgcatctgcttttatgaaaaattacctttctgcaagtggtgcacaattc
	agaatcattgaaggatcaacggattggataacactgtgagatatgctgctactgaaactgaaataa
	ggaacgcaactgtactcattgaatgcgcgggttacactggttccggaaccctgatcgcagcagaca
	aggtccttacggctgcacattgtgtagtatcggatgatcctgagacaccaattacagtgacatttt
	ttggtgcggatgaagacgtctgtgtcaatgcgacaatttcagaaatagatacatcgtgcgatgcct
	gtctgctaacactttctgactctgtcgacattccgcctattacacttatgacacagccggagcgag
	agggaagccaatggaaagcctttggctatccggcatcacgcaatgggccatcacattatcttcatg
	gcactataagtcagattttaccaaggcttttccatggcgttgatatggatttgtcggtcagtgccg
	attgtgttctggaagagtacagtggagtttctggtgccgccattctatcagaaaataaatgcattg
	cgatggtgcgcatcaggatggatggtggactaggtgcagtaagtcttgataagttaagcggtttgc
	tgattcgaaacggcctcatcccagatgacattgcatccctgccagattcatcactgtcgggtgaag
	ttgtcctgaaccgcacagaatttcgcgacaactttgaatcgttcgtcctggagcacaagggacgtg
	cagtgcttttggaaggtagtcccggctctggtaagactaccttctgccgccattatcagccccgta
	gtgagcaactcgcagtggcgggtgtctatgaatttacaccggaagacggtgctggtacgacattca
	aaattcttcctgaggtatttgccgattggctgcataaccaggtttctatactgctttcaggtaggc
	ctgctcgcagggaggaaacagaaaagatcaatctgacccaaaaggtgtctgaccttctacatactt
	tctcagattactggaagcacaaaggaaaatatggcgtcattttcattgatgctgtgaatgaggcaa
	gcgagtgcggggatgaggcagtatcgcgctttacagcattactgccggtgacacttccggagaacg
	tcaaacttgttttcaccgcaccatcattatcatcagctggtaaggctttccggcactggctcacac
	ctcaggattgtatcagcctaacgcttttaagccatagggaggtgttacagctaacagctcgagagc
	ttaaaacttccgccccttctttgtcactactcacacgagttagtgatatagctcagggccatccac
	tttatctccgatacattcttgggtatctgaaagcgaatccggatcaggttaatctggagatattcc
	cggttttcagtggcagcattgaaacctactacgaaaggctctggcaggggctggttaaggatgaga
	gcgctgtaaatctgctcggtattctctcgcggatgcgctggggcattgatatttcatcactgatcc
	ctgttctaacaccgcaggaacagacggtgtttgttccaacccttgaccgtattcagcatctgcttc
	ttaatgataaatcatcagcattgtgccaccaatcatttgcggcgtttatcaacagtaaaacggcgg
	taattaactcgctgctgcacggacgccttgccgacttctgccttaccagtggagagagttatggcc
	tgattaatcgcgcttatcacctgctcctagcctctcacgacagacatcctgaagccgcattggtgt
	gcacgcaggaatgggctgacgcctgtatcgtcaagggggctcagccggatattctaattcacgata
	tccgtcagaccctgaagaacacgcttattcgtgccgatgcagtggcatcgattcgtctgttgctgc
	ttttccaacgcatgaccttcagacaccattttttgtttctgcagtcagcttatcactcaggccttg
	ccctggctgcacttggcagaccggatgaggcccttgagcagctcataccatctggaagcctcgttg
	ttgatgcagttgatgcaattgtcagcgcacagactctcgcgcgtatgggaaacagtgaacacgcgc
	tgaagctattggaaaaggtgaagtcagctgtcgaccaagaatttgaacgcaatcccgtcaatctat
	ctgattttatcggcctttccctggcttgggtgagagctgagctgatggctggggtggttgatggcc
	acggacgcacacgcgaggttgttgagtatttgtacggttgtgggcaagtcgttcgcgataattttg
	aacaatcagcgcatagtaaatcagcatatacacgcgctttttatcctcttcaggcagaaatggaag
	ccgtgaacatagcctttaatgaccgctccgtatctttacggacggttaaagaaaagtttggtagct
	taccggaaaatattcttgatctgatgctcagttcagttatgcgggcacatgacatcattctgcaac
	atcagttgccgatgccccagcatgctttgcaacccgtttggtacaatctggacagattacttcata
	ctgatattccgtattcgaacgaaattcgttttaattcattaagtagccttatttttttcaatgcgc
	cttctgctcttattatcaggatggcgggggtattttctttcgaagtagtacccgaaataacgttgc
	tcaatgaagaaaatgagatagcagcagacagcattgacgttagtgaacagggacaactctggctgg
	tgagcgcctaccttaatgaaacgcaaccctgtcccgatattaaacatccgagtcagggatgttctg
	aatggctcaagacattgactgaggctatttttttggtacagcgggcaggcgcgccgggcagttattg
	acggcaacgatgagaaaaaagaactgcttttagtcaaggtgcagaatgatattctccctgctcttt
	cgtactcgctggaagagcgcatggcatggccgaattcatgggcaatgcctgaacagattatcccca
	tgatttacgaagagttagtaaacatgttcggcgcatgctggcccgataagatatcagtgatcactg
	atttcattctggctcatacgcctcagcaatgtggactttattccgaggggtacaggcgtttactga
	acagagttattcagactcttctaaatgagcatcggtttttggggcaatctgatacgacatttcaac
	tacttgagacgttgcatgcgtttgtttctgcttttactgagaatcggcaggagctggttcctgaat
	tactgaatattattccagcttatattagccttgatgctcctcagctggcacaggacacttacactg
	agcttttaggtgtgtcgatgggccctgactggtacaaagaagaccaatttgccctcatgacaacta
	tgctgcgcgtgataccacagcatacagacacaaatactacactttcacaagttgcaggattccttg
	aacatgcttcgggtgaaatgacatttaggcgttatgttaggcaggaaaaatcacagtttattggcg
	aacttattcgtcgtgggaattatgcacacgggtttaactattatcgtcagcagtcctgcggatccc
	atgaggaaatgctcacccaacttagccacccagctgcagatagccctcatccattgaaaggcatgc
	ggttcccggggggagcgctggatgaggaacatgctgtagaatgcattgtcagtgaactgcgaaaca
	gagtcgactggcggcttcgctggggacttcttgaaatattcagctttggcagtattggtaatcttg
	cagtgccctttgctgaacttatcaatgaattttctgcagacactgaagaccttaatgaaataccca
	aaaggttgcacaacattttacatggtgatgtgcctttctcagaacacagaaattttatcaaaaatt
	tcacagagcaccttgcagacaaccataagccactctttgctgaatttatcagtttgctatccgaag
	acactagcgataacgacgttaagcctcccccctctggtgatgctaaccagaagggtactgatacct

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.		Cloned Sequence
		cagatgatgtggcaatgcagccaggactttttgggaagcgttctgcgatcaatagggctgaagcctgcatggaaaatgcccgaaaagccgcagcacgcagaaacacagttcgtgcaagtgagttagccgttgaaagcctgcatataattcaggatggtgactggtcagtctggagaaagaacaaccatctggcggaacttacacggacgtacatattggacaactctgcggatgcaggttcggtcattcgtgcttatgcttcgcttgtagaaaaagaacgttatgccccggcatgggtaattgctagtcatctcatcgaaatagcagccagtaaattctctgatcaagaagcccaagctattaaccagatcgtacttgaacacaaccgccacatgcttgggaataccgaagcggatgctgcgcatttttcttttcttaatgaacctgatacctcagatgcaggtgaagaaacactctattttctgttttggctgctggaacacccactgaaattcagacgcgaacgggctctggaagtactgaagtggcttgcatcagacgatgataagattctgggccaatgcgtgacggaggcactcgtttcagacattgcctcacgagctgaagcactaatggcattgacagactgggtgtcagctagatctcctcagcgaatatgggactttatagttaaagagcgcagcctttttgaatggcttgaaggcactactgcactaagccaagtccatctcctggagcgagtaaccagcagagcgggatttgttttaagaaatgagattgccgcatttgagcgaccccgaaagcttttactgacatcagaagcctctggacaacggaatattccagaaaatttaccaacatgggtgcaatccttgtcgcagacccttgccgtgatggaaaagcagggaatagatatcccagctttgcttaccttactcgaaaaacgggttttacagcagagtggattggctgatatcacggtggcttttgagctggaaaagttacttgcgcgtggttttactgtgaatagaacaccaagtcaccatcgctgggagacgatggtgcgatttgcattaaaccagatcatacatgaggcggccgcacaggatgaactgcaaaacattgaacccttgctacgtgcctggaaccccgcgtcagaggagtgtgttgagccgtgggaggtttgtaaccgggcaaaacagattatctgcgctgttatggaaggtagacatcagcaagcttcgggcatagaggatggcttttcttgcattatcttgatgaagtggaggtttcccgagaaggtcaaacgcatctggtggaaatctcagcggtgttaacgacagctcataatggtcatgagagccttagaccaggtgcagaaagcgaatttaatgcaacacagacacctgatatagagcggacgcttagtgtgcaccttacatgccagcgagtcaaaatgcagcctttgctttttgggggagctacgcctgccgcagtgtcgaaaaagtttatgcagatgactggaacgttgccttcagactttattcgcaggcaatggcgaagcgggcgttctcttagtaaaaacagatgggggaaccaataagcagaggaagtctgttactcatgaaaagaacaactaccctccctccaggactgggcttagcgtggtatgtcactgtcgatgggaagttgatgaatatattttcatatgccccgaggaggagataatgaaatacagttcaatggaaacgccaaaaacgcgagaggaatttgaggctcgctgttttcacctgctcaatgcgatcaagttaggacggtatcatggcattccgggtgaaggtaacaaagagcaggttcctttctccctaacggacgagttgatctggcaaacattgataccatgactcgcctctcgatgaactcgttatatgatttccactataacagggataattatccgcagtttgatctctctgaaaatgacgagaatgaagaggctacggattga (SEQ ID NO: 284)
36	13	gggatttccaccacctcccaccgaccatctaagactttatgccactgtccctaggactgctatgtactaggagcggatgttaaactcagactcgtttcagctacattgcgttttgaataatattccatcataataactctttgaaaaatgtgatcttttcatttataacactgatgacttgcttatctcattgggatatcggaggagaatacttaactatgacaagcccgattattatgacactggctatattatatagattgatattaaaatgtaggattaggttcttgccaaggtgtcaagatttacagataggtttaaaaccatataaatatgttttacggtgagatacaatacatattgtaaggcataaacgcttggtaaaattttaattattggaagaagctaatcatggaacccatatcaattacagtggcaacttatgtagcaactaaacttattgatcaattcatctctcaagaaggatatggttgtattaagaaagcattattcccccaaaaaagatatgtggatagattatatcaactaattgaagagacggcaattgagtttgaagaaacatatccagtagaaagtggagcaataccattttatcattccgaaccattgtttgagatgttgaatgagcacatctttttaaagagttccctgacaaagagatattattagacaagttcaaagaatatccaagtatcactccccaactcaacaacaactcagcctttttatgagatgttatcattaaaaatcaataattgttcgaagttaaaaaagctacatatcgaagaaacgtataaagaaaaaatattcgatattaatgaagagctcattcaagtcaaacttattttacggtctatagatgagaaactaacttttcacttaagtgatgattggttaaatgaaaaaaatagtcaagcaatagctgacttgggaggtcgatacacacccgaactcaacgtaaagctagaaatagcagagatatttgatggcctcggtagaactaatgatttttctaaaatattttattcgcatatagatagctttctggtcgctggaaagaaattacatagttgcgatgtaatttcctcagaattatttgaaataaaccagtccttaaaagaaatttctgatatatatcaggagattaatttttctaaattagatgaaatccctataaataaatttaataactatgtttctagctgccagacagctattggcggagcggtatcaatattgtgggaactccgagaaaagtcagagcaagtaggtgaaaccaagcattacagtgataagtattcatctactctgcgaatgcttcgggaatttgactatgcgtgcaatgaattacgtatattcattaattcaacaacagtgaagttggctaacaacccattcttacttctcgaaggaaaagcaggaattggtaagtctcatttactggctgatgtgattaaaaatcgaattgcttctgggtatccttcactactcatactagggcaacaacttacttcagatgaatctccatggtcacaaatcttcaagagattacagcttaaaatcacttctcgtgaattcctagaaaaactgaatttatatggcaaaaaaacaggaaaaagagtcttagtttttattgatgctattaatgaaggtaatggaaataaattctggaatgacaatattaacagttttgtcgatgaaatcagatgctttgaatggcttggtctgataatgtcagtcagaacaacatatagaaatgtaacaatttcacatgagaatgttgtgcgaaataattttgaaattcatgaacatattggattccagaacgttgagttggaagcggttagtctattttatgattattacaatattgagaggccttcatctcctaaccttaatccagagtttaaaaatcctctatttcttaagttattgtgtgaaggcattaagaaaaatggtttaaccaaagtgcctgttggatttaatgggatttcaaatatttttaactttttagttgaagggtaaataaatcattagcatcgccaaaaaaatatgcattcgatcccagttttcctcttgttaaagatgctctcaatgaaatcataaaattcaaattagagattggtcgtaatagtatttcacttaaagatgctcactcagtggttcaatctgtagttaatgattatgttgctgataaaaccttcctcagcgccttgattgacgaaggattattgactaaaggcatagtgagaaatgatgataattctactgaggaagtagtttatgtggcttttgaaaggtttgatgatcatttaactgttaatttttttattaaatgatgttgaaaatatcgaaagtgaatttaagcctgatggtcgtctgaaaaaatattttcatgatgaatgtgattttttatataaaatcgggaatagtagaggcgttgtctattcaattgccagaaaggtatgaaaaagagctttatgaatttctgccggagttcagcaataatcttaaattactagaagcctttattgatagcttgatatggcgc

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.		Cloned Sequence
		gatattaaggctattgatttcgaaaaaattagacctttcatcaatgaacatgtttttaaatttaaa gatagttttgatcatttcctcgaggcagtgatctctatttcaggtttagttggccatccctttaat gctaatttcttgcatgattggctaaaagattattctttggcaaatcgagattcgttttggactaca gaacttaaatataaatatagtgaagactcagcatttaggcatctaatcgattgggcatgggccaga acagataaaagctttgtttcggatgagtcaatcgagctagttgcaactagtttatgctggtttta acttctagtaaccgagaacttcgagattgctcaactaaggctttagtgagtttactcgagccaaga attcctgtattgagaaaaataattgataagttttatggtgtaaatgatccttacgtttgggaaaga atatttgcagttgcattaggctgtacattgcgaactgataatattaaagaactaaaatatttagcc gaaactgtttaccaaaaggtattttgttctaagtatgtgtatccaaatatattacttagagattat gctagagagattattgaatttgctaatcatcttggattggaacttgaaagcattgaattatccaag actagaccaccctacaacagcatttggcctgacaagattccttcaaaagaggaactagagtccctt tatgataaagaaccttatcgggaactctggagctctattatggaagatggtgacttttcacgatat actattggaacaaattataatcattctgattggtctggttgcaagtttaatgaaacccctgttgac cgtaagcaagtttttaaaactttcaaatgtaaactaactgatcaacaaaaagacttgtatgatgcc acagatcctttcatttatgatgataaatgcgaaggaattaaatttggtcgtgtggtcggtagaaaa gcacaggaagaaataaaggcgagcaagaaattatttaagaattcattgtcatacgatctgttaagt gagtttgaaaatgaaatagagccatacctggatcataataataatctgctggaaactgataaacac tttgatcttcgactagctcaacaatttatattcaatcgtgttatagagcttggttgggatccggag aagcatggtaattttgaccaacaaataggaactggacgtggacgtagagaggcattccaagaacgg attggtaaaaaataccaatggattgcttattatgaatacatggcaaggctagccgataattttact cgttttgaaggttatggtgacgaacgaaaggaaaatccataccaagggccatgggagccttacgta agagatatagatcccactatcttacttaaagaaactggaacgaaaaaaataagcaataaagaaatg tggtggcttaatgatgaagtgtttgattggacttgctctaatgaagactgggttaaaagttctact actataactaattcatatgcttttattgaagttaaagatgataatggtgatgaatggatagtatta gaaagtcatccatcatggaaagaaccaaaaattattggaaacgatgattgggggcacccacgaaaa gaggtttggtatcagatcagaagttatatcgttaaagttgaagaatttgaaaattttagatgttgg gcaatagctcaagactttatgggcaggtggatgccggaatgtactgatagataccaattatttaat agggagtactattggtccgaagcatttaagtcttttaaatcagattattatggtggatctgactgg acttcggtaacagaccgggagtctggagctaagatagctgatgttagtgtcacttcgattaattat ttgtgggaagaggagttcgacaaatcaaaaatagaaactttgaatttttgaagcctagtaactta atctttgaaaagatgggattaaaaagtggggaagtagagggtagcttcaatgatgaaaatggaact atggtttgctttgcagctgaagctgtatatgcttcaaagccgcatctacttgttaaaaaagaacca tttttaacaatgttaagggacaatggttttgaaatcgtttggacattattaggtgaaaagggcgtt atagggggctcactcatatcaagtcatcattatggtcgacaggagtttagtggagcattttattat gaagacagtcagctaacaggaagtcataaaactagctttacgagataaaaatgaatctcagagctg aatatataagtagtattagaaaccgggttatacttaagaaatcaatcttaagtgtggcagtcgaat ggtagctaatatgctagcggcgctaatgcctgtttgttgctcataacaggcattcactttagttat ggcagaaaagtatacatgctgggttgggaaagtgtgaaagaaaggaagattgctgcgccgtttgtc gtcacgtttatcttcattggctatgca (SEQ ID NO: 285)
37 38	14	acaatttttgccataagacgctttcctgaaactcttctcattctcagcaggaaagcgttctcttc tcaatactctctggttatagagtattaaaaaataaggagttataatccttgtagcccaactgacat aaggacgatgctcaatgtctgacagcctgcttgttcgcaccagtagagatggcgatcagtttcatt atctttgggcggctcgccgcgcccttcgactactggaacctcagtcaactcttgttgccctgacca ttgaaggggcatcaacgacggaaatgggctctcagccagtggttgaggatggggaggagctgattg atattgctgaatattacggcagtaacgagctcgcaacagcaacaactgttcgttatatgcagctaa agcattcaacaatgcactcagatactccatttccccctagtgggttacaaaaaaccatcgaaggtt ttgcaacccgttataaggcacttatacaaaaaataccggtagaaacgttacgcactaaactcgagt tctggtttgtgacgaaccgtccagtcagtagcagcttcagtgaagcgatcaatgatgccgcgaacc aacacgttacacgccatccacatgatctggcgaaacttgagaaatttaccgggcttcaaggcgctg agttatcgatattctgccagcttttacatatagaaggtcagcaggacgatttatggagtcagcgga atatcctgctaagagaatcagcgggatatctccccgacctggatactgaagcccctctgaaattaa aagagctggttaacagaaaagcgttaaccgaaagcgccgcaaatccttccattaccagaatggatg tgttgcgtgctttgggggtggatgaaacagatctttttcctgcgccctgtcgtattgaaagaatag aaaattccgtctcaagaactcaagaggcgacgctggttcaacgtgttgttgaagcattcggcgcac ctgtgatcatccatgccgatgccggtgtggggaaatcaattttctctactcatatagaggagcatc ttcccactggttctgttagcatcttatatgactgtttcggactgggtcagtaccgtaacgcgtctt cctaccgccaccaccatcgtacagcattggttcagatggctaatgaaatggcatctcgtggtctct gtcatccattgatcccaaatgctggtactggcatatcccagtatatgcgtgcgtttctgcatcgcc tttctcagagcatttcaatactccgggcctctgagcccttggccgtattgtgtattattattgatg ctgcggacaatgcacagatggcggcggaagaaatcggtgaaacgcgttcttttatcaaagatttaa ttagagaaaagcttcctgatggagtctgccttgttgcactttgccgaccttatagacgggaattac ttgatccacctcctgaagcactcacattatccctacaaacttttaatcgcgatgagacagccgctc atcttcaccaaaaatttccagatgccagcgaaagtgatgttgacgagttccatcgtctaagctctt gcaacccccgggttcaggctctgtcattatcacaaaatcttccactgaacgacacattgagacttt tggggccaaatcccaaaacggtagaagatactattggtgaagtgctggaaaaatccattgctcgct tacgtgatacagccggaatatctgaacgtgctcaaattgatacgatttgttccgcactggcaatat tgcgtccattaattccattatctgtgctatctgccatttccggagtagctggttctgctattaaaa gtttcgcacttgatctgggacgcccgttaatcgttagtggcgagactattcagttctttgatgaac cggccgaaacatggtttcagaggcgctttaggccatcggccgctgatctgcatcagtttattacta aactgagaccactaacaaaagatagttcctatgcagcatcagttttacctgcattgatgctggaag gaaaccagctttctgaactgatcgagctagcgatatcctcacaagctctgcctgaaaccagcgcgg

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.	Cloned Sequence
	ttgaacgcagggacatagaacttcaaagattacagtttgcgttaaaagcagccttacgcacaggtcgataccaggatgcggctaaactggcactgaaagctggtggagaatgcgcgggtgacaacaggcaaagagtcctgctgagggacaatatcgatctggcagcaaaatttgtgggaagcaacggcgttcaggaactggtttcccgtaacgcatttccagatactggctggcctggctccagaaatgcttattatgccgcaatactttccgaatatcctgaactctcaggagaggcccgcagtcgccttcgactcaccatggagtggttaacaaactggagtcaattaccagatgatgagcggagcaggcaaaatgttaccgatcaggacagagcggtaatgctcattgcctgcctgaatattcatggcgcggaagcggcagcaagggagctcagaaggtggcggcctcgaaaactatcttttgacgctggaaaaattgttgccatgcagttactggcccacgcccgttatgatgaacttgatcagttggctattgcggctggaaacgatatcagcctggttatgggaattgtactggaagcaagaaaacttcaccgtccagtcgctgaacaagcaatcagaagaacctggcgcttgttaaaaagtcagcgagtcagcattaaagacagaaaccacgctaataaccagacaatagcagcaatcactggcatggttgaaatggcgcttatccaatctgtttgtactgaatcagaaagcatccagttgttggatcgttatttaccaaaggttcccccctatgctctgacttctgagtatagtaaagaaagagttgcttacgtccgggcatatgctctgcaggcaaacctgatgggctctcaattagcgcttagcgatttagcctccacagaggttaaaaaagaacttatggctgaaaaacgccacggcgaatctgatgacctgcgtcaactgaagcagtacagcggagtattaatcccttggtataatttatgggccaaagtaattcttggtaaaacaaggaaagcagacttagaaagtgagctaagtgatactcaaaaagaatcgacggctattaaaggtcattcttactctgagcattcattatcatcaaatgagatcgcaaatgtatggtttgatattctgatcgaagcaggtaatgtatcaaaagacgatgtggaaaacatcatcaaatggagtcagcataaagggaatagagtattcacaccaacgcttcaccgtttcagttctgtatgtgcagagatttcagggcttggagagctttcatatcacttcgcagaacttgccttatctttatggagggatgagcactctgatgctcagatcaaagctgacggctatatagacctttcccgttcactcatttcacttgatgaaccagaagctaaagaatactttaaccaagcgattgaagttacaaataagttaggcgatgaaaatttaagtcgatgggaagcgatacttgatcttgctgaatatgttgctggtaaaacgcaagtccctcctgaaatttcctataaactagcccgatgtgcggaactaaccagagaatatgttgatcgtgataaacattttgcatggagtgatactgttgagattttggctgagttatgtccatcttcagccctagcaataataagtcgttggcgtgaccgtacatttggcaatcatagaagcatactggcatggaccattgagcatcttgtaaagaaaaataaaattaatgcactcgatgcacttcctttaatcacatttgagaatgattggcataaatgcgacttgcttgattcagttttatcctcgtgtactgatgacaaagataagatcatggcattcgaagtggtttaccactatacaaaatttaacgtacaaaatatccaaaatcttaaaaagctggatgctatttctacatcattaggtattgaacacacagaactgaaagaaagaatttcaggtctacaacatactgagacggtttcaaaaaaatccagtctctcatcgaatgataatgagcaaggccatgaccaggaatgggagtccatttttaaagattgtgatttatcgtctattgatggtattagtgcagcatacgaaaaatttcgtaatgttcctgaattctattccaaagaaaccttcatcaagaaagcaataagccgagttaagacgggcaaagaatgtagtttcattactgccattggtgctatatttcactgggggctttatgattttaaatatattcttgaatctatacccgacgaatggacatctcgtttaagcattaaaaccaccctggcaggtttaataaaagaatattgccaacgcttctgtatgcgaatcagaaaaagtcgcgtttacgagatttttcccttcagtctggccagcaggctttctggtataagtgaaaaagagattttcggtattaccctggaggccattgcagaatcgccagagcccgcaaactctgaccgtttatttagccttcctggccttcttgttagtaaactggagagtaatgaagcgttagatgtattatcttatgccttggatttattcgacgaggtgctaaaagatgaggatggtgacggcccatggaacgagaaattatctccgccaactcatgtagaggattcacttgcaggctatatttgggcgcggctgggttctccggaggcggaaatgcgctggcaggcagcacatgcggttctggcactatgtcgaatgagtcgtacatgcgttatacaaggaattttccagcacgcaataaatgctaccactttacctttttgtgatcgcaatctgcccttttataccctccatgctcaattgtggttgatgatcgctgctgcaagggttgcgctggatgatggaaaatcgctgattcccaatattggttatttctaccattatgccactactgatcagccacatgtattaatccgtcattttgctgccagaactttacttgcactgcatgatagcgacctgatctctatcccagcacaagaagagaataaactccgaaatataaaccagtctacgactctccctgtgcttgataaggttgaagatcatagaggtgaagattcatatacttttggtatcgactttggcccttactggctaaaacctctgggacgttgtttcggtgtatctcaaaaacagttagaacctgaaatgcttcgcattattcgtgatgttcttggttttaaaggtagccgcaactgggatgaggatgagcgtaataaacgacgctattatcaagacagagataatcatcacagtcatggttcctatccacgggtcgatgactaccatttttacttgtcataccatgcaatgtttatgaccgctgggcagttattagcgacaaaaccattagttggtagtgactacgacgatgtcgaggatgttttccaggactggttaagaagacatgatatttctcggaacgatcatcgctggctcgccgatcggagagatattccccccaaagagcgctccagttggcttaatagcagttctgacaatagggatgaatggctagcgtcaatctctgaaaatgtatttaacgaaacactatgtcccagccccggactattaacgctatggggacgttggtctgacgtttgttcagatcgaaaagaatctattattgtccattctgcgttagtatcgccggagcgatctttatcgctcctcagagcattacaaacaactaaaaatgtatatgactataaaatccctgatgctggagataatcttgaaatagatcacgcacactatcagctaaaaggatggattaaagatattgctgaatactgtggaattgatgagtttgatccctgggcaggtaatgtaaggtttccaatcccagaaccagcctcatttatcattgatgcgatgaaattaactactgataaagatcatcgggtatggtattcaccttctgatgttgaaccggcgatgatttccagtatctggggccatctatcaggtaaaaatgatgaggaaaaatcacatggttataggctatgtgcttcaatacacttcataaaatcagcattagaaacattcaacatggatctcattttagaggttgatgttgatcgctattcacggaacagcagatatgaacggaataatgaaaatgagctcgacaatatcccttcaagcactcgactcttcctcttccgacatgacggaaccatccacacgctatacggcaattatagaaatggggaaaaaactagttgatgagcttgagctaaatgactctgttgatacattaagcagatggatggctcatcatatcgcagagctcatttatgatgctgaacattgtacagacgacatcgtccgtacagctaaacaagcggagattagggactctatctggtcattctggtctaacagatacgaattgccaattggtagcagaccatttcaggagctcgaacctattctaagaaccttaaaaggtcttgatcctgaaaatgagcaaccgagatttttttcaccttaccgagatctaattaatgtagaaaaagaaaccagtgaggtccaaaaatggctaaccgccgctaaggatattgattcagcagcaaaaatactgattgattactgtttatcgttagcagcagaaaatgctatcgataaatcccaagaatgggtggaattagcacagaaagctgga

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.		Cloned Sequence
		ttgaacaaagatgttgatctgcttgaaattcgtatctttcagttacgaggtaccccagccaatacagacaatcccaataatgcacaacggagaatactggaaaaaaggcaaaaaaggcttgaagcttttctcttattgggctcccagttaaacgaacaactcaaatctcagcttgaagccttaccagcaattgaggatgagccaacggatgacgacgaagacttttgatatgacttgctttagcactggagacggctcacaagacggaccacataatagcctaacccaagacttttctactagtcctaatg (SEQ ID NO: 286)
39	15	gcgcagctgacaaagattgaccgtgagcgctctgatggagaaagacgatagttgctgagtacgatatcgagggtacatttctctgtgtaggggtagttatttacaaaaaaataggagaataattaaatggtcaaaccaaactgggataactttaaagctaaatttagtgagaatcctcaaggtaattttgagtggttttgctacttgttgttctgtcaagaattcaaaatgcccgcaggtatatttagatataagaatcaatctggtatcgaaactaatccaataaccaaagataatgaaattatcggttggcaatctaaattctatgacacaaaattgtcggataacaaagctgatcttatagaaatgattgagaaaagcaaaaaggcttatccaggattaagtaaaatcattttctatactaatcaagagtgggggcaggggagaaagtcccatgaacctgaaggcgataagaacgctgataattatttggaaactgtcggaaatagtaacgatcccaaaataaaaattgaagttgatcagaaagcatatgagtcgggtatcgaaatagtatggagagttgctagttttttgaatcaccgtttgtaatagttgagaatgaaaagattgctaaacatttcttctcccttaatgaaagcatctttgatttattagaagaaaagcgcaagcacacagaaaatgttttatatgaaattcaaaccaatatagagttcaaagacagaagtattgaaattgacagacgacattgcatagaacttctacatgagaatctagttcagaaaaaaattgtcatcgtcagcggagaaggtggggttggaaaaacagcagttatcaaaaaaatttatgaagcagaaaaacaatacactcctttctatgtctttaaggctagcgagtttaaaaaggacagcattaatgagttattcggtgcgcatggcttagacgatttctctaatgctcatcaagacgaattacgtaaagtcatagtcgtagattctgctgaaaagcttttagaactgaccaatatcgatccttttaaagaattcctgactgttttaataaaggataaatggcaggttgttttcacaacccgtaacaattacttggcagatctgaactatgctttcatagatatttataagataactcctggaaacttagtaataaagaaccttgaacgcggcgagctaatagagttatctgataacaatggatttagccttcctcaagatgttcgattattagaactaatcaaaaatccattttatctaagtgaatatttgaggttctataccggtgaaagcatcgattatgtgagcttcaaagaaaagctatggaataagattatcgtcaaaaataaaccttctcgggagcagtgtttcttagcgactgcttttcagcgggctagtgagggccaatttttttgtctccccggcatgtgatactggaattttagatgagttagttaaagacggaattgtcggctatgaagctgctggttacttcattacacatgatatatacgaggaatgggcattagaaaagaaaatttctgtcgattatatccgtaaagcgaacaataacgagttcttcgaaaaaataggagaatcacttcctgttcgccgtagttttcggaattggatatctgaacgattgcttttagatgaccagtccataaagccttttatcgcagaaatagtctgtggagaaggaatatcaaatttttggaaagacgagttatgggtagctgtccttctttccgacaattcaagcatattttttaattactttaaaagatatttacttagtagtgaccagaatctattaaaaagacttactttcttattgaggcttgcttgcaaggacgttgattacgatctgcttaaacagttaggtgtaagtaattcagatctgctttccattaaatatgttcttactaagcctaagggaactggttggcagagtgtgatccaatttatctatgaaaatttagatgaaatagggatcagaaatattaattttatacttcctgtgattcaggagtggaatcaaagaaacaaagtgggtgaaacgactcgattatctagtttgatagctctaaaatattatcaatggactatagatgaggatgtctatttatccggaagggataatgagaaaaatattctgcatacgattcttcatggggcggccatgattaaacctgaaatggaagaggttttagttaaggttcttaaaaataggtggaaagagcatggtaccccatatttcgaccttatgaccttaatccttactgacttagattcatatccggtttgggcatctctcccggaatatgttctacaattggcagatctgttctggtatcggccacttaaagaaacaggcgaacgttatcacagtatggatattgaagatgagttcggtctatttaggtctcatcacgactattatccagaaagtccatatcagactcctatatattggttactacaatcacagttcaaaaaaacaatagactttattcttgattttacgaacaagacaacgatatgttttgcccactcccattttgctaaaaacgaaattgaagaagtagatgtctttattgaagaaggaaagtttataaagcaatatatatgcaatcgtctgtggtgctcataccgaggaacacaggtctctacctacttactttcatcaattcatatggcattggaaaagtttttcttgagaattttaaaatgcagactcgaaagtgttggaaagttggcttcttttcttgttaagaaataccaagtcagcttctatttctgcagtagttacgagtattgtacttgcattccctgagaagacattcaatgtagctaaagtactattccaaacaaaggacttcttccgttttgatatgaatcgaatggttctagacagaacacataaaagttcattaatctccctcagggatggctttggcggtacagattacagaaactctttgcacgaagaagatagaattaaagcttgcgatgatgtgcatagaaatacttatcttgaaaatcttgccttgcattatcaaattttcaggagtgaaaatgtaacggagaaagatgccattgaaaggcaacaagtgctctgggatattttcgacaaatactataatcagcttccagatgaagctcaagaaactgaagccgataagacgtggaggctctgcttggcaagaatggatcggcgaaagatgaaaataactaccaaggagaaagatgaagggattgagatatcattcaatcctgagattgaccctaaactaaagcaatatagtgaggaagcaataaagaaaaactccgagcatatgaagtatgtaacgctgaaactatgggcaagctataaaagagaaaaggatgaacgttataagaattatggaatgtatgaggacaatccgcaaattgctttacaagagaccaaagaaataataaaaaagcttaatgaggaaggggggtgaagatttcagactattaaatggtaatataccagcagacgtttgttctgtattactgttagattattttaatcagttgaataatgaagagagagaatactgtaaagatattgttctagcgtattctaaacttccgttgaaggaaggctataattatcaggtacaagatggaacaacctcggcaatttcagccttacccgtgatttatcataattatccaatggaaagggagactataaaaacaatattacttttgacactgtttaatgaccactctattggaatggcaggtgggcgctactcagtatttcctagtatggtgattcataaattatggctagactattttgatgatatgcagtccctattgtttggttttttgattttaaagccaaaatatgtaatcctttcaagaaaaatcattcatgaaagttatcgtcaagtagactatgacattaaaaaaataaatattaataaggtgtttttaaataactataagcattgcatatcaaatgtcatcgataataaaatatctatagatgatttgggaagtatggataaagttgatctacatattttgaacacagctttccaattaattccagttgatactgttaatattgaacataagaaattggtttccttaattgttaaaagattttctacaagcctattgtcaagtgttcgagaagatagagttgattacgctcttcggcagtctttcttggaaagatttgcctactttacgcttcatgcgcccgtgagcgatattcccgattatataaaacctttttcttgatggtttcaacggttcagagcctatttcagagttatttaaaaaa

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.		Cloned Sequence
		tttattctcgtcgaagatagattaaatacttacgccaaattttggaaggtttgggatttgttttttgataaagtggttactttgtgcaaggatggagataggtattggtatgtagataaaattataaaaagttaccttttgctgaatctccatggaaagaaaactctaatggttggcacacatttaaagatagcaatagtcaattcttttgcgatgtatctaggactatgggccattgcccttcaactttatattctcttgccaaatctttgaataacattgccagttgctatcttaatcaaggtataacttggctttcagaaatattgtcggttaataaaaagctatgggaaaagaaattggaaaatgatactgtttattatttggaatgtttggttaggcggtatattaacaatgagcgtgagcgaattagacgaaccaaacagttgaaacaagaggtcttagtaatattggatttttggtagagaaaggatcggttgttggttatatgtcacgggaaaatattctgtgatgtagttgaaaataataattttaatgagagcttttccaatttaggctccagggattggagccttttattatcg (SEQ ID NO: 287)
40	16	actagctaagcaataagggcgatcggctctcccatagatcgaggccgaatgatgttagcaatgttcactcttggctggaatctgccagaaatcgaggtcatatggtctgctttgagtgaggagcgcaaatggataaagccctcatgagttctttttcaatgacctaacttttgagaggcactgggttagatcatgtttcatgtttgcaatacaatatatatttaaacttaggtttataacttaaatgttagttcctgatctaaaccagattattaatcactcctagagtgaaatgagttaagccaagagttgataaaattaacagtttttttttacaatatctggatgtttgctagcgaacaggcatctaaaataactatgctgagctaaacttacaattcaaattgtaccgaggataaaatgcaagtacaacatcatactgaaccaaacttgaagaatgagattgtggctttatttaaggcttctcaattgataccttttttggcagtggatttactagagatattagagcaaaaaatggtaaagttcctgatgctattaaatttacggagttgattaggaatatagcggcagaaaaagaagggttaacacaaacagaaatagatgaaattctaagaatcagccagcttaaaaaagcgtttggacttctaaatatggaggaatatatacccaaacgaaaatcgaaggcattattaggtaacattttttcagagtgtaaactctctgatcacgaaaagacaaaaataataaatttagattggcctcatatttttcacgtttaatattgacgatgctatagaaaacgttaataggaaatacaaaattctgcatccaaatcgagcagttcagagagaatttatatctgctaataagtgtctattcaaaattcatggcgatattactgaatttattaaatacgaagatcaaaatctgatatttacttggcgtgaatatgcacacagtatagaagaaaataaatccatgctatcctttttatctgaggaagccaaaaactcagctttccttttcataggttgcagtcttgatggagagcttgatttaatgcatttatcaagaagcacaccatttaagaaatcaatttatttgaagaaaggatatttaaatttagaagaaaaaatagctctttcggagtacggcatcgaaaaagtaattacctttgacacttacgatcagatatatcaatggttaaataacacacttcagaatgttgagcgaaaatcccccacaagaagtttcgaactcgatgactccaagttaatgaaagaagaggctataaatttattcgctaatggaggccctgtaactaaaatagtggataataaaagaatcctgcgaaattctataactttttctcaacgagatgtctgtgatgatgcaattaaagcactacgtaatcatgactatatcctaattacaggtcgacgtttcagcggaaaatctgtacttttatttcaaattattgaggcaaaaaaagaatataatgcctcttattactcttcgactgacacattcgatccttccattaaaaactcattgataaaattcgagaatcatatattcgttttcgactctaatttctttaatgcacaaagcattgatgaaattttaaccacaagggtgcatcctagtaacaaagttgttttatgctcgagttttggtgacgcagagttatatagattcaagttaaaggataaaaagatattacataccgaaattcagattaaaaataacttgattaatgaagaaggtaactatctcaatgataagctttcttttgaggggctaccactttataaatcttcagaaacgttgttgaattttgcttatcgatactatagcgagtataaaaatagactaagtggttctaatttatttaataagcaatttgatgaagattcaatgtttgttttgattttaattgcagcttttaataaagccacatatggtcatatcaacagtcacaataaatatttgatattcagaattttatttcgcaaaatgatagattatttgaattggagtcaactaacacagatccaagtggagttataatctgcaattcaccatcctggcttttaagagttatcagtgagtatattgataagaatcctgcatcttataaaacagtatctgatttaataatatctcttgcgtcaaaaggatttcttgcagcatcaaggaaccttataagctttgataaactaaatgaacttgggaatggaaaaaatgtccataaatttatcaggggtatatataaggaaattgcacatacctatcgtgaagatatgcactactggttacaaagggctaagtcagaattaatatcggcacacacaattgatgacctcgtcgaaggaatgagttatgcaagcaaagtaagactcgatagtgccgagtttaaaaatcaaacttattacagtgccacattagtattagcgcagttgtctgcaagggctctatctataaataatgataaaatatatgcgctgagcttctttgaaagtagcctagaatccatccggaattataataataactcaaggcacataaacaaaatgatggataaaaatgatggtggctttagatatgcaatacaatatcttaaggataatccattaatagaactccttcctcgtaaggacgaagttaatgaattaattaacttctatgagagtcgtaagaaataatcatccttaaattaataaatggcaagtaactcattcccttgtcatttattaaactcttaagagccttatcccgaaaagtattaatctgagctaataagattgtttttcagctatgtcattattttattgccaatatatttacacttaagcattgacaggtagcggatagttatttttggcttgtaaataagccttttaataatagaactgtaagacaatcgctctgatttttttgaaatttatctcaatgttaaattcttccgcttttggcacaaacgggctagagcagacagatttaatgagataagggtatagatgaattctccatacccttgaacgattacttcccagttgatttgcttggtttcagtcctggggtattaccgggtgtatccttattatcacgtctgcgttgatcgggttttcctgttgattttgcaattggttttggaccaggtttaagccccataatcgtactccttagccatgtcagaggttattcctcagtgtggatataaggggagcggtaagaattatcaagcttggatgggcggtgaaaaatgactacttgactattatgtgagcaatgtcagcttttgacatttagaggccagcccattactgaagtaagccaaaaatgagtcgcgatgagccctcaacaatgagggccacctcggagattg (SEQ ID NO: 288)
41	17	tattttgcgtagctagaacgcaatcaaatctagcagtccgctttgttcggagttcggacattatgagttggcaagtaaagtagcttgctaggaagccggatttgcacggtcggtataataagatgtaaccccttgccttcatttactcgaatgaacgtgcacattggataggaggaaaaggaatgcaattcattaccaacggccctgatattcctgatgagcttttgcaggcgcacgaggaagggcgcgttgtgttcttctgtggagcaggcatttcctaccctgctggtttacctggtttcaaagggttggtagaactaatttaccagaggaacggaacaacactttcagaaattgagcgtgaggttttcgagcgtgggcaatttgacggcacattagatttgctggaacggcgcttaccagggcagcgtatagccgtccgacgcgcgttggaaaaagcccttaagccaaagctccgtcgtaggggcgctattgatactcaggcggcgctgttacgtttagcccgta

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.		Cloned Sequence
		gccgcgagggtgcccttcgattggtcactaccaactttgaccgtctctttcatgtggcagctaaacgtacaggccaggcttttcaggcctatgtagcgccgatgctgccaattccaaaaaacagccgctgggatggacttgtatacctgcatgggctgttaccggaaaaggcggatgatactgccctgaatcgtctggttgttaccagcggtgactttggcttggcttatctcactgagcgttgggcagctcgctttgtgagtgagttatttcgtaactatgtggtctgcttcgttggctacagcatcaacgacccggtactgcgctacatgatggatgcgcttgcagcagatcggaggctcggtgaagtcacaccacaagtatgggcactggggagtgtgagccggggcaggagcaccggaaagccatcgagtgggaggccaaaggggtcactcctatcctttacaccgtaccggcgggctccactgatcattcagtgctgcatcaaacgttgcacgcttgggcagatacttatcgagatggtatacagggcaaaaaggctatagtcgtcaaacatgctctggcccgcccgcaggacagcactcgtcaggacgatttcgttggtcggatgttgtgggccttgtcagataaatcaggtttaccagcaaaacgctttgcggaactcaatcctgcaccgccgctggattggttattgaaagctttctcggacgaacgatttaaatacagcgatctgccacgcttttgtgtatctccgcatgtcgaaattgacccgaaactccgattcagtctggttcagcgtcctgcgccctatgagctggccccgcagatgtcgctggtttctggatgtgtcagtgctagcaaatgggatgacgtaatgtcccatatagcccgttggctagttcgttatctgggcgaccctaggttgatcatatggattgctgaacgcggcggacaaatacacgaccgttggatgtttctgattgagagcgaactagatcgcttagcagcactgatgcgggagcgtaagacttctgagttagatgaaattctcttgcattcccccctggctattcctggtccacctatgtctactttatggcggcttctgcttagtggtcgtgtgaaatcgccattgcagaacctggatttgtatcgttggcaaaaccgcttaaagaatgaaggcttgacgactacattgcgcttggagttacgcgggttgctttctcccaaggttatgttgaggcggccgtttcgctatagtgaagacgattcgagcagcactgatgaacccttgcgaatcaagcaattggtggattgggagctggtgctgactgctgattacgtacgttcaaccctgttcgaccttgctgacgagtcatggaaatcgtccttgccatacctgttggaagattttcagcagttgttgcgtgatgcactggacttgttgcgggagttgggagagtccgacgatcgtcacgaccgctcgcattgggatttgccgtccatcactccgcactggcagaaccgggggttccgcgattgggtgagcctgattgaattacttcgggattcatggttagccgttcgagccaaagacagcgatcaggcctcgcgcattgctcagaattggtttgagttgccatatcccaccttcaaacgtctggcactgtttgccgcaagccaagacaactgcataccacctgagcggtgggttaattggttgttagaggacggttcatggtggttgtgggccacggatactcggcgagaggtattcagactgtttgttttgcagggacgacatctgacaggaattgcacaagagcgtctggaaactgctatcttggcagggcctccgcgcgagatgtacgaggataatttggaagcagacaggtggcattatttggtggctcattccgtctggttgtgtctagcgaagctcaggggagcgggccttgttttgggagagtctgcggctacacgtttgacggaaatatccacagcatacccaaaatggcaactggcaaccaacgagcgtgatgaattctctcactggatgagcggaaccggtgatccaggcttcgaggagagtatagatgtcgacattgcgccccgtaagtggcaggaattagtgcaatggctcgcaaagcctatgccagaaagactgcctttctatgaggacacttggagtgatgtttgccgtacgcgctttttttcacagtctgtatgcgttacgtaaactatcacaagatgatgtgtggcctgttggtcggtggcgtgaagctctgcagacttgggctgaaccagggatgattttgcgttcgtggcggtacgccgcaccgttggtgcttgacatgcctgacgcagtacttcaggagatttcccacgctgtcacttggtggatggaggaggcttcgaagaccatcctctgccacgaggagattctactggccctttgtcgtcgggttctgatgatagaaacaagcccagagtctagcaccattcgaaacggaattgagacctatgatcctgtttctacggcgatcaatcatcccattgggcatgtcacgcaatcactgatcaccctatggttcaaacagaacccgaatgacaatgatttgcttcctgttgaattgaaaacacttttcaccaaattgtgtaatgtacagatagagctattccgccatggtcgggtgttgctggggtcgcggctgatcgcatttttttcgcgtagatcgaccttggaccgaacagtatctattgcccttgtttgcttggagtaatcccgtcgaagcaaaagctgtgtgggaaggcttcctctggtcgccacgcctgtatgaaccgttgctgatagctttcaagtcagatttttggagagcgccaatcactattctgatcttggcgagcaccggcagcaattcgctattttcctgacttatgcagctctgggccctaccgagggatataccgtggaggagttccgaacggcaattagtgctcttccacaagaaggtctggaggtagccgcgcaggcgttataccaggcacttgaaggtgcgggcgatcagcgcgaggagtattggaaaaatcgtgtccagccattttggcaacaggtttggccaaagtcccgcaacttggccaccccacgcatatccgaatcgttgactcgtatggtgattgctgcccgaggtgaatttccggcggctttggcagtggtgcaggactggctgcaaccgctcgaacaccttagctacgacgttcgccttttgctagaatcagatatttgcagccgatatcctgcggacgctctatccctgctgaatgccgtgattgccgaacaacactgggggcctcgagagttggggcaatgcttgcttcaaattgttcaagctgctccacaactggagcaagatgttcgttatcagcgattaaatgaatattctcgaaggcgcagcgtgtgaaagtgacaggcgttggacagtgcgaactgtggagcctaacaaggtaaagacactctaactgataatgctgcgccgctcgtgcaatgcaatacagtttttatctagcggtgaattatggtgttaaaagttagcccctgacacagggtgggtagttggctctgtgtcattgatgggtattagttctgatatgagctaataccca (SEQ ID NO: 289)
42	18	gtaagacaagggttgagcaggctactaatcgttacacaggctaacaaaggcatattaagacgatttgtagcgctgtaaccttgaaaattatgtacaagcgccccgcattacgtcgttttaaaggccatcggattcaggcccgacgcggcttcacgcgattataaccgtgaaaaatccccccgcatagaacctgaattatccccgccgccgcgcagaactgacagcgcttcagaaccgttaaccctctcagaaatcccgcttttttactgtaaaaaaccatgcataaggtgcatggttttgcatgcgtttcaccgacactgaatcccccgccagcgccagcagtagcgtgccctgaggccgttaatgcacccgtattaaaagcgccctgttaagcgagcaggcggggcggggcgagcattgcgcgtcggtgttaccaattctatatggacattgagcaattcaaatataataaaggttgggtatatttcgtcctcaacgatgtcaaaaactgcaaaagcgtattataattcagatcattttcagaccacctattttaatcatgcatgcaaaatggaatatgtgatgacaaataaaaacaaaatcaaaccattattaaataatatatccgctcgcctttgggatggtcgtgcagctatattgataggagctgggttcagtcggaatgcaaagccattaacaagcaaggcaagaaagtttccaatgtggaacgacttaggtgacatttttatgaaagtgtttactgcaaaaaaaacgacaatagatattcaaatgtattgaagctaggagatgaagttcaggctgcatttggtagagcgacacttgataaattaatcatggatcatgttccagataaagaatatgaaccatccaaattacatgtttcccttctttccttgccgtg

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.		Cloned Sequence
		gattgatgtttttacgactaattatgatacattacttgagcgagcaagtgttaatgtcgactccag aaaatatgacattgtccttaataaaaatgatttaatgaatgctgaaagaccaagaattataaaact gcatggtagcttcccatcagaaaggcccttcatagttacggaggaagattacagaaagtatccttt agaaaattctccttttgtgaataccgttcaacaatcattgattgagaatactctatgtctgatagg attttcgggtgacgatcctaacttcttaaattggattggttggataagagataatcttggcacaga aaattcacccaaaatatacttgatcggtctttttttcatttaatgaagcacaacgtaagcttttaga aaaaagaaatatttccattgttgatttaagttttctaggtgattttggcaaggatcattatctagc acaccaacgctttatccaattcttatacgaatcaaaaaatcgagacaacctaatagagtggccaat agaaaccaattatgacagaattgtttttaatgatggcattgaattaaaaactgagaaaattaaaaa gtgtatcttagaatgggctcagtcaagacaatcatacccgaactggcttatttgccggaatcaaa cagaagtaatttatggcaaaacactatagattggttatctgttgctaattatgatgtcgcttggga tggttctgatgatcttgattttggatatgaaattacatggcgactaaataaagctttgctaccaat tttcaatgatacatcagaattcttatttaagttgattgaaaaatatgagatcaattacgtttcggg gataaataataaaatcattgactttgatgaaaaatactctcatataaccctcagtttaatgagatt ctgtcgacaagaaaaccttattgataaatggaagaatctaaacgatttattaattcaaaatcttga tcgattaacaccagaggtaaaatctgattattattatgaaaatatattattttcatacttcaattt aaacttcgatgaagccagaaacaaactctccaactgggaaacgaataaactcctcccccatcatga aataaaaagagcaggattacttgccgaatttggaatgcttgatgaagcaatcaatcttcttgaaga aactttatctacgattcgaagaaacagtttgctttcatctagaaacattgactattccagtgaatc tcaagaagcatatggaatctatattttgcgaatgtttaaacggagtttgcgtttagatagcaaaga tgacgattattcatctgagtataactcgcggttggctacattatcacaatatcgcagcgatcctga aaacgaaataaaatacctagaaattaaactagagtcactaccaggtaccttcaagaataccaatga cacggatttcgatcttaacaaaagaacggtgaccacttatttaggaggaagcccaacagaagtgag gtcattagatgcttttagtttctttctactggcagaggaacttggcctccctttccacataccagg aatgaacatttttagtggaatagttgagaatgcagctcgacatatttatcaatactctccagagtg ggctatttttttcaatatttagaacatttaacaaggataaggccaagagtctattcaatcgaaatag aatttcgtctcttgagcgaaaaaaggttgaagatttatttgatggatactacaaaaaatatgagca aattatcacaaaaaaaatagaagatagattaaacgataaacttgagatagaaatttctacgctatc aatcattcctgaaattctttcccggctagttacaaaagtatcatttaataaaaagaaagacattat tcaccttttgcttaaactgtttaactcggataattttcatcaatacatggagactaaagatctatt aaagcgcactatttccaatttgagcgacttacaaaagatctcactaatagatattttcattgattt cccctccgcgcctcccaatacccaattacatatgggtcaaagatacaacttccttactccatttga atgtctattaggggttacaataacccccccaaaagaaaactctaaaaaaatcgcatctgcaaaatt aaaaaaagatataaacgatttaaaaagtgataatttagacttgaggaaagctgtatcacaaaagct cataacattatataacctagaaatgcttaacaaatctgacacgactaaacttataaaaaaccttttg gtcaaagcgtgataactttggattcccaataggcagtggttactataaatttttctttataaacaa ccttaacccagataatgaaaatatagccgacaaattcatttctataattaaaacatacaaatttcc tgtgcaagaaggaaaaagagttagtattacaggtgggttagatgagtattgtactgaactcaatgg agcgctacaccatataagtcttccagagaaaaccctatctgaaataatttcaaaaatacatgactg gtatgtcaaggatcgggcctggcttgaaaaaagagatgatttagccaaggagttcactcttagatt cagaaatatcacaaatatcataacgacaattttagaacaccataaggacaaattacatgctgaatc tataaatgaaatatcaagcctactagataaaatgaaagaagacaagatacctgtaaactcagcagt aacaatgctttgtctgaaaaataaaagcacttacctcgagagaataaaagatatagagaatggact atatagctttaataaagatgatgttattgaagctatcaactcaacttatgtctttattagaaacaa tgaatttccactaaccatcattcaagctatcagcgataaaatcgcatgggatagaaaccctcgcct tcctgattgctacaatttaattgcatatataattaactcgtgtgaatttactcttccagattattt aatagagaaaatccttcgagggctggcatatcaaataaacattgatgatagagattttgttgataa caatgaatatttgaatcaccttgagaaaaaacttagtgcaacaaagctggctgcttctatgtttag aaaaaatgaaacactaggtattgaccaaccttctatcattcaagagtggaaaaacatgtgcaactc tagaaatgagttcgatgaaattaggaatgaatggaacaacaatatataaataaaggaagaacaccc aatttatattgggtgttctgttcacgaaacccttttaccataatcgaatggcaatataaattgaga ttgaaatttattctcatctaattaatcagcccaccattg (SEQ ID NO: 290)
43 44	19	tagctattgtgactatgctaaccatatgaatctattgtgtgattatgagtaatgactttttctaat atttgattttttaatgtagtaacttagctaattttaaaatttgtaaaaggatgtttatgtcgattta tcaaggtggtaacaagttaaatgaggatgattttcgttctcacgtttattccttgtgtcaattaga taatgttggcgttctgttaggtgctggtgcttctgtcggttgtggtgggaaaacgatgaaagatgt atggaaatcgtttaagcaaaactaccctgagcttttgggagcacttattgataaatatcttctggt ttcgcaaattgattctgataacaatttggtcaatgttgaacttttgatagatgaagcaactaaatt tctttctgtagctaaaactagacgatgtgaagatgaagaggaggaattcaggaaaatattaagttc attatataaagaggttacgaaggctgcattattaacaggagaacagtttagagagaaaaatcaggg taaaaaagatgcgtttaaatatcacaaagagttaatttcaaaattaatttcaaatagacagcccgg tcagtcggctccggcaatttttacaacaaattatgatttggccttagagtgggctgcagaagattt aggaatacagttgtttaatggtttttctgggctacatacacggcagttttatccccagaattttga tttggctttcagaaatgtaaatgcgaagggcgaagcaagattcggacattatcatgcgtatctcta taaattacatggctcacttacgtggtatcaaaatgatagcttgactgttaacgaagttagtgcatc tcaagcatatgatgaatatattaatgacataatcaataaagatgacttttatcgcggtcaacattt gatttatccaggggcgaataaatatagccatacaatcggcttcgtttatggagagatgtttagacg ttttggggagtttatttcgaaacctcaaacagcgttgttcataaatgggtttggtttcggtgatta tcatataaatagaataatattaggcgcgttactgaatccatctttccatgttgttatatattatcc tgaattgaaagaagcaattaccaaagtaagtaagggtggtggttcggaagctgagaaagctattgt tactttaaaaaatatggctttcaatcaagtaactgtagttgggggaggaagcaaggcatattttaa

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.		Cloned Sequence
		tagtttcgtagaacatctaccataccctgtgctctttccacgagataatattgttgatgagttggttgaagcaattgctaatctttctaaaggagaaggtaatgtccctttttaaacttactgaaatctcggctattggatacgttgtaggattagaaggggaaagaattaggataaacctgcatgaggggttgcaaggcagattagcatcgcatagaaagggggtgagctcagtaacgcaaccaggagatcttattgggttcgatgcaggtaatatattagttgtcgcaagagtgacagatatggcatttgttgaagcggataaagcgcataaggcaaatgtaggcacatctgatttagctgatatacctctaagacaaattatcgcctatgcaattggctttgtgaaaagggagttaaatggttatgtttttatatcagaagattggcgcttacctgcattgggttcttctgctgttcctttgacttcagattttttgaacatcatttatagtattgataaagaagaactcccaaaagcggttgaattaggtgtggattctagaactaaaaccgttaagatatttgcaagtgttgataaattattgtcgcgacacttagccgttcttggtagtacaggatatggtaaatcaaatttcaatgctttgttaacgaggaaggtttctgaaaaataccctaactcaagaatagttatttttgacataaatggtgaatacgcgcaagcttttacaggtattccaaatgtaaagcacactattctaggggaatccccaaatgttgatagtttggaaaaaaagcagcaaaagggtgagctatatagtgaagagtattattgttataaaaagataccatatcaggcattaggttttgctgggttaattaaattattaagaccaagtgataaaacacaattgcccgcattaagaaatgcattaagtgcaattaatcggactcattttaaaagccgtaatatttacttggaaaaagatgatggtgaaacttttcttttgtatgatgattgtcgtgacacaaatcaaagtaaattggctgagtggttggatttattaaggcgtagacgtcttaaaagaacgaatgtatggccaccgtttaaaagtttagcgactttggttgctgaatttggatgtgtagctgctgaccgttctaatggaagtaaacgtgacgcgtttggttttagtaacgtgttgccattggtaaaaatcatacaacaacttgcagaggatataagatttaaatctattgttaatttaaatggagggggtgagctagcagatggtggaacgcattgggataaagctatgagtgatgaagttgattacttctttggtaaggaaaaaggacaagaaaatgattggaatgttcatatagttaatatgaaaaatttggcacaagatcatgctccaatgttacttagtgcattgttggagatgtttgctgagatactatttagacgtgggcaggaacgttcgtatcctacggtacttttgttggaagaagcgcatcattacctgcgtgacccttatgctgaaattgactcacagattaaagcatatgaacgacttgctaaagaaggtaggaaattcaaatgctctttaattgtcagtactcagcgaccctcagagctttctcctactgttttggcaatgtgttcaaactggttttcgttacgtttgactaatgaaagagatttacaggctctcagatatgcaatggaaagcggtaatgaacaaatcttaaaacaaatatcaggtttaccaagaggtgatgctgttgcatttggttctgcatttaatttgcctgtaagaatttcaattaatcaagcaaggccagggccaaaatcttcagatgctgtttttttctgaagaatggctaattgtacagaattacgttgttaattacctgatgtacatggctagtgcaagttggtagcgcatgtctatatgcatttatttgcatgtgttttattgagtgagcgcacaagcttgatgacccgacaggtatgtatttagactgaa (SEQ ID NO: 291)
45 46	20	gtgcgccttatgtgattacaacgaaaataaaaaccatcacaccccatttaatatcagggaaccggacataaccccatgagtgcaatagaaaatttcgacgcccatacgcccatgatgcagcagtattgaaaaatataacatatccaactgattgtattgaaaatttaaaatagccatataacaaaaggttacacataagctactttttggggtttcaggcaagaaactaaaaattattaacgccatcaaattattcacatcttaataattagcattgaaatttaatgtttttggttctttgtacatgtcaatggcttgtctttgtggcagaatcataaagctatgcaatcattgcattgttattaacacagcatatttttatatacttttaacaccttacctcaaaaaggataacaaagtggacagaagtgcggttgatacaattcgtgggtattgttatcaggttgataaaacgattattgagattttttcgttaccacaaatggatgactcgattgatatagagtgcattgaagatgttgatgtctacaacgatgggcatttaactgcgatacaatgcaaatattatgaaagtaccgattataaccactccgttatatcaaagcccataagattaatgttgtcacactttaaggacaataaagaaaaaggggctaattattatctttatgggcattataaatccggtcaagaaaagttaacactcccattaaaagttgactttttcaaatctaatttcctcacctacaccgaaaaaaaaatcaaacatgaataccatattgaaaatgggcttaccgaagaggatctacaagcctttttggatcggttagttataaatatcaatgcaaaatcatttgatgatcaaaaaaaagaaactatacaaataataaaaaaccatttccaatgtgaagattatgaggcagagcattatctttattctaatgctttcagaaaaacatatgatatctcttgtaataaaaaagatagaaggataaaaaaatctgattttgttgaaagtatcaacaaatcaaaagtcttatttaacatatggttttatcaatatgaaggaagaaaagaatatttaagaaaattaaaagaatctttcatacgcagaagtgtaaacacctcaccttatgctcgttttttcatcttagaatttcaagacaaaactgatataaaaacagttaaagactgtatatataaaatacaatcaaattggtctaatttatctaaaagaacagatcgaccatattctccttttttactttttcatggcaccagcgatgccaatttatacgaattaaagaatcaattattcaatgaagatctaattttcactgatgggtacccttttaaaggaagtgtatttacccccaagatgttaatcgaaggtttttcaaataaagaaatccacttccaatttatcaacgacatagatgatttcaatgaaacactgaacagtattaatataagaaaagaagtttaccagttttatacggaaaactgccttgatatcccatcccaactaccccaggtaaacatacaagttaaagactttgccgacataaaggagatagtgtaatgagcaggaataatgatattaatgcagaagtagtatcggtatcgccaaataaattaaaaatttccgtagacgatcttgaagaatttaagatagcagaagaaaaattaggtgtaggatcttatttaagggtttcagataatcaagatgttgctcttctggcgatcatagataatttttctattgaagttaaagaaagccaaaagcagaaatacatgatagaagcaagtccaataggtcttgttaaaaatggaaaattctatcgcggtggagattcacttgcacttcctcctaaaaaagtggaaccagcgaaattagacgaaataatatccatatactcagatagtatagatataaatgaccgttttactttttcaagcttatcgcttaataccaaagtatccgtacctgtgaatgggaatagatttttcaataaacatatcgctatcgtaggttcaacgggttcaggtaaatcccacactgttgcaaaaatacttcaaaaagccgtagatgaaaagcaagaaggttataagggattaaacaattctcatataattatttttgatatacattctgaatatgaaaatgcattccctaattcaaatgtattaaatgtagatacattaacccttccatattggctattaaatggtgacgagttagaagagctttttcttgacacggaagcaaatgatcacaatcaaagaaatgtgttccgtcaggcaataacattaaataaaaagatacattttcaaggagatccagccacaaaggaaataataagctttcactcgccatattatttcgacattaatgaagtcatcaattatattaacaatagaaataatgaaagaaaaaataaagataatgaacatatttggtcagatgaggaaggaaatttcaagtttgacaatgaaaatgctcataggttattcaaagagaatgtaactcctgatggaagttcagccggtgct

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.		Cloned Sequence
		ttaaatggaaaacttctcaattttgttgatcgattacaaagtaaaatatttgataagagattagat tttattctgggtgaaggtagcaaatccgtaacatttaaagaaacattagaaactttaataagctat ggaaaagataaatcaaacataacaatacttgatgtaagcggtgttccttttgaagtacttagcata tgtgtatcattgatatctcgattaatttttgaatttggctatcattcaaaaaaaataaaaagaaaa tctaatgaaaaccaagatatcccaatattaattgtttacgaagaagcacataaatatgctcccaaa agtgatctgagcaaatacaggacatccaaagaagcaattgagaggattgcaaaagagggtagaaaa tacggagtaacccttctccttgcaagtcagagaccttctgaaatttcagaaacaatattttctcag tgtaatacttttatctcaatgcgattaactaacccagacgatcaaaattatgttaagcgattactc ccggatacagtaggtgatattacaaacctcctaccatcgctcaaagaaggtgaggccttaatcatg ggggattcaatatcaataccttcgattgtaaaaatagaaaaatgtacaatacccccatcgtcaatt gacatcaaatatcttgatgaatggagaaaagaatgggtagattcggagtttgataagataattgaa caatggagtaaaagttaatttcagaagtggattcactcttgctcaagagtgaatccactaatatca tatcctaatgatatagtttaataaaatctattctggaatcattaggctgagag (SEQ ID NO: 292)
47	21	accgtgctggcatgtttttacggagtgacgctttcattaacctgtacacgaacttctattccggca tcatgacaggcctgcagccactgcgccacttccagcggatcgccctcccggcgtaccactctgcct tctttattccataactgcagacaggtgctgccgtcgagacgcaccacaaaatccccacggcaggcc tgataggggtttgagggccaaccgtacgaaaacgtacggtaagaggaaaattatcgtcttaaaaat cgatttatgctatcacagtcgtctcttcaggtaagtacggttgcctttgcctgctttcttctcgtc tggttaagttaagaaattcagagatccatgcttgagataaaagcggaataaaaccagtaaaatgta actaaaacaacaacggaattgtatcaatgataatgtccacaccgtggctgacaccgatcgttgccg atagtgatcatgctgaggcaaatgcagtgagctatgaagcactgactccgacagaactcgactcag ataaagcaggctgttatatcagcgcgcttaattatgcttatgaacatccggatatccggaatattg ctgttaccgggccgtatggggcagggaaaagctcagtattaaaaacatggtgcaaagctcacaatg ggacactgcgggtgttaaccgtttctcttgctgattttgatatgcagagacatgtggatgaaagta atggggacagcagtagtgacgaagggacgaaaaatactggtagtgttgaaaaatctattgaataca gtattctgcaacaaatactctacaaaaataaaaagcatgagcttccctgttcccgcattgaccgta tatcagatgtgactgcgggacaaatattgcggtctgcgtcttttctgacaggaaccattttactga gtggagctgctttatttttccttgcgccggattacgttacaacaaagctatctttgccgggagcat tcgcccgttaccttcttgaatgcccgtttggggtgcgtgtgtccggtgcagtggcatctgtgatgg gatcgttatgcctgcttttgaaccagttacatcgtatcggtatatttgacaggaaagtaagtcttg ataaagtggaccttctgaaaggcgctgttacaacccgggcatcatcaccttctttacttaatgtct atattgatgaaattgtctatttttttgattcgactaaatatgatgtagtgatattcgaagatcttg accgttttaacaatggccggattttcgtgaaattgcgggaaatcaatcaaattattaataactgcc tttctgacagaaaacctgtaaaatttatttatgctgtcagagatggtattttcaactcagcagagt caagaacgaaattctttgattttgttatgcctgttattccagtgatggataaccagaatgcttatg agcattttgttaaaaaattcaaagaagaagagataaataataacttaagcgaatgtatttctcgta ttgcgacatttattcccaatatgcgtgtaatgcataatattacaaatgagtttcgactctatcaga atttagtcaatagtcgggaaaatctggccaaactacttgccatgatagcatataaaaatctctgtg cggaagattatcatggtatagatagtaaaaaaggtgttctttatcattttattcaaagctacttag accatgaaattcagaatgaattattacattctgcaaataacgaacttgaggatatggcacagtcac ttgtagcgataacaaatgaaaaactcgcaaaccgggaaaatctgcgcgaagaactgctcatgcctt accttagtaaaaattatagcggcgcgcttgtttttttatacagaaggaaggcaaataagtcttgatg atttgatacaagatgaagatgaatttctcatgcttttagataaggaaaatattcaggtcgttaccc cctataacagacaaaattttctcatgataaatcagcgggatacagaaaaactgaagcagcagtatg aaaaacgatgccatttaattgaaactaaatctgttgataatataaccagagtgaaaaataatattt ccagtctggagtcattgaggaccgaaattctttccggaactgtagctgatatagcagaaaagatga caaatgaaggctttgttgcctggataaagaagaaagaggatacaggtgtcctgacgattcagtcgg aacatgaacagattgatttatatttttctgttatcaagtggttatttatcaacagattacatgt cctatcgctcaatcttcattcccggagggctgagtgagacagataatttatttcttaaggatgtta tgtctggtaaaggtccggaaaaaacattctcattccatcttgataacgttaataatattgttgaac gactcaaaaagctgggggttctgcagcgtgacaatgctcaacatcctgctgttatcagatggctga ttgataatgaccctgataccctgaaaaacaatataatggcattactgagtcagacgggtagccagc gtgtggttagtttgctgatgttgatgcagaacgatttcacaacgtatgttcgcctgcgttacctgg agatttttatgtcagatgaacatatactgaacagattgctggcacatttatgtgcgtcagaagaac gcacacccgagcaaaagtttttttgttcaggaaatagcggcacacctgttatgcctgactgaaaaat caaatatctggcaatcggttgagattaataaacgtatcggtgagcttatagattcctccccaattc ttattactgctgtgccaaaaggatatggtgatgcgtttttttgaagtgttgaaagataatacacttt cagtttcatatattccaggtgatgtgggagacgagaagtgttctgttatcaggaaaattgcgggtg caggattattcaaatattccgtcagtaatcttaaaaatgtttatctttgcctgacgcaagacaaga atgaagaaagaatgtcattctctctttatccgtttcattgtctcgagtccctggctatttctgaat taacagaaattctgtggactaacatagaagattttattttatcggtatttattgaatcggaagaga ttgatcgtattcctgaattgctgaattcttctgaagtctcaatgactgttgttgaacagattatag ccaaaatggattttttgtataaataatctggatgatattattaatcgttcagagtgtgcggacaata atgcttcagggagaaatatctatagcatgctgttgcagcatgacaggatttttccatcctttgata atattattcatttattgcatgatacatcaattaatacttccggtgaacttgttcagtgggtaaatg agaaacactttgaatttgaaccatctgatatagtcataaatgatacaggaatatttaataattta tttctgaattaatttgctcgccagtcatttcagaagaagctttactgaaagtactgagtaatttaa acgttgttattatcgatgtgcctgaaaacattccattgcgaaatgctgaactgttatgttcagaga aaaaactggcaccgacagttaatgtctttacggtgttgtttaatgctctcagtgaaaatgttgatg atattaacaggatgaatactctgcttggtaaccttattgcccagcgtcctgagattattacccagg

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.		Cloned Sequence
		agccagaagatatttttatatcgagggtgactttgatgaagaactggcaagcgaactttttcgtc acaagctaatcggtatgaatataaaagttgccgctttacgctggttgcgtgataacaaaccgggaa ttcttgataagagctacctgctgtcattagatattctggcagaactgagtccctggatgggtgacg atgatctgcgcctgacactgcttaaacgttgtctggttgccggggatgctggcaaagacgcgcttt gcgtggtgctgaacagttttgctgatgagagctatcatggactgttaccacatgacaggttcagga aaatccctcactccgtggatttgtgggaagtggccgaattaatcagcaatcttggatttattcagc cgccaaaaatggggtcagggcgtgatgaacacaaaattgttattactcccgtacgctatgtccgtg atgttgagttttatgactgagcatcattgatacggtgttttaattgccttaaatacaaaaataaaa acagattaatgcttaatgtgcattaatctgttttagttatcaatggctgttaattattgttaattt tacattaatctttctttttcttcaggaagatccgaaaactcctggtcacggatcttcct (SEQ ID NO: 293)
48 49 50 51	22	gaaattatttggaatggatgatggcgcttgattactggaacaggtctatgacatgaaggttatgat ttgttcactgctatgaggttaacactttaacaatttcccttactattcttgtactaattccttcca aatacttctgcttgagattaggatttatcctcttgtagtgttatttacaataaagattgtgatgct gatttaacccaacgtgttgtcagttgccttgctgaactaagttcagtatctagaaattagctcttg atacatgagcgaatcagcgaaaattttcatcccgaccaattaatgaccgtaatggataggatgttg ctgctatttggcttccatgagggaacatatgtttttaaacgatcaagaaacgtccactgacctgct gtactacaccgctatcgccagcacagtggttaggcttgttgatgaaacgtcagatgcacccattac gattggtgtgcatggtgattgggggcgggaaaatcaagcgtactaaaaatgcttgaggctgcctg cgagaaaaaggataaaacgcactgtatctggtttaacggatggacgtttgagggattcgaagatgc taaaactgtaatcatcgaaaccatcgtcgaggatcttgttgcctcgcgcccgatgagcaccaaggt ggcagaagcagcaaaaaaggttcttcgtcgaattgactggttgaaaatggccaagaaagcgggggg actggcgtttaccgcatttactggcatacccacatttgatcagattaaggggatgtacgaactggc atccgactttctaagtgctccgcaggacaagctttctgctgcagatttcaaagcgtttgctgaaaa agcaggaggcttcatcaaagaggccgatactgatagtaatacgctacccaaacatattcatgcttt ccgtgaggagttcagggcgctgcttgatgctgctgaaattgaaaagctagtggtgatcgttgacga tcttgatcgctgcctgcctaaaaccgcgattgaaacgctcgaagctattcgccttttcttgtttgt agagaaaactgcatttgttatcggtgcagatgaagccatgatcgaatatgcggtaaaagaccattt ccccgacctgcctcaaagcaccgggccggtaagttatgcacgcaactatcttgaaaagctcataca ggttccatttcgaatccccgcactgggaactgcagaaacgcgtatatataccacgttgttgcttgc agaaaatgcgttgggttcggaggacgacaattttaaagcattgctcaataaagcacgggaagagat gaagcgtccttggatcagccgcgggcttgacagagaggcagtgatggcagcgttaaatggaaagat tccggaggttgtggaaaacgcgctgctattcagcctacacgttacccctatgcttagttcggggac acatggtaatccaaggcagattaaacgcttttttgaactcaatgatgttacgccaggcgattgctga tgaacgcgggttcggtagtgacattaagcgtcctgtactggcaaaaattatgcttgctgagcgttt ttaccccagcgtatacggaaagcttgttcagcttgtatctaatcatccagagggaaaaccggaagc tttggcggagtttgaagccttggtcagagggggaaaactgctccgaagagtcgcgctgacagcaa agagaattcctcagagtctgaagacgtccaaaactggctgaagattgattgggcgatcggttgggc aaaagcagagcccgcactttctggagaggatcttcgtccatatgtgtttgtcactcgtgacaaaca cagtactttgagtaatctggtcgtatcaagccatctcattcctataatggagaaacttcttggtcc gaaaattgggatggtgaaaatcaaaggggatttagagaaactgagtccaccggatgctgatgaatt attcgaaatgcttagcgataagcttttccaagaagacagtttcaatcgaaaaccaagaggatttga cggcctcgaatatctcgtagaaacacaacctcaccttcaaaggagattgattgattttgcacggcg cattcctgtaaaaaaagcaggggggatggcttgctacccgtattgcgcaaagcctagtggaccctac gttaatagaagaatatacaaaactgatccaagaatgggcgagtcaggacgaaaatctgtccctctc taaatcagcaaaagcaaccctccagttatcgggatatcaacattaatgggaacctcaaaagcttac ggggggcctgttcatggcctaatccccgatttcgtggagaatccatctccaccgaccctgccgcct gttgaccctgcggatgatagcacgctggatacgccgctcattccaccggattcgagtggctcaggg ccacttagcacaccgaaagcaaactttactcgatactcccgttcaggaagtcgtagttctctgggt aaggcggtcgctggatatgtccgcaatggagtgggggcgcaggcagggccagccgccgtatgggg gcctcacgcgctgcagcaggggggactgctcggtctcatcagcgactatcagcagggaggtgctact caggctcttgagcgcttcaatcttggtaatttggcagggcagtctgcatcgactgctcttctctcc cttgttgaattttttatgccctccaggtggttctgttgacgagggggttgcgcggcaggctatgcta gagaccatcgccgatatgtctgatgtaggagaggagaattttgatgagctcactcccgatcaatta aaagaagtctttattggtttcgtggttcactccattgaagggaggctcatggcggatattggtaaa aatgggatcaagttaccagacgacatagacgctatcgtcagtatccaggaggacctgcatgatttt gttgatggagctactcgtacacagctccgtgaggagctgaggaatcttacagggctttcaggggat gctatagacagaaaagtggaggagatttacaccgtggcatttgaattacttgcccgagaaggggag agattggaatgagccatcataccttagttgcccgtttgggcactgacgataactccgatttacagc tcagccgccaaagcacgcatctgacagaaattaattttctcaaagagaacggtaaactggatttcg gtctcgggcaggcgctgaatggtttgagtgatcttggtttaacgccaatggatgtctccgtggatc tggcactactggccgcaacggtgactgcggcggacacccgaatctcacgtgggcataacgctcaag atctgtggacgcgcgaaattgcactttatatcccggtagcttccccgacattatggaatagtcaga ctggattgctcagcaggatgttgaattttcttaccggcgaccgttggacaattcatttccgctcgc gccctgttattgagcacgggctcattcagcgatcctctaaggaacgttcggtgaaccctacttctg tttgcttgttttccgggggctcgacagcttcatcggtgccattgatttattatctaatgggggaa ccccccttctgatcagccactactgggatacgactaccagcgtttatcagcagaagtgtgctcagc tgctgtcggagcgatatggacaatcgttcagccatgtgcgagctcgtgttgggtttgaaaaaacaa cgattgagggagaagatggagaaaacacccttcgtggccgctctttcatgtttttctcgctcgcga caatggccgcagacgccctcggcgggccggtcacgataaacgtccctgaaaatggtttgatctctc tcaacgttcccctcgatccgcttcgtgtcggagcgctaagtactcggacaacccatccgttttaca

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.		Cloned Sequence
		tggcgcgttttaatgagctgctgggcaaccttggcatcagtgcacatctggaaaatccctacgcct acaaaaccaaaggtgagatggctatccattgccatgaccatgcttttctaaggcaacacgcggctg acaccatgtcatgttcgtctccgcaaagtacgcgttggaaccctgcgctgaatgagcagcaatcaa cacactgtggccgatgtgttccatgcttaatcaggcgagcatcattgtttacagctttcggcacgg acgatacgatttaccgtatcccggatctccgtagccgggtactggacagctctaagcctgaaggtg aacacgttcgggcatttcaatttgctctggcaagattggcgcgatcaccgagtcgagcaaaatttg atattcacaaaccagggccgctcagcgactatcccgactgcttagctgagtatgaaggtgtttatc tgagaggaatgaaagaagttgaacgcctgctgagtggagtcataacgaggccccttacatgaaatt agcaggacagaagcccgctccacaatgggtcgattttcactgtcatctggatctataccccaatca ctctgcactcatccgtgaatgtgacatttcacgtgttgccacgctagcggtgacgacaaccccaa ggcatggatgcgtaaccgggagttaacttccgattctccttatgttcgtgtcgcacttggtctaca tccccagctgattgcggaacgtgagcatgagatagcgttactggagcactatctcccttctgcacg ttacgttggggagatagggcttgatgccagcccgcgcttttatcgcagctttgaagcacaggagcg gatttttcccgtattctgaatgcctgtttcgagcaggggataagattctcagcatccacagcgt tcgcgctgcagccaaagtgttgggacatttggaaaacaccagacttactgaaaattgcaaggctgt cctacactggttcactgggagtatctccgaggctcgacgagctgttgaacttggatgctatttctc tattaatgaagagatgctacgttctcctaaacatcgaaagctggtgtcctttttgcctttcgaacg tatcttgacggagaccgatggaccttttgtgtttcacgaagaaaaagcgatacaccctcgtgatgt gcagcgtacggttcatgaaatcgcgcagatccaccacgtatcggacacagatgctgctatgagaat actttataatcttcgaagtttagtcaccaatagttctcacagtgagaatagttcatgaatctaatt agttggattaatacaggggaatagttgaatacttcagtcccctaaaagctaatatgctctatgtca tctaatgataagtggctccaaagagccacttatcattaacttttctaaagggaggtagaagt (SEQ ID NO: 294)
52	23	cggattgaatctgtttatgaaatttggctgctatcaactaatgggcgttaagttgattgtatgatc tgattgataaagaaggggagactaaaaatctcctcttctttgcagcagtttactgcggtcttttttg tgatgcatcagcataaaacgttttacttgtggaccctaagaaatggagaacattatgtcgactgta gatacctctacagcagaggaactcaatcaaggaggctcagattttattctgacttccctcgaggct atgcgtaagaagttattggaccttacgtctcgaaatcgacttttgaatttccctatcactcaaaaa gggtcttcactacgtattgttgatgaattaccagaacagctttatgaaaccctttgctcggaaatc ccgatggaatttgctcctgtgcccgatccaactagagcgcagctgttagagcatggctatctcaaa gttgggccagatggtaaagatatacagttaagagctcatcctagcgctaaggattgggcgcacgtc ttaggaatccgtacagattttgatttaccagatagccataaaacggttgtttctgattcagataga gagttgctggaaaaagcccatcagtttatcttgcaatatgcccaaggccagaatggaaaattaaca gggattcgttctgaatacgttaatcaaggtatagctttgtcagcgttgaaggaggcgtgctgctta gcaggctatgaagggcttgaggattttgaacgacaggcaaaggctgggaatgagattagtatatct tcttccaatccctctcatgacgataatcggatacaggctctgctttatccaaatgaactggaagct tgtttgcgcgccatctatggtaaggctcaaactgctttggaggagagtggcgccaacatcttgtat ttggcgttagggttccttgagtggtatgaaagcgattcctctgaaaaggcacgttatgcaccgtta tttacaattccggtgagatgtgaacgaggaaaattagatccgaaggatggtctttacaagtttcaa ctttattacacgggtgaagatattttgcccaatctctctttgaaggaaaaacttcaggctgacttt ggcctcgctcttcctttgttcaatgaagaggaaactccagagtcttattttgcttcggtgaagaag gttgtagagcagcacaaacctaaatggtctgtgaaacgttatggtgcacttagcttgctcaatttt ggcaagatgatgatgtatcttgacctcgatcctgcccgctggccttgtgacaagcgcaatatattg tctcatgaagtaattcgtcgctttttcaccagtcagagctgtggtcaagagaattccggcttacct ggtggcttcggtcagcatgagtactgcatcgatagttaccctgatattcatgacaaggttccacta atcgatgatgcggatagctcgcagcacagtgcgttgatcgatgctatccgtggtcaaaacttagtc attgagggccctcctggtagtggcaaatcacaaacgatcaccaacttgattgcagcagctctgctc aacggtaagaaagtcctgtttgtggcagagaagatggctgcactggaggttgtcaaacgtcgcttg gatcgtgcggggctaggtcaattttgcttagagttgcacagtcataaaactcataagcgcaaggtg ctggatgatattaatgctcgcttggtgagtcaggcgaccatgcctactatggaagagattgatgct cagattttgcgttatgaagatcttaagcagcagctcaatgaatatgccgcattgatcaataaccaa tgggcgcaaacaggcaaaacgatccatcagattttgagtggtgcaacccgttatcgtcacaaatta gatattgatgcaacagcacttcatatcgaaaacctttccgggaagcagttggataaagtgacccaa ttacggctgcgtgaccaaatagtagaatttagccgcatctacaaagaggttcgtgagcaggtgggg gctaatgcagaaatatatgagcacccttggagcggtgtgaataacacacaaattcaattgtttgac agcgctcgtatagtcgatttgctacaaacttggcagacatcaattatcgactttcaacatagctat caagaatatgtagataagtgggcgttagaaggcgaaagccttaatacgcttcaatatattgagcaa ttggtagaagatcagtcgaatcttccagtgttgtgtggttcagagcatttcccagcacttagtgag ctagattcacccgatgccattgcacgggtgcgtcactatttagataggttcgagttgctacaaggt cattatgtggccttgagccaggttatcgagcctcaaaagctacgacttttagaacaaggacaatcg tgtgactttcctcgtgaagagctggaaaaatatggtgcagcagaggatttcactttacgtgatttg gtcaggtggcttgaatccatccaatcaattcatgatgagttatcatctatttatgcgcaattaaac gatttcaaaaatgctttgccagatggtattgcttcgtatatcgatgattcgcaagctggattgcta ttctgctctgagttgttgtcgattctgggtgctttaccgactgagcttattagagttcgagatcct ctttttgatgatgatgatatcgatgcagtattgcgcgacttaatgtgtcaaatcgaaacattgcgt cctttaagagatggtctatctactttgtatcaattggaccagttgccttcccaagagatgctcgcg catgccgttgctgttatccagcaaggggattatttgcatggtttaagagtgattggcgtagtgcc aaggcactgctcatggcgcaatctcgaaagcctgacactaagtttgctgagttaaaacgctgctca gctgatttgctcaagtattcggagctgttacaacggtttgaacaaagtgactttggtaatcaactt ggtaatgcattccgagggttggacaccgactgtgaacaactcatgttattgcgtgattggtacaag aaggtccgagcttgttacgggataggttttggaaagcgagttgcgataggctctggattatttaac

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.		Cloned Sequence
		ctagatggtgagattatcaaaggtgtgcatttaatcgagaaatcgcagattagctcaagattaatgactttggttaaacgggtcgagcacgaggctaagttattaccgcgtatttctagcttgttggaagaacatgcatcttggttaggtgagcaaggtgtattgatgcaatcttaccgacaggtgcggaatactctcattgccttgcagggatggtttatcaatccagatatatcattagagcagatgactcattcctccgagattttgcaaaacataaacgatcttcagatatcccttgaaaatgactcgttacagttaggggcgttttacaattaaccccattggcttgcggtgcgtataaaaataatcaactgacgttagacactattaacgacacgctgaattttgccgagcaactggttgataagataaattgcgtatccttggctacccagatcagacatttggctagtggtagtgattacgatttactatgtcgtgatggtggagaaatagtttcgaaatggaatgaacagattaaaaatgctgagttatatgcgctagaaacaaagttagagcggagtcagtggctcaagtcgactgatggttctcttaatacattaatcgagcgcaacgaaagagcaatacagcaacccgttggttgaacgggtgggttaactttattcgttgttacgagcagatgcatgaaaatggattgcagcgaatctggagtgctgtacttgcgggctcgctcccgattgaaaaagttgaattgggtttagcattagcaattcatgaccagctggcgcgggaggttattcacatccaccctgaattgatgagagtttccggctcacagcgcaatgctttgcagaagtcatttaaagagtacgacaaaaaactgattgaattacaacgtcagcggattgcagcaaaaattgcttgccgaaatataccagaagggaattctggtggtaagaaaagtgaatatacagaactagctttgatcaaaaatgagttgggtaaaaaaaccagacatattccaattaggcaattggttaaccgtgcatgtaatgcgctggttgcaattaaaccttgtttcatgatggggccaatgtcagcagctcattacctagaacctggacgaatggaatttgatctggtggtgatggacgaagcgtctcaggtgaagccagaggatgcattgggtgtcatcgcgaggggcaagcaactagtggtcgttggtgacccgaaacagctaccaccaaccagtttctttgatcgaagtgccgacggagaagatgacgatgatgccgcggctttaagtgatactgacagcattttggatgctgctttgccactgtttcctatgagacgtttgcgttggcactatcgttcacgacatgaaaagttgattgcatactctaaccgccatttttataacagtgatttggtgatattcccttccccaaatgctgagtctccagagtatgggattaaatttacctatgtgtcaaaaggtcggttctccaatcaacacaatattgaagaagcccaagcagttgctgaggccgtacttcatcatgcgcatcaccggccgggtgagtcactcggggtagtggccatgagttccaagcaacgcgatcaaattgagcgcgctatcgatgaattgcgccgaaatcgccctgaatttaacgatgcaatcgatggcttacatgccatggaagagccactttttgtgaaaaaccttgagaacgttcaaggggatgagcgtgatgtaatctttatttcctttacctatggaccttctgagcatggtggaaaggtttatcaacgctttggacctatcaattccgatgttggctggcgtcgcttgaatgtgcttttcactcgatcaaaaaaacggatgcatgtgtttagttcaatgcgttctgaagatgtattgacgagtgaaaccagtaaacttggtgttatttcgttgaaaggttttttacagtttgccgaaagtggcaaactagattccctcacaacgcataccggcagggctccagatagtgactttgaggttgctgtaatggaagcactcaatcacgctgggtttgagtgtgaacctcaggtaggggttgcaggattctttattgatctagctgtgaaagatccaggttgtcctggccgttatttaatgggcatagagtgtgatggtgcggcttatcactcagctaaatctgctcgtgatcgtgaccgtttgcgtcaagaggttctggagcgtttgggttggagaattagccgcatttggtccactgattggttcagtaatcctgatgaggttctatctccgattatccgtaaactccatgagcttaaaacattggctccagacgttgttgtaccttcctatgaatatgtcgaaacgattgagtcaagcgctgaagtggcgtctgactcaattgattctcttatgcccaatttggggcttaaggagcaacttaagtattttgccacacatgtcattgaggttgagcttcctaatgttgatgctgatcgtcgtttgttgcggcccgcaatgcttgaggctttgctggaacatcagcctttatcacgttccgagtttgttgaacgaatacctcattatctgcggcaagcaacagatgtatacgaagcacaacgctttcttgaccgagtcttggcattaattgatggcgcagaggctgaagcgaatgatgcagcgtttgagtctgaattggcataattagttaaaggtaataagaacagtgacaactgtcgg(SEQ ID NO: 295)
53 54 55 56 57	24	atgatgaagatcacctaaaatgataggttgtttttatacagtaccaaattcaattttctctctataagatagattgcatttccgcggatgtagtttacaagggaaagacggtcaacatgcatcgcactatttctgagttttatcgcattccccctttacttattcgggcgctaaaaagtggaatttcctccgtggtggagtttcatctcaacaggggattacccaaagattcacgagattctctgggaaacagcccattgatgattgcggcccagtatggacatttcgctatttgcgaaatgttgttgagtgcgggtgttgatgttgaacatcaaaacaacctcgggcttcgcgctagtgaccttgcgcaggagcaaaaattgcgtgatctgttggcccgttatcgtcagcctctttcacttgccgaactggaacgctctgtggtttcagtcgaggactcagaaacagaggcagaattacccagcgctgaaatcccgatggattttatgctgtgggatgcagaagttgaattgaagcccgccgaagataatctgacgttaagacatgcttccgctgaagcccagcaattattatcacgctatcgcccgaaagataactctgctgagtggagcgatatcgaactcacgctccctgaaccactgacgccagtttctcactctccgcaaaattaccctcatctctcaacgttgctcattggcgcactggatacggggcgtatctctttgcgtgacatctggcatgccggggaagaggatttcggtatgcagtggcctgaattccggctcagcgtagaggcattgatcagggacttaccgctgattgtggatgacgatgatattattccgcctgacgctgctccggcgacattatcggtgagtgaacctcttgaaccctggtttgatgctttcaatgcattgcggcagttcggcatcgttgaaaactatctcgtggatatccgccagtgggatgtcgtggataaaacaaaagaagaacgactcggccagcgcatggatacggcgctaattaatctgataagaatcctggcgggtttatccgaagcggaatatatgcagttgctgcagcccaattaccttccggagccagcgcctgagatttctgaagaggaagacgtcgcagaagaagcggatgaggaaatgcctcccgtatccgatgacgatgacgataacgatgacactatcagctttatcgagcttcttgttctgctgagaagtgggaaagcaggcgagtatcaggataatcatatcccccgcccggagtatgccgacctgcaacagatagttgagcgcgcccgaacgcttatccctgatgaaggtcataaaataagtctgtatgtcagcagttacagagaggcttgggaggggctgatccacgccaacttgcgtctggtcgtcaccatcgcgaataaatatcgcgggcggggattagatgtcgaggacctgatccaggaaggtaatctgggtttgatcaaggccgttgaaaaattcgactatcgacgcggatttaaattctccacgtatgccacctggtggatccgccagaagatcagccgcgcgattgccgatcaggcgcagctcatccgtttacccgttcacttctatgagcaattcaggcgctggcgaaacagtcgggatcaattgctgtatcgccaggggataacgcccacgatcaaacggctgcaagcattgactgaccttccagaaaatcaactcaagcggatggcaaaatatgaagaacagacggtgttgattggcgattttcatgatgacgcccaggacagcgaagcggcgctgtcgggagacgcgatcc

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.	Cloned Sequence
	tgaccggaaaggatttcaccagtgctcccgttcagtctctcgagctaagagaatgtgtttcattgg
	tgctggaaacgttgttgccacgcgaaaaacagatcataaaaatgcgttttggcatcggtatgacgc
	aagatttcacgctggaagaggtgggtaaacagtttgatgtcacgcgagaaaggatacgtcagatag
	aagccaaagcgctccgtaagctccgctatcacagccgggcgtcgaaattaggcggcttcgtcgaac
	agtgggaaaccgcgttgagcgagatgcaggaagaagaagaatgacgaccatgcgccatgcgccacc
	gaatgcagccattatgatcgaagcgctgcgagggctcggttacaacactgccaccgcactggctga
	catcatcgacaacagcattagtgccggtgcccgtaaggtcgatctgacctttcactggcgtgagtc
	ggatagctatatcgtggttcgggataatggttgcggcatgtcggccgctgaactggatgttgcgat
	gcggctgggggtcaaaaacccgctgacaaagcgttcaggacacgatctgggccgcttcggtctggg
	actcaaaaccgcctcctttcgcaatgtcgccgtctgacggtcgcctccaaaaaagaggagataac
	gaccatcctgcggtgggatctggacattctcgccgccagtacggacgacggttggtatttgcttga
	aggcgctgacccaggaagtcaggaggcgttagcaaatgaggaacctgactcccacggtacggtggt
	gctgtgggacgttttagaccgaattgtcacccccggctacggtgagaaagatttcctcaatctgat
	ggatggcgttgaacaacatctggcgatggtatttcaccgattccttgaggggaacgctccccgact
	cactctcaccctcaatggtcgcaaaattaaagcttgggatccctttctcagcgggcatccttccaa
	gccctggcattcgccttcggcaatggcgccaggcgctcctgccgtgaaggtggagtgtcatgttct
	gccgcatcaggatcacctgacgacgcaggagtatcaacaggctcaaggaccggcaggctggacggc
	ccagcaaggattttatgtataccggaatgagcgattgctggtggcgggcaactggcttggactcgg
	aagcccccgggcctggacgaaagatgaaacccaccgccttgcgcgaatccgtctggatatccctaa
	tgatgccgacatagactggaagattgatattcgtaagtcgatggcccgcccaccggtttcgctgcg
	gccttggttaacccaactggcgcaatcaacgcgtgatcgtgcggtacggacatttgcaaaacgcgg
	gaaaatgaataagcgcaagcccggcgaggaacttgttcagctctggcaagcgcagaagacgccatc
	cggtgttcgttatcagatttcgttacaacatcctgttatcagcaatgtcctttcgcaggccggtga
	gttatctccacaaattcaggccatgctaagactgattgaggaaaccgttccagttcagcaaatctg
	gcttgatacggctgagacaaaagagacgccgcggacaggttttgaaactgcaccgcccgcagaggt
	gttgtccgtattgcaggtgatgtaccagactatggttggacagcaggcgatgtcaccggcgctggc
	gaaacagcacctgcaaaatatggaacccttcgataattatcccgaattaattgcactactccccga
	cgatcaacatgagaaatcgctatgagtcttaatcccttggatgacacgcaactgagtgtattgcag
	attgtgcaaacgttcctgcaaagtcaggataaaagcacgatcacgcccggtattctgcgccaacat
	attgatatggtttgtcagatgaaacctgagtggagccgccttgatagtcgggagatcctggtcgaa
	gagttgatccgccgttacagcatctggatgggagaagattcttctctgagtaatgacgaagggcat
	caaccctggctgaccgctgatgcgaaacgcgagtggcgctactggcatcgatatcgccagtggctt
	ggcaaaacgatgccttggggagtcctggatacccttgaccgttcaacggatcgtgttctgggatta
	cttgagcaaccggggcgggaagggcgttgggaccgacgtgggctggtggtcggccatgttcagtcg
	gggaagaccagccactataccggtctaatctgtaaagccgcggatgcgggatataagataatcatt
	gtgctcgctggtttgcataacaacctccgctcgcagacccaaatgcgtcttgatgaaggatttctt
	ggttacgagacgagcccactcagagaaaaagtgaccatcattggggtgggcgctattgatagcgat
	cctgtcattcgtcccaactacgtcactaaccgatctgaaaagggcgacttcagcgccggagtggct
	aagaatctggggatcagccccgagcaacggccctggctgttcgtagtaaagaaaaataagtctatt
	ttgaagcgcctgcatacctggattgagaaccatgttgccaccagcgttgaccccatcaccggaaag
	cgttttgtttcggaattaccgctgctgatgattgatgatgaagcggataacgcctcagttgatact
	ggggaaatcgtctacgatgacgatggaaaaccggatgctgaacatcagccaacggcaataaatagt
	ctgattcgtaagctgttgatgcagtttagccgtaaggcgtatgtcggatataccgctacgcccttt
	gccaatatttttattcacgagagcaatgaaacacgtgacgaaggtccggatttgttcccttccgcc
	tttatcattaatctcggcgcaccctctaactacatcggccctgcgagggtatttgggcgggccacc
	gcggaaggccggagcggagagtttcctttgattaggcgagtgagtgatcactgtagcgatgacgga
	aaaagggggtggatgccggtttctcataagagttcgcactatcccacactggatacgctaactcat
	ttcccggactcgttaaaacacgctatcgacagtttttttactagcatgctgtgtcagagaattacgc
	ggtcagggagagaaacacagttcgatgctggtccatgtgactcgcttcaataaggtgcaatcggtt
	gtttatgaaaatattgatgcctacattcaggacgtgaggcagcgactgacgcgaaggattggacac
	gaaccttttttacatcagcttgagtcactctggcaggccgatttttttgccgacgaatcaggcgatc
	cgcgaagttatgccgcagcaggttccggacgacgccttcgaatggcaggagatcgtcgacaagctg
	tataccgtgatagaaaacgtgtcggtacgaatgataaacggaacggcgaaggatgcgcttgattat
	tcggacagtgcgacaggcttaaaagtcattgcgattggcggagacaaactggctcgagggctaacg
	cttgagggattatgcactagttatttttttacgcgcctcccgcatgtatgacacgttaatgcagatg
	gggcgttggtttggttatcgccagggatatctggatgtatgccggctttataccaccgatgagctg
	attgaatggtttgagcacattgcggatgcgtcagaagagctgcgggaagagtttgacaatatggtc
	gccagcggcggcaccccacgtgatttcgggctaaaagtgaaatcacaccctgtgttaatggtgacc
	tcgcccttaaaaatgcgtagcgcgcgttcactatggctctctttcagcggcacagtggtcgaaacg
	atttcgttgtttaaagaacaggagtatcacaagcgtaactacgtggctttccagcgtctaaccggg
	cgcgtcggtgctggcgcgccgatacctgagagacgacgcggagataagattgaaaaatggaatggg
	gtcatttggcaaaatatctcccctgagccgatcatcgatttcttaacggaatatgagacccatgct
	caggccagaaaagctaacagcaaactactggcggattttgttacgcggatgaatcgcgttgatgaa
	ctcacccaatggacggtggcggtgatagggggtggcatcgatcgccatcacgatgtttgcggcttt
	tccgtaccgcttatgatgcgtaaagcgtctgaaggggtcactgaccgttattccattggccgttta
	ctttccccacgcgatgaagggattgactgtgatgaatcaacttggcttgctgcgctggaagaaacg
	cagcgtatttttcatgccgatcccggacgcaatgaagggcgagaggagcccgtcgttccaggtggc
	gtggtactgcgtcggattaaaggatttggcattaacgacattccagcacagcgtcaaaaaggttta
	ttgctcatttacttactggacccgcagcaggcattgtcggcagcggaatatcaggaagatgcctta
	cctgtggtggcttttggcatcagttttccgggaagccgcagtggggtaacggtggagtacaaagtg
	aacaacgtactatgggagcaagagtatggtgcggctgagtaaagacgatctgctggcggcctggaa
	agccttagatcgatctcagatagacgaactgcctggcgctcagggctggcgcgggattcggctttt

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.		Cloned Sequence
		tacgcaccagggctgtagctttcatgccgggcgtcgtcagcctgataatgaagaaatgctgattgccgtgtttcctcatcctctttcgcctgggtcggcggcgctgccatcttgtaaaggattccgcgttgagatggccggaacagaggagggggggcagaacggtttgatgatccgtcgccagcaaacagggaatgtggatgtctttacgacgatgattctggatattctccattcgctcctgaacgtttcgaaaccgcgcctgtttgaaactctgcttcgtcggattcgtttatggcaggcgtttatggagcgcgatacccgtccactcagtcaagaagaagaagttgggttaatcggcgaattgacgtgtctggagcggttgatcgagagcggtcttgctccgtcaacggcagtcgaagcatggataggaccgcagcatgggctacaggattttgcactcgatgaacgcgccattgagataaaaagcactacggcagcgaagggtttttgcataactatccactctcttgaacaactggactggcagcgggcaggatcgcttgtattgtgtggtttgcgcttcagcgagcatcccaccggcgcaaccctgaatgacatcattagccgtcttcgtcaacggtttgagggaaacgctacggcggcttgtatttttgagggatcactttgtcatgtcggatatttcactgaacatgctgaattctatacacgtcatttcttgctgacagaggcgttcgcactccccattgaagcggattttccctctttgacgcatgccaatgtcccgttgccggtggtgagtgcgcgctatcaactcgaactccagacacttattcctcaggcccaagattttaaccattgcttgtcagactttgcaggattaccgcatggaaattattgattttttacgtcaaacccagaatgagattcgcaaggaatatcaggatcaaatggctcagccaggggttgagtcgccttttccggagctgatttttaccgatattgttatgcgtcatatggccgatatcggcatgacattcgatgatgccgagacgtgtcactttatggcgaaagtcagtggacacaatgtgcgtctcagcggttatgccttctcagaagatggcgatcaacttgacctttttgtcagtatttatcacggtagcgacgagctctgtcacgtcccggatgctgagacaaaagcgattgccggccactgcattcaattttgcagaagtgcgttgacggtaaattatcatccacgctcgatcagtccaatgatgcctggcaactggtgacgaccatcgaacagtcctatgcggaactggagcaaatcagaatttatgtactgaccgatggtcaggtgaaaacccgctggtatcagtcacgggacgtggccggtaaaaccattaaattagaggttatggacattgtccgtttgtttaaccactggcaggaaggtaagccacgcgatgaactgcaggttaattttgatgaggtggctgggggggcgcttccctgtgtctggatcccggatgaaatgggtgagtacgattatgcgctgacggtggttccgggagagacactgcgatttatctatgaaaaatatggcaaccggattctggaagcgaacgttcgctcgtttctgagtcagacggggaaagtcaacaaggggattcgtgacactttacgtgagcagcctgagcgttttatggcttataacaacggcattgtgattgttgccgatcaggtcaggcttggtgaagcaccgggaggtggccctggtattgcgtggatgcaggggatgcagatcgtcaacggtgggcagacgacggcttccatgtttttcaccaaaaagaaatttccggcaaccaatctgcgtaacgtgcgtgtacccgcaaaagtaattgtgctgaaacagacgaataatgcacaagaagagatgttaattgcggatatttcgcgcttctcaaatagccagaataaagtcaatatttccgatctgtcagccaatcgaccagtacatgtacagctggaaaaaatggcaaacacggtgtattgcccggacggatacagtcgttggttttacgagcgagcaaatggcagttataaggttatgctggaacgagaaggtaaaacaccggcgggcattaaacggttaaaagacgcaattcctccatcccgtcggataacgaaaacggatttcgcaaaatatcactgtgcctggctccagcgtccggatttagtcagcctcggtgggcagaaaaactttgccgcattaatgacgatgattgacaaggatactgagcgttatggggatgaactgaacattgaaacttttaaaaattacattgcacaggctattatttataaaaaagcctataagttgattaattcacttttccccgcatttaaggcgaatatcgccgcctatactgttgccgcctattcacatctttatggtaacaaaacggatctggcagagatctggaatcaacagggtatcgaggaaactatggggaatcgtcttgtcagcttggctcaccgagtaaatagccttctgactgaatcggcaaatggcaggatgatttctgaatgggcgaaaaagccggagtgctgggactacgtgcgcagtaaaatctatttctccgcacagggaaaaaaggatgacttctcgcatggtgaaattgcatgatgagttcagtatcaacatgatatgtgagtattactgacgtatggcagcggttgttttgtatggatgtgctatggcatcgcatcaatatacaattaacagctg (SEQ ID NO: 296
58 59 60	25	cgtgatgaatgaagcggctaaatacattaatgataattataatttaattcattaaaatcagtaatatataaatataaaagttgtgaaatgtgatattcgtcaaagcatgtcaaaaagttttgactgttctttaggcatcattcgcaattgtctaacaacttgataggataggaacaatctcaaaaaggaaaatgacatatggcatacgaagctcaaatcagccgtactaatccagcagcatttcttttcgtcgtcgatcagtcaggttcaatgtccgacaaaatgtcttccggccgaagcaaggctgagtttgtcgccgatgctcttaatcgaactttaatgaacctaatcactcgctgcactaagtctgaaggcgtacgtgattatttcgaaattggtgttttgggttatggcggtcaaggggtttctaatggtttctctggttcactgggaggacaagtcctcaatccaatttctgctctcgaacagaatccagccagagtagaagatcgcaaacggaagatggatgatggagctggcggaatcatcgagacagcaattaagtttccagtatggttcgatcctattgctagtggcggcacgcctatgcgtgaagccctgaccagagccgccgaagagttggtgacttggtgtgatgcccatccggattgctatcctccgactatcctgcatgtgactgacggcgaatcaaacgacggtgacccggaagagattgccaatcatctacgacaaattcgcaccaatgacggtgaagttctgattcttaatatccatgtcagttctctcggaaatgatccaatcagattcccctcctcagacactggcttaccggatgcctacgctaaactgcttttccgtatgtccagccctcttccggaacatctggtgcgtttcgcgcaggaaaaaggtcatacggtcggtatagaatctcgtggattcatgttcaacgctgaggctgccgaactcgtcgatttcttcgacatcggaacccgcgcttctcagttgcgttgattcagcaatgaaactggagttcttagggacagttccgaaagatcctgaataccctaaggcgaatgaagataaatttgccttctccgaagatgggagaaggctggcgctatgtgatggcgcgagtgagtccttcaactcaaagttatgggccgatcttcttgctcgtaaatttactgcagatccgaaagtaaatcctgaatgggtagcatctgctttagcggaatattctgccacgcatgacttcccttctatgtcctggtcccagcaagcggcattcgaaagaggcagttttgcgacactaataggtgtagaggaatttgaagagcatcaggcggtagagattcttgctattggagatagcatcaccatgctggttgattgcgggaaactcatttgcgcatggcctttcgataatccagaaaaatttaatgagcggccaacactgcttgctacgctgtacgctcataacaatttcgtcggtggaagcactttctggacacggcatgggaaaactttttaccttgaaaaactcacccaacccaaactcctctgtatgacagatgcgctcggcgaatgggcactgaaacaagcgctggcagaggattctggttttatcgaattactttcgctgcaaactgaagaagagcttgcagagttagttctgagagagcgtgcagcaaaacgtatgcatatcgacgactcaacgctgcttgtactatcgttttaacgcggaaagtaaagatgccttacccatctcttgaacaatacaaccaagcgtttcagctacatagtaagctgctaatcgatcctgaattg

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.		Cloned Sequence
		aaatctggtaccgttgccacgacagggttgggtctcccccctagccatcagcggtggctttgcactgacctatacaatcaaatcaggcgctaagaaatacgccgttcgttgctttcatagagagtcaaaagccttagaacgccgttatgaggctatatccaggaagatttcaagccttcgctctccctactttctcgatttccagtttcagccccaaggggtcaaagtcgaaggaatatcataccctatcgtcaaaatggcatgggccaagggagagacgctaggagaattccttgaggtcaacaggcgttctgcacaagcaatagcgaaactatctgcatcgattgaatcacttgccgcctaccttgaaaaagaaaaaattgcacatggtgatttccagactggaaacctgatggtctccgacggaggtgcaaccgtccagttaatcgactatgacggcatgttcgttgatgagattaagacattaggaagctcggagttggggcatgtcaattttcagcatccccgtcgtaaagcaacgaatccgttcaatcacactctggatcgtttctcactaatttcactctggctggctcttaaagccttgcaaatcgatccgtccatttgggataaatcaaattcggaactggatgcaatcatttttcgagctaatgactttgtagaccccggttcatcttccatcttagggatgctatcgggaattcaacagctttccacccatgtaaagaattttgccgcagtctgcgcttcagcgatggaaaaaacgccttccctcggtgacttcattgcaagtaaaaacattcccatatcgctagcttcgatcagtatgaatggggatattccagtcagcaggctgaaacccggttatatcggtgcctacaccgtcctgtcagccttggattacagtgcttgccttcagcgagttggtgataaagttgaagttatcggaaagattattgacgtcaaactcaataagacccgaaatggcaaaccatatatctttgttaatttcggagattggcgcggtaatatcttttaaaatatcaatatggagtgaaggcattagcgctttaccttcaaaacccgatgcctcatggatagggaaatggattagtgtaatcggccttatggaaccgccttacgttagcgggaaatacaaatattcacatatctcaattacagtaacgactatcggtcaaatgaccgttctttcagaaccagatgcccgctggcgtcttgctgggccaaacgaaagtcgacaaacattaacttctactagcagtaatcaggaagccttggagcgcattaagagtaagagcaccacttcaactcctatgcccatgaacactaacgccacaactgcaaatcaggcaatccttaacaagttacgggcttctacgcaaactgtagcggcagcaagagcgcaaactcagcatgtagtacctaataaatcatcaacgcattatgtggcaccgacgggaacatcagcttcgcagccagttcaaaatattccgagccctgctagtacctcaaagcagcaaacctctcaaaaaaatatagttacaaagatttgaaatggctttttggatgattggtacttgtaaagaacaagcgcaatttcagtggccgtatcacttgcgcttgaggtgcctgcgggtatgatcttgcgacatacaccactaaaacgaattcgtggcggcacttttagcctgcccctgtgttttcccgaggatttac (SEQ ID NO: 297)
61	26	gggctgtttggttgaattaaaaatacgaactaaaaccaacaagagtcggaaaaaacttcaaaatgctgcttatggataatagtcatcttaaaaatgtacggaaaaagagactaaaatcagaaaaacatctgttatacattgacttaaagtcatcatctccgctatgagtcctcaatccaagttgacaaatgtttagccaggagttcccgtgaacgagcatctctctcatatggatgtacataccttgtttgaagaaatggacgagcaggctgatggaataacgtttaaatactcatttgatgacatagcaaagagcaacgcattggttgtcactgagtttgtcaattttgagcgtgacagcacggtagctttactcgccagccttcttactctcccggcacaccaatctcagtgtttgcgctttgagcttctgacgagccttgcactaattcactgcaaaggtcagcagatagcaaatatcgatgacgtgaaacgctggtatgtcactattggggagtcgagtagtatcgttggagaagatcctgctgaggacgtcttcgtcgcccttgttgataataaaaaaggtgattaccgtgtgctagaggggggtttgggaggcggcaggtttttatacacaattaatggtcgaaattgtatccgacatgccggatacgcaccgctatcgctcgctgaaacttgctatacaggcaattctccgtctctcagatgtcatttgtgctcgctctggcctttatcgttttcaggaaggcgcagacgaattccctgactctcttgacaccgctggtcttgatgagaaaacgctctgttcaagggtaacgttgtccgagcgttctcttcgagctgaggggatcaaacttgctgacttagcacctttcattcttgaaccttctcatataagtatgcttggaaatcaggtccctggggagggaatgcttgaacaacggccattgctccgcacacgcgatggtattgtggttgtacttcctaccgccatgaccattgcacttcgccaggcagtgataacatttgcaaagcgcacagaagaattgagcgagctagacaaagcgttagctaacgtctacagccttactttctccgagatgccggtcttcggtaatggaggaaggttaagaagactgacatgggagaagtacaaaatgagccgaacaacgatggtaacctccatcgtggatgctggtcatttgatggtacttcagttcgttttgccttccatacagcaatatgccgataccggtttcaacaacttgctacagctagatgaagagaccacgcaatttctagataactctgttgaacaaattacagttgacctcgccaaacaacccggctttcagcgtggcatcgtcgtgcgcattgcatgtgggtgggggcgggttttatggggtccctccccaactgccagatggttggggatttgaatggatgtctggtgcggactttgtccggttcggggcattacccgatatgtcaccaattgccttctggcgtgtgcaagacgcagtcgaaacgatcaggcaagctggtgttcgattaatcaatatgagcggaactctcaatcttcttgggtggatacgtgccaatgatggccatatggttcctcatgaccagttaccagatgaccgtatcacaccggaacacccgctaatgttaatgattcccacgaatttactccgtggtatacgaatagcggcagacacaggatatgaccggcatcgcattagtgacaacaatggtaaatggcatcgagtgatgaggccttcggcagaagatttctttcccaccgagcgtcagagcaagtgctacgcatcaattgatgatcttgaagcgcaacggctgacctgtgtatatgaggggcagggtaatctttgggtaacgctcgaagctccagaaatggaagattggatgctcctcgttgagcttgccaaaatggttcgaacatggattgggcggattggcgaggcactggaggtcttgagtgagcaaccaataaaaaaatcattaaaggtgtatctgcattttgatggtaacgacaatatcggcagatttgatggtgagaatttttctgatgatatgaatacattttggcgacttgaacgaatccatgagcatggggcgattcgtgtggttcttcaagatgggtatcttgcaggttttcgtctaccggataaccgtgcagaacgagctctggtgcgcgcactcggtacggcgtttgccacacttcttcggatgaaagagccagtagacaaaggggtcactgttgagcagatagcggtgcccaatgacagagcgcgcagcttccacataatgcaggcttatgacttcaaccaatatttaggccgttcactaactaaacgtcttttagctattgaagatatcgactcagccgcagcccgaattgagctagcatggcgtgctgtttcgacagatgcaccatcacgatatcagggtaaaaaggaagttggaaagctccttaatgatgtggttgatgtgctgatccaagacttactaagcgaactttcaagatttgaccgtaaacagacagtaatgcgattacttgaaaacgttgtaaaggcacgttgtgaagaggcgcactggcgtagtactgcagcagcggtccttggcttgcatgcaggagaagagggtgtcgaagagacgatagctcaagaaatgagccgttatgcgggcgcagcgttaacttcccggctaatcattgaacttgccatctgtgtgtgcccgacaagcggtggaattgaaccttctgatatggcactcagtaaacttcttgcacgggcatcactgctttttcgcataggtggtatgtcagatgccgtacgtttcggtgctttgcctgctgatattcg

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.		Cloned Sequence
		catctccccttaggtgatctcctctttcgcgatgaactcggcaaaatggtgcttgaaccaatgctttcaaaagttactaacgaacggtttgaggaacaagcggcacaattcgagcaacactatgtgaaaactgccggaggggatgatgagaatagcaaacaagatagtgttgcggctgaaaccaccgaggaccaaaccgatattttccttgcattctggaaagcagaaatgggcttcactctcgaggatggaatgcgatttatccagttccttgagtccatcggaatagagcaagaatcagcaatcttcgagatgcgaagaagccaattagcggatgctgctaaatcggctgggctcgcagatgaaactattgatgcgttcctcaaccagtttatccttagcgcgcgtccgaaatgggatgtagtgcccgatggatttgacctttctgatatatatccctggaggtttggccgacgcctttcagttgctgtacgtcccttgttacagattgaagagagtcacgatccactaattgttatcgcaccaggactcttgaatctgtcccttaaatacgttttcgatggcgcatacactgggcaatttaagcgtgacttctttcgcacagagggtatgagagacacttggttaggtggagcgcgggaaggacacacattcgaaaaaactttggagagagaacttcgtgaaataggctggacagttcgacgtggcataggctttcctgaaattcttcgcaggaatctaccaggtgatccgggggatattgatcttcttgcctggcgctcagaccgcaatcaagttctcgttatcgaatgtaaggacctctcacttgctcgtaattactcagaagttgcctcgcaactatctgaatatcaaggtgatgacataaagggcaaaccagataaactcaagaaacaccttaaacgcgtattactagccaaagaaaacatcgataattttgccaagttcacttcgatagcgaatcccgagattgtatcgtggctcgttttcagtggagcatctcccattgcctatgctcaatccaagattgaggctttggcaggaactaatgttggccgcccaagtgatcttctgaactttttgatagatatgctgtgcgataagacgccctggcaactaagttaatcgttcctactactgatagttttaaatcaagg (SEQ ID NO: 298)
62	27	gatggactggtactgtagattcaccgtggaccagcgaatctattatgtggtgagcagaacattaacacatcaatgtaacgccgtaatcattgagtctttgccggggacgcttgacatctccgaaagaattatatcgtgagtcttaaggggaatctcttgcttccggttatacatttaaccggatctagctataagactgttacatctattgggattaggtcaggacagatagcctgaaagcttttatagtgagggacttcagaaataccctagaaaaggaactgttatggtaggttcgcgctggtataaatttgattttcataaccatactccggcttcgcatgattacaaaattcctgacatcagccccagagagtggcttctggcttatatgaaacagcatgtcgattgtgttgtaatcagcgatcataacagcggagcctgggtcgacgtgttgaagggtgagctggagaatatgtcccgggacgccagcaccggcgacctgccggaatttcggccactgacactctttccgggggttgaactgacagcgaccggtaacgtacatattctggctgtgctgcacacgcacagtacaagtgccgatgtggaaaggcttctggcccagtgcaataataatagccccattccgagtgaagtccctaaccatcagctcgttcttcaactgggccccgccggcatcatcagtaatatccgccgtaatccgaaggctgtttgtattcttgcgcacattgatgcagccaaaggtgtcttaagtctgactaatcaggcagagctcaccgcagcctttcaggaaagtccccatgccgttgagattcgacaccgggtggaggatatcaccgacggaacccgccggcggctgattgataatttaccgtggctacggggctctgatgcgcaccatcctgaacaagccggcgtgcgaacctgctggctgaaaatgtcatcccctgattttgacggactcaggcatgcactgctcgatccggaaaactgtgtgctgtttgatcagctccctccggaggaacctgcgtcatatttgcgcagcctgaaattcagaacccgccactgccatcctgtgggtcaggattcggcctcggtggaattcagcccgttctataacgctgtaatcggctcaagaggcagcgggaagtccacgctcattgaaagcattcgtcttgcaatgcgcaaaacagaaggtctcactgcgacccaggggagtaagctggaccagttcattcggacggggatggaagcggattccttcatcgaatgtattttccacaaagaaggcacagatttccggctcagttggcgaccagacagtaagcatgaattacatatcttcagtgacggagaatggatgcctgacagtcactggtcggctgaccgttttccactctcgatttacagccagaaaatgctctatgagctggcttcggatactggtgcattcctgcgcgtctgtgatgagagcccggtggttaacaaacgggcctggaaagagcgctgggatcagctggaaagggaatatctgaatgaacaaatcacgttgcggggcctgcgtgccagacagggaagtgcggattcgctgcgggggaattatcggatgctgaacgtgccgtcagtcagctgcagtcaagcgcctattatccggtttgcagacagctggccctcgccagaaacgagctgtccgcagcaaccttaccccctggagcactttgagcggcgtattgcagccattcaggctctggcagaagaaccgctgcagagatccgatatcccgccggaaccttccggtctgctgatggcatttatggcgcgcctgtcatctgtgcaacagcagtatgaccagcggctcaatactctcctggcagaatatgctgcagagctcgcgggtatcaggagagagcaatcttttattgccctccgaacagcagtgagtgaccaggaaacaaatgtagaaagtgaagctgtttccctgcgggccagagggcttaatcccgatgttctcaacgaactgatggcacgctgtgagtcactgaaaaatgagctgagaaattacgacggtcttgatggggcgatctctgcctctgttgcacggtctgagcagttgctggctgaaatgcgtgcccacagaatggcattgacagataaccggaaggcgtttctctcctccctgtcgctcagcgctctggaaatcaaaattcttcccctctgcgccccttatgaagatgttatatctggttaccagacggttaccggcatcagtaattttgccgaacgtatctacgataacagtgacgggagcggattactgagcgactttatcagtgaacgtccgttcagcccgttgcctgccgcaacagagaaaaaatacagggcgctggacgagctgaaagcgctgcatcacagcatccggctggataattcagaggctggggcggggcttcatggttctttccggaatcgtctcaggagtctgaatgaccagcagctggatgccctgcaatgctggtatcctgatgacggcatccacatacgttaccagaccccccgggggggcagatggaagacattgcctttgcttctccggggcaaaagggagcgagtatgctgcagttcctcttatcctatggcaccgatcctctactactggatcaaccggaggatgacctggactgcctgatgctgagcatgagcgtgatccctgccatcatgtcgaacaagaaacgccggcagctgattatcgtgtcgcactctgcccctatagtggttaacggcgatgcagaatatgttatcagtatgcagcacgatcgcacaggcctgtatccaggactctgcggtgcactgcaggaagctccgatgaaggcactgatatgccgtcaaatggagggggggagaaaaagcgtttcgttcgcgctatgagcgtattcttagctgaagaacggaaccgtccttaaggcggccatgaccggagagtgggcctggcggctgaatgcctggataaaagacgcaaatgtcagactgatggcctctgcgtctttg (SEQ ID NO: 299)
63 64	28	atagaacgatgaaggatggaagctacatattctcggtactaagatttatttttctgacacaaaatgaccatttggcgttacataatcccaaaaaaacgtatcaaaaatctcaaaatgcgttacgattagagagtattttgattctgcgtgctcatttttttgattgctgtggctttttgttgtgggagtgttgaatggattatttatcagaagtgttaaaaatcattgaaggtgcaacaaaggcaaatgcttcgatggctagtaa

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.		Cloned Sequence
		ttatgctgggttgctggcagataagctcgaacaaaaaggggaggtcaagcaagccagaatgataagagaaaggttgcttagagctccccaggcgttggcaggagctcaaagggctggaggtgggatatctctgggctcattaccggtagatattgatagtcgactcaacactgttgatgtcagttatcctaaattagacagttcagagatttttctgcctgcagcaatcagtacccgtgttgaagagtttatcactaatgttcaacgttatgatgagtttgttaaagctgatgcagcattgccgagtcgtatgctcgtgtatggaaagccaggaacaggtaagactatgttatctaagtacatcgctacccgcttagattttccacttcttacagtgcgttgcgatactttgattagtagtttattgggacaaaccagcaaaaatcttagacaggttttcgattatgtaatgcagaggccatcagtgctttttttagacgaatttgatgctttagctggagcaagaggtaatgagagagatataggtgagcttcagcgagttgtcatttcactattgcagaatatggatgcggcatcagaggatacggtaattattgcctcaactaaccatgagcaacttctggatcctgcaatctggaggcgatttagcttcagaattccaatgcctctgcctgacatacatcagagagagttaatttggaaaaatcgtttaaagaatatgatatgtagcgatctagatttaagtgatttatcaagaaaatcggaaggattatccggagcaataattgaacaggtgagcttggatgcacgtagggatgcagttattgaaggtgcaagtgtgataaatcaccataaattgtataggcgtttgtatcttgctcaatcgcttatggaaggtgtaaatttaagcacttacgaagatgaaattcgttggttacgttctaaagataaaaaattattttctatcagagttcttgctaatttgtacaaacttacatcaagagtaatttcaaacattctgaaggagtcaggagcatatgagcagaaggggtacacagtttagtaacgcaaaagttacaaacccaatgttaagaatcccttttccagtagtgacttgggtgcaatagtaaacgctggcggtggggcaaaggtattggttgatgtaacagccgaatatagacaagggctagtaagaaatttaacaaccagtaaacattatttagaatccaaactttcagagtaccctggaagcttgggtactttggttttcaaattaagagaccagggaatagccaaaacgcataggccgaacaaaattgctcaagaggctggattgcaaaatgccggtcatgccaaaatagatgaaatgttggttgctgctcatgccggctgttttgacgtattagagtcagtcattttacatcggaatattaaagcgattttggctaatctaagcgcgattgagcgcattgaaccttgggatgagaataggaaggttccaggaggcactgatggtttgtttgaatcatcaaacatccttgtacgactatttgagtacacaggtgaagatgcaacttacaacaactatgaaaacgttatttctatattagaacaacacggagttaaatatgatgagattagacaaaaatgtggtcttcccttattaaggataatggatttatccccaaatgatagatatatattagacattctcattgattacccgggtataagaacgttaattcctgaaccaaaatattcagcattcccggttagtgtaagtgattctgttggcattgaaacaaatagctttcccgtaccatcagaagaattacccattgttgctgtatttgacactggggtaagccccatcgcggcaacaattactccttgggtagtgagtagggaaacatacgtaattcctcctgatacgagttatgaacatgggactatggtgtcttcattgatatcaggcgctcatttttaaatgacaatcatccatggattcctgatacaaaatctaaaatccatgatgtttgtgccttagatgaaaatggatcttatatatcagatttaattctgaggctagcagatgctgtaaataaaagaccagatataaaagtctggaatttgtctttgggaggcggaccatgtaatgagcagacgtttagtgattttgcgatggagttagatcggctcagcgataaatttggtattttgtttgtagttgctgcaggtaattatgtagatgaacctatacgtacatggccaaatcctgatccgcttggaggtgctgatttaatttcctctcctggagagtcagtccgagcactaacagttggttcagtttctcatatggaagctaatgatgctttaagtgaaattggaacaccgacaccatatactcgtcgtggccctgggcctgtatttactccaaagccagatataatccatgctggcggtggggttcatagaccttggaatgtaggagcaagcagtttaaaggtcgtagggccagataataggctttgctctaattttggtactagttttgctgctccaattgtggcaagtttagctgcgcatacatggcagagaatagccactaatacagactttaatgtttcaccatcattgattaaagcattattaattcattccgctcaattatcttctcctgattactcgccaagtgaaagacgctatttgggagcgggaattcctaatgaagttattgagaccttatatgatagtgatgataggtttactctgattttccaaacattcttggttcctggggtgaggtggagaaaggataactatcccataccatcggcacttattcaaaatggaaaatttaaaggtgagattgtaattactgctgcatatgcaccaccactgaaccctaatgccggcagtgaatatgttcgcgcgaacgtagagctaagttttggcttaattgagaataatactataaaaggaaaagtgcctatggaaggagaaaacggtcaatctggatatgagagagctcaaattgagcatggtggaaagtggtcaccagtaaaaattcatcgcaaggcatttaataaaggaattacttcgggtaactgggctcttcaggctaaaacaacgttgagagcgaatgaaccggccttaatggagcctttacctgtaactattgtagtaactttaaaatcattagatggaaacacacaagtttatgctgatggcgtaagagctttaaatgctaataactgggctcactatccattgcctgctcgtgtgccagtttccgtataacaactatataaatcaaacccgctgtagcgggtttgatttatttgtgggtgtgttttataaaaataccgcccatacacaacaaaatacaa (SEQ ID NO: 300)
65 66 67 68	29	cgtgattcagttcgccagactgcagcgttttccatgaatataactccatctggtttagaaagagttccaatctaacgatattgggaccagaatcacaggcggcagtggctttacgcttacaataactattctatcctgacaattttaagcctcgtttgttacgatgtaaccctataactatgtggttcctcaacctttttgcccaaaaaatgcccaatgaagtccaaagtggaaaacagatggttatccgttgatgagattgcagattacctcgcgattaagcgagacacggtatacaagtagatcgcaaagaaaggtatacctgcacacatgattggacgcctttggaaatttaaaaaggatgaagtagatggctggatacgcgatggcaaagctggcgaaaacagtaatcaagaataaaaaagcaaatttaggagcagtttaatgaaaaccgtacgtagtgcatgccagttgcaaccgaaggccttggaaatcaatgtcggcgaccagattgaacagcttgatcaaatcatcaacgacaccaatggccaagagtactttaaaaagaccttcatcactgacggttttaaaactttgctctccaagggtatggcacgcttagccggtaaatcaaacgatactgttttccacctgaagcaagctatgggtggtggtaaaacccacttgatggtcggctttggtttattagcaaaagatgctgccctctcgaaatagccacttaggatcaatgccataccaatcagattttggctcagccaaaatagcagcattcaatggacgcaataatcctcattcctatttctggggtgagatcgctcggcagctaggtcgagagggtgtattcagggagtactgggaatccggagccaaagctcccgatgaacaagcatggataaatattttttgatggtgaggaacccatcctaatcttgttggatgaaatgccaccatacttccactactacagcacccaagtccttgggcaaggaactatagctgatgtagtgacacgggctttttccaatatgttgaccgcagcgcagaagaaaaagaatgtatgtattgtagtttccgatcttgaggcagcttacgatacaggaggcaaactgattcagcgtgcattggatgatgctacgcaagaactcggacgcgccgaggtatccattacgccggtaaacctcgaatccaatgaaatctacgagattctgcgtaaacgtttgtttttgtctctgcc

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.	Cloned Sequence
	agacaaaaatgaggtctctgaaattgcgtcgatctatgcatcaagacttgcggaagccgctaaagc caaaaccgtagagcgcagtgcagaagcattggcaaatgacatcgaatctacttacccattccaccc aagctttaaaagcatcgttgctttgttcaaagaaaacgaaaagttcaaacaaacccgtggtttgat ggagttggtttctagactgcttaaatcggtgtgggaaagcgatgaagaggtgtatttgatcggtgc ccaacactttgatctttcgatacacgatgttcgtgagaagctggctgaaatttcagaaatgcgcga tgttatcgcaagagatctttgggactccaccgacagcgctcatgctcagatcattgacctcaataa cggcaaccactatgcacaacaggttggtacgctattgctaacagccagcctctccaccgcagtgaa ctcagttaagggcttaaccgagagcgaaatgctggaatgtttgattgatcctaaccatcagggtag tgactaccgaaacgcattcactgaacttgctaaatcagcttggtatttgcatcaaacacaagaagg gcgcaattacttcagtcaccaagaaaatctcaccaaaaagcttcagggatatgccgacaaagcacc tcaaaataaggttgatgaattaattcgtcaccgactagaggaaatgtatagaccagtcacgaaaga agcatacgaaaaagtactaccactccctgaaatggatgaagcacaggccacactgaggagtggtcg tgccctgttaataatcagcccagatggcaaaacaccacctggtgtagtcggcaacttctttaaggg cttggtaaacaaaaacaacattctggtattaacgggcgataaatcctctattgccagtatagaaaa ggctgcacgccatgtttatgctgttaccaaggcagacaacgaaattacagcatcacatccgcagcg caaagagttggatgagaagaaagcacagtatgagcaggacttccaaactacagtgctctctgtatt cgataagctcctgttccccggtaacaatcgaggtgaagacgttttacggcctaaagcgctggatag cacctatccatccaacgaaccatacaacggtgaacgccaagtcgtgaagactctcacgtccgaccc catcaagctttacacccagattaacgaaaatttcgacgcactgagagcccgagcagagtcattgct gttcggtactttggatgaggcaagaaagacagatttgctcgataagatgaagcaaaaaacacagat gccttggttgccaagccgtggcttcgatcaactcgctatcgaggcataccagcgaggtgtatggga ggatttaggcaatggctatattacgaaaaagcccaagccaaaaaccactgaggtaatcatcagcga ggactcatcaccggatgatgccggcaccgttcgtcttaaaatcggcgtggctaatgcaggtaacag cccacgcattcattatgctgaagatgacgaagttaccgaaagcagcccagtacttagtgataacac gctagcaaccaaagcattgcgagtgcagtttttggcagtagaccctaccggtaaaaaccttactgg aaacccaaccacctggaaaaatcgactgacattacgcaatcgctttgacgaagtggcgagaacagt cgaattgttcgttgcccccccgtggcacaatcaagtacaccctagatggttcagaagcacgtaatgg tgaaacctacaccgtgccaatccagctcgctgatcaggaagccactatctatgtctttgctgaatg tgatggcttagaagagaagcgaaatttcacctttgcggcagcaggttctaaagaaataccgatcat aaaagataagcccgccactctggtcagcccctcacccaaacgtatggatagctcggcaaaaaccta cgagggtttgaaaatcgccaaagagaaaggcattgagttcgagcagattagcttaatggttggatc tgcaccaaaggtgattcatatatcgctaggtgagatgaaaatcagcgccgaattcattgaaaccgt attaacgcacttgcaaaccgtgttaagtccagaagcccctgtggtcatgaccttcaaaaaagccta cacacagactgggcatgatcttgagcaatttgttaagcagcttggcattgaaatcggtaatggcga ggtggaacaacgatgaataaaaccgttgattttggggcaccgtcagaattcggtatgcatcacttc tatgtggagattcccgcagcgccccgtgacgctgttgtgatctatgaagactatggctttgacggt gaagattctcgccgagaaacagtagagtgtcgcctgatattagccagagagctctggactaagatc cgcgatgacgttcgccgtgactttaacgctcgcctaaagattaagaaacaaagctccggtacttgg tctaccggtaaagtgaagcttgaccgctttcttggacgtgagttgtgcgttcttggctgggcagca gaacatgcctcacccgatgaatgtctggttatttgccaaaagtggctggctttacgcccagaagaa agatggtggctttacagtaaaaccgcagctgaagcaggtcgtgatgatcaaacacaacgaggctgg cgtaaagcgctctattgcgcgctatcggatggagccaatatcaaattggaaaccaaaaagaagccc aagtctaaaaagctacaagttgaagatgagacccaggatctgtttgggtttatggaaaagggagag ttttgatggccttgcaaccgtttgaatggagagacaaaccgtctcttattgagcacctgttcccgg tacaaaaaatatctgccgagacctttaaagaacgaatggcaagccacggtcagttgctggtgtcgt tgggtgctttttggaaaggcagaaaacctctcatcttaaacaaagcgtgcattctgggctcattgt taccagcaactgacaacccgcttgaagatttagaggtatttgagctgttaatgggcatcgactctg agtcaatgcaaaagagaattgaggcttcactaccagcatcaaaacaagaaacaatcggcgattact tggtattaccctatgccgaacaaatcaggattgctaagcgcccggaagaaattgatgaatctcttt tcgtccatatttggaatcgggtcaacaatcatcttggtacttctgctcacacttttgcgcaactag ttgaggaactaggtgttgcacggtttggccataggccaagagtggcagatgtattttctggttcgg gtcaaattccgtttgaggctgctcgcttaggttgcgatgtctatgcctctgacttaaacccgatct cctgcatgcttacttggggcgctttgaacgttgttggtgcgagcgcgcaaaaaagagtagaaatag acaaagcccaacgggatatcgttaagaaagttcaaaaagagattgatgagcttgacattgagtccg atggccgaggatggcgagcaaaggtattcctatactgcgttgaggtgacctgccctgaatccggtt ggcgtgtgcctttaattccaagtttgattatcagcaatagttttcgagttgttgctgagcttaagc ccgttcctgctgagaggcgatatgatattagtatccgtgaagtatcgactgatgaggaactggagt tctataaatcaggcaccatacaagatggcgaggtaattcactcgccagatggaaaaactcagtatc gcgttaatatcaaaacaattcgcggtgactataaagaaggcaaggagaacctaaacaagctgcgaa tgtgggagaaaacagactttgctcctcgtcctgacgatattttcaggatagattattttgcgttc aatggatgaaaaaaaaacctaaaggatcgcagtattactacgaatttcgtactgtaaccaatgacg acttaaaacgcgaaaaaaaggtaatagaacatgtcgcatccaaattagatgactggcagaagcaag gtcttgttcctgatatggttattgaagcgggcgataaaacggatgagccaatcaggacgcgaggct ggactcattggcaccatttattccatccaaggcagttgctatttttgagcttggtgaacaaatatt cactcgcagaaggaaaatttaacttcttgcagtgcatgaatcacttgtccaagctaactcgctggc gaccccaggccggtggtggtggcggttctgcggctacatttgataatcaggcgctcaatactctgt acaactacccagttagagcaacaggatctatcgaaaatatcttggctgctcagcacaaccactgtg gaatcagcgagaatgtttcctttgtggttaattcacatccagcgccagagttagatgtggaaaacg acatttatattactgatccccccatatggcgatgctgtcaagtatgaagaaatcacagagttcttta ttgcctggctgaggaaaaatccgccgaaggaatttgcccactggacttgggatagtcgccgatctc ttgcggtaaaaggagaagatgagggtttccgtacaggcatggttgctgcttatcgcaagatggcgc agaagatgccagacaatggtttacaggtgctaatgtttacccatcaaagtggcgctatctgggcag acatggctaatatcatttgggcgagcggccttcaagttactgccgcatggtacgtagttactgaaa

TABLE 15-B-continued

Sequences of validated defense systems
(Cloned sequences corresponding to row No. 1-68 in Table 15-A)

Row No.	Cloned Sequence
	ctgactctgcattacgtggtggttctaacgtaaaaggcaccatcatcctcattttacgcaagcgcc atcaggcattagagaccttccgcgatgatttaggttgggaaatcgaagaagccgttaaagagcaag tcgaatcgttaatcggattggataagaaggttcgttcccaaggcgcggaaggcctctacaccgacg ctgacctgcaaatggctggttacgcagccgcgttgaaagtactgacagcttattcccgtatcgacg gtaaagacatggtgactgaagccgaggcaccacgccaaaaaggcaaaaaaactttttgttgatgagt taattgatttcgccgtgcaaacggcagttcagtttttggtgccggttggcttcgagaaaagcgaat ggcagaagcttcaagcggttgaacgcttctatctgaaaatggccgaaatggaacaccagggtgcaa aaaccttggataactatcagaacttcgccaaggcgttcaaggttcaccattttgatcaattgatga gtgatgcctcaaaggctaactctgctcggctaaagctttctaccgagttcagaagtaccatgatgt caggtgatgccgaaatgactggcactcctctgcgagcccttctttatgccttatttgagatatcga aagaagttgaagtagacgatgttcttttgcatctcatggaaaactgcccgaattacctgcccaata agcaactgcttgccaaaatggcggattacctggctgaaaagcgtgaaggtctaaaaggtaccaaaa cgttcaaccctgagcaggaagcaagcagcgcgcgtgtccttgcggaagccattcgaaaccagaggt tgtaatctatggcgattaagcgcttttcatcccgcacagaaagattagatacggaattcctcgctg aatcgttgaaaggggctgctaagtatttccggattgcgggttatttcaggagctccatctttgagc ttgtaggcgaagagattgcaaagattccagaagttaagatcatctgtaattccgagcttgatctgg ctgacttccaggtagctactggccggaatacagcactcaaagagcgctggaatgaagtggatgtag aagctgaagcgctactgaaaaaggagcgctaccagattttggatcagctattacattcgggtaatg ttgagattcgcgtagtccctagggagcggttattccttcacggcaaagcaggctcaattcattatg cagatggcagccgtaaatcttttattggctcagtgaatgaatctaaaagcgcattcgctcacaatt atgagcttgtttggcaagacgatgatgaagaaagtgcggactgggtagaaagagaattttgggcac tctggactgaaggcgtcccgctgcctgatgcgatcttagctgaaatccaccgtgtatctaatcgcc gggaagtaaccgttgatgtattgaaaccagaggaagtcccagcggcggccatggcagaagcaccta tctaccgtggaggggagcagttacagccctggcaacgctcgtttgtgactatgtttctggaacata gggagatctatggcaaggctcgcctactattggctgacgaggtgggtgttggtaaaacgctatcaa tggcaaccagtgcattagtcagtgctttactagacgatggacctgttttgattctggcaccttcta cactcacgattcagtggcaaattgagatgatggacaagctcggtgtgcctgctgcggtttggtcct cgcagaagaaagtttggctgggtgtagaggggcaaatactctcacctcgaggtgatgcctcctcta tcaaaaaatgcccttatcgaattgccattatctctaccggactgattatgcatcagcgggagaaga ctgactttgttaaagaagctggaatgcttctgaagaatcgtttcggtaccgttattctggatgagg cgcataaagcccgtattcgtggaggattaggagatcaagcttcagaacctaataatctcatggcct tcatgctgcagatcggcaggcgtacacggcatctggtactgggtactgcgacacctattcaaacca acgtacgtgagttatgggatttattgggtattttgaactctggtgctgaatttgtactaggcgatg ctctgtcgccatggcatgaccatgaacaagcgattccgttgataaccggccagactcaggtgacat ctgaggctgaagtttggcattggttaagcaacccccctgccgccaagcaatgagcaccatactgttc agcaaattcgtgactacctgtccattgataataagtcctttggatattctcatcgtttcgaagatc tcgactatatgattcagagtctttggctctccgaatgcatgacacctagcttctttaaagagaaca accctatcctacgccatacagtgctgcgtaagcgtaaacagctggaagatgacggtctgttagagc gtgaggggtgaatacacatcccattaagcgcaacctagctcagtatcagtcgcggtttgtggggct tggcattccgaccaatacaccattccaggtcgcttacgaaaaagcggaagagttcagtaagttgct tcagtcacgcactcgagccgcaggcttcatgaaatctttgatgttgcaacggatctgctcaagttt cgcatcaggcttaaaaactgctcaaaagatgttgaaacatacggtttctgacgaagacgaggatct agttgaagatgttgagcacttactttcagaaatgactcctgcggaggtcgcttgtttaagagagat tgaaacacaactgtcacgccccgaagccgttgactcaaaactgaacacagtgaaatggttcttaac ggaattccgtaccgatggaaaaacttggctggaacacggctgtattattttcagccagtattacga cacggcggagtggatagcgaaagaactggccaagtccttaaaaggcgaagtggtagccgtttatgc tggcgttggtaaaagcggcttattcaggggcgaacagtttaataacgttgaacgcgaattgattaa atccgcagtgaagacgcgcgagattctattagtggttgctacggatgccgcctgtgaaggcttaaa cctgcaaaccttgggaacactcatcaatgtcgaccttccctggaacccatctcgtttagagcagcg cctcgggcgaatcaaacgttttggtcagacacgtaagtttgtggatatgctcaatcttgtgtacag cgaaacacaagacgagaaagtttataacgtgctgtcggaacgcttacgcgatacatacgacatttt cggcagccttcccgatacgattgatgatgaatggatcgacaacgaggaagaactcaacactcgcat ggatgaatacatgcatgaacgaaagaaagctcaagatgcgttctccgttaagtatcgcggtactct cgatcctgatgctcatctctgggaacgttgcgctacagtactgtcacgtagggacattgtaagtaa gctcagcgaaccatggggaagctaattatgttgtgatgtggatgccccgctcagccaaggtcctgc acaactatgttggatgctctttttttagagggctacatcatgaattcgatcaaagttattggtacaa ttctgagtaaatctgtctctcagggtatccatttcgagtg (SEQ ID NO: 301)

TABLE 15-C

Sequences of validated defense systems (Sequences encoded by the genes corresponding to rows 1-68 of Table 15-A)

Row No.	Sequence
1	MIKNDKAWIGDLLGGPLMSRESRVIAELLLTDPDEQTWQEQIVGHNILQASSPNTAKRYAATIRLRLNTLDKSAWTLIAEG SERERQQLLFVALMLHSPVVKDFLAEVVNDLRRQFKEKLPGNSWNEFVNSQVRLHPVLASYSDSSIAKMGNNLVKALAE AGYVDTPRRRNLQAVYLLPETQAVLQRLGQQDLISILEGKR* (SEQ ID NO: 302)

TABLE 15-C-continued

Sequences of validated defense systems (Sequences encoded by the genes corresponding to rows 1-68 of Table 15-A)

Row No.	Sequence
2	MIDPVLEYRLSQIQSRINEDRFLKNNGSGNEIGFWIFDYPAQCELQVREHLKYLLRHLEKDHKFACLNVFQIIIDMLNERGLFERVCQQEVKVGTETLKKQLAGPLNQKKIADFIAKKVDLAAQDFVILTGMGNAWPLVRGHELMSALQDVMGFTPLLMFYPGTYSGYNLSPLTDTGSQNYYRAFRLVPDTGPAATLNPQ* (SEQ ID NO: 303)
3	MNIEQIFEKPLKRNINGVVKAEQTDDASAYIELDEYVITRELENHLRHFFESYVPATGPERIRMENKIGVWVSGFFGSGKSHFIKILSYLLSNRKVTHNGTERNAYSFFEDKIKDALFLADINKAVHYPTEVILFNIDSRANVDDKEDAILKVFLKVFNERIGYCADFPHIAHLERELDKRGQYETFKAAFADINGSRWEDERDAYYFISDDMAQALSQATQQSLESSRQWVEQLDKNFPLDINNFCQWVKEWLDDNGKNILFMVDEVGQFIGKNTQMMLKLQTITENLGVICGGRAWVIVTSQADINAAIGGMSSRDGQDFSKIQGRFSTRLQLSSSNTSEVIQKRLLVKTDEAKAALAKVWQEKADILRNQLAFDTTTTTALRPFTSEEEFVDNYPFVPWHYQILQKVFESIRTKGAAGKQLAMGERSQLEAFQTAAQQISAQGLDSLVPFWRFYAAIESFLEPAVSRTITQACQNGILDEFDGNLLKTLFLIRYVETLKSTLDNLVTLSIDRIDADKVELRRRVEKSLNTLERLMLIARVEDKYVFLTNEEKEIENEIRNVDVDFSAINKKLASIIFDDILKSRKYRYPANKQDFDISRFLNGHPLDGAVLNDLVVKILTPKDPTYSFYNSDATCRPYTSEGDGCILIRLPEEGRTWSDIDLVVQTEKFLKDNAGQRPEQATLLSEKARENSNREKLLRVQLESLLAEADVWAIGERLPKKSSTPSNIVDEACRYVIENTFGKLKMLRPFNGDISREIHALLTVENDTELDLGNLEESNPDAMREVETWISMNIEYNKPVYLRDILNHFARRPYGWPEDEVKLLVARLACKGKFSFSQQNNNVERKQAWELFNNSRRHSELRLHKVRRHDEAQVRKAAQTMADIAQQPFNEREEPALVEHIRQVFEEWKQELNVFRAKAEGGNNPGKNEIESGLRLLNAILNEKEDFALIEKVSSLKDELLDFSEDREDLVDFYRKQFATWQKLGAALNGSFKSNRSALEKDAAAVKALGELESIWQMPEPYKHLNRITPLIEQVQNVNHQLVEQHRQHALERIDARIEESRQRLLEAHATSELQNSVLLPMQKARKRAEVSQSIPEILAEQQETKALQMDADKKINLWIDELRKKQEAQLRAANEAKRAADSEQTYVVVEKTVIQPVPKKTHLVNVASEMRNATGGEVLETTEQVEKALDTLRTTLLAVIKAGDRIRLQ* (SEQ ID NO: 304)
4	MNTNNIKKYAPQARNDFRDAVIQKLTTLGIAADKKGNLQIAEAETIGETVRYGQFDYPLSTLPRRERLVKRAREQGFEVLVEHCAYTWFNRLCAIRYMELHGYLEHGFRMLSHPETPTAFEVLDHVPEVAEALLPENKAQLVEMKLSGNQDEALYRELLLGQCHALHHAMPFLFEAVDDEAELLLPDNLTRTDSILRGLVDDIPEEDWEQVEVIGWLYQFYISEKKDAVIGKVVKSEDIPAATQLFTPNWIVQYLVQNSVGRQWLQTYPDSPLKDKMEYYIEPAEQTPEVQAQLAAITPASIEPESIKVLDPACGSGHILIEAYNVLKNIYEERGYRGRDIPQLILENNIFGLDIDDRAAQLSGFALLMMARQDDRRIFTRDVRLNIVSLQESLHLDIAKLWQQLNFHQQVQTGSMGDMFAENNALTQTDSAEYQLLMRTLKRFVNAKTLGSLIQVPQEEEAELKVFLDALYRLEQEGDFQQKTAAKAFIPFIQQAWILAQRYDAVVANPPYMGGNYMETELKNFVSSYYPQGKADLYSSFMVRLLLQLKDNRTLSLMTPFTWMNLSSFEELRKIILTNFSIQSLVQPEYHSFFESAYVPICAFSISNTPLSWNAKFFDLSDFYGEKNQAPNFQYAIKNDNKCHWKYNRITTDFLCTPGYIIAYSLPDSALSCFKTSKKLHDVCNLKQGLITGDNERYLRFWHEISYNSFSLNEKRKKTKWFPYQKGGAYRKWYGNNDYVVDWENDGYSIKNFYNDKGKLRSRPQNIQFYCKEGLTWTSLTISSLSMRYVPNGYIFDAKGPMCFPKSSLDIWNILGYANSKVIDIFLKQLAPTMDYSQGPVGNVPFKFNDGDLNEIIKELVNIHKRDWDENETSFEFKRDMLVHFSRDINTIKGSFTLRQGENKKAINRTKFLEEMNNSFFINCFNLTDILSPEIELNKITLTHATIEIDIQKIISYAIGCQMGRYSLDREGLVYAHEGNNGFADLVAEGAYKSFPADSDGILPLMDEEWFDDDVTSRVKEFIRTVWGEEYLRENLDFIAEVLKPKKGESALETIRRYLSTQFWKDHLKMYKKRPIYWLFSSGKEKAFECLVYLHRYNDATLSRMRTEYVVPLLARYQANIDRLNDQLDEASGGESTRLKRERDSLIKKFSELRSYDDRLRHYADMRISIDLDDGVKVNYGKFGDLLADVKAITGNAPEVI* (SEQ ID NO: 305)
5	MQNQDFIAGLKAKFAEHRIVFWHDPDKRFIEELEQLKLESVTLINMTHESQLAVKKRIEIDEPEQQFLLWFPHDAPPHEQDWLLDIRLYSSEFHADFAAITLNTLGIPQLGLREHIQRRKAFFSTKRTQALKNLATEQEDEASLDKKMIAVIAGAKTAKTEDILFNLITQYVNQQIEDDSELENTQAMLKRHGLDSVLWEMLNHEMGYQAEEPSLENLLLKLFCTDLSAQADPQQRAWLEKNVLLTPSGRASALAFMVTWRADRRYKEAYDYCAQQMQAALHPEDHYRLSSPYDLHECETTLSIEQTIIHALVTQLLEESTTLDREAFKKLLSERQSKYWCQTQPEYYAIYDALRQAERLLNLRNRHIDGFHYQDSATFWKAYCEELFRFDQAYRLFNEYALLVHSKGAMILKSLDDYIEALYSNWYLAELSRNWNEVLEAENRMQAWQIPGVPRQQNFFNEVVKPQFQNPQIKRVFVIISDALRYEVAEELGNQINTEKRFTAELRSQLGVLPSYTQLGMAALLPHEQLCYQPGNGDIVYADGLSTSGIPNRDTILKNYKGMAIKSKDLLELKNQEGRDLIRDYEVVYIWHNTIDATGDTASTEDKTFEACRTAVAELKDLVTKVINRLHGTRIFVTADHGFLFQQQALSVQDKTTLQIKPENTIKNHKRFIIGHQLPADDFCWKGKVADTAGVSDNSEFLIPKGIQRFHFSGGARFVHGGTMLQEVCVPVLQIKALQKTAAEKQPQRRPVDIVAYHPMIKLVNNIDKVSLLQTHPVGELYEPRILNIYIVDNANNVVSGKERISFDSDNNTMEKRVREVTLKLIGANFNRRNEYWLILEDAQTETGYQKYPVIIDLAFQDDFF* (SEQ ID NO: 306)
6	MQTHHDLPVSGVSAGEIASEGYDLDALLNQHFAGRVVRKDLTKQLKEGANVPVYVLEYLLGMYCASDDDDVVEQGLQNVKRILADNYVRPDEAEKVKSLIRERGSYKIIDKVSVKLNQKKDVYEAQLSNLGIKDALVPSQMVKDNEKLLTGGIWCMITVNYFFEEGQKTSPFSLMTLKPIQMPNMDMEEVFDARKHFNRDQWIDVLLRSVGMEPANIEQRTKWHLITRMIPFVENNYNVCELGPRGTGKSHVYKECSPNSLLVSGGQTTVANLFYNMASRQIGLVGMWDVVAFDEVAGITFKDKDGVQIMKDYMASGSFSRGRDSIEGKASMVFVGNINQSVETLVKTSHLLAPFPTAMIDTAFFDRFHAYIPGWEIPKMRPEFFTNRYGLITDYLAEYMREMRKRSFSDAIDKFFKLGNNLNQRDVIAVRRTVSGLLKLMHPDGAYSKEDVRVCLTYAMEVRRRVKEQLKKLGGLEFFDVNFSYIDNETLEEFFVSVPEQGGSELIPAGMPKPGVVHLVTQAESGMTGLYRFETQMTAGNGKHSVSGLGSNTSAKEAIRVGFDYFKGNLNRVSAAAKFSDHEYHLHVVELHNTGPSTATSLAALIALCSILLAKPVQEQMVVLGSMTLGGVINPVQDLAASLQLAFDSGAKRVLLPMSSAMDIPTVPAELFTKFQVSFYSDPVDAVYKALGVN* (SEQ ID NO: 307)
7	MHKYPSIIVNINLREAKLKKKVREHLQSLGFTRSDSGALQAPGNTKDVIRALHSSQRAERIFANQKFITLRAAKLIKFFASGNEVIPDKISPVLERVKSGTWQGDLFRLAALTWSVPVSSGFGRRLRYLVWDESNGKLIGLIAIGDPVFNLAVRDNLIGWDTHARSSRLVNLMDAYVLGALPPYNALLGGKLIACLLRSRDLYDDFAKVYGDTVGVISQKKKQARLLAITTTSSMGRSSVYNRLKLDGIQYLKSIGYTGGWGHFHIPDSLFIELRDYLRDMDHAYADHYMFGNGPNWRLRTTKAALNALGFRDNLMKHGIQREVFISQLAENATSILQTGKGEPDLTSLLSAKEIAECAMARWMVPRSIRNPEYRLWKARDLFDFISNDSLNFPPFDEIAKTVV* (SEQ ID NO: 308)
8	MNYAIDKFTGTLILAARATKYAQYVCPVCKKGVNLRKGKVIPPYFAHLPGHGTSDCENFVPGNSIIVETIKTISKRYMDLRLLIPVGSNSREWSLELVLPTCNLCRAKITLDVGGRSQTLDMRSMVKSRQIGAELSVKSYRIVSYSGEPDPKFVTEVERECPGLPSEGAAVFTALGRGASKGFPRAQELRCTETFAFLWRHPVAPDFPDELEIKSLASKQGWNLALVTIPEVPSVESISWLKSFTY

TABLE 15-C-continued

Sequences of validated defense systems (Sequences encoded by the genes corresponding to rows 1-68 of Table 15-A)

Row No.	Sequence
	LPVVPARTSITAIWPFLNQKTSINHVECVYSDTILLSTNMAPTSSENVGPTMYAQGSSLLLSAVGVETSPAFFILNPGENDFVGVSGSIEQDVNLFFSFYKKNVSVPRKYPSIDLVFTKRNKEKTIVSLHQRRCIEVMMEARMFGHKLEYMSMPSGVEGVARIQRQTESNVIKLVSNDDIAAHDKSMRLLSPVALSQLSDCLANLTCHVEIDFLGLGKIFLPGSSMLSLDDGKFIELSPNLRSRILSFILQMGHTLHGFSLNNDFLLVEKLVDLQPEPHLLPHYRALVKEVKTNGFECNRFR* (SEQ ID NO: 309)
9	MSYQYSQEAKERISKLGQSEIVNFINEISPTLRRKAFGCLPKVPGFRAGHPTEIKEKQKRLIGYMFQSHPSSEERKAWKSFSLFWQFWAEEKIDKSFSMIDNLGLKENSGSIFIRELAKNFPKVARENIERLFIFSGFADDPDVINAFNLFPPAVVLARDIVIDTLPIRLDELEARISLIADNVEKKNNHIKELELKIDAFSEQFDNYFNNEKSSLKIINELQSLINSETKQSDIANKAIDELYHFNEKNKQLILSLQEKLDFNALAMNDISEHEKLIKSMANDISEFKNALTILCDNKIKNNELDYVNELKKLTERIDTLEINTSQASEVSVTNRFTKFHEIAHYENYEYLSSSEDISNRISLNLQAVGLTKNSAEKLARLTLATFVSGQIIQFSGSLADIIADAIAIAIGAPRYHIWRVPVGIISDMDAFDFIETIAESSRCLLLKGANLSAFEIYGAAIRDIVVQRQIHPTNYDHLALIATWKQGPATFPDGGMLAELGPVIDTDTLKMRGLSATLPQLKPGCLAKDKWTNIDGLHLDSVDDYVDELRALLDEAGFDGGTLWKRMIHIFYTSLIRIPNGNYIYDLYSVLSFYTLTWAKIKGGPVQKIEDIANRELKNYSAKISS* (SEQ ID NO: 310)
10	MEWRAVSRDKALDMLSTALNCRFDDEGLRISAVSECLRSVLYQYSISETEEARQTVTSLRLTSAVRRKLVPLWPDIADIDNAIHPGIMSILNSLAELGDMIKLEGGNWLTAPPHAVRIDNKMAVFFGGEPSCTFSTGVVAKSAGRVRLVEEKVCTGSVEIWDANEWIGAPAEGNEEWSSRLLSGTISGFIDAPGNMSETTAYVRGKWLHLSELSFNKKQIYLCRMSVDNHFSYYLGEIEAGRLCRMNSLESSDDVRRLRFFLDTKDNCPLKVRIKISNGLARLRLTRRLPRRETKVLLLGWRESGFENEHSGITHHVFPEEILPIVRSAFEGLGIIWINEFTRRNEI* (SEQ ID NO: 311)
11	MINKNKVTERSGIHDTVKSLSENLRKYIEAQYHIRDEGLIAERRALLQQNETIAQAPYIEATPIYEPGAPYSELPIPEAASNVLTQLSELGIGLYQRPYKHQSQALESFLGENASDLVIATGTGSGKTESFLMPIIGKLAIESSERPKSASLPGCRAILLYPMNALVNDQLARIRRLFGDSEASKILRSGRCAPVRFGAYTGRTPYPGRRSSRRDELFIKPLFDEFYNKLANNAPVRAELNRIGRWPSKDLDAFYGQSASQAKTYVSGKKTGKQFVLNNWGERLITQPEDRELMTRHEIQNRCPELLITNYSMLEYMLMRPIERNIFEQTKEWLKADEMNELILVLDEAHMYRGAGGAEVALLIRRLCARLDIPRERMRCILTSASLGSIEDGERFAQDLTGLSPTSSRKFRIIEGTRESRPESQIVTSKEANALAEFDLNSFQCVAEDLESAYAAIESLAERMGWQKPMIKDHSTLRNWLFDNLTGFGPIETLIEIVSGKAVKLNILSENLFPDSPQQIAERATDALLALGCYAQRASDGRVLIPTRMHLFYRGLPGLYACIDPDCNQRLGNHSGPTILGRLYTKPLDQCKCASKGRVYELFTHRDCGAAFIRGYVSSEMDFVWHQPNGPLSEDEDIDLVPIDILVEETPHVHSDYQDRWLHIATGRLSKQCQDEDSGYRKVFIPDRVKSGSEITFDECPVCMRKTRSAQNEPSKIMDHVTKGEAPFTTLVRTQISHQPASRPIDGKHPNGGKKVLIFSDGRQKAARLARDIPRDIELDLFRQSIALACSKLKDINREPKPTSVLYLAFLSVLSEHDLLIFDGEDSRKVVMARDEFYRDYNSDLAQAFDDSFSPQESPSRYKIALLKLLCSNYYSLSGTTVGFVEPSQLKSKKMWEDVQSKKLNIESKDVHALAVAWIDTLLTEFAFDESIDSTLRIKAAGFYKPTWGSQGRFGKALRKTLIQYPAMGELYVEVLEEIFRTHLTLGKDGVYFLAPNALRLKIDLLHVWKQCNDCTALMPFALEHSTCLACGSNSVKTVEPSESSYINARKGFWRSPVEEVLVSNSRLLNLSVEEHTAQLSHRDRASVHATTELYELRFQDVLINDNDKPIDVLSCTTTMEVGVDIGSLVAVALRNVPPQRENYQQRAGRAGRRGASVSTVVTYSQNGPHDSYYFLNPERIVAGSPRTPEVKVNNPKIARRHVHSFLVQTFFHELMEQGIYNPAEKTAILEKALGTTRDFFHGAKDTGLNLDSFNNWVKNRILSTNGDLRTSVAAWLPPVLETGGLSASDWFAKVAEEFLNTLHGLAEIVPQTAVLVDEENEDDEQTSGGMKFAQEELLEFLFYHGLLPSYAFPTSLCSFLVEKIVKNIRGSFEVRTVQQPQQSISQALSEYAPGRLIVIDRKTYRSGGVFSNALKGELNRARKLFNNPKKFIHCDKCSFVRDPHNNQNSENTCPICGGILKVEIMIQPEVFGPENAKELNEDDREQEITYVTAAQYPQPVDPEDFKFNNGGAHIVFTHAIDQKLVTVNRGKNEGESSGFSVCCECGAASVYDSYSPAKGAHERPYKYIATKETPRLCSGEYKRVFLGHDFRTDLLLLRITVGSPLVTDTSNAIVLRMYEDALYTIAEALRLAASRHKQLDLDPAEFGSGFRILPTIEEDTQALDLFLYDTLSGGAGYAEVAAANLDDILTATLALLESCECDTSCTDCLNHFHNQHIQSRLDRKLGASLLRYALYGMVPRCASPDIQVEKLSQLRASLELDGFQCIIKGTQEAPMIVSLNDRSIAVGSYPGLIDRPDFQHDVYKSKHTNAHIAFNEYLLRSNLPQSHQNIRKMLR* (SEQ ID NO: 312)
12	MKKVYELTSEEALSYFLRHDSYTTLELPAYINFTTLLNDINSSIHNKKIKIEPTAKELMGKDINYEVLVSKDGLYSWRRITLINPLYYVYFCRKITAPATWEIITEKFKSFESNDLFTCSSIPVRKDNSSNIAASVMNWWEDFEQKSLALALEYEFMFSTDISNFYPSIYTHSFEWVFISKEEAKKKKSKNNPGGLIDSHIQMMMNNQTNGIPLGSTLMDTFAELILGQIDIELRKKTNELKIINYKVVRYRDDYRIFSNSKDDLDIISKCLVNVLGDFGLDLNSKKTELYEDIILHSLKQAKKDYIKEKRHKSLQKMLYSIYLFSLKHPNSKTTVRYLNDFLRNLFKRKTIKDNGQQVDAMLGIISSIMAKNPTTYPVGTAIFSKLLSFLYGDDTQKKLTKLEQLHKKLDKQPNTEMLDIWFQRTQAKINLEWNKSYKSALCVRINDELTKEKTFSVNNLWNIDWIQGKETSPNKAKILSLLRKTKIVDTDKFDKMDDNITPEEVNLFFKEHSN* (SEQ ID NO: 313)
13	MSLHDKLLMHNFALANKKSPDFISELPQIEPKPYSNGHKIKWINHTLTSTEVTPPDNLIKICILIESGEIAITSVSDIANLLGYPAGQLLYILYRKKDNYRTFEIEKKNGKKRVINAPCGGLSILQTRLKPVLEYFYRPKKSAHGFIKGKSIITNAGMHIKKNFVVNIDLENYFESISFARVYGIFKSKPFNFAHPAATVLAQLCTHNGKLPQGACTSPILANIASASLDKQLTQFAGRKKISYSRYADDITFSFNQRNIDIIKKNDDGSYSLSETIDNIISKNGFKINYDKFRVQTRNTRQSVTGLVVNDKVNINRRYIRITRSMIHRWTDDKLKYALLFATEKGYQAKDNNHAIQIFRNHIYGRLSFIKMVRGKDYPGYLKLMSYMSHNDPLKTQEGLRAMKETENFDVFICHASEDKKDIAIPIYDELTKLKISAFIDHVEIKWGDSLIDKINAALVKSKYVIAILSANSVNKEWPQKELRAVLASEISSGDVKLLTLLKKEDEEVVNLSLPLLSDKFYMVYDNNPEVVANNIKSLLQR* (SEQ ID NO: 314)
14	MTKTSKLDALRAATSREDLAKILDVKLVFLTNVLYRIGSDNQYTQFTIPKKGKGVRTISAPTDRLKDIQRRICDLLSDCRDEIFAIRKISNNYSFGFERGKSIILNAYKHRGKQIILNIDLKDFFESFNFGRVRGYFLSNQDFLLNPVVATTLAKAACYNGTLPQGSPCSPIISNLICNIMDMRLAKLAKKYGCTYSRYADDITISTNKNTFPLEMATVQPEGVVLGKVLVKEIENSGFEINDSKTRLTYKTSRQEVTGLTVNRIVNIDRCYYKKTRALAHALYRTGEYKVPDENGVLVSGGLDKLEGMFGFIDQVDKFNNIKKKLNKQPDRYVLTNATLHGFKLKLNAREKAYSKFIYYKFFHGNTCPTIITEGKTDRIYLKAALHSLETSYPELFREKTDSKKKEINLNIFKSNEKTKYFLDLSGGTADLKKFVERYKNNYASYYGSVPKQPVIMVLDNDTGPSDLLNFLRNKVKSCPDDVTEMRKMKYIHVFYNLYIVLTPLSPSGEQTSMEDLFPKDILDIKIDGKKFNKNNDGDSKTEYGKHIFSMRVVRDKKRKIDFKAFCCIFDAIKDIKEHYKLMLNS* (SEQ ID NO: 315)
15	MNKKFTDEQQQQLIGHLTKKGFYRGANIKITIFLCGGDVANHQSWRHQLSQFLAKFSDVDIFYPEDLFDDLLAGQGQHSLLSLENILAEAVDVIILFPESPGSFTELGAFSNNENLRRKLICIQDAKFKSKRSFINYGPVRLLRKFNSKSVLRCSSNELKEMCDS

TABLE 15-C-continued

Sequences of validated defense systems (Sequences encoded by the genes corresponding to rows 1-68 of Table 15-A)

Row No.	Sequence
	SIDVARKLRLYKKLMASIKKVRKENKVSKDIGNILYAERFLLPCIYLLDSVNYRTLCELAFKAIKQDDVLSKIIVRSVVSRLINERKILQMTDGYQVTALGASYVRSVFDRKTLDRLRLEIMNFENRRKSTFNYDKIPYAHP* (SEQ ID NO: 316)
16	MKSAEYLNTFRLRNLGLPVMNNLHDMSKATRISVETLRLLIYTADFRYRIYTVEKKGPEKRMRTIYQPSRELKALQGWVLRNILDKLSSSPFSIGFEKHQSILNNATPHIGANFILNIDLEDFFPSLTANKVFGVFHSLGYNRLISSVLTKICCYKNLLPQGAPSSPKLANLICSKLDYRIQGYAGSRGLIYTRYADDLTLSAQSMKKVVKARDFLFSIIPSEGLVINSKKTCISGPRSQRKVTGLVISQEKVGIGREKYKEIRAKIHHIFCGKSSEIEHVRGWLSFILSVDSKSHRRLITYISKLEKKYGKNPLNKAKT* (SEQ ID NO: 317)
17	MSVIRGLAAVLRQSDSDISAFLVTAPRKYKVYKIPKRTTGFRVIAQPAKGLKDIQRAFVQLYSLPVHDASMAYMKGKGIRDNAAAHAGNQYLLKADLEDFFNSITPAIFWRCIEMSSAQTPQFEPQDKLFIEKILFWQPIKRRKTKLILSVGAPSSPVISNFCMYEFDNRIHAACKKVEITYTRYADDLTFSSNIPDVLKAVPSTLEVLLKDLFGSALRLNHSKTVFSSKAHNRHVTGITINNEETLSLGRDRKRFIKHLINQYKYGLLDNEDKAYLIGLLAFASHIEPSFITRMNEKYSLELMERLRGQR* (SEQ ID NO: 318)
18	MTKQYERKAKGGNLLSAFELYQRNSDKAPGLGEMLVGEWFEMCRDYIQDGHVDESGIFRPDNAFYLRRLTLKDFRRFSLLEIKLEEDLTVIIGNNGKGKTSILYAIAKTLSWFVANILKEGGSGQRLSEMTDIKNDAEDRYSDVSSTFFFGKGLKSVPIRLSRSALGTAERRDSEVKPAKDLADIWRVINEVNTINLPTFALYNVERSQPFNRNIKDNTGRREERFDAYSQTLGGAGRFDHFVEWYIYLHKRTVSDISSSIKELEQQVNDLQRTVDGGMVSVKSLLEQMKFKLSEAIERNDAAVSSRVLTESVQKSIVEKAICSVVPSISNIWVEMITGSDLVKVTNDGHDVTIDQLSDGQRVFLSLVADLARRMVMLNPLLENPLEGRGIVLIDEIELHLHPKWQQEVILNLRSAFPNIQFIITTHSPIVLSTIEKRCIREFEPNDDGDQSFLDSPDMQTKGSENAQILEQVMNVHSTPPGIAESHWLGNFELLLLDNSGELDNHSQVLYDQIKAHFGIDSIELKKADSLIRINKMKNKLNKIRAEKGK* (SEQ ID NO: 319)
19	MRELARLERPEILDQYIAGQNDWMEIDQSAVWPKLTEMQGGFCAYCECRLNRCHIEHFRPRGKFPALTFIWNNLFGSCGDSRKSGGWSRCGIYKDNGAGAYNADDLIKPDEENPDDYLLFLTTGEVVPAIGLTGRALKKAQETIRVFNLNGDIKLFGSRRTAVQAIMPNVEYLYTLLEEFDEDDWNEMLRDELEKIESDEYKTALKHAWTFNQEFA* (SEQ ID NO: 320)
20	MKLLDKKYYNLEPKYEYLKDSFILGLAWKKTDSFVRTHNWYADILELDKCAFDISDEVTNWSNEISKNALSKSDIELIPAPKGASWFINQGKWTTNKDNRKIRPLANISIRDQSFATAVTMCLADAIETRQKDCSLSNLGYAEHVKNKVVSYGNRLVCDWDNERARFRWGGSEYYRKFSSDYRSFLQRPIYIGRETVNKVSGIDDVYIISLDLKNFFGSIKINLLLEKIKKISADHYAAKFINDNEFWTLANRILSWDWPEESLSLLESLDIKEKNVGLPQGLASAGALANAYLIEFDESLISKLRTKIEDSQIILHDYCRYVDDIRLVISGEALESNKIKESIHALVQGILDETLAQNPSDNEPYLKINDSKTYILELSDIDNGSGLTNRINEIQHEVGASSIPERNGLDNNIPALQQLLLTEQDNFSEDVDSLFPGFKNDKSIKVESVRRFSAHRLEKSLAKKSKLISPEERKQFDNETSLIAKKLLKAWLKDPSIMVIFRKAIAINPNLDAYSTILEIIFSRIQRNRDKRDKYIMLYLLSDIFRSVIDVYRNLESEYVDDYQKLMGEVTLFAQKILSCKSFIPNYAYQQALFYLAVINKPFIASNKASFDLARLQCVLIKQHLEPLNSSDGYLFEVSAQISKDYRANAAFLLSHTNSNKVVDLIIEKFAFRGGEFWNAIWKEIVRMQDKDRINEFRWAISKYESKPNSSEHYLSSVISFKENPFRYEHALLKLGVALVELFDDTEKNVWQPDGKQYSPHEIKVKLEGNSTSWGELWRPNFSISCSIDKKGEPGKDPRYISPEWLANYPQTQNDEQKIYWVCSVLRSAALGNVDYTQRNDLKLDKAKYDGIHSQFYKRRMGMLHTPESIVGSYGTITDWFASFLQHGLQWPGFSSSYISQEDILSITNIIEFKNCLLERLGYLNKQICISSNVPTLPTVVNRPELASNHFRIVTVQQLFPKDTNFHPSDVTLANPDVRWKHREHLAEICKLTEQTLNAKLKTESREHTSTADLIVFSELAVHPEDEDIVRALAFRTKAIIFSGFVFCEQDGRIVNKARWIIPDSSESGTQWRVRDQGKHHMTSDEVALGIQGYRPSQHIISIEGHPEGPFKLTGAICYDATDIKLAADLRDLTDMFVIAAYNKDVDTFDNMASALQWHMYQHIVITNTGEYGGSTMQAPYKEKYHKLISHAHGTGQIAISTADIDLAAFRRKLQIYKKTKTQPAGYNRKH* (SEQ ID NO: 321)
21	MDTLVKLATIISPLISAGVAIWAILVAKKTISESKEIAKKTIADTAYQAYLQLAMENPQFSKGYSADCRQERDPMYDQYVWYVARMIFCFEKIIEVEVNLKDSSWANTLEKHLKFHSEHFKKTNVVEEALYIPPILDLIRCAAN* (SEQ ID NO: 322)
22	MNNDDYPWFRKRGYLHFDEPVSLKKAVKYVSSPEKIIKHSFLPFLSFEVKSFKIKKDKSTKQLSKTEKLRPIAYSSHLDSHIYAFYAEYLTGHYELLIQENNLHENILAFRSLNKSNIEFAKRAFDTITEMGECSAVALDLSGFFDNLDHQILKHQWCKVIGTEALPQDHFAIYKSITRYSKVDKNRAYEILGISKNNPKYNRRKICTPVDFRNKIRKNGLIIVNNSQKGIPQGSPISALLSNIYMLDFDIEMRDYAQERGGHYYRYCDDMLFIVPTKYNKTLAGDVAQRIKHLKVELNTKKTEIRDFIYKDSTLVANMPLQYLGFIFDGSNILLRSSSLARYSERMKRGVRLAKATMDSKNRIRENKGEALKALFKKKLYARYSHIGRRNFLTYGYRAAKIMNSKAIKRQLKPLQKRLENEILK* (SEQ ID NO: 323)
23	MLNQSFSVSNLIKLLKKTDPKRYKIGRNSAEYKKYIADKVNGSIETYSFGSISNSRINNKNVYIFKDFMDVLVARKINDNIKRVYSVKQNNRHDIIKKVNTVLSEPVNYYIYRLDIKSFYESIDKNIVFQRINNNPIISHNTKKFINGLFKHNAFSANNGLPRGMGLSATLSEIFMEEFDAELARLPEVFYASRYVDDIIVFSFYKIPDYKNYFSRILPNGLHLNERKCSEYTIEDTSTKHSEIEFLGYSFIIHHGLKNQRRHVVIRISEEKIKKIKRRIALAVKDYSNNSDAELLKKRIKYLTGNILVNSNSNKTDALYSGIYYNYQHLTDKTQLKELDIFKNRMLFSSKGEVGRKILAAGHNLLTAPKKYSFLAGFEKRLLSSFKREDIIKINKVW* (SEQ ID NO: 324)
24	MKIKISKSDYKRVLLTDILPYEVPILFSNEGFYKLISENKVLPGTFSEGLKLDSYTIPYSYKIKKGLASSRSLGIIHPSTQLRICDFYDKYEHLMVHMCTKSPFSLRYPSKIGSYYYEKDFLKSRINLKDGLVQFHNHGFDSQETSSSSHFSYKKYPFIYKFYESYEFHRLERKFRKLLKLDIAKCFSHIYTHSVSWAVKSKEFSKVNRTYNSFEGCLDKLFQDANYGETNGIIIGPEFSRIFAEIILQRVDLNVESHLNLEPGIVKDKSYAIRRYVDDYFIFADDDETFKLIEFVLANELEKYKLYLNESKKEFIERPFVTGATMAKNDIAEIIEDLYGSLIHTEKLDELTAMVNLNPDVKIQPENMNDLFPLKGVWNKKLHADKFIKRIKIAVRKNNTTFDLVSSYLLSAIKSKFFKVIRLLRMFDLSGKEDITYKFFSIFNEVIFFIYAMDFRVRQTYIISQVILEINSFANKQASDISEVIKKNTFDELLMCMKSMGNIHERPVELSNLLICMKGLGEQYKLNPDEFKDLLGISENECFYDLEYFSICSMLHYIGDDVLYLKMKEDIVLAIQSLISGRNDIKKDTETFMLFLDMMTCPYLTVKHKRHYRTYVEANTGQKRFTNAVIDSEIDSLKNNVIFFNWSGDADLEHVLYKKELRTAYE* (SEQ ID NO: 325)
25	MVIFDEKRHLYEALLRHNYFPNQKGSISEIPPCFSSRTFTPEIAELISSDTSGRRSLQGYDCVEYYATRYNNFPRTLSIIHPKAYSKLAKHIHDNWEEIRFIKENENSMIKPDMHADGRIIIMNYEDAETKTIRELNDGFGRRFKVNADISGCFTNIYSHSIPWAVIG

TABLE 15-C-continued

Sequences of validated defense systems (Sequences encoded by the genes corresponding to rows 1-68 of Table 15-A)

Row No.	Sequence
	VNNAKIALNTKVKNQDKHWSDKLDYFQRQAKRNETHGVPIGPATSSIVCEIILSAVDKRLRDDGFLFRRYIDDYTCYCKTHDDAKEFLHLLGMELSKYKLSLNLHKTKITNLPGTLNDNWVSLLNVNSPTKKRFTDQDLNKLSSSEVINFLDYAVQLNTQVGGGSILKYAISLVINNLDEYTITQVYDYLLNLSWHYPMLIPYLGVLIEHVYLDDGDEYKNKFNEILSMCAENKCSDGMAWTLYFCIKNNIDIDDDVIEKIICFGDCLSLCLLDSSDIYEEKINNFVSDIIKLDYEYDIDRYWLLFYQRFFKDKAPSPYNDKCFDIMKGYGVDFMPDENYKTKAESYCHVVNNPFLEDGDEIVSFNDYMAIA* (SEQ ID NO: 326)
26	MTSTIDFYESDFSATLYPLKTNQILLKHHSQEMSEYIYQKVINPAYPTDSFLSQQKVFSTKPKGHLRRTVKLDPVAEYFIYDVIYRNRKIFRPEVSESRKSFGYIFRNGSRIPIHVSYNEYKQSLKKYSELYSHSIHFDIASYFNSLYHHDIIHWFSSKEGVSPADVEALGQFFREINSGRSIDFMPQGIYPAKMIGNEFLKFVDLHGRLKSAQIVRFMDDFTIFDNDIETLNNDFIRIQQLLGQVSLNINPSKTTFDNVMGDVNETLTQIKSSLKEIITEYEHIPTASGVEVVETNIEIIKHLDDEQVNKLIDLLKDEKIEESDADLILGFLRTHNDSLLSQMPMLLGRFPNLIKHIYTICSGITDKSGLVKILLSYLNTNNNFLEYQLFWIGAIVEDYLLGVGEYGSVLHKLYELSGDFKIARAKVLEIPEQGFGFKEIRNEYLRTGQSDWLSWSSAIGTRNLKSAERNYILDYFSKGSPINYLVASCVKKL* (SEQ ID NO: 327)
27	MTSEIVLNLDFPEYKDDFCTDSIDEQDNELWQQQANKKLLSFLEVMGEEARRYKENNSRSTHPHYKTLSSYHHAIFISGARGAGKTVFMRNARFSWQKHYNKDLKRPKLYFIDVIDPTLLNIDDRFSEVIIASIYATVEKRMKQPDIAQNIKDNFINSLKTLSGALGKSKDYDEYRGIDRIQKYRSGIHLEKYFHQFLISSVELLDCDALVLPIDDVDMKIDNAFGVLDDIRCLLSCPLVLPLVSGDNDLYRFIAKSKFEELLNRKANSNYAKEGSEIAERLSEAYITKVFPSHVKIPLQPIDELLPYLYIHSNEDENKQHTSYSEFIKLVQQKFYFLCNGQERSTNWPQPRSAREVTQLIRSLPPSTLSKEDDSGTDLWQRFAVWAEERRDGLALTNVESYLFIKNAKAVEDLNLSNLIAFNPLLQKGKYPWAEKDFYKQQSQRRKELNAPETNSGILNTVFSEQRKDFILRSMPALELIMEPMYVTKTVAEKNDNSALIAIYTHSDYYSQQQNRRCHIFFGRAFEIMFWSVLAKTENLPQEFYEKDKFKSLFGNIFKKVPFYSIFSMNPTKVVDEENDDGSEPDFSQKLDDSINELVEDIYIWATSNKLRAFKNKNLIPLMTCVFNKVFSQINVLRKNVQDRVKFRDEHLSDLAKRFEYMFINAIFTFIREGVVVNTNVATGAAPARVRNLSEFNRYDKTLSRNMSGILSVKEDNGLTIVKESEGDIADLLFEIWHSPLFKLTTRTCYPIGKINSQNTAQENLSSDFNSFFENGINFELIKQYYWQTSNHDNIRTADVREWATSRLNEAIILFSWMKESKSIKAKIDGQSYEGRLFRGLQQALEGYEEV* (SEQ ID NO: 328)
28	MFNQDPYWLIPTLCLASDRIFYAQLRDHLGQKSSGERKKEKNGYILVQAAQDYQFYFGGRIRKEDVQNNALMWQIETGNENCLSMLDSLSAYFLTWRGNCFEVRRERLEPWLMICSVIDPAWIIAYAYQQLIKQNVVCDSELISLLTEHQCPFAFPKGRGDISFADNHVHLNGHGYSSISMLNFIDGNYKVKKGIKWPYRQEYTLFESGLLDKNDLPRWLSAYSSCLLKNVYNSFQQGKRSEVDFTCLKDAVETVLADEDKYYFLEVASLYDVVTLQQRVLYEAAQQKYHSHQRWLLYTCGIMLGTESEDYANALANLIRISNILRNYMVVSAVGLGQFIDFFGFNYRRITKPADTNNRVHYDSSAGISREYRVSPDFVLGSGVMPDIYARQLFDFYCTQARKGVPEQGHIVVHFTRSFPDKKSTYDKLLTECRERLRSQCDYFGRFLTSLTLQSIEYKNLSTDEDRSIDIRKLVRGYDVAGNENELQIEVFAPVLRVLRAAKFKGEGVNFKRLQRPFITVHAGEDYCHILSGLRAMDEAVEFCMLGEGDRIGHGLALGVDIKLWANRQKRAYLTVGQHLDNLVWAYHQAVLLSQHIVEHIPVMHELRDKIHYWSHQLYSETYTPDLLFKAWLLRRNWPDYKSIISDPANINEWVPDQHILVSTDETTAKARKIWERYLNSGLAENDVFNRIISVNCAPDTAQNFSMTFNENEDILSKGELLLYEAIQDFLIEKYSRLGLVIEACPTSNIYIGRLEKYHEHPLFRWNPPDSQWIKPGGKFNRFGLRTGPLSVCINTDDSALMPTTIENEHRLMRDCAIHFYGIGTWMADLWINSIRIKGIEIFKGNHLSQDLDNLI* (SEQ ID NO: 329)
29	MNTIYIPLDSGESAVLKDPDTLLPRNIYEQLTRFIEKAVNEVPKPHEALNETRSHKAISIDGARGTGKTSVLVNLNDYLQSNAQQLAGKIHILDPIDPTLLEDGESLFLHIIVAAVLHDKEIKTAQSRDLDKSRVFTQKLENLAHGLESVDLQQNQRGMDKIRSLYGSKHLANCVEEFLKSALELIGKKLLILPIDDVDTSLNRAFENLEILRRYLTSPYVLPVVSGDRRLYDEVCWRDFHGRLNKDSAYNRKNTYDIARDLAIEYQRKILPLPRRLSMPDVSDYWQQDGIEVTLDKNGIPLRNFMAWLKIFITGPVNGLEGSDLPLPIPSIRALTQFINHCRDLIRELPEPFRKKVSTLALRRMWQMPDVPLDVLESFAEKHRELSKEAKREYGEAYKLFYDGLKNFTAWDSKAYLEDDKQSAWLDRLCEYFRFEPKAGAVFLTLQAKQFWVSWAQGDNRNQSILATPLFQPLLHNFREYDVFERYDDLSDWESQLRTRLPESWLTAIKGQKTLLPYPVAEAGINTSLKWRYWEELENYGFDPALESKANFLLSTLMQRNFYTNSKQSVVINIGRVFEIIIASLVSDLELADLQRIRQRSPFYSASALAPTKTLDLEEDFTKKNTRFMNNRSETDRDISDDILVDVPDKNEDAWKKICDEINHWRKTHNVASTNLSPWLVYKVFNKTYSQVANNVFVPSGMQNVDAALNVFGRVFYAVWSAFGSFEKGELFGLSDVVATTNIISAKNFYNHDNFRVNVGPFTPEQNQNSDSDREAYQHRKMYGEKTRAVSYVLATHPLKKWIDEVLRTEFKQKQNAQIQTERKMPIQAEKIIDISPAREFITRKLSLNSHSRLVKTRIIKQLKMLYPNYDKAKDFIDEVTNHFPQNDPAINTLQKAFAELYPDGDK* (SEQ ID NO: 330)
30	MLTRSLSEHAAGCFFTDERLSQRFLDILLSPPKDFETWSSLQEESFKLLVKSIDSRYPRTYRLTDVRQLVGNICDNGLLTSPTLPWLDVIADQLLLRNGDLLYYRENKVQDYVRIAAELDPALLVGWRLGDWLLQSPPPRLTDITRVVMAQNPFFAPPANAGKPFAEGHVHLGGVTAGDTILDGYLFEEIELPKSKDMLLWAHKEHDELTPLINRAKSLLTVLLSAPPQTVSEQTQNGFDQRKTVSEKYKALQNPMDSIHRLPDWLLLAKKNRGTESVSPGWFLNQLAHASEKKHPSRWLWLQLYLCHSYQLKDTHPLERTAILCFWLTVNALRRHIIMDGQGLACFTERYFNGALRAGKKADSSNMRYLFAGKDDVAEVKASPKAFDHEMVTGFSSTLLKTLGIPAVFPPYIFGEHEIKPDERVLRYIGALERWQFCGHFSRSKTASRGKRAKADLQANWTEAERLLQKLYSHNGWNHPVFLGGKRNPHFHFQPSNWFRGLDVAGDENVLKIAGFAPMLRWLRSGLYPVPEGLRASMSFHFSIHAGEDYAHPASGLRHIDETVRFCEMREGDRLGHALALGIEPALWAKRHGEMILPLDEHLDNLVWQWHYATLLSASLPLAQAVLPLLERRIARFIARCEWCKKRPPQIDNSVVGKQACSDDKPLENITPDTLYRAWLLRRNCSYRLQQLHGGSPLTSQEKCALPDWATLSDKGNVAAQLYQQRHSSLLDDMPPQLVVVRVADEWGTQELIGLGNPGKLRQQALDGKDILQDIDTPVELQFMHALQDYLLDHYDRKGLIIETNPTSNVYIARFKKHVEHPIFRWNPPDEELLKPGAEFNRYGLRRGPVRVLVNTDDPGIMPTTLRTEFLLLREAAIERGVSRTMAEYWLERLRLYGLEQFQRNHLNVFEVIE* (SEQ ID NO: 331)
31	MSGTFPYLQYTDVNGLQPKLKEELKNLRRKEYLSYWPRFLIRRISLYALPFLMFFTFFFCLSLTKKVGAEEVTNILGTVSISFSSCLLLGIIISGVVLLLQWTCFNCKYSPQDTNGVVGARKLNYKLLAHVVFVIACVLLFVFIYCTNNKVFYGFIVFLGLTLLPLVIDRTLGVTRQNERHKLYIRRLERLDELNILREKMNIKFEESHFIEYMKLVDEADHGKNQDTVSDTSYFMTLIENKLKV* (SEQ ID NO; 332)
32	MKIVSNTVWDGLKLPDYRARFFIEVWKEILYVNTPSFYQSKMINTMSGAEELVEAIDDYIQDDKSKKSLLSMIEDYKGNLKKDSIAKDTFKNLHATLLKKIETVPDPISSNYILELKTIVKLVLSKESDYYHELKKQLKSSILSNADLNKKARLMDSIYQLTKS

TABLE 15-C-continued

Sequences of validated defense systems (Sequences encoded by the genes corresponding to rows 1-68 of Table 15-A)

Row No.	Sequence
	FIGYLLWKGYSPTYLYNRMEYLTRIKNYGSRDFSAQFNSCLDKLTIRIHDYTVYFLITPLSKYLIELNNILDVSFINREGIINEKNYNKISQGVESSVLAKIVVNTTDYVSAAWQANEKLDKVIDYLEIEKPEYNIRYSPVCLTEFSNGRFTHRQTINIGRLKQFITSKNYSILENIPNESKVLLRESIKLDRYDVLTRSLRYLRVAKESTSLEQKLLGVWIALECIFESTSGNIISGITNHIPTFYSTQSLEIRIRYSKDLLEARLKPISDSLLEITANQKSKFRDLSLKEYFDIVKIEKNRNKIFDELVSKGDEFAVFRLIKIFESFGTSKKINDRFNDTKKDVESQLYRIYKVRNKITHRAYYGNIRPQLVDHLYSYLLSAYSTLIYSLRYNAINKFEPQDMFNAYIISCESLIFNVEEEKKLENITMDEIILS* (SEQ ID NO: 333)
33	MVAIKMYPAKDGDAFLIICDEEKSAFLIDGGYAETFRQHILPDLRELSFNGYRLRLVMATHIDSDHIGGLVDFFLVNGHAAEPAVITVDRVWHNSLRAMTRPENNAQKVDSREITDFLRRRYHVEADKAKPHEISARQGSSLAASLLAGDYHWNEGKGYQCICTGTSIPNLMCDNSLTILSPSKERISALCLWWRRQLASLGFSGRSSSSEAFDDAFEFFCKREASQVPLPHVINARTPLLERDYARDTSPTNGSSIAFSLVLNKKRILMLGDAWAEEVVTSLGASGASHHFDIIKISHHGSIRNTSPNLLKIIDAPVYLISTDGKKHARHPNLAVLKAIVDRPAAFTRTLYFNYANSASAFMKNYLSASGAQFRIIEGSTDWITL* (SEQ ID NO: 334)
34	MRYAATETEIRNATVLIECAGYTGSGTLIAADKVLTAAHCVVSDDPETPITVTFFGADEDVCVNATISEIDTSCDACLLTLSDSVDIPPITLMTQPEREGSQWKAFGYPASRNGPSHYLHGTISQILPRLFHGVDMDLSVSADCVLEEYSGVSGAAILSENKCIAMVRIRMDGGLGAVSLDKLSGLLIRNGLIPDDIASLPDSSLSGEVVLNRTEFRDNFESFVLEHKGRAVLLEGSPGSGKTTFCRHYQPRSEQLAVAGVYEFTPEDGAGTTFKILPEVFADWLHNQVSILLSGRPARREETEKINLTQKVSDLLHTFSDYWKHKGKYGVIFIDAVNEASECGDEAVSRFTALLPVTLPENVKLVFTAPSLSSAGKAFRHWLTPQDCISLTLLSHREVLQLTARELKTSAPSLSLLTRVSDIAQGHPLYLRYILGYLKANPDQVNLEIFPVFSGSIETYYERLWQGLVKDESAVNLLGILSRMRWGIDISSLIPVLTPQEQTVFVPTLDRIQHLLLNDKSSALCHQSFAAFINSKTAVINSLLHGRLADFCLTSGESYGLINRAYHLLLASHDRHPEAALVCTQEWADACIVKGAQPDILIHDIRQTLKNTLIRADAVASIRLLLLFQRMTFRHHFLFLQSAYHSGLALAALGRPDEALEQLIPSGSLVVDAVDAIVSAQTLARMGNSEHALKLLEKVKSAVDQEFERNPVNLSDFIGLSLAWVRAELMAGVVDGHGRTREVVEYLYGCGQVVRDNFEQSAHSKSAYTRAFYPLQAEMEAVNIAFNDRSVSLRTVKEKFGSLPENILDLMLSSVMRAHDIILQHQLPMPQHALQPVWYNLDRLLHTDIPYSNEIRFNSLSSLIFFNAPSALIIRMAGVFSFEVVPEITLLNEENEIAADSIDVSEQGQLWLVSAYLNETQPCPDIKHPSQGCSEWLKTLTEAIFWYSGQARRAVIDGNDEKKELLLVKVQNDILPALSYSLEERMAWPNSWAMPEQIIPMIYEELVNMFGACWPDKISVITDFILAHTPQQCGLYSEGYRRLLNRVIQTLLNEHRFLGQSDTTFQLLETLHAFVSAFTENRQELVPELLNIIPAYISLDAPQLAQDTYTELLGVSMGPDWYKEDQFALMTTMLRVIPQHTDTNTTLSQVAGFLEHASGEMTFRRYVRQEKSQFIGELIRRGNYAHGFNYYRQQSCGSHEEMLTQLSHPAADSPHPLKGMRFPGGALDEEHAVECIVSELRNRVDWRLRWGLLEIFSFGSIGNLAVPFAELINEFSADTEDLNEIPKRLHNILHGDVPFSEHRNFIKNFTEHLADNHKPLFAEFISLLSEDTSDNDVKPPPSGDANQKGTDTSDDVAMQPGLFGKRSAINRAEACMENARKAAARRNTVRASELAVESLHIIQDGDWSVWRKNNHLAELTRTYILDNSADAGSVIRAYASLVEKERYAPAWVIASHLIEIAASKFSDQEAQAINQIVLEHNRHMLGNTEADAAHFSFLNEPDTSDAGEETLYFLFWLLEHPLKFRRERALEVLKWLASDDDKILGQCVTEALVSDIASRAEALMALTDWVSARSPQRIWDFIVKERSLFEWLEGTTALSQVHLLERVTSRAGFVLRNEIAAFERPRKLLLTSEASGQRNIPENLPTWVQSLSQTLAVMEKQGIDIPALLTLLEKRVLQQSGLADITVAFELEKLLARGFTVNRTPSHHRWETMVRFALNQIIHEAAAQDELQNIEPLLRAWNPASEECVEPWEVCNRAKQIICAVMEGRHQQASGIEDGFFLHYLDEVEVSREGQTHLVEISAVLTTAHNGHESLRPGAESEFNATQTPDIERTLSVHLTCQRVKMQPLLFGGATPAAVSKKFMQMTGTLPSDFIRRQWRSGRSLSKNRWGEPISRGSLLLMKRTTTLPPGLGLAWYVTVDGKLMNIFSYAPRRR* (SEQ ID NO: 335)
35	MKYSSMETPKTREEFEARCFHLLNAIKLGRYHGIPGEGNKEQVPFLPNGRVDLANIDTMTRLSMNSLYDFHYNRDNYPQFDLSENDENEEATD* (SEQ ID NO: 336)
36	MEPISITVATYVATKLIDQFISQEGYGCIKKALFPQKRYVDRLYQLIEETAIEFEETYPVESGAIPFYHSEPLFEMLNEHIFFKEFPDKEILLDKFKEYPSITPPTQQQLSLFYEMLSLKINNCSKLKKLHIEETYKEKIFDINEELIQVKLILRSIDEKLTFHLSDDWLNEKNSQAIADLGGRYTPELNVKLEIAEIFDGLGRTNDFSKIFYSHIDSFLVAGKKLHSCDVISSELFEINQSLKEISDIYQEINFSKLDEIPINKFNNYVSSCQTAIGGAVSILWELREKSEQVGETKHYSDKYSSTLRMLREFDYACNELRIFINSTTVKLANNPFLLLEGKAGIGKSHLLADVIKNRIASGYPSLLILGQQLTSDESPWSQIFKRLQLKITSREFLEKLNLYGKKTGKRVLVFIDAINEGNGNKFWNDNINSFVDEIRCFEWLGLIMSVRTTYRNVTISHENVVRNNFEIHEHIGFQNVELEAVSLFYDYYNIERPSSPNLNPEFKNPLFLKLLCEGIKKNGLTKVPVGFNGISNIFNFLVEGVNKSLASPKKYAFDPSFPLVKDALNEIIKFKLEIGRNSISLKDAHSVVQSVVNDYVADKTFLSALIDEGLLTKGIVRNDDNSTEEVVYVAFERFDDHLTVNFLLNDVENIESEFKPDGRLKKYFHDECDFYIKSGIVEALSIQLPERYEKELYEFLPEFSNNLKLLEAFIDSLIWRDIKAIDFEKIRPFINEHVFKFKDSFDHFLEAVISISGLVGHPFNANFLHDWLKDYSLANRDSFWTTELKYKYSEDSAFRHLIDWAWARTDKSFVSDESIELVATSLCWFLTSSNRELRDCSTKALVSLLEPRIPVLRKIIDKFYGVNDPYVWERIFAVALGCTLRTDNIKELKYLAETVYQKVFCSKYVYPNILLRDYAREIIEFANHLGLELESIELSKTRPPYNSIWPDKIPSKEELESLYDKEPYRELWSSIMEDGDFSRYTIGTNYNHSDWSGCKFNETPVDRKQVFKTFKCKLTDQQKDLYDATDPFIYDDKCEGIKFGRVVGRKAQEEIKASKKLFKNSLSYDLLSEFENEIEPYLDHNNNLLETDKHFDLRLAQQFIFNRVIELGWDPEKHGNFDQQIGTGRGRREAFQERIGKKYQWIAYYEYMARLADNFTRFEGYGDERKENPYQGPWEPYVRDIDPTILLKETGTKKISNKEMWWLNDEVFDWTCSNEDWVKSSTTITNSYAFIEVKDDNGDEWIVLESHPSWKEPKIIGNDDWGHPRKEVWYQIRSYIVKVEEFENFRCWAIAQDFMGRWMPECTDRYQLFNREYYWSEAFKSFKSDYYGGSDWTSVTDRESGAKIADVSVTSINYLWEEEFDKSKIETLNFLKPSNLIFEKMGLKSGEVEGSFNDENGTMVCFAAEAVYASKPHLLVKKEPFLTMLRDNGFEIVWTLLGEKGVIGGSLISSHHYGRQEFSGAFYYEDSQLTGSHKTSFTR* (SEQ ID NO: 337)
37	MSDSLLVRTSRDGDQFHYLWAARRALRLLEPQSTLVALTIEGASTTEMGSQPVVEDGEELIDIAEYYGSNELATATTVRYMQLKHSTMHSDTPFPPSGLQKTIEGFATRYKALIQKIPVETLRTKLEFWFVTNRPVSSSFSEAINDAANQHVTRHPHDLAKLEKFTGLQGAELSIFCQLLHIEGQQDDLWSQRNILLRESAGYLPDLDTEAPLKLKELVNRKALTESAANPSITRMDVLRALGVDETDLFPAPCRIERIENSVSRTQEATLVQRVVEAFGAPVIIHADAGVGKSIFSTHIEEHLPTGSVSILYDCFGLGQYRNASSYRHHHRTALVQMANEMASRGLCHPLIPNAGTGISQYMRAFLHRLSQSISILRASEPLAVLCIIIDAADNAQMAAEEIGETRSFIKDLIREKLPDGVCLVALCRPYRRELLDPPPEALTLSLQTFNRDETAAHLHQKFPDASESDVDEFHRLSSCNPRVQALSLSQNLPLNDTLRLLGPNPKTVEDTIGEVLEKSIARLRDTAGISERAQIDTICSALAILRPLIPLSVLSAISGVAGSAIKSFALDLGRPLIVSGETIQFFDEPAETWFQRRFRPSAADLHQFITKLRPLTKDSSYAASVLPALMLEGNQLSELIELAISSQALPETSAVERRDIELQRLQFALKAALRTGRYQDAAKLALKAGGECAGDNRQRVLLRDNIDLAAKFVGSNGVQELVSRNAFPDTGWPGSRNAYYAAILSEYPELSGEARSRLRLTMEWLTNWSQLPDDERSRQNVTDQDRAVMLIACLNIHGAEAAARELRRWRPRKLSFDAGK

TABLE 15-C-continued

Sequences of validated defense systems (Sequences encoded by the genes corresponding to rows 1-68 of Table 15-A)

Row No.	Sequence
	IVAMQLLAHARYDELDQLAIAAGNDISLVMGIVLEARKLHRPVAEQAIRRTWRLLKSQRVSIKDRNHANNQTIAAITGMVEMALIQSVCTESESIQLLDRYLPKVPPYALTSEYSKERVAYVRAYALQANLMGSQLALSDLASTEVKKELMAEKRHGESDDLRQLKQYSGVLIPWYNLWAKVILGKTRKADLESELSDTQKESTAIKGHSYSEHSLSSNEIANVWFDILIEAGNVSKDDVENIIKWSQHKGNRVFTPTLHRFSSVCAEISGLGELSYHFAELALSLWRDEHSDAQIKADGYIDLSRSLISLDEPEAKEYFNQAIEVTNKLGDENLSRWEAILDLAEYVAGKTQVPPEISYKLARCAELTREYVDRDKHFAWSDTVEILAELCPSSALAIISRWRDRTFGNHRSILAWTIEHLVKKNKINALDALPLITFENDWHKCDLLDSVLSSCTDDKDKIMAFEVVYHYTKFNVQNIQNLKKLDAISTSLGIEHTELKERISGLQHTETVSKKSSLSSNDNEQGHDQEWESIFKDCDLSSIDGISAAYEKFRNVPEFYSKETFIKKAISRVKTGKECSFITAIGAIFHWGLYDFKYILESIPDEWTSRLSIKTTLAGLIKEYCQRFCMRIRKSRVYEIFPFSLASRLSGISEKEIFGITLEAIAESPEPANSDRLFSLPGLLVSKLESNEALDVLSYALDLFDEVLKDEDGDGPWNEKLSPPTHVEDSLAGYIWARLGSPEAEMRWQAAHAVLALCRMSRTCVIQGIFQHAINATTLPFCDRNLPFYTLHAQLWLMIAAARVALDDGKSLIPNIGYFYHYATTDQPHVLIRHFAARTLLALHDSDLISIPAQEENKLRNINQSTTLPVLDKVEDHRGEDSYTFGIDFGPYWLKPLGRCFGVSQKQLEPEMLRIIRDVLGFKGSRNWDEDERNKRRYYQDRDNHHSHGSYPRVDDYHFYLSYHAMFMTAGQLLATKPLVGSDYDDVEDVFQDWLRRHDISRNDHRWLADRRDIPPKERSSWLNSSSDNRDEWLASISENVFNETLCPSPGLLTLWGRWSDVCSDRKESIIVHSALVSPERSLSLLRALQTTKNVYDYKIPDAGDNLEIDHAHYQLKGWIKDIAEYCGIDEFDPWAGNVRFPIPEPASFIIDAMKLTTDKDHRVWYSPSDVEPAMISSIWGHLSGKNDEEKSHGYRLCASIHFIKSALETFNMDLILEVDVDRYSRNSRYERNNENELDNIPSSTRLFLFRHDGTIHTLYGNYRNGEKTS* (SEQ ID NO: 338)
38	MAHHIAELIYDAEHCTDDIVRTAKQAEIRDSIWSFWSNRYELPIGSRPFQELEPILRTLKGLDPENEQPRFFSPYRDLINVEKETSEVQKWLTAAKDIDSAAKILIDYCLSLAAENAIDKSQEWVELAQKAGLNKDVDLLEIRIFQLRGTPANTDNPNNAQRRILEKRQKRLEAFLLLGSQLNEQLKSQLEALPAIEDEPTDDDEDF* (SEQ ID NO: 339)
39	MVKPNWDNFKAKFSENPQGNFEWFCYLLFCQEFKMPAGIFRYKNQSGIETNPITKDNEHGWQSKFYDTKLSDNKADLIEMIEKSKKAYPGLSKIIFYTNQEWGQGRKSHEPEGDKNADNYLETVGNSNDPKIKIEVDQKAYESGIEIVWRVASFFESPFVIVENEKIAKHFFSLNESIFDLLEEKRKHTENVLYEIQTNIEFKDRSIEIDRRHCIELLHENLVQKKIVIVSGEGGVGKTAVIKKIYEAEKQYTPFYVFKASEFKKDSINELFGAHGLDDFSNAHQDELRKVIVVDSAEKLLELTNIDPFKEFLTVLIKDKWQVVFTTRNNYLADLNYAFIDIYKITPGNLVIKNLERGELIELSDNNGFSLPQDVRLLELIKNPFYLSEYLRFYTGESIDYVSFKEKLWNKIIVKNKPSREQCFLATAFQRASEGQFFVSPACDTGILDELVKDGIVGYEAAGYFITHDIYEEWALEKKISVDYIRKANNNEFFEKIGESLPVRRSFRNWISERLLLDDQSIKPFIAEIVCGEGISNFWKDELWVAVLLSDNSSIFFNYFKRYLLSSDQNLLKRLTFLLRLACKDVDYDLLKQLGVSNSDLLSIKYVLTKPKGTGWQSVIQFIYENLDEIGIRNINFILPVIQEWNQRNKVGETTRLSSLIALKYYQWTIDEDVYLSGRDNEKNILHTILHGAAMIKPEMEEVLVKVLKNRWKEHGTPYFDLMTLILTDLDSYPVWASLPEYVLQLADLFWYRPLKETGERYHSMDIEDEFGLFRSHHDYYPESPYQTPIYWLLQSQFKKTIDFILDFTNKTTICFAHSHFAKNEIEEVDVFIEEGKFIKQYICNRLWCSYRGTQVSTYLLSSIHMALEKFFLENFKNADSKVLESWLLFLLRNTKSASISAVVTSIVLAFPEKTFNVAKVLFQTKDFFRFDMNRMVLDRTHKSSLISLRDGFGGTDYRNSLHEEDRIKACDDVHRNTYLENLALHYQIFRSENVTEKDAIERQQVLWDIFDKYYNQLPDEAQETEADKTWRLCLARMDRRKMKITTKEKDEGIEISFNPEIDPKLKQYSEEAIKKNSEHMKYVTLKLWASYKREKDERYKNYGMYEDNPQIALQETKEIIKKLNEEGGEDFRLLNGNIPADVCSVLLLDYFNQLNNEEREYCKDIVLAYSKLPLKEGYNYQVQDGTTSAISALPVIYHNYPMERETIKTILLLTLFNDHSIGMAGGRYSVFPSMVIHKLWLDYFDDMQSLLFGFLILKPKYVILSRKIIHESYRQVDYDIKKININKVFLNNYKHCISNVIDNKISIDDLGSMDKVDLHILNTAFQLIPVDTVNIEHKKLVSLIVKRFSTSLLSSVREDRVDYALRQSFLERFAYFTLHAPVSDIPDYIKPFLDGFNGSEPISELFKKFILVEDRLNTYAKFWKVWDLFFDKVVTLCKDGDRYWYVDKIIKSYLFAESPWKENSNGWHTFKDSNSQFFCDVSRTMGHCPSTLYSLAKSLNNIASCYLNQGITWLSEILSVNKKLWEKKLENDTVYYLECLVRRYINNERERIRRTKQLKQEVLVILDFLVEKGSVVGYMSRENIL* (SEQ ID NO: 340)
40	MQVQHHTEPNLKNEIVALFKASQLIPFFGSGFTRDIRAKNGKVPDAIKFTELIRNIAAEKEGLTQTEIDEILRISQLKKAFGLLNMEEYIPKRKSKALLGNIFSECKLSDHEKTKIINLDWPHIFTFNIDDAIENVNRKYKILHPNRAVQREFISANKCLFKIHGDITEFIKYEDQNLIFTWREYAHSIEENKSMLSFLSEEAKNSAFLFIGCSLDGELDLMHLSRSTPFKKSIYLKKGYLNLEEKIALSEYGIEKVITFDTYDQIYQWLNNTLQNVERKSPTRSFELDDSKLMKEEAINLFANGGPVTKIVDNKRILRNSITFSQRDVCDDAIKALRNHDYILITGRRFSGKSVLLFQIIEAKKEYNASYYSSTDTFDPSIKNSLIKFENHIFVFDSNFFNAQSIDEILTTRVHPSNKVVLCSSFGDAELYRFKLKDKKILHTEIQIKNNLINEEGNYLNDKLSFEGLPLYKSSETLLNFAYRYYSEYKNRLSGSNLFNKQFDEDSMFVLILIAAFNKATYGHINSHNKYFDIQNFISQNDRLFELESTNTDPSGVIICNSPSWLLRVISEYIDKNPASYKTVSDLHSLASKGFLAASRNLISFDKLNELGNGKNVHKFIRGIYKEIAHTYREDMHYWLQRAKSELISAHTIDDLVEGMSYASKVRLDSAEFKNQTYYSATLVLAQLSARALSINNDKIYALSFFESSLESIRNYNNNSRHINKMMDKNDGGFRYAIQYLKDNPLIELLPRKDEVNELINFYESRKK* (SEQ ID NO: 341)
41	MQFITNGPDIPDELLQAHEEGRVVFFCGAGISYPAGLPGFKGLVELIYQRNGTTLSEIEREVFERGQFDGTLDLLERRLPGQRIAVRRALEKALKPKLRRRGAIDTQAALLRLARSREGALRLVTTNFDRLFHVAAKRTGQAFQAYVAPMLPIPKNSRWDGLVYLHGLLPEKADDTALNRLVVTSGDFGLAYLTERWAARFVSELFRNYVVCFVGYSINDPVLRYMMDALAADRRLGEVTPQVWALGECEPGQEHRKAIEWEAKGVTPILYTVPAGSTDHSVLHQTLHAWADTYRDGIQGKKAIVVKHALARPQDSTRQDDFVGRMLWALSDKSGLPAKRFAELNPAPPLDWLLKAFSDERFKYSDLPRFCVSPHVEIDPKLRFSLVQRPAPYELAPQMSLVSGCVSASKWDDVMSHIARWLVRYLGDPRLIIWIAERGGQIHDRWMFLIESELDRLAALMRERKTSELDEILLHSPLAIPGPPMSTLWRLLLSGRVKSPLQNLDLYRWQNRLKNEGLTTTLRLELRGLLSPKVMLRRPFRYSEDDSSSTDEPLRIKQLVDWELVLTADYVRSTLFDLADESWKSSLPYLLEDFQQLLRDALDLLRELGESDDRHDRSHWDLPSITPHWQNRGFRDWVSLIELLRDSWLAVRAKDSDQASRIAQNWFELPYPTFKRLALFAASQDNCIPPERWVNWLLEDGSWWLWATDTRREVFRLFVLQGRHLTGIAQERLETAILAGPPREMYEDNLEADRWHYLVAHSVWLCLAKLRGAGLVLGESAATRLTEISTAYPKWQLATNERDEFSHWMSGTGDPGFEESIDVDIAPRKWQELVQWLAKPMPERLPFYEDTWSDVCRTRFFHSLYALRKLSQDDVWPVGRWREALQTWAEPGMILRSWRYAAPLVLDMPDAVLQEISHAVTWWMEEASKTILCHEEILLALCRRVLMIETSPESSTIRNGIETYDPVSTAINHPIGHVTQSLITLWFKQNPNDNDLLPVELKTLFTKLCNVQIELFRHGRVLLGSRLIAFFRVDRPWTEQYLLPLFAWSNPVEAKAVWEGFLWSPRLYEPLLIAFKSDFLESANHYSDLGEHRQQFAIFLTYAALGPTEGYTVEEFRTAISALPQEGLEVAAQALYQALEGAGDQREEYWKNRVQPFWQQVWPKSRNLATPRISESLTRMVIAARGEFPAALAVVQDWLQPLEHLSYDVRLLLESDICSRYPADALSLLNAVIAEQHWGPRELGQCLLQIVQAAPQLEQDVRYQRLNEYSRRRSV* (SEQ ID NO: 342)

TABLE 15-C-continued

Sequences of validated defense systems (Sequences encoded by the genes corresponding to rows 1-68 of Table 15-A)

Row No.	Sequence
42	MTNKNKIKPLLNNISARLWDGRAAILIGAGFSRNAKPLTSKARKFPMWNDLGDIFYESVYCKKNDNRYSNVLKLGDEVQAAFGRATLDKLIMDHVPDKEYEPSKLHVSLLSLPWIDVFTTNYDTLLERASVNVDSRKYDIVLNKNDLMNAERPRIIKLHGSFPSERPFIVTEEDYRKYPLENSPFVNTVQQSLIENTLCLIGFSGDDPNFLNWIGWIRDNLGTENSPKIYLIGLFSFNEAQRKLLEKRNISIVDLSFLGDFGKDHYLAHQRFIQFLYESKNRDNLIEWPIETNYDRIVFNDGIELKTEKIKKCILEWAQSRQSYPNWLILPESNRSNLWQNTIDWLSVANYDVAWDGSDDLDFGYEITWRLNKALLPIFNDTSEFLFKLIEKYEINYVSGINNKIIDFDEKYSHITLSLMRFCRQENLIDKWKNLNDLLIQNLDRLTPEVKSDYYYENILFSYFNLNFDEARNKLSNWETNKLLPHHEIKRAGLLAEFGMLDEAINLLEETLSTIRRNSLLSSRNIDYSSESQEAYGIYILRMFKRSLRLDSKDDDYSSEYNSRLATLSQYRSDPENEIKYLEIKLESLPGTFKNTNDTDFDLNKRTVTTYLGGSPTEVRSLDAFSFFLLAEELGLPFHIPGMNIFSGIVENAARHIYQYSPEWAIFSIFRTFNKDKAKSLFNRNRISSLERKKVEDLFDGYYKKYEQIITKKIEDRLNDKLEIEISTLSIIPEILSRLVTKVSFNKKKDIIHLLLKLFNSDNFHQYMETKDLLKRTISNLSDLQKISLIDIFIDFPSAPPNTQLHMGQRYNFLTPFECLLGVTITPPKENSKKIASAKLKKDINDLKSDNLDLRKAVSQKLITLYNLEMLNKSDTTKLIKNLWSKRDNFGFPIGSGYYKFFFINNLNPDNENIADKFISIIKTYKFPVQEGKRVSITGGLDEYCTELNGALHHISLPEKTLSEIISKIHDWYVKDRAWLEKRDDLAKEFTLRFRNITNIITTILEHHKDKLHAESINEISSLLDKMKEDKIPVNSAVTMLCLKNKSTYLERIKDIENGLYSFNKDDVIEAINSTYVFIRNNEFPLTIIQAISDKIAWDRNPRLPDCYNLIAYIINSCEFTLPDYLIEKILRGLAYQINIDDRDFVDNNEYLNHLEKKLSATKLAASMFRKNETLGIDQPSIIQEWKNMCNSRNEFDEIRNEWNNNI* (SEQ ID NO: 343)
43	MSIYQGGNKLNEDDFRSHVYSLCQLDNVGVLLGAGASVGCGGKTMKDVWKSFKQNYPELLGALIDKYLLVSQIDSDNNLVNVELLIDEATKFLSVAKTRRCEDEEEEFRKILSSLYKEVTKAALLTGEQFREKNQGKKDAFKYHKELISKLISNRQPGQSAPAIFTTNYDLALEWAAEDLGIQLFNGFSGLHTRQFYPQNFDLAFRNVNAKGEARFGHYHAYLYKLHGSLTWYQNDSLTVNEVSASQAYDEYINDIINKDDFYRGQHLIYPGANKYSHTIGFVYGEMFRRFGEFISKPQTALFINGFGFGDYHINRIILGALLNPSFHVVIYYPELKEAITKVSKGGGSEAEKAIVTLKNMAFNQVTVVGGGSKAYFNSFVEHLPYPVLFPRDNIVDELVEAIANLSKGEGNVPF* (SEQ ID NO: 344)
44	MSLFKLTEISAIGYVVGLEGERIRINLHEGLQGRLASHRKGVSSVTQPGDLIGFDAGNILVVARVTDMAFVEADKAHKANVGTSDLADIPLRQIIAYAIGFVKRELNGYVFISEDWRLPALGSSAVPLTSDFLNIIYSIDKEELPKAVELGVDSRTKTVKIFASVDKLLSRHLAVLGSTGYGKSNFNALLTRKVSEKYPNSRIVIFDINGEYAQAFTGIPNVKHTILGESPNVDSLEKKQQKGELYSEEYYCYKKIPYQALGFAGLIKLLRPSDKTQLPALRNALSAINRTHFKSRNIYLEKDDGETFLLYDDCRDTNQSKLAEWLDLLRRRRLKRTNVWPPFKSLATLVAEFGCVAADRSNGSKRDAFGFSNVLPLVKIIQQLAEDIRFKSIVNLNGGGELADGGTHWDKAMSDEVDYFFGKEKGQENDWNVHIVNMKNLAQDHAPMLLSALLEMFAEILFRRGQERSYPTVLLLEEAHHYLRDPYAEIDSQIKAYERLAKEGRKFKCSLIVSTQRPSELSPTVLAMCSNWFSLRLTNERDLQALRYAMESGNEQILKQISGLPRGDAVAFGSAFNLPVRISINQARPGPKSSDAVFSEEWANCTELRC* (SEQ ID NO: 345)
45	MDRSAVDTIRGYCYQVDKTIIEIFSLPQMDDSIDIECIEDVDVYNDGHLTAIQCKYYESTDYNHSVISKPIRLMLSHFKDNKEKGANYYLYGHYKSGQEKLTLPLKVDFFKSNFLTYTEKKIKHEYHIENGLTEEDLQAFLDRLVININAKSFDDQKKETIQIIKNHFQCEDYEAEHYLYSNAFRKTYDISCNKKDRRIKKSDFVESINKSKVLFNIWFYQYEGRKEYLRKLKESFIRRSVNTSPYARFFILEFQDKTDIKTVKDCIYKIQSNWSNLSKRTDRPYSPFLLFHGTSDANLYELKNQLFNEDLIFTDGYPFKGSVFTPKMLIEGFSNKEIHFQFINDIDDFNETLNSINIRKEVYQFYTENCLDIPSQLPQVNIQVKDFADIKEIV* (SEQ ID NO: 346)
46	MSRNNDINAEVVSVSPNKLKISVDDLEEFKIAEEKLGVGSYLRVSDNQDVALLAIIDNFSIEVKESQKQKYMIEASPIGLVKNGKFYRGGDSLALPPKKVEPAKLDEIISIYSDSIDINDRFTFSSLSLNTKVSVPVNGNRFFNKHIAIVGSTGSGKSHTVAKILQKAVDEKQEGYKGLNNSHIIIFDIHSEYENAFPNSNVLNVDTLTLPYWLLNGDELEELFLDTEANDHNQRNVFRQAITLNKKIHFQGDPATKEIISFHSPYYFDINEVINYINNRNNERKNKDNEHIWSDEEGNFKFDNENAHRLFKENVTPDGSSAGALNGKLLNFVDRLQSKIFDKRLDFILGEGSKSVTFKETLETLISYGKDKSNITILDVSGVPFEVLSICVSLISRLIFEFGYHSKKIKRKSNENQDIPILIVYEEAHKYAPKSDLSKYRTSKEAIERIAKEGRKYGVTLLLASQRPSEISETIFSQCNTFISMRLTNPDDQNYVKRLLPDTVGDITNLLPSLKEGEALIMGDSISIPSIVKIEKCTIPPSSIDIKYLDEWRKEWVDSEFDKIIEQWSKS* (SEQ ID NO: 347)
47	MIMSTPWLTPIVADSDHAEANAVSYEALTPTELDSDKAGCYISALNYAYEHPDIRNIAVTGPYGAGKSSVLKTWCKAHNGTLRVLTVSLADFDMQRHVDESNGDSSSDEGTKNTGSVEKSIEYSILQQILYKNKKHELPCSRIDRISDVTAGQILRSASFLTGTILLSGAALFFLAPDYVTTKLSLPGAFARYLLECPFGVRVSGAVASVMGSLCLLLNQLHRIGIFDRKVSLDKVDLLKGAVTTRASSPSLLNVYIDEIVYFFDSTKYDVVIFEDLDRFNNGRIFVKLREINQIINNCLSDRKPVKFIYAVRDGIFNSAESRTKFFDFVMPVIPVMDNQNAYEHFVKKFKEEEINNNLSECISRIATFIPNMRVMHNITNEFRLYQNLVNSRENLAKLLAMIAYKNLCAEDYHGIDSKKGVLYHFIQSYLDHEIQNELLHSANNELEDMAQSLVAITNEKLANRENLREELLMPYLSKNYSGALVFYTEGRQISLDDLIQDEDEFLMLLDKENIQVVTPYNRQNFLMINQRDTEKLKQQYEKRCHLIETKSVDNITRVKNNISSLESLRTEILSGTVADIAEKMTNEGFVAWIKKKEDTGVLTIQSEHEQIDFIFFLLSSGYLSTDYMSYRSIFIPGGLSETDNLFLKDVMSGKGPEKTFSFHLDNVNNIVERLKKLGVLQRDNAQHPAVIRWLIDNDPDTLKNNIMALLSQTGSQRVVSLLMLMQNDFTTYVRLRYLEIFMSDEHILNRLLAHLCASEERTPEQKFFVQEIAAHLLCLTEKSNIWQSVEINKRIGELIDSSPILITAVPKGYGDAFFEVLKDNTLSVSYIPGDVGDEKCSVIRKIAGAGLFKYSVSNLKNVYLCLTQDKNEERMSFSLYPFHCLESLAISELTEILWTNIEDFILSVFIESEEIDRIPELLNSSEVSMTVVEQIIAKMDFCINNLDDIINRSECADNNASGRNIYSMLLQHDRIFPSFDNIIHLLHDTSINTSGELVQWVNEKHFEFEPSDIVINDTGIFNNFISELICSPVISEEALLKVLSNLNVVIIDVPENIPLRNAELLCSEKKLAPTVNVFTVLFNALSENVDDINRMNTLLGNLIAQRPEIITQEPEDIFYIEGDFDEELASELFRHKLIGMNIKVAALRWLRDNKPGILDKSYLLSLDILAELSPWMGDDDLRLTLLKRCLVAGDAGKDALCVVLNSFADESYHGLLPHDRFRKIPHSVDLWEVAELISNLGFIQPPKMGSGRDEHKIVITPVRYVRDVEFYD* (SEQ ID NO: 348)
48	MFLNDQETSTDLLYYTAIASTVVRLVDETSDAPITIGVHGDWGAGKSSVLKMLEAACEKKDKTHCIWFNGWTFEGFEDAKTVIIETIVEDLVASRPMSTKVAEAAKKVLRRIDWLKMAKKAGGLAFTAFTGIPTFDQIKGMYELASDFLSAPQDKLSAADFKAFAEKAGGFIKEADTDSNTLPKHIHAFREEFRALLDAAEIEKLVVIVDDLDRCLPKTAIETLEAIRLFLFVEKTAFVIGADEAMIEYAVKDHFPDLPQSTGPVSYARNYLEKLIQVPFRIPALGTAETRIYTTLLLAENALGSEDDNFKALLNKAREEMKRPWISRGLDREAVMAALNGKIPEVVENALLFSLHVTPMLSSGTHGNPRQIKRFLNSMMLRQAIADERGFGSDIKRPVLAKIMLAERFYPSVYGKLVQLVSNHPEGKPEALAEFEALVRGGKTAPKSRADSKENSSESEDVQNWLKIDWAIGWAKAEPALSGEDLRPYVFVTRDKHSTLSNLVVSSHLIPIMEKLLGPKIGMVKIKGDLEKLSPPDADELFEMLSDKLFQEDSFNRKPRGFDGLEYLV

TABLE 15-C-continued

Sequences of validated defense systems (Sequences encoded by the genes corresponding to rows 1-68 of Table 15-A)

Row No.	Sequence
	ETQPHLQRRLIDFARRIPVKKAGGWLATRIAQSLVDPTLIEEYTKLIQEWASQDENLSLSKSAKATLQLSGYQH* (SEQ ID NO: 349)
49	MGTSKAYGGPVHGLIPDFVENPSPPTLPPVDPADDSTLDTPLIPPDSSGSGPLSTPKANFTRYSRSGSRSSLGKAVAGYVRNGVGGAGRASRRMGASRAAAGGLLGLISDYQQGGATQALERFNLGNLAGQSASTALLSLVEFLCPPGGSVDEGVARQAMLETIADMSDVGEENFDELTPDQLKEVFIGFVVHSIEGRLMADIGKNGIKLPDDIDAIVSIQEDLHDFVDGATRTQLREELRNLTGLSGDAIDRKVEEIYTVAFELLAREGERLE* (SEQ ID NO: 350)
50	MSHHTLVARLGTDDNSDLQLSRQSTHLTEINFLKENGKLDFGLGQALNGLSDLGLTPMDVSVDLALLAATVTAADTRISRGHNAQDLWTREIALYIPVASPTLWNSQTGLLSRMLNFLTGDRWTIHFRSRPVIEHGLIQRSSKERSVNPTSVCLFSGGLDSFIGAIDLLSNGGTPLLISHYWDTTTSVYQQKCAQLLSERYGQSFSHVRARVGFEKTTIEGEDGENTLRGRSFMFFSLATMAADALGGPVTINVPENGLISLNVPLDPLRVGALSTRTTHPFYMARFNELLGNLGISAHLENPYAYKTKGEMAIHCHDHAFLRQHAADTMSCSSPQSTRWNPALNEQQSTHCGRCVPCLIRRASLFTAFGTDDTIYRIPDLRSRVLDSSKPEGEHVRAFQFALARLARSPSRAKFDIHKPGPLSDYPDCLAEYEGVYLRGMKEVERLLSGVITRPLT* (SEQ ID NO: 351)
51	MKLAGQKPAPQWVDFHCHLDLYPNHSALIRECDISRVATLAVTTTPKAWMRNRELTSDSPYVRVALGLHPQLIAEREHEIALLEHYLPSARYVGEIGLDASPRFYRSFEAQERIFSRILNACFEQGDKILSIHSVRAAAKVLGHLENTRLTENCKAVLHWFTGSISEARRAVELGCYFSINEEMLRSPKHRKLVSFLPFERILTETDGPFVFHEEKAIHPRDVQRTVHEIAQIHHVSDTDAAMRILYNLRSLVTNSSHSENSS* (SEQ ID NO: 352)
52	MSTVDTSTAEELNQGGSDFILTSLEAMRKKLLDLTSRNRLLNFPITQKGSSLRIVDELPEQLYETLCSEIPMEFAPVPDPTRAQLLEHGYLKVGPDGKDIQLRAHPSAKDWAHVLGIRTDFDLPDSHKTVVSDSDRELLEKAHQFILQYAQGQNGKLTGIRSEYVNQGIALSALKEACCLAGYEGLEDFERQAKAGNEISISSSNPSHDDNRIQALLYPNELEACLRAIYGKAQTALEESGANILYLALGFLEWYESDSSEKARYAPLFTIPVRCERGKLDPKDGLYKFQLYYTGEDILPNLSLKEKLQADFGLALPLFNEEETPESYFASVKKVVEQHKPKWSVKRYGALSLLNFGKMMMYLDLDPARWPCDKRNILSHEVIRRFFTSQSCGQENSGLPGGFGQHEYCIDSYPDIHDKVPLIDDADSSQHSALIDAIRGQNLVIEGPPGSGKSQTITNLIAAALLNGKKVLFVAEKMAALEVVKRRLDRAGLGQFCLELHSHKTHKRKVLDDINARLVSQATMPTMEEIDAQILRYEDLKQQLNEYAALINNQWAQTGKTIHQILSGATRYRHKLDIDATALHIENLSGKQLDKVTQLRLRDQIVEFSRIYKEVREQVGANAEIYEHPWSGVNNTQIQLFDSARIVDLLQTWQTSIIDFQHSYQEYVDKWALEGESLNTLQYIEQLVEDQSNLPVLCGSEHFPALSELDSPDAIARVRHYLDRFELLQGHYVALSQVIEPQKLRLLEQGQSCDFPREELEKYGAAEDFTLRDLVRWLESIQSIHDELSSIYAQLNDFKNALPDGIASYIDDSQAGLLFCSELLSILGALPTELIRVRDPLFDDDDIDAVLRDLMCQIETLRPLRDGLSTLYQLDQLPSQEMLAHAVAVIQQGGLFAWFKSDWRSAKALLMAQSRKPDTKFAELKRCSADLLKYSELLQRFEQSDFGNQLGNAFRGLDTDCEQLMLLRDWYKKVRACYGIGFGKRVAIGSGLFNLDGEIIKGVHLIEKSQISSRLMTLVKRVEHEAKLLPRISSLLEEHASWLGEQGVLMQSYRQVRNTLIALQGWFINPDISLEQMTHSSEILQNINDLQISLENDSLQLGAFLQLTPLACGAYKNNQLTLDTINDTLNFAEQLVDKINCVSLATQIRHLASGSDYDLLCRDGGEIVSKWNEQIKNAELYALETKLERSQWLKSTDGSLNTLIERNERAIQQPRWLNGWVNFIRCYEQMHENGLQRIWSAVLAGSLPIEKVELGLALAIHDQLAREVIHIHPELMRVSGSQRNALQKSFKEYDKKLIELQRQRIAAKIACRNIPEGNSGGKKSEYTELALIKNELGKKTRHIPIRQLVNRACNALVAIKPCFMMGPMSAAHYLEPGRMEFDLVVMDEASQVKPEDALGVIARGKQLVVVGDPKQLPPTSFFDRSADGEDDDDAAALSDTDSILDAALPLFPMRRLRWHYRSRHEKLIAYSNRHFYNSDLVIFPSPNAESPEYGIKFTYVSKGRFSNQHNIEEAQAVAEAVLHHAHHRPGESLGVVAMSSKQRDQIERAIDELRRNRPEFNDAIDGLHAMEEPLFVKNLENVQGDERDVIFISFTYGPSEHGGKVYQRFGPINSDVGWRRLNVLFTRSKKRMHVFSSMRSEDVLTSETSKLGVISLKGFLQFAESGKLDSLTTHTGRAPDSDFEVAVMEALNHAGFECEPQVGVAGFFIDLAVKDPGCPGRYLMGIECDGAAYHSAKSARDRDRLRQEVLERLGWRISRIWSTDWFSNPDEVLSPIIRKLHELKTLAPDVVVPSYEYVETIESSAEVASDSIDSLMPNLGLKEQLKYFATHVIEVELPNVDADRRLLRPAMLEALLEHQPLSRSEFVERIPHYLRQATDVYEAQRFLDRVLALIDGAEAEANDAAFESELA* (SEQ ID NO: 353)
53	MHRTISEFYRIPPLLIRALKSGISSVVEFHLNRGLPKDSRDSLGNSPLMIAAQYGHFAICEMLLSAGVDVEHQNNLGLRASDLAQEQKLRDLLARYRQPLSLAELERSVVSVEDSETEAELPSAEIPMDFMLWDAEVELKPAEDNLTLRHASAEAQQLLSRYRPKDNSAEWSDIELTLPEPLTPVSHSPQNYPHLSTLLIGALDTGRISLRDIWHAGEEDFGMQWPEFRLSVEALIRDLPLIVDDDDIIPPDAAPATLSVSEPLEPWFDAFNALRQFGIVENYLVDIRQWDVVDKTKEERLGQRMDTALINLIRILAGLSEAEYMQLLQPNYLPEPAPEISEEEDVAEEADEEMPPVSDDDDDNDDTISFIELLVLLRSGKAGEYQDNHIPRPEYADLQQIVERARTLIPDEGHKISLYVSSYREAWEGLIHANLRLVVTIANKYRGRGLDVEDLIQEGNLGLIKAVEKFDYRRGFKFSTYATWWIRQKISRAIADQAQLIRLPVHFYEQFRRWRNSRDQLLYRQGITPTIKRLQALTDLPENQLKRMAKYEEQTVLIGDFHDDAQDSEAALSGDAILTGKDFTSAPVQSLELRECVSLVLETLLPREKQIIKMRFGIGMTQDFTLEEVGKQFDVTRERIRQIEAKALRKLRYHSRASKLGGFVEQWETALSEMQEEEE* (SEQ ID NO: 354)
54	MTTMRHAPPNAAIMIEALRGLGYNTATALADIIDNSISAGARKVDLTFHWRESDSYIVVRDNGCGMSAAELDVAMRLGVKNPLTKRSGHDLGRFGLGLKTASFSQCRRLTVASKKEEITTILRWDLDILAASTDDGWYLLEGADPGSQEALANEEPDSHGTVVLWDVLDRIVTPGYGEKDFLNLMDGVEQHLAMVFHRFLEGNAPRLTLTLNGRKIKAWDPFLSGHPSKPWHSPSAMAPGAPAVKVECHVLPHQDHLTTQEYQQAQGPAGWTAQQGFYVVYRNERLLVAGNWLGLGSPRAWTKDETHRLARIRLDIPNDADIDWKIDIRKSMARPPVSLRPWLTQLAQSTRDRAVRTFAKRGKMNKRKPGEELVQLWQAQKTPSGVRYQISLQHPVISNVLSQAGELSPQIQAMLRLIEETVPVQQIWLDTAETKETPRTGFETAPPAEVLSVLQVMYQTMVGQQAMSPALAKQHLQNMEPFDNYPELIALLPDDQHEKSL* (SEQ ID NO: 355)
55	MSLNPLDDTQLSVLQIVQTFLQSQDKSTITPGILRQHIDMVCQMKPEWSRLDSREILVEELIRRYSIWMGEDSSLSNDEGHQPWLTADAKREWRYWHRYRQWLGKTMPWGVLDTLDRSTDRVLGLLEQPGREGRWDRRGLVVGHVQSGKTSHYTGLICKAADAGYKIIIVLAGLHNNLRSQTQMRLDEGFLGYETSPLREKVTIIGVGAIDSDPVIRPNYVTNRSEKGDFSAGVAKNLGISPEQRPWLFVVKKNKSILKRLHTWIENHVATSVDPITGKRFVSELPLLMIDDEADNASVDTGEIVYDDDGKPDAEHQPTAINSLIRKLLMQFSRKAYVGYTATPFANIFIHESNETRDEGPDLFPSAFIINLGAPSNYIGPARVFGRATAEGRSGEFPLIRRVSDHCSDDGKRGWMPVSHKSSHYPTLDTLTHFPDSLKHAIDSFLLACCVRELRGQGEKHSSMLVHVTRFNKVQSVVYENIDAYIQDVRQRLTRRIGHEPFLHQLESLWQADFLPTNQAIREVMPQQVPDDAFEWQEIVDKLYTVIENVSVRMINGTAKDALDYSDSATGLKVIAIGGDKLARGLTLEGLCTSYFLRASRMYDTLMQMGRWFGYRQGYLDVCRLYTTDELIEWFEHIADASEELREEFDNMVASGGTPRDFGLKVKSHPVLMVTSPLKMRSARSLWLSFSGTVVETISLFKEQEYHKRNYVAFQRLTGRVGAGAPI

TABLE 15-C-continued

Sequences of validated defense systems (Sequences encoded by the genes corresponding to rows 1-68 of Table 15-A)

Row No.	Sequence
	PERRRGDKIEKWNGVIWQNISPEPIIDFLTEYETHAQARKANSKLLADFVTRMNRVDELTQWTVAVIGGGIDRHHDVCGFSVPLMMRKASEGVTDRYSIGRLLSPRDEGIDCDESTWLAALEETQRIFHADPGRNEGREEPVVPGGVVLRRIKGFGINDIPAQRQKGLLLIYLLDPQQALSAAEYQEDALPVVAFGISFPGSRSGVTVEYKVNNVLWEQEYGAAE* (SEQ ID NO: 356)
56	MVRLSKDDLLAAWKALDRSQIDELPGAQGWRGIRLFTHQGCSFHAGRRQPDNEEMLIAVFPHPLSPGSAALPSCKGFRVEMAGTEEGGQNGLMIRRQQTGNVDVFTTMILDILHSLLNVSKPRLFETLLRRIRLWQAFMERDTRPLSQEEEVGLIGELTCLERLIESGLAPSTAVEAWIGPQHGLQDFALDERAIEIKSTTAAKGFCITIHSLEQLDWQRAGSLVLCGLRFSEHPTGATLNDIISRLRQRFEGNATAACIFEGSLCHVGYFTEHAEFYTRHFLLTEAFALPIEADFPSLTHANVPLPVVSARYQLELQTLIPQAQDFNHCLSDFAGLPHGNY* (SEQ ID NO: 357)
57	MEIIDFLRQTQNEIRKEYQDQMAQPGVESPFPELIFTDIVMRHMADIGMTFDDAETCHFMAKVSGHNVRLSGYAFSEDGDQLDLFVSIYHGSDELCHVPDAETKAIAGHCIQFLQKCVDGKLSSTLDQSNDAWQLVTTIEQSYAELEQIRIYVLTDGQVKTRWYQSRDVAGKTIKLEVMDIVRLFNHWQEGKPRDELQVNFDEVAGGALPCVWIPDEMGEYDYALTVVPGETLRFIYEKYGNRILEANVRSFLSQTGKVNKGIRDTLREQPERFMAYNNGIVIVADQVRLGEAPGGGPGIAWMQGMQIVNGGQTTASMFFTKKKFPATNLRNVRVPAKVIVLKQTNNAQEEMLIADISRFSNSQNKVNISDLSANRPVHVQLEKMANTVYCPDGYSRWFYERANGSYKVMLEREGKTPAGIKRLKDAIPPSRRITKTDFAKYHCAWLQRPDLVSLGGQKNFAALMTMIDKDTERYGDELNIETFKNYIAQAIIYKKAYKLINSLFPAFKANIAAYTVAAYSHLYGNKTDLAEIWNQQGIEETMGNRLVSLAHRVNSLLTESANGRMISEWAKKPECWDYVRSKIYFSAQGKKDDFSHGEIA* (SEQ ID NO: 358)
58	MAYEAQISRTNPAAFLFVVDQSGSMSDKMSSGRSKAEFVADALNRTLMNLITRCTKSEGVRDYFEIGVLGYGGQGVSNGFSGSLGGQVLNPISALEQNPARVEDRKRKMDDGAGGIIETAIKFPVWFDPIASGGTPMREALTRAAEELVTWCDAHPDCYPPTILHVTDGESNDGDPEEIANHLRQIRTNDGEVLILNIHVSSLGNDPIRFPSSDTGLPDAYAKLLFRMSSPLPEHLVRFAQEKGHTVGIESRGFMFNAEAAELVDFFDIGTRASQLR* (SEQ ID NO: 359)
59	MKLEFLGTVPKDPEYPKANEDKFAFSEDGRRLALCDGASESFNSKLWADLLARKFTADPKVNPEWVASALAEYSATHDFPSMSWSQQAAFERGSFATLIGVEEFEEHQAVEILAIGDSITMLVDCGKLICAWPFDNPEKFNERPTLLATLYAHNNFVGGSTFWTRHGKTFYLEKLTQPKLLCMTDALGEWALKQALAEDSGFIELLSLQTEEELAELVLRERAAKRMHIDDSTLLVLSF* (SEQ ID NO: 360)
60	MPYPSLEQYNQAFQLHSKLLIDPELKSGTVATTGLGLPLAISGGFALTYTIKSGAKKYAVRCFHRESKALERRYEAISRKISSLRSPYFLDFQFQPQGVKVEGISYPIVKMAWAKGETLGEFLEVNRRSAQAIAKLSASIESLAAYLEKEKIAHGDFQTGNLMVSDGGATVQLIDYDGMFVDEIKTLGSSELGHVNFQHPRRKATNPFNHTLDRFSLISLWLALKALQIDPSIWDKSNSELDAIIFRANDFVDPGSSSILGMLSGIQQLSTHVKNFAAVCASAMEKTPSLGDFIASKNIPISLASISMNGDIPVSRLKPGYIGAYTVLSALDYSACLQRVGDKVEVIGKIIDVKLNKTRNGKPYIFVNFGDWRGNIFKISIWSEGISALPSKPDASWIGKWISVIGLMEPPYVSGKYKYSHISITVTTIGQMTVLSEPDARWRLAGPNESRQTLTSTSSNQEALERIKSKSTTSTPMPMNTNATTANQAILNKLRASTQTVAAARAQTQHVVPNKSSTHYVAPTGTSASQPVQNIPSPASTSKQQTSQKNIVTKILKWLFG* (SEQ ID NO: 361)
61	MNEHLSHMDVHTLFEEMDEQADGITFKYSFDDIAKSNALVVTEFVNFERDSTVALLASLLTLPAHQSQCLRFELLTSLALIHCKGQQIANIDDVKRWYVTIGESSSIVGEDPAEDVFVALVDNKKGDYRVLEGVWEAAGFYTQLMVEIVSDMPDTHRYRSLKLAIQAILRLSDVICARSGLYRFQEGADEFPDSLDTAGLDEKTLCSRVTLSERSLRAEGIKLADLAPFILEPSHISMLGNQVPGEGMLEQRPLLRTRDGIVVVLPTAMTIALRQAVITFAKRTEELSELDKALANVYSLTFSEMPVFGNGGRLRRLTWEKYKMSRTTMVTSIVDAGHLMVLQFVLPSIQQYADTGFNNLLQLDEETTQFLDNSVEQITVDLAKQPGFQRGIVVRIACGWGAGFMGVPPQLPDGWGFEWMSGADFVRFGALPDMSPIAFWRVQDAVETIRQAGVRLINMSGTLNLLGWIRANDGHMVPHDQLPDDRITPEHPLMLMIPTNLLRGIRIAADTGYDRHRISDNNGKWHRVMRPSAEDFFPTERQSKCYASIDDLEAQRLTCVYEGQGNLWVTLEAPEMEDWMLLVELAKMVRTWIGRIGEALEVLSEQPIKKSLKVYLHFDGNDNIGRFDGENFSDDMNTFWRLERIHEHGAIRVVLQDGYLAGFRLPDNRAERALVRALGTAFATLLRMKEPVDKGVTVEQIAVPNDRARSFHIMQAYDFNQYLGRSLTKRLLAIEDIDSAAARIELAWRAVSTDAPSRYQGKKEVGKLLNDVVDVLIQDLLSELSRFDRKQTVMRLLENVVKARCEEAHWRSTAAAVLGLHAGEEGVEETIAQEMSRYAGAALTSRLIIELAICVCPTSGGIEPSDMALSKLLARASLLFRIGGMSDAVRFGALPADIRISPLGDLLFRDELGKMVLEPMLSKVTNERFEEQAAQFEQHYVKTAGGDDENSKQDSVAAETTEDQTDIFLAFWKAEMGFTLEDGMRFIQFLESIGIEQESAIFEMRRSQLADAAKSAGLADETIDAFLNQFILSARPKWDVVPDGFDLSDIYPWRFGRRLSVAVRPLLQIEESHDPLIVIAPGLLNLSLKYVFDGAYTGQFKRDFFRTEGMRDTWLGGAREGHTFEKTLERELREIGWTVRRGIGFPEILRRNLPGDPGDIDLLAWRSDRNQVLVIECKDLSLARNYSEVASQLSEYQGDDIKGKPDKLKKHLKRVLLAKENIDNFAKFTSIANPEIVSWLVFSGASPIAYAQSKIEALAGTNVGRPSDLLNF* (SEQ ID NO: 362)
62	MVGSRWYKFDFHNHTPASHDYKIPDISPREWLLAYMKQHVDCVVISDHNSGAWVDVLKGELENMSRDASTGDLPEFRPLTLFPGVELTATGNVHILAVLHTHSTSADVERLLAQCNNNSPIPSEVPNHQLVLQLGPAGIISNIRRNPKAVCILAHIDAAKGVLSLTNQAELTAAFQESPHAVEIRHRVEDITDGTRRRLIDNLPWLRGSDAHHPEQAGVRTCWLKMSSPDFDGLRHALLDPENCVLFDQLPPEEPASYLRSLKFRTRHCHPVGQDSASVEFSPFYNAVIGSRGSGKSTLIESIRLAMRKTEGLTATQGSKLDQFIRTGMEADSFIECIFHKEGTDFRLSWRPDSKHELHIFSDGEWMPDSHWSADRFPLSIYSQKMLYELASDTGAFLRVCDESPVVNKRAWKERWDQLEREYLNEQITLRGLRARQGSADSLRGELSDAERAVSQLQSSAYYPVCRQLALARNELSAATLPLEHFERRIAAIQALAEEPLQRSDIPPEPSGLLMAFMARLSSVQQQYDQRLNTLLAEYAAELAGIRREQSFIALRTAVSDQETNVESEAVSLRARGLNPDVLNELMARCESLKNELRNYDGLDGAISASVARSEQLLAEMRAHRMALTDNRKAFLSSLSLSALEIKILPLCAPYEDVISGYQTVTGISNFAERIYDNSDGSGLLSDFISERPFSPLPAATEKKYRALDELKALHHSIRLDNSEAGAGLHGSFRNRLRSLNDQQLDALQCWYPDDGIHIRYQTPGGQMEDIAFASPGQKGASMLQFLLSYGTDPLLLDQPEDDLDCLMLSMSVIPAIMSNKKRRQLIIVSHSAPIVVNGDAEYVISMQHDRTGLYPGLCGALQEAPMKALICRQMEGGEKAFRSRYERILS* (SEQ ID NO: 363)
63	MDYLSEVLKIIEGATKANASMASNYAGLLADKLEQKGEVKQARMIRERLLRAPQALAGAQRAGGGISLGSLPVDIDSRLNTVDVSYPKLDSSEIFLPAAISTRVEEFITNVQRYDEFVKADAALPSRMLVYGKPGTGKTMLSKYIATRLDFPLLTVRCDTLISSLLGQTSKNLRQVFDYVMQRPSVLFLDEFDALAGARGNERDIGELQRVVISLLQNMDAASEDTVIIASTNHEQLLDPAIWR

TABLE 15-C-continued

Row No.	Sequence
	Sequences of validated defense systems (Sequences encoded by the genes corresponding to rows 1-68 of Table 15-A)
	RFSFRIPMPLPDIHQRELIWKNRLKNMICSDLDLSDLSRKSEGLSGAIIEQVSLDARRDAVIEGASVINHHKLYRRLYLAQSLMEGVNLSTYEDEIRWLRSKDKKLFSIRVLANLYKLTSRVISNILKESGAYEQKGYTV* (SEQ ID NO: 364)
64	MSRRGTQFSNAKVTNPMLRIPFSSSDLGAIVNAGGGAKVLVDVTAEYRQGLVRNLTTSKHYLESKLSEYPGSLGTLVFKLRDQGIAKTHRPNKIAQEAGLQNAGHAKIDEMLVAAHAGCFDVLESVILHRNIKAILANLSAIERIEPWDENRKVPGGTDGLFESSNILVRLFEYTGEDATYNNYENVISILEQHGVKYDEIRQKCGLPLLRIMDLSPNDRYILDILIDYPGIRTLIPEPKYSAFPVSVSDSVGIETNSFPVPSEELPIVAVFDTGVSPIAATITPWVVSRETYVIPPDTSYEHGTMVSSLISGAHFLNDNHPWIPDTKSKIHDVCALDENGSYISDLILRLADAVNKRPDIKVWNLSLGGGPCNEQTFSDFAMELDRLSDKFGILFVVAAGNYVDEPIRTWPNPDPLGGADLISSPGESVRALTVGSVSHMEANDALSEIGTPTPYTRRGPGPVFTPKPDIIHAGGGVHRPWNVGASSLKVVGPDNRLCSNFGTSFAAPIVASLAAHTWQRIATNTDFNVSPSLIKALLIHSAQLSSPDYSPSERRYLGAGIPNEVIETLYDSDDRFTLIFQTFLVPGVRWRKDNYPIPSALIQNGKFKGEIVITAAYAPPLNPNAGSEYVRANVELSFGLIENNTIKGKVPMEGENGQSGYERAQIEHGGKWSPVKIHRKAFNKGITSGNWALQAKTTLRANEPALMEPLPVTIVVTLKSLDGNTQVYADGVRALNANNWAHYPLPARVPVSV* (SEQ ID NO: 365)
65	MKTVRSACQLQPKALEINVGDQIEQLDQIINDTNGQEYFKKTFITDGFKTLLSKGMARLAGKSNDTVFHLKQAMGGGKTHLMVGFGLLAKDAALRNSHLGSMPYQSDFGSAKIAAFNGRNNPHSYFWGEIARQLGREGVFREYWESGAKAPDEQAWINIFDGEEPILILLDEMPPYFHYYSTQVLGQGTIADVVTRAFSNMLTAAQKKKNVCIVVSDLEAAYDTGGKLIQRALDDATQELGRAEVSITPVNLESNEIYEILRKRLFLSLPDKNEVSEIASIYASRLAEAAKAKTVERSAEALANDIESTYPFHPSFKSIVALFKENEKFKQTRGLMELVSRLLKSVWESDEEVYLIGAQHFDLSIHDVREKLAEISEMRDVIARDLWDSTDSAHAQIIDLNNGNHYAQQVGTLLLTASLSTAVNSVKGLTESEMLECLIDPNHQGSDYRNAFTELAKSAWYLHQTQEGRNYFSHQENLTKKLQGYADKAPQNKVDELIRHRLEEMYRPVTKEAYEKVLPLPEMDEAQATLRSGRALLIISPDGKTPPGVVGNFFKGLVNKNNILVLTGDKSSIASIEKAARHVYAVTKADNEITASHPQRKELDEKKAQYEQDFQTTVLSVFDKLLFPGNNRGEDVLRPKALDSTYPSNEPYNGERQVVKTLTSDPIKLYTQINENFDALRARAESLLFGTLDEARKTDLLDKMKQKTQMPWLPSRGFDQLAIEAYQRGVWEDLGNGYITKKPKPKTTEVIISEDSSPDDAGTVRLKIGVANAGNSPRIHYAEDDEVTESSPVLSDNTLATKALRVQFLAVDPTGKNLTGNPTTWKNRLTLRNRFDEVARTVELFVAPRGTIKYTLDGSEARNGETYTVPIQLADQEATIYVFAECDGLEEKRNFTFAAAGSKEIPIIKDKPATLVSPSPKRMDSSAKTYEGLKIAKEKGIEFEQISLMVGSAPKVIHISLGEMKISAEFIETVLTHLQTVLSPEAPVVMTFKKAYTQTGHDLEQFVKQLGIEIGNGEVEQR* (SEQ ID NO: 366)
66	MNKTVDFGAPSEFGMHHFYVEIPAAPRDAVVIYEDYGFDGEDSRRETVECRLILARELWTKIRDDVRRDFNARLKIKKQSSGTWSTGKVKLDRFLGRELCVLGWAAEHASPDECLVICQKWLALRPEERWWLYSKTAAEAGRDDQTQRGWRKALYCALSDGANIKLETKKKPKSKKLQVEDETQDLFGFMEKGEF* (SEQ ID NO: 367)
67	MALQPFEWRDKPSLIEHLFPVQKISAETFKERMASHGQLLVSLGAFWKGRKPLILNKACILGSLLPATDNPLEDLEVFELLMGIDSESMQKRIEASLPASKQETIGDYLVLPYAEQIRIAKRPEEIDESLFVHIWNRVNNHLGTSAHTFAQLVEELGVARFGHRPRVADVFSGSGQIPFEAARLGCDVYASDLNPISCMLTWGALNVVGASAQKRVEIDKAQRDIVKKVQKEIDELDIESDGRGWRAKVFLYCVEVTCPESGWRVPLIPSLIISNSFRVVAELKPVPAERRYDISIREVSTDEELEFYKSGTIQDGEVIHSPDGKTQYRVNIKTIRGDYKEGKENLNKLRMWEKTDFAPRPDDIFQDRLFCVQWMKKKPKGSQYYYEFRTVTNDDLKREKKVIEHVASKLDDWQKQGLVPDMVIEAGDKTDEPIRTRGWTHWHHLFHPRQLLFLSLVNKYSLAEGKFNFLQCMNHLSKLTRWRPQAGGGGGSAATFDNQALNTLYNYPVRATGSIENILAAQHNHCGISENVSFVVNSHPAPELDVENDIYITDPPYGDAVKYEEITEFFIAWLRKNPPKEFAHWTWDSRRSLAVKGEDEGFRTGMVAAYRKMAQKMPDNGLQVLMFTHQSGAIWADMANIIWASGLQVTAAWYVVTETDSALRGGSNVKGTIILILRKRHQALETFRDDLGWEIEEAVKEQVESLIGLDKKVRSQGAEGLYTDADLQMAGYAAALKVLTAYSRIDGKDMVTEAEAPRQKGKKTFVDELIDFAVQTAVQFLVPVGFEKSEWQKLQAVERFYLKMAEMEHQGAKTLDNYQNFAKAFKVHHFDQLMSDASKANSARLKLSTEFRSTMMSGDAEMTGTPLRALLYALFEISKEVEVDDVLLHLMENCPNYLPNKQLLAKMADYLAEKREGLKGTKTFNPEQEASSARVLAEAIRNQRL* (SEQ ID NO: 368)
68	MAIKRFSSRTERLDTEFLAESLKGAAKYFRIAGYFRSSIFELVGEEIAKIPEVKIICNSELDLADFQVATGRNTALKERWNEVDVEAEALLKKERYQILDQLLHSGNVEIRVVPRERLFLHGKAGSIHYADGSRKSFIGSVNESKSAFAHNYELVWQDDDEESADWVEREFWALWTEGVPLPDAILAEIHRVSNRREVTVDVLKPEEVPAAAMAEAPIYRGGEQLQPWQRSFVTMFLEHREIYGKARLLLADEVGVGKTLSMATSALVSALLDDGPVLILAPSTLTIQWQIEMMDKLGVPAAVWSSQKKVWLGVEGQILSPRGDASSIKKCPYRIAIISTGLIMHQREKTDFVKEAGMLLKNRFGTVILDEAHKARIRGGLGDQASEPNNLMAFMLQIGRRTRHLVLGTATPIQTNVRELWDLLGILNSGAEFVLGDALSPWHDHEQAIPLITGQTQVTSEAEVWHWLSNPLPPSNEHHTVQQIRDYLSIDNKSFGYSHRFEDLDYMIQSLWLSECMTPSFFKENNPILRHTVLRKRKQLEDDGLLERVGVNTHPIKRNLAQYQSRFVGLGIPTNTPFQVAYEKAEEFSKLLQSRTRAAGFMKSLMLQRICSSFASGLKTAQKMLKHTVSDEDEDLVEDVEHLLSEMTPAEVACLREIETQLSRPEAVDSKLNTVKWFLTEFRTDGKTWLEHGCIIFSQYYDTAEWIAKELAKSLKGEVVAVYAGVGKSGLFRGEQFNNVERELIKSAVKTREILLVVATDAACEGLNLQTLGTLINVDLPWNPSRLEQRLGRIKRFGQTRKFVDMLNLVYSETQDEKVYNVLSERLRDTYDIFGSLPDTIDDEWIDNEEELNTRMDEYMHERKKAQDAFSVKYRGTLDPDAHLWERCATVLSRRDIVSKLSEPWGS* (SEQ ID NO: 369)

TABLE 16A

Additional tested homologs of predicted defense systems

System #	Name	Observed Activity	# Genes	Source Organism	Strain	Promoter	Codon	Gene A	Gene B
1	Retron-TIR	+	1	*Escherichia coli*	NCTC9024	Native	Native	STF89551.1	—
2	Retron-TOPRIM	−	1	*Escherichia coli*	NCTC13441	Native	Native	WP_000476153.1	—
5	RT-nitrilase (UG1)	−	1	*Escherichia coli*	N1	Native	Human	WP_001121606.1	—

TABLE 16A-continued

Additional tested homologs of predicted defense systems

7	RT (UG3) + RT (UG8)	−	2	*Escherichia coli*	NCTC9091	Native	Native	STJ76581.1	STJ76580.1
7	RT (UG3) + RT (UG8)	−	2	*Salmonella enterica*	NCTC6026	Native	Native	WP_001530977.1	WP_001185451.1
7	RT (UG3) + RT (UG8)	−	3	*Acinetobacter calcoaceticus*	NCTC7412	Native	Native	WP_000227776.1	WP_000620968.1
8	RT (UG15)	+	1	*Escherichia coli*	STEC66	Native	Human	WP_032207424.1	—
10	ATPase + adenosine deaminase (RADAR)	+	2	*Escherichia coli*	NCTC11116	Native	Native	WP_096949333.1	WP_001538182.1
13	STAND	−	1	*Escherichia coli*	NCTC10650	Native	Native	SQB54359.1	—
21	Transmembrane ATPase	+	1	*Escherichia coli*	NCTC8620	Native	Native	WP_048228060.1	—
22	ATPase + QueC + TatD DNAse	+	4	*Escherichia coli*	ECOR10	Native	Native	WP_000269401.1	WP_000537316.1
23	DUF4011-helicase-Vsr-DUF3320	−	1	*Citrobacter braakii*	NCTC9067	Native	Native	WP_115191085.1	—
28	ATPase + protease (ietAS)	+	2	*Escherichia coli*	ECOR12	Native	Native	OWD36540.1	OWD36541.1
28	ATPase + protease (ietAS)	−	2	*Escherichia coli*	NCTC9008	Native	Native	WP_001460375.1	WP_020244573.1
30	Retron-protease	−	1	*Proteus mirabilis*	127_PMIR	Native	Native	WP_161800346.1	—
30	Retron-protease	−	1	*Yersinia aleksiciae*	404/81	Native	Native	WP_054888011.1	—
30	Retron-protease	−	1	*Yersinia bercovieri*	3016/84	Native	Native	WP_054872116.1	—
30	Retron-protease	−	1	*Yersinia enterocolitica*	ST5081	Native	Native	WP_050337179.1	—
31	RT-nitrilase (UG5)	−	1	*Escherichia coli*	NCTC4169	Native	Native	WP_001521910.1	—
31	RT-nitrilase (UG5)	−	1	*Klebsiella pneumoniae*	KPNIH39	Native	Native	WP_023301376.1	—
32	TOPRIM-RT-nitrilase (UG10)	−	1	*Pseudomonas rhizosphaerae*	DSM16299	bla	Native	WP_084139843.1	—
32	TOPRIM-RT-nitrilase (UG10)	−	1	*Vogesella indigofera*	DSM3303	bla	Native	WP_120809745.1	—
33	RT (UG7)	−	1	*Escherichia coli*	NCTC9069	bla	Native	WP_000064054.1	—
34	RT (UG9) + PolA	−	2	*Photorhabdus* sp.	CRCIA-P01	lac	Native	WP_118986603.1	WP_118986604.1
34	RT (UG9) + PolA	−	2	*Pantoea* sp.	B40	lac	Native	WP_042677494.1	WP_128574327.1
34	RT (UG9) + PolA	−	2	*Vibrio litoralis*	DSM17657	lac	Native	WP_051241322.1	WP_083962817.1
34	RT (UG9) + PolA	−	2	*Pseudomonas brassicacearum*	Wood1	lac	Native	WP_080587824.1	WP_027911782.1
35	DUF4297-STAND	−	1	*Escherichia coli*	NCTC9036	Native	Native	WP_060615938.1	—
36	DUF4297-STAND	−	1	*Salmonella enterica*	NCTC10718	Native	Native	WP_115407481.1	—
37	ATPase_GHKL + Helicase_SF2	−	2	*Pectobacterium wasabiae*	CFBP3304	bla	Native	WP_005974598.1	WP_005974600.1
37	ATPase_GHKL + Helicase_SF2	−	2	*Vibrio harveyi*	ATCC43516	bla	Native	WP_061066216.1	WP_061066217.1
38	ATPase_GHKL-DUF3684-DUF3883	−	1	*Raoultella planticola*	NCTC9528	Native	Native	WP_112150151.1	—
39	TerY-P + helicase + HEPN + ATPase + DUF2357	−	7	*Obesumbacterium proteus*	DSM2777	Native	Native	WP_057631338.1	WP_057631339.1
40	Kinase-helicase	−	2	*Escherichia coli*	NCTC13919	Native	Native	WP_000877066.1	WP_001294844.1
41	Helicase-DUF559 + SMC + McrB + DUF2357 + ATPase	−	5	*Plasticicumulans lactativorans*	DSM25287	Native	Native	WP_132537919.1	WP_132537920.1
41	Helicase-DUF559 + SMC + McrB + DUF2357 + ATPase	−	5	*Yoonia sediminilitoris*	DSM29955	bla	Native	PUB10544.1	PUB10545.1
42	GTPase + GTPase + TM	−	3	*Pantoea cypripedii*	DSM3873	Native	Native	WP_084873987.1	WP_084873988.1
43	TM + GTPase + GTPase	−	3	*Escherichia coli*	NCTC10962	Native	Native	STI27515.1	STI27516.1

TABLE 16A-continued

Additional tested homologs of predicted defense systems

44	Dcm + HerA + Vsr	–	5	*Pseudomonas aeruginosa*	NCTC10727	Native	Native	WP_031690635.1	WP_004363346.1
44	Dem + HerA + Vsr	–	5	*Aquimonas voraii*	DSM16957	Native	Native	SDD97145.1	SDD97170.1
45	RecQ	–	1	*Klebsiella oxytoca*	NCTC11696	Native	Native	WP_032728854.1	—
46	Histidine kinase + phosphoribosyltransferase	–	2	*Pseudomonas aeruginosa*	NCTC13717	Native	Native	WP_003450792.1	WP_003450790.1
47	PH-TerB-DUF726 + TM	–	2	*Klebsiella pneumoniae*	NCTC11357	Native	Native	WP_126494466.1	WP_023316678.1
48	TerB + DUF2791 + Lhr helicase	–	3	*Escherichia coli*	NCTC9024	Native	Native	VDY98671.1	VDY98669.1

System #	Gene C	Gene D	Gene E	Gene F	Gene G	bp
1	—	—	—	—	—	2393
2	—	—	—	—	—	2569
5	—	—	—	—	—	4154
7	—	—	—	—	—	3648
7	—	—	—	—	—	3818
7	WP_000837118.1	—	—	—	—	4236
8	—	—	—	—	—	1951
10	—	—	—	—	—	5533
13	—	—	—	—	—	4781
21	—	—	—	—	—	4037
22	WP_000192874.1	WP_000020778.1	—	—	—	4891
23	—	—	—	—	—	6502
28	—	—	—	—	—	3678
28	—	—	—	—	—	3917
30	—	—	—	—	—	2009
30	—	—	—	—	—	1946
30	—	—	—	—	—	2032
30	—	—	—	—	—	1996
31	—	—	—	—	—	3679
31	—	—	—	—	—	3479
32	—	—	—	—	—	7494
32	—	—	—	—	—	7656
33	—	—	—	—	—	3894
34	—	—	—	—	—	3208
34	—	—	—	—	—	3211
34	—	—	—	—	—	3196
34	—	—	—	—	—	3382
35	—	—	—	—	—	6514
36	—	—	—	—	—	6261
37	—	—	—	—	—	10166
37	—	—	—	—	—	10210
38	—	—	—	—	—	5918
39	WP_057631340.1	WP_057631341.1	WP_057631342.1	WP_057631343.1	WP_080376085.1	12191
40	—	—	—	—	—	6873
41	WP_132537921.1	WP_132537922.1	WP_132537923.1	—	—	11931
41	PUB10546.1	PUB10547.1	PUB10548.1	—	—	11041
42	WP_084873989.1	—	—	—	—	4789
43	STI27517.1	—	—	—	—	4577
44	WP_004363343.1	WP_003131012.1	WP_071534163.1	—	—	11911
44	SDD97192.1	SDD97211.1	SDD97232.1	—	—	11635
45	—	—	—	—	—	5424
46	—	—	—	—	—	4088
47	—	—	—	—	—	3637
48	VDY98667.1	—	—	—	—	6037

TABLE 16B

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
1	Retron-TIR	atccctgaattccccgaaggtgaacaatccactgttcacccttcaccgtatattaacccgttatcacactgaaattaaaagagaaaaatgaaaggtgaacagtgtgaacaatcaaatcaaaaaaactttctactcccactatagcctgactggtcgtctccaaaacgagcggaaaagcatcaacaatgaatagttaactgttaactccgcgccaactcattaccacttaactcaatgatattaaatggaaaactatcgaaatgaatactctgcaaaattaaatgcaaaaaaatatatgccagtcaaatttcgttacgcactctcttccaagaaagagataaatgctttatacgtccaccatactatgttattttttaatacggctctgccttaaatctgtgaggttgtttcgcctcgaagtatcttatgttagcacatcacgctaccaatcagcggttagttacttgacgtaactgttaattggctaaagtttgcatagagtgattgggcggagccgtaaatttagtccataaatacagtaacgaggtagagagtgtctttacatgacaagctactgatgcttagtctcaattcggcgaataaagaagaagatgagacaatcccggagttacctaagttagagcctcagccctatcaagctggaaataagttgaaatgggataataaagagctgaaaaatcagcccatcacttcaaagaatgacattaatgtaatatgcaaaaaaattgaaacaaaagcattgtaattacatcagcaaacgatgtagccaatctgttagaagtcccggtcggacaattattatttattttatataataaaaaagataactatagaacttttgaaataaaaaagaaaaatggaaaaagtagaatcataaatgcacctcaaggcggtttatcaattctgcaagagaaattaaagccagttcttgagtacttttatcgccccaaaaaaccagcacatggatttattaaggataaaagtatattaacaaatgcagaaaaacatacaaagaaaaaatatgttgttaatgtagatttagaaaattattttggttcagtcactttcgctagagtatatgggatatttaaaagtaagccatttaatttctctcatcctgcggcgagtatattagctcaactatgtactaaggatggaaaattacctcaaggagcatgtacctcccctgttctagcaaatttagcatcagcctcactcgataaacacctaacccaactggcacgtagaaaaaacatcacatatacaagatatgcagatgatattactttttcatttaatcaacgacaagtcagagaaatcataacgctagataatgaaaataattttgaattgggcgaggcgattatctctgtgatagagaaaagtggcttcagcataaacacaagtaaattcagagttcagaaaagaaatgaacgtcaaaaagttactggtctagtggtaaatgaaaaagtaaatgttgagcgtaaatatcttagagttactcgttcattagttcataaatggagagaagacaagttaacatcagcattgttgtttgttactaaaaaaggttttaaggcaacaaataacgaacatgctatatcaatttttcgcaatcatatttatgggcgattgagttttataaaaatgatccgtggtgaggacttcccgttatatcttaaattaatggctgaaatgagtcatcatgatcctttaaaaacaaaagaagggcttagagcaatgaaagaaactgaaacttacgatgtatttatttgtcatgcaagcgaagataaaacatccatcgcaattccaatttacgaagaattaattaaattaaatatatcaacattcatagatcatgttgaaataaattggggcgattcattaatccaaaaaattaactcagctcttgtaaagtctaaatatgtaattgccattctttcggctaattctgtagataaacattggcctaagaaagaattgcattctgtgcttgcaagagaaatcactgaaggtgaagtaaaattacttactcttgtaaaagaagcagatgaagcaatagttgctgaatctttgccgctcttaagtgataagctttatatgacctataaagataatccggcagaagttgcagataaggttcgtgcgcttttaaacaagtgacagctactgtcaaatgtgtataaagtcattgatattttatataaaatcaatggattgcaatccatataagattccttatgcatcagtgacccggtgctcgcccggtcactgcttcagtcccagcagaactcagacgaggcgcttaacatctaacgggatgccaacccgacgtttggttttatcggctatctagcctatatagaagca (SEQ ID NO: 370)
2	Retron-TOPRIM	cacgtaaatatgaaaactgttagcccacatagcccaacaaaaatatttgatagttaaccttctgttactaaagaaaacaggaaagtaaaagtgggctaaagcttatgcgccctcgatgttgggctagccccaaaaacggtaaatttagcttaagtgcataattggttagctcaaaagcattatttttcatttaaataaattagttaattggtcttgtttagatgattcaactgggctgactactttctttgtatatactccggataaattttcccagctaacttgcctaatcatcactctgatgccagaaatgaacagaacgcaaaccatctataacttattgaggattttgaaaaaaattgattgggggcttgagttatatgatgactatgctaatttaatacggcacatgcaggtagatttgttggttgtggtatcgcaatcagtgttaacaaggtcgggagtattcgccctctgactgccgtcaagtcatcttggcgtcaccgttaaatgcgtaagagtacctgcatgtgcattaacataatcaataatggaatttactgttatgtttaaacctacctatctggcaaggctgcaggcttgttgtaacaaatttgaactggctgatttgcttcagattaaagttacatttctgactaatgttttgtatagaataaggccagaaaatcaatacaaaaaatttactataaagaaaaagtctggaggagagcgggagatctttgctcctgatgaaaaactgaaagatattcaacaacgactttctgaacttctatatatatgccaggaagaaatttgggcaaaaaataatattaaacaaaatgtatcacatggttttgagaagaataaaactataattacaaatgctgagaggcatcgagataaaaatattgtatttaatattgatattgagaatttcttcccatcctttaattttggtcgcgtgcgaggatattttattgcaaaccaaaatttcaagttacatccaaatgttgcaaccattattgcgcagatagcctgcctggatggatcgcttccgcaaggaagcccttgttctccagtaataactaatcttatttgtaggattttagatttcagattatcaaagctagcagtcacatatggttgtagttacagccgctatgcagatgacattacgttttcaacaaacaaaaaaaacatccctgatgcattagtttctaatgagaaagaaaacgaaccaggtaagatattggtagaagaaattcatcgtgcaggcttcactttaaaccataataaaaacagagtgtctaggtgtacatcaagacagcaagttacaggtttaactgtaaataaaaaaaataaatgtaagcagagagtatataaagaatacaagagcgatggcgcattctttatactttgaaggttcgtatacacttattgagaaagatggaaaacatagaaagggcacccttagtgaattagaagggcgatttgcatttatcgatatgcttgataaatataataatgtggaagcaaagaaaaatgcgcgtcctgagagatatgtggttaaaggatttgggttggattttaagcagagacttaactccagagagaaagcatacagcaaattcctatactataaaaatttctatggaaatgagcaaataacaatcttaacagaagggaaaactgacccggtttatcttaagtgtgcaattgattctttgtttttggattaccctcagttagttagagaggaaaaaaacacaaagaatagagtgttaaaagttaatttatttaaaaccaatgacaagaaaaaatattttctcgatttgtctggtggagctgcagactattcgaggtttttcagacgacatggtttactttgtaaagcgtatgaaaaacagcctcctaaaaatccagtgataattttattagataatgacacagggccatctgacttcataaatcaaataataaaggattattcgcatctaccaaaaaaagcggaggatgttagaaaaggggcgttttatcacttagagagtaatttatatgttctttttactccgttattaccaggggataactattcttcactagaggatttttttgaaccaaaagttttgcaaatgaagtataatggaaaaagcttcgataaaagcaataatcatgacagttctactacatttggaaaagatagatttgctacttatatagtaagggaaaatagaaaaactatcgatttttcattattcaaacccatacttgattcaattattgaaatcaaaaaacattttatcaatctacacccatcaaagtgatggttatgaaaagagataaaaatgctgatgtcaaaagaggcttatgctcggcacagtggagtgagctgccaaactgtcgatgactgggtagccggtggggcggaagtagttatgtcccgtagcaaggttaagatttgctcttgtgtgtggggaaccttagtcaattactttcctggcgcactgtgttagattttgtaaaattttaaaagactaaagatttaatatcacttctccatggaggttgtg (SEQ ID NO: 371)

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
5	RT-nitrilase (UG1)	gtggcaagattataccccatcaggcataagatgctttgacttataacgcatcagtttgaaacacaatggtgatgggggtcacaggggctgacatgtacttttaagattaaaaagcattaacatctactttgaagaaaacagaaaaaaacaatcacaaacctttaaaaacaaaaactatgccaattattaataaaaagtatcaagagcttcagttaacagatgagtacattaccgatccactgctcatggccctagcctggaagaaaagccatcactacatacgtaccacaaattggtatgctgacaactttgaactagacctgtcggctttggacctaatgcagcactgtaaagattgggtcaagagaatgcaggacaaaaaagaatttaaattttcagagctacaacttgttcctgtaccaaaagcctgtaaatgggagtttaagactgtcgaaaataaggttctatggcaaccttgtgatgaaaaagaacttaccctacgccccccttgcccatatacccatagctgaacaaaccatcatgacattagtcatgatgtgcctagccaatacaatagaaaccaagcaaggaaacccagacaccagctatgacatcgtccaccagaaaggtatcgtcaattacggaaatagactttattgtcagtatattgacgataaagcagagcacagcttcggtgcaacagtgacatatagtaaatacttcactgattatcggaaatttttaaataggccttatcattttgcgtcaaaagcgcaaggtgaaatttcgccggacgaagccgtttacatcatagaactagatcttgcgaagtttttcgatttagtaaacaggaagactctaattcaaaagataaaaaaccatatcagtgagtcaataaacaataaagaaaacccactcgccaatcatttatttaaatgttttgcaaactgggactggactgcatctagcataaaaaattatgacatatgcaagtcagacgaagtaacagaaataccaaaaggcatccctcaaggattggttgcagcagggtttctatcaaatatttacttacttgaattagatcaattcttgcataataaaattaacacagacataactgatgacattaaatttgttgattactgtcgatatgtcgatgacatgcgatttgtggttaaggttaaaaaatcaaaaaataataataccgcattcataaatgatgtaataaccaatcttcttaaaaatgagatagatataatcttggactgataattaatcctaaaaaaacaaaagtagaaatttttagaggcaaatccgcaggcatctcgcgtagcttggaaaacatccagaccagattaagcggcccaatatcaatggatagcgccaacgaacaacttgggcatcttgagtcattattaagtctgacaaaaaccgattttgaaccaccgaaaaatggtaaatcaaatagattagctgagattgaaaaagaccgtttcgatgtcagggaggacactcttaagcgcttttctgccaataaaatcagtaagatactaaaagagttaagacatttcatctcgcaggatatagatactgatggggaggttattgccggggaatgggattatctgcaagaacgtttggcacggcgttttattgtctgttggagccatgacccgtcactggcactgctactcaagaaagggctggaacttttccctgatcctaagctattagaccctatacttgaacagctttgctcactcattgaaagcgataatgaaaaacaaagtgcagtagctacttattgccttgctgaaatatttcgacattcagcaatgactattcataaaaaagacacctatgcattccctgcacaagccaatgtggatgggtactttgaaaaaatacaacattgcgccgcgacattcattaataagcgcagcgcctctgacaacgaaacttggaacctgttaattaatcaggctagttttctgttgcttgtgcgtttagataatacattagaaaaaaatggcactgatgccaggcatgatcttatcttaaaactggcatcaggctttagaacaattacacttcccactaaaatggatagcaagactatagcctcatgtattttgttggctagtcaattagttaaagataacaaaccatttattcgctcctgcgcttctttgtgcgaaagaatttatgacaaagaacacgtcataaaattgaagaaaatagttagcataatatcacatcaaaacttatcattgtttaaatccttagtttatcattcacgacctttacaacagaagtggctaaactcagactccgtgaaaataataattaatgaatgccatatagatatacaacctttggcgacttctttaggcatgataaaaagtagtcactcattacttagaatcatatcaagacctgataacccatttgccaatgagataatggcattaaaactgatgcaagcccttttattggacaggattgtttgcctggataataaaaaagattatcaaataagtgtagcaaacaccaaagtgacgtttcataactactccaaccctccaacatcgaatgtcttcgatgcaggaatggatatggatgcaaaattattcaaatcatcgggatgggtcgattctattttcacggatgatgcagacactcaaatattgtatagagttgccatgtgcatccgttcagtactactcggcaaacaagactggacagattttggtcaagcaatttcccccaaacagggttatcggggtattaaaactagtagagacaaacgtcaattggggatgatgacaacacctgagtccattgccggtgagaactctcaggtttctggttggcttaccacactcttatccaagttgcttgcctggccgggaatttcagtgggtgataatggatatcaatggccagcaatttttacagtagatgctgtcagaaaactagttgatgctcggctgagtaaacttaagcaggattactgcaaactatcaggaactccgggacttacagaaaaaatacagttcaactggtctgactcgaaaaaagccctaacagttgctatggtccagtcaaaactgcctgcaacgaaagattttgtcagccatggacttcttttaaactccgcaaagtatagagtgattcatcgcagacatgttgctgaagtggctgatttagttgtaaaacacacgcttgcacaaaaaacaactcaacgaactcatggtgaaaaaatagagaacattgatttaatagtatggcctgagctcgctgtacatagtgacgatttggatgtactcatcgccttatctagaaaaacgaatgcaatcatatactcgggcctgacatttattgagcaacctggaatcaaaggaccaaataattgtgccgtttggattgtcccacctaaaagcaatagcagccagaaagaaatgataagacttcaaggcaagcataatatgatggaagatgagaaaggccgggttgaaccctggagaccataccaattgatgcttgaacttgttcacccccaatttactgataaaaaaggatttgttctcacaggctccatttgttatgacgcaaccgacatcgcgctaagtgcagatctcagggataaatcaaatgcttatcttgtagcagcattaaacagggatgttaatacattcgattccatggttgaagcactgcattatcatatgtaccagcatgttgtgctcgttaactcaggggaattcggaggatcttacgctaaagcaccttacaaggagccgtttaatcgtttgattgctcatgttcatggcaatgatcaggtagctataagtacgtttgaaatgaacatgtttgatttccgtcgtgataatataggaaaaagtatgcaatccgggttagataaaaaaactgctcctgcaggaatcataatgtaataaatattagatatttttatattagaggtgaggagatggcgtcacctctaatattttcgctgattgtatttagcatcaaataataaaggtacaattaatttaagtgactatcatgaaaaaattagttccgccatatcaagtaaccccggcacaaatctatcgttccgttgccagttctacagccattgaaaccggaaaac (SEQ ID NO: 372)
7	RT (UG3) + RT (UG8)	gcgttgaatggtataactatggcacggttaccgcatgttttgagctgtaatcgaagttatgaaaattgctatataaagcggtcgctgttgtggagatacgattgcgggaagtgatggaaagagctataaaaagtacagaggatagtttaatgagggtattatgaaccgtcagccgtttacttcagcagcacttaaacgaaacttaagtgaaagtgagaaggcttattattttaaaaaaaataatgttgctgagttagaatcattaattagtgatgccgttttaattgctaatgagaattttcgctctggtgtgagtgtaaagaaactaaatattaaggacgctgcgtttacactgcttcatgtttgaaggaaaaaataatacttagacattgcaatgcaaatttaaaatgccttgaatcgcttcgtcccaaacaacgaaatacaataattagtgagcttaaaatttatttggaagaaggtactccattcaaaatatatcgtttggatataaagtctttctttgaatcaattgatttaccgcagctttttcagctcttacataacgaaacacgactgtctagacatacaaaaaatttgctagaatggtatcttaaatcgtgtgaaaggcttcactcttcgaaaggattacctagagggttagaaattagtcctatgttatcagaattgtacttggcacaatttgataatagtattcataggcatccagaagtatttattattcaagatttgtagatgatatggtaatcgtttcaagtggttgtgaatgtgaagcgtcctttatggaatttatacaagatgtattaccaaagggattggctttaaataaaaataaattaaaaatatctccatgcataccaaagagaagtaagggtttaaataaacaggataaattgcttcatgaatttgactttctagggtactcgttttctataatagacacacctttgagcaaagatggtgagattaatagctgttacagaaaggttgttgttaatttatctaaatctcgcctgaagaaaattaaaacaagaatagctaggtctttctactcttatcatattaatggtgattttaaactattgctagacaggatttctttttgactagtaacagggatttaaatcgcaaaataaaatcgttaagttctttagaaaaa

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		agcaagataagtacaggtatttattacagtaatgcgaagttagatgttgactccatatccctaaaaaaattagatgacttttgctatattgtgtgcaatctaatactgggcgtttgaatagtgttgcaaaaaaaccttttaatttgaagcaaaaaaaagaactgctaagaaatagttttagaaaaggctttgtggatagagtatatagaaagtataactttaagcgctatactgagattacaaaaatatggttataaagaaaaacattaaacttgataagaaagattatctcagggctttactatgtgatacactgcccggtgattgtccaattattttttcaaatgatggcttatatataaacttaacagaatatgatagagtttgtaatgatttgttacattttactccggtttcttctttcttaaaaaaaatagttaaccctaatttagactcttctattagtgtcgcagatcgccaccgagaaaagaagaaacaaagctccccatttggctattgtatagtaaaagatgcctttagccaaagacatctttctttaattcacccaagatctcaaattaattattcggaattttataaaacatactcatccgttatcacattaaatactttaaaagtaattttctattcgctacccacgtaaggtcgctaactctttctttttatatgaaaataatgctttggaaaaatataaaggggaagatatcgaaacaacaaaggatgagttaatgaggaaatattcatcctcttattttagttatggcggtttcaacaggatatataaactatttcaaagtaagatgtttattgagcttgagaaaagattctcggtgatgtggatgttagatgtatcacattgttttgatagcatatatacgcattcggtttcttgggcattaaaaaataaatcatatatcaaaaaacatgttaaacacagcaatcaatttggacaagaattagatacactgatgcaacgtagcaataataatgaaacaaatggaatacctattggttcagagtttagcagggttttgcagaattaatatttcagcgaattgattgcaatattgagtcatgccttcttagtgaacatggatgggttaataataaagattatgttatattgagatatgtagatgattttattgtttttgtaatggtgagtcaagtgccgaagttattacaaaaataattaatgtgaagttaaatgaatataatctacaattaaatgtaaacaagcttaagaagtattctaggccattttgcactagcaagacaagtttgattgtcaaagttaatgaattaattcgcaatttagaaattaaactgtatgaaaaacgtgatagtggctttactttaaataaaataagaagtaagcatgatttaaagatatatgtaattaatcatgtcaagtctatatgcattgaaaatcaagtgtcttattctgatgtttcatcatatataatatcatctctttccaaaagattaatatcaataattgatatattacgagttcaagaaaatgaagatgatgtagatgtaaaaaaaaggattaaggacttaattttcacaataaccgatattatgttgttctttttcagtgttaacccaactgtttcatcatcttataaattatcaaagacaatggttgttgttaataactatttgaatgaaatatctagtgactatagtagtatttttatgactacgttagtgaatgctgcggaaaacattaattttggtgagaatgataatgggctgtttattgatgatttcatttcaattgaaaaggttaatttaatcttggctgctactttttttggagataattatcttataagtgacagtttttttcatggagttatacataaaaagaaattggactacttactataatctcactgctattctattttagaaacagaagatcattccgaaaattgaagtgtataatagagggtgaaataaaggaaatattaagttctaatatggatttgctgcaatcatcggaaaaggcacatttattttggatgtcatgtcatgtccatttgtctcaatagagacaaggcgtttttatatagaaaatatctcaagagctatgagccaaagctgaacagaagtcatctggagattgagaatgatttgcaatctctgcttcaaacatattggtttgtcaagtgggatgagttagatattgtgaaaatgattgagaaaaaagaattgaaagaaagctattaatttgataaatatgagtcgtggtcagtttcaaaatacttacgtcatcgtcgtcggtgtattttatatcgattatgaagacgatttcgctggaactgaaatcggcttgaatgcttaaacttaagctaaaaaaacagtttgagaccaaagcctaaattattaggctttggattttcaggttcagttgagagtaattgctgtctg (SEQ ID NO: 373)
7	RT (UG3) + RT (UG8)	agatacagtctccatcatactcagaggcgcataccccttacatatctcaggtttatctggcttaggctatgacgctaacccactagagaatcggagaaaagtaaagactgtttgatttgtgagcttgattgattgcaatttaagcgctcgacacagggcaggatgccaaacaccttcaacagagaggtcggtagctccagcatatgcaagctaacgttgctttggaacttcaactaagtaccaagagtggacggttccttagtatcaggcaagtatatgattgcacctagcggtgtaaagagttataaaaaagcataaaacgttgtattgtgagactttaatgaaccggcagccatttacttcatcagcacttaaacgtaatttaagcgaaagtgagaaagcctattattttagcaaaggaaatagcgaaaaattagaatcattaattaacgatgcagtattaattgccaatgaaaattttcgttctggagtcagtgtcaaaaaattaaacatcaaggggcgttgtgtttattccgcatcgaatttaaaagaaaaattaatactgaggcattgcaattccaatctgaagtgtctggaatcacttttgcctaaacaaagaaataaaataattgatgaattgaagctttatcttagagaaggcacacagtttagggtttatcggctagatataaagtctttttttgagtccatccagttgccccagctttttaaatatatgcatgatgagtcgagactatccaggcatactaaaaacctgctagaatggtatcttaaagcttgtgagcgtattcatgccacacaaggcttacctagagggcttgaaattagtccaatgctatctgaattatatttgtcagagtttgatcgcaatatcaatcgacatccagaagtattttattactccaggtttgtagatgacatggtgattatttcaagtgggaatgaagaccaaaagacctttatgaaacaggtagtggatttccttcctaacggtttgaaactaaataaaaacaagctaaacatatcccctttaattcctaaaagaagtaaaggggataataataatgataaattactccataaatttgatttccttggttattcttttgcagttatagatacaccattagcaaagaatacagtaaacatcatatatagaaagataattattgacctatcaagcggtcgattgaaaaaaataaaaacaagaatatcaagagccttttatgcatttaagaataatggtgattataagctattactagacaggatttcttttctaactagcaatagagatttaaacagaaaaattaaatcactgagttcaactgagaagaccaaaattagcaccggaatatattatagcaacgctcggcttgacgaaaactccaagacactaaagcaactggataacttttttaatttattgtgtaatgtcaaatagagggcgtttgaatagtgttgccaagcattctttaagtataaaccaaagaaaggaattattgcgaaataattttacgaaaggtttttctgcaagaatttataggaaatataattttcaacgttatacagagattactaaaatatggctctaaaaaagaatattaaacttgataaaaaggattataccagagctttgttgtgtgatacccaaccagcagactgtccgattattttctcaaatgatgggctttatgctaatttggcatattttgatgttaactataaaacatcaacagattttactcctctttcatctttcttaaaaaaaaataattaacccatcgttggacttgtctattacggttgatgaaagagagcagaaaaggaaaaaacagagcttccctttcggttactgtattgttaaagattcttttagcttgagacgtctttctttaattcatccgagatctcaacttaattattgtgagttttacaaaaattattcatcagttataacctacaattcatcaaagagtaattattcaataagatatcctaagaaagttgccaattcattctttttatatgagaagaatggagcggaaagatataaaggggaggatattgaaactactgaggatgaattaatgaggaagtactcttcttcatattttcgtatggtggtttcaatagaatatataaattattccaaagtaaatctttctttgaacttgaaaaaagattctctataatgtggatgctggatgtatcacattgttttgatagtatctatactcactcagtgtcgtgggctttaaaaaataaagcttacattcgcaagcatgtaactaacagtaatcagtttggtcaagaattagatacattgatgcagcgaagtaataataatgaaacaaatggcatcccaataggctctgaatttagtagaatatttgccgaattgatcttccaacgaatcgacaataatattgagttggatcttatggatgagcatgggtggaaaaataaaaaagactatgtgatattaaggtatgttgatgattttattgtgttttgcaataatgaatcgaatgcagaaataatttctaaaactattaatgtgaaattaaatgagtttaatctccaactaaataaaaataaattcaaaaaatattcaagaccattctgcactagcaaaacaggacttattatcaaagttaatgagttaattcaaaatttggaatcaaaattatacgaaaagcatgacggcaatattgttcttaataagataagaaataagcatgatttgaaagtatatatgattaataacattaagtctatatgcttagatagtcaggcttcttattcagatgtatcgtcctatttgttatcctcactgtctaaaagattaatagcacttatccatcacttttcttttgagaaaaataaagatgaagaatttaaaaaaatcaagatgtaatatttacactatctgatttaatgttattctttttagcgttaatccaacagtatcatcctcgtacaaattatctaaatcaatgatcattattaatgattatttgaaagggatttcaagtgattatagt

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		aatattttatgacatcattggtaaatactgctgaaaatatcaattttggtgataatgacaatggattatttatagatgattttatatccattgaaaaggtcaatttaattttggcagcaacgtttttggggataaccacctggtaagtgaatctttttttgatgggattttgcaccaaaagaaattagattactttacaatcatatccttattattctatttcaggaatagaaattcatttcaggcacttaagagtatagttgaaagaaaaattatagaattactatgtccagatatggatttgttacagtcttcggagaaggcacatttatttttggatgtaatgtcttgtccatttgtatcaataaaaacaagaagatttatatatataagatatctaaagtcttttgagccaaaaaatctaagaacccactctgagattgagaatgatttgcaatcaatgctccaatgctactggtttgtcaagtgggatgagttagatcttttaaagatgatagagaaaaaagaattgaaggaaacttattgatctgataaaacattaatgtggtcagtttcgaaatacttacgcattattggtaagataaaatcttatgttaccaataatgtgatttcgctagatttggaatcggcttaactgcttaaacttatgctaacagaattgcttaagacctaaccattctttggaatgagatggggcttccaggtccagttgagagtagtcactta (SEQ ID NO: 374)
7	RT (UG3) + RT (UG8)	cttgagtttgcgtaagataatttcgtgaaaattaaagcaattaatataaaaaatgtaattactagtgtgtacagatatgaaaaatgatagttataaaaccatatgaaaattgaagaaagagttcaatttttgccttgtcagtaacaaataggtagcttattgaaaaaagataaaaaattaacaaaaaatcaataaattcatatagaataaaaatattaaagaaatgaaataagtgtttgcttcatcagttttagggatacattaaagtggttgataaagaaaaatattatactggattaataaaagatataaaaatagtagcttatgcaagattcaataaaatacgtcgtttaaagagaaataatttttttaggattgttatctatttcggtagtttctatcttagttattatattatcaattgtagaaaaaaatttataatataaaaacaatgagtttaattccattgtttgaaccaaatatagaaatatggttcttttgtatacttgcttcaataattattctttgtatatctattgcactctctactatgaagattgatattgaaatagaaaggttaaataaaagtgcagttgaacttaatgaagtaaggcggaaaattgaatttaatattgagaatagtaattatcaaaatagtacattgtttgataaatatcttgaaataataaagtcagacttaataaatcatgatgaggttgattataaaataaataagtatttagtcagtaaagttggtagtaagtttgcttattatcgaatgtattttattgatcagaattttacatcaatattttatcttttataacattttaagcttttcttcaattatttcaattattttgcaggtaatgttgaagtgataagacaagattttagtgtaaattccctgttgagaatcacaactaaaaatgaaattgttaaatttaacttgggtcgtaataaggaagagtatgctattgcattatctcaagtttctaattatctattagagggcaatgaaataatagataatttaagctgtagaatagaaagaaataaagttatatttagtactaattcaattaatacttttatgctttaaaaaaaatttctaaagatttaagccgattgtataaaattgagcctcctaatagagatgatatttctgaacaaatttatagaattttgaacactctacaagctatagtattgtaaggttagacattaaaagtttttatgaaaatattcaatataatgaggtaattaaaaagctggatagagataaaatactagttgcaaaatctattaaaattcttaaggatttatataactttattgataatggtttaccacgaggtttatctataagtcctattttgtcagaaatatttatgaaagaagtcgatcaacaaattagaaatatagatcatgtatactattatgctagatatgttgatgacataatagtaatttcaacagataagagtgattctatatatgaaaaaacaattaaagttttagagaaatatgatttaaatgttaatagtaagagatatataaaaaatattcctgctgtgaacaataatgaaatctcaactttatataagtttgattacttaggatataagtatattatagatacaatttcatataaaaataaacgaatagttaaagcggaactgtcagatgataaaaaaagaaaaattaaaactagaataatacatagtcttttagatagagtttataatacaacgcattatgatcgggaggagttgttaattaagcgattaaaagtgttatcctctaactactcaataacatataatgaattgtcaaaaactaatttaaaagctggtatgttttatagtcataggttagtaaataattatggtatttttagtgaatttaataaattttatctaaagctatctactgtcaacaaaacaatttctttggtaaagctatgtcgcagattcctagtaaagaaaaagaaaatattattaaaagtatttgttttgttagtggatttaaagataaaaactttattgagttagagagggttgaaatggaacgagtaaaaaagtgttggaaaaataaacgatataagaagctttgaggtaaaaatgaaaagtaagatttatttagataaaaaggattttttatagagtattgttaactgatgtattaccctatgaagtaccttttattttaagtaatgaaggtttttatagaaacttaaaaagcaactcatttcattcagttactaaaaaaatattagaattaactttatttacttcacaagtaaacactaatccttttaattttaaaatctctaaagatgatagtaattttaggaagttatatttagttcacccaagttcacaaataaaaatatcaaatttatataaaaattattatcaattaattacgcatttgtgtagtagaagttctttttcacttagatatccaacttatgttgcaaaagctttttatagtatagaaagagatagatctaattccgaaaattataaagatgaagatattgaattactgtcacaaaaaagccctaaatatgcaagtacttattttgtatataaagatatcagtttttatataaattctatgattcttatagatttcaccgtattgaaaaaaagtttaataaactattaaagtttgatattgctaaatgttttgactcaatatcaacatttcaattacctagatcagttaataaaaattgtagctttgaaagtcatacagatatacatagttttgaacatttattttcttcaattatgaaaggtgcttatcatggtaatacacatggtattgtaataggaccagagttttctagaattttcgctgaaattttattgcaatctatagatgtagcaataaaaaataagttaagaaatgaaatgggaattaaggagggtgttgattatgttataaaaagatatgtagatgattatttttttattttataataatgagcaaacttcaaatttaatttttgaatgtattgttgaagaactttctaagtatagactattttgcaatgaatcaaaaagtattaggactactattccttttattacaggtattactattgctaaacatgaaataaggaagagattagaaactttttttgaattatttgagtcaataaataataaagatgattatattgggctaaaattaaatcattattataaaatatcaaatcaattaattagtgatattaagtgtattgttttaataataatgtaagttattcaagtatttctggttattttttttactttaatgaaaaatcatgttttgcatataaaaaatagtttttcttttgaggataaatctaaagttgaaaatttaagtaagttatttcttattattcttgatgtttcgtttttttgtttactgtatgaattttaaagttagaagcacatatttaatttctcaaattatagttttgattagtactattgctgaatcatttgatttaaatttgatagatttaattaataaaaaaatatatgatgaggtggatttggttttaaagataaagtcaaattcaaacttattgaataatattgaaattttaaatctattaattgctgttagagatattgatcttaattatcagatcttagtagatgatcttatgttattgttttcttcagaaaggattaataagtataattatttctctttaatgactttttttattttatgttcaaaggaaaaaacagtatcagcctatcagagatagaatttatgcaataataattcaaaaatttaatcagaataatctaaatgtctcaaatgattctgagttaattcacatttttttgactcacttagctgtccttatttaactaaaaatcaaaaaattaatataactaactctgcattaaattctattattaaattaaatgataatgaaattgatgtttttgtagaagaaatgagcaaaactaattggtttattgactggaacttgcaaacaaaagatgcaattcagcgtttgctgatgaaaaaagaattgaaatcaccctatgaaaattgagataattaagctagaaactagatatacctccgacatttgttggttgatttttacacactatataactcctagtttctataaaaggatgtttctaacatccttttattttttttgagatttaatttttcttttagtgacaactaagttttactataactaatagc (SEQ ID NO: 375)
8	RT (UG15)	aattccccgaaaatccgcccgtttttactgaaaaaagccatgcatcgataaggtgcatggctttgcatgcgttttcctgcctcattttctgcagaccgcgccattcccggcgcggcctgagcgtgtcagtgcaactgcattaaaactgccccgcaaagcgggcgggcgaggcggggaaagcactgcgcgcaagctatgtgaggtgatgtgtaatacatatcacgaatagcgtaggtagctgttggctttgcctgatcaaggtgacagtatacatatcttaaaatataaatatttatgattatttatttgaaagaggttgaataatgattttttgatgaaaaaaga

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		catttatatgaagctctgctgcggcataattattttccgaatcagaaggggacgatttcagaaatcccaccatgtttttcttcaagaacttttacaccagaaatttgtgaattaatagt ttctaatgagccggggaaaagaaaattacatggatacgattgtgtcgaatactcatcgactaggtataataactttcccagagtattatccttaattcacccaagagcatatgcac agttagcaaagcatttgtatgagtcttgggatgagattcgaaaaatcaaagaaaataaaaacagtatgattaaacctgaaatgcatcctgacggtagactttttatcatgaattat gaggatgcagaaacaagaactgtaagggagttaaacgatggatttggaagacgatttaaagttaaaactgatatcgcaggatgttttaacaatatatattcacactcaattcctt gggctgttgtcggtgtgaataaggcaaagacatcaatgaataagcataaaaatagccaagatgttcattggagtgatagattggattattatcaaagacaaacaagacgaggc gaaactcatggtgtccctgttggacctgcaacgtcaagtattgtatgtgagataatattaagttccatagataatattcttgagaataaaggattcttattcagacgttacattgatga ttatacatgttattgtaaaactcatgatgaagcgaaagagtttctccatgttttaggtactgaactttctaagttaaagttatctctaaatttgcataaaactaaaattaccagtcttccc agtacattgaatgatgattgggtgtcgttgcttagtattaactctccatccaggagagtattcaggaataatgactcggatatattatctgcatctgaggttataagctttttggattat gcggtacaacttcatctgacgaatgggggcggtagtatattaaagtatgctatatctttaattattaataaagtagatgaggcgtcagcaagagagatgtacgactacgttttaaa tctgagttggcactatcctatattaattccatatttagatgtattgcatccaaagattaacattaatgatgaggtcaggttaaaacttaatgaggttttgaattcctgcatagataataa gttttctgatggcatggcttgggtgttgtattattgcttaaaatattccattgatattgacagttgtctcattagtaagatttttgaaaacggtgattgcctaagtatttgtattttggataa aactggaagatatgataaggaaatagaagaattttctaaaaatataatttcattggattatttgtatgaggttgataaatattggatattgttttatcagcgattctattcagggaaagg atataatccttacaatgatgattgttgtttcgatataatgaaaacatatggagttaattttatgcctgatgatggttatcaaacgaaagctgaacactattgtaatatagtaaatagtcc atttcttgagaatgatgaacaagtaataagttttaacgattattgttcataatttataattagcctccg (SEQ ID NO: 376)
10	ATPase + adenosine deamnase (RADAR)	actgctcgacaaaacgaaccgttcattcgcgaggatggtggcagtgaatgaggtggtcagttttatcagcgcttcaaggtagctttataggatggattgtagcgaagtgccca acaaattgattgaagctaagggcattgagcattgcatgcatcatgctcagactgacaaaaaatcaaaataaatggattgatacggacatgacagacagcgtacagactgaaa ctaccgagggaaaaatcatcatcaacttgtttgctcccaatcttcccggaagtaccaaagaagatgatctcattcagaaatctctgcgtgaccagttggttgagagtatccgaaa ctcgattgcttatcctgacaccgataagtttgctgggctaacacggtttattgatgagtccggccgtaatgtatttttttgtggatggtactcgcggtgcgggtaaaactacttttatc aatagcgtggtcaaatctctgaacagtgatcaagatgatgtcaaagtcaacatcaagtgtttgccgaccatcgaccccaccaagttgccgcgtcatgagccaattttggtcact gtgactgcccgtctgaataaaatggtgtccgacaaattaaaaggatactgggcgtcgaatgactatagaaaacaaaaagaacaatggcagaatcatcttgcacaacttcagc gtggtttacatctgctgacagacaaggaatataagccggaatatttcagtgacgctttgaaactggatgcccagcttgattactccattggtggtcaggatttgtcagaaatcttt gaggagctggttaaacgcgcgtgtgaaattctcgactgcaaagccattttgattacttttgatgatattgatactcagtttgacgcgggttgggatgtacttgaatctattcgtaaat tctttaacagccggaaattggtggtggtagcgacaggtgacttgcgtctatattcccaattgattcgcggtaaacaatacgaaaattacagcaaaactttgctcgaacaggaaa aagagagcgtccgcttagcagagcgaggctatatggttgaacaccttgaacagcaatatttattaaaactttttccggtacaaaaacgtattcaattgaaaacaatgttgcaattg gtcggcgaaaagggaaaagccggtaaagaggagatcaaggttaaaaccgagccaggcatgcaggatattgacgccatagatgttcggcaagcaattggcgatgctgtta gggaaggccttaatttgagagagggatcagatgctgacatgtatgtaaatgaactgctgaagcagccagtgcggttgttgatgcaggtgcttcaggatttctatacaaaaaaat atcatgccacatcggtaaagcttgatggtaaacaaagcagaaatgaaaggcctaatgagttatcagttccgaatttacttagaaatgccttatatggctcgatgctaagcagcat ttatcgtgcagggttaaattatgaacagcatcgatttggtatggattcgctctgtaaggacatttttacctatgtaaagcaggatcgtgattttaacactgggttttatttacggcctc agtcagaaagcgaagcattaagaaattgctctatttacttagcgtctcaggtgagtgaaaactgtcagggcagtctgtcaaagttcctacagatgcttttggttggttgtggctct gtcagcatattcaaccaatttgtgaccgagttagcacgagctgaaaatgatagagaaaaattcgaacagcttattagtgagtatgtagcttatatgtctgttggcagaattgaaag tgcctcacattgggctaatcgatgttgtgcggtggttgcaaacagccctaatgatgagaaaattggtgtttttcttggcatggtgcaattaaatcgtaaatcacgacaacacatgc ctggggggttacaaaaaatttaacattgatactgagaatggcctagcaaaagccgcaatggcgtcttccttgagtacggtagcttcaaataatcttatggatttctgtagtgtttttaa tctgattggtgctattgcagatatctcagcatgccgttgtgaaaggtcagccattactaatgcttttaataaagttatagctcagacaacatgtattgttccccccatggagcgaggc tgctgttcgtgcagaaatgaaaggctcaagtaaaagtgcagataacgatgctgctgttttggatgtagaccttgatcccaaggatgatggcgtgattgatgaaagtcagcagg atgacgcaacggaattttctgatgccattactaaagttgagcaatggcttaaaaacgtaaacgaaatcgagattggaattcgtccgtcggcacttttgattggtaaagtatggag tcggttctatttcaaccttaataatgtagctgatcaacataaaaccagactctatagaaatgcagagcatggacgaatggctagtcaatcaaatgccgcgaaaattatgcgtttta atgttttagcatttcttcatgcggtattggttgaagagagtttatatcattcggttagtgatagggaatatatcggtgaggggttaagactaaatccagttacttcagttgatgagttt gagaaaaagataaaaataattggtgagaaattaaaagcggataataaaacatggaaaaatacccatccattgtttttcttattaattagctgtccaattctacatccgttcatttttcc tgttggtgggattaattgttcagtcaaagcactgaacaaagaaacaagtttcaataagctgattgatgaaattgttggcgataaattactttctgatgaagaatgggactatctgac taaaaataatgatcaaaaaacaaacactagacaacaaattttcaaaatactataacatcgctgaattcctccacaatcgtcggagcatcatacgataaggatacaccagccag gaaaaccaagtcacctttattaggtgatagcgaagaaaaatgataatggccttcgtataaggattgggtatggaaaggtttcttcttaactcaacagttctgttatataggctaag cacagtctctttggatgaggtatcacttgatgagagagtggagtcatctgtattccttgctcaatacgaacaggctcgtagtttacctgatcatgtagctaaatctgcttggtcatat ttagtgcaacaaatcaaacagcggaatatgaaactcggcccagtagcaatcttacgcctgatagctgaaaagtttattaaaaacgagaaaggtggccccaaaatcgatctac ctatgttctcggaatggcaaacgctgatgagtcgagtatcgtgtctaccaattatagcgtgtcatcaggtatttaatccagggccagccagtcaggaatatagttttcgctggcct ttatacccatatcacccgacggttgaagactacattacccgtgaatgcttacatgaaactcaccaacacctaaatggcagtaccagtgcagaagagtgttggctggatgcact caaacacccagaagcatgcctcagagattttgagaagggctgggcatctcaagagatgaaacaactctgcgcccagattgatccatctctgacacctagaatcttcaaggat cgtttgcaaatcgcctgtaatattcgcgaaattctttgtcgggttgctcagggcgtggaattgccagagtggatagcatcaatgcaaaatccgcagcaactggcgaatagcac

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		aattctgcataatggccgggagtatgggtttgcgacagtttggccaattgacgacaaatacagtcaggagtctgagttttgctggctaaccggattgttggaaaaatggcggttt aatgcgccagaagggttagaacgattgctttggatttacctgctgattcaaaatcagtacttgaccttactggttcagcgagacgatttttcggatttgaacagttccagaattac accatgacggagttgagggaggaaacagagaaatcttatttgtctcgttttaaacatgctcatggtgcaggagtgtattctcaggtgcgttatctggaaggacgttttgctccga agagcgaccccaacaaaatgcaaaagctgctcttcagtgtgttaagaggatattgggaatatctgagtgctcatatgtccatggaatgggtgcatgaaaagcctctgactatat cgcaagtgctcgataacctcgaactggttgaacctcatggcaagtgtgtagagctggcgctagtgccgcactttatcaaaagaaagcccaaaaatggtgaggcctatcctca cgcattactattcaaagacctgaaaaatcaggcagctattctgatggacatgctgaagtctgaaccgcgtctgacaggctggattcgaggagtagatgccgcagctaatgag atgcacgcaccacctgagttattttgccccttgttccgggtactagccaaatcaggtattgctcattttacctatcatgttggcgaggactttccgcatctgatcagtggtattcgct ccattgatgatgccttgagattttaccattgcgtaatggcgatcgtcttggtcactgcacggcgattggtattacacctagcatctggaaacgctctttgccattgtccttatccat gaccaaagagacgagattgctcgatttggtgtttatctggcgggaacttcgaagtcatccggaactgctgcgttacgctagtgatgcagcgattgaagctgttcgcttggctca taaagtgttttcgctggaagaggaagtctcgattaccacccttgatcaggtatttgaaatgcgggggctgttggccgaatcggaaggcctactgagtgagctaaatgaaccatt aaaacccaaatccctctggttggaagagtatgagcgcgccagagagttggttaaaacaacgggtatgaaaaggccgttgaagttgtataagcaatggctaacatctgacaat gtgcgaaagcagcgtgctgaatatgttgaagttgccctagaatatttgccggatgaagcagttgttgcattacaacaagctgtaatggcaaaaatggcagaccgaaacattgc gatagaatgcccaccgaccagcaatacacgtatcagtcagtaccgaaacgtcagcgagcatcatatctttcgctggatgggcttgccgggtgaggcgattgaaggtgatgtt cctatgtctatttgccttggctctgatgatccggggatcttcgctgcggacttgaaatccgagttctatcatctgttcgttgtgttaacccgaaagttcggtttgtcgccagcagatg ctttgagaaaggtagctgaggtgaacgagaatgggcgcatttatcgctttcatgatgtcagctagcctgtatacattgaggattctgtaattgttcaagaccagcagtgctcattg ctaactatctat (SEQ ID NO: 377)
13	STAND	aaatctctttcgcgtcaatagtggtaatatttttttatcattgtcctctttctactgacatactgattgtccgacagtggagccagtcgaaattgttgacagctagtcggggctcgtct ggtctttctagcagtaagaaacgtattaatattggatcgccactagtttaacagatacctcagaattatttatagactgacaccaccccggcagacgatcctgccctataggaag ctaagtggaaacttatccagtaacagcttgtcgattttatcccagagggtgttcctcaggatgtatcgctgaaatcaaatccagcactaagaatgaggggtgagaaaccatttcc ttggtgggtctttgaccatttctgttgaactaatgtttttgggttatcaaggatacaaattcaaggcagtgtttcactaaaccttacctcgcttcaataccaatacatttttaatgggtat aatatgtgactgcttttgccgcattattgacaggaacaaggactggtgatgaatattgatttcagtttaattcgtagcgcccccaaaagccgtaacgatagctttgaagcactcgc cgtacagttatttaggaaaacctgtcgagtaccgacaaattcaacatttattagtctgcgtggagatggtggagacggtggcgttgaggcatatttccgctcaccggacggtgc cgtattcggtgttcaggcaaaatacttttttccagcttgcttccgcagagcttacacagattgatagttcccttaaagctgcgctaagcaaccatcccacactaaccgaatactgga tttatataccgtttgacctgaccgggcgtgttgctgcgggaaagcgaggaaaaagccaggcggaacgctttgaagaatggaaaagtaaagtcgaatcggaagcgtcagcg aaagggaagtcactttctattgtcctttgtaccgctgctgttatctgcaatcaattacttgagatagacccttacggagggatgcgcaggtattggtttgatgacacgttgctgaca acagctcaaattcaacaatgtctggaggacgccattgcttttgccgggccaagatatacttcaatgctggatgtggtgacgaatgctcatgtcggcctggatttctttggtggga ctggtgacttttgcgagtggtacgaaacatcattaacaccaatcgttcgagagttccattcactgaatggatacggacgcaaatcgctggatatactcggcgaaacccgtgcta catctgccacggcattgattgaagaaataattgcctactgtgagagcatgagagataacaatgtcacggccacatcggttacagatctttccgtcgctctgtcatccctattgac acttttcgctgatgcccgccatgctcaagaagataaattttatgaaaagcatggcaagcatagtgatacagaatcgttccgacagttccacgcagagtatatgtgtgcatttcct gccggagatatggatgcggcgagaaaatgggaagagcaggcgcagcaactgcaaaatttgctgacttctcaggtcattggtgccgcaacagcacattccttactgctggtt gggccagcgggtatcggcaaaacccacgcgattgtcagcgcagcattgcgtcgactggaacatggtggtttttcactggtcgtctttggagacgactttggcaaagcagag ccttgggaagtgctacgcagtaaaatagggctgggtgccgccatcgatcgttcgacattatttgaatgcatacaggcctgcgccgaacatactggcttaccttttgtcatttatat cgatgcattgaacgaaagcccgcgagaagtgcgctggaaggacaagcttcccgaattgctcgctcaatgcaagtcttatccagacatcaaaatctgcgtttcaacccgagat acctatcgcaatcttgtggtcgattcacgctttccagggtttgctttcgaacacatcggttttcaggacatcaattcgaagcggtacaagctttcgcagcctactatgagctggat gcagagattacaccactttttttcacccgaactcggtaatcctttatttttacacttggcctgtaaaacgctaaagggcgaaggccgtgacagtctggatatttctttgccgggtttta cctctctgtttcaaggacatctcaaacattgcgatgttttaattcgagaacgcctccactacgcaaaccctcgtaatctggtaagggctgcaatgatggcactcgcgaaaaccct gacacatgagttgccgcagaaccgaacgtgggaaacctgttgcgaagcactgagcaaaatagtgggaactgagaccacacctgaatcctttttaaatgcattggcacatga aggcctcattatcctttctgttgtagatgaggataccttcctgatccgtctgggttatcaacgctacggtgacatactccgtgctatcagccttgtggaaactcttgattcggataca gtaaaactagcggagaaaattgcagcgttaacagaagaagatgctggattgctggaagctcttgccgccgtgctgccagagaaaactgctcttgaaattactgctgaagaag taggattaccatccgaacaagcccataagctgttcatccagtcattggtttggcgctcccgacaaagtgtagtggaagaaattgatgaacacatccatgcagcactgcataca cctggattatgggagtcggtttatgaagcgctgttttcacttagtctggttcctgaccatcgtctaaacgcaactaactggctggggccatttttacggcagtcatccttagctgaa cgtgacacctacttgtcattagctgcgctgggatcatttgataataagactgctgtctattcactcatccatgcagcactatttgctgacataacccattggcctgctgaaagccg gaggctggccagtctaacacttgcctggctcacttcgtgtgctgaccgccgaatcagggatttatcctcaaaagggctaagcagaatcctggcaaactacccggagaactgc caaacagtaatcagtgaatttgcatattgtgatgatgattacgtattagagcgtattagccttgctatctacagtgcatgcttattgtcataccaacgcagaaatgcgtttatgccag cgctccctggtctattaagcattgcgtcagatagcaagaatattctgctccgggatacggttcagctattagtaaacttgttgaaaacaggagaatttcccacagccgtaacaag ccaattacagcattaccagacaaacgtatcattaccatcacgatggcctgtactggcggatgtcaaacccctcctagatctggaacatttaccatcaaacatggtgctctgggg agaatccatggccccggatttctggcgttatcaggtggaatcgaagatttccggctttgacttggagagcgccaatatcagccatgaaaacattgcctgttggttaatgcgaga

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		agcacttaatttaggatatcccggttataaccactgcgcgctcaattatgatcgccatatcgggagtcagtatggctcgggacggggtagaaaagggtatgctgaccgactcggtaaaaaatattactggatcgccttacatcgactactgggcattctggccagtaatgttcccgcactggaagacccatattccgactacgaacctacaagtgatcttctatggtcagtcgacgtccgtaaagttgacctgaccgatgtacgcgatatcaccgcagaaggtgtctatccagtactgatggaggaaacaaattatgcattccctgaccacaattcagatatcaaaggttgggttaggaccgatgattttccaccttatgaagcttgtcttattcgaactgacgaggaaggagagcagtgggtagcgctttcacatagctattgggatgacgataaagcgccgaatgaaaatagctgggattccccgtacttgggagtgcgtgcttcctactcaagcgcactcataaatgaaagcatccagaactttaaacagaaaagatcacgcgatattttccaatataatcagggaagtagttgttatcgcggttatcttgctgaatatcctgacagcccggtatacaaacaacttcttaatagtgatgaagatagtgaagcgtttaattttacagaagtcagtttactgcgcggaaacgaatgggaatacgactactcatataccatgcccgagcgccaggataacctcattgcgccatgcctgggaattattcaaaaactcgaactttttatgggattgtcaaagcggttgggttgatcattctggcaaacttatcgccttccatcaaaaaggtgtaaaacaacgcggacttttcatccatcgttcggcattgaacgcctatctgtccataacaggtgaagagcttatacatcgccgttttgctaacagaggatattttgatttagctggtcgtaatagcacgcaaatagacctgaaaacttggatccagtaccgggcagacaaggcaccggtagttttacgagaagaggaactgccgtttaactgctgacaacgatacttattaagtaatcaactggctgccttggcatcgaatgccagaagagccatttcgcactaccaatttaagtagactgaaggaatacttggtacaagcaaacgcacgccatatcggatagaggggact (SEQ ID NO: 378)
21	Transmembrane ATPase	attatctgccaaccgataagatggctgcctaagtcgtagcgattcagcactgttttagcggcgctcgattgcaaagtcgtgctttgctgacttgcgattgtgctctttacgagcaaagctttcaggtatagtaagtgctaactgtagtgtaaaattatagggatagatgaagaaaacaacgaggctttagctaatctttgcagttgtgtctgctataataaggcgaaattttatctgcatgattttgtttgattaactccgaaagccagctctctcggtgaagattgggaagggatatcaatgagtgatgatagctataaatttcaaaagttaacgccgttcagcgatgttgagctgggtgtatataaaaatgcgatagattttgtttttgccaataacgatctaaaaaatgttgcgatatcagggcaatatagcgcaggaaaaagtagtcttatcgaatcctataagaaaagtcattcaaatataaagtttgttcatatctcacttgctcatttcagatcgattgaggaagctgaaactaatgaaccaagtaaagatataaatgaaaccgcgttagaaggtaaagttcttaaccagttaattcaccaaattaatgctgatgatattccccagacacattttaaagtaaagaaaaaaataaaaactaacaacattgtgataaacaccatctttacggtgttatttatcgccatgatactacatatcacgctatttaataagtgggaaaagtttgtttcacttttatctgaaggtaatataaagacactacttacattatcaactaaatacgatacgctttttaattagtgggtttatatgtactatcctatcttgtattttcatttacaagttaataaaaacccaaaagaatcgtaatgttcttaagaaaataaatttacagggtaatgaaatagagattttttgaagaaagtaacgagtcttatttcgatagatatttaaatgaagtattgtaccttttcgagaacgttgatgctgatgccattgtttttgaagacatggaccgttttaatagtaataacatctttgaacgtcttcatgaggttaacagactggttaatattcaacgggacacagcagggcacaagaaatcgacgttacgttttatttacttgcttcgtgatgatatcttcatttcgaaggatagaaccaaattctttgattatatcattccagttattcctgttgttgatagttctaactcttacgatcagtttatcacacattttgatggtggtggtattctcaagttgttcaatgaaagatttctacaagggatgtctttatatattgatgatatgagaatattgaagaatatttataacgaatttcaaatttattataacaaattaaacacgacagaacttgactgtaataaaatgttggccattattgcctataagaatattttcccaagagattttagtgagttgcaacttaatcaaggtatggtttataccatatttagtgaaaaagacaaccttattattgaagaaataaagaaaatagaaaaagatattagagatagaaaaaaagagattgaggcaatcaatgatgaaatactcaactctagtcaggaggttgatgctatatacgataaggaattatctagatataataatcatcctcactataatcaggctgagaaagctgatatagcaaagagaagggcggctagaaaagaaagtgttgaaaataaatttaatggtaaaatagaagaaattaatgagcttatatcaagatcaagagaaagtttggttgattctagaaacaaaagacttaaagaagtaataactagagaaaacattgatgaaatatttaaactcacctataccaatgaaattggagaggaaagagactttaatgaaataaaaagcagtgagcattttgacttgcttaaataccttattcgtgatggttatattgatgaaacctataccgactatatgacctatttttatgaaaatagcctgagtcgaattgataagatgtttttacgcagcattaccgatcaaaaaggcaaagagttcacttatcaactcaagaaccccaagctggtcgttgcccgccttcgagaagtggattttgaacaggaagaggcgcttaattttgatttattagcttatctgcttcaaacgccagcccaggtaaacttaataaaacgtttattcaaacaactaagaaaagatagaagagttgagtttattcgtggttactttgaaactgagagggctcagcctgtcttcattaatcgattaaatacacagtggcctgagtttttttcttatgcgctgacagagagtgaattttctgctgattgggttaaactctactctataggcacgttttattattctgccaatgacgccatcgaggccattaatattgatgattgtctgactgattacatctctgattcggcaggttatttagcaatatcagaaccgaaggttgacaaattaattagtggttttaagttgcttaacgtctcttttgtcagtattaaatttgaaaacgcaaataaagtactctttgatgcggtttaccagcattcactttatgatattaatttttccaacctgaccttaatgctgagtaaggtttacacgcttaatagtgaagatgatattcgccataagaactatacactagtgatgtcacaacctgattctcccttggctagttatgttaataaccatattagggactatctggatatggttttatctagttgtgatggttcaatcgtggatgatgaatccattgttttatccgttcttaataatgagggaatatctgatgaacaaaaaggccagtatataaacgctttgcaaactttcgtgacatctctgagtgaggttgagagcgaatctttatggtcatctttgttggataaagatagagcagtgtgctctgaggaaaatattgtctcttattttgaacatgttgatggactggatgactcacttatcgaatttatcaatagaactgatgtagacctgaattttcaaaatattaatattgataacgagcttaaaggtaaattatttaaatcgattgttatctgtaatgatttatcaaatgataaatatgaaaaattaatttgctcactaaatattatttgtaaaacatcctttagcgctagtaatatcgcgagtgataagttcaaaatattagtggataaaaatattattcgtatgaatgttgcgccacttaatttcatacgagataactattcagagcaactttcctattatattcataagaatatcagggcatacgttgaattaatgacgattgataactttattttggatgaggctatatcaatactttcttggaaagttgatgatgatttgaaagttaagctactcgagtttgttaaaactccgttggctatttatagtaagaattactctcaggtcgttaatgactatatttagaaaataattttaaaccagatgaacttctaatcttgacgtcatcttataaaacttggggaacctctactcagtcgctcatcttgagtcgagcaatacaggatatatcagcattgatagcaagtcctaatgatgtttctgaaccgttactaaaaaacctgtttgtcgcagagggactgaatatgcagaataaaatagcactgctaatcgctttgttgccgggtaaggatttgagtaagacgacttgcaaagagtatcttgatctgcttggtttatcggagttcagtaaaatttggggcgaggcaaacctaaaattgaagttgattcaactaatcaaagtttattaacagcattaagagataaccacttcttctctgattttgaggtggataatgaaaatcccacttattataaaataacaaggcggcgctctatgtttggctcagatacatagcattatgtatttttctacagtttgggcacttttatagtgcccaatttttacgctgaaacttacgcagataatctgactttttcccagttgacgagtacacctag (SEQ ID NO: 379)

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
22	ATPase + QueC + TatD + DNAse	atctatagcagtcatcatattggattattggtgaagtggtacactgaatttgcccacctgaacagagttggttttatcaaacctgtagtttactcaatgacgtaaaaattggtgatgtaaaggatataaaaatgtggtcagacaaagagtcatcagaagactacctaaattttggtgaagtatctcagttagccgtggatgtacttaccacgaaagatatgttaccagtatctatcggaattttggaaactgggggcaggtaaatcctctctgttaaaactgatagagcaaaaacttgagcaagacgacaaagattggattgttatcaattttgactcttggctctatcaggggtacgacgacgcccgtgccgcacttcttgaagtcatcgctacagaattgacaaaagctgctgaaggtaattctacccttatatcaaaaactaagagactccttagtcgagttgatggttttagagctatgggattactagctgagggtacagctttaatggcaggattacctactggcggtttgctttctagggggattggtgcattaagaaatatcaccgatggcatccagagccaggaagagtatgaggctttaggcaatatagctaaagaaggtaaagaaactgcttgtggtttgattaaaccacaaacaaaaaaaagcccccctcagcagattgatgcctttcgtaaggaatatggggaaattctagaagaacttggaaagccactcattgtggtaatagataacctagaccgctgtctccctgccaatgctatccatacacttgaagctatcaggctattccttttcttgactaatacagcctttattattgcagcagatgaggacatgattcgctcttctgtggctgattacttcaaaggggcatcacagcgccatcaaatagattatctggataagctaatccaggttcctattcgggtgcctaaggctggggtccgtgagatccgttcgtatctgttcatgctttatgccattgaacatggcttagaaggcgaaaaaataactatgctccgtgagggcttagagaaggcgttacagcaatcctggaaagatgaaccaatctcacgtcaggaggccttaaaaatgactggtgaagcggatgatagcaacctcgcgctggcgtttgcgcgtgctgaccgtattgctcccattttagccaactctccaattattcatggtaatcccaggatcgttaaacgcttgttgaatgttgtgaaaatgcgatctcaaattgcgaagcgacgagcaatgcctttggatgaagcaattattactaagctagtaatttttgaacgctgtgttggagtggatggcaccgctgatttatatcatctcgtggatattgaacaaggtgttccccagatacttaaacagcttgacgataatggcggtcaaatacctactgatgcaccaaagacatggactgatagtccaacgactaaatctttcatcagtcaatgggcccaacttgaacctcgtcttggtgggattgacttaagggccgccatatatctgtcccgagaaactatgccaataggtgcatatgtggttggtttatcgccatctggacgggaagtactaaatgcactaattgaattgaaaaacactagttctcctacagcagaaaaccttttgaaagcacttcctcgtgaggagcaaatacctgtaatggaaggtttaattaaccagttacggcaggtatcagattgggatcgtaagcccagaggcttttccggcgcatgtctgttggcccgctactcaacagatgcagccagcatattaattcgttatctacaggaattacagttggggatgaaacgaccagcgtggatgactgcagcattaaaagatgaacaatggaataaggacgcttaatgggaacatcacaatcaagtaaaggtccaggaggtggctctccgctggttccaccatgggctgatgatcagccacagcaaccgttaccctcgccgcaagaaaggaggtttgcgccatttcgagaatcgttgggaaatgcggtatcaaatggaaatcgagcagatttcagaaaagccatagggcactacgcgcgaaaagcctccggagggagcagtaacgctgctcggcgattagggagtgtcacgcaagctggggccgaattatttggggctttagtgggaatgccttcggctcccggagaaccaagcatcgatttgggcagtttggcaggccttccatgcgaaatagcaatatcaactattgctcaagctttaacatcacaggatggtgactcagaaaagatctgtgcggccatgaaccatgctttagtggaggctcttgatggcgtagaaattttcgatcctcaaaaaataactgatggtttgattgttgacacaatgattggttatctagcggaaagtattttccttcagatggtaatggattctaatagggcatggaacaaagcagatacaccttcaaaggcaattcatgcagaaattgaactccgggaattgattaaagttgttgttgataaacatatggcaccaaaacttgccggtaacataagatcgttcacacgaaaccaaatggtaaaaattgaacgtcaggccattattgaggcctggcaagaatgggaggcataccagtgacacaattagttttccatcataaacatcaccatttgccgccagcaagtgagaaagtgttacctgttcagctatatggattaagtggtcagaggcgcggagatatatctgttatcgggaatcctgcgattgatcggatcagacgtttgggagtacagcttccagctaaggtcatggattttctgagtgttgcattagcagtaactgcagcagatactttcgttcagcgtgaaagttccgaggatggttggacccgccaattgtcgttacgactccccccttcatgaaccatccagatggattagtctaaagaaagaacttgagagtgctttgcatttcttagtggagacatctgggatttcgaattttgtgacgatggttatgcaccgccagagccttatagccagcattcaaggcatcgtctgattaagctaaaagggcttgactgtgtcagcttatttcaggaggtctggattcagctattggtgcaatagatcttctggctgcagggcgcgctccacttttggttagtcatgcttataaaggggataagtctcgtcaagatcagattgctgaaaaattaagtggccaattttcgcgctttgagattaatgctgacccacacatttatcaaggcgtgactgatattacgatgcgaactcgtagcctcaattttcttgcccttgcggccgtaggtgcttgtgccgtacaagagatatctcaacaagaaaagattgatttgttcgtacctgaaaatggatttatctcattaaatgcaccacttactccacggcggataggttcgctgagcacacgaacaacacatccacatttattacgagcatacaaaagatctttgatgcgctcggtatttcttgtcaaataatcaatccatatcagtttaagacaaaaggaaaaatgatctccgaatgttcaaataagcagctcttatctaaaattgtggaaagtacagtatcctgcagtcattggaaacgaatggggcagcaatgtggggtatgtataccgtgtatcattcgacgagcatcacttcatgcaggggggaattagtagagatgttgaatatattttccagtccttagctaaagtaatgaatgaaatagatcgcagggacgacctgatcgcccttaggattgcgatcacgcagaaatcgactttgaaaataggtacatggattgccaaaagtggccctttgcctacggcagaatttgataatttcaagcaagtatttaaggatggcctagatgaggttgaaagctatttactgagtgagaacatagtatgagcatcgatatgcactgtcatctagacttatatcctcggccagacctcgtggctgaagaaagtaaacgtcgagggacttatattctgtcggtgacaacaacacctaaagcatggcatggtacttctttattggctaaagaaagtcaacgaatccgaactgctcttgggctacatcctcaaatcgcgcatcaaagatcgcatgagttagacctgtttgattcattgctttcggaaactaagtatgtaggggaaatagggcttgatggtggacagggatttaaagaacattgggatattcaattgaaagtgttccgacacattctcaacagtgtaaatcgggctggtggcaagattatgactatccatagtcggggaagtgcatcagcggtgcttgatgagattgaaaatatcgatggggtggcaatattgcattggttcactggaacacctaagcagcttgaaagggcaattgatttaggatgctggttctcagtggggcctgctatgctcgatacaataaagggtaaggccttagttttgaaaatacccaaatcacgcattcttacagaaacagatgggccatttgctaagtttcgtaatgacccactaatgccatgggatagtgggattgcagagaaacagttagccgcattatggggggattagtcagatggaggttaatgctcagctagttgataattttaaggtattatgtacatcataagaatgaaaaacttagatatgcatttacagttcaattcatttttcgtcatcagttaattacacataaaattaaaagtaagaatatatctaccctgtgaatgagcaaggcggatttatatagtttgtaattagtttaaatgtaagcagttcgtcagagtgcgtattccgctctattcgatcacggattggccgttatgaccc (SEQ ID NO: 380)
23	DUF4011-helcase-Vsr-DUF3320	gctatcctacctcagattactgggctgacctaatctatagatcaggttctctttatactttatgttagcgaaatactaagatgcttcttagtgacgacctcttgacggtagaggacgcgtgcatagattttacaatcactgcctttcgcccccctaacctaatccgcgaatgatgcatcctgaacttgcgcgccagttcttatactcgccgtcagagcaatcaaattgctgatgctttctgcctgttcaaggcatctcctgtcgtcagcaatactgtgcatatttgattgatttcctcttaaggagaattagtttcatgggtattaaagcgcaggtgagtatcgcgcacaagctggggttcacatcacaccaaaatgcagttccgctgttacgtgagcttatcttgcataatgagtccgaagagacatttcaggatctgacactgcatctgaggaccgtgccagctg

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		tgctcgaagaaaaaaaatggaatatcgatcgcctgcttcccggtacttcacttgatatcagagatcgggatatcaaacttaatgctgaatggctagccgaactgactgaaagcgtactctgcgaagtcacgctaagtttgcgccagggtgaggaagaactcttcattacccattacccgcttgaggcactggcgaaaaatgaatggggcggcagtgcaatgattgaattgctcccttcatttattattcctaatgatccggctgtggatcgtgtactcaaggcaacctctgatgtccttcgccgtgcaggcaaggatgacgctcttaatggttatgaaagcaagtcgagaactcgtgtctgggaaattgcctcagctctctggactgctgtttgcaacctcaatatcagttatgcccttcccccagccagttttgaacgcaatggccagaaaattcgcactccaggagccattctggaaggaaaagtcgcgacctgtctggatacaacattattatttgcttcagcactggaacagattggtctgaattcactgctaatgctcagtgaaggtcatgcgtttgctggtgtctggttacaaccgcaggaattttcgcagctagtgacagatgacgtctctgcggtgcgcaaacgtgtcgacctgaaagaaatggtcgtatttgagacaactctcgcgaccagagctcacccgccttcatttactcaggcatctgatgaagcgttaaagcatcttaacgaggatgtttttcacgcagccattgattcccgtcgcgcgcgtatgcagaaaattcggccactggctctggggggcactcgccttgaagaccagtcggatgcctgcgaggttatttgcatgggtttgaggaagcccccctatatccccgatgttgatattgatatcgagacaactggcgaaaaagaagccgggggggcggctggtacagtggcaacgaaaacttctggacttaaccacccgtaaccgcctgttacacctgtctgaaagcgctaaaggcattcgtttgatctgtgcgaatccgggccatcttgaagataaactggctgaaggcaaacgcattcgcattgtcccgctccctgatctcgaaagcggcggccgcgatgccgaactttatcagcagctcacaaatgagaacctgcaggaagaatacgctcagattgcgctggaacgcggtgaagtcgtctcctcaatggaaaaataccgcctcgagtcatccctgatcgacctctatcgaaaatcgaaaagtgatctcgaggaaggtggtgccaacactcttttcctcgctgttggcttccttaaatggaaaaaatctgctgatgaccccaaaagttactctgctccactgatactgctgccgattcaacttgaccgtaaaagtgcactttcgggcgtgaccatgcgtttgctggaagaagagccccgcttcaaccttacactgcttgagctgctgcataatgactttgctctgacaatcaacggcctcgatggtgatctacccaccgatgaaagtggtgttgatgtggatggtatctggaatatggtacggcgtgctgtacgcgacatacccggtttcgaagtcacccgcgatgtcgtgattggcacattctcttttgccaaatatctgatgtggaaagatctcatcgaccgggcacctcagctgatgcaaagtgcgctggtaaagtatcttatcgaacgcggccaggaaaatgccgttctggataagagcggagaagtcatcaacgctcatgaactcgatgacaacatcaatacgcaggatcttttcttgccgttgcctgcagattcctcgcaaatcgccgctgttgtagcctctgcaaaaggcagggatttgttctggatggcccacccggtaccggtaagtcgcaaaccatagccaatatgatcgcgcataaccttgcgctaggcaggcgcgtactttttgtcgctgaaaagaaagcggcgctggatgtggtctatcgtaggcttgaggcccagggactcggtgaattttgtctggaactgcactcgagcaaaacgtccaagatggattttctgaaacagctcgagcgggcatgggatgcgcgtgatctactaaccaccgaggagtggaaggaagaagcggccaaggtgcagcacctgcgtgacaaactcaatgaggttgtccgtttgctccatcggcgctggcccaatggcttaacactccatcaggcaatgggcacagttatcagggatgcaagtagcgccacgccgcactttagctggcctgcatcgactttgcattcttctgcagagatgacacagttcagagagatagtaaaacgtctggagctgaaccgtgatgcatggaaacagcacggcgatcattttgaactcatcgcgcaggctgactggaccaatggatggcagtcctctctcattgctgcagcaaactcattgcctgcaaccatcgatcaccttgaagacgcgaccgaggcgttactgaaggcgacgggagttactctgctctctaccgagccggagagactgtcgcagttaacttcattctgtgaattattgtcggaagcttacggcattgatctgagtttcatgttcgcaccggatgccgcaagccgtatagagtcagcgaataaagccgttcacctcctgaaagagattgaagcgacaaaggctaatctgtcagttacctacccttgtaacagttggcagcacgttaatgtcccacagatcagaaacgcacttgacgtcgctgacaaaaaattctggttctttgcgaccagtgcccgcaagaaagtcattggtgaagttatccgacaacactcgctaacgtcagcccccgacttatccgttgatctccccattgctgaaactctgcagacattgctgcaacgtctgaccgagcttaactctgctactgtatctctgccgggatgggttggactggataccaacgttgcacagttgcagaccaccctgcaacttgccgaatctatccgcaattcgcttggtggtttcgcttcttcgccacagcagttggccgagatccgcactgcggtaaaaaacctgattgttgatgccaatgaccttctcggttcgcagggcgttatctccgcactaacccggaaactgcgcacagcgatcgccgatttcaatgatgcacaggttagcttctgcaatctgataaaaccatctgaggataaaccatcgctcccggcactgcgtgactgcgcactcaatatcctgcaacatcagtccgctcttaaagcctggagtgactggagccgtgtgcgtgaggaagcgatttcacatggcctgcaaccagtgatcaacgcgctggtccatcttgactcaggagacatcagcgcggcagagatttttgaaactgcctattgccgctggtttgcatcgtggatgatcgattcagagccgctgctgcacaattttgtgccggctgagcacatgagtgatattgaggcttaccgtacgcaaaccgatcgtctgtccaaactggcagtacgctacatccgtgcccgtttatgtggcgtcattcctgcaaaaaatgaggtcagcaagcagggtggttttgctctgcttaaacatgaactacagaaatcccgtcgtcataaaccggtacgtcagatggcagcagaaatgggagatgccatggccaaacttgccccctgcatgcttatgagtccgctttcagtcgcccagttcctgccctcggaccaggacttgtttgaccttgtgattttcgatgaagcatcgcagattgccccgtgggatgctatcggcaccatggcgcgtggcaaacaggtggtaatcgctggcgatccccgccaaatgccgcctaccagctttttaatcgtgcagccaatgacactgacgatgatactgaagaagatatggaaagcattctggatgagtgtcttgctgccggcctgtataaccacagcctgagctggcattaccggagccgtcatgaaagcctgattaccttctccaaccatcgctactatgacagtagcctgattacgttccccgcttcggaaacaaagcaaagtgctgtccagtggtgcaaggttgcaggcgtctactctaaagggaaaggacgtcataatcaggccgaggcagaagcgatcgtcgctgaaacggtgaagcgactgactgataaagagttcgttgcatcaggcagatcgataggcattatcacgctgaataccgaacagcaaaagctagtcagcgatctgctggaccgtgccagacagcaacaccctgaaattgaacccttcttccagtctgaactggaagaacctgttgtggttaaaaacctcgaaacggttcaggggatgaacgcgatttgatcatactctgcatcgggtacggcccgactgaaccgggcgcaaatacaatgtcgatgaattttggaccgcttaatcgcgagggaggctggcgccgactgaatgttgccgtcacacgtgcgcggcaggaaatgatggtcttcagctcgttcgatccttcctttatcgaccttaatcggaccaacgcccgcgcggttgctgacctcaaacactttattgagtttgcccagcgcggccctgtagctcttgcccaggcagtacgtgggtctgtaggcggttatgactcaccgtttgaagaggcagtggcaaatggcctgagaagaaaaggctggcatgttgtcccgcaaattggcgtatcccgtttccgtattgatttggggatcgttcatccggataagcctggcgactatcttgtcggtgttgaatgtgacggcgccacttaccatagcgcagcaacagcacgcgatcgcgataaagtccggagctccatcctgcagggcctgggctggaaattactgcgcctctggtcaacagaatggtggattgataaagaaggcgcactcgacaggctggatgcagcaataagtcgcctgctggaggactccagagcagcggaagccgcactgattgctgaagcagaaaaacaaaagcagattacgccagtcatcgctcccgtaaccaatgatgtcagtgatgacatactggtttctgaaactacacctgtcgctaatgatgcggaaatatccgcgtcagtaacccctgtcatcccgcttactgccaaagtaagcgaagatgatggtaacactgggctgaggtatgcatctttagcttctcagaataacgacaagccagtgaatgtcggtaagtatgtcgttaacgatcttcaggaatggtgcgacaggacagatgcagaacaattctatatcgctgaatatgatgagacacttaaaaccctcattgaagcggtggtgacaagtgaatcaccggtcctggatacaacgcttgtgcaacgcatcgcacgtatacacggcttca

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		ctcgcgccggcagactgatacgtgaacgcgtaatggaaattgtggatcaacactatcaccttgcaaccgatcactcaggtgaagacttcgtctggctgtccgcagcgcaacg tgctgactggaatgtgtttcgtttgccagccacggataacgacattcgtcaggttgacgcgatccccagtgaggaattacgcgcactggcgctgagtattgaaggtgacaata agatacaggaaatgacccgctcgcttggcattaaacgcctgactagtcaggcaaaaaaaaggattgaatcagtacttgatgttgtttgaaggtcaaccgtgtggaaaacctctt ttagagactaacagtctgaaatatagagtcttattcgatcatcttgagaccgaatgtattagagtcgatttctgacacctcttatcgtggttttctgcatcaccaacatcgaccagtt gggcgtaatcaaggaggacgtctggaaaacgaatctatggtcactcccgttttgcaacaccgattttgacaataagttggtttgcttgaatctattcggcatcagaatggaatttt ttttccacgcctcgatgagttccgcgcctgatgaa (SEQ ID NO: 381)
28	ATPase + protease (ietAS)	ttaatgcaaacgcatcaggaagggcagacctagtcacatgtagaatacgatagcaataaaaaagtctaattagaatgcaaattgatgcaactctatgccctccaagaactcca aacctgaaagatttatgtaaaacatagtgttcgtttcaccaaaatacatataaactacattaaaatagaaatttgtctcacctataagccatttagacaacagattaatgaggtttgta tcacaaatgaccacaaacgagatactttcgcagcttatcagtcttggactcaaaggggataaagttgcttttgttcggcaggcttcgaaactcgcgcgttcctatgattctatggg gctgcctgagcttgcttcagccattagaggtagtattcaagataaaaacacgtttaacttgcagaaagtatcacgcagtacatcacctattttgaacgtcttgatacattacctgt agataaagaaactaaatttgatttagcagacgtaactcaaccgtcttctgaaattcaactcccattgttgaaagatagcactctgaaaaaaattaaagaattttgactttcactgaa cgagctaaagaattaaaggatgccggtcttggcgtgacatcctctatgattttatatgggccaccaggttgtggtaaaaccttgacatcaaaatatattgcatcctgtctaaattta ccgcttcttactgcaagatgtgactccttagtctcatcatatctggggtctacttctaaaaatatcaggcagctatttgagtatgcaagtaaagcaccatgtgttttatttctagatga actagattctctagcaaaggctagagatgatcagcatgagttaggtgaactgaagagggtggtggtttctttattgcaaaatattgacaatctacctgaagaaacaatattgattg ctgcaagtaatcatgaaaatcttctagatagcgcagtttggaggcgctttgagtatagaatatctattggattgcctgattttgaagtcagaaaacaactatttgaacaatattcaaa cataaaagctacatatgacgattttgttgatgaccttgcggaaatatcatcagggctaaactgctcatttatagaacaatgctgcttaagatctgagcgacatgctctggtttacaa taataaacaaatcgatacccgatttttagtcgaggctatcttagaagcgaagggagttacatttgatgaagaagataatttacttataaagattgtgaccactctcagagaataca atcccaaaagatttacaatacgaaagatagcaaaaatactagggctttcaaatgctaaagtgtcaaggctaactaagaactatagagagatattatgagtaacaaagaaagac caataaaaataattgaggcgacacctcaagattttactgaaaaaacatataatttcggaaagaaacaacctatccgaacagtaacaactagtctaaaaaatagactcaaacaa gaagtcgatgacgttaaaaatttttccagagctcatttaaaaaatggcccaatataccggcggtggctagagttactcttcatgaaaaagctcttgctaagtcacatcgcccatc aagcctattaggtgataatacatgtcccgtaataggcagtgataattttggagaattacttataagtgttactgaaaaagggttagcacaacttcgcaaaaaaattgaaaatagca ctaattctcataatgggacagtacatattgctgtaattgaaaagatcgaaccttttagtcttaaccatgatgttatagataaaaataaatcagatagttttcttctgaaactctttgacc ataaagatagaacaactaaccgcagtatcgacaaagaattaatggaatttgcagatgaactaggaatacaaaaacccaaaaagtatgatatcagttcagatttgagtatatatg aagtaaaagggaatgataacatcgcccaactggcaagttttattggcatacgaaaattagaacctatgccaacatttggtcttactcatacagtatcgcaatatattcctgctgaa actctagacctagatgatttttcccttacctcaagaggataaacattatccactactcggaattatagatagcggagtcgatcccaataacaacatacttaggccatggatttggga tagtttagatttagtaaaaggagaacacgactattctcatgggaacatggttgcaagtttagcaattaatggaagatggttgaataactatgctggttttcctcaatgccaagctga aattgttgatgttgcagcctttcccaaagatggtacgctcaaattgccacaattaatgaaagctatccgagaggctgtgaccacctatccagaagtacgtgtatggaatctgtca ttaggttgtcaatccccatgttctgaagacagcttctctgaattggggcatttttttaaatgcacttcatgatgagcatgattgtcttttcgtcgtagcatccggcaactacatttatgat cctcaacgaacctggcctcctcaagaattaggtgggcatgacagaatatcagcccccgcagattctgttcgttcattaactgttggctcagttgcccatttagaatcgtctgactc tgtggtcaaaagatttgaaccttcatctttttctagaagaggtcccggcccagcctttatacccaaaccagagataaatcactttggaggtaattgtgacagtaaattaaactgtg aacataccggaatcatagctattggcgaggacaatgctctttgcgaaagtattggcacaagtttatcagcaccgttaatctcaagtttagcggcatcactgtggcatgaactaga tgttaatggttctatttcaccatcgcctgaacgtatcaaggcactattaattcattctgcgttaaaaaactcaccagccaaaacggagcattatgcgtttaattatcaaggatttgga cgcccaagcgatcatataaatgatattattggttgcaataaaaatgagattacatttctatttgaaatagatacccgagaaggtattgaattcagtagaacgccatttgtaatacca cagtcattacgtactgaggatggaaaattcacaggtgaaattattatgacactcgtttattctccaccgcttgattatgactacccatctgaatattgccgttctaatgtggatgtgtc attcgggacttacacttatgatccagttaacgctaaatggatacatagcggaaaaattccacaaataaaagaaaagagtgaattatttgaaaaggtactgatagaaaatggcttc aaatggtctccagtcaaagtttatagaaaacaatttccgcaaggtataaatggggagcaatggagacttaaacttgatgttcagagacgagcagagcaagagcctctatcttc acctcaacgtgctgtattggctattacgttaagatctcttgccaattctactacagtctacaacgaagccgaggttgaaataaataatcttggttggaaagaaactgatattgttgtt cgtgaacaaccaaaaatcaggattcgtcaaaaataagcattatggtcaccttttataggtgaccattta (SEQ ID NO: 382)
28	ATPase + protease (ietAS)	gggacactcaggttacataacaatgagtgatacagttcacgtagtgaaggtactatgcctaggtgtttgattacactttgatcattgatgatacgctcatgaaggtattactttcct gtaatgagcaggtaggtaacgatgtcgaactaaatgaatttatagtaaactttgcaacaagagaacaagggagtatgaggggttatggctactgcagagcagatcaaagcttt attgaaaagccacgttgatcgtgatgatcagcgtttcttttctattgctttgcaggtggcagctaaggaagcaaggcaaggtcatcataagcttgctaatgatataaaaaacttag ttgataaaaatcagaaaacaacgagttctgtaggtttagttgaaaaacgacttacaccatttgttaagcagcctgatggtgatcttaaggggttacttgagcaaacgaacaagcc agtacatcttcaagatctggtgatttctggaagcgttagggaaagattgaatcaggttctgcttgaacaaaaacagaaagataaactttctgagtttgggcttattccaagaagaa aaattcttttcactggtcctcccggtactggtaagacaatgtccgcatcagtcattgctacagagttaaagctaccactttatacagtcgtcttagataatctaatcactcgctatat gggtgaaactgcagctaagctgcgtttaattttgaccacatacggcaaacaagagctgtatatttttttgacgagttcgatgctataggaactcagcgtggcgctcagaatgac gttggagaaattcgtagggtcttaaattctttttaatgtttgtagagcaggatgattctgagagcatagttttagctgcaaccaatcatccagagcttttagatcgcgccttatatag

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		acgatttgacgatattataccgttcacaaggcctgaggataatctaatcaggaatcttattgaacagagactcgctgtctttgacctcggtaatttattttggagtgagatcattgatagtgcttcaggtctaagtgcagcggagatcacgcgagcaagtgaagatgctgccaaagaatcagtgctttataatgcaaacaatattacaaccgatttgttagtaaaggctataaagcgtaggcaagaaagtagacaataagggatgaaatgactaccaacaagaggcatattttattaaacggctatgtttcccccgaaaactatcgctctaggagcaatggtcgtagtccccaagtcccagctcgtgatcgagcggtacatggtatatcattactaaatcagtatagccgtatattgaatcattatgatgaaagaccgaggcttcccctgttactgatgaaaaagggatttatgttaggctaatcagttttgaacaatgcgatcttcctatagataaaatcgataatacttatttcaagctttgttctttagttaaatcaaataatcgtgaaactgcgattatatacattaatgaaaatgacagaactaaattcactaaaaaaataaatgactatttgaatccatcgaaggatggtatcgagttccctagaaatcatttgttaattgatagcatacaaaatatcgagttagcagatataacttctttctggacagataaaaaagatcttattccggatgatcacggtgttgaaaagtggtttgagctttggcttaagggtaataaggaggatgtgctaaatattgctcggcgtttatgcgaaagaattaatggaaggctcgggaatacttctattaatttttcgatactactgttgttcttatccgtacgagtctatcgagattaaaagtttgtcctgaattaatatctaatttaaaagagataagatcagcgagggatgatatatcagttatagttaattccttacctacagaacagcatcagtgggcagaaaatgttgctgcaagaattacgcgtaacaatgaagctgatgtttctgtttgtatattagatacaggtgttaactacaataatccactattatctagatttactaactcatcactggcagctgcttgggacatatcttggccacttttcgatgattataatcaaaggccttataatgaccacggttccagacaagcaggactatgtgtttatggagatttcctgtctgttttattgaacgatcaggacatttcgattccgtacaatatcgaatcaggaaggatactacctccaagagctactaatgatcctaatctttatggagctattactacaggaacgtcaagtcgtctggagctggaaaacccgaactggcgcagagtttattcgcttgctgtgacagcagagcctaatactcttggaggccaaccgtcctcatggtctgcagagattgacaagtttagttttggtttagaggatgatatccgcagattatttataatttctgcgggtaactctcaacctacaaatttagaattagattattgggattcagtgactcttgctgaaattgaagatcctgctcaatcttggaatgcattaactgtaggggcgtatactgataaaacaacccatacagaccgcgaatatgatggttggtctcctttcgctatgtcagaagatattgcaccgtcatctcggtcatcggtatcctggggatggaaaaagcatgccccatataagccagatttagtagaggaaggcggaaacaaacttatatcacctagccgtgatgaaatcacaaatacaattgaattatctttgctcacaacctctggcagggcaacaaatcaattgtttgaagttaattcagatactagcgcagcctgtgctctagtatcaaaacatgctgctatgctaatggctcagtacccagaatattggcctgaaactattaggggattacttgttcatacagcaagatggactagtcgtatgcacgaacgatatagaacagaacgtgcacaggggacaccaaaatcggctaaagaaagcttattaaggatggttggttatggagtacctaatttaaatcgagcaatgcatagtgcggaaaatgcacttacattaatatctcagtcggaaatcaccccatttaaaagagatggttctactgatcctacattgaatgaaatgcatctgttttcactcccttggcccgtagaagctcttcgcttactaccaccagaaacaaatgttattttaagaatcacattgtcgtattttattgaacctaatccaagtcaaaaaggattcagacgacaatattcgtatcaatctcatggattgagatttgcagttattagacctaatcagacccttgaaaatttccgtgcttcgataaaccgtaatgcgaataatgaagaatacaatggacctgaaggagatgcgtcaggatggtttctggggcctcaactcagagttagaggttcattacactcagatgcttggaaaggcagtgctgcagatttaacagagatgaatactatcgctgtctatcctgttggtggatggtggaaatatcgtactgcgcaggatcgctatattaacaatgttaaatatagtttattggttagcatagatgtaccagatgagaacattgatatttacagtgagattcaaaacattattcaaattgataatcaaatagatattgaacattaaggttttatgcctaaggtttaatgagtttgaaatgaaaaatcctttactaattggctgggtcgatgataaagacctggccatctttttatacggaaatgatttatgttttattttactaaatttatattagaaccatcgtgcagattgtgataattccttcatactgatttttttacctattatagttgatttttgttgcttgatatctctctttaatacaacggcgtagtac (SEQ ID NO: 383)
30	Retron-protease	tctatctaaaagtatacatatagtatttcaatgaaggttatattatattttgtggctgttttctaattttatcaataagattattgcaaaaggctgataaatataatagctttattatatcggaggagttgatttaactttcctatactatctgtataggctaataccaatggcaattttgccctcaaattggtctccttaatgtttatcaacgtgttatacggtagtgataaaacctcctccgatatttttctcatgaattgggatattttaaatatgttttgctcagtaaccaagttgcatgaatgtaaaaatgttgaacaattatactatttttttaggatgtgaagaggctgaaattagtaggtttttatatagtggagtaattaaataccgctctttttccatacttaaaaaaaatggtaattttagaaatataagagcacctgtaaagtatttaaaagaaattcagtataagataaaggatgagctcgaaaaatattataccccgaaatcatgtactcatggttttatagctggaaggaatataatcacaaatgcgaaacctcatataagaaaagaatttatttaaatatagatttaaaggatttttttgattcaattaattttggacgagttagtcgtttatttcaaagccaacctctaaacttgccagagaatgttgcccatgttttggcacatatttgttgctataatagagccttacctcaaggtgctcccacatccccaattatatctaatatgatatcttatcgtttagacagacaattgaaggagttggcaagaaataatgcgtgtacttataccagatatgcagatgatataactttttcttttactaaaactaaaaagtatcttccaaaatcaattgtttctttaagtaaagataataacattatactaggccatgaattaaaaaaggtaattgaagataattggtttgaaataaatgaaggaaaagtaaggttacaacataaaacacaaagacaatcagtaacaaatattacggttaacactaaaattaatataagtagaaaatttaaaaaacaaacttcagctatggttaatgcattatttaaatatggagcatctaaagctgaaagagaatattttagtaagtatcacaagggttatatagcagaaaggcaatataataagattaaagaaaaaccaggtttattatttacacaaaaagtaagaggaaggttgaattatatccgattagtttgtggtaagaataatgaaagctggagaaagctcatgtataaatatactgtggcaataggacaacctaatgaggagtacaatagaacattgtgggatattgctggtgattcaacgttcattctttggtcgaattcctcacaaggaagtggtttttttcttgaaaatattggtttagttacaaatgagcatgtaatcgaaggaatagaaaacagcaatattaataatgatctaataatactttggttaccaaatgaaagaaaagaatatattgagttacacttagcttggaaagatgataatactgatttagctgtaattacttctaatatatcttttcttgacataaagcctttacaagtagagccagttcctatttatgatataggaacagaagtatatgcagttgggtatcctaattatgacgccagaggctcaattggaaaacctactattattacagcaaaaataacgagtataattactcgagaaaggcaagaaagaatcgttatagaccaaccaatagtacatgggcatagtggtggggtcgttttaaatgctgatggacgtgtaataggcattgttgcaaatggaaatgccgaggggggaattaagagtagttcctaatgcttttattcctattgaaatattattaaatgagcacaagttacgaactaaatcataaaattattattcttaaaataattaaatatttttttaaaaccactagtttgataactagcggtttttttatttttggagtacat (SEQ ID NO: 384)
30	Retron-protease	ctttaaaatgtttcatacagcatacttgtataaaaaaaaactttatgctataaagacataagtggcggcctttgagtttaactttcctacgactatctgcgtaggtcatttttcaacggcagttttgcactctaagtttgccgataagtttgtcgcgcagctggcaatagagaaaacatggccgccactcttccatataaggatttttatgccctcattttcattaaaagaatgtaatg

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		acgtttggaaattatgtgatttactgggagttaacttcgaaaatctatctaaaaaagtatatccaagtaataataggttatatagatgtttctttattccaaagaaatctggtggactaagagaaatatactgccctattaaatcacttaagaacttacaaaagaaaataaaaatagagctagaaaaagaaataaaatacagatcgcctgcacatggatttattaaagggaaaagtataataacaaatgctgaacaacatataggaaaaactatagtacttaacttagacctcgaagatttttcaaaaatatacattttggcagaataaaaaaattatttgaatcaagcccattaaatttaaaacactctgtatcaactttccttgctcatatctgttgtagaggtggtgtattaatagctggttcgccaacatctccgattatatcaaatatgatttgttataaattagatggtcaacttcaacgtttagctaaaaaaaaccactgtacatacaccagatatgcagatgatataacattctcttttacttgctcagaaagaaagttgccgagagggatcgtacatatagatgaaagttcattattaggttttaaattaggcgatgagttatctgaaattatttcaagtaacaacttcactctaaatgaatctaaaataagattaagtcgaaaatcacaacgtcaagaagtaacgggtttaatagtaaattcaaaagtaaacgtaaaaagagacttcattcgcagaacatcatctatgattcatgctctaaaaattcatggtgctgaagacgcagaaaaagaacattatttaaaatataaaaaaacttatataccagaaagacaaaataaaagacaaaaggataaacctggagatctatacacaaaagtaatcaaagggagactaaactatcttagaatggttagaggtgaggattgtaacttgtggcgtaaacttatgtatgattttactgttgcaatgaagaatccagatgagtcttataaacgaacatggttagacgatgcggcagagtctactgtgatatttaacacttacgatgggtgcggcagtggtttttttaataaatcatgatatcaaaaaatatcccaatggactcattattactaattatcacgtgattcctgagataaatagtgataatatttcaaacattgaagttcatacatggatgaatccttctaaaggattttttattacttaaatttgtagcttcaagtaaagacttagatattgctatattaactgcggacataccatttccagttagtaagttttggttgtaaattcatgtcctaactatagacctggaattaaaattcataccataggatatccagattattcatctggagaggatccaacttttatatctacaaaaattaaaggtaaaactacatatcatggtcaattgagatatcagatcatagatgaaataaaacatgggaatagcggaggccctgtctttgattcagatagaaaagtcataggcattgtgtctaatggaaacgaaaaaggtgcaccaaaaaacaataagagtagcttcataccaatcgagaccttgcttgattttataaattgtcaaaagtaaatgttttaaaaaaaccatacattgataactatattttttacacagtaaaaaacaccataatcttatatggatatcagatta (SEQ ID NO: 385)
30	Retron-protease	aagaaaaaggaatcttctaaattaatgaaactataattatacgaatcagtaataccacagttattgacatattttgtaataagctttattttttactaaagcacagtacatcatacaaatttaattttctactgacttatcagcggtagccataaacgtgtatcttctgcctcagctatcctacagtttcttgtggattgtcgtcattgcaaaagagaaaactagatgatgtattgtgctcccctttttaaaggactcgcatacaatgtttgacccattcaaagtagcgccgccaaaattgaaactacatcaatgtgtagacgttcatgagctttctgcaatattaggaacgaactacaatcagttatcaaaattaatatatcctaccactcaaaattcttattattgtttcagtattgataaaatgaacgggaacaagcgagttataaatgcacccaaaaataaattaaagtcgatacaaagacgattagcatatttacttaatgagtattatcctgtcagggatgttgctcatggttttattaaaaataaaagtattgtgtcaaatgcagaacagcatgttcttaaaaactgcgtattcaatatagatttagaaaacttctttggtcagatccatttcgggcgtatacgtaatttattattttcaccgccatttaacttttcaacttcggtatcaacagtaatttcacatatttgctgtagtgatggttttcttcctcaaggtgcaccaacatctcctataatatctaatttaatatgttataaattagataatgaacttaggcgattggccgtttatcataaatgtacttatacaagatatgtagatgatataacattctcttttacatgcaaagcaaatagaataccatcacaaatagttgtatcttcaggaaatacggtaacgccaggtaatgagataaatgcaataataacaaggaatggtttctctataaacgacaaaaaaaccagactgcaacaaaagaatgaaaggcaaatagttactgggatagtggtaaataaacggacaaatgttcaacggagttttgtccgaaaaacaaactcaatgctgtatgcatgggaaaaatttggagctatcttagctgaaaaggattactttgataaatacaatagcaagattaaaactataaaactaaaagatttcattgataatccgggagagttatttaagagtatcgtaaaaggaaggataaactatataaaaatggttagagggaaagatgatgtaatatatagaaaattcgcccataggatatcttgtttattcggcaagtttgataataggtatcttaaaacaccgtatgattttgctattgaatctacatttgtactcgaaaatagatgtgatgactcacaaggtactgcatttttactagagagaatagggttggttacaaaccatcatgtcgtagaagatatctgtgatatcacagatgagtttattgacttattcttatggaatgaaataggcaatattcgaaagacaaaattcataatgtcaaacaaactgtttgatattgccgttttcgaaagaacatccgacttcgacaatataacaccattaaaaattggtgatgatagtggaataaaaaatggtactgttattacagtaattggtttcccacaatattctcctggtgaaagcgcttatgtgaatacaggaaaggtaattcaatcgaaaactatgtatggtaataaattttggcttattgatatacctgttattcatggaaatagtggtgggccagtattaaatgacaaatttgaagttataggtattgctagcatcggtacagcgaagaacgatagttcatctaaacttcatgggttcattcccatatcgactttattaagatatacgggtgaagataagccttaatctctctttctctaagtgatttttaaagcgcctacagtccatactgtctggcgttttttttgttaccggtcatacgtgccattctgatgctgagaatatgacattgggcat (SEQ ID NO: 386)
30	Retron-protease	ttacattactatataatatgcaattaaaatgaataatttatactattgacatattttgtaatacgctatatttttttaacggcacagcgcattttatcacaatttaactttctactgactatctgcggtagccataaacatgtaacttctgcatcggccgactttccgtatctcgcatgtttgccgaatttgcaaaagagaaaatagataaagtgcactgtgccctatttaaaggaatgataataaaatgtttaatccaaccaatatattaccaccaaaaataaaattaaataaatgtggtgatgtacatatattagctgcgttatttaatttaacttatgaagatctatctaaattaatttatccaactccaaatagatcctattatcaatttgctatcgataaaaaaaatggtagtaaacgggtgattagcgctcccaaaaagaaattaaaaatcgttcaaaaaaagatagcagatgaattacttacactttatcctattcgtgatgtttctcatggttttattaaaggaaaaagtattgtttctaatgcggaaaaacatgttcttaaaagttgcgtacttaatatagatctcgaagatttctttggaagtatacatttcggaagagtaagaaatttgttaacttcaccttcatttaatatacccttacctgtagcaacagtgatttccaatatatgttgttataacggatccattccacaaggagcacctacatctcctattatttccaatttaatatgttataagttagataatgaattacgacaactcgctggtaaatataattgcacctatacgagatatgtcgatgatataacattctcattcacatataaagccaaaagaataccatatcaactagttacctctgatgccaacataataaatataggagttgaattagaggaaataataactagaaatggtttttcaattaacaaaaacaaaactagattacagagtaaaaatgaaagacaaactgtcacaggaatagttgtaaataagaaaactaatttacagcgaaaattcatacggcaaacctcatccatgttgtatgcatgggaaaaacatggcgtagtagctgctgaaaatgaacactttgttaaatataacaaaaaaaaataagctaataaaattaagggatttcgtagataaaccaggagagttgttcaaaagaatagtaaaaggtcgaataaattatataaaaatggttagaggtgaagacgatataatatatcgtaaatttgctcacagaatatcttgtttatttggcaatgtaaataatagatatttgaaaactccatctgattttgctattgattcgatttttatcttagaaaatgaggtggatatatcacaaggtacagctttcctcttagaggatgttggtattgtaactaattatcatgttgttccaagtatagatgaatataatgatattgacttatctcttttttcgatataatgaattggataataaaagaaaagtaaagttcataatgtcaaataagttatacgacttg

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		gcgatattcgatactaatggcaattttgatgatataaagaaattttccataggggatgattctaatttaaaggtaggttcagaaatatctgttattggcttcccacaatataccacgggagagtacccttatataaataccggtaaaatagtccaatctaaagctcttttcaataataaaatctggcttgttgatatacctattattcatggaaatagtggtggtccagtttttaatgagaaatttgaaattattggcgttgcctcaaatgggacggagagaaatgatcagtcatcaaagttacatggcttcataccaatatcaacactaataaaatttattagcagtaaatgattttaatattaaagtgataagcgcccctgttacgcacacagagaggcgctttttttatttcacctctcatgatgaatcgtttcgagccaaaaaggcagagt (SEQ ID NO: 387)
31	RT-nitrilase (UG5)	agggatacgccacagcaagaaatagtttacttattcctcattttgtcgactaaaaatcgacattaaacaaaaaattcaaacttaatcactttcgggaaaaatgtgacaaatatatgctcggactggttgcggggagcgtgtaacatggatacaaatcaaaattattgccagcctcactgatggattactggtgtcaagagccccccttcgggcatgaaacggctggctaattctgtacagactgtaatctaaggacgataacgcatgacatatcaggcaattttcactggctgggatgatctgacgattgaagaccttctggtcgcttaccggaaagcaaaagccgatagcttctttgagaatacatttcctgttgctatcaaatttgccgagtatgagcaggaattacttgaaaacctgcaaaaactcttagatcttttgcagagcgaagatggattcagtagcaataagaagttgattggcaaatttcgtttgttaccgaaaaattaaccacaaagaaaaaacatgaatcccaaaatggacacgtccactttctaatcctaaacgagcagccgaccatttatttaataattttgatctgataccagagtttcgtattattggtgacttcccggttgatagtcacattatctctgcactatggattaacatggtcgggcataaatttgatgccagcttagataactgttgctatggcgcgcggctaaagcgtattcgtaatgatgaattatttagcaatgagcaggataatccattccatatcagtgccgtgggttcttttagcccctacttccagccctaccaaaaatggcgtggtgatggcttaaaagctatacgtgacgagttggaaaaagatcgtgacattatcgccgcctcactggatttaaaaagttactatcatttttattgatccactggctataacctctgatgatctctataacacactaaacataaaactgactgaggatgaaaaagcgtttactgcacagttagcagtattcttaaagcactggtctgacggcgcagcggcatttggaaagaaaatagcgtacaaaacacctgttattaatggtggtctggtcattggattaacagccagtcggatcatttcaaatatattgctacaccattgggataaattagtcattgaaaaactatcaccaattcactacggtcgttatgtcgatgatatgttccttgtaatacgcgatacagggacaattactaataatcacgaatttatgttattgctgcaagataggcttggcaatgattgcgtttatttgaaaaacgagcaaaaacaaatatggcaaatacagcagggcgagcatttccagggtaagaccaccatccagttacaatccgataagcaaaaactttcgtgcttcaagggagggctggaatagacctgctcgacagtatcgaaaaggagatctacgagctttctagtgaacaccgcttgatgccttcaccggatcaactggaacactccaccgcagctaaagtcctttccgctgccggtagtgtaggtgaaaatgccgatactctgcgccgtgcggatggattaaccattcgtcgtttgggctggtcactgcaattacgctacgttgaaacactggcacgagatctgcctccaagtgaatggaaagaacagcgggaagagttttatcagtttgcctacaaccatattcttagggctgataatctatttgcacattttagttatctgccaaggctgcttggctttgctatcagtatgaatgaatggcagcacgcggaaaaaattgtacttaaagcttacgaatccatcaacctgttggcatcggtgattacttcaggtaaggaagtgaatataaatggttgcaaaactcgagcagtaaatgatctttggcgctgtataaaaggcacattaagctggctatttgttgatgcagcgacacgatattacagtcctgacagattatttcttgataaacgttcaaagaaagaagagtgccttgcggatacatttttttaatcatatttcacaaagtctgacgaatctaaaggatttactggatcttcgctttgattcagcagattttttatttaaaagcgccattggtagctcgagctgatttagcaaaggaaccttataaacagatcgtaaagagtcagtcggcagaaaaacttgttaatcagcgtgatagtaaaaaagaagttaaaatactgaaattaatgagcgactcatcgcttattgatattgacgttattaagctatttttgaaatcaaccaagaatacccgactggaaaaagtggctaaaggaaatcgtaagaacgaaagttacctaccttacattttccctacacgtcctttaacacccgctgaaatatcagaactggcccccgaatgtgttggattaccctccacatccgacaaaaaaccagatgagagaccgtccaccatttgggcaaaatatactcaagcattacgcggagtatggatcaaaccgacgttgctagcatcggagcaggactcagatgaagcgacaaaaaaagctcggcctaagaaattcattcatattggcacagacaggaaacataaagttgtcgttgcgctaaccagcattaaaacagaggaggacgactgggctaaaatggcctgcaataaatctaacttgtcccgttcaaggtaccagcggatttctgaactggttaatgcaacattgaaactatctcctaaacctgattatgttttattccctgagctttcaatcccgttacgctgggttaacagtattgctgatcgtttgagttcggcgggtatcagtctaattgcgggaacagaataccgccacttagacgataatcaactgaagagtgaggccgtacttgtcctttcagataacagactcggctatccagcgagtgtcaaaatatggcaacccaagctggaacccgccgtaggtgaagatgaggcattattttcaatttatggtaagtcttgggattcgacacttaatgttaaacaacgtaagccggtatatattcatcacggcgtcaattttggcgttatgatttgctctgaactccagaatagtaaagcgaggatccgttttcagggcgcactcgatgcattaatggtattgagctggaataaagatctagatacgtttgcatcgttgattgaatcagcagcgctggatattcatgcctatactattttagtgaataaccgaaaatacggcgatagtcgcgtacgttccccggcaaaagaaccctttatgcgtgatattgctcgtgtgaagggcggtgataatgactttgtggtcgctgcaacgctggatatcgactcgttaagggcatttcagagcagggcaaaacgctggcctaaaggcggcgataaattcaaaccgttacctgaaggattccagttggcaaagaaccgcaaaaagctaccgccaaaataagaaactgattttcgctattaataatcagggtatttttgcgtgagatgttggtaaacatgatgtagcccttgccactcatgaccaatcgcagtatctttctcccgcgcctgcaaaatcaggcgtcgggattagcctcctgaagaaatcttatcggcgacacatgacgcgccagcgtcttttttttgtgttgttcgcacggttacatc (SEQ ID NO: 388)
31	RT-nitrilase (UG5)	ttttcaaaggagtttcgctttccaaatatacaagaaatcattatttctaaaggtatctataagtggatgattcgttttattggaacagttgcattctcgttaattaaagcggctgcttccgaccggcgaatggtcattcagaagctgagaatgtggttatttttttaaagaggaattggcatgattattagccttgaagagcttggccttgcctaccgaaaagcaaaagtcgatctgtactattcatcccatgtttcgctggaagcaattgcgtcttacgaagagtccctacatacgaatctgacggttctgcaggaaaaaatacaaggtgacgacgaatcatgggtggaagagaatgagttcactggcaactggtttctggccacaaaatctgtagacatgtcttgctgggaacagcagcgagaaccgcaagctaacggtctcatatttcctcacctgctgaaaagtgggcatatgcttgcaacccaatggctgataaaaacgaacaaaaaaaaatcaaagccgagtttcgagtaatggctcaatgcagtctggattttcatgttctctcgactctttggatgttaaaagtcgggcatctttttgatgccaaattatctacctgtgcttacggtaaccgcctgcgccgtactctagatggaaaagacatcaatgcactttcaattggttcttttcaaccttacctcagaccttttcgtgattggcgtgacaatggcattaacgccatgcggagcgcgctaagtgaaagcaaaaaaatcgtggcactcactgctgatgttagttctttctatcacgaactgaatcccgggtttatgcttgatccaaccttcgtcaaagatattttggagttggaactcactgctgaacaaagcaagcttaatcgattattcattaatgcgttaaaagca

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		tgggcaattgagactccgttgaagaaagggttaccagtaggtctccctgcttcagctgttgttgccaacgtagccctgatcgagctggatcgcgttattgagcagcaagtcgc acctatatattacggacggtatgtagatgacatcattctggtcatggaaaatggtgcgaatttccgttccatggcagagctatggcaatggttgttcgcccgttcttccggcaaac tggactgggtaaagggcgaggaaaacaaacagatcagttttcaaccaaactacctgcatgacagccagattcgttttgcaaatgcgaagaataaagtgtttatccttgcgggt gactccggaaaaaccttagtggaagctattgctcatcagatttatgaacgagccagcgagtggcgagccatgcctcggttaccgcattcctcgaacaatgttggaactgattt gcttgctgcaactcaaagtaatggcgaagtcgctgacaatttgcgtaaagcagatgcactgactatgcgtagggctggttttgccatcaaactacgcgactttgaagcctatga gcgtgacctgcaaccgggcacatggaaaggccatcgccaggcatttttcgggcatttattgatcatgttgtggtgctgccacaattctttgatttatcagtctacctaccccgag tgatccgactggccacggcctgtgaggactttgtcgaactgcgcaaacttatcttagcgctcgagaatatttgcgatgaagttcgagaaaattgcctccttaccatcaaggcgt gtcctgatgatcacctccctttttgaagcagagattattggcaaatggagggctcagctttttagcagtgtgcttgaagctatcgttgcggcatttcctccgcgtatttccaaggtgg gtaagcaaacctggaatgaccatttaaaaaactggcacgcccggtgtgggctagacattcaatattcgggtcgtgatttttcattaaagggctaccaagaacagcaggcgag attattctctttcgacttagcgcacatgccattccgctttattggtctaccaaaagagatgattgctcaacggggcatacccgctccgaaaacagtagcccactgtgcggaagc agcagaattactgcctgatattgtcgttttgggtaatcaggttgtagcaaaatggtgcaaatttaaaatcattccacatggactgctatttgccacccggcctttcagcctgccgg aactctttatcctaaacaatgaggcttatacagcttcagctcagcaagaaatgcgagctattattttcgctgttcgcggttttgtactcggtaataaaacaccttgtgtcgataaaca aggcatattgcaaatccctgacggccaatctgctggaaaatatggggttgccatatctagctggaaaacgtccatgtcaagctggactgcggcggtcatgcgttcagccgat ccggatgcaaaccgttacgctcgcttatgtcgcttgcttgatggtgtgatagcccaaccacataacagtcgttacttaattctgccggagctctcactccctgcgcactggtttatt agaattgcccgtaagttacaaggtcgcgggatttcacttgtcaccggcattgaatatttacatgccagtaaagcaagagtacgcaatcaggtatgggcttccttgtctcatgatg gattgggttttccttcactaatgatttaccgtcaggacaaacaacgcccagcactgcatgaagagcaggaattacaacgaatagcagggctagaaatgaaaccagaaaaga aatggacaacgcctcccatcattcaacacggtgattttcgtttttccttgttgatttgtagtgagctgaccaatattagttatcgcgcagcgctgcgtggcaacgttgacgcgctgt ttgtgccagaatggaatcaggatactgaaactttcaatgccttggtcgagtctgctgcgctagatatccatgcttacatcatccaatgcaatgaccgccagtatggcgatagcc gcatccgaggccctttcaaagatagctggaagcgtgatgtattgcgagtcaaaggtggtattacagattattgtgtaataggcgaaattgacgtacattctttacgacaatttcaa agtagctatcgttctcctggtaaaccctttaagccggttccggatggatttgagatagagcactctcgaaaaatgttgccagaagcataagtaaaattggaaaaaaatatcgatg caggttattaaagatgaggcaacatgccatagtcaatcataacctgcagatgtaatttgaaactgcatgttgagaattacggatttatttgtgtattcaccctcgcataaaaatgaa gtagctttcatattccacactactgataccccctgaaaatatataactaaaaaaaacaattttaaaacatgaggtaggaatagcaatctgactgtgatgtagttatttttttgatgaag ataattaggtgctcgttgttc (SEQ ID NO: 389)
32	TOPRIM-RT-nitrilase (UG10)	atgccccgtatcaacgttgagaaactgctgcttgagatcgaaatcgacaaggtggcagagcgattgggtatggcgcttaggagcgaatcagctacgcgcaagctcacgctg tgcccgttccatgacgataaaactccttcccttctaattgatacgagcagagataattctggacagcattaccactgctttgcctgcggtgaacatggagatgcaatcgatctgg tgaagggagttcttcatatcgatttcaaaggtgcattagagtggctgtcaccaaactctactaccacccctgtaaatagggcgagaaaacagaaggctatgcagcctgagca gccagaaggctcagggcttgcgcaagcttataagttatacctgttaagcaatgacaagcaacgactagctaactgggtgactgatcgcaagcttgatatttttttgatggaagat gcaggattcatatacgcacacaaaaactcactatctaaacaggtttcctcaagaaaagattttggaacgaagcgtgaattagcagcaacattggaagaagcgaacctaatac gcaaaatccttccaagctcggggttccaaaactactatttaaatctacagtcaatccacgacaacaactatatagactttttttcaggggatcgaatcgtattcccgataagagac gatcagaaaaaaactactaggccttgccgcccgggcggtagatgagcaaccagcaaaatacctattctcaaaaaactttccaaaatccaaagctatttttagaatagagcaagc tacaaccactctacgagcattggctaagcgaggcgaaacagatctacgcttatatatctgcgaaggatttttttgacgctctaagattggaaagcttgggatttcctgcagtagca gtaatgggaacatcaattagcaaagaacaaattaagattatgaaagggcttagcgacacgctcccttcaaagctagcctctttgacaatctgtatttgttttgatcgcgatgaagc gggattaagaggagcatccgaggctgtactaaaattcttaggcgctaatctcgacgtggtatttgtatggcctactactgctcagcttacaagcgcagaccattcaaacacaag cataaaagatcctgacgaatatttgagaaatttgtccgcgccgcaggccaagtcacttatcgatgtttccacctatggacctgtagtagcagtactagcaaatcagtttggtgtg catgccgacgaactgcttgaaaatctaaagtggaacagtgccagtcgctctcgaaaatacaggtcatttgagaaaactcgtgctgaactcaggaaagttgtagccaaccccc atctccaatcaagcgacctttttttaaatggccgaacagatcttgactcggcggctcaaatagaatggattgatttttttaagtgtcgacattgcgactgaagccgctccatcggaa tgttatcttaccaactcaggcaccagactaaaccacgcccgactgctcgcctatatgggctcacgaagaggagagttgccctgcgaagaatcaaaatgggagcggttagat attgcggcaagtgcattcaatgtgttgctcgctgaacgattggctaatgaaatacatggacccatcgacccgttcgaggccgtatgggtgccgaggtccttcggcgcagaag agccgagattaaaggtgatgcctcaacctgaggatttaatagcgcatcagtacttactaaatgagctacttacagaacgctgggatgcttccgctctcggtgttacagcattca gccagtgcataccagctgtccgctattaccgcgaagaaagaaaaactgttacgacaggaatatctacccccctcagataacacccaacctattatacttgaacagacgctaagt ttcgcctatcaaattgatatggaggttattgagggcaggcagccagcttcagatcagggaatgtttcgtccgttcctagactgctggcgagactttatgcagtcccttaaaaatc aagccaaatctataaattacgtgcatgttatccgcctcgatgtcagtcgatattacgaccgcatccgcagacacgtcgtaagagacagcattcaaccatttatacaacaagctct ggaaactgtcgctgataatgcaccggcgtttgctgaactgatgaaaatacaagcatctgcggatgaagcagcggacaaatccgcaataattgtcgagcaattatgcgacatg ctctttggctacccataccttagccctgataacgggagaattaataaatcagatcccttacgcggtattcctcaaggcccagtaatctcagcatggttaggctcagtggctttgttt ccagtagatctcgcggcactggaaatgatgaacaaatacaatgtagacggggaaactcatctagggtatgcaaggtatgtagatgacatagttttactagctagcagctccgt acttcttgaggaactgagagagctagttgatcaaaaaactcggagcttagacctggcgttggtcgcgaaagctgacgctattccgccaatgtctgctgaggaatttgcagatta tgcaaatcaagggcgagctttagaagcatctggtccagcgtgggaaccaccgttggctggcgatggtgaagcggggtgggagttttggtcaggcactccccctcagata

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		gacaatctgccctgcaactgctatcaaattgggagatatacaaaagcccaatagaaataatcttgcaaacagtgaaaacgtccttcctagctatggatttacgttctagcgagct tgcaaagggagcaaggctaatatggtacgttgtagcatccgacctcctctcagctgacattgatccaagcgatgcggcagatttagcgtgggaaatttatgatcgctattggaa ggaatgtactgaggagtgtgggtggcagttaaacccggatagtttcggatgggaggcaccgaatctgttcgcacttgagggactggaaaagcttatagatcataaaaatagc ctccaatcgggtttaactgctttagaaaataccgttcggcacaaacgcatctctttcctagctagaaccgtgcttggggagcggttcaaactgcatgctcttgaaagcagctcta cgcttaagcaccagatagataaaagactagatctcctcgaatggaaagcgtcaaaatcgtgcggaatgcccgttcgtagaactaaatcctacgcagagcgatcaatgtatatt cgctcctggcaacccttcaactggttccatgccgcagtagaagatttcatgctcgcggatcagtccagcggatccgacccattgagttcatatgtcactcagttccaatctatag aaaagagcatcagacctaatcacgccgcttcttatgagttcttccggtatttactgccatccgatggcagcgatagcgatcttgagtttttctcaaaaacagagaatcgatactcc ggcttagcaattcagattttggttgcattagtccctcgggaaagcataatacagattctctcaaatagagcgcgcttactttgtcctctagaagctggtaaaaaactattagtcatg ccccctcttcctggcgtcaatcagcaacgtatagttgcttgccagatcgatagctcctcagaaaacaaaatcaaaaaaatcagctcgtttgagtgctatgaaatagattcaacta aaaccaataccacatctctagactttttggtgcaaactctgcgggcgtagttgtgcttacacccacatggaacaccgaagcccaacctcaatccgccatacttcgatcaaact cagaagtcccgaaaaatcttttgttggaggtatttgagaaaccgtcaaccggtttcccttccgctattcagggattgaagcacgtagcctcactatatagagccattgtggtaata atggctgaatacgagaggcaaaatgatggtttagagcttatacccgcttggccataccttgccacagatatgacctctgggaactgctacctaatttgtgagggcgtaacgaa aggagaagtaggaaaccgagcatttgtaagagacggtgggcgggccctaagaaccattgagataccgatatacgaagcccagttgtggcgagccggggttgcgctaagc gattacataggcctgcacgacgatattgctaaatttagctcctccgaatccgaaatacctttggatgcgacaacgcttgccgccccgtcacagtacgtgctacgaagccaactt cgtaaactgaggggtgcctttgctaactcacaaatagggcggcgcgttatgcccccaagttttcttccggcaagtgttgaacgtgcgcttgagttattggagcattttccggaa gactcagatagtacaaagatgcagctaatgcatctgcttgccactgaaaccgaaactgcgggaatgcgcgtccgctatgagaaaaatattgaggtcacagagctcacggtat ttctacgtgcggtcgccgacagggttctaacgaaactacccttaagcataggtgaggtcattgctgcaccgactacagcagtcagtggcctgaggagagacctgagtgggg tcttgacccttgccagaagcatatggtcgatggatgaagaagaaaaactctctccaattttttgcgtggaagatttttcgagctggaattgtaggtattggtatcgctgttgctctac gggggattatagcttcactaagaagccacgggggggtttgcacgctttgagggatttgattttccagcggaatgggagcttccccctgccacagcagttttatccgaaccggcg acaacagataaaaccactgatgaaaatgtaagcctcctcgaccatttccgggtactcgtatcacatctcggacaccgaatgaggttggacgacaacggcgagccacaaatc ccagaagaaatcagcacagaaataagaaaatacgctacagcattagcgggcctcactactaaagactcaactgcggtggacgcaagcgactggcctttctttgatatcagc gaaaaagttttgataccctaaatatagaattattagagaacgtcagcaatctaatcaaaaacttagattccgcgcttggtctccaggtaattttggttacgcaacaatcatacggc ttcaatgctcaaaccaaacgcttcactgactcaagaggacttgcatgggatataaagccatggatgatctcgcaatacccattgcgtgctcgccacgttgaggagtgttttgatc aagaccgtagaatcgtacgtgtatggagcgagatttacgaaaaaaacagtcaacgcctgctttctatatcagtactaggcgagcctttcgcatcaattgcactatgtaaggactt ggaatcgccttatgccgagactaaaaatgtagacagcaagcacaacactgtattaggtcctagcgagcagggttctgaaagcgcacccatagatatttcaccgattcttgaaa ctgctgagcctgaggccgagactgccttagcagacacacaattaataccaaccccaaaccaaactagcactgaagacagctttgataaaatagatactgagcgtaatacaac acacaataaaaaactaccgcttaccgacgcaacactcaacgcccgaaagaattcatttagaaatagccagctaacagcctggagcgataggaagtccaataaaaaccctgc ccatgttcgggtagctctatttcagtgggaccaagagctgagctatgcacaccctatggtggaggccaccccacaaaaatggcctttcagttccgtctgtaaaccagcagtttt aaaagaacttaaacgcctatataactctccctatcaagcccttttgaatgcaactgaatctgccggtcaacaccacctatggaaaaacgaaaatatttccctacccagctggggt gagcttcgtcgtcggcgattattgctcaacgcagtgaacgcatgccagtcatttggcgtggacttattgatacttcctgaatactcagtccgtgcagaaactgttaagtggttaaa agaagagtgcttacccggaaagacggtagcggttttagcaggaacatttttagctttcgactccggtccgccccccctaaaacaaagcgcgagcctcaacctcttgtggccc gtaccgcgtgatattgccgaatgcctcaaaccgcttgcacccaaaacaaatgaagatgctatgtccttgagtgacaagattgacaagggcattgtattgcaatggggcagatc aaagaaataccgatcagtagctctaaatgagttcatccggcctggaactgatcctctcacccccctgttcatgcccggaaaaataatagatgaattgagacgtgcaaattggg atctggacgctgatggtgttgttaagttgctagccaacacagagttgccacttgcgaatttcatggagctgatatgctctgagattttcctgttcacgagcccaaccaacattcca gagatggcaagagattatgtttcaatgtgtgcaagatttggcttcggcgctgcagaagctcaagtctgggcggatctcaaactactatctaaatggctttcggtctgttccaagc ctggtggtgccgactctagacgatcaattttgatcgtacctgccgcgaccactcgtactgctgattattggatagcaggccaagctggcttgcttgccgccggcactacaactg tatttatcaatggcgtaggatctgggcttaagggtggcagttgttttattggcagagagagctggaaaacaggggctggttctcacggttacattgagaccattacgccatacc atggctggtcaaaggaatttactataatagcaaacatgacccactgagcgaaattgatcaagcattggtgatcgcagatatcgatcctcataacatgcttgaaggcaaaccta gacctcagatgctgccagttcccttacagctagtggcatacctaccaatcgttgaaactgtcgacgaaacaagcttggaccaaactctctgtgacgcagttcaggttgaccata acaatattgcaagaattaatcagggtcagcgattgggtggacgacttaaaagtcgaaatgagttctggcaacttatcacgcaaagtataaataatgatgtcgacaacgactttat cattaacttcagtaaatactttactgatgggaaagcgattcttgagcgagcaaactctttcttcaacaatggacaccaacagccttttttcatcggtagttaagctagacctgctctg ctctccggcactttacgactggctagaggccgatatgacgttgcgggagggtgaggcgttacccaacatctcagtcccttcatggaccaaataa (SEQ ID NO: 390)
32	TOPRIM-RT-nitrilase (UG10)	atggatcggtttgacattggtgaggtacttgcgaagtcgcctttagatgaagtagtacggcgcctcggcatcgagaccgagaggcggggaaaccaactcagtgcaatctgc ccatttcaccaagacactcgaccgtcgctgcgtttttttccagcggacagcagatctcccgagcattttcattgttttgcgtgtggcgcacacggccatgcgatcgacttagttaa gcaagtccaaagtgtagatttcttgccggcggtgcaatggctttcgcagagctttggcatcaaagacatccggcgacagccaaagaatcagccagatcgcaaaggcgccat cgaaggcgcacaggcattcgcgcttcggatatttgatgagcaccacgatacacaacgattccggacttggtgcgaagagcgagccttcgaggctgatttcctgtaccgcca gggggtgcgctgtgtgcctcactcggttctcgtgcaagagttggcgtcgagaagcacaggcgagcgtgttgagctgatcgatggcctgcttgctctcggcctgattaagcgc

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		ttgcaacaagcatcccattcggatcagtacaagcttagctttccagatcaattccaattgcagttccaagactacttccacgacgggcgtgttttgatcccgatctatggtggtgc
		cgcaaagcgaccggaactggttggcttcgcgggacgggcactgctggctgtgccgccagaaggagtccccaaatacttgttaagcccagggtttcaaaaagccaaatacc
		tgttcaatgcgccgagtgccttttcgtcagcaacgggggaactgagggacggcgacactgcaacgttatatctcgtggagggcttcctagatgccctacgcctgcaggcgtt
		aggcttgaacgcagtggcgcttatgggcacctcactcagcaatgggcagttagagctgctgaagcacttcgttgatggcctgccacagggcaaggctgagtttgtacttagc
		atcttcctcgacaacgataaagctgggtttgcagggacggatcggttggtgcgacgcctgctgggtttgtccggagttgatctgcgctggattggccttgatggctataccaac
		cgtccgcttggcaaggatccggacacttgtctaaaagtgctttcgagccgagtggaggcaacggactggttgcaggacttcaatcggccggccgaggcagccttgctggta
		tccgaattgggagacattgatgcctccgaactgccgaacgaacgctgggctgaactgaattccagtgctcgggagcgggcggtgtacaagactgcgacgactattcgaca
		ggttcgtggctcgcggcctttacagggcgtgattcagcgactgaaggctacagaagagagttgggctaccgaactttgtgaattgctgggtaccgttgaaggaacacagcg
		gaatcggagttccgtgttgtttctccagggcttggaagagcgcctctctcatgcccgaaatttggcgtatcacggatcgcgccgtggcgagctcccatgcgatgaagaatctt
		ggctgactttggatttgagtgcgcgcctgtttgatcgcattgcccaacaacgattggcagagcgtggctggatccaagccgccccatatgatgcagtccacctgccgcgcaa
		gcttacggctaatactacggtactggatgacccgcgtcgcaaggttatgccacacccggccgatttgcacttgcaacagttgctgctgaatgaactgctgacgcagcggcac
		gacttgctgagtgtcgaaggcaagaccttctcggaatggattcctgctgttcgctggttttctgccacccgcaaagtcgaagtgactgggccgtttgacgacctccccgctgc
		agaaggggaggagaccacattgagttttggctaccaagtagatatggatgtgctggagggcagcaagaccccgtcagaccaaggcatgttcaggccctacgggcagtgtt
		ggcgcgacttcatgagcagtttgagcaggcagtgccacgctatcggcggtcgagtgcatgtgcttcgactggacgcccagcgctactacgactccattcagcgttatgtggt
		acgcgatgcactactggactcgatcaaaggggctttgacgggaaccggggcgggcatcttcggcccactacttggccggagcgaaacagctagcacgcaggaggtcgc
		agaggctctggtcgacaaggtttgtaacttcctctttggccaccaataccggcccccaaatacaagagctgtcggctctagtctggatgcgattgggattccgcagggtccgg
		ttctatctgcatatattggtaccatcgccttgttcccggtggatgctgcggcgcgcaggttcatgcgtcgcaacgtccgaccggggcaggatggtatgaacctgccccgcgtg
		ggctatgcccgttatgtggacgacatcgtgctgttcgcagacagcgaagcgctgctggccgagttacaagaggtcctccagaccgagtcagctaagttgtctatctcactgat
		aaacaagggcgaacgcattagatccggcacgccagagcaggtgatgcaccagctcaatgagggacgcagtctggcagcttcggtgccggcttgggaaccaccattcgtt
		ggcgatggtgagtctggatggggtctcggcggcgatctgccagacgtagaccggcaatgcgctttgaaaatgctgcgacatcccgcactgatggacgagccgaaattgat
		tcaggagcaggtcaggcaagccatgcaggctcctgacctccgtccaaacgatctgggcctgtgcgcccgatggttgtggtggcaggtggccactgaactgtccaacgaat
		ctccgcaaaacgacccaagctcggcttggagtcgctactggcagttgtggcgacatgtttgcgaggggcacgactgggccggggagttcgaacgaaggggctacgcac
		agctatacgctgtggaaggcctggacaaattactcgattccaacccttggatggagaatgaacaaacccatagcgaagtaccgcagaaacgggcaattcgtattgggcttgc
		gaagctggtcatctcggcggggttcttctcggaggtgcaaccttctgagaataacgtgcatgtccagcggcgcgcgcgtcttgtggccggtaaggcgcggcagctttccgg
		cgggctgtcgaccactctactaagtcagccacaagacacgcagccggttacgacgatcgagtggttgtgcatggctgctgaattggtacgtgcggcccctgtcgatattgct
		ggcgctgaaggtacgcccccgattctagcgcccatcaagaatcgggttgctcttggcaccgtggatgctgtggcatcgcaggtctgcgaagtgctacggcttgcggatact
		caggatgggaagcttggtgacgtattacccaacccagtgcaggatgacgtagcgcggctagcacttggtttggtgatagataacgcgacccccaatcagcggctggctgtt
		ctgaccaagttcccgggactgctgagtatccgcagtaacggtgacgagctttccttggttcagcgtttacctatcacggagataacgtcactgtgggccttgggtgagccgca
		aaacggggctcgatatctctaccggttctccttgccccttcgccccttgcgtctcgagacctggcctgcgttgaacttgcgagcgatggcatgccagaggccaggttggag
		gcattgagcttcgaatctacgtcgctcggcccccaatcgtgccctcaccaattggtaagagagaagagcattgaaagtgtttcatgggcgaagtttgacttggattcatcgccc
		aatttgagtcggactgaactggcggttcgcctgtacgtcgcgctagtggccatgcagcggaaggacacaagcgatgctgatctaatgtacgttccttttgcaccacagctattc
		cgatcaggcgatgccacgcagccaacgctgcacttggttgcagaacctgtgaagcgccatacgctaggtgtgagcgcctggtaccgggattgcgatgggcgggtgcgta
		cggttagtgttccacacgtcggtgctgacctatggcgtgcgggctgggcggtggccgacgcattgggcatggcggtagacatgtcaggagaaaccggtctgcgcgatga
		gcaactgtcggacaagacgccgatctcggttgagcactatctactccgtcagcagttgcgcaagctgcagggtgtttacttgtctgaggcccagacattgcgcaaagatgaa
		cagaccggcctgccgcgcacagtaatgcgggcgctgcagcttctgggcgaattcgatggtcgtgcggaacctgaccagcaagtgcgacagttactggttatggaggcgg
		aaacacgggcgatggccttgcgtctacagcagcagggggcgagagtttgcacgcgctgttgcatcaggtgtttccagccgtgctgaacaaactgcccttgtgggccatcg
		attgcttggccctgcctaaccagcccgccgaacaccaaccgctgcggccagatttggcactcatgctgtcgttgtgcacggccatggagggttattggggccaggggggg
		gcagcgcatcaccatacaaccactccggctctgcgtgcggcgctagctttggcaacagcgggagcagggttgcgtgggagcgttgccgcgctatggggtctgacacagg
		cgcgtggtgccctgcggatgcccgagcgccttgacctgccagccgcttggccgttgcctgatatggtgcgcacggatccgcagtcggactacaaagccatgcgccaatgg
		ctcatcgaaggcgattggccagcgctgtgccgcaccagcccttggcactggatgctcgcgctgaccggtctgttgggtgccaacttcccacaggcttttgaactgcctcagtt
		gcagcaggtctttaccgcgttggcagcttggcagagccaactaagcgctgaggacggcgcctccgtatggccttatgatgggctgccagtactggatccgcagcagtggg
		cgacatttctcgacgcattgcctctggcgatcaggcaaatcgacgatttgcttggcatgcgggtggcccccctgtactgccccacggtatcgccgcaacccccataccggcga
		gttcaccgatgccagcaatcaagattggctgcttggcaagtcgcagttcacaggactaggtgctgttgaccgcattgcacggcgtaccaccggcggacgcattctaaacgtc
		tggacagagacccggagaaaggctgacgatgagctactggcagtgcatacgctggatcgaaagctgggggcctggttggaacgcgccgatcaccccgagacagcgta
		cgacggcacgggcgctcctgtggccatgccctcggagaagcctgctggcgaaatcgtcgagcaggtattggctacctttgtgccggatgtcgctgagtctgcctcagacct
		agcccaaagctctactgacgaactgacggagaagcctactggcaaaatcggtgagcagatattggctccctctgtgccggatgtcgccgagcctgccccagaccttgccc
		aaagcgctgctgacgaaccgacggagaggcctgctggcgaaatcgtcgagcaggtaatggctccctctgtggcggatgtctccgagtctgccccagatcttgcccaaagc
		tctactgaaaaaccgacgatgcagcctgtggctgagatggacggcggagccaatattgagtacagcaaggatgttgatcgcttggcggagcacctggacatttcacagaag

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		cagtcccgaaagagtcgtgctgatcacaagaattcgaaggcccatttccgcgttgcattgttccaatggcaggtcgaggacacctatacacatcctctgagcgaagtcggtttgcgaggcctgcccattggtgaaggggctaaggccgaactgcgtggaatggtcgctgccaatggtgacctctcggtcgctgacaaggccgccaaacggggtgaggagcaccaatggaccaacaacgtgaaggtcatgtcctggcatgagcacagacgccggacattgatacgtcaggcattgaatgcttgcaaggatcttggcgtgcaattgcttgtgttgccggaggtctcggttcggagagacacggttgagtggctcgaaggcgtactgaaagactttgaagggttggcggtactggcgggtacctatcgccacttttccaccagagcggaagaccgcgaccaccttcgcgcaccgctgacgttgctctggcggcccgagaccgaaatggccaaggcgcttgggcttgggaatgagaacacgacattcaagttcgaacgcggcaagaagtatcgtgcggtggctgctaatgagttgttccggcccgatttgagtcagctctctccgctctacacagaagtgaagctgatggaggaggtcaagagggaactcaaccgtcgaggacgaagcatgcttgggccagatcaactgcctgagctggctcatgcactggtgcatttgtcgccacccctgcgctattgtatggaactgatttgctcggagctctttctgctgaccagtccggccaattttgaaccactgaggaaagaggtgaacatgctcttgcagcggttcccttcgtactctgaggatacgaagaaattgattcgggatgacatcgaggcggtcggtgagctgctgactgttgcccagagaaaccgggagcggcgttcggtgcttctggtccctgcatttacgagccgcagtaacgactattggcacgcagggcaggccagtgtgttggcttccggcacggccactgtgttctgtaacgctgcccacaagaacagtgctggtgggagctgcttcattggcattaattcagtgagtcgctcgtcggagaccgcagggattgttaactctttgacgccttatcacggctggcaaaagggcatcctgcaggcgaactctgaaggggcgctttcgaagcatgatcaggcgcttgtggtcgtagatattgatccagtacatgtggtgagtggtaaaccgaggccacagctgttaccagagcccatgtccttggtggcctatctgccagtgatcgaactgatggacaaggaccaaaccgctgatggtgtagtgcgtgcattggaggcggaacttgaggatccaggcatgggggtaaagccagggagctgcttgcggcaacgggcttccatgcgcatgacaagttttacagggcttaccagacgcttctcaatgaaaaagggtctgacatcagcaaagcgcacggcgcaaaggcgttggatgattttgtgaagttcttcgcagacccggatgcgttgcgcaagcgtttcttagcttggcaagatgaacgacatcagcagccgagtctcgtgtccggaagcctgcagttggagccggcatggctcgatttcttggttgcggatatgacatgcatcgatcagatggccaaagtgagggtgccgccatggaaggagaacttgggaataggtgggccttctctagcgagtgactcgtga (SEQ ID NO: 391)
33	RT (UG7)	tctccacttcttcaaacatccgtatttatccataaccgcactgttttataaaagatttttgtttttactgttcgtattagtccataactttccagtagaatccagtactaaatgtgtataggattatgtatatgttcctgttcgattttggaattctatacacatgcccctaaatgatatgcagattcgccgtgctaaacctgaagctaaagcctatacacttggggatgggcaagggttgtctttacttgtagagccaaatggaagtaaaagctggcgatttcgttatcgctatgccggtaaacccaaaatgatctcgcttggtgtttacccaacgatcactcttgctgatgctcgttcccgtcgtgatgaagctcgaaaacttgtggcagaaggaaagaaccctagtgaggttcgaaaagagcaaaagctggctctgcaaacagagtcagagaacgccttcgaaaagatagccagagagtggcatcaacagaagtctaccaaatggtcggcgggatatgcatcagacatcatggaagcgtttaagaacgacatttttccttatgtgggaacaaggccagtgggagagattaaaccgctagaactgcttaatgtgctgcgtaaaatcgaaaagcgcggtgcattagaaaaaatgcgcaaagttcggcagcgatgctcagaagttttccgctatgccattgctactggaagggctgagtttaaccctgctgcggatctttcaagcgccctcaatgtacaccaatcaaatcatttcccgttcttaaaggctaatgagatacctgattttcttcgcgccttaaacggatataccggaagtcggcttgtcctgattgccacgaaattgctcatgattacaggtgttagaaccatcgaattacgtgcggcattatggtcagaatttgatttagataacgctatttgggaaattcctgctgaaaggatgaaaatgcgcagatcacaccttgtgcctttgtcgactcaagcgttagatttgctaaatgaactcaagatgatgacagggaagtatagttatgtttttccggggcggaacgatccgaacaagcctatgagtgaggcgagtattaaccaagttatcaagcgtattggttatggtggaaaacttactggtcatggatttcgacattccttatctactatcctccacgaaaaaggatatgattcggcttggatagaaatacagcttgctcatatagataagaataatattagaggtacgtataatcatgctcaatatattgataaacgccgtgatatgatgcagtggtattcagattatatttttattaaggagaatgtgaatgagtaacgagtttgatagtagtaaactagaaaattgctttgagcttgcattggaaaatattataaagcacggcgatacagatattttcccttacccatttgaaagtcggttatttgaagatgataaggagaaagtaaaaactgcattaatgcaaacatttaatgactttgaaaataaaaggatcgagattccaccaaacataattaatagcttttcaagtattggttattatggttacaggtgggcgacccaaattgatccattctggaatgctttttttcttgggttagttttaaaaatcgctgatgatattgaaaggaatagatctactaaaacgcaggtttattcatatcgctttaaaccaaaccttgctgatggttctctttttgataaagagatctcttggagaaaatatcaagaagacagtatctctgaatgttctaacgatgaaataaagtatgtacttacatgcgatatagcagatttctatccgcgtatttatcaccaccgtttagaaaatgcgttagatagagtcgaccccaataaagattactctgggaaaatcaagaaattactacagacatttagtgaaacaaaatcatatggagtaccagttggatgtcctgcctctagaatattagcagaactagctctagattctattgataaattattgtctatgaatagaatcaactataagcgttatgtcgacgactttgttatttttgtaactctagagaggatgctcataagattttaactttgcttagtaaaaaactgatggaaaatgaagggctaactttacagaaacaaaaaccaatattgttactaaagaagagttcctttcagtaactaaagctaagttgcatggtaatgatgaagatgaagaatctcctatgaaggctaaatttatgagtcttcctataagattcgatccttactcagcaaatgcgatagaggaatatgaagagataaaggaatctttaaaagattttgacttgttagctatgctgagtagtgagttacaaaaatcaaaaattaaccaatcttttagcaagcatttgataaaggcattctcagcaacatcagatgaaataataagtagtgctttcaaagtaatgtttaataacttgcatgagttatatccaatatttacaactataattcaagtagctaactccaactggcaaaaattaagcacagaaaccaaagatattattcttgataaaataactgcactaattaaacaagattcatatattttgagtactgagctcaacttagcctatgtagcccgaatgctctcaaaagaaaattcagaaaaatccaccctaatccttagtgaaatatacaataacaatccagaaagcatcttagtcaagaacatagttacacagtcaatggcaaaaattaattcttacgcatggctttctgatatcaaaaaaaatttctctgcaatgcatccgttgcagagaagactattgatcgtttccagttacatcttaggtgatgaaggacggcactggagagagcataataagaaaacattcaactttgtagaggtgatttacagggattgggcaagtaaaaggcataccgcaagaaatcttgaggatgcgctatgatatctgaattaacgttttctcgaaaattcacttcattttggaatcaattgcttccaaatgctaataatttcatacgcatcattaacggcagtctcatcgaggacgtttatcctcctctagatgactgcgctaataggtcaaataacgtctttgttaatgagtgcgcatttaatttatatagggcaatacagaatgattcgttagacagaaatattctttcagcacatgatatcttccataatgctgattttcaggttgtttttgaaaaaacaaaagaatatctacagcggttcgcttacggttctaacttcaagctacccttaagcatggttgagtacaatgccataagggaaatagcaagaaacattttgtctcgatatggaatggaaaaccaaattgaagtgtctccacaattcgatggatgcggagtaataaataattcatatggcgatatttattattcaaatgttcttgtggaaataaaatcaggagataggaagtttagtgtttacgatcttagacaggtgctaatatatttcactttaaacttttactcaaaaaacaaaagaaacatcaagagatttgagcttttcaatcctcggatgggtatcacttatagtgataccattgtcaaccttagcaaagagttggcgt

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		ttattcaacctgaagaattgtactttgagataatgaattctattacagaagaaaatttcatagtaactgaaatgcaacgctagatatcatgcagaccgctacaatccattgtagtggtctatttctaaacgttccttctgacgaataaagccaaaataccaaatagaattaaagaaaattataatatcagccttagcgcgcaatgctccccccgccacgcccgcccgctttgcggggcggttttaatgcagttgcactgacacgctcaga (SEQ ID NO: 392)
34	RT (UG9) + PolA	atgtcatataatgaaaatgactgggataaagaacatctactatcgtttccaataaatgtgaaagcggtgattgcacatatgcgtcaggacatgagagacgattggtttcctgatcctctatcctataatgacctatttgaaaaagcggatgatctcagagaagtactaatggagttgctgcttgaaggtaatgggcgctatgaagggaatctacgaaatttatgtaacatacccaaaaaaagggcttggcataagatattctctagaaactgattttacgatagatttatttatcaggcaatttgttcattttaattcctttttttgatccattactttcgcctcgagttttagggcatcgatataacaaaaaaagaactaaggaaaagtaccttttttaagtctaggattgaattatggcaaacttttgaaggtgtaacctatactgcaatcactagtagtaaagctttgatggctacagatgttcttaattattttgaaaatatatctatcgataaagtcaaagaaagctttgagttactaatcccccaggtgaaagcaaatggcgcggaaaaattaaagatcagaaatgcaatcaatacactctgtgaattactttgcaagtgggggttcagtaaatttcacggattaccacaaaatagagatccttcttcattcatagctaatgtcatgcttaattctatcgatcagaaaatggttgttttaggttatgattattatcgttatgtggatgatattagaataatttgcccagatataagtagtgctaggcgttcactaattgagttaattggtgcattaagaactattggaatgaacatcaattcaagtaaaacaaaaatacttacatctgattcagataaggatttggtagcagaatttttttccgtcacttgatgatagaagtataactatagataatatgtggaagtcacggaatcgaagaattattgccagatctgccaagtatattcatgcaatgattaaggattgtatagagagacaagaaacacaatctagacaatttcgatttgcagttaaccgcttgataaaacttgttgatgcaaatgtttttgacgtacattcttcattaggtgaagaattgcttgatatgattataagtacctttatcgatcacccagcctctacagatcaatactgtagattaatttgtgctttgcagccgttagataaacattttgaaaaaataacagatttcctatgtgatcatgattctgcgatacattcgtggcaaaactatcatatctggttgactttagcctaccataattttaaatcagaccagttaattgagacggcatgtgagcggttgaatttaatttcaaatgatccagaggttgcggctgtatttatatacttgtcttgtattggtgagacggaaaaactcattccggtaatctctcaatttgatgccagttggcctaacaggcatcaacgaagttttcttcttgcaactaaagatttgcctcaagactcattaaaaaaaatagttgaaaaattgacaattaagcttaggaatacggctagaagggctacgccacactattataataatcgcccgttagcagaacggaagtttcctaagattgttgatctatatgacgaggttaccacctatgattgatgctcaacctaaagtatttttatttattaaagattattctgagttaggtgaagataggtattttctattaaatgggaatgtcttctctgaggtttgtgcagagcaaatagtatcacaaacagagctgattgtttgccacgattattggttaatcgctccgtcaatttggatgtctattgggtcactcccatctttgattgtagatgtagatgaattccaaattattgtatctggaatgaagaaagaaagattgttaagagactgcaaggatatcacgagaaggtcgaatatatatgaaggtaatgaggacttatgttctaggtattttaaaatatttaaccgaactttaccttttgaagaggcggtttttagggactttagccttttactaagggaacattatctttcagttaaaaattatgcatctttaaatgatgagttatatcggtttgaaagtatagagattcctgtttcgagatatgttataaattcaatttgcaggggaattaaaataaatcagggccaacttttaatacataaaaaaaccttgagcatgatttctacactgcattgaaagaatactcagcaaaatataatgtacctcttgaagtacctgatgatcaagatgttatagagtatttagagcctatgggatatgattttacgggtgtagacgtggactatatccttaaatttgtccctatggaaagtaattacgcgaaagatgtattgtcgcttaggaaactatctcgatctagaaacgttcttaattctatacctttaagcacgcgccgtgcttatccgatggttgatactttgggtctattacttctaggatttatttaagagacccatccttgcagaatcttgcaaaaaagcatcgtaacatactaattccggacgatagaaagcgatttgtatatgttgactatgatcaatttgaggctggaataatggcagctctttcacaagatgaggagctgttatcattatactcggggaaagatatgtatgtgggtttcgctgagaaacttttcaataatataaatatgaggaaggacgcgaagaggttatttctgtcatatgcttatgggatgtcgatgaaatcattgatagatgcagcggtaggatttggtgcgaatagaaaggtggctaaggaaatattcaaaagctttgtctattttgaaaaatggaaagaagggatatggagtgattttgccagaagtggcaagattgggactgctaatggtaattaccttatacgtgatagagaggggccattagatggaaaagagaaacgttcatctgtaagtcaagtgattcaaggaacagcttcattaatatttaaggaagccttgatgtcgctggaagctttgaaagctgtagaattattattgcctatgcatgatgctgttttggtacaggtgccgttagatttcgaggataaagttatagcagaattgcttgcaaatgttatgtctgaccattttggacaaaagattgtaggtaaagcttctatcgacactttctttgaagattaa (SEQ ID NO: 393)
34	RT (UG9) + PolA	atgtcattatctaatttagagaataaaaaagacgatggtctatttcatttcccaattgatgttgatgctgtgcttcttcatttgaaacaggatatgcgagatgattggtttcctgactgtcttcagtatgaagacctttttttataagaaaaacaacattaccgaaaaagtagagggcaagattgtttctggacatggtgtctacgatactgacattcggtttatccacgatatccccaagagtactttggggttaagatattccctcgaaacagacttttacgatagatttatctatcaagcgatttgtagtttttttaatgccttattttgacccattaatatcgaatcgagtttttagtcatagatacaatgaacatcgaaccaaagaaaagtatatttttaaaaatagaattgacttatggcaaaatttcgaaggcatcaccaagctagggatatgtgatgataactatcttttggtcaccgacttacttaactattttgagcatatttcaattggaaatatccaaaaatcctttatagatttacttcctaaagttaaagcgacaggaaaagtcaaaagccaaattagaagcgccatccacactttatgtactttacttgagaagtggtgttttaataatcttcatggattacctcaaaatagggatgcatcatcatttattgcaaatatagtattaaccgccgtcgataaagctatggttcaaaaaggctatgattattttcgctacgttgatgatataagaattatatgcaaaaatgaatttcacgcaaaaaaagccttgaatattctcatatttgaacttcgaaagcttgggatgaatattaactctaaaaagacaaatatatactcttcgtcatcatcccaaagtgataaagaagaactattccctggtttcgatgaaagaagcattgccattgacaacatgtggaaatcaaggagtaaaaacgtaataatcagatctattccagaattaactaatatgttaatcgaactaattgataaaaatgaaactcaaagtcgcaggtttagattctgtattaatagaattataaaactagtctcaactggattatttaaaagtggttcaattctatcaaataaagtagttggcgcattgattaaggcattatatgaacaaccggcctcttctgatcaaatatgcaagcttttggttgatttaaaattcacaaaaaaacataaaatcgctttagaggaatttataaccaatgatgagctatgtatttacggatggcaaaaccatcatatttggatattattgtctctaaagaatatttccacaaaaaaaataattgaccgtgccaagtgcatttgcaatatacaacccataccatctgaagcatccgcatgcttcatatttttagccatgaatagtgaatttaaatacctagataccttagctgacaaattagacaggacatggtcatttcagctgcaacgccattttctccttgcaattagaagctcaaaaaaaacttcatcaccagagcttataaaacatgtactgccagcgatacaaggaaccgtaagggggggttaaaatgaacaaaaaattaaaaaatatttttattcatgcaaacccaaaccctgtctcttttttctgaaatctacaatgagttaagtccttatgattgatcaatacaacattcttttatatctaaaagactttcaagctaaagggaaggatcgctattttctatttaaagaaaacttgctatcggaagtacaagca

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		gatgaattgtttaatttagactcacatttaatcactcatgattatacaatcatttctgagagtatatttaaaaaatgccataaactccctaataaagttgttgacattgtcgattttaagaa atttctattacaagaaaaaatcaccgaaaaaaacaaagattcctttaagataaaagaaatcattaaagacgaattccaagacaaaaatgacttaatagaatactttgagatattttta taagaagaagcctttcaatattgatacctatctcttatttgctcataaaatatcagatggatatgagcgtttactcgctgaatcgttggcattaggagagcaggatagatatttcaac attgaaattccatgctataacgcattgtgcactcatctggctgctggcataaaaatcaacaacgaaaaattaaaagaatataagaacgagataaattatgattattttaaaaaaata aagtcatttagtgaaaccttcaacttcatgtatgaaatgccttctaatgaaagcatcaagcgatatgtcacagagaagggatatagtcttagcgaagagtctttagattatataatt gagtttattccaatgcctgatgattttggcaaaaaagttcgtgagttacaaaaaataaatgcaactagaaatacattcttgagcatgcctcactcaaggaacacaatttacccatc agttgatgtaaatggctccgtaacttcaaggatatatttaaagtcacccaccattcaaaatatatcaaaaaattacagagacatattcattgctgataaaggatgcgcgttgagtta tgttgattatgaccagtttgaagttggcattatggctcactttagcgatgacgagaaattaatcgaaatttattctgatgctgacatatacttaaaattctctgaggatgtatttggaac cgctgagaaaaggaaaattgccaagcggttatttttgtcttttacctatggaatgagtaaagaaaacctcattaaggtcgtcgaagaaaatcaaggcaacattagaaaagcaag agaattcttttcttcatttaaaaagtttgatgaatggagggcgcgtactgtacaacagttttcagacgaaggtagagtcgggacacttcatgggaatttcttgaagataaaaaacg caggagatctctcaaatagagaaaaaagatcgtgcattagtcaagttatacagggcacaggttcattaattttaaaaaaaccatcatcgaaatatctaaaattaaagatttaaaa ataatcatccccatgcatgatgcacttttgattcagcatcctgatgactttaatgctgatataattattaaaatatttgaagatgtcatgagcgatacattaaaaaatgaaaggcttat cactaaggcttcattgggaacttttatttaa (SEQ ID NO: 394)
34	RT (UG9) + PolA	atgaatacattcaaagcagaacaacttctaacatttcctattgatacaaatgcaacattaaagcatctacgacaggacatgaaagatgactggttttatgatgcaattaggtatga agatctactctctaataagactgacttgcaacgtgttttagctgaaaatcttaatatcaaccatggtaattataaatcaggtgacaaagctatttatgatgtgccaaaacgtgcattg ggtctacgctatactttagaaacagattttatgaccgctttctatatcaggctatatgtacttttttaatgccttatttcgatcctcttttatctaatcgagtttttagccatcgatataata aatatggtaattcaaagtatctttttaagcatcgtattgaattgtggaatacatttgaaaatattagctatgtttcactaattgatgataaaacacttttaataacagaccttctcaattatt ttgaacaaataaatattgaatcaattgaaagttcattcattagaatgatagcagaccttaatgtatcaggggcagaaaaaaacacgattagaagtgctattagcactttgaaagttt tattagagaaatggtgttataacgataagcatggattgcctcaaaatcgtgatgcttcatcatttattgcgaatgtcgttcttgattctgttgacaaaaaaatggtaaagaaaggata tgattattttcgttacgttgatgatattaggattatatgtaatgatgaaatggaagcaaggagagctttgaatgacctgatttttgaattaagaaagttagggttgaatataaattcca aaaagacagaaatactcaataaacatagtggaaataaagaggatttttttcctagtaaagatgacactatgactttaattgatactatgtggagatctaaaagtaagaaagttatc gcaagatcgattccaattctttttgagtttttaaaaaatcagatcgacgagggaaaaactcaaagtagacctttccgttattgtataaatagatttaagaccttgatatcatctaattta tttgaggctaaatcagttttagctagagagattgcagatacattaattggggagctagggaaacagccggtttccacagatcaattttgtaaactcttaatggatttggacttgtca aatgagcaaaataaagtcatatctaattatatagtaaatgaaaatgtagcgatatatggttggcaaaattataatttaatactacttatggctcataataaatattttgatgataatttga ttgatttttgcaagctgaaaattgaaaagaaaattaaaagcccagaaacaccagcatgttttatttatttggcatcaattggcttgcagaatgaggttgaaaagtttattgattctttt gataacacttggccatatcaacatcaacgatacttttttaatagcacttcaagacacatcaccaaaaaaattacaaccaatgtttggtaaggtaggatatcgtctaaaagggaccg ttaaaagattaaaggaaaataaactatttaaaggcgagtcaatataccttaaggattttaactcgactttaattcaagaaatatatcatgagatatcaccatatgagtaaaggaaaa gtggtttttcttgtttatcaaaaagacttttcagaaagtggaaaagaccgatattttatatttgataatgaaagtctttttgaggtaacagtacaagaactcgttagttataaatgtttca ttgttacacatgacttttggttgatttcaagctctatatataaaagtgcaaatgtattaccgaataagattattgatgttgtacttttagcaaagattgtatctggagttaaatctgttact agtgatactcaaccatgggatatatcaaaaactatcaaaccaatattctcaaaatctgaggactttaattattatatggatgtgtattataggaggaaaagttttgattttgacatatat cttctttttgcacataagctctgtgaatattttgaaagtttaagtgaaacttcctatcaacaagaggaaacgagtaggttttatagtttagaattaccagtatataatttaatgactttag ctgtttgtagagggataaaaatagataatgaaacttttcgagagcacaaggaaaacttacaattagattttatcgagaattaaaaaagttttctgagaagcatgatgtattgtatg agttaccaaaagaaggtgatattcgggaaaagttaattacattgaattatcatgttgatggcgtgtctatagattttctacttgatttcataccctccatagatggatatacggatgat cttcgccgtttgcagaagataaataaaagctatcaaatatttaattcaatatcgagctcctctaatagattgcatcctatagttgaatctcattggacatcaacatctcgaatttattat aaatctcctgcaattcaaaatattgctaaaaagtatagggatatttttataccagatgcaggtaagatattgagttacgtcgattatgatcaatttgagatcggagttatggcttatat ttcaaaagatcctatgatgattgaaatatatacgagaacagatgcttatagtgattttgctattaaagtttttaacgataaaaataaacgaaaaagtgccaaggtaatatttctttcat atgtttatggtatgtcaatggataatataaagaaatctacaataagcatgggagggaactctggcaagcttcaagattactttgaaaaatttgaggtttttgaaagttggaaacaa agtgtttggaaagaatttgagagtgaaggtcgaattggtactatcaagtctaactatttaaaaagggcaggtgaaggtaagttaacagaaaaagaaaaaagaatttctgtaaat cacgttattcaaggtacagcaacttatattttaagcttgctctgttagaagtttcaaaagttgatgatatagatatattgatcccaatgcatgatgcggcacttattcagcatactga aaaagtaagttctgaaaaatttaaagaaatatttgaaaatgttatgacagaagtattaccaggtattcaaggaaaagcttcattagaagatttctatatttcagaataa (SEQ ID NO: 395)
34	RT (UG9) + PolA	atgagtgaacaattcgtgtccgaggcggcaggaactccgcatctggcagagcaggatgatggtcttaaaaatctgaagttattgattgaatccttcaatacagacaaactgaa ctccagcgaacaaaagaaactccaagaactccggtccattctttcaccactactaaaaaaaggtggcgttttagcagacttatttcaagacgggaaagacgttttagcatttcc gatcgacgtcgacagtgtcctgcaacatttaaaccaagatatgagggatgactggtttactgacacacttcaacacaaagatcttctctcgaacaaacaatcccttcatgaagtc ctacatgaattgttaaatgaaggaaatggacaatatatcggctctttcaggagtgtttacaatataccaaaaaaagggctagggattagatactcgctagaaactgacttttacga cagatttatatatcaagcaatctgtaccttcctaatacaattttatgatccactcttatctcatcgagtactaagccacagattcaataaagatagaaaatcagagaaatacatattta

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		aaagccggattgatttatggcaaactttcgaaggggtaactagaacggcactcagcaataatcaatcactactagcaaccgatctaatcaattgctatgaaaatattacaattga aacaatccgcacagcgtttgagcgatcaattgaacatataaatacttccggtccaaataaagtattaattaggaatgcagtgcaaaccctctgcaaccttttgtcgcgatgggga tacagtgaacgtcacggcctgcctcaaaaccgcgacgcatcgtcattcatcgcaaacgttgtcttgaatgatattgaccatgaaatggtgcgattagggtacgattattatcgat acgtggacgacatcagggtaatttgtcccaacacgagagtcgcaaagaaagcgttgaccgagcttataaatcagctcagaaaggtcgggatgaatataaattctggaaaaa caaaaattttaacccaagactcgactgctaatgaagttgatgagtttttcccaacatctgacgatcgaagcctcacaatcgacaacatgtggagatcaagaagcagaagggtt attgcgcgttcagcaaaatatatatttcaaatattgaaagagtgcatcgaagaaaaacaaacacagtccaggcagtttcgattcgcggtaaaccgactaatcaagctgaccgat gcaggcattttgatattcatgcaaccatagcaacagacttaaaagcactcttaattagctcacttgaggaccatgcggcttcgaccgatcagtactgcagacttcttgggattct agacctcaacgagcacgagctcaatgatatttacaaccatctcagtgatcatgagcgctcggttcactcttggcaaaattttcatctatggttacttctagcaaatcgcaaatataa aagcactaatttaataacgctagcaactgcaagaatagagtccgacatacttcaaccagagatagcggccatctttatttatctaaagtgtgttggtgaagcacaagttttaattg ataacatttccaaatttgagtctgcctggccatattaccatcagcgaaattttctattagcctgtagcgattttgatcataatcaactgaaacctttaatttctaagctaggccctaaac ttaaatggaccggtagcagagccaagccttattttactaatggtatgcctttggtcgaacgagacaaaatagccatgcttgatctttatgatgagatcacaccatatgactgaatc caaaaaagccttactttttatagctgactatacagaccaagggcaagacagaatcttcttatggtcagatggcactttaggtgaagtcaccatatctgatttagtagatcaaaagc atgagcttgtctgccatgacttatggttaatcgccccatcgctctatcgggcgacaaacaaactaccatccaacatcacagatattgaagaacttcgaatcctcacttctggaaa gaaaaaagaaagagaatcgagagacaagaaagacatatcccaactcctgtcctcgtttgtttccgaagaaactattgcaagatataaagagatttttaaccgtaagatacctttta gatgaagctgttctgtcttcaattggcgaagccctattaaaatgctcagaagttgtaaaaagcgatgcaaatactgccggtgaatgggagagattcatcacaatcgaacgccc cgtaaacgactatctaataagatcaacatcagaaggtatttctatttctgaagaaaaacttagataccataaaaacaaaatagaattcgaattctatatggcattgaagagtttttct tccgactacgatatgcctctagaggttccctccgatcaagccgttatcgaatacctagagcctaaaggctttgactttaccggcctagacgtggattacattttaaatttcgtccct atgcaatcacattttgcagaggacttaattcgcttaagaaagattcaaaattcacgtagagtattagcagccattcccttgagccaaagtagaatttatccgatagtcgatagcttt ggatctatcacctcaagaatctacttcaaagacccgtcgttacaaaatttggcaaaacaccatcgagacattttaattccagataccaacaagcagttgtcctacatagactacg accaatttgaagcaggcgtaatggccgcactctccggcgatgagaaactattagagttatataacagtagcgatgtatatgaaattgctgcaaaagaaatatttgacgacaag agcaagagaaagcaagccaagaggctatttctttcttatgcctatggcatgaagcgacaacacatccttgctgcagcgcagggctttggtgcagatcgccaaaacgctaaga aattctttgagcaattcaagacattcgaagcttggaaagtcttagttcacgaagagtttcaccgtacgggaagaattggcactgcgcttggcaattatatgcaccgtgagcgaa aaggagaactaacaagcaaggaaaaaagatctgctatcagccaaattgtgcaagggactgcctcgttaatattcaagaaagcattactatgcttgagttcaatatctgaagtaa aactaaaactgccaatgcacgacgctgttttgctggaacatcccgcagactacgacatggatcgggtaatcaatatttttcagaaataatgtctgaacattttcaaaataagatt caaggcaaggcgtcattaagccaattccatgaagatctataa (SEQ ID NO: 396)
35	DUF4297-STAND	gaaatttcgcgacagagatccttaacggtgcgtcgagcttcgacggaattcagaataatgatggtctggtgttcggtgaatcgtgctttgcgcatggcgatctcctatcagaac aaaaccagtatgccggatgatctctaaaagtgaatggaccgatatgcagggatgcttacagtgggtcttcgacctttataagcatagtaaagaatagaatatgccaatgtacga taatctgtgcactctattacctgcgcaaaaaagtacaccagaattgtttgtctggtttggcaaattgagatcattaggcggcatagcgaatgactttaaatgaaaagcccgattca tcaataaagattgttaaaacaaaaaccttgcccccagcagagggcgagcgccgggcaatgcgtggctatatgggccaatatgaaagagccggtgcagccatttatgctgaa ttagagcgtgggcaattggagtggataggcgtagcggaccgcagtgcgggtatcgttgatgatttagtacttggatttaatggccttatcgttgggcaccagttcaaaacgtcc cgtttccctggtacatttacagtacagacactcttagtagggtctgatggtctgcttaagccattagtttgcgcctggcaaaatctttgtagtgctaacccaacgtctcaggtagaa attcgtttagttgtcaacgattatccatcagttaacgacgctcccggaatggaagctccagctcatagcgctgccttccttgatgagtttgaacattatcccaaacgcacgcttga ggaatggcgctacagtaactggggccgtttagtcgaaatattatttcaacattcctgcctaggtgacgatgatttcgagagatttttcatgcgttgcgcataattcatggttctgca gcagattttatacaattccataaactcagtgcagaacaagcgagactggcgtctgatatagcaaaaatattacctcgactggtctccgataaacgagatagggatcgatggtcc tgtgaagaactattatatgaactagggtggaaagatcccaccaaaacacgccacttacatcgttttcccatcggtgctcacgtccaacgcaaccgcgatacggaactacaact tctccagacgatacgcaacacaatccagggctatgtggcattgattgggcctccaggttcggggaaatcgaccttgctacagacaaccctagctaccgagtataacactcgg gtcgtgcgctatctggctttcataccgggcgctgcgcaaggtgtagggcgcggggaagctgatgatttcttcgaagacatttctgcccagttacgcagcagcgggctgcctg gacttcgccttcgagacagcagccaatttgaaaggcgcgaacaattcggtgaactgctcaaacaagctggcgagcgttatcaacgtgatacagtaagaaccatcattattgtt gatgggctggatcatatcccccgcgaagaactaccagcccattcgctgttaggggaattgccgctgcctgcagccatccctttgggcgtgacatttatacttggcacccagcg actggaactcaggcatctcaaacccgcagtacaggaacaggctgggcatccggatcgtctcgtaacaatgcatccacttgagagagtggcggtcgccaggatggcagac gttttaggtcttgattcaaccatttcgcgtgtaaaactttatgaacttagccgcggtcatccgctggcggccaattatctcattaaggcactgttatcggctgatgaacaggacata tcatgcatcctcgccggagggatggaatttaatggcgatattgaatcagtttacgcatctgcctggagagaaatcgcaaacgaccctgatgttatgcatgtactgggtttcattg cccgtgtcgaagctccgatgccgctgaaattgctggcaacaatcgtagatgctcaggcgatagagcgtaccttaaagaccgtccggcatttactcaaggaaacctcaaagg ggtggactgtattccataacagcttccgtctatttgtgctctccaaaccaaagataacactgggcagtatagatgaaacctattcacaacatatttatcgtgaattagctaaactat ctcgtcatgcaccagaacattcattacagtcctggctaacactgcgctatctcgcccggtcaggagagcgtgatgaacttctggcactcgcaactccagcatattttcgacacc agtttgcacatggacgttcctgttcagagattgatgcggacattcacttggctctgattgctgcgcgttccacgtatgatggtgtaattgccacacggttattactttgccgtgatg agatatccagacgaactcaagcactggagtatgccaatgaacttccgcgcgcgatgttaaaagttggcgatattgatgcggcgatctctttcgtccaggactttcccaatgcg

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		ggctatgaagttgttgaccttcttttggaacagggtgattttgaccgcgcgaaagaactgtttgagcaccttgagccattatctcaattgcataccccccagattcgagcactatgg ggattcgcataatctacaagaattcaaaaaatgggcaaaacgagttgttcacttccgcgacgctgagcaaattaagcaggcaatagactatttgaccgttgaggggtttaaac acgccacaagtgtatcaaccgatgaaaatatttcctctattcgcgaacagttaaagtggacagtggtcgaggcaattgttaactggcaatcagacgttaatattcaggatacctg caatcagtatggcattcatgtgcaagagataccggttttgatgactcaggctggatttattgctagagacagaggaaataacaccttagcatcggaattatttaagactgccatg gcattgtctgattttaatgatgtttctaatggggggcgaagatcgattgcattattttatgccacatcaggctgcaccgatctggcttcaaaattattcgaaaacctttttgcgcctgc aatttcgatgggagacaatgaattagaatcaacaaaagcactgacgcttgcagccatggaacatgcgcaactttgcgttttgctcggcaaatccttgcccgacgtagtcacctc aacacacgctatcttacgaccgctgcagacacatgcttcagaaacgggacgcttgttggggctgtccataataaatgcctcatgtattccttctggaaatattaaaatggtctgtc gcatggtgatgagatatgtaatgcaactcaatagctattctggaaacgatacctatcaggctcaattggcattgacagctacatcaccactgatttgtacattaattaaaatttctg cgctgtgtggtaaggttgaatattattcagtaataaatgaaattgataatgcaatgcctgctttaatattaaaaggcaatacactactccggcgtgaaatagcattggcaatgtatc aggctgacggtgaccgtgaaagggcggccgccagatttgagcctatggtaaacgagttggtagaaaatacacctagcgagcaactcgagactctgtcagttctggcaaac agctttgctgcaattggcgatgttgaccgggcactaaacttacttgcttcgatacatgaccactgtttaggctacgctctggcagcgcgtaaggaccctttatactctgtttggaa agacatattgattttggccaatgcggcagacccagaacaccgtgctcaacgaataggtcagttgatacgacaggttgatggtatgaaggaaaccgagggagcatctgccgc atatcgtttgacagaagtgttaatcaatgaagcaatgcgtatgaatgcgcacagtggttataccgtggcacagaaactcagcaactgggggctgattccatggccaaatcagg taaatgaactggtaattggtatgctagatcgccgtcctgaaatggtgtttctctgtacacaaatttggtgcgggctatgccttccattctacattgaaccctattatcgtgaccctac acatgtaggcaattatattgacgttgctgcaaatgcagcggggccttcatcaattgccaaactggtatcaattctattaccggcaatccaggttcatagtcgagctcacgagcga ctcacgctaataaatcgcctgagcaaggcggcattaagacacggttataccgataaccaacttgataatgccattactcgatggacttcagaggcccccgaagcccgccgct cctacacgccacaaacgtacgacgaagcttcaacccttgacgaacttcaacaggcatttgaatcaaatgattccgaacctgagtatcatgcgccttatcgtttttgtgagcttgc agagtccgccgcattagacaaggtggtgaaaatgtatgagtgctggcattgcctgcagtcggatgcacgttgtcgtttttttggttgcagagcggctagttaatgcgggggaca cgacgttagccagaaaattagttgatgattacgataccagtagtgaccgggagatgtcatggagccaatggttaggaggaaatcgattccgtctcttccacgcgcgtaagcta ctcgatggagcagcaattcatcatgaagcatatgaagacttcatcagttcaattgtggctgggaaagagagcaccatgtcgttgctaacagatatggcagacattcttcctgtg atctgtgagtcgccagactggcccgccgtctggtctatcctggcagagcagatgtctttcactcgcgaacaccgtattggtgaacttttcgaatttggaaatgaaaatatgaccg acgaagagttacttgcggaattgctccatttttcattacgattgcctatcaccgaagctcgacgacacgcagagaaaactgcactaattctggcggtacattcaacaggaggg caaatcgtatttgagaacaccataacacgactcctgaacggcacccttgatgaaccattccaggcattgcaaattttgcttttgctaaaacagaaccactttgctgctaaatttggt gatttagtctctggccttacgaatcatcgtgatgtagctgttgctgaagctgcgtgcttgttagcacaatattggcagctacctgtatcgattgattttcatccgttgccgttgaccta tcgattggcactcgacggagaccctgatcatgaaaatgctctgttagatcctgtgagtggggcaatgcgtattgaagtcgacttaggatggacacaaatgcttcgtcccgttgc acggagacttgcagagtttgctgattgtgacgaaatgaacatacgccagcgtgccgcaacgtttattcagcaatggggagggctggcagcctttggccctggagcaacaaa aaaaatcgaatctcagttacgcacactctcaatgcaaatcacctatcttaagccccatgcttacattggcatactggcacttcgtcatgtcgctggagagctgagcttggcaggc ttgctctcgccaagggataaaccatcgctactggaacaaatggatgcagtacttccgccaactcctcgccctgaaatgcaaatccggccaactggcattaggcgaccgctta aagtcaaggatgccccgtggagtgaagctgaagaaatgtggacaaatttggttgacgaggatgttaaaccctggataggtcgtgccgacgaattcgtaatagccgaggtttc acaattcaaaatgcatgatacccggcgtgctgaatatcaggtctatcgtattagcgcacctcaaattcatatttctgatgccaaattcatggcatggtatcaaagtttgcccgctgt cgtttggctgggaaaaatgatcccacttgacgaagacctcgcaccgacaatagtcaggcgtgtagtaagctccatcgggacaatgtcttcgccgggatatgccattgcattat gtcctaatatccagatgcatctgggatggcatgaatgctgcgagatgcctaatatttataccgaccagaactcaacaatcgtagcaagattagtgaactggcgagacgccgg gccagtggatattgatgatgattatatatgggggaaggttgctatctgacgctttccaatgcaggcctgatacaagtcaagactctgttcggcgaattcaccgtgcgtaatttc gcaagcagggctgttcggcaattgcgacaaggcgaagcgcaaatgataaagacagctcagaatcagttcccgatactgtagcgagacgatttcacaacacggttcgattac ctgacttctccaaccatggtctgaagaagtcagggagtgtagatcatgccggcattctgtttctgaatggcgcaggatttcgggtcagggtcaccacaacaggcttgtccttttc t (SEQ ID NO: 397)
36	DUF4297- STAND	ttgtgcgtagcacttctccagtttttgttgaaacagataaagagactaaatcgatcattcgaacccaaaaatggccgatttgatgcagacaacgatttaagccatatctggtagcg caatcgtcacctatgacaaaagttacatacttgtaatattctgaattcaatattcttcgtgaaattcattcaatgcttctttgagtagtgttttggcgttatgataatttcctaaatatcata aggttatcaggcggtgatgtatgaggcgatttgtctatggcgattaaaaacagcgcaatcatttatgcaggctatgattatcagacactccaaggtgtcaggctactggcggatt ggctcaatacaccaactaaatataaccgaatagcatttgaggctgatgcgaaacaagttgatgctccacaaggcattgatgatattgtctgcgaacgtcaggatggtaaaaca gatttttggcaagttaagtttacgccagataccgacaaagaagacaatcaactatcatgggaatggttactgaaacgtagtggtcatagtattcgagctcgttctatactgcaaa aaatagctgatgctgttgataaagtacctgcggaaagaaggggagatattactcttttgaccaataaaatacctaatcgtgagatagcaacttgcttgcgaaataacaaaatag attggaatcaggttccaattgctaagcagcaaagcattattcttcagttaggtacccaggaaagagcaaagcaatttttcgatatattacaaatatgtcatagtgatcaaagttata cgcgattaaatagtattgtcccagaactacttcgcaaacataccaacgaggagggggtatatcgcctgattgaacgagctaaacgttgggctatccagcgtaattcaccttcg gatggtggatggatatgtcttgaacatattcgtgcagtgatttcaactaatagacctgaacctattccgcagacttttgtcttgccagataactatattgttcctgatgcagattttca cgacaaattcattgattcacttttaatcctactaatcgattagttgtcttaactggtgctccaggaaagggtaaaagtacttacatcagccatatttgtcagatattacaaactcgcg agtttccttatattcgccatcattattttcttgggttagatgatcgtacgacagatagattaagtcccagaatcgttgctgaagacttgatgtgtcaggtcaaagcattttgctcacaa

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		atcgaaatgaaaaattatcatgcagagcacctacataaagtgctggctgaatgtgggcagatatataaagaagaaggtaaacgatttttcatcattattgatggtttggatcatgt ctggcgtgataacggcaaagataaatctccactggatgagctattttgccaattgttaccgttgcctgataatgtaacattattggttggtactcaaccagtagatgatgagctatt gccatcaagattgttacagaacagtccaagagaagaatggttgcacctaccaaatatgtcaggcgatgctattcgtaaatatctatcgggacaagttgaaagtggccgtatcgt attcaattttcatcaaagccagtatgaagaagttttatcacagtgtgctgagttgttgactactaaaactcagggatatcctcttcatgttatctactcatgtgaaaaattacatgttga aggtaaagggttatcgcactgggaaatagaaaacctgcctcgctgcgaaggcggaaacattacaaattattataatgaattatggaaaatattaaattacgagcaacgcgatat tcttcatctctgttgtgcttttcctttttttatggcctgccacatcatttctgagatttttctgagaggactgaaactataccgaatgttaaggctgtaatccatttgctttatgagtccatt gctggattaagaccgtttcatgaaagcttgattgtttttacccgtagcacaactgaacatgagaatagaataaaattattattgccagcgctaatttcatggctggagaaaagcgc acccaaaccgataaaaaattgttggtactggtcatgtcttgcttacaatggtgatccatatcctttaagaaatggcttaactagagactggatattggaacggttggctgaagggt atcgacaggatgagtttattcgattactcactcaggctgaaacttctgctttagccgaagggcattttagtgaggcctatcagcatcgttcacgcaagactcgactacttaatgct aggttgcaaatctgggatatgtcgacgttgggcgtttgcagtatgattaatgcttctgaagcattgcttaaacaatatcaatctacccagaatgtcagttcaccaaagatactggc aactttggctatcgctttatggtttcgtaatcatttcgatgaagcaaagcgcattacaagattggcgttacaacgctactcaaatgaatcatccgtatataccaataaaaatagcga tgagtcgcgtgctgacattcgtttattaatcaaagctgctgttttgactgagtgtttcgatgaaaaatggttggcaaccggttcagtacacaagtggagtgatagtaatattaatct gcttatcgaatgtgcggaatataaatcagatataggattactattttcattacatgatgttttaagcaaactgtcataaaaaataaaatagtaaatgcgattgtcagagttgggattg ttgaacaaatagatttagaatactggccacatttttctggtcttgactccgctctgctgcggttatacagtcatttatccactgcacatccatgttcacttataacagagcaaggtga aagtgaaatcggtagatatcatgttcatccagaagtatcctacgatgaatggttctatgacagcctttttttatcgtcttaatgccagtggagattattgttggctaccggttagcacg ggggaaggacaggaggaagtcagcagtcatttctccatttaaatgatttctcagatattattgctgaaagtatggctctaaatattcaacaaagcttcagcgattttgttcacttat tgctttggtatcagatcttaaagatcatcaaatgcaaatccaacagaagcgaatgtttttaaaactgattgggtaagcattgctttaaatttacacttaatcatgcattgcaagccg gttaatacggaagaaattgatattattcttaattctgagcatacagccctgtatcggctgcataaaactattcttaactttcatagtagagccttcgaatctgatgcaatagcaaactt tctggtatttgaggatgggaggcagaaggaaaaactacaagagacaaatgaatatttggcgaataatcttgagttgtcagagattgcgcttcattatgatctcaatcaatcaattt tttttgagcgagtcaagttatgttgggactatggtctgggatacggacatcataaagatatagctctgaatcaggtgctgactgcaataaaaactattgcaactgttgagcctaaa tatgcattaacgcagcttgagcgtgtgagtccattggttcataatatttgtgacttcacagatggtgaccatactcaacattccgtaacggaattgtctgcgctatatgctcatctttc tccccttactttaagtagtatctatgacagttatgttagcgagggtgagtggtatgatgcggataatgcattaacgcaatacttaaaacatgctgatctatcatcacctttcgttgag agtttatgccggacattactagatgatgggcaaattgaaataatacagaatcgtgctaaagacaatgccatattgactacgttttggccggaaatattaccacgaaaaatggatt atagtagtagcgcaaaacgttcattaaggggggactgaaaaatttgatccagcaaaaatcagccctgctgatgtaactaatttactcaatgttcggtcaagttatgaaaatattcct aagtggtatcattattggaaagaccaaggaaaagttacagaagtaattaacgtattgctgccaatcattaataatggcttgccagaatatagtgaatttcgttatatattatctgattt atttgaagatacattgcgtttgaaaggtaaaaaatatgcttttcccattttagtgcaggaacatattcagcgaaatggttggggtgaatggggggagtctgatgatcaaacatatg ctcggttagataaagttatcagattgtatccggataaaattgatgactttctttacaagacgactcgacttcatcactataaaactaaagaagagaacttggtaattcccgggaata agctaacatatttattagtaaatgtaggccgagtggatgaggcgaaaagtctatgtgaagcgatgatttcggaggtagaggcagaaacccagaatcttccgttgtgcaaacct caatggcaatgggagggagaattagataacgatatgatcgccgttaaattcatcattcgtcgtctttttttggcctgttcaatgtgtaaaacatcttgtcgctgatcaattgtctcatct cttagttaatggtcaatgtgctgaagaaattgaaaatttacttgtagttgagatgggaaatcgtcaactggagtcagaggtggtagatattttaactgttctctggttagctagtttg aaaggttataaggttcagaataatatatcttcctttatttatgctcgtagctttctttcagatgcattgctggaggctatcgttccaaatttaccaaacctcagtcgctatcaagtgctg tataaacatcctgatgatgatggtaatcactatggctttgaaaaaacacttggcaatgaacttccccatatattttgggatgaagtaaaaaggcttgaggagaaatctggagctc cggctaaaatattaatgaaaaaagaatggaatgatatttgttataatcatgttcaacgatgggaaagggttgattatttcttcggttcagagcgtgatggttttactatgagtttttcc acaaggaatacacgatttggtatatctgcatacttgagaaccattaaccggcttatcaacgaatttagaatgccaaagcattatgcagaacattattcgatttgtttaatgtcagcc aacccattattttattccgtatctaatcaccgacctggttggttacctttatggcaatatggggagattaccacaaaggaaaatgtaaaaacatatgttgaggaatgcctgaatgc attcaaaaatgaacaggaaaattcaatattaggagcattgtcattacctgtacgcatcgatgaaaataattggttagatattacggctgttatggggatacaaacagaagaatatg cctcttttaagatacaacatgccgactgtggtcatagtgtagatagtttacttcaagcttatagaaatattaaattttcatttgcaaaatgggctgaataccaaaattgtgtaccactat tgggaagtacacgcgaattactgagaatagcacggtgggatataatgtacgaatttcgtgggcttttctcattcggttgccaggaacaggttactgcctacccggctaaaaatc gtattaacttcgattatcagggtaaaaccatcggctatagtgacttctggcaagcaataccattatcaatttatcctaaggatatacgctcacctgttgctacttacactgcttatgat aaggaccttgcctgtaactggaaaaatcatagcgtactgaaaaagcctaatatcatgttatgtgattgtaaggtactaaagagagaaaatagttacagtccttttgaaatatcaga tattcgttttcactttgaatctgagccgttatagtaaggattattttgcgataattaatcaacggggagctggtcaaagtgcctgctcccatattgactaatatacaaatgtgtttgtta agacctttccaaaggtaggggggaattatgaatttccgctcctcgctcatagccgcctgccagatttaaccccaccctaccacagggccccctcaagccaagccgccgccaat acaattttcccccacaccaaaacgcctccctccctagagcacgtactcacaacgccga (SEQ ID NO: 398)
37	ATPase_ GHKL + Helicase_ SF2	atggctaaagcgcactccacgccgctcaacgatattgcgattatcgctgcgaatttaaaagaccgttataaaaatggcttccctgttctgaaagaaattgtgcaaaacgcagat gacgcacaagcgtcatcattaatctttggctggagccctggtattgctggggcagatcaccctttattgggcgatcccgcgcttttctttatcaataatgcgccgctgacactcg aagatgtagaggggatcctctccattggcattggcactaaaccgggtgatgaaaatgcggtggggaaatttgggctcggtatgaaaagcctgttccatctcggtgaagtatttt tttaccagtcctttgactggcatactgcttcggccaaatcagacgtttttaaccccctgggacagttacagatcttcttgggccgaggtgagcgagcaggataaagttcgtattga

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		ggatgaagtccgcgcaattacccaaaatgcgtgtgatgattatttcgttgtctgggttccgctgcgttcagagagtatctatcaggcgcgccaggatgatgaaaactttattattg
		tcggcgaagactatcgttatgaggtgcctgattttatttcagacccgggactcggggataagctcgccagcctgttaccgctgatgaaaaccttgcaggacattgagctggtc
		gtgaaaacagggcaggggtatcagcgtcaaatacatatctcgctgcctgaaaaggcaactcgcccacaatttaccaatcttaatggtgctggggaatggcaaggccacatta
		ccgttcagcgtgctggattgccggaccctcagcaaaaattctacgtcgggcatgaggttttgctgaatgctcctgagttttctgccctgaaatcacaacgcgcctggccattca
		gttattcacgagaaggtaagaagactgcggataaagcgctgcctcatgccgctgtggtgatgctggcggagaaagtaccagaaggagaggcaacgctggcggtggaatg
		ggcggtgtttttacctttgggtgagcaggacaccgcgcagcatgcgcagaaacaaacattctctatttctggtcagtactcgtatcaaattattctgcacggttactttttcatcgat
		gccgggcgagtgggtatccaggggctggctacactcaccagcgccacgccgttattcaatgccccagattctccaggccaggaacaactggttcaggaatggaaccgctg
		tcttgctactcagggaacgttgccgctattaccgaaagcgcttgcctctcttatgtcgcttattcacgccagggatgcggaaaaagcggcaatttcggatggtgtgcgtagagc
		tttacgcaacaataatgcctggttccactgggtaacgttgtaccatctgtgggtatgcgaactaacgcgggatggaagtcagtggtgtttagttgatgcgaacactcccgttcgt
		cgattgcctgccacaccttcaggtgaagcgcatcgcccctgggaagtgctgcccgctctggaaagtctgggtgtaacgcaccgatttatcgatgaaacgcagcagaatatct
		acaacgaatttaaaagtaagtggcagttgtcggagattcaggtgttgctgcatagcgtacccgaaatggtgttcactagcttaaagcttacaaattatctcaatcaattgctgaaa
		gaactgccgattcagtcagacagctttgtgcttgacctgattgcattgctcagaaaaacgttatttagcgtgccgctggttgagctctcacgtaaccaggcggcgatcggagaa
		ttgatggcgttcattcgtccgacctggcgttacaggattgccattgaccgtcaggagcaggccctgtgggaaacgcttgggcgtaccgctatggataggttgttggttcctgct
		tttctcgataacagtaaagaacctgccagcgcatctctgaattgggagacggttggcagcctgctgcaagcgatgcagaaacaggcttctgccagcgataactttgaaaaatt
		ggtgcgggattttattggcaagctctcatctcccgatcgtcaggagctataccgtcggtttgataccttgaaggtctttaaggtttcacagccaacggggatatcttacctggag
		acgcgctgtcacttgcttgaactaaaacaaaagcgaaggatattcaaacttggcgggagcgctaattttggtatgggtttaagcgcattgttgcagcaggcattgcttgaaaaa
		gaaatcgtattgatcaccaatgatattaaccagaccttatttggtggttctgaatattcagaagcaaaggagtgtgacagcgaaggggttatccatctgcttgagcttcaccctcg
		tctggattcgccgacaaaacgtatcgatttactcaataaaatggctgcggacggggacaaatttagcgccggagatcggcttgtctatcgctatctgatgcacggtaattcgga
		tgatactggtgaagctgaattgtggaaggcgggtaaagcgcatcccgtatgggcaaaaattctttctgatgccgattcggagcaggtcaagtggactattatttcgccagaaat
		tgagcagaatcttggactgactcccggattcgagaaggcgcttaggcttgatagtgtaacgccggatcatgtgatccaccgcttcaaagaaagccttgaatatctggagtttga
		tgacttatctgcagaagatgcggaagaagttctgatgcacattggccgctctatgggcgaaacaatgtggcggcagatggctcttcatcgtagggaaggcaaagaggggta
		tatatcccttgatgatcgttgtttcttgcgtgggggcgcattgaactgcccactgaattgaatgacaacgtgacgttcatccaacccgccagtcagccagagatgcaggatca
		gcagcgcaaatatctgacaatggtgaacgccgaacatgcggtcatgctggctttatccgggccgaacccggaacgttactgcgactttatcctgcaattgttaatgcaaccga
		cgaatgatttgtcttcagagagagcattcaataacctgcgccgccaaaaatggctattgcaccgcggtgtggcgatggcaccagaaaatattctggatattagcgcggcaga
		ctatccggagatcgcgaagctgacagaagcgacgccgctcatcgctctgcttgaggatattgctctcccagatgaggctaactgtgcgctgagttcattggtcgtgcgaggc
		aaggctgcgttttacaaggcgctcactgtagcaggtacacttccactttatgcaatcggtagcagcttacgtctcactgatacgattattcttcaggccagtgacaggtcgtacg
		cgtttgagagctttgacggttggttgctcttaattgagtgtctcaaaggtgctgagtcgcttgagggtaatgaggctatcaatgcgctgagtttttcgcatccggttacagacaag
		atagttgctagctaccggcatctcgttgacagcatgaatccaacccaaagtggtgaattgcgtaaagcactgttaagcacgctgtgtcatacccattcagatcccgccagcgta
		ctgcgttcaatcccgctcagaacggctgctgatacctgggcgttagccaccaatctctgttatggcgtaacgggagcagaacgtagtgctgtcctacatgacgacgactggg
		cgtatttgtccccttggctgcaggctaatgacttgtcggtagacagtactgagtccgaagggcatctcagtcatgttgagcattctgccaatgtcttaagggaatactttgcgccc
		tgggaacgctgggttccacgtaaggcaattgctgcactgctggctttgctggcggggaatcgtaaggttcataagctatgtgagagctacctggggttgcaaagttatgccct
		gttcgtgaatgaactgtcgcaagacagcaaacccttaactaaccatgacgctcactttgcagagttaacgctcttacagtgcattgagaaatatgcctttgccgtgaaggtttac
		gaagaaaacacgttgcaggttcattctctgttccaggaacgtttgaccgtggcgctggcaactgacctggatacgatctttgtgggtcagcacggctacgctttttataccggtc
		aggcaccgcaaatcttcattcgccgattttccccagaccagtatacgcctcagcaacttttggcgattctgaaacgcagcaccagctggctgcaggaaggtatttatctgcaga
		aggcaaggctagacacgctctggcaatcctttgagcaggccgagcagttggatgtgaatatcgcgcgcgtcactatcctgaacagcattgttgagcgcctgaaaacactgg
		gccttaaaaactctcagcttaacgttttaatgagagcctatgagagtgagcttcactctcttgctgaaagtagtgacggcaagttgctccacagctcgaggctcactgaaattgt
		ctatgacattgcaaatgctatccaggatcgccctgaactgcaggctgaaatattaacggcggtcagaaagcgtatagaggatgctcagtatcagccatcaagcgttcctttga
		gctgttccagaatgccgatgatgcagtagaagagttgttcaagctggatagcgatgcccgtcatgagcgggtacaccagaaatttatggtgaaagagcaaaacggcggatt
		gtcattcttcaactgggggagagaaattaaccgctttcagagcgtgaaaaatgagcaagtcgagaatgtacatgatggctacaaaaacgatctgaaaaaaatgctggcgcttt
		accagtcggataaagagcagggcgttaccggcaagttcggtctcggcttcaaaagctgtctgctggtgtctgatcatccttacctattgtcggggcggctggcgactaaaata
		gcgggtggaattgtgcccgaatcctgtgatgctgaaagttataaacaactaaaccaactcactgaaagtgccgcgacaaatggcctgtcacctactcttgtgtatttgccactg
		cgccagcatatgcaagcggaagtggtgttaaaagattttactctgtatgcaggtttgctaagtctttatgcacgtaacttgtgccagattgtcattgatgagcatgaatggcgctg
		ggagcctgttcagtatgcacgtattcctggtctgtcattgggcaaggttatgctgcctaacggcaagggtgctcagtcgccagtgcgggtggtggtttaccagactgaaatcg
		atgatgagcgctgccatctggttttccaggtcacgcgtaggggcctgagaagttttgatactcatattccgcgattgtggaacttgtcgccattgatgagtgatacccggcagg
		gctttttgattaacgctggatttgaggttgatattggtcgacgccagttggctattgaagctgaccgtaatcggggcattatccagaaagcgggagcaaaagttcattcgctgct
		ggaattactttggtgggaaacggagcataactgggaggagctggttgttgagtgggaactgagccctgaattgacccatactcagttctgggaaagcttctgggacgtgatgt
		ctacaggcattagtaacgatattaacgcgatggaaaacgaaaaattgctacagcagctttacgaaagcgaaaatggcatcatgagcttctatcgctcatatcccgcgctgcct
		aacggatttaaagagcaggctgccggactgataacgtggagcgacagagtgcgtagcgcggatgaactggtttctcgtctggcgagttcactgattcatctccctgcgtttca

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		ggcattgcacagtgcacagtgcctggtggcagacacgacgggaagcaaacttaaagtcgaaagtaaactgtcgcttgaatcattaataagctcgtcgttgccggataaacagggtgttgatatccagcatctgtcaccgcgggatgctgaaaagctggcagtcgtatttaacgaagagttcgacaagcgactgggtgaactgacaggctggcaggacaaaattgaggctttcagaaaacagctgataaacctgcatgtgcaaacacaagcaggctctacacgcccgattagccaaattttgctcggtaacactccttgtgccgaaaaaaatgaacggatgatctctgggtttgcacctaccgatgccatcatttcatcatcatattctaagcaggcctgtgaatttattgtttattgcaaacgcagaagtcagggatatgtttttgaggatttagtcaaatgggcaaagcgcaaaggcctggcggctgataatcaaaagcggcaggcattttgtcgttttctgattgaaggactggaaggggagaaactggcgggtatgctgatggaagagataccaccggactggttgcttgaacttaagctgcgcccaggcgccttcccggcagactggcactggagcaataatgatattgcctctctcctgcaggggcggttactgactaacattgacagaacaaaggcatgggagcgcgagattcgggagacaccggaagaatacgaaccgttggtgacaccaggtgaagccgtacaaaaaatacacacctggtgggagaggaaccagcaggaagagttggtgaaatacaatgctcggctctaccctgaaggctggtttgactgggaagctttaagaaatgcctctgacgatcagcgttcacgcctggcgttattgaaactcctgtatctaggctcatgccagaccattgggcggactcaggaagagcaacacagtgccgcaattgagtattttgaggacaaaggctggtgggaaacctttatcaaccctgatgcagcgcagcaatggctggatgtgatggacaattatctggaggattctttgtacggagatacctaccgtatctggctgcaaatattgcctctgtatcgtttttcaaagcatttagattcctatcgcaaactactggatatgtcggaagcgttccttgaggatattggggatttgctgcgaccggcatccagtttcaatctttcgggaacgggcgtgggaactgtagtcccggagttacgtgcaactctgggtactggggtgaacttcatcttccgtgaattggtgcgtaataacgtatttatcgattccagcattcatcgatattgtttctctgcgccggaacgcgtcaggcgtctgttactggcgatggagttcgacgaaatggatgttaagcaatccactgccagtgactcgcttctgctgtggacgtttttccgcgaacatctcggtgaggaagatgcgacctttaatcattgtttcgacataccgctgcgcattttaaccagcgaagggaaacgctcacttcgtattgagatatttggacaggatcccctggattacgtatgaaaatgatctttcagcagggccagcaggtacgacatgaacgctttgggctggggacgattgaactcttgcgggaaaacactgcactcattcgtttcgagtcgagttttgaagaacgtccactttccgaactggagccggtgcgcagtgctcaggatgctttggcagaaggaaattatgacgatctgcgtgaagttctggcgcgcagtcaggcgcttgcgatccgctccatcaatgatagttgggggggtgttctctacttcacgtatcaacctgctgccgcatcagttatgggtatgtcaccgcgtgttacggcaatggccggtacaaaagctgattgctgatgacgtagggttggggaaaaccgttgaggcggggctaatcctttggccgctgctggctaaaaagcgtgtgcagcgtctgttggtttagcgcctgcatcgttagtaccgcagtggcaggagcgtttgcggcagatgtttgatattcgtttgtccctctactccgcggaaattgatactgagcgatcagattactggaatacgcatccctgggtggtcgcttcattgccgacactgcgaaaagatattaatggcaggcacgagcgaatgctcaaagcagacgactgggacttgctgatcatcgatgaagcacatcaccttaactcgctagaagattcgggggcgactcagggctatcgatttgtgcagaagcttatcgatcacggaaagttcgcctcacggctttttttcacagctacccccccatcgcgggaaaaattacggcttctttgctctgttgaggcttttacgtccagacttatttgacgtgaataagccatttgaaactcagcagcatcatgttcgggatgttgtgattcgcaataataagcaaaccgtcacgaatatggacggtgagcgtttgttcaagaccgtcaacgtgacctcacagacctatcattttttctgaggctgaacagtcattctatgaccggctcacacgatttattctttcagggcaggcctacgcttcgtcgctaagctctgcaaaccagcaggccgtgcaactggtgttaacggcaatacagaaactggcggcaagttcggtagcggcaatttatgccgcaataaatgggcgtatcgccaggctcggggaaaatcagaaaaagctgcaggcgctgaatgatgaaatgaatgccatcatgagtgattctcaggccccggatctcgatgatgcctacattgcgcttgaaagcgaatatgttgaaatgtctgcttcggttcaacttatgcaaaatgagctgcccatgcttgaagagctgcaggcgcttgcggggaatgtggaatcggaaacgaaaatccagaccttgcttcatgtgctggaaaacacgtttcttaatcgcaccgtcgtattctttactgaatataaagcgacacaggccctgctaattaatactctgaatgctcgctttggctatggttgcgtcagctttatcaatggcgaaggacgcctggaagggatttacaataaacagggcgtcaaaacgtcatggagtatggatcgctaccatgctgcggagcaatttaaaagcgggcaggtacgctttattgtttgtactgaagccggtggtgaaggtattgatttgcaggacaactgttattccatgattcatgttgatctgccgtggaatccgatgcgtcttcaccagcgtgtagggcgactcaaccgctatggtcaaaaaaatcaggttgaagttattactttacgcaaccccgatactgtagagtccagaatatgggacttgttaaacagcaaaataaccacagtcatgcgttctttgggcgacgcgatggaggaaccggaagatctgttgcagcttattcttgggatgagtgataaagtttttttcaattcacttttttgctgatggcctgacacaaaagccagaaactctaaatacgtggttcgattctagagcagggaccttcggtggtcagtcagccgtcagcgtggttaaaggtcttgtaggccatgcggataagttcgagtatcagaacttagatgaggttccgaagcttgatcttatccatatgtatggtttcctcgagaacatgctgaaattgaatggacaccgtctggacaatgataagggtgttcttagctttgtcactcccaaagactggatcacacagtttggtatcaagaagaaatataacaatatgacttttgaacgtgttcctacagagaaatcgttagaagtgcttgggatagggcatgtgattattaatatgctattaatcaggctgagaaatttaacgcctctacggcagtagcaaggggtatttcctcagctttactgatttacacattgagagaccagattactggcgatagtaatgtacaatcattttcagttgttggagtggtactggaagataatattcaaattttggtcaacgctgagttagtcaataaactggcttttatatatgacaacctacctaaaggttcgacggtgattaagcttgacagtgcattccatgttaattttgagagggatataaagcgtgctgaggccgcattagatctctttattcctgggttgaatttaccctatgagcaagtagtatggcaacatacagcaactttttttgccacagtaa (SEQ ID NO: 399)
37	ATPase_GHKL + Helicase_SF2	atggcgggtgcttcaatagacgctattggtgtgattaaccaaatcaaagacaacttaacagaccgatacgaggatggctttcctgtccttaaagagatcattcaaaatgctgacgatgcgggtgcgaacgaattaactattggttggagtaaaggtttctgcaatgcagaaaatgaactactcaatgcgccagcgctgtttttttatcaatgatgcaccactggcagaggaacaccgtgatgccattttatcgatagcgcagagctcgaaagctacatctaaggcatcagttggaaagtttggtttgggaatgaaaagtttgtttcatatgggtgaggcattcttctttatgtccgatcaatggcgaattgagcattgggcgtcagatgttttcaatccatgggataagtatcgtgatgcatggaatgaattcggtgaaaatgacaaatgccagatcgcaacaaagttaaaagggtttttaagtaccgataagccttggtttgttgtttgggtcccgttgcgtacaaaagcgctagctaaagcacacaataactacattatcatcaacaactttagtggtgatgaaaaactccctagtttctttaatcaggctcacttatcagagaaaacttctgagatttgcctcaactcaagaatctcaaagacatcggctttttctgcgagtctgacaagggtgtgtttgatgaagtgacctccatacagttacatgaagattcgtctcgaagctctttttgcggtgaaccgcgattaaataatggagactcttttgcagtcttctcagggaaaatctattcaaattcgaatgaagagcgttgtgcactggactatgcaggatgcgagcgagtcatctttgatgagcgtttaaatcaattaaaagacgaaaatatggggtggcctaagagttatcagttcgacaagaaagcgaacttgcctgttgaggctctcgacaaagctgaacagcatgcttctgtaacattttcgcgttttaaaacaaaggggcaagcgtacctcaaagcc

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		aactgggctgttttccttcccttaagccaaaccaaggaacttgttgctgtgcctatcgagggggagtacgactacaatctctatttacacggctacttctttgttgatgctgggcgt
		aaggggttgcatggccacgacaatcttgggttttctacctccctagagcatgtaaaaaatgatgagaaaaagctgcgtgaggtttggaacatcattctagccagtgaggggac
		attcaacctcgttttaccggctctaaatgagttttgtcagaagttaaggctgccacatcaaataaaaactgttttgaccaaggctttgtacgatctcctcatagaaagatatagaaa
		agaagtatccaagagcgccaattggataatcaatatcgatgacaaggggctgcttggtctttacttgataagaatgcccaatgcttaccgatccctcgtccagagaatagtga
		ttactctcgaatttggtcaacgttgcctggtttgagtaagttactggataaaaagtcactgtatgaagccacgggtaatgaatttttaaccgagcagaatcaacgtgatagttgga
		atattacgctcctggaagaagcgttaggaagtggtgttgtcaacgcattttacagatcaatcaatattgaatatctgcttcagttccttcaactagctaaggagcagtgcacgacg
		gaagattttgataacctgattattccacagttccgagaggtattgtctactcataagcttgctgaactttcattgaacaaggctcttaacacgcaagtttttgagcttgttagcgcac
		ctaaaaccgtcgtactaccaattgataaagatgatcaatctatttgggaacttgtctgcaagatcattcctgcaaagctactgctccctaaatttctgtctactcacaataagccaat
		tcatgacaatgtcactgaagaagagctcttcgcacttttaaccctagtagatagctacatcaaaaaacagggtgaacgtttatcctctgatgaatcgtctgcctgtgagcgtctca
		ttacatttgttattgattgtgtaaatgcaagtgaggtaatccaaaaaagcgatttttatcagaagagtgggcatttaaagcttctaaaagtggaagctcttggttcgcaacagagca
		caaaatatcgctccttaaacgaactcatagtgttaaaagaaaaataccagctgtttcttcgtggaggggagcggaactttggtaaagggttggggaaagagctagttgcagtc
		gtgcctggcttggagctttgttttataagcaaggattttgaaattggtggcctatatgaagggcttaccgcttgttctgaagccgcgtgcctacgactgctttccacgtacccaaat
		cttggttcaaattcggcaagactagcgctcactaaagtattctctgccgagctctctacagatgaggagaaaagaggtttccggtatttgattcacggcagcaaagaagacga
		cttgagacaaacgctttggaagccaaacagggcaactaacccagtatggatgaaaatttggcgtatgtgtcagccagaagatttccctggatggtgtgagttagatgaagagt
		tttctaatgctttgacaaaccagtacgaacattttattggcgttaaagagcagttctataaagacattatctctgaatacagaacaatactgcctgaatgcaattttgataactttgat
		gactgggaagtggagcaactgctcgcagatattggtagtcaaggagatgaaaggctatggaaagcgttgcctgtccataggacagctcataacactagagtcgcgattacg
		accaaatgcctgatggaaggaagtgcaacagttccaagtgaatgggatgttcaccttattcaacattcagccattgctgaagtcgccgcttgccagcataaatgggtgaatcat
		ggtctacctaaagagctgatcgagattgcgcttacccaatcaagtccagctcagtattccgcatttattttggaccagctctgcgctattcgtattgcgaatgaaggaattgagca
		tgagttggaaggcaagataaataataccaagtggctgcgattagcgtcaggaaccgaggtttcaccggaagctattttatctttctctgccaatgagctgcctgagtctgcaaa
		gttctgcgagttaaaagagtcaaacatttacatgttctctcaactcgatggaaacatgtttgagcacgatcaagcacgtggtttcttgagagagtgggtcgcaaaaagtaacag
		ctcagtttgctcgtgcattttggcagaagccgcgcaacatcaaagttatgtagttggtaatttttccaacatttctgctcaggtgctagaacagatttcatgcatcccgccattgatg
		cagctatctgcaggctggggcttactggttgagctctaccaaagccaatatctttcagtgaatgaaaacaagcaagtgatgctatgtaaggaaacagaaccacaatcattatg
		gtgggcgctggagcgtattgctgatgatgatattcacggtcagtcaaaggaacttcggaaagcatttttagaagcgttgtgtaacaccgagggaggcgttgattatcttcctaa
		actgagatttcgcaatgagaacggaagttatgtatcgggcaacacactggtatcgaatgttgctcaggtagttgctgataacttaatttcgccacaagaatacgcagtcattgag
		agttattgcagtaaatctgctctcacgaatggtaatacgtcaaaaatcattgagttagcgggcgataatgcgccagtacttagtgattacttcgatgactgggaagggatggttc
		cccctgatgccatagcgacatttatagcactgtttgctaaatctggtggcgtcgagaaattggttaacaattatctaagacagtcaacgctggagtcgataaagcaggggtatg
		aggaaaagtggaactccggaaagggacgtagaggcgaattttcacactatccgtatagctcgttatataaaagtgttgattttgaactggcaatttgtgcagaaaatgcggcgt
		acatgacgtcgattttcggcgaaagaattcaagttaaattacaaaaaacaccagattcattgcttgttcaccaagcgaacaagtccaagacgaaaaggatagagcttcgccga
		gttgatacaaagaatgtatcaaaagaccaacttctccgcatgcttgccaaagctgtagaaacgatttttactgatgtgtttggtgcagagtgtattcgatttgaaagtgaatttttga
		agaggtttggtgcttcagaacaggtagatattcagattacccgacagatagtcttggagaatgttgtcccccctacttgaaaggcttcaagtgcgagaagaaggactttgtgattt
		acgttcagattacaaacgtgaacagcgtgttttggcgagcagtgatccttctgtactacaagatcgctcacgccttaacagcgtccttacgaagattaaagagactcttgaaaat
		aacgaaaaagtgcaatctttggtactcgaatctgtacgaaaagagatgagtaaacatttccaatactcgcctttcagcgtgccatttgagctgtttcaaaatgccgatgatgcttt
		gtgtgaacttattgaaatgcagggcgactcaaccaatgtactgactcgatttgatgtggtttctggcagtgatgggactcttaacttctaccattgggggagagaggttaactact
		gtaaaagttcatatgtcgcaggcaaaaaccaatttgaccgcgacttagaaaagatggtgagtctcaacgtttcggataagtcagatggaaaaacaggcaagtttggactggg
		ctttaaaagttcattgcttcttaccgacattccacgtttggtgagtggtgatatttgtgcagaaattcatgctggcgtattaccgagtgttcctagcaaaccagtgatgacggaactt
		aatcaaaatgtcgatgagtataaaattggaaatcgtaaaccgacattaatccagttgcctaaatgtgataagaagcgggcagatttgaagttggttttgggacgtttcaaaagta
		acgctggcattctcacggtttttcacgacaaattcgagaaatcaatattgatgagcagcgatttgggtggtcgggacaggctctccataatatccctgaagtacttgtcggtga
		agtgaaactgccaacaaatacttctgaagagtctaacgttatccttcgaagtaatagagtgcttattatcaataccgagtccggtcagttcctttttgctttggattctaacggagtt
		gtttctctttcgaatcgaaaaaacctaagtagcttttgggtgttaaacccgattgacgaagatctgaaattgggtttctgcatcaacgcgccatttgcggttgatattggtcgctctc
		agcttgctgtagataacggagacaatatcgatctttccagttcactcggcaaagcgttatcagctgtgttggtcaaaatgtttgcagcttcttcgaataattggaatgaatttgctg
		aagaggttggcctgggacaaagcagcacatttatcaagttttgggcgtcactttgggatgtaataacagcccattggccagcaaggcttggagagacgaactctaaagctga
		actgattaaacaaatgttcacagtggaagatggtctgcttgcgttttaccagagatgtgcggctcttcctcgaaatcttggtgtaaaggaagattctcttgttcaacttaaaaacgtt
		gatactggagcgaataaacctttgaccaaggcatttaataccttgggaaatcacccgatacttcaacggctatataaagaccaacaactcgtcgggcatgacacctttgagtttt
		tgaagagtatcgattttagaccgaataatggtgcgttaactaagctcgaattgatcgatttgattggacaggactttcctcacaatgaagtaaaccacgacagagcaagtttctat
		ggtcgcctatttggtaaaaactttgaaaagttaatgtcgaattttgaaatgacagtgactgagaaaaaggtgttggaagagcgtttttctgaattgaagtttctcaacaaaaccggt
		gtatacgtgactgcaagcaaactgattgttgaggggagccctgagagagacttgctatccaagtttgcaccagacagcgcgaagttaagtgaaaaatatgaccaagcatcaa
		tggacttggttagcttcattcgtcgtgacgtaagctatgacattcattcatgggctaagcaaataagatctgaagaatctaacaggggaggaaagcaggaagggttgtgtagct
		tccttgttgaaggcggctatttagcatcatcgcttctcagaaaactacagacggatcaccccgcgtttcttacaaagggacgttttgatccgagcgtattaacagaaaaatggcg

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		ttggagttcttcaaaggcttcggctttcattagcatttggattgatacagaggaagataaagcaaggcacgtacgacaagcgcaaaaagagtttattccgaatgtgaccaatggtgagcagatcctcgaaaacatcacgaactggtggaatcaatgtcgtaatcaaagcttaattgattatgacaaacagctctatgctcaaccaatgccttggaaggcaatgacagaggacttcgagcttgaaacgttagaggttcgtaaaggttggttgaagttgttctatttagggagttgccaaacattaggtttcaataacgatgtagctaatcggaatgttgtttcttggttcgaggacaaggggtggtgggataaactagccgttgccaatggtcctagccctgaagtatggaaagaattaatggaagaatatcttcaaacagcacgcgttgatgagcgttatagagtttggattcaagttcttcctttgtatcgctttgctactaagctcaaggactatgtcgctctcttcatgaacgcttcctttattgataatcttgatgatttgttaaaaccaaatagttcaaacaagttatcaggctctggcatccaagtatctgagttaaaaggaacgctcggtattgggattaatttcattttacgagagttgcaaaggcaccaagttttggagcgtgagtattgtgaagatatccaaaagtacgcatttgttttgcctgctcgattacgaaagttactcaaaaaaatgggagcaggtttaagctttgacgcagagccagagaattcagagcgagcttacgactatttcgtttcggcattaaatagtgaaacccaccctcttcttaaggactttgacatcccatttagagtcttgttggctgataagcaagcgtttgaacgttgttttaattttgctctagatgagcagtttgaggaagtatatggataacattatacgcgttattcacccaaaattcggtgtcggtaccgtcgaattcgaaaaagctgagacatctcttgtccgatttgaacatggttttgaggagtgtttgaaaagtgagcttgaggcggtcgctgatcttaagtccgatcttgtttctggacagagtgtcgctgcctctgaacttgcgttaaaaacattagcgcactcactaaaaagtgttaatgaaaattggagtgtttttctaaatcgaacattaatttacttcctcatcagttatgggtatgccatcgagttctaaggcaatggccaacaaatcaactgattgctgatgatgttggtttaggtaaaacgatagaggcgggcttgattttatggccccttatcgagaggaaaagagtcaagcgtcttctgattttgacgccagcacctttggttgagcagtggcaccaaagaatgcttgatatgtttgatattcgtttgagtatgtatgcaccagaaaatgatacctcgcgcgtcaattactgggactcaaacaatatggttgtcgcttctctacctacgctaaggaacgacaagaatgggcgtttagagcggatgttaaatgctgagccgtgggatatgctcattgttgatgaggcgcaccatctaaattcaacggaagataagggtggaacgttaggctttcgctttatacagacgttgattgaaaatgataagtttgaatcgaagttatttttacagcgacgccgcatcgaggaaaagaacacggattcttctccttattgcagttgctgagaccggatttgttcaacgttaagcaaatggatgagcgagaaatgcgcccatttgtgaaagatgtgttgattcgaaacaataaacaatttgttacggatatgaatggtgagaggttatttaaacctctgtctgtgtcctcaagaacttacagttacagtgaacaagagcaacatttctatgacctcttaaccaagtttattgtatcgggtcaagcgtatgcatcctctttgaattcaagggatcaaagagcggttatgttggttcttaccgcaatgcagaagctcgcttctagttcaattgcagctatcgagagagctctaaaggacggatagagaaacataaactaggtaagcaacgtcttcaggatattgaagttcaacaggctgctttattagaaaagcgtgaggagtcagaatcgcagtctgaaagcgagatatacagtgatgaattagcgcaattagaactggaatttattgaaacgacaacgcgggttcaattgatggatgatgagctccctagaattatggagttgttgtctgcttgtcagaaagttggctctgaaacaagaattttaacaatattagatatcctagaaacggagttcaaagatagaactgtcgtctttttactgagtataaagctacgcaagcgctattaatgggtgctttgaataaaaagtatggtgaaggctgcgttacttttattaatggtgaaaatcgtcttctgaatgtagagaatggctcaggagtatgtgttgattatgtcaccgatagatacaatgccgcgaagcgttttaatgaaggcaaagtacgatttataatttctacagaggctggtggtgaagggattgatttacaacaaaattgtttttcaatgattcatgtcgacttgccttggaacccgatgcgacttcatcaacgtgtggggaggttgaatcgatatgggcaagtcaaaaacgtagaagtaatcactcttcgaaatcctgataccgtcgagtcaagaatctgggatttgctgaatacgaagatcgatttaatcatgcgttcggttggcggtgcgatggatgagccagaaaacctaatggagttgatattaggtatggcggatagcacattgtttaatgagttgtttacagaagcagccaatcgtaaaaactctgaatctctctctgcttggtttgaccataaaacaaaaacattcggtggcgagtctgtagtgcaaaaagtgaaagacttgattggtagagcagaaaaatttgactatcaagatcttgaggctgtaccgcgtttagatcttggagatttaaaaccgtttttactcagatgctttcatttaatcaaagacgttgtaagtatgatgaaaatggtggtttatcgtttttgacacctcacgcatggttggggcaatttggaaccagacgctcgtatgagaaattgcattttgaccgcaaagctaaacagcttgattcagaagctgacatcataggctttgggcatcccatgttttcaaaagcggttaatcaaggagagcaaatccctggaagttacgcgtttcttaacggtatagagaaagatcttgtagtgtttaaggttcaagatcaggttacgggaaccgatgcatcagtaaaagtgagtattgttggactggtgctcgatgataatggcgattgtgaattggtcaaggacgaagaccttatcgggtatttaaacgagtatcttaaaatttccaatgatgttgactctaaacgtacaccagaggatttagtgtctgttattcaaactgctaatgattatctaatggagaatgtgtcatcaattggcttaccatttaggctgcctaattctgaaccattaacggtattctacaaagcaagtaactaa (SEQ ID NO: 400)
38	ATPase_GHKL-DUF3684-DUF3883	gtcatagtcccttacggagataattcattgaaattaatatcttatacagcacatgtaaatagccgtggtgtatttttatccaatgaatcgttacaaaaataagatgcatgcccaccctgttctgtgtgaacgctacgaccagctacggatttataccaaaagtaggaattctatatgtcacgtattaccatcaacgttttatggttaaccgtaccaatagcgcggaagtgggcatgagcgaagtagcagatcaacagcaattggaaactcagccagcgggtgatgacctcctgcaaggtgtcaaacgcgttctcaggcatgccgttcaggcgtacggggatgggttaaaggtttatcaaagcctgcaaaatctcaacgaggtgattggcacggagtacggtaatcgggtcatttatgagttgattcaaaatgcgcatgatgcgcatacgtccgaagaacgtgggcggatagctgtcagcctggtgcttgaaaacctttcacggggaacgctctacatcgctaatgatgggcgagggtttcgccatcaggatgttgaagcggtcaaaaacctggcgatcagctccaaagagattggcgaaggtattggcaataaggggcttggatttcgcagtatcgaggcgctgacgcaatccgtgaggatctattctcgctcaaatacgaacggcaaggaccgatttgagggttactgtttccgtttcgcagatactgacgaaatcgcgcataatattcgcgatctcggtgttgatgacgcgatcagcaacgaagttgccaaaacgcttccccgctatcttgtgcctgttcctctagatgatcaaccggaggatgtccgcacttttgcccgcaacggtttctccaccgttatcgtggcaccgttagaaactgaagcggcagttacgcttgccagaacgcaggtgaaggagctgaccaatcgcgatgttccactgatgcttttcctcgatcgtattaccgaaatcagtatcgaaattttatccccggatgagaaagccgaaaagcgcaccatgcaacggcaggaaaaggcgctgggaagtattcctgacgcgcctgatgtcagtctctacgaagtcgatataggtcagcggaaacgctttttagtggccagaagcaatgtcgataaagcgcgcgtgcagcaagcggtgagcgatagcttattgactgcacctcagctaaagcgttggctgaactggcaagggataccggttgtttctgtcgccgttggcctgaacaaatcaacagtaacttctggaagactctacaactttttgccaatgggcactgaggccgcttcaccgatttgcggctatatcgatgcaccattttttaccgatattgacaggcgtaacacgaacatgagtttgcagctgaaccggctgttaatggaagtggctgcggaaacctgtgccgctgctgctttgtccgtcgtatcccgtgagctggatataggtgcatctgcggtttttgatctgtttgcctggacggggggaacatcgtcgcatgatgcaaacagcactggaacggaaagatacttcgctcagcaaagcccgcctgattccggtgatggctccgccaggaaaacagcaatggtcgagtcttgaagaagtcagtatctggccggaggtgaaatttgccatcctgaagccgaaagacgttgccagatacag

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		tggcgcgcagttggtttctagcgaattgaatacgccgcgcatagtgcgtttgagggagataacaaaatttccctatatgtatcagtcattagatccttcggcgcagacactggtgaaatgggcagaagcctttgccctttcgctggtggaacggaaattctcccctgccagttggaccaaattctatgatgatttggtcaccttgtttgctgcggtaaaagtgaaactcaacacacttgagaactgcctgatcctgtatgaccgccagggcaaactccggcccgcaggcgggcataacagtaatgaacacaatggcgtttttgtacgtcggcatgtatccagaggcgacaaaaagaaagataagcgtaccgggattccgttgccgccagcgattgtttctcggcgctaccggtttctggatgaaaaaatcgtgcttagtgcggcgacgttcaatgcgtttaccgtcgccgacctgataagagagtacgatccgatcaaagccctgtcagggctgaatacggccctgagtaataaggcgacagtcagacagcgccaggatgcactattgtgggcatttgaggtctggcgcagcagtagtgtcgttgtcgatgtggagctgaaaaaagccgatctccatattcccgtgcagtcgggttggtgtgcggcaagcaaggctatgttttcatcctcctggacgccaacagggaaggttgtggaaagctatttaaccggcgcgatggggatctcgcctgactgccgtctggcagcgggtttgttattgattgagctgcaagactggccgggcgtcgtgcaaaacagcaaaaccgactggattaaattcctccgcgtgcttggcgttgcagatggattacagccggttgaatctaaggtaagagcgcgagcatatggcgatagttggaatagcttttacgcaatggcgacgagcatgaggggtttgatagcgactggagggcagaagtaaagcgggcacatataagtttctaccatcctcagacggtctatacctcggaaggaaaaacatggcgattgcccgggcaacttgagcacgcaacattgccagacgatctgagggagctgttgtgtacgctgattttcgcctttctgaagtcgcagactacggagtttttacctttgaggtcggtcgttttgagcgacagaattcgcaaacagactcccgtacgctgccaacgccgcttggcactttttacgcactaaagcttggcttgccagcactagctcactatctgaaggattgcattttagccgtccagatgcgtgctgggcttcgcgggagcggcgcaataaacctccgcgtttcctagaccatttgattgagcacaacgttgatattattgaagagagtcaactagcggagcgcttgttttctgcgaaaattggcctacgtgattggaatcataccgggacggcgttggatcgcattaaagaactggtctacattgttccgcagttgaacgctggcgataaggcggatttacagcgggaatatcaacgaagctggcgtgatatcctcgacagcgacgaagctcttcccgacggattggacctgattgtttttcgccgtgggcagcatgaagtgctgcgcggcaacagcgatctgcctcctgcggtgattgtcaccagtattgcacaaaaaattgaagcacaaatgcttgcttctgcaggctacgcaatactcggtattggcctggatgagaccgatacactcgtctcctgcctcggtgatacgggacgattttcaccccgtaagattaatgacggcggagtgcaactttacctcgatggtaagccgttttatcccgatgagagcgatccgttgcttatctccttcgacatgaactggttaccggaaatcctggttattggtctggcgttactcggggaaaacttagagcggggcgttcacgccaccaaggttgataagcagctgcgcgcaatcagggtacgccgttgtaagaccctctcttttgccgtgcagggcgatgatgccacccaacggagtcgttcgtcagctattcctggccccatgaaacgatgccgacgctgattattgaagaggggctggtgtttaactggcagaccttagcgaagatttcccgcaacctctcacggctggtggataaccggttacgtttcattgaaaccttacttttgcgcctcgcagttggtcgcgataatggctcgttgagtaaaccggatgacgttaccctggcttgggagatgaattgcgatgttcaaacgatccgtgatcattacgcccgactgcgcacggacatcactcatgtgatagacatgctacttcctgtggtgacgtatctcaacggtattgagcttgctcaggttctcaagcgggaatatgccttatctaggtcagtatttgatgtgcgtagttggatttcatcacatctatctgatagtgatatacctgctgaaaagctgctggacgtgtgtgaaacagcaaccgatcgggttgaactccgtaaaatgctgtcgtttgattttcagcaatttaacctggctctggaagcgttaggggaaacaccgctgtccaatgaggatgctctgcgcagatttttttacggcctttgtcgggcagaggcgttcacatattatcgatcggttacgccgacactatctggcgacctttgataccggcggagatttgtcacaatacgttcagcataaatctttgggcttcatttccttcaactctgaatggattttgacacatgaaaccttggaaaaggagatggtggactcgcaggttgacacgcaacttttgagtgcgttaggaccggacaatggtgaagagctgtctgcacttaatacgttattagacgcgaatcgtaaaaatgtgcgcgaatttgccatgcaggctcagccgcgagtttccgcctggtgcagacaaaatgatgtcccggtgaatgctcactggcagtacaacgatcctcaggcgttttgccgacagctcgaaaataagggctttcttgatttccggctctttgagccggattcactaccggattactgcctgcgcgccgggctatggccaccaacgatgccgcccagcctagatcaggatgtgctgaatatcgacatgaggaaagtttcccaggaaaaagaacgcgctgagcaggcaaaacggcaacaggaacttgagcgtcgcagtatcttttttttccgggcagtcgcttgatacagccagcccgctatttgccgatcaacttcgggaactggcgagtaccgatagtagttggcaggtgcgcagccagcacaagacgcaggccttgatggattttggcgtggtgacaatgcgtcaggcgagcggcggaggttgcggaaaaagaaccgggcgtgcgtatcgggagcctcgattgacacctgcacagcagcaagccatggggctggcgagcgagtggctggcttttcagtatctgcgcgatcgctttccggattatacggatgaaacttgctgggtatctggtaatcgggcttcgttttgcgggggcgaggaaggagatgattcggccgggtatgatttcatagtgaagacgccgaaagtggaatggcttttcgaagtcaaatccaccctcgaagatggtcaggagtttgaactgactgccaatgaacttcgtgtggcaagtgcggcggctaaagacgcaagccgacgttaccgaatcctctacgtcccttatgtgctttcgccggatagatggtgcgttatcgaattaccaaacccgatgggcgataaaacacgcaatcacttcagcgttgtggggcatggatctttgcgtttgcgttttcagcggcaggagaactgacagcaaccctgctcagggaaacctgagcggggtttttaaatatggcctctatggatagggggacactttctgcagtaaatggataataagaaagctaacgttgaagtctgattctgccattttccacgacagctaaatgctggatcttctttttaggatcccaacatacctagcagtaggacgtaagtatgcttgagttcatctcgatatccttgtttctgaatgacaggcattactatttcgtgggtgtgaaccgatgaaggggggtgatgtcattggaaaataatgaggtagtagcaaggagaagttctgctcttatcatagtgaaaaagcggtttgggaacaaatcggaactgata (SEQ ID NO: 401)
39	TerY-P + helicase + HEPN + ATPase + DUF2357	accttcttcgctaactgatggctaatgaggccgtaataaaacttaccttacctgtaaatacttttactactcattcagatcagaatgaagaggtttattttatttcattgaaaattaataaataaaaatattggcacggtatgtgcttatacagaatgccattttactaacaaggaatttaccgatgtcggaattaaaaaaatttcaggtacaaacagcacgtgcattgccggtgattgtgttggcggataccagtgggagtatgtcaacagatggcaagattgatgcacttaatctggggctcagggaaatgcttgatagttttaaacaagagagccgcctgcgcgctgaaattcaggtcagcgttattacgtttggtggtcaccaggctgaagttagcttgccattgacgcctgctcaccagttgcaaagtattacctccctggaggcaaatggcatgactccactgggtggcgcactatcgctggcctgcgagattattgaaaatccaacgcgaaaatttcagccgattatcgtgcttatctccgatggctaccctaacgacgactgggaagccccttttgctcgcctgattcacggtgaacttactgccaaggcctcccgttttgccatggctatcggtgcagatgccgatgaatcaatgctcaacgaatttgcaaatgatcctgaggctcctctcttccacgcagaaaacgcgcgtgacattcgccgtttttcagagcggtaagcatgagcgtcagcgcacgaagccgttccgcaaccccgaatcagtctacaccgttgcagatcccgagtgctgatgatcaggactgggagttctgatgcgcctgtacgcttctggcacctcggtacgtggtcccgcacaccaacaggatgatgaacccaatcaggatgctgtagggatttacggtctgcgtggtggctggtgtattgccgttgctgacgggttgggtagccgatcaaaaagtcatttgggttcccgtaaggcagtcaatctgctgcggcagatc

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		atgcgcggtgcggagatgctggtcgctgccgaagtgactccagcgttacgtgaagcttggctaaaccactttggtactgactatcacgattacgaaactacctgtttgtgggc ctgtgtcgaggcgtcgggccatggcgtgatcggacaggtaggcgatggcctgctgctggtcagaagtgctggggtgttcaacgtaatgagcacaccacgacggggttaca gcaatcacactgagactctggcacagcgtgcacatttagatagttgcagtgccagagtggcattaacccaacccggagatggcgtactgatgatgaccgacggtatcgctg atgaccttatcccggatcagctggagtcattctttaatgctatctaccaacggatacggcaatgcagcaagcgtcgtacacgtcgctggttaacacaggaacttaacggctggt cgactccaaatcatggtgacgacaagagcctcgctggaattttcaggatggactgaccacatgacatcaatagtaaaaacgcaaccaaaacgcgtggtgaaggataccag gggatcaagttacgagctgacagaggtaattaaccgtggtggacaaggcattgtttaccggacgacctatccgcaaaccctggtgaaaggttttactaatcaggacccacag gaacgccagcgctggcgcaaccatattacatggctgctcagccaggatcttagcgacctcaaacttgcacgtccattaatacttctggcggagcctcgctttggttacgtaatg gagctgatggatggcctggttccattggatagcctgttgaacagctttataaacgcaggggaggagtctctggcggattatctgcgtcagggaggactccgtcggcggattc gtatcctttgccagctggcacgcacactcaatcagcttcacgcacgcggcatgttgtatggtgatctctcccccagcaatatttttgtttcagacgatccaagacacgcggaga cctggcttatcgactgcgataacatcagcctgacagcccatcacaatctgactctgcataccgtggactatggtgctcccgaagtggtcaggggagaatcgttactgtccagc ctgaccgatgtatggagcttcgccgtcattgcctggcaactgctgactcataaccatccgtttaaaggggaactggtcagtaatggtcctcctgagatggaagaagctgccat gcgcggtgaatacccgtggatcaatgacgcacaggatgacgcgaatcactgcttcgtcaatctgccaccggagctgattgcacatagtgcactgccaactctcttcgctcgc tgctttgaacagggaaggtttgaacctcatgagcgtccgggtatggctgaatggcttgaggcgctgagtgctgtggatgagcgtctgtttacctgtgacagctgtgggggaa gcacgctcctggcagaggaagcagaaagcgcgaacgatgccgtttgcttttactgtgacagtcccgccgaccgcctcctggtccggtttagtgaatatgtgactgagcaaca agacggctcgaatccagacaccaaaaccttgattgccacagggcgaaatgtatggctgcagccaggtcaccgtgttgagttaaagcgcctgttgccaagttttatctatgacc actggccatcagatcatctgcagattgattacaccgcccgcgggattgggatccatccgttgcttggcggagagctatacctacaacgcggtgaaactatcaaaccactgcg ggggtttcagggactcaaaaacgagctgcgcggaacaggtggggagccttggcagatccatatcggcgatcctggccagtcgcatgtaatctggcagttcacgtggtgac aatatatgaaaattaacgaatttccactgatgtccaaagatattctgctgctggaaacggataaaggaaccaccgggttccggccaaagcaagctatcacctttcaggcgtatg gtgagaattggctggcggtacagggggatcattgcgtaagtgtccagtgctcccctggtgatcacgaactctttagccgtctggtgatgagggatcaggttcgttggttgctga ccagtaaagcggaaaaacagttgcgggttcaatattgcacgcctgttgaagtcacaccaatgcagctcgagttgggaattgatgagcgaattgcggaagaccttttcgcgaa aaaacagatcaataacaacgatattgagcttgcctgccgctggtttgaagagacttttattgtccatagcgagtcagaaagtgactggttaacggttggccgttttagcaatcat gcagccaaaggtggttttcagctattgggaaacggctggcgtgcggatgttgagcgcaacccggaccacggctttcttatcagacgtattactggtcatttaagccatgatac aggcttctcgttgctggttggacacttcgccttccgggatatgtcagttgctgcggtgctgaatagtgcaacccagcaggcaatgctcgatgccgcactgcgagacagtgcc agctaccttgagctctggaatctctacaacgataaagagtggcagagcgagttgaaaaaggccgaaacgctgggtgttctgcgctttgttgcgtgcgagggcaccgaagct ggccgggaaaatgtctggcatctgactccccgaactcctgaagaatacagagaatttcgccagcgctggcgcgcgctcgatctgcccgcaggcactcaggttgacctggg cgctgaaactcccgactgggcagaagaactcagtaccgaagaggatacggtactgaaaacgccgcgcgggaagatcgagttcgctgatgaatatgtggtctttacttcagc ctcgaatcgccgagacgtgcgccccgcaaagcctgaaggatggctctacctctcgttggcaggatatcgcacagtcggcaaacgtcgcctggcggcaaaacgtgccattg attccggtaaacgcatgccacagttgaagtggctgctggaaggggtcgttgttcctgctgctcggcgtcgcaacatccaggggatgacaccctacgcccgcgaaatctttaa gggtggcaaaccaacgggcaaccaggaactggctgtgtttaccgctctgaacacacccgacattgctatcgtaattggcccgcccggaacagggaaaacccaggtgatc gctgcgctacagcgacgtctggcggaagaggcccaggaaaagaatattgctgctcaggttttaatcagcagttttcagcatgatgccgtcgataacgcgctggaccgcagt gacgttttcggtctgcctgcatcacgtgtgggcgggcgtcgtgcttcagtagaagacgagtcaccactggatccctggttgtctcgccacgccagtcatctgcaggagaaaa ttgctgaccagtatcaacgctacccggagttgaaaacaattgccgacctcacttcccggcttgccctgcagcgattggcaaacgacctgcctcaacaacgggcagaggcttt ttcgcatatttatcaggacgtcaattccctggcagagaaagggctggtcacggactcccggcttgagatacgtctgcaggactatattaagcatctgaaacaggatggtgttgc tgaggtcagtacggtgatgaatgtagcagtattgcgccgcattcgcgcgttacggaccactcagactgctttctcagatgatggtgccgatcgtgcctgggatttgctgcgatg gttgaagcggaatgttcctgacatcgacgctgagctgacctcggtattggaaatagctgccgatgccagagaagttcctgtggcactcgtcgagtgccagcaacagctgctg gagcgttttctgcccgattatcgacctccggccctcaaaaataagatcgatgatgaaggactggctctactgaatgacctcgacaagcatctttccgacttgatgcatcggcgt aagcagggtgtggcatgggtgcttgaacaaatggccgatacgctggagatggaccgccgtgccgcacaggaggtggtggatgaatacgccatggtggtgggagcgacc tgccagcaggccgccgggcaacagatggccagcctcaagtcggtttcaggagtcaagagcagtgacattgagttcgataccgtagtcgttgacgaggctgcacgcgcca accctcttgacctgtttgtgcctatgtcgatggccacgcggagaattattctggtcggcgacgaccgccagcttccgcatatgctggaaccggatattgaaggccagttacag gaggagcatcagcttacggcactgcaactggctgcctttcgttcaagtctttttgagcgcatgaggctaaagctactggacctgcaaaagaaagataatttacagagggttgtg atgcttgataagcagttccgcatgcatccactgctgggagatttcatcagccagcagttttatgaaaaagaagggctggggagagtggaaccaggccgtagcgcagagga atttgtctttgacgaaggtttcctgagagcgctggggccactggcgtcggcctatcgtgacaaggtctgccagtggatcgacctgcccgcttctgctgggctggcagaaaaat caggaaccagccgtatccgcaccattgaagcggagcgtattgctcaagaggtggcacagttactgaaagccggaggagaaaccctctctgttggggtaattactttctatgc cgcacaacgagaactgattatggaaaagttatccgaaatcaggctggaaggcgtgccactgatggaaaaacgtaacggaacctatgaaccgcatgaaaactttcgctgggt gcgcaagtaccgtgctgacggttcgttcagccaggaagagcggttacgagtaggttcggtggatgccttccagggtaaagagttcgatgttgtactgctatcctgcgtgcgc acctggcgtcagccgaggtcctcatctgccgccgatgatgcagctgccagggaacaaatgcttaatgaactgttcggtttcctgcgtctgcctaaccgcatgaacgtcgccat gagccgacaacgacagatgctgctttgcttcggcgatgcagcactggccaccgctcccgaagccctggaagccgcgccagcactggcagcatttcataccttatgcggag gcgttcatggcactcttcgctgaaacaggtatttatattcaatctgccccacggccgcagggtgaagcgcgcccgatactctggccagtcaggatacatagggtgctctaccc

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		ggaaagctatcaggctcagatcaatgtcttccaacgcgcaattctcggattggtacgagcgcgcgtcgtacgtccgaccgaactggcagaactgaccggtctgcaccctaa
		acttattacgcttatcctggcacaaagcgtcagtaatggctggcttgagtccggtgaagataccctcacttcagcgggtcagcggttgctggatgatgaggatgacggtattg
		gcaaacaaaaatcaggctatgtattgcaggatgctgtaagcggaaagttctggccgcgtctggtcagcacattgaagcaaatcgaaccggtcaatcctctggataaatatcc
		gcaatttatactgaccaggaaaacaggagcgacactgcgacctttcctgatgaatgccagccgatcgccactgccgcctctggaacgcaaagaactgaagcgtgcctggc
		gtgactatcgtgacgactatcgtgccagtcagcaactgggcgtcagccgtttgccgccacacattaacctgcacggtctgcagcagctagaggaaccaccgcagtgcgca
		cgaatactggtgtggatcaccactgatcgagagagtggacagctatggagtgccgcggacccatttgctctgcgcagtaacgcatggtggctggacctgccttcaatcgtg
		gaaagtgactcccggttgcaaaagatactggaaccgctggttgtggtgccacgcgccgcagaacaaacctaccagcagtggcttgaggctatcgcgcacgaaactgatttt
		aagatgatgagtcaatacccttgggccgaacgtttaccggatgtgaaacgttatttggtggcgctattggtacatagagggaggatcgagcagggtgataacggtcaaagtg
		agctggatgccgcactgaacgagtgccagaagctgctggaggttgttatgcagtggctgattcgtcgtcatccagccaacgcggaattattacccaagggccgcctggata
		aaattaatacggccaacttgctcaaggatatgaaaataccagcatttaccccatcagttattgatggcctatctggccagataatacgtcaggtgcgctacgcatgtagcaacc
		catccggctcattgaaggcactacttttttgcagcggctgtcggtgcgaaccaggatccacagcacccatttggtcactggatgactcagcgttacaactgccaatgctgctgc
		aactggcggatcgtcgcaacaagagtagtcatggacagagtaaatatcttgataagccggtacaggaactcactcagcagatggttgaggaaagtatcagttatgcattgag
		ttttaccgaacgttttaaggaatggatgtaatgtcaaaacgagcacaacagaagtatacctcacctattcccaagcagagaaatggctctgctgcggcatctgccatcaccaca
		cttcagaggtctgcaatgacaaccgagtcgcagattattgccgcagcccatcacacagctcagagtgaaaagcttccaaaagatatcgattttgatgtgacatggctggaacg
		tatcagtcaacgtcttcagcaggaaggagatgatcaatttgtctcctggcttcagacatttactcttttctgccagaaactggcgcaaagggatgaagagacgcaagcagcag
		cacagcgtattcaacagctggagctgacgctggaggagcaaagcgaaaagttagaacaggaccgtgttgaacatgacattcaagctcgggaactggcggaaaagaaag
		ccgggatcgtgagcaaagaacgagagctgaatgaacgtgagctcaacgccaaagcgggcttcagcgagcagaatgcagcatcgctgcgaaacctgacccagaggcag
		cagttactcgaccagcagcatcaggaggatattcaacagctcatcacacaaaagcaggggttaatgcgggaaatatcgcaggccattgtccagttgacccagttacaaatcc
		agcaaagcgacgcggaggcacagcgcagcttgtcactggaccagcgcgaagaagacatcatcaggaaagaggaggatctgaagcgcgccagccgtcgtctggaacg
		agacgagcggtctgtagaggcggagagacaggcgctgaacgaatgtttggctgaagcaatgcaaacagaacgccttgagtttgaaaagaagctggatcagaaagagcgt
		cagttcgacaaagctcaggaacgggtgcaaaacctcagtgaacgcctcatggaatgggaggaacttgatcaggcgctcaatggccaatccgcttcgcaaatgctgaatga
		gctggataagttacgcgatgaaaaccgcgaacttaaaagtcagttcgcgcacactaacctagcagagctggagcgcgagaacaaatctctggccaacagcaaaagcgctc
		ttaaaaatcagctggaaaatctgcttgcagagatggacaagctacaacgcgaggtggatcttcagcgagtggctgcgacccagcttgagacagtggcacgggagaagcg
		gcttcttgagcagcagaaacatctgcttggtcaccagattgatgagattgaagctcgtattggcaagctgaccgatgccagcaaaacccagacgccgttccctgccatgtcac
		aaatggacgagaagaatgggctcaacgcaaaacgtgatcatcgagaggtcggtgacctgaaaaatttgccagtgagcttcagcagcgtattgctcaggcggaagagagc
		gtgcagctattctatccactggaaagtatccagctgctgcttggtggtctggcgatgagccaactgcacctgttccaagggatcagcgggaccggaaaaaccagcctcgcc
		aaggcctttgcaaaagcgatgggggatttgtaccgatatttcggtgcaggctggctggcgtgaccgcgacgatcttctaggccactataatgccttcgagcggcgctatta
		cgagaaagactgccttcaggcactctaccgtgctcaaacaccgtactggcaggacacctgtaatgtcattcttctcgatgagatgaatctttctcgaccggagcagtattttgct
		gagtttctctcggccctggagaagaacagccacgctgatcgaaaaattgcccttaccgaaacagctttactcaatgccccggaacggctcgttgaaggacgccatattctggt
		accaggtaacctgtggtttattggcaccgccaaccatgatgaaaccacaaatgagctggccgacaaaacctacgatcgtgcccatgtgatgacactaccgaagcacgacac
		tcgctttcctgtcagggagatggagaaaaccagctattcgtggcggtcactgcatgaagcctttgctaaagcaaaaacgcaacatgcggaaacggtcaggaacatgctgga
		gcaactgtccggtcatgaatttactcacctgctggaaacagattttggcatcggctggggcaaccgttttgacaagcaggcgatggatttcatcccggtgacgatggcctccg
		gggcagaagctgggcgcgcgctcgatcatctgctggcgacccgtattatgcgctcaggtaaggttaccgggcgctataatattggcttggaatcggtcacacgactcaaag
		aagaacttgaatttttctggattcaggtcggtctgcaaggcgatccggttgaatctatggcattgctggaggcagatatccgccgtctgtcaggtgcgcgctgatgtggcacga
		tcgtttaactggtaggcaacatgcacatcttccgcaacggattgatcacgggcgttactcaatcgaggcttcccctctgacgctaaatggacatacaccgaatttttttcggattg
		ctggtcagcgacggcggagcaaattgtcggctggacgatacgctgcataacttcattcagcctccgcccggccatgaagaggaaacccggctgctggaggaagccatca
		ccacgatcggtgccgcagttgatgatgacatcagtgtgctatcgccgctgatgccagcagctattgtcgataatcaaagccttttgctacattcgaacgtgcactgctggagg
		tgatacaaaaaggacatttacagcatatatcacagcggccgcggctggatttacgttatgacgatgaggtggccgacgttgcccgcgtgcgtcgtctggcaaagggtgcact
		ggtacatctggcgtcacactccgaatgctggcagcgtcagacactcggcggcgtggtacccaagcagatactggcacagtttagcgaagatgatttcaatatctacgagaat
		cgggtttatgcgcgattactggataagatcgaacgtcatttgtatcaccggctgcgcactttgagaagcctgcaatctactcttgcccaagcactggacttctatcaatctcagg
		aggtgaattaccgcctgcgcaatgctatttgtcagttgtgggggatgacttacgatgaggatgcgactgatggcgcatctcggcagctcaacgccacattggcgacgctgga
		gcaaattttccgcatcatttccggtctgcgacaaagcggcctctatctgcgggtaagtcgtactgcgcaagtgacaggtggagttcatatgacgaatattttaagtcacgatcct
		cactatggtcatttgcctttactatgggcacagttggctgacggggctcagcccgaaaatttgcctcaacaacgcctcagagtgaaccagagcctggcagctgcgtatagca
		gctatgccgggttggtgttacgccatgcgttgcagccctggttacacggtaagagtgaaggaagctgggctggtcgcactctgcgacttcgccagcaaggcatggaatggc
		tgctgagctgtgattccaatgacagtgccagtgaagagacgctgttgtctctggtgccatttctgaaccaccagcaggtagcggtagacctaccggaaaatcggtatatcgcc
		tggccttgcgtggggcatttacagcaggcattacctgataaagagggctggattcggctttcacctttagatatgtactgtgtagagcgttttggcttactgatagataaaattctt
		agccgggaattattgcgaaactttgcccgtccggttatccgtattccccggtgcgtattaccacttgctacaaaactgtcttcactgacagttgatcaacagttaaatcagataac
		actgcatggggatctgactaaagctgagctggaacaattaacctctcatttaatcaacaacaatgctagcacacaggcagaggaaattacgctgcgataccgggaatggcg

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		agcattgcaacagtgccctgtctgcgaccatacaaccgaactggtttatcaatatcccggtggatttaaaaccctctgtaaaaactgcaataccgctcgttatttcagccagcat gaaaatgcacactttttttgaacaaaccagaacagtagaaagagaaagtaaaaccttcctggctcaggggcggagagtttttaactttcagttttagcagggtttttacgactcgc tgcatttttaaagagttaagaataatgaaacttcagggcatcttttatatatcggtattacgcaaatcagtagtttcggttgcgcgttttgtatacataccggcaagtgtccaatcaca gtgaatagccaaaatcgccgggagcacgttcggtcagcctgcggacatggtttttatcacgt (SEQ ID NO: 402)
40	Kinase-helicase	ggattcaccattatagtgacatgttcaagatgatgatatatctttgaaaagtgttctctttgcgaacggtatagaatttctagcgttacttttcataattacactttttagggttaggcag gcacaatctatgcgctgtcttagataactacatccatttttactggactaccaccaacaaaaatttagtggtgcaggagaaaacgtgaagtatcagatagtaggtggtgctggcc tgcaccgcagcgaaaccaaaacagttgatatgatggttaagcagttaccagatagttggtttggctatgctggcttagttgttactgatagccaagggtcgatggaaatcgatat gctaattattactgctgaccgtctgctattagtcgagcttaaagagtggaatggtaacatcacatttgaagggggaagtggctgcaaaatggtaagtcacgaggcaaaagtc cctatcagatcaagcgtgagcatgcactgcgactaaaagatttgttgcaggaagagttatctcgtaagctgggttactttttgcatgttgaggctcatgtagtgctgtgtggcaca gctggtcctgaaaacttgccattaagtgagaggcgctatgttcatacccgtgatgaattcttgactataggtaacccaaaaaattacgaaaagctggtgcaacacactaactttttt tcatctttttgaaggggaaagcctcgaccaaattctgatgaggcattacctataattaagtccttctttgaaggaccaaaagtcaggcctttgccactaaaagaaagcggttatc ttgcgaacgataagccattctttagtcaccctcacatggtctacaacgaattcagggctacccacaaagacaatagtcaacacagaggtctgctacggcagtggaactttgat gccttgggtgtagcaaacgcaatgcaaacattgtgggctgagatagctctgcgtgagactcgagtcggtcgcctagttcgtcatggcagcgcaactatgcaggattatatgtt gcgtgctgtaagggaactatccgaggaggatataactgatgatgcccgtgagctgtatgagttacgccgtagttttagccgattagatgagattctagatagcgaagctgacg gatggagtaaatctgagcgtattgatcgcgttcgtgcattattagctccattctcggaattacatagcttgggtatcagtcattgtgatattgacccgcacaatctatggtacgcag gggatcagaagagcattgtcgttactggctttggcgcagcctcactggagggacataatagcctagagtcattgcgtccgacattgcaaagtgctccatatattttgcccgaag atgcttttgaagaagcagttgagccctatcgcctagatgtattcatgttggctgtaattgcttatcgtatttgttttgcaggtgaatcattactgactcctggacagatgcctgaatgg agagctccattaactgatccttttagcggtattctaaatagctggtttgagcaagctcttaaccttgagccaagtaaacgctttccacgtgcggacataatgctcaatgagtttaat gcagctactaaggaacatagccaagaatttgatgaagctaaccagatttatcaagaattaaagcaaaacaaattctttcgcgaagggatgaacagcgttggtgtgttaattgag tttcctccacttcctgaacagttgtctatggtttactctgctcttgctgctattgctacgactggcagcatcagttatcactgtgaacaaggtgggaaagctctgcaggtaaaattgt gggatggtgttattttgacccctcaacaacctggtgttaaccgccgtatccacgcttttaagcaacggatcgataagcttacgcatataaatctgccaactcctaaggtgcagtc ctatggactattaggacaaggcggcttgtatgtagtgagcgagtatgtggatggcctaccgtggtcacagtttattgctgagaacgtgttagtacaatcccaacgttttacaattg cggaaaagttgatcaacaccattcatgcttttcatgaaaagcagttacctcatggagatctttgcccagagaaactgctggtacaagtcggggagcagacagtaattactctga ttggattgcttgaattcagtgatgaattaactgcagataatcgctaccagccagagaatcccgaaagtactgatgcttttgggcgagattgctttgcagtatatcgtatggtggag gagctatttagtgaagatatgccagtactggtgcaggctgagctagaacgcgcaaaacaaaccgttgacggtatacctatcgcgctcgatcctttgctgcagtcaattcgagc accggaacaagctgagattaatcaagttgtggcgtctgagtcacaggataaggtaattcctgtttgctggggcacagatgattggccgcaagaagtgaagcttctagaacaa aatgatgggatctattattttcaatgtaactggtcatctaacccacgctttgcgcatgaattgcgttgttacatcactggcctaggagagcggctattgatagacttagatcctgata atcgcactattaatagaatagtgtatgaaaaaggattatcgatcgaagaaagtataaaggctggtaaatattcccaggctaaaattaatactcaactttcattacaacgtggctca cttaatcagcgtaatacttttattgaactactgtttaacctcgagccagtaattgatgccatcattgagcgagctaatcctaatcaagagatggatgaagatgacttcgatagtagt gagtcaagcccaattgagttatggcaggcattatctgatacagaagtagacctacgagatatagtcaacatcgactctactgactttcaggaatcaccgagtggttgcttactct acccatatactacggaatccggtgctgacctcagctttgaacttgatgataagatcattgtttatattaaagataagcgtgaatcagtgcaattaggggaattgcagctaagtgag actacgccgagtctattggctattcgctttgattttgatgctgctcgtaagcgaattagtagcggcagccagctacaattggaatcgatccgtgacaaatcatcaagagagttgc gtcaaagagcccttcaacgggtaattgaaaacaaagcagagatccagcatctgccacagtattttgattaccaccagaaaccctgcatgcagcaaatgcaaccgcggccat ccgcggagacattacgcgcactttatgatcagcctggacaacgttttaatgaacagcagctaatggcatttcaacagttggtcgagtttggaccagttggagttctgcagggac cacctggaacaggtaaaacaacatttatttcaaaatttattcactatctgtatcaacattgcggtgtgaataacattcttttggtcgggcaatcccatgcctctgttgataatgtagcc atcaaggctcgagagctctgccatacgaaaggaatggaactggatacagtacgtattggtaatgaacttatgattgatgagggtatgctaagtgttgcaactaaagctcttcag cgacagattcagcataaatttcaccgtgaatatgatctgcgagttagctccctaggaaagcgcctagggatggccccattattagtccaacagttatgtcagttacatcgtacgc tgaatcccttgatggtgacatatggccaatatagccgtgagctggataaagtagaacaaataaagagtagtagtattagtcatcaagagcgactggctgaattattagaacaaa gcaatcagcttaaactgcgaacacaagaaattattaactcaatattcgatgacagcttgctgaaaactcttgtctatgatgaaaccttgataagacagttggctgagcaagttgc catacaatacaattataacaatccagagaaccttgaacgttttatgcagctattggaaatgagccaagagtggatggatgtattacgcggcggcgaggctggatttgatcgattt atgttcaaaagtaagcgattggtttgtggaactcttgttggtgttgggaatcgtcgactagaactagctgagtccagctttgattgggtaatagttgatgaggctggccgagcac aagctgctgaattgatggtagcgctgcaatcaggcaagcgggtgctgttggtaggggatcataaacaattgccaccattctatcatcaacagcatcttaagttagcctctaaga aattagaactcgggaaagggatcttttatgagtctgattttgaacgtgcttttaaagcaacaggcggcgtaacactcgatactcaatatcgaatggtagaaccaattggcgagtt agtatcggagtgcttttacgctcaagatatcggtaaactgcattcatcgaggaaagtctcgccagattggtattccaagttaccaatcccttggaacaaaactgttacttggatcg atagttcgagccctaatgaagcaggtgcagaagaacataagggtaatggtcgttactataatcaacgagaagtccggctactgctagaggctttgcagtcattgtcgagtgat ggctgcattgcacagcttgagcaaactattaccacagaacagccatatcctattggtataatcacaatgtatcgtcagcaaaaagaggaaattgacaatgctatcagtcgggct gaatgggctgcatcgttacgtggtttgatcaagatcgataccgttgattcatatcagggccaggaaaacaagataattatcctcagtctggttcgcgataatcccaacaaactac

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		aaggtttcctgcgcgacgcgccgcgaataaacgttgctatttcgcgagctcaagaaaggttattgattctgggagcaaggcgtatgtggtcaaagaccaataatgattcagcacttggaaacgttcatgaatttattagtaaacaggttgcagtagatgaacccaactaccaaatcctgtgtggtcaaagtctgcttggagataacaactaatgtcagaaccacgtctgggtaatctgattaccgttttactacctgcgcgtagttacaagatcaactgcgctttgaccactgaaaaactgatgcctggaattgaacagtttgcatgtcgcttgctgctgatttttgatcaactctatcccagcgagttacagaattactttggtctaactgatcgtgagcgagaggtattgcttgatgggttgctggctaacagactgatcaacattaatcctgatgggcatattgaggctagctcattcctacgtaagcatgcagctaataatggtgggaagccaagtttagttaaatatcaagaatgtacggaggaagttgcattcgatctactaactctttcgatatgtaaaccgcaaccaaatcgtcgttttacttctggactgccagagctattgccgcggcatcagatcgggggagatgctgctgcggtaacagaggcttttagttcccagtttcggcaccatcttttgctcagccgcaacagcgagtatgagcgtcaacggactaaattatataagataatgggctgtagttcgcatgagatggtgcagctcccaatagagatagaggttagctacggtgtttctgctgggagcattgagccgcagaaatttactcgttcctatgaatatttaggtaacacccggctgccgctttcaaacgagctggaagctcatatcgcagattttttgggagaacataaactagatgaattcggtatcgactgtgaagatttctgtaaactagcaaatgataaagtgttgttacaatttgctaatggttataagttcaactattccggctggatagaggctcgtgaacaacgtaaaactggctacggtacttcattgactaccggcatgttaggggctgtttatttgccgcacaattctaagctgttcattagtatgttgcataatgcattacgtgattatataggtaaaacagctccaaaagcgctgtggtatagcagtaaagtaccactgtggggagctaatggtagtcaactttcgcgttttactcgcgctctaggcgatatacttggcaattatgccgatgataagattgctcgcatttcgcttttacactcaagtgcagatgaaggtgaaaaacgtcaagagcgtaagcggcacttaggtcgttttcctaccggtattggccttacttcagaggctaaatttgatcgtttggagatcctcttaattcctgatgtgattgctttggtgcaataccacggtcaacctaattctgatagtgcattaaccctgccgattggttatataactgttgagccagagcgtttagaattacttaaaaaactaatgattaagcgaactgaaggggctgttgcaaccattacttggtctgaatcaaaatttgaaaatttagcttcgctattacctgttgagtttctgattaaactgaataagaaaagcggtgaagatgtggatgctgcaataaaaaaaatgcagatctataaccgtgctgaaaccgcacgggcaattttatcgctacgcaagtagcatttatattgcaacgaataaatttttctaggttgctatgaactagctaaagggcaacaaatagataaacggcgttattcatgtcaaatgagataatgttaaattgatagggatttataccccgccggccattttgaatggtcggagttgttataaacgtta (SEQ ID NO: 403)
41	Helicase-DUF559 + SMC + McrB + DUF2357 + ATPase	ggggcgaaaaggggaatgccggtcattgccggacgagtgcaccttaaaatgtgcggcagggggcgcccgcgggctgatccatttggcagaatggccgtgcatgcgacgatcgagcgcgggagacggctgaccctgatggacaaacgcgctttgagcgagcgggacatctgcactaagttcatcacgtcgtggcttgacagatgtttgccttgaccggtcgaatagccccattcggggccgtgtactttgcaaatggccgaggtgcccgaaaaaccggtctggagccaggacaagaattacagtgcgcgaaccccaccggttactcacagcccgcttattggagttgatcgaaacccatcccgaaggattgcgactcgacgaggttcaggcgcgtacgcgtgttgaagggtgtcgcgcgggagtcgatgatctcgcagcagcgctactcgatctccagcaccaaggtcttgcacatataaacgcagcccggcgctggtttccgaagcgggcggcgagtgtacgaccatcctccgcagtcactggttcggatgacgtggcgggtgcagggctggtgctgcaggcgctaccggcgcgcatcactggcaacgatatggcggtagcaccagcacctgcattgagtgctaccggcacctcgctcaagccgacttggggcctgttacgcagcctgctgccgtattacgccgaggcgctagcccgcaatgaacgggcgttgctactcggaacgcctgagcgctacggcgagcagttcctgctcgtggcaccacgcggccgatggtggccagcagcagggttaggctacgggctagaactctcgcgtacgcatctgccggttgcttttctcaccgcgttagcccgacgcacgcgcgaaccgattcatgtagcctaccccatcgcgctggtgcggccccgcgacgccgcgcgcagcccctttctgttaccagtggcaactgtggcagcggactggaccctcgacgccgagaaactgcgcctgaatctgccggcccaaacgccggcgatcgaatggtcgtgggtgcgcggacagcgccagcgcggacgccagattcgcgagttgctcgatgcacttgatgtcaatgctgacgacgaagtctggcgggcaggctccttcgtcgactgggcgaccttcgtcgatcgtctcgctgcaaccacccctaccgaggtgcgcacaccgctcgatctcgctcagcccaacaatgagttggattgtggccaggcgggcggtatttacggggcgttggggctgttcctgtcgagcgaattgcagttcgcgcgcggggcggtgcgtgatctcaagtccatgacgcagtggtcagatgacgagctggccacaacggcgctggctgcgtgcttcagcgatgccatccacaaggcaccgaatccggtcatcgttccggtgctggagccgcttgtgcttggcgaggatcagcttgcggccgtgcgtgccgggctaaacgatcggctgaccgtggtaaccgggccgcccgggaccggcaagtcacaggtcgccgttgccctgatggctagcgcagcgcttgtcggtcgcagcgtcctgtttgccagccgcaatcatcaggcgatcgacgcagtcgtcgggcggctggccgaagtagttgaagaccggccgctggtaatccgtgccaatgcgcgcgaaagcgatgacagcttcgactttacccgtgcgatcgaagccatcctcgcgcggcccggtggtgagaggcccggcgaagggctggctggctcgatcgaagtgctgacgcggctcgatgcggcacggaccgctgcgatcgaacaggccgccactgctaaccaagcgatcaacgaactcgggcggctggaagcagcgatcggagatctgacggcagcccttggcatcgacgcagccgctccactaccgcgggatctgcccgctgccacacgacccttgcatagttggctagagcgcctgtttgcgccttgggtacggtaccggcgactacaacggctacggcgtctagcgctgggatggggccagcttggttttggcgagtgcgacgaatcgacgctggagctacacgaacaacgtctactcgacctgcaggagctggctgcgctgcgggtcgagcgggatcaggcagaggcagccgtgcgtcaactccgttcaaccggcgatccgatcgcgctcggagagcggctgtgcgcttcatccaaattgcgtctgcaggggctcgccgaactgcttatcgagtgtgcgcctgaagatcgccgtgcgttgaccgcgttgcgcggcgatctggctctggcgcgcggtgatggcgccgccggtgctgcccgtgctcgggaactctggtcggctcagcgagccctgatcctcggccagatgccgctatgggccgtgtcaaacctcggcgcagccagccgcattccgctggtacccgggttgttcgattatgtggtgcttgacgaggcatcgcagtgtgatatcgcttcggctttgccgctgctggcccgggctcggcaggcgatcgtgattggtgatcccgcgcagcttacgcatatctcccaagtgcgccgggagtgggaagccgaaaccctgcgcaatgccggcttgatgaggcctggcatcggcagctatttgttctcgaccaacagtttgttccatcttgctgctgctgccgccggcgaccatcacctgctgcgcgatcacttccgctgccatgaagatattgccgactacattagtgccacattctacggcaatcgcctgcggccattgaccgacccgcgtagcctgcgggcaccagtcggacaggcagccggttttcactggacgaccgcgcccggtccgatccaaccagcccgcaccggctgctttgcaccagccgagatcgaagccatcgtgcacgaattgcattggttgctgggtgagggcggcttcactggaagcattggcgtagtcacatcgtttcgcgaacaggccaaccgtctacgcgaccgcatcgagcattgtttgagtgccgaggcgattgcaagcgcacgattggaggttcacaccgctcacggcttccagggcgatgcgcgcgatgtgattctactcagtttatgtatcggtccggatatgccggctggggcgcgagccttcctgcacgacacgggaaatctcgttaatgttgcggtgagccgtgcccgcgccgtttgccatatcttcggcaacctggagtatggagctcactgcggtatccggtatgtcgaggcactgctggcacggcgccatcgaacaggcgatg

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		ccactgccagtttcgaatcccccctgggaagaaaagctctggcgcgccttggctgagcgcggtatcgagacaacaccacaatacccgattgccggtcgccggcttgatctgg
		cattgctgaccgacagtgtgcgtctcgatattgaggtcgatggcgaccgttttcatcgcgacctcgacggtcggcgcaaggtgggtgatctatggcgagatcatcaattgcag
		gcgctcggctggcgggtcgtgcgcttctgggtttacgaactgcgggagaacatggatggttgcgtcgaacgcatccttgtccacatccgaagcaccgattactgagcatcac
		cgttccccaccagcagcagccgtgccaccagcgaattggcggcgaatgcaactcgtgctcgggctggccggggctctggcgctggctagcctcgtcactgtattggtggg
		tgtaatcggcgacgccaccgaacgcgagagttggcgagtacggcgtagcgagcatcaggaggtgctgggcgcgctcagcaccgcacgtgcccagcttgatgaggaagt
		cgccaacctacgccgtaatcgtgctgcgctcgatgcagacctgaatcgtctccggaccagcgccgaagctgagcagggcggcgcagcacggctgcgtgaggaagtcgc
		cgcactacgccaggagctcgccgccggccgcgccgagttggctgtggctacgcagcggcgcgacaccctgcaggctgcagtgaagacggccgatacgacgctggcg
		gaactgaacgcgcgccgcgatgaggccgagcgtcagaccggtgaggcagcagaacgccggcgggtcgcggccgaagccgagcgggccgcgaaggcccagcaga
		gcaaggccgaacaagcccgcgacagtgcggttgcacagcagaaggaggctgagcggcgcatcgagcagatccttcaggacctgaaaaccgccgaagaacgagtagg
		tggactgcgcacgcaagaggctcaactaaaagcggctacaactgcctccactgccgaacgtgaccggctggatgctgaagccaagcggctcggactggagcttgtcaag
		ctcgatcagcagcgccagcagcttgagcgcgatacccgtactaccgccgaaactcgacggacggccgaggggctccagcagcagctcgaccaagcgaaccgggatct
		cggtaccgtccgcgaagccctgaagaccgcgcaggggcagctagccgaaacgcgcggccagcagacccaactcgccgacgaactggcccggctgcgcgcacagaa
		aaccggcctggatggcgtgatcaccgcggctgctaacgctcaagcggaacttgacaaactgcaggctcagcagaaacgggcggagcaagcagcagaaacgacgcgtc
		tcgatgttcgtcagctcgaatctcggaaaacggcactggaagccgacatcatcaaattcaccgccagcggcaaggatttggaaaagttccgtgccgaactggctgatacca
		atgcagaactcgaacgtctgcgtcagcaattggttgaggcacggagccggcgcgagactatcgcgattgaagtggaacgcctaacgcaacagcgcggcgaactggagc
		gcaccatcggttcactaacgccgcgagcgcaggaggccgaagcgctacggatccggctccagcaagacaacggcactttgctcgccctgcgcgagcagattgaacgctt
		gcgcactgaacgtgacagcttgcagcagccggtcacatcttccatgcatgtccccggcgacaacgccgcggcacgctgatcaaggatcgcgctgatggacacgaacacc
		ctggtctggcttgcatcgggtggcacgcttgccggcatcgtcagtgttatcaccgcattggtgtgcggcatgcactacggtgcggcgctacgccgcataccggctgcggcct
		ttttggaagatatcgtcgcacgcgtcgcaactcgtcgcgaggaactcgaacggctggatgcccaattgggcgagcgccacaacggcctccagggcctgcggggcgaaa
		cggagatgctgacggcccgccgggatgccttggcagcgcaactgcgcgaactgcaggaggacctggttgcactcgatgggcgccgggccgacatcgcttcggtgcgc
		gatgagttggcggaagcacggacgcaacttgccatgctcgtcagtgaactgaccgaacggcggacgcagcaggagcaactcgaacgcgcggccgaacgtgcccgtgc
		acaactgtccctgctcgaagaacgccggagcgagatcgaggcaatcgatacagccgagcgcgaagcacggatacggctcaccgaggcgcagacggaactgggcacc
		gtcgtccaggcgcgggaagcggcacggcgtgaagccgaggcggcagcgcgcgacagggagatgctggcaacgaacatcgaccggctcaccgatgagcgcaacga
		actgcgcgctgacatcgccagtctccaagccgaacgcaatccgctgtcgactgaagttcagggcctgcgccggcacttggagcagttgcatcttcagcagcaggcactcg
		acggcgatcttcaacgcctgcaatccctacagccggtactggaagataaaatcagcggcctgcaacaggaagttgttacccggaccgctgaactcaaagaccttcaggcc
		gaacgtgatccgctgtcgactgaagttcagggtctgcgccggcacttggagcagttgcaccttcagcggcagacactcgacggcgatcttcaacgcctgcaatccctacag
		ccggtactggaagacaaaatcagcggcctgcaacaggaagttgttacccggaccgctgagctcaaagaccttcaggccgaacgtgatccgctggcagcggacattgatg
		gcctgcgtcggcaactcgaaccgctgcgtacacagtgcgacgaagtcgaagcggaactcgcccgccgccgcgccgaactcgccgcgatcgagcaggagatccgtacc
		aaaggcggtggtagcgtcggcaacccggaagacgtgctcgccgatctcgaacaggcaccggcttgtctggtcggcgacggcggcaggggaccgttgatgccgaatcc
		gcagcgcgacgacgacgaaacagcaatgctcggccgcgtgcggacacaccttgatcggctccgtctgcactttcccgagcgcactctttatgcttttcatactgcgctcaag
		acggcaacgattagtccgcttacagtgctggccggcatttccggtaccggcaagagtcagctgccgcgccgctatgccgaagcaatgggtatccatttcttgaaactgccgg
		ttcaaccacgttgggatagcccgcaggacatgctcggtttctacaattatttggagaagcgctacaaagcgaccgaatttgcacgggctctggtgcatttcgacacgtacaact
		ggccgcttgcccggcctttcaaggatcggctactgttgatcctgcttgacgaactgaacctcgctcgcgtcgagtactacttcagcgagtttctgagccaactcgaaggccgt
		cccgccccgggcgatcgcgatcctgagcacatccgcagttcggaaatcgtgctcgatactggcggcgttggcggaccgccgccacgcatctatcccggccacaacctgc
		tgttcgtcggcacgatgaacgaggatgagtcgacacagacactttccgacaaggtgctcgatcgcgccaacctgctgcgcttcccgcgccccgaaaaactggccggaga
		aacgctggcgagcggcggcgagccggcggaaggcttcctgccggcctctcgctggcatgcgtggcggcgcagttttggcacgctgccggcaacgctgcgcgaaccag
		tcgaacgttggatccacgatctcaatgagcatctagacgggctgcatcgaccgttcgcgcaccgtgtcaatcaggcgatgctcgcctacatcgccaactatccgggtgtcgc
		cgagccgatggcgcaaaccagtcctctggatcaggcccgcattgcctttgccgatcaactcgaacagcgcattctgccgaagctacgaggcattgacctgggtgactctgg
		agtcacccagcacctcgaccgcatccgtgcgttgatcgacaacgagttgcatgatgcaacactggctcgcgcctttcagcgcgccgcgcaagatgacggcagcggcagg
		ccgttcgtgtggaaaggcgtacgccgtgaatcgatatgatcccgctggtgctggctatgccatggggactactggcacagactccgatcgccggccagccgacgcgccga
		ccgttacatgacggtgaaacggtcgaactcgatgggcggtacggtgccatggtggcgctacccgagcggaccgacctgcaactgggcagtcggcgctggccggtgcag
		gtggaaggtgccgcctttgcctggttcgagggatcctttcggttggtgtcgctgccgactgcagccttgaccagcgaacgtcagatccggttcgatcttctaacggcgggcg
		agtctgtgctgagtgtcgggctcgtgttgcgtaatcatctactgcgtccgcgcggagccggacgtgacgatccggccgccgatgcattgcacacctttgtgttgcaggttctc
		gaccgcatccgtgaggccgaaccgtccggtgccggagacgattgggatgatctcggcaccggttgggcgcggctgcgcaccgcctggcttgagcgcgatgcgcagatc
		gaagaagcgcgccgcgatctgatcgtcgaacatgctgaacaactcccggcccacatcacagaaatcgctatccacccgcgtcgggtgctcaaacgcacccgcgagttgct
		gccgatcgatcgtatccaggaactcgacaccgcctgtctcgaatggctgatccggcagcccggcgttaccgttgccgaaaaggccggtccgcgccagcgactgctcggc
		atcgcgcgcgaggagcatctcgatacgctcgaaaaccgggtgctgaaagatttcctgcgtctgagcgtcgaggctgccagcgtctggcagcgggagaaccggcgttttca
		caacagtgagcgcgcccggctggtcgggcgttatctcgcgctgtgccgcatgcatcatcgcgaactgtgcgcggctggcatcggtgaccccatgcccccggtcgctccga

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		atttcgtgctgcaacaagattcccgctaccgcgtgatctggcgcgcgtaccgcgaactgttgagcgctgagcagcgtatggacgatctctggcgctggcagtgtcggttgtg gagcgacttcgctcggcttgtcgtggtgatgggggtgcaagagttgtgcgacaagccgagtgcgctctcgcccctcttcgtgcgcagggaacaggcaagcggacgctggt cggacacgctcggcctgctcggtgtattcctgatcgacctgaacggcaggtcgtatgtggcggaagtctgtgatgcgagccagttgccccgaaacgacacgtcacgagcg aagctggcgtcctggcagtatgcactcggttgcacagcactcatccgcctcatcgatttgtggagtgggcattgtgcgagcctgtgtgtctgggccatgcatagcgctacagc cgagacgcttccgttgaccgagttggtcgcttcagccgatgaagccctgagtacggccatcagacaggaaggtctgcgcaacggcgagcaacttcgggcacgtggactg gtgatccgctcggcgccgccgggaaagaccgagtacgccacccaggctgggcaggtctacggactgacgctggccatcgggtcggaacatatccgcgaggcgcttgg cgagtgcactttgatcctgcaggacagtctggagcgcctgtttgcatgagcggagtgcacggcattgatctcaatggtgtgctcgattgcgtggtgcgcctcgatcgggcac cgcgaccagcgccgacaccgccggtgatcgtctccggttcaccacagggcctgctgacgggagccgcggcactgcaatcgccctgcggccgacctggcatggaagcc gaggaaggtatccgcctgccagtgctggccctgctgcacgcgctcagtggtgaggggcggcacgatacgcacgatacggccgtgctgctcggccgacacctgcgtagc ctgttgtccgatgatacgcatgctgctgtcgtcgcagtgcctgacacacctggtttcgacgaacgagctcgcacccggctgctggatggcgcgctacgcgccgggctcgat ctgcacctactatggcgcccggtcgcagcgttgcttggttggggcgaaacactgggaaacggcgaactccaagccctgcacggccggacggcctgcgtcgtgcagttgtt gccggacggcatctcgattggcgatttcggcctcgaatgcgtggtgcagggtggccggccgacgttagtaccggtgcgccggcgcgacggcgaacgtcaattttactcgt ggagcggtggtggactggttgcactgctcgcgcgcgaagctggaaccgacgaagccagtctgtgggtcggaccgtgggtatggaaggtcttgcttgggcagcctgcaga acgcgaggtgctggccgacccgcatgcaccgggtggttggcgactcgccagcggtccttccacactgtgcggcgccttagccgcggagttgcgcacaggcctgcgtata gcactcggagccgcgcgctcggcactgcgcaatgcagcggtcattctgatcgaggggcctatcgccgatgcaccgcttttggacgcaatgcagccaacactcgcgctac gccagatcgtggctgcggaactgaccgtggtgctcggcccgacggtgtccgcaagactcgtcgccatgccgctcgccgatgctctaattgccagaggggccgctatctgt gctgcgcgtcaagcggcgcggcagatcacgtattacgatttcctgccgatgctcgaaatcaatgtgctgcaggccggagagcatgcgttcgttgaactcatcggtcgcgaa gagcgcatcgcggggggcatgagttacacgaatacgttggccgatcgcttcaccgttgccgcaagcacgcgctcgctcgagttctacctgctgaaagaggacgaagcag gcgctcgtcacagcgaaacggtgctgccggtaccgccggcagccgacgtggaaatcagcctgcacgtcacgcagacacccgctcaaggctacgcacgcgtggagata ctctcggccgtccggggcgcgctcggtgaagcaccgatcctgctcgattggtcagcgatgacagagattgaaggctcgcgcgaggatattctgcgcgaactcgaattcga ggggctcggctatccggacatcgtaccgcaacgtgcacatcacctgctctgggattaccagcgcagtgacggcatgactatcgctgccgcgatgcgggccttcaattgtaa gcctatcctaagttcaccgcgcaaccagtacaatcaattggttaaacaaacgcgcgcactcgtcgggctgcgcagcaatctgtttttttctgacaaagggcaccagttctgatcg tagtgcttacaccgccgtcgattcggatggccaattgccacctggaatcgcgccgacaatccaacaggaattcgaaaactttcgagtgcggctcgacacggattttgccgca atcaccagcgtccgtaatcgacaagatatcgcaacccggcgtgaattggcgcgactgggcgcctgcttgtatgcagcgtgtcctaatgcaattgttcattacttccaacgcatt gtcgcacgtagcgccgatgacctgacactggtgttgcatgccggcaaagtgctgagcaccgaaccagatcttgacagtcttttccattattgcgcgtctcgctacgatgaagc catccgcgctgtcaagagactgtcggtccacgtggtacgcgcggcaggcgatgctttggcttatcatgaaaaagctggaggcattcttgataaccgaagcgctgacaagttg gctgaagctgcgctcctattgctaaaggaggaaatccaggcacataattacaaaatacgattccgtgccgccgcgcgactcggcctatttctgttacgccaccggcagcggc ggcgcgatttcctgcatccgagtagcgctgacacggctaatcgtcggcgtgccaaagagttcgatgccctgttgatccaggctatcgcatcgaagcgccttaaccaagatct ggaaaatgccttggaagaaatccgtgcacaaatccgatatcgcggtacaaatgcgatcgttgatatcgatcctgacgaagatggcgagattaacgagaacgaagtggagta gaggctgttgggcacccgctcgccatccctgtcgagcatcccggcttcgcgggcgcccatcccgtgcctttacggcgtgttcaacggccccggttcgccctgcgtatcggg ctcctgctacgcccgtcgagacgcgctgcgcagactcgacgctcaaatggcttgacgccattctccctggctacc (SEQ ID NO: 404)
41	Helicase-DUF559 + SMC + McrB + DUF2357 + ATPase	atgtctctggttttgaagacggttcgggcttttccgagaggtcggactaccgaagaattgcttgttctcgtcggtgcggctttctcaaatgacaagcggcttgcggctctcagcg aactggagacgctatttcgcgatggtttgatagtgaaaggcaaggacggtcgctggcgtgcaaaggcagatggtttcaaacccagacatgagagcgtgtcggcttcgaga ggtggagggcctgagggcttcgttgatgtcattcacgctgccaatgcattcttctcctcggaaccgacggcggccgaactacctgatcaagaagacgaaagttcagatgctc ccgatccgcaagcgctactgagatattggcgctcggccttgcgtgccgatccacgaggagccacgacccaggttctcgacaaacatggaatcgagtgggccttgatctctg ggcgtggccctatcggtccagaagaagggcaaacgctgactgtttcaatcgaactcgacgcgattgatcctgcctttcgagaggctctggtgcgaagggaaggtcacgag aacgcgcttgcagtgggttggccgatggcggtcggacgacgtggcggagttcctgtctttcgacccgttggcatgttagcagcagcttgggatcgtaaggatgaccgtctaa tcctgacgattgatgccgatgacgttttggtaaaccctgattgggtcaaaagtgccgctcgtgccagcggctggaagcgcgacgacctcgctgaccttttttcgtggacgatg ggctggggctgcgggctcaggattttgtggagaaggtaaggattgccgttgccagtcagatacgtggtcgcgttgtcggcgagaatctcgccacacagctcgatgcctcgg ctcaagggattttgacagcgccgcgatcttcctaccgactgactcttctttcaccgcgggggctgctcgtgacctggatgccattgcgacatggccgaaggaccgccttgag agaactgcgcttggcgcggtattcgggtttgaccttcaagacggcacggacaaggctgctgcaatcgacgcagttccgctgaacaaggaacagttgcgcgcggttcgatc cgcatgccaagcgcctttgaccgtcgtgaccggtccgcccgggactggcaaaagccaagcgatcgtatctatggccgcgtcagtgctcgcagatggtggcagtgttctcgt cgcctccaagaaccatcaagcgcttgatgctgtggaggaccgtcttggctctcttgctccggacgtcccattcgccatccggacactgaacccgaatgacgaggcggatac gggcttcaaggacgccctcaaacaactcatcgacagcgaaaatgtgacgcgcaacgcatctgtcgacgaattcgcattaggcgagctcaaaagcgacgcgatcgcgaga agcgaagtggttagcgtgatcgataagatcacggaaacggaatgcgaaatttccgatattctggaccggattcaagtccgagaggatcgcgggcgccctgacaaccaaga ctctgaagacgtggatccgagacaaagtctcttactccgctttgtctcttggtttggatcgcttttcgccaagcgtccccccaaagtagcgccagtgacagatcattcttcgtccc gccgcggaatgaacgtcaaagagcttcattgcgcgctggcagaaaaaagatatgaacgcgatgcgctcgggacacctgacgatccgatcgccttaggcgagaagatccg

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		ggaagcgaccgagaatcttctgcctcgcattctgtccgcccggacacatctcccagaggatgagaggcgcgaaatcgcagaactctacgatgactggacattcgacgggg gacggggacatcccccctactgatctttcgcgcgtcctcatttcgcatcggcctttgtggcttgcatcgatcttgggcacgcctcgacgcatacctcttgatgacgggctgtttga cctcgtgatcttcgacgaggcgagccaatgcgacatcgcgacggccgttccgttgctggcgcgcgcgaagcgggccgtcgttgttggggatgatcgacaactgtcattcat ccctcaactgggtcaggcgcaggatcgcaatctcatgcaggctcagggcctaccggtcgccagaatgggccgtttcgcccagagtcgccgttcgctattcgatttcgcatcg cgcgtgtctgttgccgacaacaggattactctgaggcaccagtatcgttcagcaggccccatcgtcgattacatcagcgagaacttctacggaaaccagttgcagacctcgta tgacccgaggcgactgaacgtgccagatggggtgcgccctggcctcgcatgggaacatgttcctgctcccgcggtcccgcaaatgggcaacgtcaatccgtcggaagta agcgcgattgttaggcacctgaaaaagctgatcgttgaagacaaatacactggcagcatcggtgtcataacgccgtttcgcgctcaagtggccgctatcgagaacgcggtc gatgccgtcctggatgaaccgaagcgcattgcctgcgagctcaaggttggcacagttgacggttttcagggacaggagcgggatctcatcatgttctcgccttgcgtcggtc cacgcagcccgcagtctggcttgaccttctttcagcgagatacgcgccgtttgaacgttgcgatttcgcgggctcgggcggtcgcgatgatcttcggcgatcttgattttgcac gttcagggcaatcaaaagcgctggccaagctcgcttcgagggcgacggaagcgcggacgaaacggggcgaaggtgtgttcgacagcgattgggaacgcaaagtctatc acgctctgaaggcccgaggtctggatccgcagccgcagcacgaaatagctgggcggaggctggacttcgcgttgtttggagcgaatgatgtaaagctcgatctcgaggtc gacggacgcagatggcacgaaagcccagacggtcgtcgaaagacgtcagacctgtggcgcgatcatcaactgaagtccatgggatggcgggtgcgccggttctgggtg gacgaactttcaagggatatggagggttgtcttgaccgagtcgaacaagacctatcgtaagtcgagcaggaacaccgcggttgcgttggggctgggtggcgccgccatcc ttgcctcgggctttctcgtcctgcaagtcaactcgctcgatcgccgatatggtcgtatcgaggaaaatctgagctactacaccggggaactccaatccgcgcagcagcaact ggcttttgctcgtgagcagtttcgcgaactttctgaccaaaagcaaagcttgtctcaggaagtcgcgagcgcccgaacgcagccttcaaagcgcggctcagagagaggcgg atgcgcaggctagtgtcgaagcaagccaggccaaattgactgctgagcgggaccgtttggccgaagcccaaaaaacgattgcggatgcgcagcgaattgaacgtgaaac tgctcaagctttgctgcgaagaaatggcctcgaaacagaggtggtcaaactgaaaggcgatgtgcaggcccttaaggagagccagcaagagttgtctgctggtgttgacca aacgcaatcggctgtcgatcgcctcgaagagagaagagctgaacttcaacgtgaagtggatagactcgcgcccgccgttgaagaccttcgtgcacaggagcggcttgtcg aacaactgcgaggtgacgaggatcgtctcgaacagagcctcgacgatttgaatgcgaacattgcaattgcacggactgaattggcgaccagcgcggaaaaggtcgatgc ggccgaggagaggctgcgtgcagggcaggaacaaatagcatccacagaagctcaacttgaaacactgaatttcgaagtcgatgacctcgagtcgagacagggcgaact gcaggcaagtgtctcgggagcagagacgcgtcttttttcattgcaaaatgaactggagatcgcacagaacgcggtgacgcgagctgatgcgcagcgcgctgaaactaca gaagcactcaacatcgctcaggaacagttttcgacgcgaagcgctcagctctctaccctccagtcgcagattgcatcggcagaggaagagcttgccgaacttgaagagag acgggcggaattcagcagattgcaggctcaaatggaccagctgcaagcacgtcgaacgacactagaggaggttctccccgatcttgagaagcgagttcaagcagagcgg gctaatttgggttctatcacgacagaagtggagacagagctcgggcgagttgctgtactcaaaggccagggttccagtctggaggccgacatcgagcgcctccaagagcg tcgcgacgaactcgggctggaaacgcagtccgccactgctgaggcggaggccgcgcgcgcatcccttcaagctgagcttggtcaacttgcggaaaccgatgccctttcaa gagcgcggactgccgatttgaggcgcttgagagaagctcttggagctgctgaaagagagctttccgaacttgaagagagacgggcggaattcagcagattgcaggctcaa atagaccagctgcaagcacgtcgaacgacactagaggaggttctccccgaacttgagaagcgagttcaagcagagcgggctaatttgggttctatcacgacagaagtgga aacagagctcgggcgagttgctgaactcaaaggccagggttccagtctggaagccgacatcgagcgcctccaagagcgtcgcgacgaactcgggctggaaacgcagtc cgccactgctgaggcggaggccgcgcgcgcatcccttcaagctgagcttggtcaacttgcggaaaccgatgccctttcaagagcgcggactgccgatttgaggcgcttga gagaagctcttgctgctgccgatgatgagctttccgagacacgagcggaactgatggacggacagtctgtggaacaggaaccagtatcaaccattagtgaaggcgctggc gcccgtgaaaacgctcagtctgacaactccgcgccatcgagcaccgacaattgaggtaaccgaaaatgcttacggacaatacaatacttgtgctggcgattgcgggtgtcct gatactgctcgccgtggttcaacttttctggccgcccgccacgaccgggcggttacggcagcaggcccgatcgaagagcttgccgtctacgagaagcggctggaagaaa aacagcggctcatggacgatcttgaagctgaagtggaaaaacgtcgggaggcaatggccgtcgttactgacctccgggctgaggtcgacggtctacggcgtcagaagga ggagctccttacagaatgggagagtctccgtgaacgtcgcgacgaagttgcggcagttcgcaaggagactgaggacgccgttgtcgaacgccagcaactcgaaacgga gatcgccccgcttcgtgcggagtatctggagataaaggaaaggctggaaaaggcggaggagctcattgagcgcactgacgccttgagacgagagcacgacgaaatctc cacacaggtcaaagatcttcgggacaagaagaggcaacttgaagaggccgaggaacgggtttctcgcctggaagagcgttccttcgaacttgagacatcgaatgctcggc ttgagggacagaagtcttcgcatgaaagcgagttgtccgccttggaagcgcggatcgcctcggaacacggtgggttggcatctgcccaaaccgaacatgctcgcctcgat gcagaggttgcggctctgaaccaggaaacccgccgctccaggggcgaaatcgagacgctccaggacactcgaagcgcgcttgatgctcgattggcacacctcaaggcc gagatagctcgccgagaaggtcgaaccgtcgacggggaaaccggcgaaacggatccgcttcgcgagctcaatgaaacaccaccggtcattacggagatgaggacctg ggacaacgcgccccgcgagaacgaggcggatgccatcaaacgcgtcgaacgccgcctacgcgcaaagggtctcgactacccggctcgcacgcttcgcgcttttcacac cgccatgaaagtaaatgaaacaacgcagatggcggtccttgccggtatttccggaacgggcaagagccagctcccgcgtcaatacgcggccggtatgggcatcggtttctt gcaagttccggtgcagccacgttgggatagtcctcaggatctgatgggattttacaactacatcgaaggcaagttccgacccacagacatggcgcgtgcgctttgggcggtc gacgggcttaacaacgacgatgcggaacaggatcgcatgatgatgatcctgctggacgagatgaacctcgcaagggtcgaatactatttctcggacttcctcagcaggctg gaaagccgtccgcgtcccgatgacgtcgacaatgaaaacgaacgcaaggacgctgtgatcgagcttgaaatcccgaacatggaacgcccccccaggatttttccgggcta caacctcttgtttgcgggcactatgaacgaggacgaaagcacgcagtcgctatccgataaagttgtcgaccgtgcgaatatccttcgtttttccgccccgaagaaaatcaagg acggacaggcagaaggaacggtcgagccgattttggccctttcgcaacagacatgggagagctgggggcggtcgagtgcgtctgtcgatggcggtcggcgtgtcacca accggattgaacaaatggttgatctgatgcgtgacttcaaacggcctttcggtcatcggctcggacgcgcgatcatggcttacgcggcgaactatcctgaggttgaaggcgg ccgcggtgtcgacgacgctctcgcggatcaattagagatgcgccttctaccgaaactcaggggcgtggaaaccgacatggctggccctcagttctcgaggttgatgaccttt

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		gtggaacgcgagctgggggacgacgccttggcccaagcaatcggtgagtcaatgtccctcgccgaggcaaccgggcagttcgtatggagtggagtcacgcgttgatgcg gtttctggcccgtccctgggcggcgaaagcccttggagaggacgaagcctttgggcccgaagactgtctgatcggtagctaccagggggcgaacccaggcggctacga atacgtgacgctcttgaggggaaacgtccgaggtagcgataccggaactgttctgtttccctatccaaagcgtgaggaagctgtcgggcccgcgcgtaagggcttcccggt gcgcccaaggtcggggcacgatcctgccactccggacgaagaagaaggcgcagaggcccttcgacacatgaacgaagttcttgcacgtatccaagaactggaaggtgc gattgaagacccaagcgatacatgggggcgcctgagggatgcttggaagcgcgccgaaaatgaagccgaacccaaaatggctgaaatcgtccggcaggcgcggggca tgcttccggtgcttcgcgatctggaaaaacgcatccgccgggttctacgtaggcacagggagctaactccccttgatcgggtgcaggagatggatcggacctctatggtgtg gctcagccgacagccagggcgaagcatcgcggaacgtgcaggttcttcgcaacgaattcttgcgacggttcgccgtgagaatttcgatacgctcgagaaccgtgtcctgca tgcctacacgcgtcttgccgcagatgttgcacgcgaatggacccgtgagcaccctcgtgcgaaggacagtgttcgctacaaacaggttgaggcttttaggaaggcctgtcg agtattgtcgcgaacactcagtgacctcggtgtcatgatcgcgtcggccggcgtccagccaaactatgtgctcatgcaagatcgcagctatcgagaggttcatgagggatgg ctgaggcttctcttacgccgaaaaattgtagatgatctttgggcttggcaggccgaaacttggacggatttctccgttctttcgatcattcttgccatcgacgaattggaagaggc tgaacttgtcgctcagtcgccgatttcgtggagcggtgaggcaacaggcggacgctggttcaatcaggatcggccaatcgccgtcttttggctgcgcgacaccaaccgcatt gttgaagtccaagcacgccctgagcgaccaggaaccatgttgagcgcggcacaagcgcacgtcgccctcagaatttccgatcccaaacgggctgaccttccgcgcagga tcgctgtctggacgccacatgccatgcgtagaattgatctcgaggatactgtgcgggggcagttcaactgcttcaccaaatccagcccctcgctcagacggaagttttgcg gaatgggttgatcatgaccccagcacgtggtgtcgcagctgaagagagcgcaactcacggaagagcgatcgttacggcaatcgccataggcccagccggtgaagaccta gcgaagggattccaggccgtgcgcgacttcattcgcagtgagctatacgaggtcgcaacatgatcgaccgaaaactatgcggcttcgatctcaacggatggagagatttcg ttgcgaagaactggcgctccgtgccaggtgaagacgaggtcattggtccgaccgatatcgtcacaagtggccctattcgtcgatcgtgcggatcggggaaagccgcctcg caggttggatcggaggaccgcaggctgacattgctccgcacggtcgcggtggtggttggggtgatgtcgggtcagaacaaagacgcattcccgttcggtcactgctggaa atgcgtgatgacggggtcgaaaaactcgcccaggcacttgtgggatctgcgagcggttcggcaaacacagtcgtttcgatcgatgagggcccggatggcgatgaagccgt ccaagagcaccttctcgaagcacttgcccgagggaagttccgaaatggctcattggtttggcgaccagttcttgccgccttgttcgccattcatcgcgatcaggtttcggagg ggcagcttgtaggcgtcgtctcccatcagcgccaaggcttgtcagttcaaaagctgcgtattcgtagcgcaaggaatgtgctcgccccggagcgacgcgaggccgctgcc catataccgtgcgacgctggttacgagtccctattccgaggtgcccgcaacgccgctgtcggggcagagggtttttcggcgcgcacagctcatcgtgcgatcgcaagctcg gtcggaaaagctggtttagggatggattgcaatcctgagatgctccgcatgcccaacggcgattgggagctcttggaccttaataaatttgacgcgtcggaagtggtgagtgt cccgagttccgagctcgatctggccgattgcgacgtcgttcttttcgagaccctttgtgaaggtcggctcaaaaaatgcctgagtgatgctatccaaagagcagctccagtcg aggtgctctctcttcccgcaacggctgttgcggaaggtgccttggaagcagcacgccgagccggggacggggaaccgatcttcttcgactttctaccacgattgtccaccat cgtgttcggatcggatggcgcaaagaatttcgatctcatacggaaagaagaaacgctcgaagcaggccggacctacagaagccctgaagcagcatctctcgcgataccg gcagggcaggagagcgtctctgtctacctgaggaaagaggaagctccctggcctcgaaaggcaagggtgtcgcttggagctcctctgaagcatcaagctgccgtctcgct gtgggtcgaacagaaaccggccgccgggcgagcgcggatcctcatggaatcgccggacttggggcggaatttcgcggtggattgggatgaagcactggaagaggaac ggccctggtctgagatcatcgagagcttggatacgcaagtgtcaattcccaaacgtctggttcttccctgcggcatggaggcatggcatgacagcgatcgatccgcaggtat gctaactttgctcgaatccgagcctaatcgcagccgcacggattgggcgacccttcggcaaaaactttcacagcgtccctttggcaaatactgcatctcaagtgacggcgac gtgcctccggagatcgcggcagaaaccctcgagcggtttgaaattctgaccagcaaagcgcttgaggttactgaaaagcgcctgaggggcgaaagcggctacggaacg gaagacaatgaggctctcaaattcttgagttggcagttccgccgatgcccgcgcgatgtcgcgacgtggctgatggactgtattgaagcgtccgggcgcaaccatccgttcg tcaaacatcaagcaagttgggttctcgtatatcagggccttggccgcatcgtcggaaacgaagaggacgaagcgagagcaatgcggttgcttctgacttcgtccattgagga ctgggtctggaaccgacaaagcgcggccatggcgttcatgctgtctcgttctgacagcgctccatcttacctggaacgagaagacgtagagaagctgaccaagaggactat cgcggacttccaacgtaatatcggcggccaatatacaatgtttaactacgcgcctttcttacttgcaggcctgataagatggcgtctcgttgatcctaaagctttggtgatcggg gccgacccgttggcggatgacctcttggctatcattgagaaaacagagcacgacctgaaggcccgttgtgggtccaatatgaatttccaaaggcggcggtcgaagttcttgc ctatcctccaagacctgaagtcagagctggcgggagaaggttcgaatcctgacctgttgttggatatctatggagcgagcggaacgtga (SEQ ID NO: 405)
42	GTPase + GTPase + TM	tcgcgatcaaggggtgagcaggggataaacgcaaagacattgaagttgaggagaatttagttgccttacctgcgaaaaatctgagcgatcttgcattaaagattttctatctca ggccgatgctcataagagcatttcctgaatttcaccctttttttgctcgccatccctctgcgaataaggacaccgcgccagatatgtcactcatcacccatacattagaaaacctc acaaaagccttgcgtactgcgttgcgtgtctcaattgaatgcaatgagcgcagcgaaaatacccataaaattttaaacgtgttacgtcaggttgagctgacgctgatgctgcat caacaacctatctatgccattgccggtacgcagggagcgggtaaaaccactctggcaaaaagcctgctgggcattgacgatagctggcttgaggcgaatccgggacggg gcgagcagataccgttatttattgagcaacggcacgatgttcagggtgattatccgcaatttatttatgtctgtgctcaccacaaaaccggtgaaatttttgacagccagccgcg cagtggcgatgagctgaaacagatgctgcgtgactggtcgcaaatggtgaatcaggagatagaagggggcaaaatcctctatccgaaattaatcattaataagtcagacagt tttattgatgaagagatggtctgggcgctgttgcccggctacgagatcagcaacagccagaatcatcgctggcagggcatgatgcggcatgtcatggtcaacgccagaggc gtgttgctggtcactgacccgacgttaatggcaaatacgaaccagagcctgctggtgaacgatctgcgcagtgtgttcgccgatcgttctccggtgattgtcgtgaccaaaac agaaagcctgaacgatgcggagaaggccgaggtaaaagcgagcgctgccgcactttttcatgagacctcctcaccggtggtcgctgccggtgtcgataatcaagcgcagt ggataggtgagctccgcactgcatttgctgagggtatccataatagcgccgcgtcagaagcggccgcgatcgaacgtttgatgactctggtcaatgacgatgttgcggatatt attgataacctgaatctgctgtacgcggagcaggacagtggcgaggaacgtaccgtcgctattcttgaagcgttcgataaagcagccgagcgctatgaacagcaactgcgt

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		aaagccatcaaacgagaaactgacgggcatcggcaaaaagccactgaatcttgccagcgccgttatcaggaagaagaagaagggccggtcaataatttaaaaggactcg gtcgtcgtctgatgtttcagggggcggagattgatcgtgaacgcaaaaatcgggtactggacgcctggcaaacccgctttgagcagcaatctctggccgatcacaatatggt cgcgctggaaacgctcaaccgtcgtgagttgaggcattacggtctttcacaggagacgctgtcaccccaacggttgacctcgcccgcggcgacaatgggatatttgtcggt ggctgaggaggataatttttcctcgctggcccctttgcgccatctgctgggatcggctgcaacaagggatgcgccgccgcagttagaccagctttccacggtattaaaagtgc tgcctgccatgacgatggaatatgcgcgcggttgggtggcgatcaaccaggcgatgcccgcagcgtcagagctaaccagcgagttgcggccacaacaaattctcgacgc gatttttagcgcgcagagtagcatccacccggtgaaaaccgcgctgatggcgtttatcggtgccgacgccgcggacggcacgctggatggcgaagtgggcactccgcag aatgaagatagcggcgtatttacgcctgtcgcgatagcaggcaaagcgatgctggtcggtgcggcggtttatgcgttgtatcaggtggcgggcgtggtgagtgagagtgat aaagctcaggcctggtatattgaacggatgatgaaggaactggcgcaatataatgaaaacgtcatcatcgagcgttatcaggacacgatgggcgatctgcgtcagctgattg aaatcaacctcaaccgtttatttggcgtgcaggatgtcctcacgcagaaaagctatctctggttagctattcagggactcacgacggtacaaaaggaagcccggcagtatgaa gccagtatcaaacaatatctggcgtgatatttgccatgagcgttatcgatgggcggaaaatagctacatcaacctgctgcgtcaggttgatgccgagcggttaatccagcctc atgcagacatctcccgccagatatcggtcattgtctatggtccgacgcaggtgggaaaaacctccctgattctgaccctgctgggcgtcagggatgactgttttaaagaactta accagctgctgcgtggtgggcaggcattaggtcacgcgtcaacggcgcgaacttaccgttaccggatatcacgggatgatgcctggtattttagccacaaagaccagggaa caaccgcctggtcggatagcggggcggcagatattttcgccagcctgcgtgcagaggttcaggcgggcaggcgctactttgacagtatcgacgtatttattccgcaacgttt cttccatcctcagcagcggcaaaatggtttgttaatccgcgacctgccgggtattcaggctgcggatgacaatgaaagggaatatgtgactcagcttgccagccagtttattcg ttctgcggatgtgatcctgctgaccggcaaagcggattatttaggctttctgaaacccgaggagttgggtaatgacctactggctgactggttctggcagccacatcgctacaa aattgtattaacccggacttttagcaacagttccattcgggaaatgttgcgccgtgtttcccccgataaatcctggctgcaggcttatttgtttgagcaaatcaatacgctggaatt gcaacttccggcggagatgcgtcaacacatttatccgctcgaatgcggtcactcctggcaaaccctgattgagggggtgacgattatgctgactattgccaacggttgcgtg agcagatattaaccgacctgcgccatcatatgttgcaggcggtccatccactttctcgtttacgtacgggatacgccttacctgaattaattatccgccaccgggacaagttgca gcagcagtacacagcgctgcacagcacgctggacaaagaacaggaatattacctgcgtaaaaaagagcagctgtcgtctgtgcagactgaatattcccggcatctggcaa agagccagacacgactggacagattgcagcggctacgggaacggctgaataaaagacaggcgcgcaacgcgcatcaatccatcgctgtgccaccgatgggcacaaga acggtcagtgccttactgaaaatgattgctgaggcaagagaagagatggcgcttcatccggcgttaaagcaccttcctgcccatttcgctgcgcaacagattaaccaccatgc cttcacggcgattgagcaaaagctgcatggctatcatgcggataattatctctttgccagcaactataagcatgactatcaggaaacgatcaacgcgatcaaacaacacctga aactgatcaccacattagccgctaatttccagcgtagtgagctggagagacacatcaaggaacatcgtcgtcgccagcaacgtttacaacaccacaccacccggcgagac aaactcctgacggcagtgaccaataagcttacgcgcatcaatacgcagcaacaggaattaacgcacagccatatgcgtgacgaggatcattatcagcagctgattggcgag agccgtcgctttcaggaactgatcagagtggcgaaaaatgaacgagccaccctgattgaacaacacattaggcgtacggatattggtcaggctgagcgactggcctggcta ctcgctgcccgtgcgttaaagaaagactacgaatatgtcagagcattaggagagtagtgcatgtcagtggaacatgacccggttattgcgcaggataatgacgagcggatg ctggatgaattggtgcaggaactgtttctgaccttgctgacgcgtgagctggcgcaacagaaagcggttatcgaaaccattaatgacaacgtctcgtatcaggctggtgagtc attaaaatcgttgaaacgggagatcaaactttccatcagcaccctgtcgaatgcgcaacagcaatatcaggaagagcaggccatcgccagggaggaatacgagaagcgg ctggagcagcagactcaaacatttgccagtgatgcggaaaaaaatcaccaacagtcacagcagcagatggcagcacttcggcaaggtgagcagcagctggctgcacagt taacagatttgcagcaacagcatgccacacttcatcagcgctcaggtcagatgctgaatagcattaaatggctggtggtggggctgggggcgtcaacctgctgctgtttgc ggctgtcatcatgatgtttttctcgggcatcgataatcatccgcgcatgcaggtttgtccggatatggtgcgcctggtgcaccatgacttttctctggcacggataaacggacg cacaggcagcgaatgacgcgccctgaataaactggcacaacttctgcattcatttcctcaggcttgtatacaaggccgcataccg (SEQ ID NO: 406)
43	TM + GTPase + GTPase	atcagggcaaggaccgttgcccatatgtgactggttttggtgtcggctatgtggccaggctgcgtgaaagctactgatcgcttttaatctaagtggtggatttatatgatcaatc attattgataaactcatgaagaaacctaatttatttaataaaattaaaaagtatacgattagatattgcgggtgtagatatgactcaccacattaaaggtcaaggcagacatcaggt gacgttgctctctgacgtgcttgatgattttgtcacagaagataaaaacacgttgaagagagaaaaatgaataccgcagaagactttaaccgcctctatgccgacgtttcacgc aatattcagcagacgctgactgatatcgctgcacttcatgttgaaaatgaagagggaaagcagcagctacaatcgatggtcactcagttgcaatccctgcaggatggctttaa ccagaagctcacgtggctgcaaaagcatgccgaatgggacaaatttaccctggcattctttggcgaaaccaacgccggtaagagtacgataatcgaatcgctgcgcatcttg tttgacgaagaatcccgccgccagctgctgcaaaaaaaccacaacgacctggaaaaagccgagctggaattacaggaaatctcggaacgactgcgcagcgacttagggc ggatctatagcgatgtagtggataaaatcaccgatatcagtttttccgctctgcgtctgatgcaaattctcgacaatgaaagcgccctgcgtcacaaacgggaagaggaagag agcaaggaacgcctgctggttgaaaagacggaaagccagtcgcgattgcaaattctgcaaaaacacaccagcgccaaaacacgattaaccctgtgcattgccgccgtcat ctcttttgtcgcaggcgcaggcgcgagcgccgccgtggtgttcaatatgatggcggggcaataggatgagtaacgcactagatcttcaggctagtaccacgtcagtacgttc gcaacgaaagtcctcattgaatattcaggagctcctgaataaaacgctgcctcacctggttcagaccataatcaggaatgagagattaaaaaacaccctacttcaggttgatg gtctcattatcggtaccggcgaggcggattttaccaaagggaatacccgctacgccttacatattgacgataagaccttccatctgctggacgtacccggcattgaaggcaat gagtcacgctatatcagccaggtgaaggaggctatcgccgaagcgcatatggtagtgtacgttaacggtaccaacaaaaagcctgaaaccgccaccgccgaaaagatca aatcatacctcgaatacggtacgcaggtttatccgctggttaacgtgcgtggatatgccgacgcctatgaattcgaagaagatcgccacgatctgatgcagcaaggaggcgc aggagaagcgctgaagcaaaccgtcggggtactgcaaccggtgctgggctccgatgtgctgcttcccggtaactgcgttcaggggctgctggccttctgcgggctagcct atgacgatgcgacgcaaagcaccactatccacccctcgcgcgcgcacaacctcgccacgcaacagaaacgctatttccagcacttttcttctcgtcgggagatgcaggaatt

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		tagccagattgacgccattgcccgcgtcattcgcggtaaagtcgccacttttcgcgaagatattgttgaaagcaacaaaggcaaagtgcgagagtcactgggtcagtatctac aggtactaaacacgcaactcaccaatcatcgcgcattIctaaagaaaacagagccggaatttgacaaatgctgcgtcgcctttgctaacgccattgcagcctttgaacgccga atcatcaataaccgccgtaaccgctggaacgactttttcaatgatctgatggaaaaaagcgacgacattgttgaagacgattttggtgataaagaggcgattgcccagcgtatt agccagcagtttaaatcgcgtcgcgtcgaggtgaaaaaattaatgctccaggacactgaggagggcgttaaggccttacaggagcagatgattcaagcggtggctcgtttgt tgcaagatattaagcacattgagttccagcagcatgtcgatttcgcccacggcggtgaattcgaatttggtcgcgagatcgcgctgggttatgaccttgggttaagggatttcg gctcaatggcctttaaaatcggcagctacgccttaagcggcgccacagtcggtagcgccttcccggtgatcggtacggccattggtgccgtagcaggcgctttagtcggcgt cgtcatgaccgttgtcggtttctttaccagcaaagcgtcgaaagttcgcaaagcgcaggggaaagtgcgcgacaagctagaaagcgccagagataaagcgctggacggt attgatgatgaggtccgtaacctggttgcggctatcgagaatgaactgaaaagcagcctgctgcaaaaagtgaatgccatgcatacggcattgcagcagccgatcgccatttt cgaacagcaaatcacgcaagtcacccatttaaaaaatcaactcgagaacatgccttatggaacaattcaaacagttcagtattgagaagcaggctgccattaactcgctgcta cagctgcgcggcatgctggaaacgctgggcgaaatggagatcgatgtcaacgacgatctgcaaaaaatcgcgtcggccatcacagccgttgagtccgacgtgttgcgcat tgccctgttgggggctttttcggacggtaaaaccagcgttatcgccgcctggctcggcaaaatcatggaagatatgaatatctcgatggacgaatcttctgaccgtctgagcat ctataagccggaaggattacccggagaatgtgagatcgtagataccccggggctgtttggtgataaagaacgagaaatagacggcaaacaggtgatgtatgaagatctcac caaacgttttatttccgaagcgcatctgctttttttacgttgtcgatgccactaatccgcttaaagagagtcacagcgccatcgcaaaatgggtgctacgcgatctgaataagctgt catcgaccatcttcatcatcaacaaaatggatgaagtgactgatttaaccgatcaggcgctgtttgcagaacaggcggccatcaaaaaagagaacctaaagggcaagctac agcgcgcggcaaacctgaatgcgctagagcttgaacagcttaatattgtttgcattgcttcaaatccaaacggtcgtggccttcccttctggttcaacaaacctgaacattacga aagccgctcacgcatcaacgatctcaaaacagttgccgctgagattctgaaaaccaatgttcccgaagtgctgctggcgaaaactggcatggatgtggtgaaagatatcgtc acccagcgtatcaccagcgcccagctgcatctcagcaaactcagcacgttcgttgcgaaaaatgatgaagatacttcgcgttttacatgcgatatccagcaaagccgtaacg aggtcaaacgtctggctggcgaaatgtttgaagaacttagtttgctggaaaagcagctgatgagccagctacgcccgttggagctggatggcattcgcccctttatggacga cgaactgggctataacgatgagggcgtcggctttaaattacacctgcgtattaagcatattgtggatcgctttttttgcgcaatcctccgccgtcacgcagcgactgtcggacga tattactcgtcagcttaattccagcgagagcttcttaagcggagttggcgaaggggcatttaaatccctcggcggcgtgtttaaagggatttccaaaattagcccggagacgat taaaaccacgatttttgctgcacgcgataccattgggcaattaacgggctatgtctacacctttaaaccgtgggaagcgaccaaactggctggcggcatcgctaagtgggctg gtccggccggggccgcatttaccatcggctctgatctatgggatgcctataaagcgcatgaacgtgagcgagagctggaagaggcgaaaaatgagttgacccggatgatc aaagatccgttcagcgatatctatagcgtcttgagttcagatgaaaagacgttcgctttctttgccccccagattcaagagatggaaaaagtcatttgcgatctgacagaaaaaa gcgacaccattcggaagagccagcaaaagctaagcatactccagcagaagctcgagcagtttaaccgttcgagcgagcagcaagtgtcctgatacacaaacggcagccc gcaggccacgtttagttataaatcaaactaaacgtggccaggtgacatgcccccccgttgattaacacacgttatcgtcgggtggaaaggacaacctcctacgtccgcttcaca gcggacactcaggtttaacagtccagtacgtttagcttacggataaatcattttatgatgatgtggagaatgggggat (SEQ ID NO: 407)
44	Dcm + HerA + Vsr	gacagcttccagggtatcgtggacgcgtcatgcaaagagatggggatgagggattttaatattctaccccttgtaccccatgccagtggtcgacctcataaatcattgattttaa aagcctcacttagggcgctcgctgccaccgatgccccacgatgcctgacgatcttcaacgactccccgcaaaagtccctatgcctcggaaaagccgccaaccccaacaac accacctaacaacaagaaacaggacctcgtgccgagcttgttagcgcgactgactagccgtccgaaagcaaaaacaccgcgagccaaacaaggcaatttcttgcccccct aaggaaccacctgaggattgaacaccagcgcagcttactgtatataaaaacagttaaagtcctgttctcaggctgcatctggatcacacagccgccgttactcggaaacacg gcggattagcgcgcacgctcaggccctccagccctaacggaatatgaatatccagaaaatcaaacacatatcagcctcacgcagcgcatagcgccctgccagaacacag caggaagtcattgcgtttgcgttcctggcaatccatcattcacggttagggcccctataagacctgcagaagcagcgcgccatgggcagacccggcaaaagcccccaaac gggtgtggagaagctttatggagaaggaaatcccccacgaaggattcacaggctctagtaaagagccgctccagacgctccttccctttaatatcgatgaacccgggcagg agcccatgaaaatccaagatttcccccactccccgcctccgaacagccgttgatgtttgcagacttgtttgcaggctgtggtggcctgtccctcggtctctcactttcaggcat gaacggcgtgtttgccatcgaacgcgacaagatggctttctcgaccctatccgccaacttgcttgaagggcggaaggtgccggctccgcagttttcatggccctcatggcta ggcaagaaagcctgggcaatcgacgaggttctcgaaaagcacccgattgagctcagtcagctaaagggcaagatccatgtcttggcaggaggaccaccctgccaaggttt cagctttgcaggaaaaaggaatgaatccgaccccccgcaacaagctgttcgagaagtacgtcgaaatggtccaggccatccgaccatcggcccttgtcctggaaaatgtccc tggaatgaaggtggcgcacgccacaaagaaatggaagcaactaggtatctcgatcaagccccagtcctactacgacaagctggtagagagtctggacaggatcggatac cacgtccagggcaatatcgtcgactcctctcgcttcggggtacctcagaagcgcccacgcctgatagtaattgggctcagaaaggacctggcccagcacctcgaaggcgg ggtagcccgagcctttgtgctgctagaggaagcccggctcaagcagctacaagagttcgaccttcccgaggccatccatgccgaggatgccatctcggatatggagatag gtcacgcgggaacgaggccctgcaatgaccctgactcccctaggaaattcgaagagattgcctataccggccctcgaacggcgttccaaaggctcatgcatcgaggctgt gatggcaccatcgatagcttgcgcctcgccaggcacaagccagagataaaggctaggttccaggcgatcatcgacgaccccaactgtgccaagggcgtacggatgaacg ccgagatacgccaagcatatggactcaagaaacaccgcatctacccaatgcaggccagcgctccggctcccactatcacgacactgccggacgatgtcctccactacaag gagcccaggatactgaccgttcgggagtctgctcgactgcagtcattcccggactggttccagttccgaggaaaattcaccactggcggtagccaacggacgaaggagtg cccgcgctacacccaggtgggcaacgcggtaccaccttatttggcacgcgccgtcggcttggctatcaaggcaatgttggatgaggccgtgatgctcgccggccaacagg cagagcgagaacaagaagagaaaatgatagccatcgcttgaacacataggagtcgaggggaatggatagctcccaactggaagggggcgcaatacccggccgcgcttgt cgactgggccggccatcactcaggaggcgtaaaaaggctgctggataaaaatagcggccagcctaacaagcagctgctacggacgaaccttttgtcccgtctccaggcct

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		gggctaacaggcttcccaccgagacctcagctgtccccaggattgtcctgcttgtgggtggtcccgggaatgggaagacagaggcaatcgagtgcaccatccgctggctc
		gacgagagcctcggctgcgatggccggttggtcgaggaactctcgaaagccttccatccctcaaccggctccgcagtcccccggctggccagggtagatgccggcagcc
		ttgccaagctagatagcagactgagcctcgacattgtccaggatgcctctgctaccgccgggcatgagggaagcaccgcccccgtccttcttatagaggagcttgccaggc
		tactggatggacctccgacccaagcctatctctgctgtgtcaatcgtggtgtcctcgatgatgccctgatccacgcaatagacaacaatctggaacaagcacgaactcttctcg
		aggcggttacccgggctgtaagcctggcgtacaacgcgccttcatgctggcccctcgagggtttcccatccattgcagtctggccgatggatgccgagtcgctcttggtaaa
		gccggacgacgagcccgtagcccctgccgagatactcctaggccaagccactgctcccgatatgtggccagcgaaaggggaatgcccagcaggcgacaaatgccctttt
		ctgcgccagccaggccatcctcgcgcgggatgagaacagggcatccttgctgaagatattgcgctggtatgagctcgccagtggcaagcgttggagtttccgggacctgtt
		ctccctcacctcgtacttgctagcaggccaccatcctgtagtccacgatccctcagggactccccaccagtccactccttgccaatgggctgcgaaccttgtcgacctcgacc
		aaaaggccctaacggcgaaaaggcatggcaagcagtcgctaactgccattttccacctgtcgacttcgagctaccaacatgcgctcttccatcgctgggacaaggacgcag
		ctacctcgctccgccgcgacctcaaggatcttggcctcgagaaggaactcgagatggaggaagggcgaaccctaatggggcttgtctatttcctttcggagcgcaaaagcc
		actatctcccagcgaccatcgcccctctgctggaggggctggtcgaaacgctagatccagccttcgcaagcccagacggagaagttgcagtcagcagtcgaaacacaata
		gtcctcggcgacttggatatgcgtttcagtcggtccctggccggaggtattgaattcgttcgtaagtaccaggtgctatcgccaaacgagctcgatttactccggcgcctatcc
		gcatcagacgccatgctttcgttaccgagcatacggcgcaagaggccggtggccgccagccgagtccagcacgtcctccgtgatttcgcatgtcgcctagtacgcagaag
		catatgcacccggacggccatcgtggcggacgctcccattctcgaggcattccagcaggtcgtcgaggacagcgacaagcaccatcacctcttcaaggtggtaaggcaa
		gtaaaggaattgctgaacactgggaaggagttcgaggtgtcactaaccactacctttggccaaccactccccccctcgacaacgccaggcaacgctggtcgtcccgcagag
		cccggtccggatgtcccccagaacaacaagggacgccctcacccaccgatttgctatctccatgtcggccaagggcaatcagtccagccagtcccactgacctacgacc
		ttttcaaagccgtgaaggaactggaaagagggctctcacctgcatcccttccacgcacagtcgttgcactgctggacacgactaaggcccggctttccggcccgattgtccg
		cgaccatgaactactcgatgatgcccggatccgcatcggcgcagatggcacggtggtcggccgctcgtggaatggttttgctgaaagccgggaggacgacgtatgagcct
		tgcggatttcaagcagaccccgtggagcaaatcacatccgaactaccagaagtcggccctggcaatcagccctgcccctgagtatgcgagctcggaagtcctgcttgcctc
		gctctaccgaaccataggcttcgcaacagccagcgagggcggcgtgccgcaggccgggcgagatctagacaagcgtatccagaaactccgcgagaaacgccaatccc
		caccaacaggagcggtagtcggtgtagaggcttggaatactgtgcttcacgggatcctggagagcccgaagcttcccaaccagtcgtccaagcgtttcctccaggtaacgc
		ccatcgtacccggggccgcactcttctccgggtctgcccgtctgagcagcaactcgtggcccgcaggcagcttgattcgccgcatggtctgcctgggatcgatggatgggg
		agacggcgcaacgactttggcaacgcctcttcgctgcattgaacgtggacgacgaggacgatgtcttcgcacgctggcttgaccaagagacatcggcgtggaacccggg
		agcaagcaactgggcactctcgccaatacccgcggacgagatggtcacgttggagacggcagatttcctggggatcccctttctccccgcccggcgatttaccaaggacct
		acaggccatcatgcaggccaagggttcaatgacccgccggcagtggactagccttctcgaggcattgcttcgcctggcagccgcatcccacgtgacgtggctgtgcgacg
		tccacgccaggacttggagctgcctgtgggccgcactaacggatggcattgctccttccagtgaactggaagcaagacgggcgctgttcccggaagccccgcagtacatg
		acgtacgggggaaaagccctccaaggcatcaaggacaaggtgtctagctacctaaatgcccggctgggaatcaatgccctcctctggtctctggcgcagataggagctcc
		ctattctggcaacctctcctcgagcgccggaattgctgcactttgccagcatattcgtcagcacaaggccgagcttactcgcctaggcacgcttgagacgattgccgatgtgc
		gcgagcaagaagcccgtgcgcttctttgcaagaaaggcatcggctctaacctgctggagtttgcgcggcacgtccttgggcaacgccaggctgcagtcccattgctgagg
		gggtacgaccagggatacatcctgaagaagaaaggcagcagcccgtccagcccatgggttgtctccctcggccccgtcgccgtgcttgccttggtccactgcgcccttgc
		aggaatgggcggtccccgctcggtccaccggcttggacagcacctagaggcttatggcatggccgtggacaagcatgacattggcaggaacgacctgggccaccagttg
		cgaatgctcggcctagtgctagatagccccgatgccgaaagtggcatgctgctactcccccgttccccataaaccaagccagccagggcccggaacatgaatagacttg
		cacactggcttgccgccactgtccacgagaaagtcaggggctcgacacaagggttcggaggtaccagcctagaatatcggcttatcttccgcggcccacccctcgagcta
		ctcgaaccggcctacgacgagctggcccgcaacggagggatccaggtgccaagcggggcagacggaggactggtgaccctgccggtactgctccagtatccagccgg
		ccagctgcagggacccaggccacgcatcggagcatccggtaagtgtgacaacgaccacttgcttgatatacgcaacgaccctgccaaccctagctttattgccctggtccc
		gccgggactgcacaacaacctctcgatcgagtcaaccaccgacgaattcggattgggggcagccaccagcacggggcatgcatccttcgaacaatggtgggaggatgg
		ctttgtccagcaagcagtcaacgaggcgttgatcgctgccggcataacggacgcccagagggatgacgccaggggcctggtccgcgcaaccgcagcctcggtcgacga
		ggtggatccagacaagggaggtcatcgcgcggcctggcgcctactctcgcgcatctactcgatagcaaacgtgaatcaagggttgcctgcaggaacagcgctatcactgg
		catgtggtcttcccccaatgaaggagggaggaatttccgccaagactcagctttcggtcctgggaaaaatcgccgacgagcttgcggacggtttcaagactggcatcgagc
		gcctggcacaaggcgtccaacaaggggttgcgcaagcgctgcgcgaactgctttcccatctccactcgaattgcgacgtacctacggccttcgagcgtgccacagcggctt
		tctacctgcccagtgccgatattgaactggcgcctcctccatcctggtggaccacgctcaccaccgagcagtggacggaactacttgccgacgagcctgacgaggtcgtcg
		gcgagctaacgatccggtgtaccaatagtttgatccctatggggaaaggcttgccggccgtagtacgggacaaagtcgagctattgatttccacaagcgaagagagccaac
		caaaggagctcctgttgacaggcggatcctacggcaaggttccgacgtcattgccagcgggccctaatgggactaccagccacattgacctatttccctcctcccacaaagc
		gccaatgagctacaaggtttccgcggacggctgcaagcctgcgagcgtccgggtcatctccctcgcgagctggaagcccggaatactcgttacctgcaggcttgcgacaa
		agctctcgccaccgaggaagccccgcaagaactcagctgcgatggactgggaaacatccctgtcgctgccgggctccggtcgttatgagctccagctccaccttgctccg
		ggggcgagcattggaaaggtagaaggcttgccggacgatgccaccgaattcgaggagcagcgggagacaatcgaaccacggcaagttggggaatacgagtatctaata
		gaggtcgaggctgatggcaagtaccagctggacatcgcctttactgaagccggcgagcaagttccgaaggtctgccgggtatacctgacctgcgaagaggcaaaggagg
		aaggttgcaggagcgaattcgagcggctcatcaagctcaaccgacggcatctcgagaagttcgataccaaggctgttgtccatcttgaccggaacgcacgctcctccagcc

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		tgcagtcgtgggtgctggaggatcagaacgtatccaattccttcaggccactggtgatcgcggacgactatgcgtcccggtgggcccctcctgactgggacgccccgcacggccctgtactctcgaacgggcgtttccttcatgaccccgccccgaggccacgagcttccaacctcccaagggcttcatcgaggctcggcaggggatcgcccggtacatacgtggtagcgacgaccaatcggggctccttgagtcagcgccgcttggtgcctggctatccgaagaccctgggttccgctcccttgtcgaggactaccttggagcgttcatgtcttggctggacgccgacccgggtatcgcctgctggatcgacaccattgccgtctgctccctggagccggatggtcgtaccctgggaaggatcccagacgccatcatcctttcccccctgcacccattgcgcctcgcatggcactgcttcgcccagaaagtactccgtgacgaggccgagggcgaagccccgtgcccggcagcaagcatcctcgatccggactgcgtccccgatctactgaccatctcgctgcaggcaccgggaggagtggatcaggtcgacttcctttccgtcgaatgcagctccgactactggtccgtgctttggaacggatcccggctgggacaaatacccgatcgcgctcgccgggccccgttcgacagtagcttcgggctggcagttggagggatatcgagcgggttcagccccgcccaggtctcacgagcactcgacgacgtcaccgacctcctggcagccaagcctatcgtcagcctggtagtgtccagcgcaggtggcaccacggatgcatgcaacgaagggttggccacctggtgcaccaagcgattcggcaacggggaccatgacaccccgcggcacggtgtcgggccaaggattgtggaggtattcgataccaggcaggctggccggcccgaccaggcgacgatcgccaacctctccgaggacacaggcaaccacgtccgctggtatgacaagcaaccaactgggtccaagccagacctgggcatcattgcccaactagattcggcccaacccgaatccaaggaggtcggaatgctttcgccgatgggaaccggcggactgatcaggcaccgcgtcaggcgccaactccaagcctccttcctaagtgaatcccggcagggcctgcagatgccaccctccggcgaaccgttcgcagataaggtttccgcatgcatgctcatgatggaaaggctcagggacggcaaggtcggcctgcagttctcccctaatgtccatgcagtgtccagcatgctcgaggaaaacagcgctgggttcgtcgctgtatcgtcgtcagcaatcgaccccgcctgcttcctcggaggctggatacaagggacgtatctatgggactacgacctcccctcgtactcgcatcgcgcaggcgacacaagcggctactacctgttatcacaggtcaagcaggctgatcgcgatgcgctacggcgagtcttgaagccccttccgggatgcgaggatctggacgatgatcaggtcgagcaaatcctcctcgaggttgcgcggaggggggattcctacggtgcgaggcctctccggggacgatacgggggcgacgggcgaccttggcctgttcctcgctgtccggctcctacaggatcagttccgtgtgacaggcaacaaggaaagcctgctgccggtgcttgccggatcaccggaggactcgacgatagcaataatcatccccgtcgaccccttccggggttacctttccgatcttgcccgctcccttggcaaggagcgcaaggatacctccctgtcgcgtcccgatctgctggtagtgggcgtgcgcgcatgcagcgacaagatccacctgcaccttacgcccatagaggtcaagtgcaggcaaggagtagtcttcggtgcaggcgaatcaaccgaggcactctcccaagccaaggccctgtcgtcattgcttcgtgccatcgaggaacgtgcaggtagttctctggcatggcgccttgccttccagcacctgttgctctcaatggttggctttggcctgcgagtctacagccagcatcaggcagtaggtgggcatgccggccgctgggctagctaccatgaacgtatcgctgcagccatactcagcccaacccgccgatcagcatcgatgagaaggggcggctgatcgtggtggacgcgtcgctccagagcagcccgcatgatcgcgatggcgacaagtacacagagaccattgtcatttccagccgagatgccggtcgtatcatcgttgggaatgacgcacagtccttctatgatggcgtacgtgcaaaggtcgacgactgggggctgctaccctgccaggcaagtgcggccggcaccccaatcgtgcagcccgacatcactcccccggacgatgtccagacgggcgaccccatagtagtcccagcagaagatatccccggggcatccaccagtctggtcgatcagacatctaccggcgtagcggaaccaggggcaagccctgcccccccaactgacgagccagggacagggatcattctctctgttggcaagactgtggatggtttcgagcctcgatcactatccctgaacatatccgacacccggctcaaccagttgaacattggtgtcgttggcgacctcgggacaggcaagacccagttcctcaaatcgttaatcctgcagatatccagggcccgcgaggccaaccgcggaatcacgccaaggttcctgatcttcgactacaagcgcgactacagcagccaggactttgtcgaggccacgggcgccaaggtggtgaaaccctatcgcctgcccctgaatctcttcgacaccacggggatgggggagtcctccgcaccatggctggacaggtttcgcttcttcgccgacgtactcgacaaggtgtattccggcatcggccccgtgcagcgggacaaacttaagggtgcagtccgcagcgcctacgaggtggctggtgggcaaggccgccagccaacgatctacgatatccatgccgagtaccgagagctgctcgcagggaagtcggactcgccgatggctatcatcgacgacctagtggacatggaggtcttcgcgcgctcaggggaaacgaagccgttcgacgagttcctggatggagtcgtggtgatatccctcgattccatggggcaggacgacaggagcaagaacctgctcgtcgccatcatgctgaatatgttctacgagaacatgctacgcacgccgaagcgccccttccttggcacgtccccacagctccgggccatcgactcgtacctattggtggacgaagcggacaacatcatgcgctatgagttcgacgtgctccgcaagttgctactgcagggccgcgagttcgggacgggcgtcatccttgcctcgcagtacctgcggcatttcaaggcaggggcaaccgactaccgggaaccattgctgacctggttcatccacaaggtacccaacgcaacacccgcggagcttggagtactcggcttcacctcggacctggcagagctatcagagcgagtgaagacccttcccaaccaccactgtctctacaagtcattcgacgtggctggagaggtcatacggggactgcctttcttcgaactcaccaaccaagcctgaccaacgcccggcctgcgaatacaggccgggcaaggaggctcctaatgacagacttcctttctcccgcagaacgctcggacaggatgtcacgtatccggggcaaggacacgcagcccgagctagcattacgcaaggtccttcaccggctcggactccgataccgattgcatggcgcggggctactaggcaagccagatctcgtgttcccgcgatacaggaccgtggtattcgtgcatgggtgcttctggcataggcacaagggatgcaatatcgccacgatccctaagagcaacacacccttttggctggagaaattcgaaaagaatgtcgtacgtgacgcgcgagtagcaacagatttgcaggccttgggatggacggtacttgtcgtatgggagtgtgaactgacatctgccaaaaaagcccagaagactggcgaacgcctatatgaggttatccgtagtcgtagccacggaaagtatcggtaatcgactgaagcagccctgcggcctgtagtggtctactgatcccggacaccgatttaggcgaaaatcctcgccgtgagagaggtgtccg (SEQ ID NO: 408)
44	Dcm + HerA + Vsr	cgaacggagcaggtagatccgcgctaactgacttgcccaatctggctgcattcgtccaacgctaggcggcttcgcaggaaaagcgaaacggagggagattctacgcgcacctttgtgcagacctgaggctccaccagacctgagagcccggcacgattgactgatcataggagtaaggccaagaagcgacttgatgcgcttgtaaggtaaattctcagcgaatcgaagtaatgacaccgaaacacgtgcggtcgacaaccgtgtaagattgctgataaaaagagcaggacgtcacaagaaatgaacttggaagtagtgccggcgagccggactttcatcgacctcttctcgggatgcggaggtttgtcgctgggactttgccaggctggatggaaaggactcttcgccatcgagaaggccacggatgcgttcgagactttccgggagaacttccttggtgagaactcccgctttgcctttgattggcccagctggttggagcagcgcgcacactccatcgatgacgttttggcactgcgcggtctacatttgtcgaaaatgcggggtgaagtcgacctcatcgcgggtggtccgccatgtcaaggattctcgttcgcgggcaagcgaaacgcgaaggatccccgtaaccagctctcccagcggtacgtcgatttcgtcgagcgactccagccgaagtccctagttctggagaacgttcccggcatgaacgtcgcccataagtatgagcacgggaagagtcgcaagacttactacga

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		aaagcttctgcattcgctttcaatagccggctacgtggtgtcggggcgtgtcttggacgcggctgacttcggcgtcccgcagcgccgcactcgactaattgccgttgggattc
		ggtcggatatcgcggataagcttgcatgcgcggctagctcgactcccgcagacgtgctcgagggcatcttcgatgcaatcaatcaggcaggcaagcgtcagctcgtccgat
		atggccagggcgcccatgtcacggttcgggacgcgatctctgatctcgcgattgggccggccgatcacgagaacaccgaagactacgtgggaagcgagcgatgtgcag
		gctacaggcaggtcaggtaccaggggccgaacacgccttaccagatcgccatggcttctggggtcaccccatccgaaatggacagcatgcgacttgcccgtcatcgtcct
		gatgtagaaaagcgcttcaaggcgatccttgaaacttgcccgcgaggggtcaacttgagcgccgagttgagggcgcagcatagaatgctgaagcataggacggtgccga
		tgcatcccgaaaagccggcgccaaccctgactaccctgccggatgacgtcctgcactaccgagacccgaggatcctgacggtccgggagtacgcccgaattcagtctttc
		ccggactggttccgtttcaagggcaaatacaccacgggcggggcgtcccgtcgtcatgagtgcccgcggtacacgcaggttggcaatgcggtcccgccgctgctcgggc
		aggccattggctcaggattaatggcgtgcctctctttgagttcaacgcgagtgataagggccagtgcgcccagtctcgcgatggccgagaaaaaggcttttgccgtatagca
		attagtcagctgcaagaatcgaacaggtggatagacgatgacgaaataccccgatggattgcttgattggtcgggcaatcgggctggaggagtcaagaaactcttctacgg
		cggcagcggccgccccgtcgggaaggtgatagagactcctctactcacccgtctctgggaatggtcggatagcgtcgtccagttcgagccgggcattccgcgggcggtgt
		tgctgttgggagggccgggaaacggcaagacagaggcaattgagcagacgcttcgccgaattgactcaaggcttgcgctgagcggagcgctcatcgacaagcttgcgg
		ctgtcttcgagtccaaggatggagtcccccaggacgccttgtggaggtggatcttggggcgctticagggggggcgctcgagcgggacaatctcgattgtccaagacgcct
		cggaggggaatccgggctctcctgatcttccggcgcaattgctctgcaacgacctagcaggactcgtcgaagacaacgtgtcaaagcgcatctatttagcgtgcataaatcg
		cggcgtcctagatgatgccctgatacttgcgacggaaagaggtgacacagaaattggtgctttgctgaagcaaatcatccggtcggtgtcgatggcggcccatggcgtctca
		tgctggcctctgcagggatatccgggcatcgcagtctggccaatggatgtggagaccttggtcgcaggcgtccagggtcaaccttcacccgcggagcaggttcttcatattg
		cggccaatgccgaccattggcctgatttcggggcatgcgaagcgggtcagtattgcccgttttgcacaagtcgcaggctcctttccggcgagccccatgcgggatctctcgc
		caagctgctccgatggtatgagctggcgagcggaaagcgctggaacttcagggacctgttttcccttgtcgcccacctgttggctggaacccctagcaatgccgatgcgtcc
		ggttattcgccctgcaaatgggcggcaaaacaactgaatcccccggcggcgacccgcgcaaggccgatgtactccgaaagcgcggagtctttcggttgctggcttccca
		ataccaacacgcgctctttggcgactggccaatcgagcatgcgtcgggtctccgaagagacatcgccgacctagggcttggtgatttcccggcgcttgtggctatccagca
		gttcctggcgctggataagcggcgggagtcgacggcaaccctccgtgcccagctctccggcatgtcatccgtattggatccagcaaaggcaagccccaccttcgaggtta
		gggtaagcgctaatactgttattcgttacgaagacttggataggcggttcagcctgtccatccaaggaggcagagagtacctccaagaatatcagtgcctctcggagatcga
		gatttcagcactcaaggtccttgaggaggccgacaataagttgtctgatcacttagtcaggcgatctcggccggcgacagcaattcgagtccaggcgcttctgagggccatc
		gcgtgcaggctggcaaggaggtcgattggcgtcaggtgttgtgtcacaaaggatgccgacgtcctcgaggagttccaccgcgtcaccaatggcgattcgtcggcgctgca
		gcaggcgatcaggcaggtcgaggcacttctcaacgtcaatcgccggttcgttgtttgtctcaacaacacctttggtgagccgctgcctcccccagagcggcgcgcgatgctt
		accacggacattcagcgcgttaagccggtgcccgccttggagggtgttgagcggccgagatcgccgatgcccttcctgagggtcggcgcacaaggcaacgccaggccc
		atagccctgaccttcgatctcttcaaggcgacgaaatcccttaggcgtggcatggtcgcgtcgtcacttccgaggtcggtggtcgcgcttctcgatacgacccgagctggtct
		tgcgggagcgatcgtgcgagacgaagacgctctggaaggtgcggagatccggatcggaatcagggatgaggtcatagtgcggacctttggaagtttcgtcatccgccag
		gagggtgcttgatgtccatgcaggagtttctcgcttcaccatggaagaaagaagcctcgcaccgagccttcaacgaatcctcttttggtatgaggtctgccccggagttcgca
		actggcgaggtcgtcctgtcttcgctctaccgcgccgtcggctttgacggggtttccgaggagaaagtgccctcgcttggcaatgatttcaggaaggcgctggacaaggaa
		cgcagaaagcagaacgcagctggtggtctgagcccagaagcctggcgcacggtcgtggatcgtgtcgtgcaaagtcctaaggttgcgcagcaatcctccaagcgattcct
		atcgctgtccccggtcgttcccgacgcggccatctactcgggcgccgcgcgccttggaggaaactcctggaacccggggcggctgatcaagcaaatggtcggaatcggg
		tcggagaccatggagggcgcggaaacgctttggggcgaactctacgatgctttgtccgtgacggaagcggatgatgtctgggcaagatggctccaaacagaatttagtccc
		aggcgcccagagcaaatagcgtgggccccaagaccgatggatcaaccagatttgcttccgcaatccgatagacggggagtttcctatcccgctcggcagttcgtggtgga
		cctgcgaggaatcttggatgcgaagtccgccatgacgcggcggcagtggatcacactgctcgaggcgctacttcgaattggatcggtcagccatgtgctgtggctgtgcga
		cgtcaatgaccgcttgtggcgtgcgatgcgtgcggcgctcgagggcgaggcgagtggcgtgcccgccgatgccgccgccataagaaccgacattctggccgtcaggcg
		gcggacgctctcgttcgggaatcccgctgtcccagcgattcgggacctggcctctcgatacctatccgcacgcctgggaatcaactgtgtcctttggacgctggacgaactt
		ggcgtgggctcaagtcgactttgttcgtccgaagaaatccttgacttcatcaagagcgttcaggccaacgcagggggggctcaaggcccgtggcgtcatggatgccttccatt
		ccctgcaagacaaggaagtcaggaccattggctgtaagaaaggagtcggagcaaaccttctggaattcagccagtacacgcttggacagaggcagacgatggaccagg
		cactccgcgggtacgaccagagctatttcctcaggaagaacggggatgccaggaacgcgccatgggttctatctctagggcccgctgccgtacttgcgatggtccactcgt
		gcctacatgcggtggatggaccgcgatcgatacaaaggctttcatcccatctcgggagctacggcatcgagtttgatctccacggcgtcaacgatagcgtccttggaaagca
		actccgaatgctcggactcgtactggatagcccggatgccgagagcggtatgctccttgtgcccccgttcgtagcctgaggaaggaggcaatgatgagcacgctagccaa
		gggaattgcaagctgggtcgaaaaagccatggcgcgtgagatcgcgacgctggtggccgggaatatggagtgtcgcgcagtcttctgcggcccgccaaagcacatcctg
		aatcaagtatttgggcatcttatccacggtcgatcgctgatcgaagcgacaagggccgatggtcaggcggttcagtatcccgtgatccttcaggtcgaccgcctccctacag
		ggtttcccatcggctccgccacacagtcgggatgccttcagttccatggactcgctgccgtcaggaacgacaggaatggtgttttcctagttcttgtcgagcccggtgctcaa
		gcgagcgatacgcatgaatcaactcgaacttcgcttggactcgagccatcggtaaacgagggcggtgcctcgatcattgcctggtggtctgatccattcattcagtcgcttgtt
		gattctgccctctcagaactctccggtcgcgacgccgcggctgccaaggatctactaaaggaggcgatgatcgccgccgacgcggcagatcagcacgaagtagcgaga
		gttggagcctggcgcgtcatcgaacggttgtgggagctaaaagaacgcggcttgtctcttgaccaactcgttagcttggccgccggattcccgccctctagcgacggaagta
		ttgaaccgagatccaagaccgccatcctttcagccatcgtggacaggatcgaagccgagaacttcggtggcttactgtcgtcccttctgcaaaaagccagggacgatatcga

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		aaaagaacacatcaccgcgtgcctctcgaatatgaggggcaggtgcgatgtggttactgcggttcggcgatgtgcgccatatgcgtacatgccttcggacgccatcgctgg
		cgaagtctggtggaagtcgctcactgtcgagcgctgggaagagttgctcgatgatggcgctctacccgatgcgggcggcgacatcattattcagtgtgccaatccgatgattt
		cgcaccttaagggcatggttcccgtcgtcaagggatccgtgcaacttaggatcgaggttccagagaagtacgtgggcaggcggttggaggttatccgcgaggtcccgggt
		gcgaaggcggcgacgaaggtttggacagttgacgcggaacgcatgatccacgtcgaggacgacgagatcccccccacaagagtccgatgaagtactcggcaagcctc
		gaaggatcagccggaaagaaggcgagcgttcgaattgtctcaatggatggctggctccctggggtggttgcctctgcgacgacggcgacaaaaggttccctcccgaaac
		gctcaaaagcagcgaagttagaggcgtcgctgtctctctccgggcaggggaggcactaccttgacatctacttaaggccgggcgtcgagctcgcgtcaatgctcgccacc
		ggtagtgacgaggaaggaaatccagacccgtccatcacggcgccaatcggcatggtcgcggagggcgagttcggggtcgaaatcgaaatcgaaggggaatgcttcttc
		gacatcacgctcagggttccggaggttgcggatgatcaggtcatccggatcgaattgtcggcggagcaatcaagcccggaagagtgctcaagccacttcgaattgcagctc
		cttaagaactctagcggtcggaagcccagcgcggtccacgttaatgctcagctaagaagtgcgcagcttcaaggttggatgctggagcaggggcgcgctggtcgctcctat
		tatcccttcgttatggccgcggactatgccgccgactggcacaggcgggactggactggcgcagatgacacgatcttctcgaaggctagcttcctgtgcgatccccggccc
		tcgccggaagaaatggcgccgccgcaggctttcatagatgccagagccgcactggccgccaggatcaggggtggtgacggaaatggcttggtcgaaggtgtgccgctc
		ggtgagtggatggcaacggatcccgatttcgccggggaaatagacgtctacttgaaatcctacatgcactggcttgcgagcgatccagatggggcggtttggtgtgacgta
		gggttggtcgcgcggctcgagcctaacggacttaccttggtgcaagagccggatgcggtgatagttagcccgatgcatccggtaagacttgcttggcactgtgtggcccag
		cgagccatgttccttgccgcacgaaagagaccttgtccagccgccagcatcctcgatccggattgtgtgcccgatgcgatcactctcccactgagaaacgccatgggtggc
		aagaccaacgccactttttctcggtcgaatgcagttcggactactggtcgattctttggaacgcggggcgcttggaagccctttcttcacatggggcgacagccccgcttgac
		cgggagtttggcctactcgtcggcggaatctccggtgggtttagtgtttcgcaggtgcacaaagcgctcgaggacatctgttcgatgctggtggcgaagccggtcgtcggc
		gtcctggtgtccagtaccgcgagccagaacaatgcgtgcaatgaaggtctgctttcctggggcaggaagtacttcggcggcggggatagggcggcaggcttggacgcct
		gggtcggggccagcgaggtcaggatctacgacgacagaccggaagatgcccggcctgatgatgcggagatttcaaatctggccgaggatacggcgaacgccgtgcact
		ggtattccggcacggtggccggcgaggctcccgatctagcgatcatcgcccagcttgagacctccaatcccggtgcactcccaaccaaactaaattctccgttgggcttcgg
		tgggctcgtgaggacccgaattcgggagccttccagcatggcggggggtcaactgctccgtgagtcgcgcatgtctggtcccgcggcgcccactggcgacgggctggc
		cgacgctgtagcaagtgccatctcgtcgctcgagaacatctcggagcaacgccttggttacgtattcgcccctagcattcatgtgatcaaggggcgctggagagcgcgga
		atttgccgcagtttcctcttcgagcgttgacccggcctgctttctcggaagttggttggagggcacctatctttgggactacgagctcccgtcgtactcaggtcgtgccggaga
		cagcaatggctactacttgttgtcacggatcaaggatctcgacctcgaaaccctgagaagcgtggtcaagaggttccccggttgcgaggagatgccggaagccgtgcttgc
		tggaatagtcgaggaggtcgcacggcgtggtattccaaccgtcaggggcctcgccgcaggtgattctggcgcgacgggtgatttggggctactcgtggccacgaggctg
		cttcaggatagcttccgggcggccgaatcaggcgctggtctcctgacgccttggcgcagggagggagacatcgaagagcttgctctcgtcattccggtggatccattccag
		ggctatcttgacgatctcgcgaaggcgctaaagcgccctacgctccaccgcccagacctattggtcgcgacggtgcgaatcagtgacctgggagttcaggtccgactgact
		cccatcgaggtcaagaaccggggtgctggagcggcgatgccgcaatccgatcgagaagccgcgcttgcccaggcacgctcgctggcatccctgctagatgcaatgctg
		gcaacgtattctgaggatcaagagatggttctctggcggattgcgcaccagaacctcttgacctcgatgatcgggtacgcattccgtgtttacagccaacgtctggcagccca
		aggcaagtcgggagactggtcgcgcctgcacgcacgagtcatggaagcaatcctgagctcccaggccgatgtgcgggtggattcgagaggccgcctgatcgtgatcgat
		ggctctagccaaagtggtccgagggatacagatggagatggtttccacgagactatcgagctctcgcacaaggatgctgcgcttttcatccgtggcgagcacgatgcgctct
		gcacggccatgaagcagaagctaggtggctgggaaatgttccctgaagggagggatgccggactctccaatcaatcgccgcccgtggcccatgagactgcgcccttggt
		ggatggcggcgttgaggtgccgtcccttcacgcgctccaagcaacggcggggcccgagggcagctcgctgccgtcttcgggagtcgaagccatgggcgcgtcgcagc
		cggcctccccgggagccatcgacgtggatggcggcatggcccagtccgggctgatcattcgggtcggtgaaacgatcgatgggtttgagagccaaattcggcggctgaa
		tcttggcaacacggccctgaaccaaatgaacatgggagtcgtcggcgatctggggaccggtaagacgcagctgctccagtctctggtttaccagatagccaaggggaaag
		atggaaatagaggtattgagccgagcgtcctcatcttcgactacaaaaaggattactcttcgaaggagttcgttgatgcggtagctgccagggtcattagccctcatcaccttc
		ctctcaacttgttcgatgtttcaactgcatcgcagtccatcaatccaaagctcgagcgctacaagttcttctccgacgttctggacaagatctattcagggatcgggccgaagca
		gcgagaccgccttaagaactccgtcaaggacgcatatgtgcaagccgccgaagggcagtatccaacgatttacgacgtccatcgaaattacgtagaagcacttgatggag
		gcgcggactccctgtcgggaatcctaggcgacctcgtagacatggagctcttcacgccggatccaagtgtcgttgtttcgtcggccgaattcctgcgcggagtggtcgtgat
		atcgctaaatgaacttggttccgatgaccggaccaagaacatgctcgtggccatcatgctcaacgtcttctacgagcacatgctgcggatacagaagcggcctttccttgggg
		agaaccgcaatatgcgtgttgtcgactccatgctgctcgttgacgaggccgacaacatcatgaagtatgaattcgacgtcctgcgtcgggtcctcctgcagggacgtgagttt
		ggcgtcggggtgatcctcgcttcgcagtacttgagtcacttcaaggcaggtgcgacggactaccgggagcctttgctttcctggttcatacacaaggtcccgaacgttcgtcc
		gcaggagctttcggcgcttggctttagtgatgcggtgggattgccgcaattggcggagcgtatccgtagccttggcgtccatgaatgtctctacaagactcatgacgtgcaag
		gtgagttcgtccgcggcgcgcccttctacagacggggtgagtgggccaaggaatgacttttcgtcgtgtcgatttatcgcctagttacgcttttggtcttaagttgcgttcctaag
		agaggtgggctgtgtccgacaatgcgtattacgtttatgcgctgaaagatccacggatggcgcccgcccagccgttctacataggtaaaggaaccgggacgcgctcccatg
		accatcttgtaaggccagacgattcaaagaagggaagcaagatctccgagatcatggcctcagggcgtcaggtgctggtaacccggctcgtggacgggctcacagaaga
		gcaagcgttgagaattgaggccgagcttattgccgcttttggcaccctcgatactgggggatgctcctgaattccgttctgccaagcgggttggtaaacaagagccgtagct
		cgctggttgtcccgtctggcgtaagggagaaggctcagattggtctggcccttctaaaggacgccgttctggagctggccaaggcgaatccgactggtatctcgaactccg
		atgctgcgagcatgctcggcctgcgtagcgactacggcggaggatcgaaggactatctgtcgtacagcctcctcgggctgctcatgcgggagggaaagctcgctcgggtt

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		gccggcactaagcggcacgttgctcaagtgagctagctgtggggttccggatcgggctggcccgctcggcgctgcgctacgaagctcgcttgcctgccaaggatgctgc ggtcatcgaacgcatgaagcactacgccgcgctgtatccgcggttttgctatcgccggatccatatctatctggagcgcgagggcttccatctcggctgggaccggatgtt (SEQ ID NO: 409)
45	RecQ	atttgcctgagacttatttcccgtggcgcttagctagctaagagtgggcatcgtgagcaccattgatgatatgaaatgacggtatagcaatttaaccgtctggatttcaccagaa attagtgattcaataggaaattaaatacgttttatatttcaatgtgtatcaaaatcattcctgaaatttcctggtgctatatttgatgaaaacggataaacattctgttgattttaataaaa ttctgtctttcgatttagagcttacgcgtgatgaaaagttaaggcatatgggggccgtgctggcggaacgcacgttgagtttgaagataaatcaggatgaagcgattcatcaatt ggatgaaatggcaggcgatgcagatttaatcctcggtcataacatactggatcatgatttaccctggattgccaaacaacgcgtacgtgctcaaatattattagataaaccaatc attgataccctttatttatcaccgctagcttttcccgcaaatccataccatcggctgattaaagactataaactggtaagagatagcattaacgatccagtgaatgacgctaaatta tcgcttcaggtattcaccgagcaaatatgtgcgctgcaagaaaagccgctggctcagttgcagctatatcagtatctttttgagcacggcgttgccagccatttcagtacacgtg ggatggccagcatttttccgcactgacgggtcaggcgtccatatccgccgtagttttacctacgctagttaaatcggttgctcagaataaagcatgccctaaccagcttaatcg ggttattggcgatgctcttaaacagcctttgcgcttactaccattggcttttgcctgtgcctggctccccgtatcgggagggaattctgttttaccgccctggatatggcgccgtttt cccgtcaccgctgatatcatccgcgaactgcgtgagcaaaaatgccagtctgaaacttgccgctactgctgtgaaaaccatgatgctcgtcggcatttacagaaaattttcgag ctgaacgattttcgtaaacttcctgatggctcgccgttacagcgcaatatcgttgagtacggattagctagtcgttcactgcttgggatattaccgactagcggagggaagtcttt atgttatcaacttcctgcgattgtcaggaatctgcgaaatggttctttaaccattgttatttcgcctttacaagcgctgatgaaagatcaagtggataatttacgtcataaggcaggt attaaaggcgttgaggccatttcagggatgctaactttacctgagcgcggcgctattcttgagcaggtccgtaagggggatattgcgattctttacctctctcctgagcaattac gtaaccgcgcggtaaaacaagctatcaagcaacgtcagattagtggatgggttttgatgaggctcactgtttatcaaagtggggccatgattttcgtcctgactatctgtattgt ggcaaggttattgaatctttggcgcaggagcagtctgtgcagattcctccggtattttgctataccgcaacggcgaagttggatgtgattaatgatatttgtcggtattttgacaaa aaattatcgcacccattagctcgtttttcaggggagtagaaagaattaatcttcactatgaaatcattgcaagtaatggcttgagcaaaattagtcagattttgaatttgctcgata aattttttttctaatgatgatgaaggtgcatgcattatctattgcgcgacccgccgttcggtagatgaaatcagcgatgtgttgacccaacagcaacctttaccggttgctcgtttta tgcccggcttgaaaatagtgaaaagaaagaaatccttgaagggtttattgctaaccgttatcgagttatttgtgctactaatgcctttggcatgggaatagacaaagaaaatgtac gtttagtaatacatgcggagatccccggttctctggaaaattatctccaggaggcagggcgtgctgggcgggatacgctggacgcgcattgtgtgctattatttgatgagcag gacattgaaaaacagtttcgccttcaggctattagtgaagtaagctttaaagatatttatgcaatatttaagggaatcaaaaagaaagttaatgaaaataatgaagtcgttgccac aagtattgagctaattaatcatcctatggttaaaaccagtttctctatcgatgataacaatgcggatactaaagttaaaacggggatagcgtggctggaacgtgttggttatgtgg agcgacttgataatataactcaggtttttcagggaaaagtggcctttccttctctggaagaagcgcaaagtaagatggcagcgctgcacttgaatcctgcggcgatggttctct ggaatgctgttttacaggcgctattaaatgctaatgacgatgacggacttagtgccgacagcattgctgatgaggttgcccaatttcttccgcataaagaaaataatacgtcagg aattgaagcaaaagatgttatgcgcgtattgacacagatggctgatgttggcctggtcaccaggggaatgctgctgaccgtacgtatgcgccccaaagggaaagataatgc gaggatcacaactgagttaattcacaatattgaaatcgccatgttagggctgctgcgcgaagctcatcctgatattgaactggggatgccatggcctctccagattgcggttat gaatcaagagattattcagcaaggctatgatagaagtaataccacgttactacaaaatatattatttagctggtctcaggatgctcgagcaaacggtcataaagggcttattgatt ttcgttatggtacaaggaacagctaccagattattatgtatcgtgactgggcatatatcgaaagagccattttacaacgtcatcgtgtgacaagctccgtactgaattttatttatca attggcattggatagtgatgaaagcagtatcaaaaaagtgatgctttctttctcactggaacaggttatcgattatttaagaaaagatgttgatattattccaatgatccaacagag acagggggggatgagcagcagtggctgatggctggtgcagaacgtgctctactttatcttcatgaacaacatgccattgtgctgcaaaatgggctggctgttttccggaca gcgatgagcttgaaattgcaggctgaaaaatcgcaacggtatgtcaaagctgattatgaaccactggctctccattatcagcaaaagacgcttcagatccatgtgatgaatga atacgccaggcttggtcttgaaaaacctaactatgcccaacggctcgtacaggattactttgctatggatgccgagtcatttgttccactttattttaaagggcggcgaaaaattct cgatctggcaaccagcgaaagctcatggaaacgcattgttgaaaatttgcataatcccgatcaggagcaaattgtgcaggcgagccttgaacaaaatacgttagttcttgccg gaccaggctcagggaaaagtaaagttattatccatcgatgcgcctatcttttacgcgtgaagcaggtcgacccgcgtaaaatcctgttgctctgctataaccgtaacgcagcg atttccttaagacgcagattgaagtcgttgcttggtaaagatggcgccagcataatggtacaaaccttccacggattagcattgagccttacgggataccagattgagcggaa agataatgacgaaatcgattttgataacctgctctggaaagcaatagctttactcaaaggcgatgaaacgcagctcgggttagaagttgaagaacaacgtgaatacctcctcg gcgggcttgagtatttactagtggatgaatatcaggatattgatgagccacagtatcagctgattgccgcgctggcaggtaaaaatgaaagtgaagatgatgctcgtcttaatct catggcggtgggtgatgacgatcaatctatttatggtttccgtgatgccagcgtgcgatttattcgtttgtttgaaagcgattactccgcccgtactcatttttttaacgtggaattacc gctctacggccaatattattgcatgttcaaattatcttatcagtcataatcaggggagaatgaaatgcgagcatccgatcgtaatcgatcgcgctcgccagatgcttccgccagg cggagagtggagcgcacttgaaccttcggaaggcaaagttgttatccagcattgtaccggcgcggctcagcaggcggcagaagtcgtgcgccaaattcagtatattcaacg gctgcagccggaatgccctcttgagaaaattgcggttattgcacgcaatgggctcgacaaaaaggagcttatttgggtccgttcagcccttgcggatgcaggtattccttgcc gctttgcgctggagaaagattatggtttccccattcgccactgtcgggagatcgccaattatctgctatggctacgagaaagagcgctcgagtcgctgacgccagcagagct gtgtcagcaactaccggggcgagaccaggcgaaccgttggcacgatattatttatgaattaattgagcaatgggagctaagccagggaggcgagccattacctgccgctta ttttgaacatttcatactggaatatttacatgcccagcacagccaggttcgctttggcctgggggttttgctgagcaccgtacatggcgtaaaaggtgaagagtttgagcatgtc attatattagatggaggttggcgtagttcgcactctctgcaacctgaaaataacgaagaagaacgaaggctcttttatgttggcatgacgcgagcgatatcccgacttgttattat gcatgatgatcgtgcgccaaatccctatatcgaacagttagatccagcggtcatcagccatactgctgcacaagccgttgcgcctgggatcttacgtcgtttctcgatcatcgg

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		attgcgccagctctatatcagttttgcaggtggacatccggctggtcatcccattcattcgttacttaccgatatgcaggttggggatagcgtccaactggtctctgtcgggaataccatcaaggtgaatgctaatcaatcggcaattgcgcagctttcaagtgccggaaagagccagtggcaattttctctttccgggatccgcaaaattgaagtgcttgccatgctacagcgcagcaaaacactaacagcagaggattatcaagttgcggtgaaagtggacaattggtatgtaccgatattattggttgaaacccgtgaagaagccgcttatgacaatattacttgaagcagaatac (SEQ ID NO: 410)
46	Histidine kinase + phosphoribosyltransferase	aactcacccgctctgaacgagccccttgaaacacaagacaccgtttttcccttaccataagggataggcaaacgactgtgtttatgactaccagcagagacaaaaccatcgaagtgctcggccacccatttgcgcctctaggttgctacgagactgcagaggatccatgtagcagattacctcggccatgaagctgctaacggaagcgaagccatagaccgtaggcgatacacgtacgtatggctttccggaagggcgatcctagtcaactgtctgatgtccgccaaatctttctcaatactggtcattcaccttttccttgaccggctgtcaggcccaacgtgcattcagatcgtcgcctaaatttgttgcatcacgtagagtctgccgcgtgctcgcccctatgccagactagtctgatgtggcggatgagataggtcacgacggtggtggctcggtagagtcggcatcgccgagtcaacgatggaacgtaaggggcgtgaatgcaaatcagccgtaagctcaacctttatgagatcgaggatctctaccagtcgcttggtacggattccaatctcaggcttcctatcagcatgagccacggcgggggttgggcgtggatgcttcgctggcccagttcatcgtcacctgggcacgtgcttgcgaaaaaaccgtccttcacctatatgccccgctggcgacgacgccatgacgcaaatcacgcagttggcgcagagtgcttctgggttcttcgcgctgatcatgtgcagtgaagtccacgctcagaatcatcaactgatcgatcggcgggaagcgcttctggcgatcaggccccttgtcgatgcgatgttcgcaggcgaccttcgtaacacctccaacatccgaggcgcccgtccaacggccatcaatctgttctgcgtgaacaacgcaaagcgtgagttcatcaagccgttttacttcgatcacgccgtgccgaaagtccagccgagatcttggttctcgactctcttggagacgtcatcgaagctgatgaatgctcgcagtggacaaggggcactgcttaggtcaggtctcccggcattgggcagcgtgctttgggagttgatctccaacgctgaccagcacgctgtcactgatgtaggcgggaacaagtacaagaaggcgctgcgtggcacctccatcaaactcaaccgaatgagtcgtcaggatgcgctgatgtattcagaccaagagccggagttggcgcgctttatcctgaagcatttcctgagagctgaggtactggacttcctggaagtctcggtcatcgacagcggtcctggactggcacggcggtggctgacggcgaaggaggggcggccagtagaaagcctggaggagctgagtcttgaggctgagcttgaggccacgctcgattgcttcaaaaagcacattacatccaagccgcagtctccgaactcgggtatggggctgcataacgctgttcaagcactcaacaagctcaaggcgttcgtacgcgttcggacgggtcggctttcactgcatcaggcttttcagggaagtgatgagattatggagttcgatccgtcgattcgatacggtggccgtgtgttggccgctgtggaaggcactgtcttcaccatctgcattccggtgagctgacatgttcgatctcatggatttgaagtcgagttgcgtcagtcaggtaagccggttcatgtggtggttttcttcactggccctgatctcctcacagacacgcaagcggctcacgctctacagcaccaattgtcgggttacgtcatgcctgacctagtggtgtttctgatgcctggttacaccttggatgaattccgagcacaccaggcaaatgctacatcgcccctgatggcggagctaagccgtaaaggcccaggctcgcctcgcacctacgcgagtgcgttctatgacgtgaatggtgccattaccgagtacgtcaatatctctggccctgaggagcagttcgaggaactcatcaagcacaactctaacgctatcgcgaggactggcctgacccacctcgtcgaacgctccaacgtgctgaagaaggcgcctgcaggcttcttctactcaaagccctcttctcgggcttcgaactatttcattcgggcggaagacctgctctctgagaccttgcatgcccactacctggcgtttgcatgcctatctctcatcagtaaggcaacggaagatgggatggggacgcccgataccctgtatctggacacaatcgcattgctgcctctggcgctgtccatgcaggtgtacctcatgcgatttgagcagccgggctttgcgaatatccggtcattccattcgcacgaaggcctaatcaagggtgggcctttgcccaaggcagtttccgccctgtgtctcatttccgcatcgacccagtgcggcctcgcgcagcaatgggtgaaggtaaacagtgctccgccgacgcgcgtggccaccattctttcatttgagcgctcatcggactcctgctccgtcttgcacacactgaagcagcccgaagactttgaaatgttgggggagggtgaagcgagcgggattcgtctaattcggatccatggcgagcggttcgttgctgagcacagtgaaaccaagctgctgaacatcggcactgatcatgcgccgcccctgctgcaatccaagttctactcgttcatgggggccaacctgttcagctgcttcacccatgaccggccaggactgaggcctcggacagtgcatgtcgataaagataacctggtggctgccagcgatttcggtgaatggttcgacagggtactgcttgaggaagctgtcgcgtcgacccgttggatcatccacgatgacgacgctgccagtgcggccctggccgatcgagcgatcgcttacttagggatgtgtggcgtcaaggtcggtaacaaggtctccttcgatgacttcgatgccaacacgaattttgacgggtctgtcatcgtcattgccgctgctgccgaacgtggctcacgcctgcagagtgtgagccgacgcctgcgtaccgctcagcaatcgggtaccaggctttacattacgggggcactcttcgggcgcagctatcaactgatgaaggatctgcagagcaacctgacgcaacctgccaaggatcacagccggtatgttttcaagacgtacatggagatcccggcagcggagcttgcctgcacgagtcattggccgaagagcagcggctgctcatctccttgcattcatttgcggaaactttctcgccagcgattacgcagcgcatggaagtatttgatcgcgcctctactgggggcttggtctgaacccattttggccgagcagtcacaccgggcagccgatgacacttagccgaggctttgcgtttgtcgacggtacgaaggatgtgaggggcgcgacgtcaacggatatttacctaaccatcttgtggattctgcagaatgcccggtacagcggtaaggtgcagaacgccaagcggcttgagtccggtgagcttcagcaggtgctcctatcgccggatgtgttctcgcgcttcgacgatggcgttatccaggccgcattcttgcgcgcagcggtgccggcggagcttgactacagggctcatgaaacccacagcctggccatatcggacatcattcagcgcatcgccgcagggtacggacatgaacgtggtgaagccgccatggagtttgtcatggccttggctatcgggaagatacgactgcacaaggatgtcgataaccggctgcggagtaacttgatcaatatcttgacgccgcacgttcaggagatccgttatctgctggatccgaattacgaatcaccgttgtgatcaatttccgctaacccgttgcatgcgaggtatccagttaccggcaactcagctcatggctgagctgaaccctggttgctcttctagtttcgatggcttgccgattgccgggatcacccacctgcgtcggttctgcgacgaaggtctaagggcagggtggtggcacctggcttgctcattccgtttgacctcgccaccat (SEQ ID NO: 411)
47	PH-TerB-DUF726 + TM	cgctcagtccggttggtggttttggttggtttggcgattgctcagatcgcacaatccgggctgagttccctttcagtgatctactattccgcgcagctatttagtggatataatcacgctttgaaaaaaaaacgggtcaattactcttcgccccacagcaacgaataaggagaaatttgtgagtaacgtcaacactttccttaaggaaaatttatcttcagtaagtaagaatgtttttgtggctcctggcatccctgaaaaaaaactgaataatgtcgctaaagcatttaatgttgtggataacttgaatactgtgctagccatttatgacaatacggtatttggtagcgcaaaagatggcatcgtttttaccggtgaaaaactggtcataaaagaagcttttgaaagtccttatgacttgttctacagcaatattgaagcagtagaatatatagaagatgtcacggtaaatgataaaggcaaggagaagcgaacagagtctgtttccctcaaactaaaaaatggcgaggtaaaacgaatcaaaggcttgatggagtgcaactataagaagttgagc

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		gacattcttaagcataccatcagtgactttgatgagttcaaagaagaagatcagctcatcactcttgccgaaatgtcagaagctctcaaagtggcttatgtcaaaatcattgtgaa catggcgttctcagatgatggtcaggttgataaaaaagaatttgccgaaattctcttgttgatgacccgacttgagttaacgactgaatcccggtttacactgcgtagttatgtcg gttcagaatccagtctgataccggttgaagaattaattgcgatcattgaccgggaatgtgtcccaagccataacaaatcaataaaagtctctcttgttaaagacctgattagcattt tcatgagtgttaatgaaggtgaatataaaaaattcccgtttcttcagcaagtgcaacctttgctgggcgtaactgacgaagaaatagaactcgcagtaatggctattcagcaaga ttttaagatgttacgggaagatttttccgatgatgcgctgaaacgcagtatgaaagaacttacggcaaaagcaggtgcggtaggcgtgccactcgctgctgtctatctctctgg ctctgtcatcggtatgtccgcagcgggcatcacttctgggcttgcaacacttggacttggtggcgtgctgggtttttcaagtatggcaacaggtatcggtgttgcggtgttattag gtgtaggtgcctataaagggattcgtcatcttacgggtgccaatgaactggataaaaccaagcgccgggaactcatgcttaatgaagtcatcaagcagacacaatccacattg tccgcgctaattaatgatctaaattatatttctggaaagtttaacgacgccctggatgcgcataatcggcaaggagaaaaaattctaaaactccagaagatgatgaatgcattga ccggtgcagcagatgaattgaataagaaatctaataaaatgcaaaacagtgcactcaaacttaagtgccctgtttatcttgatgaggccaaactcagttcgctgacccgagag cccatcaaaaaacaattccatgatgttgttctttcattctacgaagaatatcttgttgaagagcaaaacgatgggaagagtgttgaagtgaaaaaacttaagatcaaagaaaacg cttccactcagcaattagagaaacttgccgcgatctttgaaggcatcggctatttcagagcgggggatgttattaaaggcaaactaactgggctattctcataatgaaaaaacc agatactcaggtatcggccttgctggtgcagaagcaccagcttgaacaaagcgagcatcaattgggtgaccttgatgctgctctagaagcgcttaacgctttgcaaactgata ccgaagcttctttagatgaaatgattttggctatggatggtgttctggaacactcaggtatcacgtttgatgaggatatccacacaacggtttctagtgaattcagcgattaccttg aatcctgtttgaccacgtcatcgtccagtatcagtaaactgtcgatgatagaaacaatagcgttcaccagcgatatggactgggaaacctattcccagtccatatcgcagtatgc ccataaacacaatatcgatttaatagtcgatccgtttagcgccctgatgtctccaatccaaagaattgctctggaaaaacgtattcaggaagacttgaccttaaagactgcccgc tgcgacaaatatgattacatgatcgctggcacctgtggcgttattggcggacttatcgatatttttctggtaggcgtacctggagcaggaaaactgacccagcttgcagataatg cagtggacggtgccgttgagaaattcgcttcagcctttggatggaagggcagttcagaagcaagcgattcgacaaaaagcgctatcggtttttctggagagaaaattcaaaat caattatgaccatcggcatggcggagatgttgacggtttgttcaggatgaacacgaagaatcaccatattaaaagtctcgcccactccccggacttagtcggtttatttttctcga tcctggatcaatttaccagtacggcacattttgtggcagacggaaaattggtttccgtagataccgagacttttgagcttaaagggaataacgttgtctctaaggtatttagtggttt cgtaaactggctgggccaccttttctctgatatggcaggttcttccggtgcagcagggagaggctccggtatccccattcctttcttttcattacttcagtttattaatgtgggtgaa tttggccagcatcgccagtctttcgcaaccgtcgccgtccaggtttttgagaaagggtatgacttacggcatggattagcgatggcgatccccgtcatgattactgagttgcttg tgcgaatcacctggacggttaaacaacgttgctatcataagaaggactggggtgaatgtattccttcagcaaataaccctgaactcaggcgaatgttgcttgtggcgcatgga accttgtgtctgatggatgtaggagatgcggcacttcgttcaggaggcgaaatgattcagttcctcctgagaacgaacctcatcggctggacgaggtttggaattctagcgatt aaagaactccatgtctggtataaagcaggcggaattgatgccaatgctgtagatgaatatatggatcatgaacttcggcgaatgctaaaagcggggtagcgttacggctttgtt gaataacattacgtttgggtgcttggctgtaaaaagctaggcaatggcgtatctgtcgacgcaatgcagaaaaggcaacttaattgcgaaacagaaatgttcggtgagttgctt gaccgtcctatggcagctaagtgccagaagtcgacgttgctaacatcagtatgtactcatcggcacagtccatgtcagagctattaactatagataaaaattcaataattaataa aataagaaccatctttctaggtggttcttattattaacaataaatattacgatttcaacgagggttagaatg (SEQ ID NO: 412)
48	TerB + DUF279 + Lhr helcase	cctggtcctgccaattgctcccccagccatatgacataatccttttgaataatagggtttttatgcttgtactctagcccattcgcggtatcattttacgatctctcttccagttttatgc ttaccgcctttgcctatcgtagaacaatgccgggaagcgttatcagcgattaagggcaaggaatgggcttctggatatttgttattatgctggcggttatctggcttctgttttcca aaaagaaaaaatcgccgcccccagagtaaacaacaaaatcatcaccaaaataaatcattcatctcgacagaaatctctcaataagccagataacagcatgacaaatatgca ttctcaggcctccgatgatgacgaactggcaacctttacttttgtgaacgggcagacggttgaatacagcaccagccgccagccgtcacgagaaaacgccgcccgtagcaa taccactccagcgcgatgggtcaaaccgggagaaagcatcaccattcaaaatgtcgtcattaatcacggttatttttatttcggcgggcggttaaaaacacattcatcaggaga atatggatatctttataacgatgactccgacgcttcgctggttaatgacgcttttcccatcgagcctggttcacggcattattatgatgagtcactgggatactggcccagctttgc cacactctcccctcgctgccgtggcgcctatcttgactggctggcaagcgatcgcagcgatgcgagctgccccgttggctatgtttttatctatttttacggtctagaacgccgc gtactggccgatggcacacaagaagccatttctgacgatgaattcaaagcattattcgaagagatatcgcgcctgagaaccgtatttcaggcaagcggttccttccggcattat gcaacgcagttgctggaaatgatgatcgttctccgaccgaagttgctttctatatataccgaaaacgaatatttctcatcgaggagttcattactgttcagattaaatctagcgact gtggtcgataaaggacaacctatttgtgccgctctggcactggcatggatatactattttcctgattacaccctgcgcacgcctgcccgtcgatgtcatgctgaattttccgcatt attcaaacagcgttatactcaaaaatacggtgacggtattgtcgtcaaacccaataaaacacggttgtatttaagctataccccccgccagtggtacgcttcgggaacttcaggta aaaaaacagatggatcttcccgatcccagcgttttaaaagccccagttcagaaattaatttctgttgcagaatcctgtatcaacgcgctggatgcctacagtcgctatctcggta aaaaagatgcctcaccaagtgatgtcgccgccatcatgctgcttcccgatgaaatactgaccgaagatgcagaacgtctatttgctgaatttaaacactgggcagatgagaaa atccgtgaacattcaggactggcgacagtggctgatttctgggccagactgggtatgcctgtaccggataagattaataagaaagaagccgagctgatgcaaaatttcgccc ggcgagcaggctacggcattgcgccggatatgcgctatcaccttgtcagaccggatccagaaggtcatcttgttttatttcctgaagggcatgcggaattctacgtaccgtcg gcggaatttacgtcagtctctgtggcgcttcggttgggtgccatgattgcacaaatggacaagcgcgtggatgttgctgaacaggccgcgctggagaaaacgattaatcata acgatgcgctgtcgccaacagaaaaacgttcgctgcacgcctacctcacctggcggctcaatacgcctgcaaatcaggctggtctgaaaggtaaaattgagcaactcagcg ataaagataaatccactattggcaacgtgattatcagcgtcgcctgcgcagatggaaaaatcgatccggctgaaatcaaacaactggaaaaaatctacgccagcctcggtct ggacagcagtgccgttaccagcgatatccaccgactgtcaaccgcagaaacaactccgacagctacgttacaaaccccatcagcgacgagcggcgcgttttctcttgatga acggatccttgcccgtcatgaatccgacacaacggacgtacgccagttactgaacaccatcttcaccgaagatgaacccgcagacgaatccccagcggagatcccgccac

TABLE 16B-continued

(cloned sequences of systems #1-48)

System #	Name	Cloned Sequence
		acgctggcgcaggtcttgatgaagcacatcatcaactttaccaacgtttgcaggaaaaagaacgctgggcgcgaaacgaagtcgctgagctatgccagcagtttaatttgat gctaagcggcgcgattgaagcaattaatgactggtctttcgaacaggttgacgccccggtgcttgatgatgacgatgatatttacgttgacctggaaattgcacaagaactcaa aggataatttatgtctggcattcgtattcgtctcaaagaaagagacgctattattcagtcactgaagtcaggtgttacgcctaaaattggtattcagcacattcaggttggccgggt caacgaaataaaagcgctgtatcaggatattgagcgtatcgctgatggcggcgcaggattccggctgattattggggaatatggctcaggtaagacattctttttaagcgttgt gcgctcaattgcgctagaaaaaaagctggtgacaatcagcgccgatttatccccggacaggcgcatccacgcgacgggtgggcaggcgcgtaacctctactccgagcta atgaaaaatctatccacccgaaataagccggatggaaacgcattattaagcgtggttgagcgctttatcacggaagccagaaaagaagcagaaagtacaaatgtgtcagttc cgacgattattcaccaaaagctcgccgccctgtctgatatggttggcggttacgatttcgccaaagtcattgaatgttactggcagggccacgagcaggataatgagacattg aaatcaaatgccatccgctggctaagaggtgaatacaccacgaaaaccgacgcccgtaacgatctgggtgtgcgcaccattatttctgatgcctctttctacgattcgctaaag ctgatgagcctgtttgtccgtcaggccggatacgcgggtctgctggtgaatctggatgagatggtcaatctgtataagctcagtaacactcaggcccgcgttgccaactatga acagatactgcgtattctgaatgactgcctgcaagggacggctgaatatatcggtttttttacttggcggtacgccagaattcctgttcgatccgcgcaaggggttgtacagctac gaagcgctccagtcccgactggcggaaaatagcttcgctcagcgggctggtgtcattgattattcgtccccttccctgcacttagccagcctgacgccggaagaactctatatt ctgttgaaaaaccttcgtcacgtttattccggcggcgatgcggataagtatctggttcctgatgatgctctgacggcatttttacgccactgtagcaacactattggcgatgcctat ttccgtacgccacgaaacacgattaaagccttcctggatatgctggccgtgctggaacaaaacccatccattcagtggtcacagttaatcgccggtgtcgcgatcgcggaag aaaaacccagtgatatggatgaaataacatcggcagaagatgccgatgaggacggtctggccgacttcagattatgatgaacgaataccagcggctggatccacggatac agaagtggatataccggcagggatgggccgatctcagggaactgcaaaaaaaatccgtttcaccgatattagcgggcgatcgggatgttctgatcagcgccgcgactgcc gcaggtaaaacagaagcgtttttcctgcccgcctgttctgccattgcggatattcagggcggctttggcattttatacatcagcccgcttaaggccctgattaacgatcagtatc gaaggctggaaaacctcggtgatgcgttggagatgccggtcacgccctggcatggtgatgttgcgcagagcaaaaagctgaaagcaaagaagaatcctgccggtattttg cttatcaccccggaatcgctggaagcgatgctgatccgcaatgcgggatggttaaagcaggctttcgcgccactggcatatatcgccattgatgaattccatgctttcatcggtt ctgagcggggtatgcagcttctctctctgttaaatcgagtcgatcacctgctgggaagaatcaacaatccagtcccccgagtcgcactcagcgcaacgctgggggaactgg aacaggtgccgttatctctgcggccaaatcaacgtctgccctgtgacattattaccgacagtcagactcacgccacgctaaaagtacaggtgaaaggttatctggaaccgctg accacctcgggccagcaatctccaccgtcggcagagacgcaaatctgccatgatatctttcgcctctgtcgtggtgattcccatctggtgttcgctaatagtcgcaaacggac cgaaagcattgccgccacgcttagcgatctcagtgaagcgagcatcgttcccaatgagttctttccccatcacggatctctgtccagagatctgcgtgaaacgctggaacaga ggcttcaacaaggcaacttacccaccaccgccatctgtacgatgacgttagagcttggcatcgacatcggtaaagtcagctccgttgtgcaagttaccgcccccattccgta gccagcctgcgtcagcgaatgggacgctccggtcggcgcgactcgcctgccgtattgagaatgctgattgccgaacatgaactgacgccaacatcaggcattgtcgacca gctcaggcttcagcttgttcagtcgctggccatgatccgcttacttatcggcaacaaatggtttgagccagctgatacccggcagatgcactattccaccctgttccatcagatc ctggcgatcgtggcgcagtggggaggcgtgcgtgcggatcagatctggtcacagctatgcctgcaagggccatttcagaaagtccggatctatgacttcaaaacgttattga aacatatgggggagcaccagtttctgacccagctctcaagcggcgaactggttctgggcgtcgagggcgaacgtcaggtaaatcaatacaccttctacgccgtgttcagca cgccggaagagtttcgcattgtggcggggagcaaaacactgggctccattcccgttgattccccactgatgcctgatcaacacattattttcggcggtcgacgctggaaggta accgatatcgatagtgataaaaaagttatttatgtcgaggcgacaaagggtgggcagccgccgttatttggcggacaagggatgtccattcatgatgtcgtccgccaagaaat gctcactatttatcgggaaggcgactaccgcatcaccgttggcaatcgcaaggccgattttgccgataccacggccaaaaacctgtttgatgaagggctgcactgttttcgca acaataatctggcttcggaatgttttattcagcagagacagcatgtctacattcttccctggctaggcgatcaaaccgtaaacacgttgtcggcattacttatccaacgcggtttc aaggcgggctcatttgctggtgtggttgaagtagaaaaaactacggtctcggaggttaaacaagcgttattcagcgcacttcaggaagggctaccttacgaatcccgtcttgc cgaaagcatcgttgaaaagtgcctcgaaaaatatgatgagtatttacccgagacgttgctgacgcaggaatatggattacgtgcttttaatattgaacgcgtgacggagtggtt gcaggggcatttatattaaggggaagaaga (SEQ ID NO: 413)

TABLE 17

Genome coordinates of RADAR editing sites in Figure 27

Site #	Gene	Position in genome (Genbank: GCA_000005845.2)	% A-to-I RNA editing
1	ffs	476502	82
2	dinQ	3647752	88
2	dinQ	3647753	57
3	ftsI	92547	90
4	lpp	1757597	52
5	rpsB	190414	76
6	ssrA	2755713	61
6	ssrA	2755714	56
7	(intergenic)	3647944	69
7	(intergenic)	3647945	97
8	hokB	1492029	95
9	mgrR	1622894	87
9	mgrR	1622895	87
10	ptsI (1)	2534135	80
11	secY	3443842	78
12	atpC	3915927	69
12	atpC	3915928	76
13	rbsB (1)	3937080	76
14	rpoA	3440833	74
15	rplI	4426356	73
16	(intergenic)	2002020	70
17	pflB	951380	68
17	pflB	951381	58
18	ptsI (2)	2534211	68
19	rplA (1)	4179468	66
19	rplA (1)	4179469	68
20	(intergenic)	127818	68
21	skp	200777	67
22	(intergenic)	2518138	51
22	(intergenic)	2518139	66
23	rbsB (2)	3937116	65
24	infC	1800153	65
25	rplT	1799499	64
26	gapA (1)	1863658	64
27	sodB	1735694	62
28	gapA (2)	1862864	61
29	rpsC	3449386	61
30	leuW	697012	61
31	rpsA	962878	60
32	ibsC	3056901	60
33	ahpC	639397	59
33	ahpC	639398	56
34	oxyS	4158372	59
35	rpmG	3811305	58
36	(intergenic)	780980	57
37	iscU	2660065	57
38	ryfD	2734233	56
39	deaD	3306635	56
40	hns	1292675	56
41	(intergenic)	4392565	56
42	tig	456390	56
42	tig	456391	56
43	rplA (2)	4178970	56
44	tsf	191433	51
44	tsf	191434	55
45	rnpB	3270434	54
46	(intergenic)	781019	54
46	(intergenic)	781020	52
47	eno	2906708	52
48	(intergenic)	3071334	51

TABLE 18A

Description of phage T2 fragments in FIGS. 28C-28E

Fragment #	Length (bp)	A93% editing		A121% editing		Gene #	Accession	Gene	Description
1	2392	28	23	37	32	1	AYD82599.1	rIIA.1	hypothetical protein
						2	AYD82598.1	rIIA	protector from prophage-induced early lysis
2	1818	5	5	6	6	1	AYD82600.1	gp39	DNA topoisomerase II large subunit
3	261	6	6	8	9	1	AYD82601.1	gp39.1	hypothetical protein
4	1423	8	5	10	8	1	AYD82606.1	—	hypothetical protein
						2	AYD82605.1	cef	modifier of suppressor tRNAs
						3	AYD82604.1	goF	mRNA metabolism modulator
						4	AYD82603.1	gp39.2	hypothetical protein
						5	AYD82602.1	—	hypothetical protein
5	3570	6	9	7	11	1	AYD82613.1	srd	anti-sigma factor
						2	AYD82612.1	dda.1	hypothetical protein
						3	AYD82611.1	dda	DNA helicase
						4	AYD82610.1	dexA.2	hypothetical protein
						5	AYD82609.1	dexA.1	hypothetical protein
						6	AYD82608.1	dexA	exonuclease
7	1339	38	44	49	56	1	AYD82628.1	—	hypothetical protein
						2	AYD82627.1	dam	DNA adenine methyltransferase
8	201	4	2	5	3	1	AYD82629.1	—	hypothetical protein
9	442	1	1	2	2	1	AYD82635.1	dmd	discriminator of mRNA degradation
						2	AYD82634.1	gp61.4	hypothetical protein
10	2956	22	20	29	27	1	AYD82638.1	uvsX	RecA-like recombination protein
						2	AYD82637.1	gp40	head vertex assembly chaperone
						3	AYD82636.1	gp41	helicase
11	2697	2	2	3	3	1	AYD82644.1	gp43	DNA polymerase
12	687	3	3	5	4	1	AYD82648.1	gp45	sliding clamp
13	588	85	85	93	92	1	AYD82650.1	gp45.2	hypothetical protein
						2	AYD82649.1	rpbA	RNA polymerase binding protein
14	1203	52	46	59	53	1	AYD82657.1	a-gt	DNA alpha glucosyl transferase
15	545	27	22	48	40	1	AYD82664.1	gp55.2	hypothetical protein
						2	AYD82663.1	gp55.1	hypothetical protein

TABLE 18A-continued

Description of phage T2 fragments in FIGS. 28C-28E

Fragment #	Length (bp)	A93% editing		A121% editing		Gene #	Accession	Gene	Description
16	3394	60	57	69	67	1	AYD82674.1	gp49	recombination endonuclease VII
						2	AYD82673.1	nrdD	anaerobic ribonucleotide reductase subunit
						3	AYD82672.1	nrdG	anaerobic NTP reductase small subunit
						4	AYD82671.1	—	hypothetical protein
						5	AYD82670.1	gp55.8	hypothetical protein
						6	AYD82669.1	nrdH	glutaredoxin
18	2329	3	2	5	3	1	AYD82686.1	nrdC.5	hypothetical protein
19	528	5	5	8	8	1	AYD82689.1	nrdC.8	hypothetical protein
20	303	2	1	3	2	1	AYD82690.1	nrdC.9	hypothetical protein
21	2659	30	31	33	36	1	AYD82699.1	mobD.2	hypothetical protein
						3	AYD82693.1	nrdC.11	hypothetical protein
22	902	6	6	7	7	1	AYD82706.1	rI.1	hypothetical protein
						2	AYD82705.1	rI	lysis inhibition regulator
						3	AYD82704.1	rI.-1	hypothetical protein
23	2602	4	4	6	7	1	AYD82724.1	ip4	hypothetical protein
						2	AYD82721.1	vs.7	hypothetical protein
						3	AYD82720.1	vs.6	hypothetical protein
						4	AYD82719.1	vs.5	hypothetical protein
						5	AYD82718.1	vs.4	hypothetical protein
						6	AYD82717.1	vs.3	hypothetical protein
24	495	6	5	10	8	1	AYD82725.1	e	lysozyme murein hydrolase
25	594	7	5	9	8	1	AYD82730.1	e.6	hypothetical protein
26	177	3	3	4	4	1	AYD82731.1	—	hypothetical protein
27	264	3	2	4	3	1	AYD82732.1	e.8	hypothetical protein
28	351	7	6	10	10	1	AYD82733.1	—	hypothetical protein
29	402	5	4	8	5	1	AYD82734.1	trna.1	hypothetical protein
30	991	2	2	6	4	1	AYD82737.1	trna.4	putative membrane protein
						2	AYD82736.1	trna.2	hypothetical protein
						3	AYD82735.1	—	hypothetical protein
31	309	6	5	8	9	1	AYD82738.1	ip7	hypothetical protein
32	255	20	19	26	25	1	AYD82739.1	ip5	hypothetical protein
33	1423	28	27	36	36	1	AYD82742.1	gp1	deoxynucleoside monophosphate kinase
						2	AYD82741.1	gp57A	chaperone for tail fiber formation
						3	AYD82740.1	gp57B	hypothetical protein
34	1277	54	54	69	72	1	AYD82745.1	gp50	head completion protein
						2	AYD82744.1	gp2	DNA end protector protein
35	8107	2	2	3	3	1	AYD82755.1	gp9	baseplate wedge tail fiber connector
						2	AYD82756.1	gp10	baseplate wedge subunit and tail pin
						3	AYD82757.1	gp11	baseplate wedge subunit and tail pin
						4	AYD82758.1	gp12	short tail fibers protein
						5	AYD82759.1	wac	fibritin
						6	AYD82760.1	gp13	neck protein
						7	AYD82761.1	gp14	neck protein
36	5149	33	37	46	50	1	AYD82762.1	gp15	tail sheath stabilizer and completion protein
						2	AYD82763.1	gp16	small terminase protein
						3	AYD82764.1	gp17	large terminase protein
						4	AYD82765.1	gp18	tail sheath protein
37	492	4	4	6	6	1	AYD82766.1	gp19	tail tube protein
38	1284	2	3	3	4	1	AYD82773.1	gp24	capsid vertex protein
39	1476	35	33	45	40	1	AYD82863.1	gp24.3	hypothetical protein
						2	AYD82775.1	gp24.2	hypothetical protein
40	1807	17	23	23	30	1	AYD82776.1	inh	inhibitor of prohead protease
41	832	1	3	2	3	1	AYD82781.1	uvsY	recombination, repair and ssDNA binding protein
						2	AYD82780.1	uvsY.-1	hypothetical protein
						3	AYD82779.1	uvsY.-2	hypothetical protein
42	1025	1	1	2	2	1	AYD82783.1	gp26	baseplate hub subunit
						2	AYD82782.1	gp25	tail lysozyme
43	6240	1	1	1	1	1	AYD82784.1	gp51	baseplate hub assembly protein
						2	AYD82785.1	gp27	baseplate hub subunit
						3	AYD82786.1	gp28	baseplate hub distal subunit
						4	AYD82787.1	gp29	baseplate hub subunit tail length determinator
						5	AYD82788.1	gp48	baseplate subunit
						6	AYD82789.1	gp54	baseplate subunit
44	291	1	1	2	2	1	AYD82790.1	alt.-3	hypothetical protein
45	4155	2	2	3	3	1	AYD82792.1	alt	ADP-ribosyltransferase
						2	AYD82791.1	alt.-1	hypothetical protein
46	366	6	7	8	9	1	AYD82801.1	gp30.7	hypothetical protein
47	177	6	6	9	9	1	AYD82802.1	gp30.9	hypothetical protein
48	249	2	3	3	4	1	AYD82803.1	rIII	lysis inhibition accessory protein
49	336	1	2	2	2	1	AYD82804.1	gp31	head assembly cochaperone with GroEL

TABLE 18A-continued

Description of phage T2 fragments in FIGS. 28C-28E

Fragment #	Length (bp)	A93% editing		A121% editing		Gene #	Accession	Gene	Description
50	1698	4	3	6	4	1	AYD82809.1	cd.2	hypothetical protein
						2	AYD82808.1	cd.1	hypothetical protein
						3	AYD82807.1	cd	deoxycytidylate deaminase
						4	AYD82806.1	gp31.2	hypothetical protein
						5	AYD82805.1	gp31.1	hypothetical protein
51	276	3	3	5	5	1	AYD82810.1	cd.3	hypothetical protein
52	3683	5	6	7	8	1	AYD82823.1	td	thymidylate synthetase
						2	AYD82822.1	nrdA.2	hypothetical protein
						3	AYD82821.1	nrdA.1	hypothetical protein
						4	AYD82820.1	nrdA	ribonucleoside-diphosphate reductase subunit alpha
53	1448	45	62	58	69	1	AYD82827.1	frd.1	hypothetical protein
						2	AYD82826.1	—	hypothetical protein
						3	AYD82825.1	frd	dihydrofolate reductase
						4	AYD82824.1	—	hypothetical protein
54	366	1	2	2	3	1	AYD82828.1	frd.2	hypothetical protein
55	228	11	11	16	16	1	AYD82829.1	frd.3	hypothetical protein
56	909	2	3	3	4	1	AYD82830.1	gp32	single-stranded DNA binding protein
57	2162	40	48	51	67	1	AYD82834.1	rnh	RnaseH
						2	AYD82833.1	dsbA	double-stranded DNA binding protein
						3	AYD82832.1	gp33	late promoter transcription accessory protein
						4	AYD82831.1	—	hypothetical protein
58	4997	3	2	5	3	1	AYD82835.1	gp34	long tail fiber proximal subunit
						2	AYD82836.1	gp35	hinge connector of long tail fiber proximal connector
59	417	42	48	46	54	1	AYD82859.1	—	hypothetical protein
						2	BBC14887.1	ndd.6	putative outer membrane protein
						3	AYD82858.1	ndd.5	putative outer membrane protein
60	1166	26	27	29	31	1	AYD82862.1	rIIB	protector from prophage-induced early lysis
						2	AYD82861.1	denB.1	hypothetical protein

TABLE 18B

DNA sequences of fragments #1-60 in Table 18A

Fragment #	DNA sequence
1	atgaaatcatatagagtaaatttagaactttttgataaagcagttcatcgagaatatagaatcattcaacgctttttcgatatgggagaagccgaagaatttaaaaaccgctttaaggatattagagataaaattcaatccgacaccgcaactaaagatgaattactagaagttgctgaagttattaagcgtaatatgaattaatgaggaaattatgattatcaccactgaaaaagaaacaattcttggtaatggttctaaatcaaaagcatttagcatcacagcatctcctaaagtatttaaaattctgtcatctgatttgtatacaaacaaaattcgcgcagtagtccgtgaattgattactaacatgattgatgcccatgctctcaatggaaatcctgaaaaatttatcattcaagttccaggacgattagatccgcgatttgtttgtcgagattttggtccgggtatgagtgattttgatattcagggtgatgataattctcctgggctgtataattcatacttcagttcatctaaagctgaatctaatgatttcattggtggatttggtttaggttctaaatctccgtttagttatactgatacgtttagtattacttcataccataaaggtgaaattcgtggttatgtagcttacatggatggtgatggcccacagattaaacctacattcgtaaaagaaatgggtccagatgataaaactggcattgaaatcgtagttccagttgaagaaaaagactttagaaactttgcttatgaagtttcttatatcatgcggccgttcaaagatttggctatcattaatagtcttgaccgtgaaattgactattttccggattttgatgattattacggcgtaaatccagaaagatactggcctgatcgtggtggattatatgctatctatggcggtattgtttatcctattgatggtgttattagagaccgcaactggttaagcattcgcaatgaagtgaattacattaagtttccaatgggttcacttgatattgctccatctcgcgaggctctttcacttgatgatcgtactcgtaaaaatattattgagcgagttaaagaactcagtgagcaagcatttaatgaagatgtaaaacgatttaaagaatctacatctcctcgtcacacatatcgtgaattgatgaagatggggtattctgctcgagattatatgattagtaattcagtcaaattcacgactaaaaatctgtcatataagaagatgcagagtatgtttgaacctgatagtaagttatgcaatgcaggagttgtgtatgaagtaaatcttgaccctcgactgaagcgcattaagcaaagtcatgaaacttcagccgttgcatcaagttatcgtctgtttggtattaatacaacaaaaattaatattgttattgataatattaaaaatcgtgttaatattgtccgtggattagcacgtgcgttagatgatagtgaatttaataacactttgaatattcatcacaatgagcgtcttctgtttattaacccagaagtagaatcgcagattgatttgcttcctgatattatggcaatgtttgaaagtgatgaagttaacattcattatttgtcagaaatcgaagctttagttaaaagctatattccaaaggtagttaaaagtaaagctcctcgtcctaaagctgctacagcatttaagtttgaaattaaagacgggcgctgggaaaaagaggaactatttacacttacgtcagaagcagatgaaattactggttatgtagcgtatatgcatcgttctgatattttctctatggatggtactacatctctttgtaatccatctatgaatatttgattcgtatggctaatcttattggcattaatgaattttatgttattcgtccgcttttacagaaaaaggtaaaagaactcggtcagtgccaatgtatttttgaaactctacgcgatttatatgtagatgcttttgatgatgtagattatgataagtatgtaggttattcaagttcagctaaacgatatattgataaaattatcaagtatcctgagctagattttatgatgaagtacttcagtgtagatgaagtttctgaagaatatacacgactcgctaatatggttagttcattacagggtgtatattttaatggtggaaaagataccattggtcatgacatctggacagtaactaatctttttgatgtattatcaaataatgcttcaaaaaacagtgataaaatggttgctgagtttaccaagaaattccgtattgtttccgacttcatcggatatcgcaactctttaagtgatgatgaagtttctcaaatcgctaaaactatgaaggcccttgcggcctaa (SEQ ID NO: 414)
2	atgattaagaatgaaattaaaattctgagcgatattgaacacatcaaaaagcgtagtggcatgtatattggctcttctgctaatgaaatgcatgagcgctttctgtttggtaaatgggaaagtgttcagtatgtacctggtcttgttaagcttattgatgaaattatcgataactcagtagatgaaggtattcgtactaagtttaaattagcaaataaaattaatgttactattaaaaacaatcaagtaacagttgaagataacggtcgtggtattccacaagcgatggttaaaacacctactggtgaagaaattcctggtccagttgctgcatggactattccaaaagcaggtggtaactttggtgatgataaagaacgcgtcaccggtggtatgaatggtgttggttctagtttgacaaacatttttctgtgatgtttgtcggtgaaactggcgatggtcaaaataatattgtagttcgttgttcaaatggcatggaaaataaatcatgggaagatattcctggaaaatggaaaggaactcgtgttactttcattcctgattttatgtcatttgaaactaatgagctgtcccaagtttatcttgacattacacttgatcgtctccagacgcttgctgtagtttatcctgatattcaatttacctttaatggtaaaaaggttcagggcaattttaagaaatatgcacgacagtatgatgaacatgctattgttcaagaacaagaaaattgttctattgcggttggtcgttcaccggatggttttcgtcagttgacgtacgtcaataacattcatactaagaatggtggccatcatattgactgtgttatggatgatatttgtgaagaccttattccacaaatcaaacgtaaattcaaaattgatgtaactaaagcacgtgttaaagaatgtttgactatcgttatgtttgttcgcgatatgaaaaacatgcgatttgactctcaaactaaagaacgacttacttctccttttggtgaaattcgtagtcatattcaacttgatgctaaaaagatttcacgcgctattctaaataatgaagcaattttaatgccaattattgaagcagcattagctcgtaaattggcggcggaaaaagcagcagagacaaaggcagctaaaaaagcttctaaagctaaggttcataaacatatcaaagcgaatctttgtggtaaagatgctgatactactcttttcttgactgagggtgattctgctatcggatatcttattgatgttcgtgataaagaacttcatggtggttatccattgcgtggtaaagttcttaatagctggggtatgtcatatgccgatatgcttaaaaacaaagaactatttgatatttgcgcaatcactggtctagttcttggtgaaaaagctgaaaacttgaattatcataatattgctattatgactgatgctgaccatgatggtctaggaagcatttatccttctctgctcggatttttagtaattggccagaattgtttgagcaaggacgaattcgctttgtcaaaactcctgtaatcatcgctcaggtcggtaaaaaacaagaatggttttatacagtcgctgaatatgagagtgccaaagatgctctacctaaacatagcatccgttatattaaaggacttggctctttggaaaaatctgaatatcgtgaaatgattcaaaatccagtatatgatgttgttaaacttcctgagaactggaaagagctttttgaaatgctcatgggagataatgctgaccttcgtaaagaatggatgagccagtag (SEQ ID NO: 415)
3	atgaaatatattaatcgttctatcgcagcattagtattagcagtgtctttagtaggatgtactgatgctgataatgcaacaaaagttttgtcttcaagtggttttactaatattgaaatcactggatataattggtttggttgctctgaaaatgatttccagcatactggatttcgtgctattggacctaccgggcagaaagtagaaggaacagtatgttctggtttattcttcaaagattcgactatccgttttaaataa (SEQ ID NO: 416)
4	atggaaaacttaattatcatcgagcaatctttcaacgattatggtatggcttatggttatcgtgcgataatggaagattctcgtggatgtgttatcgatattgctgaatgtaaagatttactgcagcttttgaagattgttcgcaaaaattgggattgtgaaaatattaaagttcgaattgttacagaagaagaaactgtttttcatgatgtaaaattcgctaaaggtgctgctactcttctgaaacgtatcgctccactgttcaattaatgaggaaattataatgaaacgtaaaattgttcagaactgcactaatgatgaatttgaagatgtattattcgatccagatttggtagtagttcaaaaggaacacaccatcaagtttactcacttgacttcggtttatgtgtatgagaaagtcggtgataaacaaccaatttacggtgtatttcgtgaaattactgaagatggcacaacttactggaaggaaatttattaatggctattaaattt

TABLE 18B-continued

Frag-ment #	DNA sequences of fragments #1-60 in Table 18A DNA sequence
	gaagttaataaatggtatcaatttaaaaataaacaagctcaagaaaattttattaaagaccatactgataacggaatctatgcacgccgtttaggtatgcatccttttaaaattttagatgttgattatctttggcgtcctactaaaattgtgacatctactggcacagttggatatgcaacacacggtgatatccttgacgaaaactttatctggctttctactaacgaagctgggttctttgatgaagtggaaaatccatatcaggcagttgaagagcaagagcaggaagagaaagagcaagaacaaatagaagatttcacagaattcccagtaatgaaagttactattgaaaataatgaacaggcatggtccttgtatcaaatgctgaaagcacactttaaggaataattatgccaatgtatgattataaatgccaatccgaagattgcgggcatgaatatgaaaaaattaaaaagatttctgaacgagaaaatgatgtttgccctaaatgtcatcgtttgtctactcgtcggccttctgctcctaagcatgtgaatggtggtttttacgacttacttaaagggtaattatgtttaaaatcggtaagaaatattgcattcgtgaaggtgaagaacagaaatatctactttctgctagtaataggaatagttctattaatgctgtaatattgactagtgaatttatcgttgaagatatgaaaggtcataatgttacaatgattagtacagcatctggaaatgatggaaaaattcttcatagttgtcagagtaatgttctaatttatgatgaagaatttgacttcttcaaagaagtttccgaagattttgattttgaatgtactattactatgaaatctggtgaccctctttcttttacagttagatga (SEQ ID NO: 417)
5	atgaagctgcataatatgtctaataatcaaattcgtaaaattaaacgtcgtttagagcatactcaggcatctgctaaaagacgttctaaagattttaacttagacttcaattacattaagaacatttttagaccaaaaagtttgcgcttactcgggagaaccttttgataatcgtattgaaggagagaaattatcattagaacgttttgataataacgttggatacattaaagggaatgttattgcagtaaagaaaaagtataatacatttcgttctgattatactttagaggagttaattgaaaaacgtgatttgtttgctttgcgaattggtcgttcatctgcgaaaaaagttcataaactaaatttagatgaaaagaaatgggctaaaatcaaaaagacttataatcaaattaaagctatacagaaaaaacgtgaaaaccgaattgaacacatttctcagattctaaatcaaaacagacctctgacattaagctaagaattatagcacttaaagctcgtattgatggttctcgtatagcagaaggcgctgaagttgttaaattgaacgttcttcttaaaggctcggattggaaaactgtgaaaaagttgtcagaagcagaaatgcaatatgatatgtgtgataaaattattcaaggtgtagagcggtatcaaaacttgtcttttattgataaacttaaactgaaaagaggatatccgctaaattgttcaatttttaaacttatccgaggataatatggtttatgtatatgcgatagtttaccgagacaaagacggatttacggcgccagttccgcttgatgaacatcgtcctgctgtatttttgaatggaagattgctgataaagtatttaccactcttaaagagcagtatcaactagctttaggtaagggaattccaagattagttgagactccacgcaagttttggtttaataaaatagaagttaaacatgttaagcctgatgtagacacacaaagattatatcggcgaattttagatactgggcgtattgttagtataccaattgcagggaatttacgatgacatttgatgatttgaccgaaggtcaaaaaaatgcctttaacattgttatgaaggctattaaagaaaagaaacatcatgtaactattaatggacctgctggtaccggtaagactactcttactaagttcatcattgaagctttaatatctacgggtgaaactggtattattttagcagctcctactcatgcagctaaaaagattctttcaaaactatcatggaaagaagcgagtactattcatagtattcttaaaattaacccagtaacatacgaagaaaacgttctttttgaacaaaaagaagtaccagatttagctaaatgcagggtattaatctgcgatgaagtgtcaatgtatgatagaaagctatttaaaattctgctttcaactatcccgccgtggtgtactataattggaataggcgataataagcaaattagacctgttgacccaggagaaaatactgcttatatcagtccattctttacacacaaagatttttatcagtgtgaactcactgaagttaaacgcagtaatgctcctattattgatgtagctactgacgttcgtaacggtaagtggatttatgataaagttgttgacgggcatggagtacgtggatttactggtgataccgctttacgcgattttatggtaaattatttttcaatcgtcaaatctttagatgatttgtttgaaaatcgcgtaatggcatttacgaataaatctgttgataagttaaatagcattattcgtaaaaagattttttgaaactgataaagattttattgttggtgaaattattgtaatgcaggaaccattaattaaaacatataaaattgatggaaagcctgtgtcagaaattatttttaataacggacaattagttcgtattatagaagcagagtatacatcaacgtttgttaaagctcgtggtgttcctggagaatacttaattcgtcattgggatttaacagtagaaacttacggcgatgatgaatattatcgtgaaaagattaaaataatttcatctgatgaagaactatataagtttaacctatttttaggtaaaacagcagaaacttataaaaattggaacaaaggtggaaaagctccatggagtgatttttgggatgctaaatcacagttcagtaaagtgaaagcacttcctgcatcaacattccataaagcgcaaggtatgtctgtagaccgtgctttcatttatacaccttgtattcattatgcagatgctgaattggctcaacaacttctttatgttggtgttacccgtggtcgttatgatgtattttatgtatgattaaatttgaggaagctattcgtggaaataactaaagatcagttttatcttcttcaagataaagtaagcgaaatttatgaaattgctcatggtaaaaatcgtgaaactgtaaaaattgaatctagtaagttgatgcttcaattagaagaaattgaacgagatttaattgcgttagaattcttttgtggcgaagtgaaaactgttacaattaatgattatgttttaggcgaaattagctatctttatgaggcgattattaatgattgaattaagttggtgccagtttaaatctcttatgacaaatgttaaagctgtcattgaagaaaatcagggtcctgaaaatattactattcgcgaaaaagctttaaagatagtatacagtcttgaagaaatacaaaaagatattgaatctatggcaaaatttattgatgagcctattaataaagtttatattcaagactatactgtaggtcaaattcgcgatttagcgaggaaagtttaatgtttgattttattatagattttgaaacaatgggaagtggtgaaaaagcagctgttattgatttggctgtaattgcttttgaccctaatccagaagtcgttgaaacattcgatgaattagtttcacgtggcattaaaatcaaatttgatttaaaaagccaaaaaggacatcgtctttttactaaaagcactatcgaatggtggaaaaatcaatctcctgaagctcgaaaaaatattgcaccatcagatgaagatgtaagcactatcgatggtattgcaaaatttaatgattacatcaatgcacataatatcgatccttggaaatctcaaggctggtgtcgtggaatgtcatttgattttccaattttagtcgatctcattcgtgatattcaacgccttaatggtgtatctgagaatgagcttgacacatttaagttagaaccatgtaaattctggaatcagcgtgatattcgtaccagaattgaagcacttctgcttgttcgtgatatgaccacgtgtcctcttccaaaaggaactttagatggattcgttgcgcatgattctattcatgactgtgcgaaagacatcctgatgatgaagtatgctttgcgatatgctatgggtcttgaagatgctccatcagaggaagaatgcgatcctctatctcttccaacaaaacgataa (SEQ ID NO: 418)
7	atgattaataaaattgtgcatgaaatggctttaaacggagattcatataaaatatctgccgtagttgaaaatttcatacttaataaagtaaaagaatatttcactgattgttcagttagttatcaagaaaaaatggttttaattgatgatactgaaaaatcaaataatttgttttgctctaattttataactaaaaagcgtactagaagatttgatattgttatttctcgcaacggtaaaaagcatataattgaaattaaacaccaagttggtggaggtacagctattgattcggttggaatatatttagaagataaagagaaattaaaagaatacacaaaaactgaaaaccctgtgtcattgatgatattagatttttttgccatgcggatattatccacgtaataaatggacaaaaagagaatcatttactgataatccaaccatccaagcaaggtttaatgaatatgctaaatcacaaaacgtgttagtattattatcaaatacatatgatgaagaattgtataattcattttttgctgcaataaatgagagaatataatgctaggagctatcgcgtatacgggtaataaacaatcattattacctgaacttaagcctcactttccgaaatatgacagattcgtggatttattttgtggaggtttatcagtgtctttgaacgtcaatggtcctgtattggccaatgatattcaagaaccaattattgaaatgtataagcgtcttattaatgtatcatgggatgacgttttaaaagtaataaagcaatacaaactatcaaaaacatcaaaagaagagtttttgaaattacgtgaagattataataaaactagagatcctcttttactttatgttcttcattttcacgggtttagtaatatgattcgtataaacgataaaggaaattttactactccgtttggaaaaagaactataaacaaaaatagtgaaaaacgctttaatcactttaaacaaaattgtgataaaataatctttagttcattgcattt

TABLE 18B-continued

DNA sequences of fragments #1-60 in Table 18A

Fragment #	DNA sequence
	taaagatgtcaaaattctagacggcgattttgtatatgtggaccctccgtatcttataacagttgctgattataataaattttggtcagaagaagaagaaaaagaccttttaaatcttttagattcttaaatgacagaggaataaaatttggactgtcgaatgttttagagcatcacggaaaggaaaacactcttcttaaagaatggtctaaaaaatataatgttaagcatcttaataaaaaatacgtctttaacatatatcattccaaagaaaagaatggaactgatgaagtatatatttttaattaa (SEQ ID NO: 419)
8	atggtacaaaaattaatggcacttgttaatgccatcaaaggtaataaaaagcgtatagcttttactatttctgctatggtaggaattttactctggaactttattttatcacctgttgcaattgcacatggtattaatattccaatagttactcttgatacattcgtagatttagcatttgctttagttgggttaatttaa (SEQ ID NO: 420)
9	atggaattggtaaaggtagtttttatggggtggtttaagaatgaaagcatgtttactaaagaaaccacaatgatgaaagatgacgttcaatgggctactactcaatatgctgaagttaataaagcattagttaaagctttcattgatgataagaaagtgtgtgaagtggattgccgaggataatatgcatattgttttatttaaacctactccgtataacgtcaggaaaaatacgcaattcaaagcacttattgcagatacgtgggaattggtgttagatattccagcagaagaaagtcctccatttggtcgagtggaatttattaagtttgctgttcgccctacgaagcggcagattcgccaatgcaaaagatactttcgtaagatcgttaagctagagaaacagtttgtaacatgtgattacgcaaaagttttaaaataa (SEQ ID NO: 421)
10	atgtctattgcagatttaaaatcccgtttgattaaagcttccacttctaaaatgactgctgagctgactacatctaaattctttaatgaaaaggatgtaatccgtacaaaaatcccaatgcttaatattgctatttctggtgcgattgatggtggtatgcagtctggtttaactattttcgcagggccttctaaacactttaaatcaaatatgtctttgactatggttgcggcatatttgaacaaatatcctgacgcggtttgtctattctatgatagcgaatttggtattactccagcttatttgcgatccatgggagttgacccggaacgagtaattcatacgccaattcagtcagttgaacagctgaaaattgatatggtgaaccagcttgaagctattgagcgtggtgaaaaggttattgtattcatcgactcaatcggtaatatggcttctaagaaagaaacggaagatgccttgaatgaaaaatctgtggcagatatgactcgtgctaaatcactgaagtcattattccgtattgttactccttattttagcattaaaaatattccgtgtgttgcggttaaccatacaattgaaacaattgaaatgtttagtaaaaccgtgatgacaggtggtacaggcgtaatgtattcggctgatactgtattcattatcggtaaacgtcagattaaagatggttctgatcttcaggggtatcaatttgttctaaatgtagaaaaatctcgtaccgttaaagaaaaaagtaaatttttattgatgttaaatttgacggtggtatcgatccttattctggattgttagatatggctctagaattaggattcgtggtaaaacctaagaatggttggtatgctcgtgaatttcttgacgaagaaaccggcgagatgattcgcgaagaaaaatcttggcgtgcaaaagatactaactgcactacattctggggtcctttatttaagcatcaaccattccgagatgctattaaacgtgcttatcagttaggtgctattgatagtaatgaaattgttgaagctgaagttgatgaattgattaactcaaaggttgaaaaatttaaatctccagaaagtaaaagtaaatcagcagctgatttagaaactgacctcgaacagttaagtgatatggaagaatttaatgaataaagatgatttagatttagatctagaaattatcgatgaatccccctcttcggagggggaagaagaaagaaaagaacgtctttttaatgagtctcttaagataattaaatccgctatggaaaatgttatccaggagattgtcattaaactagaagatggttctacacatatagtgtatgtaacaaaactggattgggttgatggaaaggttgtaatggactttgctgttcttgaccaagaaagaaaagctgagttagctcctcatgtagaaaaatgtattacaatgcaattacaagatgcatttaataaaaggtcaaagaaaaaatttaaattcttttaaggagtaagtgtggtagaaattattctttctcatctcatatttgatcaagcttatttttcaaaagtttggccatatatggattcagaatattttgaaagtggtccagctaaaaatacattcaaattaattaaatctcatgttaatgagtaccatagcgttccatctattaatgcgttaaatgttgcattagaaaatagttcatttactgaaacagaatattctggtgtaaaaacacttatttcaaaactagctgattctccggaagaccacagctggttagtaaaagaaacagaaaaatatgttcagcaaagggcgatgtttaatgctacgtctaaaataatcgaaattcaaactaatgctgagcttcctccggaaaaacgaaataagaaaatgccggatgttggtgctattcctgacatcatgcgccaagcattatcaatttcatttgatagctacgttggtcatgattggatggatgactacgaagcacgttggctatcttatatgaataaagctcgtaaggttccatttaaactcagaattctaaacaaaattactaaaggcggagctgagactggaacactgaacgttttaatggctggcgttaacgtcggtaagtcattaggattgtgttcattggcagcagattatttacagctcggacataatgttctttacatttcaatggaaatggcagaagaagtctgtgctaaacgtattgatgctaatatgcttgatgtttctcttgatgacattgatgatgggcatatttcttacgctgagtataaaggaaaaatggaaaaatggcgtgagaaatctactctcggtcgtttaatcgttaaacagtatcctaccggtggagcagatgctaatacatttcgatcgcttttaaatgaattgaagctcaaaaagaatttgttccaacaatcattattgtcgactatctaggtatttgtaaatcttgccgcattagagtttattcagaaaatagttacacaactgttaaagctattgcagaggaattgcgtgctctggctgttgaaaccgaaactgttctttggactgcagcacaggttggtaaacaagcttgggactcttccgatgttaacatgagcgatattgcagaatctgccggtcttccagcaacagccgattttatgcttgcagtcattgaaaccgaggagctagcagctgctgaacaacaactcattaagcaaatcaaatcacgatatggtgataaaaacaaatggaataagtttttgatgggtgttcaaaaaggaaatcagaaatgggtagaaattgaacaagattctactccaactgaagtgaacgaagtagcaggttcacaacagattcaggctgagcagaatcgctatcaaagaaatgaatccactcgagctcagttagatgctttggcgaatgaattaaaattttag (SEQ ID NO: 422)
11	atgaaagaatttatatctctatcgaaacagtcggaaataatattattgaacgttatattgatgaaaacggaaaggaacgtactcgtgaagtagaatatcttccgactatgtttaggcattgtaaggaagagtcaaaatacaaagacatctatggtaaaaactgtgctcctcaaaaatttccatcaatgaaagatgctcgagattggatgaagcgaatggaagacatcggtctcgaagctctcggtatgaacgattttaaactcgcttatatcagtgatacgtatggttcagaaattgtttatgaccgaaaatttgttcgtgtagctaactgtgacattgaggttactggtgataaatttcctgacccaatgaaagcagaatatgaaattgatgctatcactcattatgattcaattgacgaccgtttttatgttttcgaccttttgaattcaatgtacggttcagtatcaaaatgggatgcaaagttagctgctaagcttgactgtgaaggtggtgatgaagttcctcaagaaattcttgaccgagtaatttatatgccatttgataatgagcgtgatatgctcatggaatatattaatctctgggaacagaaacgacctgctatttttactggttggaatattgaggggtttgacgttccgtatatcatgaatcgcgttaaaatgattctgggtgaacgcagtatgaaacgtttctctccaatcggtcgggtaaaatctaaactaattcaaaatatgtacggtagcaaagaaatttattctattgatggcgtatctattcttgattatttagatttgtacaagaaattcgcttttactaatttgccgtcattctctttggaatcagttgctcaacatgaaaccaaaaaaggtaaattaccatacgacggtcctattaataaacttcgtgagactaatcatcaacgatacattagttataacatcattgacgtagaatcagttcaagcaattgataaaattcgtgggtttatcgatctagttttaagtatgtcttattatgctaaaatgcctttttctggtgtaatgagtcctattaaaacttgggatgctattatttttaactcattgaaaggtgaacacaaggttattcctcaacaaggttcgcacgttaaacagagttttccgggtgcatttgtatttgaacctaaaccaattgctcgtcgatacattatgagttttgacttgacgtctctgtatccgagcattattcgccaggttaacattagtcctgaaactattc

TABLE 18B-continued

DNA sequences of fragments #1-60 in Table 18A

Fragment #	DNA sequence
	gtggtcagtttaaagttcatccaattcatgaatatatcgcaggaacagctcctaaaccaagtgatgaatattcttgttctccgaatggatggatgtatgataagcatcaagaaggtatcattccaaaggaaatcgctaaagtatttttccagcgtaaagattggaaaaagaaaatgttcgctgaagaaatgaatgccgaagctattaaaaagattattatgaaaggcgcagggtcttgttcaactaaaccagaagttgaacgatatgttaagttcactgatgatttcttaaatgaactatcgaattatactgaatctgttcttaatagtctgattgaagaatgtgaaaaagcagctacacttgctaatacaaatcagctgaaccgtaaaattcttattaacagtctttatggtgctcttggtaatattcatttccgttactatgatttacgaaatgctactgctatcacaatttttggtcaagttggtattcagtggattgctcgtaaaattaatgaatatctgaataaagtatgcggaactaatgatgaagatttcatcgcagcaggtgatactgattcggtatatgtttgtgtagataaagttattgaaaaagttggtcttgaccgattcaaagagcagaacgatttggttgaattcatgaatcagtttggtaagaaaaagatggaacctatgattgatgttgcatatcgtgagttatgtgattatatgaataaccgcgagcatctgatgcatatggaccgtgaagctatttcttgccctccgcttggttcaaagggtgttggtggattttggaaagcgaaaaaacgttatgctctgaacgtttatgatatggaagataagcgatttgctgaaccgcatctaaaaatcatgggtatggaaactcagcagagttcaacaccaaaagcagtgcaagaagcactcgaagaaagtattcgtcgtattcttcaggaaggcgaagagtctgtccaagaatattacaagaacttcgagaaagaatatcgtcaacttgactataaagttattgctgaagtaaaaactgcgaacgatatagcgaaatatgatgataaaggttggccaggatttaaatgtccgttccatattcgtggtgtgctaacttatcgtcgagctgttagtggtctgggtgtagctccaattttggatggaaataaagtaatggttcttccattacgtgaaggaaatccgtttggtgataagtgcattgcttggccatcgggtacagaacttccaaaagaaattcgttctgatgtactatcttggattgactactcaactttgttccaaaaatcgtttgttaaaccgcttgcgggtatgtgtgaatcggcaggtatggactatgaggaaaaagcttcgttagacttcctgtttggctga (SEQ ID NO: 423)
12	atgaaactgtctaaagatactactgctctgcttaaaaatttcgctactattaactctggtattatgcttaaatccggtcaatttattatgactcgcgcagttaatggtacaacttatgcggaagcaaatatttctgacgttattgattttgatgtagcgatttacgatttgaacggttttctcggtattctgtctctagttaatgatgatgcagaaatttcccagtcagaagatggaaatattaaaattgctgatgctcgttcaacaatttttggccagcagccgatccgagtacagtagttgctcctaataaaccaattccattcccggtagcatctgttgttactgaaattaaagctgaagaccttcaacaactgttgcgtgtatctcgtggtctgcaaattgatacaattgctatcacggtaaaagaaggtaaaatcgtaattaacggttttaataaagtagaagattctgctctgacccgtgttaaatattctttgactcttggtgattatgatggtgaaaatacatttaatttcattatcaatatggcaaatatgaaaatgcaaccaggaaattataaacttctgctctgggcaaaaggtaaacaaggtgctgctaaatttgaaggtgaacacgcgaattatgtagtagctcttgaagctgattctacccacgatttttaa (SEQ ID NO: 424)
13	atggaatattcaactggacagcatctattaactattcctgaaataaaacgatatattctgagaaataatttttctaatgaagagcatatagttactgaatctatgcttaggaatgcatttaaagcagaatatacaaaaataatgtccaatagaaatgaagcttggactgttactgattattatgactaaaggtgtattatgactaaaattactgtgaattatactgttgatgtaaaagatattcagccaaaacacgtgcgttctgaatcaaatccacaaaaccaaaataaaattcgtcgagcatgggttttgtctctttctgataacgcaatggaagttattcagaacaaaattaaatctgcacctgctcgtcatgcgtattatgaagctatcgatcgtgaagtaagtaataaatggattgaactaatgcgcaaacatactacagaatccctaaacgccggtgctaaatttattatgacttcatgtggtgaacgccttgaagatgattattgcggtaatgcagatgaacgtctaattgttgctgctcaaattgttgcggaaacaattgcggctgattttaatcgttaa (SEQ ID NO: 425)
14	atgaaagtatgtatttttatggctcgaggtcttgaaggttgcggtgtaactaaattttctcttgagcaacgtgattggtttattaaaaatggtcatgaagtaactttggtttatgctaaagataaatcatttactcgtaactgtgcgcatgattataaatcattttcaattccggttttattggcaaaagaatacgataaaacacttaagctggtaaatgattgtgatattctaattatcaattcagttcctgctacttcagttgaagaagacactattaataactataaaaaaattattgataacattaaaccttcggttcgtgttgtagtttatcaacatgaccattcttctctttctttgcgtcgaaatttgggattagaagaaactgttcgtcgagctgatgttatttttagccattctgataatggtgattttaataaagttctgatgaaagaatggtatccagaaactgtttctctgtttgatgatattgaagaagcgccgacagtatataactttcagcctcctatggatattgcgaaggttcggtcaacctactggaaagatgtttctgaaattaacatgaatatcaaccgttggattggtcgtacgactacatggaaaggtttttatcagatgtttgattttcacgaaaaacatcttaaacctgcaggactaagtactattatggaaggtctggaacgttctccagcgttcattcctattaaagaaaaaggaattccatacgagtattatcgtcttcatcaagtagaccaaattaaaattgctcctaatttaccaacgcaaattcttgaccgttatgtaaatagcgaaatgcttgaacgcatgagtaaatccggatttggttatcagttgagtaagttggacaaaaaatatctacaacgttctttagaatatactcatctcgagcttggtgcatgtggaacaattcctgttttctggaaatcaacgggtgataatttaaaattccgtgttgataatactcctttgacctcgcatgatagcggtatcatttggtttgatgaaaatgacatggaatcaacattcgagcgtattaaagaactgtcatctgaccgaactctttatgaccgcgaacgtgaaaaagcatatgaatttttgtatcagcatcaagattcaagcttctgctttaaagaacagtttgacattattacaaaataa (SEQ ID NO: 426)
15	atgactattcaaattaaaaacgccatcaattcttacgcatatgataaagtagtttctttgttagaaaaaggcgatattgtaactcctcaaattttggataaatgggaaaaagagcttcatcagacgatgaaacagaatgatcagaagattggacgcaatactgtccgtgaattgttggttcaatatatcttgtcagaatttgatgttaaagcttttggtgtagaatctaaagcttatcaaaagcatgaaatttccgataaaactattcgtcgcatgaaaaatcaacgcaagaaaaaatttgcagacctgaaaattactaaggtataattatgaacgaagctcttattaacgatttgcgtcttgctggatatgaagtaaatacaaatggcattggtttaattcaaattgaaggaaacggattcatccttgagtatgaatttagccaatggtggttatacgctaattacggtgaattaattgaatatgttgaccaatttgattcactagatgcagctcttggagcggctaagctgatgaattcttga (SEQ ID NO: 427)
16	atgttattgactggcaaattatacaaagaagaaaaacaaaaattttataatgcacaaaacggtaaatgcttaatttgccaacgagaactaaatcctgatgttcaagctaatcacctcgaccatgaccatgaattaaatggaccaaaagcaggaaaggtgcgtggattgctttgtaatctctgcaatgccgcagaaggtcaaatgaagcacaaatttaatcgttctggcttaaagggacaaggggttgattatcttgaatggttagaaaatttacttacttacttaaaatccgattacacccaaaataatattcaccctaactttgttggagataaatcaaaggaattttctcgtttaggaaaagaggaaatgatggccgagatgcttcaaagaggatttgaatataatgaatctgacaccaaaacacagttaatagcttcattcaagaagcagcttagaaagagtttaaaatgacaattgaaaaagaaattgaaggattg

TABLE 18B-continued

DNA sequences of fragments #1-60 in Table 18A

Fragment #	DNA sequence
	attcataaaactaataaagaccttttaaacgagaatgctaataaagattctcgtgtttttccaactcaacgggaccttatggctggtattgtgtctaaacacattgccaaaaatatggtcccgtcttttattatgaaagcgcatgaaagcggaattattcatttccatgatattgattattcccctgctcttccatttactaattgctgtttagtagatttaaaaggaatgcttgaaaacggatttaagcttggtaatgcacagattgaaactcctaaatcaattggcgttgctactgcaattatggcacaaattactgcacaggttgcttctcaccaatacggcggaacgacttttgccaatgtagataaagtactttctccttatgttaaacgcacatatgcaaaacatattgaggatgcagaaaaatggcaaatcgctgatgcgttgaattatgctcaatctaaaacagaaaaagacgtatacgatgcattccaagcttatgaatatgaagtaaatactctctttagttcaaacggacaaacgccttttgtaacaattacatttggtacgggaactgactggactgaacgaatgattcagaaagcaattctgaaaaatcgcattaaaggtcttggccgtgatgggataactcctatttccctaagcttgttatgttcgttgaagaaggtgttaatctttataaagacgatccgaactatgatattaagcagcttgctttagagtgtgcaagcaaaaggatgtatcctgatattatttcagctaagaacaataaagctatcactggttcatctgttcctgtttctccaatgggttgccgtagtttcttgggcgtatggaaagattcgactggcaatgaaattcttgatggacgtaataatcttggtgttgtaacactgaatcttcctcgcatcgcgttagattcttatattggaacacagttcaatgaacagaaatttgttgaattgttcaatgaacgaatggatttatgttttgaagctttgatgtgtagaattagttccttaaaaggagttaaagctactgttgctcctattctttaccaagaaggtgcattcggggttcgtcttaaacctgatgacgacataattgagttatttaaaaacggtagaagttcagtgtctttaggatacattggtattcacgaattgaatattcttgtcggtcgtgatattggacgagaaattttaactaaaatgaatgctcatcttaaacagtggactgaaagaaccggatttgcttttagtttatattctactcctgctgaaaaccttttgttatcgcttctgtaaactcgatacagaaaaatatggaagcgtaaaagatgttaccgataaaggctggtacactaacagtttccatgtttcagtagaagaaaatattactccgtttgaaaagatttctcgtgaagcgccatatcatttcattgcgacaggtggtcacatttcttatgttgaacttcctgatatgaaaaataacttaaagggtcttgaggccgtatgggattatgctgcacaacatttagattattttggtgttaacatgccggtagataaatgttttacatgtggaagtacccatgaaatgactcctactgaaaacggatttgtttgttctatttgtggagaaactgatcctaaaaagatgaacacaataagaagaacgtgtggttatttgggaaatccgaacgaacgtggatttaatctcggcaaaaataaagaaatcatgcatagggttaagcatcaatgaattatgatagattttatccttgcgattttgtgaatggccctggttgcagggtcgttcttttcgttacaggttgtttgcataaatgtgaagggtgttataataaatcaacatggaatgctagaaatggtattccattcactggtgaaacactagaacaattaattgaatgtttgaataatgattatatagaaggattgactataactggaggagaccctctctatccggataatcgagatgtcattcattgcattgttcaaacagtaaaaaatctttatcccaataaaagcatttggttgtggacaggatataagtttgaagatattaaacaactagaaatgcttaaatatgttgatgttattattgatgggaagtatgagaaaaatcttccgactaaaaagctgtggcgaggatcagataatcagcgactttggtcaaataccgatggggtgtggaaacatgattaaattgaattacattatggatactataaatgatatgatttttcattttggtccagaattttattcccaatatagtttagtgcttatcaatgcttggttaattaattaagggtaaaatatgtataaatttcgtaaaggtttagctgattttcttacaactgtaacattctttctgtttatggcagttggagctattttccttattccttttattgctatattttcgtgattagtttaatttctccagaaaagggcttatcttccagtgagttcaatgagcgcctggataaaattactaacaagctgaatgctgctcttagtaaggaatagttgtgaaacaaaataagattgaagtctatggaattccagatgaagtaggtcgttgtcctggatgtcaatcagttacaaaacttctaaaggagctcaatgctccttttactttctataaagttcttacaaataatggtaagattgagtatgatcgtccactgattgtatctcttgctaaacgcgctggattcacatctcttaacattcgttatccagtcattttcattaatgattctagacaaaagaacattaaacacttcaaagaaactctcatttcacttggatatgatagagatatcatagaagattaa (SEQ ID NO: 428)
18	atgaaacagttgataattaaaagattgaatttattgatatgttgtttatgtatagtaattgcatatggttattacgcaattaatgattatatgcattataaagattatgatgttactgtagttaatacccttacaggaactcaaggaaaggggtctagtttatcgtttattgccgtatatgaactcaaagacggttatagatttagcgaatatatttcgccagagatgtattcatcaatagaaaaaggcgataattactgtaagtttacgtcctttcgacgtaaaacagacattgtttgataatattgtttggttctttggaatggtattagttcaatctatatgtggtacttatatagtctgttcaatcttattccgcgtaattagtaaaattgagtgaggaaaatatgtcagtagtaattaataatgtcaatgcagtaattaaatctttagttaataaaaaaatgatgaatgaatggactgtacttcgtcgtggagagccagataaatttttttcatagatttaacccaactttggatttgaatgttattgacagagatgttcatgctgaaattttagataaatttaaagttgatattggatttggattagaaaaacatttacagcgaacaaacgggtctggaatgagtttatctaatcgcatcatgaaagcccttaataaaattggagcattgtctcgtattaacgcgagtgaaatccttcgtaattataataaaggatatgacctttatggccgactaatgccgaaattatcattcgatcaaatgattgcggatttgtgggaaaatcaacgacgattattagcattaggcgctcgattagctaaaggtctagataaacaaatgattttaaaactaataatacagaagaccttaaatgctttaaatttagtactcgtggagatgattattacgtcagagctcgctctacagattatgtcaatatggggcatcatctctgtttagcttttgaagttttaaaagaagctTgaacgttagaatattcatctggtgctaaatgcccgattggttcaaattgcattttaatttatcgcccgaatgaatccagttcaactaaattgcctacaaaacctgtaccagttcgtagtaacgaaaaacattctgaacaaattgattattttaataaacagattgaagagctgaatatttctattcaacaatatgacgatgaaatctttagactatctggattgagtagtaaagctaaatctgaacgtgaaaaattaattaaaattgttgatttacttaaatcttaaggaacaccatgaaaactcgttctcaaattgaagatatggttcgtaatgccagctatactcgtgatgctatgacatttttgtgtgaaaataatttagaccttaataaagttaattgttccattcacgcctttaaacatctgaacagcagtgaatgggtgcgtaattttaatgaagcagggtatattacacaaatgactgctcgtgagcagctcgttgatttctgtaaaactattgattataaaaatcctctatttgttcaaggcgttggtcagagtaaggttgatttatcaacaggatttttaatccaaatcattatcgtcttgaatggagatttattgctctattccgtaaacaattaaagcaaattttgtcgactgctagtcgattaaaaggttctgatattaacttaaagaatctgaaatttgatggttatactcttcagatggaagtaagaccattaaaagaaaataatagaactgcacgaattagctttaaacctaatacaaaaaattctctttcaatttgtgaatgccttaaatcacagttgatagaagcatttaagtatatggatgttgttgctagtgttcagtctaagatttcacagcatttcgaacgatttaaattaggcacaacaacgtatgaacttgatatggtcgttttatttaaatacgatttttgagaaaggacgaagttgtacaagagaaaaagcaggaagtgcaagataacttaaatttatctaattacttatcaaacgatcctaaattttggatgtatagttcaggtaataaagatgcatggaaattcaataaagtgaattttcttcctattgaaaatccgagtcttaaacctgttgaaaaatggcacgcggatgcgattgagaagtctatcaaggcagtagatgatgaactcgttaaagcaactaatgaagtgttagaagctgaaaagatgctagaaaaagcacaagaaaaagtcaaaaatctcacgaagcaacgttctaaactgaacaatgcactaaatgcactgaactag (SEQ ID NO: 429)
19	atgaacgctaaagatattttcaacctggtaaattacaacgatggtaaatttaaatctgaagcacaaagcaagttctttaatgacatctcaatcggaggtgaaatcacagttgatggaggacaaatttacaaatcccgttggaattggatcgttattatcgatgagattggtattgtagaaatttacaaaaatacgaataaaaatcgtacattacactggtctcgtgatactaacgaacagtacaaaaaggataaagcatctaagttatctcgtgtaactcaagaagatattgagttcatcaagaaagatattttgatgtatgataacttaattgctgaagagcaagctgttattgataaatttgacgagattaaagcttc

TABLE 18B-continued

DNA sequences of fragments #1-60 in Table 18A	
Fragment #	DNA sequence
	tcgtgaaattcctgattttatgaaagaatcagtaaatgaacgatacactctcatttcagagcgtattgaaacttacaaaaagcaaagagctgaacgccaaaatactcttcggaagtttgaagaacggttaaagacggtactcgcataa (SEQ ID NO: 430)
20	atgttatactcaaaggctcgtgaaatttacgaaactaagattaaagaagctgtatttcaattcgcaacaacgatgcgatggacaaatgattgggaatattcaaaaaatcataagaagcccctggtgacaagaaaggctcatatgttagtgttaatagaccgtgagcagattaaagcccgagaagccctccagaatcataaaaaggctgcctttgaatggtttatggataacactgctcctgagactaagaaagcagtgagcgcgtggttcagtggaaaaaattgtgaaagaagtttcttttag (SEQ ID NO: 431)
21	atgaaagttttgtttgttgtgtatgtgatgattcaatataattacccaatgtttacttataatttggtgaataacattattgatatgattcagaggagtatgtaattatgagtgagtcgaagagaatcaatatgaaacgattagtattagaagatagtgtgctttttggtgaattagcgatcgaaaaagtaaataacatgtatcgtttgacgcaagaagatgatatgttatattacgcctagtgaaattgttcgtttaacccaaattgaatatgcttacactgataaaattgtaagcattaatgatgagcataaaattcatttttattcttcatgcccaggatttaatattaaaagcgagtcaatgtgcttatcaattaataattgggataattttataactaacattaaatatttttatgattctactaaaagaaaacataatttaaaatggtttaaaaatgtaatgctattattactaactcctgtaatcagaatgatgaaactatttttaaatgtttcaaaatgctatgaagagggagatgtagtatctattcgtcaaattgacgattttcgatcgcatatcattacattaaacaaagacgaagctattgcactaaagacttatcttgattctgttattccaactatgatttcaaagtgaggaaatatgtttatttcaagtggaagtggtttaattcgtgttgaatttaaaaatgacatcttccttagtcaaggagatgatattattaaaatgagttatgacgaaatcaagaaaatttgtcatactcttgaaagccgtggaaaagtaaatgctgttttgacattggtgatttatgggtaacgctttatgaagtatccgaaggatttaacattgaagatgaaaataacatttttagctattgataaaagaactgatttgcttgatgtattaaaagcctatgaacagtcaaacggtggaagaaaagctgtattgatttatcaaaaaccgcattcatgtggaactgcttcaatcatttcaaatattgaaggcgaagttgatacttatatgtgttttaaaagctggtggtgaccgtcatccggatttttatttctattcgtcaaaacaatggagaaatttcattatcaaaatcagaagctgaagctatgattaagtatttaacaaccgttacgccttcaatgaaaggataattatgattattaatgaaaactcttggcactataaattattcaaactgtttaacgatgaatggcaacgacctaagacactatgcgcatatttttggtctattgcctcctacatttttcgtttctatttttgggtgtgctatactcgtagggctaacaattatttgtgcagaaagcctacaacgttggcttattttcggtagtttatggactcttcttccatcggcatttatacttgcgcttttggttgttttacttattatcggttcatttgttattcctgcacatttgcgtgaaaaatataaagattataaatggaaaaaggattatgctttacacgtagaaaatattgatagggcgtataaaggtttacctcctattcaacctaagaaatcgattatcgttgaatttttaaaagttcgtaaagctaaagtatgtcctgttattgaatataaggctgaatgatgaaaacagtaatgaaaagctatttggtagtcatctttatggaacttctaccccagaatctgatgtagattttaaagaaatttttgttcctcctgctcgcgatattcttatcggaaatgtcaaagagcacatgagcaaaaacactaacaacacatcatctaaaaacactaaagatgatattgaccatgaactatacagtcttaaatatttctttaaattagcagcagatggtgaaactgtagcgttagatatgcttcacactcaacctgaactagtggttaaatctgatttgcctgatgtgtggaagtttattcaagacaaccgttctcgtttttatacgactaacatgaaatcatatttaggatatgtccgtaagcaagcttctaaatacggtgtcaagggttctcgtttggctgcattacgtgatgtattgaaagtagttaatcaaatccccgagcaatgggttgattaccaagaagatggttctattaagcagcgtcgtactaaagttgaagatattaagcatcgtcttccagaaaacgaattctgtgaatgggtgttccataatcatgagaaaacaggcccacaaacgttctacactgtattgggtcgtaaatatcagacaacgctttctcttattgagcttaagcagtcactgaacaaattagatgctgaatatggtgaacgtgcccgtaaggccgaagccaacgaaggcattgactggaaagctctgagccatgcttgtcgtggtggacttcaactattggaaatttacaaaactggtgacttggtttatccacttcaagacgctccatttattctcgacgtgaagttgggtaaacatccatttaaaacggttcaagagtttttggaagatgtggtcgatcaagtagaagcagcatctactgaagcttctaagaacggtatgcagcaaaaagtagacatgggtttctgggatgacttccttgagaaggtttatcttgaaaaccaccgaagttattataaatga (SEQ ID NO: 432)
22	atgctacaattaactgaaaagcaacttcgcaatcttactgtgcttcaattagatgaaattcgtagggaagttggaaatatcatttcagctttgcgtcgagaagtatcacttaaccaatctccggcagactatactagattgcgaaattttgaaaaataccttgataaagttaaggccgtgcatcggcataaagtaaatacaggacaaaaatgataggaggcctttatggccttaaaagcaacggtactatttgccatgctaggattgtcatttgttttatctccatcaattgaagcgaatgtcgatcctcattttgataaatttatggaatctggtattaggcacgtttatacacttttttgaaaataaaagcgtagaatcgtctgaacaattctatagttttatgagaacgacctataaaaatgacccgtgctcttctgattttgaatgtatagagcgaggcgcggagatggcacaatcatacgctagaattatgaaaattaaattggagactgaatgaaattcagcgacttttcacaaagtggaaaaccttcaaaggcagatgaatacttaggtttattaatggctgcacaagcttattttcattctgcacattttgaaactaaaagttatgctagacacaaagcatacgatttttatttttttccgagttgccagatttgattgataaattttgtgagcaatatttggggtattctggtagaaaatacacaccttcaattccagatgccagtaaacttcctaccgacacaattaaaatgattgatcgcatactagaccaatctaacagcatttataaagaaatgcctccagccatccaaagcacgatagatgatattactggaatgttttaccagagtaagtatcttctttccctcgaataa (SEQ ID NO: 433)
23	atgaaaacctatcaagaatttattactgaagcagctattaattctcaaattattgctgaatcttttactgatcttttgaaatttaaaaaaggtcagaaaatcactgctgtattggatgatggtacagaagttgagatggatgtacagggatataattatgcagtagatggaaaactgtataataaatctcatgctaaatttgattcatttgacgactttgttaatacagttgaagatgaaaaaactcgtcgatccattgcaactggtgatgctaaggttcttatggcacatggtcatgaacgcattcgcgctaaacagaataaaatgggtgaagataatttcgcattagttggttatcaatctggtaaacaaacttatggctatcaacgtactgctaccatgtataacaaaaatggtaaaattgcctttgtgaatagtaaaggttctattcagtacgttaaatcgttcaaataacatgggaacaacctggacctcatgattctgtgagggattcccgccaacctgtaataatgtcgagcccaagcgcggtaatgggtaaatacagaaatggacaattcatgcgccatggaatggcccaaatttagagagaagaaatgagaacatttttaactggtccttatctatccctgatgaatgcttttacacaccattctgatgctagagtagaagaaatttgtaaaaacgaatatcccgccatttgaagacttacttaaacagtattgcacacttcgactagatggtggacgtcaatctggtaaatcaattgctgtgactaactttgctgctaattggttgtatgatggcggaacagttattgttctttctaatacttcagcttacgctaaaatttctgcaaataacatcaaaaaggaattttcgcgttattctaatgatgatatacgttttcgtttatttactgattctgtgcgcagttattggtaataaaggaagcaagttcagaggtttatcgctttcgcgaattttgtatataattgatgagcctgtcaaatctcctgatatggataagatttatagtgtccatattgacactgtacactgctgctgtaatattaaatgttgtattggtggtattactcgtccacagtttttcgtaatcggaatgcaatga

TABLE 18B-continued

DNA sequences of fragments #1-60 in Table 18A

Fragment #	DNA sequence
	tgacagacactcagcttttcgaatatctttatttttcgccaaaaactattaaaaataaattggtgaatcattttgaaattttggcaaaaaataacattttaagcgaattttatcctaagcaatacaaattacaaaaaggcgtattcaaaggatgcagagttttgtgtactgctcctaatgcacggctaatgaataaaattccatattttaccatggaatttattgatggaccttttaaaggattaattacccacagtttaatggcatatgattctgagccatttttaattaaagaacaatcttggataaatttattttctaattgaggtttatatgaaagcatatcaaattcttgaaggcacacataaaggtactatttattttgaagatggtattcaagcacgaattattgtctctaaaacctttaaagaggactcttttgtagacccagaaattttctatggtttgcatgcccgtgaaattgaaattgagccacaacctacagttaaaattgaaggtggtcaacacctgaacgttaacgttctgcgtcgtgaaactctggaagatgcagttaagcatccggaaaaatatccgcagctgaccatccgtgtatccggttatgcagttcgctttaactctctgactccggaacagcagcgcgacgttatcgctcgtacctttaccgagagtttgtaatggcaaagataattattgaaggttctaaagatgtgataaatgctttcgccgagtggtttagtaattcaggcgaacagcaatttaatgaagcctggaatatgggtgatattgatggaatttatcctacgacagaagtttctgttcagggatatggcattcatgaacctattcgtttagttgaatatgatttatgtactggtgaggaagtcaaatatgattgaagatattaagggataaaccacatactgaagagaaaatcggtaaagtgaatgctatcaaagacgctgaagttcgtttaggacttatctttgatgctttatatgatgaattctgggaagcactagataattgtgaagactgtgaattcgcgaagaattatgctgaaagcctcgatcagttaactattgctaaaacgaaactcaaagaagccagtatgtgggcttgtcgtgcagtgttccaaccagaggaaaaatactaatggatcaattaagcgcagggtttggttatgagtattatactgcacctcgtcgtgtatctgttgctcctaagaaaattcaaagtcttgatgacttccaggaagtagtccgtaacgctttccaggactatgcacggtatcttaaagaagattcgcaggactgtctcgaagaagatgaaattgcttactatacgcagcgtcttgaacagctcaaaaatctacatgaggttcgtgcagaagtttcaaagtctatgaataaattgattagatttaaagaataa (SEQ ID NO: 434)
24	atgaatatatttgaaatgttgcgtatagatgaaggtcttagacttaaaatctataaagacacagaaggctattacactattggcatcggtcatttgcttacaaaaagtccatcacttagtgttgctaaatctgaattagataaagctattggacgtaattgcaatggtgtaattaccaaagacgaggctgaaaaactctttaatcaggatgttgatgctgctgttcgcggaattctgagaaatgctaaattaaaaccggtttatgattctcttgatgcggttcgtcgttgtgcattgattaatatggtcttccaaatgggagaaaccggtgtggcaggatttactaactctttacgtatgcttcaacaaaaacgctgggatgaagcagcagttaacttagctaaaagtagatggtataatcaaacacctaatcgcgcaaaacgagtcattgcaacgtttagaactggcacttgtgacgcgtataaaaatctataa (SEQ ID NO: 435)
25	atgaacacactgaagaaaattgttgagtttattcgcactaaacttggttctgctatggctaaaaatctatctgttgaagaacagtatactgccgcagcagcaaaactgcttgataaaattaaagacctaaaaactgcttctgttaaatctattaatgaagaaaaacgtattcgtgaacttattgttgaaaagaataaacaggctgaatcaaaagagcgtgaaattcgcaagcttctttccgaaggtcaagatgtaacaatgcatgctaaactcggtttgctatatcgtcgaacagctgaacagctgactactaaagctgatggttatgctgaaatgcgaattgaaatcgctaagaaagtagttgagttagatgatgctcgccaagaacttgcagttaaattggaatatatccgtgaaactcgtgcagcaaatgcccttggaattagtactgctgatgatgtagttgaaattgcagcactgactaaggttgatattgaagatactcttgctcgagttgaaacctttaatggcaatatttctggggttgaaactacctctgccgatgttcaggaatatattaattctctgaaataa (SEQ ID NO: 436)
26	atgactactttaattatttggttcgacgaaaatgaagaaacatattgcgtgaacattggcgaaagcccaatgccagaatttgaatcttcagataaaaactcggttgtatcttgggctgaaggttataaagcagcaaaaggcgatgttgaaatagtttacaaactatccggagtataa (SEQ ID NO: 437)
27	atggataattacggtgaactgttcaacttctttatgaaatgtgtttcagaagatttcggtcgtacagtgaatgatattaaagttatcggtcctgaccatccgatgtttgaaacttacgcagtaatgggtaatgaagatggtcagtggtatactgtaaaggtcgtgattaacatgttcactgctgaaggttatgttaaactgtcttctaaagtttaccatgataacgacgaaatcgcagaagaatatttcaataatatgaaataa (SEQ ID NO: 438)
28	atgaaaggtaatgtttatttagtcgttcatgatttaacattctattttaatcataatgacactgttatttctgaacgtgtaattaatttgctttatcagcatgcagactatgtttatgtcgaaaacgaatttaggcattggcaatttctcaaaaatcgttcatttggtttagatggttacgaatactttgaacgtaaagaccttttagataaaattccattatctacacaataccaaaatcacaagtctttacataaatgccggctaattcgaaatgctgaatccgcgtatgaagcaattgatttatggcgtaaacgccgtgaacagattgatgctttaaaagaatattaa (SEQ ID NO: 439)
29	atgaatggctattggtggaaatcaacgggaaaatatgataagcgtggaagaaagggtcatgaatactgcatgtgccgtttcggtgataaaggaccatattcattaaataacatatattgcgcaactaataatcaaaatacaaaagatgcgagactaaatgatagatttcctccaaaatctaaaaattttaattttaatggtcgaaaacactcggcacagtccttagaaaaaatttctaaaaataatgcaagtaccttaagcaaagatgagataactagacgattaaaaatattagaaaattttaatatggatgaacgaggttttattaaaaattatgcaaacgctataaatgttagccatactcaagctagaaagtttttaaataaatattacataaaataa (SEQ ID NO: 440)
30	atgaaacgttgtgaattaattcgaaatgttgctattgcaatttctgcttccgcttttagtttttcaatgtttgttggatttatatgcggattattgactacagcagaaaatgtgttttcacttgtagtagcatttttaattggtttaatcgctatcgttatggataaaatttctaaaggttaataatgattctttatgcgaaagtatcgtccgttgaaaatggatataaatatgatcaagatgcggctaaagccttgattgatgattatggcattttaacatgttttgaagttgaaaaggtttacattgaccgttcatcttctcaagttaaattagtgaaggaagaccgtaaatttaatacagtaaattttgatttctttattgaaacagaaaaaggtcctcttgaatatgatattttcaagaatcctttgggtcttgaatgtattaaatatacttacattaatatggtgaacaaatgtatattcgtttaggcagcacaattcctaagggttacgtaattgatgtcactacctgggaaaatgatggtgataactataaaaccaaaacactgtttggcgtagaagagcatgagctccaacaatttaaatatcttttgaagaagtttaagagtcgtcattctagcactaaagctgaccgttattgtggtaatgggttgttcagcgagcaagagcttttatatatgaatatttggttgaaggactgttctcagaccaactttatccagaattcattaaaaaggtctttgatatagaagttga

TABLE 18B-continued

DNA sequences of fragments #1-60 in Table 18A

Fragment #	DNA sequence
	ccttggtaataaatccgaagaagatgaagaacgtgtatttgacttattctttgtgaatggtaataagatatttgaaggcctcattgatattcttggtcatgcttctgaatactatgaatatgatttcttgc gtgtagttgaacatgtagaatttgcttatatcgaagaagaaattgttttgccgactgttaaaatggttgatttgctttaa (SEQ ID NO: 441)
31	atgaaaacatttaaagaatttatcaatgaagcggctgcgccaaagacattcgttattaatactcagacgagtcttgacgatgagtatgcagaggcaattctgaagtcacttgctaagaacggcg ttgaagtaatcgcctcggactttaagaaaggggcttccgagatgtttatttctataactaaaggatctaaagctaagatcaaatcatcattcggagttgctcgtaccgatcaaatcgacaatcatg actttaaacaaactggtgctaaacggcagaatacaattgcatcacgcggaataaaatag (SEQ ID NO: 442)
32	atgaaaactttcaaagagtttgctacaaaaactactattactgaatcttcccatggtatggaagtaaaacttggaatggctttagctgaagctgagcgtcttttctctcgtattaaagaacttgctgc tgttgatccttcatcttttaaaggagaccaaactaaagttaaagcgcttttagcattatgctctgatgcaggcgaaatcgctaagaacggttctaagatgaagaaacgattagaagatttaaaata a (SEQ ID NO: 443)
33	atgaaactaatctttttaagtggtgtaaagcgtagtggaaaagatactactgctgattttatcatgagcaattattctgcagttaaataccaacttgctggtcctattaaggatgcattggcttatgca tggggagtatttgcagcaaacactgactatccttgcttaactcgtaaagagtttgaaggaattgactatgatcgtgagactaatttaaatctgactaaattagaagtaatcacgattatggaacaa gcattttgctatcttaatggtaaaagcccaattaaaggtgtgtttgtttttgatgacgaaggaaaagaatcagttaatttcgtagcatttaacaagattactgacgttataaataatattgaagatcaa tggtcagtccgtcgtctgatgcaagccctaggtacggatttgattgttaataacttcgaccgcatgtactgggtaaaattatttgctttagattatcttgataaatttaactcaggttatgattattatat cgttcctgatacccgtcaagatcatgaaatggatgcggctagggcgatgggtgctacagtaattcatgtagttcgtcctggtcaaaaatccaatgatacacatattacagaagctggattgcca attcgtgatggcgatttagtaattacaaacgatggttctattgaagaactttttctaaaattaaaaatacactaaaggtactataatgtctgaacaaactattgaacaaaaactgtctgctgaaatc gtaactctgaaatctcgcattcttgatacacaggaccaagcggctcgtctgatggaagaatccaaaattctgcaaggaactttggctgaaattgctcgtgcagtaggtatcactggcgatacc atcaaagttgaagaaatcgttgaagctgtcaagaatcttactgctgaatctgcagatgaagcaaaagatgaagaataatggaatttaaagacttttcaacgggtctttatgtagcagctaagtttt cagaattaacacttgatgcgctggaagaactccagcgctctttacgtgttcctaatccagttcctagagaaaaaattcattcgactatatgttattcaagagtaaatgttccatatgttccatcgagt ggaagttttgaagtagcttcttctggacatttagaagtatggaaaacacaagatggatcgactcttgtacttgtgctagattctgaatatctgcgctgtcgacacatgtatgcgcgggcattaggt gctacacatgattttgatgattacacaccgcatataacattgtcttataatgttgggcccctatcatttagcggtgatgtacaaattccggtcgtattagatcgtgaatacaaagagcctcttaaact cgattgggcagatgatttaaaataa (SEQ ID NO: 444)
34	atggcatattctggaaaatgggttcctaaaaatatatcaaagtatagaggtgaccctaaaaaaattacgtatagatcaaattgggaaaaattctttttttgaatggttagataaaaatccagaaatta ttgcatggggtagtgaaacagcagtaattccttattttttgtaatgcagaagggaaaaaacgtagatacttcatggatatttggatgaaagattcttctgggcaagaatttttttattgaaataaaacct aaaaaagaaacacaaccaccggttaaaccagcacatctaacaaccgcagcgaagaaaagatttatgaatgaaatttatacatattctgttaataccgacaaatggaaagcagcacaatcttta gctgaaaagcgtggaataaaatttagaattctaacagaagatggattacgagctcttggctttaagggggcataatggctatttttcaaataattaatgaaagcactccccaagttccaaaggtt aagcaatcattaaacgaaaagaaatggattcagataggtcttgaatacaaaaaggccaaagcaaaaggaatgacaggaaagcaatttgctgaagaaagaggaatcaaatactctacgttta cttcagcaatgtcaaaatatgcttcaggaattaaaacggctgaaaagattcaaaaacttgaatcaaaaccaatgaataaactcaataagcaagaaagacaactgcttatgataaattcattcag acaaacattgcgtgataaaattcgtaatgaaggtgcagcaattaataataaaaccagaaagtggtttgccgaaactattaagcaagtaaaaggacataaagttgttcgcccgcagccgggac gaatatatgcttttgcttatgatgctaaacacaaggaaactcttccttattgggataaatttcctttgataatttaccttggtttaggtaagcataatttaatgtacggattgaacttgcactatattccac ctaaagctcgtcagcaatttctagaagagcttttaaagcaatatgcaaatacacctactattactaataaaacgaaattaaaaattgattggagtcaagtgaaaggatttagaggtgcagatcaa atgattaaggcgtatatacctggtaatattatgggtagccttgttgaaatcgccccgaaagactgggcgaacgttgtgttgatgccacttcagcagttcgtttcaaaaggaaaacgtttctctgc aaacaaagtctggtcaaatatctaa (SEQ ID NO: 445)
35	atgttcattcaagaaccaaagaaattgattgataccggcgaaattggtaacgcttctactggtgatatcttattcgacggtggtaataaaattaatagtgattttaacgcaatttataatgcgtttgg cgatcagcgtaaaatggcagtagcaaatggcactggagcagatggtcaaattatccatgctactggatattatcaaaaacactctattacagagtacgcaactccagtaaaagttggcactag gcatgatattgatacctctactgtaggtgttaaagttatcattgaaagaggcgaacttggcgactgcgttgaatttattaactctaatggatcaatatcagttactaatcctctaacaattcaagctat tgattcaattaaaggtgtttcaggtaatttagtagtaactagcccatatagtaaagttactttacgctgtatttcatctgataattctacatcggtttggaattattctattgaaagtatgtttggacaaaa ggaatcaccagctgaaggtacatggaatgtttctacatccggatcagttgatattccactatttcaacgcactgaatacaatatggctaaattgctagttacgtgccaatcagtagatggaagaa aaattaaaacagcagaaataaatattcttgtggatactgttaattcagaggtaatttcttctgaatatgctgtcatgcgagttgggaatgaaaccgaagaagatgaaatcgctaatattgcatttag tattaaagaaaactatgtaacggcgactataagttcttcaactgtcggtatgagagcagcagttaaagttatcgctacgcagaaaatcggggtggctcaataatgaaacaaaatattaatatcg gtaatgttgtagatgatggtaccggtgactacctgcgtaaaggtggtataaaaataaatgaaaactttgatgagctttattatgaactcggtgatggtgatgttccatattcagccggtgcctgga aaacttataatgcttcatcaggacaaacattaacagcagaatggggaaaatcatacgctattaatacatcttctggaagagtgactataaatcttccaaagggtacagttaatgattacaacaag gtaattagagctagagacgtatttgctacatggaacgtcaacccagttacactagtagctgcttccggcgatacgattaaagggtctgcagtaccagttgaaattaatgttcaattcagcgattt agaactagtgtattgtgccccaggacgttgggaatatgtcaaaaataaacaaattgacaaaattaccagttcagacattagtaatgtagctcgtaaagaattttagtcgaagtccaagggcaa

TABLE 18B-continued

DNA sequences of fragments #1-60 in Table 18A

Fragment #	DNA sequence
	acagacttttagatgttttcagtggaactagttataatgtaaataacatcagagtaaaacatcgtggtaacgaattatattatggcgatgtgtttagcgaaaacagcgattttggctctccaggcg
	aaaatgaaggagaactggttcctcttgatggatttaatattcgattaagacagccttgtaatattggtgacactgttcaaattgaaacatttatggatggtgtatcgcagtggagaagttcatatac
	aagacgtcaaattagattgttagattcaaaattaacgtcaaaaacttctctagaaggaagtatttacgttactgatttatcaacaatgaaatcaattccattttctgcttttggattaattccaggagaa
	cctattaatcctaattctcttgaagttagttttaatggaattttacaagaattggctggaacagttggaatgccattatttcattgtgttggtgccgattcagacgatgaagtagaatgctctgttttag
	gtggaacttgggaacaatctcataccgattattcagttgaaactgatgaaaacggcataccagaaattttacatttcgatagagtatttgagcatggtgacattatcaatatcacctggtttaataa
	tgatttgggtacattattaacaaaagatgagattattgatgaaactgataatctctatgtatcgcaaggaccgggagtagatatttccggtgatgtaaatttaacagactttgataaaattggttggc
	caaatgtagaagcagttcaatcttatcaacgcgaatttactgctgtttcaaatatctttgatacgatttatcctattggaactatatatgaaaacgctgttaatccaaataaccctgttacatatatggg
	attcggctcatggaaattatttgggcaaggaaaagttttagttggatggaatgaagatatttcggaccctaactttgctctaaataacaacgatttagattctggtggaaatccttcgcatactgca
	ggcggaacaggtggttctacttctgttacattggaaaatgctaatcttcctgcaaccgagacagatgaagaagttctaatagttgatgaaaatggatcagtcattgttggtggatgtcaatacga
	tccagatgaatccggtccaatttatactaaataccgtgaagctaaagcatctactaactctactcacactccgccaacatcaataactaacattcaaccatatattacagtttatcgttggataagg
	attgcataatgagtttacttaataacaaagcgggagttatttcccgcttagccgattttcttggttttagacctaaaactggcgacattgatgtaatgaatcgtcaatcagtcgggtcagtgacaat
	atctcaattagcgaaaggattttatgaaccaaacatagaatcagctattaatgacgttcataatttttctataaaagacgttggtacaattattactaataaaactggtgtttctcctgagggtgtttct
	caaactgattattgggcattttctggaactgtaacagacgattctcttcctccgggttctcctgttacggtattagtatttggtcttccagtttcagcaacaactggaatgacggcaattgagtttgtt
	gcaaaagttcgtgttgcccttcaagaagctattgcatcatttactgctatcaactcatataaagaccatccaacagatggtagtaaattagaagttacttatttagataatcaaaaacatgtattaag
	cacatattctacatatggaataactatttcgcaggaaattatttctgagtctaaacctggctatggtacatggaatttattaggcgcacaaactgtaactttagataatcagcagactcctacagtat
	tttatcattttgagagaacagcatgagtaataatacatatcaacacgtttctaatgaatctcgttatgtaaaatttgatcctaccgatacgaattttccaccagagattactgatgttcaggctgctat
	agcagccatttctcctgctggcgtaaatggagttcctgatgcatcgtcaacaacaaagggaattttatttcttgccactgaacaggaagttatcgatggaactaataataccaaagcagttacac
	cagcaacgttggcaacaagattatcatatccaaacgcaactgaagctgtttacggattaacaagatattcaaccgatgatgaagccattgccggagttaataatgaatcttctataactccagct
	aaatttactgttgctcttaataatgtctttgaaactcgtgtttcaactgaatcatcaaatggggttattaaaatttcatctttaccgcaagcattggcaggtgcagatgatactactgcaatgactccat
	taaaaacacaacaattagctgttaaattgattgcgcaaattgctccttctaaaaatgctgctacagaatctgagcaaggtgtaattcagttagctacagtagcacaggctcgtcagggaacttta
	agagaaggatacgcaatttctccttatacgtttatgaattctactgctactgaagaatataaaggcgtaattaaattaggaacgcaatcagaagttaactcgaataatgcttctgttgcggttactg
	gagcaactcttaatggtcgtggttctacgacgtcaatgagaggcgtagttaaattaactacaaccgccggttcacagagtggaggcgatgcttcatcagccttagcttggaatgctgacgttat
	ccaccaaagaggcggtcaaactattaatggaacacttcgcattaataatacgcttacaatagcttcaggtggggcaaatattaccggaacagttaacatgactggcggttatattcaaggtaa
	acgcgtcgtaacacaaaatgaaattgatagaactattcctgtcggagctattatgatgtgggccgctgatagtcttcctagtgatgcttggcgttttgccacggtggaactgtttcagcgtcaga
	ttgtccattatatgcttctagaattggaacaagatatggcggaagctcatcaaatcctggattgcctgacatgcgcggtctttttgttcgtggctctggccgtggctctcatttaacaaatccaaat
	gttaatggtaatgaccaatttggtaaacctagattaggtgtaggttgtactggtggatatgttggtgaagtacagaaacaacagatgtcttatcataaacatgctggtggatttggtgagtatgat
	gattctggggcattcggtaatactcgtagatcaaattttgttggtacacgtaaaggacttgactgggataaccgttcatacttcactaatgacgggtatgaaattgacccagcatcacaacgaaa
	ttccagatatacattaaatcgtcctgaattaattggaaatgaaacacgtccatggaacatttctttaaactacataattaaggtaaaagaatgacagatattgtactgaatgacttaccattcgttga
	cggccctcctgcagagggccagagccgcatttcctggattaaaaacggcgaagaaatattaggagctgacacgcagtatggaagcgaaggttcaatgaatagacctacagtttctgtacta
	agaaatgtcgaagttctcgataaaaacattggaatacttaaaacatctttagaaaccgcaaatagtgatattaaaacaattcagggcatcttagatgtatctggtgatattgaagctttggcccaa
	ataggtatcaataaaaaggatatttctgacctcaaaacgctaaccagtgaacatacagaaatattaaatggacctaatagtacagttgacaacattcttgctgatattggtccatttaactctgag
	gccaactctgtatacagaacaatcagaaatgatttactgtggataaagcgtgaacttggacaatacgcaggtcaagatattaatggtcttcctgttgtaggaaatcctagtagtggaatgaagc
	atcgcattattaataatactgatgccattacttcacagggaatacgtttaagcgaattagaaacaaaatttattgaatctgatgtaggttctttgactattgaagttggtaatcttcgtgaagagcttg
	gaccgaaaccaccatcattttcacaaaacgtttatagtcgtttaaatgaaattgacactaaacagacaacatttgaatctgacattagtgctattaagacctcaataggatatccaggaaataatt
	cgattattactagtgttaatacaaacactgataatattgcatctattaatttagagctaaatcaaagtggaggtattaaacagcgtttaaccgttattgaaacttctattggttcagatgatattccttc
	gagtattaaaggccaaatcaaagataatacaactttaatcgaatctctaaatggaatcgtcggtgaaaacacttcatctggtttaagagcgaatgtttcatggttaaacaaaattgttggaactga
	ttctagcggtggacaaccttctccttctgggtctcttttaaaccgagtttctacaattgaaacttctgtttcaggattgaataacgatgttcaaaacctacaagtagagattggtaataatagcgcag
	gaattaaagggcaagttgtagcgttaaatactttagtaaatggaactaatccaaacggttcaacagtcgaagaacgcggattaaccaattcaataaaagctaacgaaaccaacattgcatcag
	ttacacaagaagtgaatacagctaaaggtaatatatcttctttacaaagcggtgttcaagctctccaagaagccggttatattcctgaagcgccaagagatgggcaagcttacgttcgtaaaga
	cggcgaatgggtattgctttctaccttttatcaccagcataacatggggccgcaaggccccaaaggatttaaatgtcaggatataattctcagaatccaaaggaactcaaagatgtcattcta
	agacgtttaggggctccaattattaatgttgagttaacacccgatcaaatttacgattgtatccagcgtgccctagaattatacggtgaataccattttgatggactcaataaagggtttcatgtgtt
	ttacgtaggggatgacgaagaaaagtacaagaccggagtcttcgatttaagaggttctaacgtatttgcagtaactcgcattttacgcacaaatattgggtcaataacatctatggatggaaac
	gctacatatccgtggtttactgactttcttttgggaatggctggtattaatggcggaatgggaacgtcttgtaatagattttatggaccaaatgcctttggtgccgatttggggtattttactcaactt
	accagttatatgggaatgatgcaggatatgctctctcctattccagacttttggtttaattcagcaaatgaacagctcaaagtcatgggaaacttccaaaaatatgatttaattatcgtagaaagct
	ggactaaatcatacattgatacaaacaaaatggttggaaatacagtaggatatggaacagtcgttccacaagataactggtcattatctgaacgatataataaCccagacaacaatttagtag
	gtcgtgttgttggtcaagacccaaatgttaagcaaggtgcttacaataatcgttgggtgaaagactatgcaacagctttagctaaagaattaaatggtcaaattttagcacgccaccagggaat
	gatgcttcctggcggtgttacaattgatggacaacgcttaatagaagaagctcgattagaaaaagaagcactgcgcgaagaattatacttacttgaccctccatttggaattttggtaggttaat

TABLE 18B-continued

DNA sequences of fragments #1-60 in Table 18A

Fragment #	DNA sequence
	atggctacttacgataaaaatctttttgctaaattggaaaaccgcacaggttattctcagaccaatgaaactgaaatactaaatccttatgtaaatttcaatcattataaaaacagccaaatattagctgatgtattagtagctgaaagcattcaaatgcgaggtgtagaatgctattatgttccaagagagtatgtttcccctgatttgatattcggcgaagacttgaaaaataaatttactaaagcttggaaatttgctgcatatttaaattcatttgaaggatatgaaggagctaaatcgttctttagtaattttggtatgcaagtacaggatgaagttactttgtccattaatccaaacttgtttaaacaccaagtaaatggaaaagaaccgaaagaaggcgatttgatatattttcctatggataacagcttatttgaaattaactgggttgaaccatatgatccattttatcaattaggccaaaacgctattcgtaaaattacggcaggtaaattcatttattctggagaagaaattaatccagttctacagaaaaatgaaggaattaacattccagaatttagtgaattagaattaaatcctgttcgcaatcttaacggtattcatgacattaatattgatcagtatgctgaagtagatcaaattaattctgaagctaaagaatatgttgaaccctatgttgttgtcaataacagaggcaaatctttcgaatctagcccatttgataatgatttcatggattaa (SEQ ID NO: 446)
36	atgtttggttatttttataattcgtcttttagacgatatgctaccttgatgggcgatttgttttcaaatatccaaatcaaacgtcagttagaatctggtgataagtttatacgtgttcctattacatatgcatcaaaggaacactttatgatgaaattgaataaatggacatcaataaattcacaagaagatgtagctaaagttgaaaccattctacctcgtataaatttacatttagttgatttagctataatgctccatttaaaacaaacatttaaatcagaatttactgcaaaaaggtgcaacttctgtagtatcgcagtataatccatctcctattaaaatgatttatgaattgagtatctttactcgctacgaagatgatatgtttcaaatagttgaacagattcttccatatttcaacctcattttaatacaactatgtacgagcagtttggaaatgatattccatttaaaagggatatcaaaattgtactgatgtctgctgctatagacgaagctatagatggggataatttatctcgtcgtagaattgaatggtcattaacatttgaagtaaatggatggatgtatcctccagtagatgatgcagaaggattaattcgtactacttatacagattttcacgccaatacaagagatttgcctgatggcgaaggtgtttttgaatctgtcgatagcgaagttgttcctcgagatattaacccagaagactgggatggaacagtaaaacaaactttcactagtaatgtaaatagaccaacaccgccagaacctcctggcccaagaacatagaggttattatggaaggtcttgatataaacaaacttttagatatttctgacctccccggaattgacggggaggaaatcaaagtatatgaacctctgcaattagtagaagttaaaagcaatccacaaaaccgtactcctgacttagaagatgattatggagtagttcgtcgaaatatgcattttcaacaacaaatgctaatggacgcggccaagattttcttgagacggcaaagaatgctgattctcctcgtcacatggaagtatttgcaactcttatggggcaaatgactacgacgaacagagaaatactgaagcttcataaagatatgaaagatattacatctgagcaggttggcaccaaaggcgctgttcctacaggtcaaatgaatattcagaatgcgacagtattcatgggttcaccaacagaattaatggacgaaattggtgatgcttacgaggctcaagaagctcgtgagaaggtgataaatggaacaaccaattaatgcattaaatgatttccatccgttaaatgaagctggaaaaattttaataaaacacccaagcttagcggaaagaaaagatgaagatggaattcattggataaaatctcagtgggatggaaaatggtatcctgaaaaattcagtgattaccttcgtctacacaaaatagtaaaaattccaaacaactctgataagcctgaattatttcaaacttataaagataagaataataaaagatctcggtatatgggtcttcctaacttgaaacgagctaatattaaaacacaatggactcgtgaaatggttgaggaatggaaaaaatgccgagacgatattgtttattttgcagaaacatactgtgctattactcatattgactatggtgtcataaaggttcaattacgtgactatcagcgtgatatgctcaaaataatgtcatctaaacgtatgactgtttgtaatctatcgcgtcagctcggtaaaacaacggtagtagctattttccttgcacactttgtatgttttaacaaggataaagctgtaggtattcttgcgcacaaaggctcaatgtctgcggaagttttagaccgtactaagcaagcaattgaactgcttcctgactttttacagccaggtatagttgaatggaataagggttcaattgaactagataatggttcttcaattggcgcttatgcttcctctcctgacgcagttcgtggtaactcgttcgcaatgatttacattgacgaatgtgcgtttattccaaacttccatgattcctggcttgctattcaaccagtaatttcatctggtcgtcgttcgaaaattattattactacgactcctaatggattaaatcatttttatgatatttggactgctgctgttgaaggtaaatctggatttgaaccatatactgctatttggaattcagttaaagaacgtctttataacgatgaagatatttttgacgatggatggcaatggagcatacaaaccattaatggttctactttagctcaatttcgtcaagaacacaccgcagcgtttgaagggacttctggtacattaatttcgggaatgaaattagctattatggatttcattgaagtaactccagatgatcatggttttcatcgatttaaaagccctgaaccagatagaaaatatattgcaactctagactgctcagaaggtcgtgggcaagattaccacgctttgcatattattgatgttaccgatgatgtgtgggaacaggttggtgttttgcactcaaacactatttctcatttaattctacctgacatcgttatgcgttatttagtagaatacaatgaatgcccagtttatattgaattaaatagtactggtgtgtcagttgcaaaatcgctttatatggatttagaatacgaaggtgttatctgtgattcatatactgatttaggaatgaaacaaactaaacgcacgaaagcagtaggatgttccacgctaaaagaccttattgaaaaagataagcttattattcatcaccgagcgactattcaagaatttagaacgtttagtgaaaaaggcgtgtcttgggcggctgaagaaggttatcacgacgatttagtaatgtctttagtaatttttggatggttatcaacacaatcaaaatttattgattatgcggataaagatgacacgcgattagcatctgaagtattttcaaaagagcttcaagatatgagcgacgactacgcgccagttatatttgtggattcggttcattctgctgagtatgttccagtatctcatggtatgtcaatggtataaatatattaaagcatattaaagaggattaaaaatgactttattatctccgggcattgagctcaaagaaactacggttcaaagcaccgtggttaataactctactggtacagcagctttggccggtaaattccagtggggtcctgcttttcagattaaacaggttacaaatgaagtagatttagttaatacttttggtcaaccaaccgctgaaactgctgactattttatgtctgcgatgaatttcttgcagtacggaaatgacttacgagtagttcgtgctgttgatagagataccgctaaaaactcatcaccaatcgctggtaatattgaatacacaatttctaccccaggtagtaactatgcggttggagataaaatcacagtcaaatatgtttcagatgatattgaaactgaaggtaaaattactgaagtagacgcagatggaaaaattaagaaaattaatattcctactgcaaaaattatcgctaaagcgaaagaagtcggtgaatatccaacactaggttctaactggactgcggaaatttcttcatcttcctctggtttagctgcagtaataactcttggaaaaattattactgattctggtattttattagctgaaattgaaaatgctgaagctgctatgacagcggttgactttcaagcaaatcttaaaaaatatggaattccaggagtagtagcgctttatccaggcgaattaggcgataaaattgaaattgaaatcgtatctaaagctgactatgcaaaaggagcttctgcattactcccaatttatccaggtggtggtactcgtgcatctactgccaaagcagtgtttggatatggaccgcaaactgattcacaatacgctattatagttcgtcgcaatgatgctattgttcaaagcgttgttctttcaactaagcgtggtgaaaaagatatttacgatagtaacatctatatcgatgactttttcgcaaaaggcggctcagaatatatttttgcaactgcacaaaactggccagaaggcttctctggaattttaactctgtctggtggattatcatcaaatgctgaagtaacagcaggagatttgatggaagcttgggacttctttgctgaccgtgaatccgttgatgttcaactgtttattgcgggttcttgtgccggtgaatctttagaaacagcatctactgtccaaaaacacgtcgtttcaattggggatgctcgccaagattgcttagtattgtgctctcctccgcgtgaaactgtagttggaattcctgtaactcgtgcagtagataatttagttaactggagaactgcggcaggttcatacactgataataactttaatatcagttcaacctacgcagcaattgatggtaactataagtatcagtatgacaaatataatgatgtgaatcgttgggttccattagcagctgatattgctggtttatgcgcaagaactgataacgtatctcagacttggatgtctccagctggttataatcgtggccagattcttaacgttattaaacttgctattgaaactcgccaggctcagcgcgaccgtttataccaagaagctatcaacccagtaactggtacaggtggcgatggttacgtattgtatggtgataaaacagctacttctgttccttctccatttgatcgtattaacgttcgtcgtctgtttaatatgttgaaaacgaatatcggacgtagttcaaaatatcgtttgttcgaattaaacaacgcgtttactcgttcatcattccgcacagaaactgcccagtacttgcagg

TABLE 18B-continued

DNA sequences of fragments #1-60 in Table 18A

Fragment #	DNA sequence
	gaattaaagctctcggtggaatttatgaatatcgtgtagtttgcgatacaacaaataacactccgtcagtaattgatagaaatgagtttgttgcaacattctacatccaacctgcgcgcagtataaattatattactttgaatttcgtcgcaacggctactggtgcagatttcgatgagttaactggtcttgcaggttaa (SEQ ID NO: 447)
37	atgtttgtagatgatgtaacacgcgcgtttgaatcaggtgattttgcgcgacctaacttattccaagtagaaatttcttatcttggacaaaattttacgtttcaatgtaaagccactgctttaccagctggtattgtagaaaaaaattccagtcggatttatgaaccgtaaaattaacgtagcaggcgatcgtacattcgatgactggactgttacagtaatgaacgatgaagctcatgatgctcgccagaagttcgttgattggcaaagcattgctgcggggcaaggaaacgaaattactggtggaaaacctgcagagtataaaaagagcgctatcgttcgtcaatatgctcgtgacgctaaaacagtaacaaaagaaattgaaattaaaggtctgtggcctactaacgtgggtgaacttcaattagattgggattcaaacaatgaaatccaaacatttgaagtaactcttgctctcgattattgggaataa (SEQ ID NO: 448)
38	atggctaaaatcaacgaacttctgcgcgaatcaaccacaacgaatagcaactcaatcggtcgcccaaatctcgttgctttgactcgcgctaccactaaattaatatattctgacattgtagcaacgcaaagaactaatcaacctgttgctgctttttatggtatcaaataccttaacccagacaacgaatttacatttaaaactggtgctacttatgctggcgaagctggatatgtagaccgagaacaaatcacagaattaacagaagagtctaaattaactctcaataaaggcgatttattcaaatataataatatcgtttataaagtattagaagatacaccatttgctgatattgaagaaagcgacttagagctggctcttcagattgcaattgttcttttaaaggttcgtctattttctgacgcagcgtcaacaagcaaatttgaaagctctgatagtgaaattgcggatgctagattccagattaataaatggcaaaccgcggttaaatctcgtaaacttaaaactggcatcacagttgaattagcgcaagatttagaagcaaatggattcgatgctcctaatttcttggaagatttgcttgcaactgaaatggcagatgaaatcaataaagatattctgcaatctttgattacagtgtcaaaacgctataaagttacaggaattactgatagtggattcatcgatttgagttatgcgtctgcacctgaagctggtcgttcattataccgaatggtatgtgaaattgtttcgcatatccaaaaagaatcaacttatacagcaacgttctgtgttgcttctgctcgtgccgctgcgattcttgctgcatcaggttggttaaaacataaaccagaagatgacaaatatctttcacaaaatgcctacgggttattagctaatggtttaccgctttattgcgatactaacagcccattagattatgtaatcgttggtgtagtagaaaatatcggtgaaaaagaaattgttggatcaattttctatgctccgtatacagaaggtctcgacttagatgaccctgaacatgtaggcgcatttaaagttgttgttgatccagaaagcttacaaccgtctatcagtttattagttagatatgctttatcagcaaatccttataccgtagcaaaagatgaaaaagaagcaagagtaattgatggtggagacatggataaaatggcgggtcgttcagatttgtctgttttattaggtgttaaattaccaaaaattattattgatgaataa (SEQ ID NO: 449)
39	atgagaactgaggttgtggtgtttactcttcatgagtctggaaagtcattcattgaaattgctcgtgaattaaacttacatgcaaaagaagtggctgtattatgggctcgagctatgactgctaagaataaatttgaaactcgagaaaaagttgtctatagaaaaagacatatcaataaaaaggtgaaaaatggaacagtatgaactttatgaaaatgaatcttttgctaatcaattacgcgaaaaagcattaaaaagtaaacagtttaagctagagtgttttattaaagatttttcggaacttgctaataaagcagctgaacaaggtaaaacatatttagttattatactgctcgcgataaattgattactgaagaaattggtgattggctgagaaaagaaggatttaattttaaagtcaatagtgatcagcgtgatggtgattggttagaaattacattttgaggattaattatgtttaaaaagtagcagtcttgaaaatcattacaactctaaatttattgaaaaactttacagcttgggattgactggcggcgaatgggtagctcgtgaaaagattcacggcacaaatttctcattgattattgagcgtgataaagtaacttgtgctaaacgtactggaccgattcttcctgctgaagatttctttgggtatgaaattattctaaagaattacgctgattccattaaagcagtacaagatattatggaaacctcagcggttgtatcttatcaagtctttggcgaattcgctggacctggcattcagaagaatgttgattatggcgataaagatttttatgtatttgacattattgtcactacagaaagtggtgatgtgacttatgttgatgattatatgatggaatcattctgtaatacatttaaatttaaaattgctccacttttaggtcgcggtaaatttgaagagcttattaaattgccaaatgatttagattctgtcgtccaagattataattttacagtagaccatgctggattagttgatgcaaataaatgcgtttggaatgccgaagcaaaaggcgaagtatttactgctgaaggatatgtattgaaaccttgttatccttcttggcttcataatggaaatcgtgtagcaattaaatgcaagaattccaaatttagtgaaaagaaaaagtctgataagcctattaaagctaaagttgaactatcagaagctgataacaaattggtgggaattttagcttgttacgttacactgaaccgtgtaaataacgttatttctaaaattggcgaaattggtccaaaggattttggaaaggtgatggggctaactgttcaagatattttggaagaaacttctcgtgaaggtattactctaactcaagcagataatccttctttgattaaaaaggaattagttaaaatggtaagatgtacttcgtccagcttggattgagttggtgagctaa (SEQ ID NO: 450)
40	atgatagataaagattatattgcagagctgaaggctcttgatgataacaaagaagctaaagctaaattagctgaatatgctgaacagtttggtataaaggtcaaaaagaataaatcttttgataatatcgttgttgatattgaagaagccctccagaagctcgctagtgaacctatgccagagactgatgggttatctattaaagacttaattgatgctgctgatgccgcagagggattaaaatatgacgatgaagaagtcaatccagaagcagcacttctgattgattctccggttaaatctgacattaaaattgaagtagtagaaacggataaaattcctgaaaataccgatgttttgattgaagatactcctttttgttgaagaaaagtttgaacaagctgtagctgagattattgaatctgaaaagccgtctgtatttactcttccggaaaactttagtccgaatcttcagctgattggaaaaaatccaggattctgcactgttccttggtggatttatcaatggattgctgaaactccggattggaaatctcacccaactagttttgaacatgcgtcagcacaccaaactttatttagcttaatttattacattaaccgcgacggatcagttttaattcgtgaaacacgcaattcttctttcgtaacattaaaataaggataacttatgacttttacagttgatataactcctaaaacaccgacaggggttattgatgaaaccaagcagtttactgctgcacccagtggtcaaactgaaggtggaactattacctatgcttggagcgtagataatgttccacaagatggagctgaagcaacttttagttatacctgccggtcaaaagactattaaagtagttgcaacaaatacaattccagaagctgaagctgaaacagcagaagctactacaactatcacagttcaaaataagacacaaacgaccaccttagctgtaactcctaatagccctgacgctggagtaatcggaaccccagttcaatttactgctgccttagcttctcaacctgatggagcatctgctacgtatcagtggtatgtagatgattcacaagttggtggagaaactaactctacatttagctatactccaactacaagtggagttaaaaaatcaagtgtgtagctcaagtaaccgcgacagattatgatgcactaagcgttacttctaatgaagtgtcattaacggttaataagaagacaatgaatccacaggttacattgactcctccttctattaacgttcaacaagatgcttcggctacatttactgctaatgttactgatgctccagaagaagcgcaaattacttattcatggaagaaagattcttctcctgtagaagggtcaactaatgtatataccgttgatacttcatctgttggaagtcaaactattgaagtgactgccgtcgttactgctactgattatgatagcaaaacagttaaaacaacaggtcaagttcaggtaactgataaagttgctccagaaccagaaggtgaattaccttatgttcatcctcttccacatcgtacttcagcttacatctggtgcggttggtgggttatggatgaaatccaaaaaatgactgaa

TABLE 18B-continued

DNA sequences of fragments #1-60 in Table 18A

Fragment #	DNA sequence
	gaaggtaaagattggaaaactgaagatccagagtaaatactacctgcatcgttacactcttcagaagatgatgaaagactatccagaagttgatgtccaagaatcgcgtaatggatacatcattcataaaactgctttagaaactggtatcatctatacctatccataa (SEQ ID NO: 451)
41	atgagattagaagatcttcaagaagaattgaagaaagatgtgtttatagattcaactaaattacagtatgaagcagctaataatgtgatgttatacagtaaatggcttaataagcattcaagtattaaaaaggaaatgcttagaattgacgcacagaaaaaagttgctcttaaagctaaattagactactactcgggacgaggagatggtgatgaatttagtatggatcgttacgagaaatcagaaatgaagacagttctatcagcggataaggatgttttaaaggttgatacctcgttacagtattgggggattttattagatttctgtagcggagctcttgatgctatcaaatcacgcggatttgctattaagcatattcaagacatgcgggcatttgaggctggaaaataatgagatatagcattgatgatgcttttaattatgaagaagaatttgaaactgagattcaattcttaatgaaaaagcataatcttaagcgtcaggatattcgtatcctggccgatcacccgtgtggtgaagatgtcctttatattaaaggaaaatttgccggatatcttgatgaatatttttattctaaagatatgggcattgatatgcatatgagagttgtataaatagatataattcagaggagacaatcatgtcagataagatttgtgttgtctgtaaaactccaatcgattctgcattggttgttgaaacagacaaaggtcctgtacatcctgggccttgctataattacattaaagaactaccagtttcagaaagttcggaagaacaattaaatgaaacacaactttgctatag (SEQ ID NO: 452)
42	atgtatgaatacaaatttgatgtgagagttggttctaaaataatcaattgtcgcgcattcacgcttaaagaatatctagaacttattactgccaaaaataatggttccgtagaagtaattgttaaaaagctaatcaaagactgcacaaatgcaaaagatttaaaccgccaagaatcagaactattgctgattcatttatgggcgcattctcttggagaagttaatcacgaaaactcctggaagtgcacctgtggaactgaaataccaacccatataaatctattacatacacaaatagatgcaccagaagacctctggtatacactgggtgacattaaaattaaattccgataccctaaaatttttgatgataaaaatatagcccacatgatagtatcatgcatagaaacgattcatgctaacggtgaaagcattccagttgaagacttaaatgaaaaagaactagaagatttatattctatcatcacagagtcagatattgtagctataaaagatatgcttttaaagcctaccgtttatttggctgttccaattaaatgtccagagtgtggaaaaacccatgctcatgtaataagaggcctcaaagagttctttgagttactataatggcaaatattaataagctttattctgacattgacccggaaatgaaaatggattggaacaaagatgtttccagatcacttggattaaggtcaattaaaaacagtcttttgggaattattacaacaagaaaaggttcaagaccgtttgaccctgaatttggatgtgatttatcagatcagctttttgaaaatatgactcctcttactgctgacacggttgagcgcaatatcgaaagcgcagtaagaaactatgagccacgtattgataaattatcagttaatgtgataccagtttatgatgattatactctgatagtagaaatacgcttttcggtcatcgataaccctgatgatattgagcagataaaactgcaactggcttcgagtaatagggtataa (SEQ ID NO: 453)
43	atggcaaacattattcgttgtaaattaccagatggtgttcatcgttttaaaccatttacggtagaagattatcgagatttttgttagttcgaaacgatatagaacatcggtcaccacaagaacaaaaagaaataattactgatttaattgatgattattttggagactatccgaagacttggcaaccatttatatttttgcaggtatttgtagggtcaataggtaaaactaaagtaccggtcacatttgtatgtccaaaatgtaaaaaagaaaagacagttccatttgaaatatatcaaaaagaattaaaggaacctgtttttgatgtagctaatgttaaaattaaattaaagtttccttctgagttttatgaaaataaagcaaagatgattactgaaaatattcattctgttcaagtagatgaaatatggtatgattggaaggaaattagtgaatcaagccaaatagaacttgttgatgccatcgagatagaaacattagaaaaaattctcgatgcaatgaatcctattaatttaactctacatatgtcatgctgtaataagtacattaaaaaatacactgatatagtagacgtgtttaagctgttagttaacccagatgagatatttactttttatcaaattaatcacacactcgtaaaaagtaattatagcttaaattcaataatgaaaatgattcctgccgagcgcggattcgtattaaaactgattgagaaggataaacaataatgagtatgttgcaacgccccggatatccaaatctcagcgttaaattatttgatagctacgacgcttggagtaataatagatttgttgaattagctgctactattaccacattaactatgcgggattctctttatggacgaaatgaaggaatgctgcagttttatgattctaaaaacatccatacaaaaatggatggaaatgaaataattcagatttctgtagctaatgcaaatgatattaataatgttaaaacacgaatttatggatgtaagcatttttccgtgtcagtagattcaaaaggtgataacatcattgctattgaattgggaactattcattctatagaaaatcttaaatttggtagacaattttccctgatgcaggtgaatctataaaagaaatgcttggtgtcatttatcaggatcgcacattattaactccagcaataaatgctataaatgcttatgttcctgatattccatggactagcacatttgaaaactatttgtcatatgtaagagaagttgctctagctgtaggaagcgacaaatttgtatttgtatggcaagacatcatgggagttaacatgatggactatgatatgatgataaatcaagaaccatatccaatgattgtcggtgagccatctttaataggtcaattcatccaagaattaaaatatccattagcatatgatttcgtttggttgactaaatcgaatccttacaaacgtgatccaatgaaaaatgctactatctatgctcattcattttttagattcttcactgccaatgattactacaggaaagggtgaaaactctattgtagtgtcaagatcaggtgcttattctgaaatgacttataggaatggatatgaagaagctattcgtcttcaaactatggcacaatatgacggttatgctaaatgttctactgtcggtaattttaacttgactcctggtgttaaaattatttttaatgatagtaaaaaccaatttaaaacagaattttacgttgatgaagttatccatgaattatccaataataattcagtaactcatctatatatgttcactaatgcaacgaaactggaaacaatagacccagttaaggttaaaaatgaatttaaatctgatactaccactgaagaaagtagttcttccaataagcaataaagaagtttctattcctaaaatgggtcttaaacattataacattttaaaggatgttaaaggtcctgatgaaaatttaaaacttcttattgattctatttgtccgaatttatcaccggcagaagttgatttcgtttctattcatttattggaatttaatggaaagattaaatctcgtaaagaaatagatggctatacttatgacattaatgatgtttatgtatgccaaagattagaatttcaataccaaggaaatacattttattttagacctcctggaaaatttgaacaatttttaacggtgagcgatatgttatctaaatgcttgcttaaggtcaacgatgaagttaaagaaattaattttcttgagatgccagcattcgttttaaaatgggcaaatgatatttttacaactttagcaattcctggccctaatggtccaataaccggaattggcaatattattggattatttgaatgaaaaagccacaagaaatgcaaacgatgcgtagaaaagttatttcagataataaaccaacacaggaagcggctaaatccgcttctaacactttatctggacttaatgatatatctacgaaattggatgatgctcaagctgcttctgaattaatagctcaaactgtcgaagaaaaatcgaatgaaatagttggagcaattggtaacgtagaaaacgcagtgagtgatactactgccggttctgagttaattgctgaaactgtcgaaattggcaacaatattaataaagaaatcggtgaatcactcggaagcaaattagataaattaacaagtttactagagcaaaaaattcagacagctggaattcaacagactggaactagtttagccacagttgaaagcgctattcctgttaaagtcgttgaggatgatactgctgaatctgtgggtcctttattaccggctcccgaagcagttaataatgatcctgacgctgattttttccctacccctcagccagttgaacccaaacaagaatcgccagaagaaaaacagaaaaaagaagcatttaacttaaaattatctcaagctttagataaattaacaaagactgttgattttggatttaagaaatccatttcaattagtgataaaatatcaagcatgttatttaagtacaccatcagtgctgctattgaagctgctaaaatgactgcaatgatattggctgttgttgttggaatagacctgttgatggttcactttaaatattggtcagataaattttcaaaagcctgggatttatttaatactgactttactaaattctctagcgaaaccggaacttggggtcctttattacagagcatctttgattctattgataaaattaaacaactttgggaagcgggagattggggtggatt

TABLE 18B-continued

DNA sequences of fragments #1-60 in Table 18A

Frag-ment #	DNA sequence
	gacagtagctattgttgaagggcttggaaaggttctttataatttaggagaacttattcagcttggaatggctaaattatctgcggcaattcttcgagtcattcctggcatgaaggatactgctgat gaagtagaaggaagagcactagaaaatttccaaaattctactggagcatctctcaataaagaagaccaagaaaaagtagcaaattatcaagataaacgaatgaatggagaccttggcccaa tagcagaaggactagacaaaatctctaactggaaaactcgtgcatctaactggattcgtggtgtagataataaagaagcactgactactgacgaagaacgtgcagcagaagaagaaaaatt aaagcagctttcacctgaagaaagaaaaaatgctttaatgaaggccaatgaagctcgtgccgcgatgattcgttttgaaaaatatgctgattcagctgatatgagtaaagactcaacggttaaa tcagttgaagctgcctatgaagaccttaaacagcggatggatgacccggatttaaataattcgccggcagttaaaaaagaacttgcttctagatttgctaaaattgatgctacttatcaagagct caagaaaaatcagcctaatgccaaacctgaaacttctgctaaatcaccagaagcgaaacaggtccaggttattgaaaagaacaaagcacaacaagctcctgttcaacaagcatctccttca atcaataatactaataatgttattaagaaaaatactgtcgttcataatatgacacctgttacgagcacaactgctcctggtgtatttggcgcgactggagttaattaaggaataatatggcaattgtt aaagaaataactgctgatttaattaaaaagtccggtgagacaatttcagccggacagagcactaaatcagaagtaggaattaaaacatacacagcccagtttccaactgggcgtgctagtgg taatgacactacaggggacttccaggtaacagatctatataagaatggattattatttactgcatacaatatgtcatctagggattctggaagtcttagatcgatgagatctaactactcttcttcat cttcgagtattttacgtacagccagaaacactattagtagtacagtatcaaaactatcaaatggattaatatcaaataataattcaggaacaataagtaaagctcctgtcgcaaacattcttttacc gagatctaaatctgatgttgatacatcatcacatagatttaatgatgttcaagaaagccttatcagtagaggcggaggtactgctactggagtgctaagtaatattgcttcaaccgcagtatttgg ggcgttggaaagtataacacaaggtataatggctgataataatgaacagatttatacgacagccagaagtatgtatggtggtgctgaaaatagaactaaagtgtttacatgggatttaactcca cgttcaacagaagatttaatggctattattaatatctatcaatattttaactatttttcttatggtgaaacgggtaaatctcaatatgctgctgaaataaaggggtatttagatgattggtatcgttctac gttaattgaacctttatctccggaagacgcagctaaaaataaaacactatttgagaaaatgacatcgagtttaactaacgttctagtagtttcaaacccgacggtttggatggtgaaaaactttgg tgcaacatctaagtttgatggaaaaacggaaatatttggtccatgccaaatacagagcatcagatttgataaaacacctaacggtaactttaacggattagctattgctccaaatctccctagtac atttactctcgagattactatgagagaaattatcacgttaaaccgtgcttctttatatgcggggactttttaatgtattctttagaggaatttaataatcaagcaataaacgcagatttccaacgtaata atatgtttagctgcgttttgcaacaactccatcaactaaaagctcttcgttgataagttcaattagcaacttttcttataataacttgggcctaaattcagattggttaggattaactcaaggtgatatt aatcagggaattacaacgctaattacagctggcacacaaaaactaataagaaaatcggggttagtaaatatcttattggtgccatgagtcaacgtacagttcaaagtttattaggctcatttac agttggtacatatttaattgacttctttaacatggcatataactcatctggattgatgatatactctgtaaaaatgccagagaatagattatcctatgaaactgattggaactacaactctcctaatatt cgtataactgggagagaattagaccctttggttatttcatttagaatggattcagaatcgtgtaattaccgtgcaatgcaagactgggttaatgctgttcaagacccagtaactggattacgtgct ctgccacaagatgtcgaggcagatatccaggttaatcttcattctcgtaatggattgcctcatactgcggtgatgttcaccggatgtattccagtgtcagtgagcgctcctgagttatcatatgat ggagataaccaaataactacatttgatgttacttttgcgtatagagtcatgcaggctggagcagttgataggcaagctgcgcttgaatggcttgaatctgctgctataaatggtattcaaagctct tctggaaataatggaggtgttactgaactatctagttcgctttcacgacttagtagattaggaggaactgcaggaagcatttcaaacattaatactatgacagggattgtcaattcgcagagtaa aatattaggagcaatataa (SEQ ID NO: 454)
44	atgaaatcttctttgcgctttttaggtcaagaacttgtagttgaaggcgttattcctgctgataatgcttttaacgaagcggtttacgatgaatttattaaaatttttggaacagataaaaagttcggaa tttttccttctgaaaatttttcaaagccagaacagactgaaagcattttccagggtgtagtaacaggtaaatttgagtcagaagctccggtaaaaattgaagtttatattgaagacagtttagttgct tcagtttctgctttcatttcattccgtaaataa (SEQ ID NO: 455)
45	atggaactcattacagaattatttgacgaagatactactcttccgattacaaacttaaatccaaagaagaaaataccacaaattttttcagttcatgttgatgatgcaattgaacaaccaggctttc gtttatgtacctatacatctggaggtgatactaatcgcgatttaaaaatgggcgataaaatgatgcatattgttccttttacattaactgctaaaggttcaattgctaaattaaaaggtcttggtccaa gcccaattaattatatcaattcagtttttactgttgcaatgcaaacaatgcgtcagtataaaattgatgcttgtatgcttcgtattcttaagtctaaaactgctggtcaagctcgacaaattcaagttatt gctgatagacttatccgtagtcgttcaggtggcagatacgtccttcttaaggaactctgggattatgataaaaagtatgcatatattcttatacatcgcaaaaatgtatcactagaagacattccag gagttccggaaattagtaccgagctctttactaaagttgaatcgaaggtcggtgatgtttatatcaataaagatactggagctcaagtaactaaaaacgaggcaattgcagcatctattgcaca agaaaatgataaacgtactgaccaagctgtaatcgttaaagttaaaatttcccgtagagcaattgcgcaaagtcaatcattggaatcttctagatttgaaagtgaattattccagaagtatgaatc taccgcagctaatttcaataagcctgctaccgctcctttaattcccgaagcagaagaaatgaaaattggaattaattcattagcttctaaaacaaaggcagcaaaaattattgccgaaggaact gcgaatgaacttcactatgactataaattcttttcaaaaagtgaggttgatgaagtttctgaaaaaattaaagatgtaatttttaacgcgattaaaaatgaaccaactacttcaataaaatgtttaga gaaatacgcggcagctgtcaatcaattctttgaagaatataaagataattggcttgataaacataataaaactcgtaaagggcagccagatgaagtctggggagaaataactaaaaatgcctg gaatgcagcaaaaactaaattcctcaaacgaatgatttatagtttttctggaattggtgctggtccaatgattgatattactattgcttgtgatggttctaaatatacaccatcacaaaagcgcggta ttagagagtattgtggttcaggatatacagacattaataatcttcttttaggtcgttacaatccagaacgatatgatgtaatgagtgaaaaagaaattgaatctgctataaataatttagattcagctt ttgaaaatggtgaccgcataccggaaggcattacagtttatcgtgctcaaagtatgactgctcctatatacgaagcgctagttaaaaataaagtgttctatttcagaaattttgtatctacttctttaa ctcctatcatttttggacgttttggaattacacatgctggtattggtcttttagaaccagaagctcgcaatgaattaacagttgataaaaatgaagaaggaataactattaatccaaacgaaataag agcgtataaagaaaatcctgaatacgttaaagttcaaataggatgggcaattgatggagctcataaagttaatgttgtatatccaggaagtctcggaatagcaacagaagctgaagttattcta ccgcgcggattgatggtcaaagttaataaaataactgatgcttctaataatgacggaaccacgtctaataatacaaaactcattcaagctgaagttatgaccacagaagaactcaccgaatcg gtaatctatgacggagaccgtttaatggaaaccggcgaagtagttgcaatgacaggtgatattgaaatagaagacagagttgactttgcatcatttgtttcatcaaatgttaaacagaaagtag aatcatctctcggaattattgcgtcttgcatagatattacaaacatgccttacaagttcgttcaaggataaatcatggaacttattacagaattatttgacggcgcttcggcgccggttgttaactta aatcctaagcataaaataccacaaatttttgctattcaagccggcgaagaaagcgtgcttcctggatttagattttgtacatacacctctggtggtgatacaaataaaaacgttaagccaggcga

TABLE 18B-continued

DNA sequences of fragments #1-60 in Table 18A	
Fragment #	DNA sequence
	taaaatgatgcatatcgtaatgataggtgtcaacgagaaattatcgctggttaagcttagaaacttgggtggaaatccaattggcgtcattaatgctgtttttgatactgctcttcaaacaatgaaacagtataaaatcgacgcatgcttattccgcgtactaaaaagtaaaacaaatggcgcagctcgtcaaatgcaagttattgctgaccgtttagtacgtactaaaggagcaggtcgatatgttcttttaaaggaaatctgggactatgataaaaagtatgcatatattatggtttaccgtaaaaatgccaatttagaagacattccaggtgtacctcctatttcaactgagttattcgcaaaagttgaatcgaaggtcggtgatgtttatgtagatgttaaaacaggtgatgctgttcctaaagctgtcgctgttgctgcttctattgctttagaaaatgataaacgtactgaccaagcggttattcagaaaactaaaattagtcgtcgattagcagcacaagctcaatattctactgtcgatgcttcacttcagggtgatagcttcgctgccaagaaatatcaagagtttgaatctaaagttccggtatataaagcagaaggaccaatgaactctggcgttattcagattggttcaaacttcagcaaaggagctatcggtggtatgagaagtgcttctcgttttaaatctagcgattatgaactagaaaacttccgaaatcatattgcattagcccatgcacgtttacgtgatccatctatcaagttacagagcgatataacatatcaaggttctcaagaatatttaaagaataaagaattctttgattataaaactgataaaattttaagtgatcttgctgatattaatatttctaatagctttgatgttattaagaaaattatcaatgatttggttaaaggttctaaagctacgccagatgaaaagacagttattattcaatttgtcatgaatggcatttataaattgattaatgaatctgctgcccaggcatatgaatatgcaagcactgaagtaactccaaaaggactgactcaggctgagtctgatgtaattgaagattattgtgcagattcatatgttgaaatgaactcgttccttttgggtaaaccagattctacccgtgaagaatatatggaacgtgctattaagcacatcgagacgttggattctgcattcgctaaaggttcagttcttcctccaggaactacgctttatcgcggacaagaagttacctttaaaactttgcgtcacaacattgaaaacaaaatgttctatttcaagaacttcgtatcgacatcacttaaaccaaatatctttggcgagcatggtaaaaactatatggctctagatgattccggtgcagtattttctggagaaggagaaggttccgttgatgcagaagatttgatgcatatgggtagtcattctacatatgctaatgaagatgctgaaactagcgtgggtatggtaattaaaggagctgagcgaatcaaagttatcgttccaggtcatttatcaggatttccatcagaagctgaagttattctaccgcgtggaattttactgaagattaataaagtaagtacgtactttatgaaagaaactgcttataacaagtatctaatcgaaggtacaatcgttcctccttctgaacaattagaagaatcagtatatgatggagaccatttaatggaaactggtgaagttcgtccaatggctggatttaatcaattccttgtagaagaatcaaaagaagaggaaaacgaagtttctcaaatattagcttctttggttaacatcaacggaatgtctaaaaagttcaaaatgtag (SEQ ID NO: 456)
46	atgaactacatcaactttgaacgtaaatatgtttctaatggtattgcaggttctattgatactatctgcctttggaaacatcaaaatggatcagtatgcgaaattgaacagtatatgactcctaactatgtttatatgcgatttgaaaatggcatcacggtttcaatcacaatggaaggttccaactttaaaatcgctctggatgatgattttcgtcaacgcgatttagggactcatccttgctggaatggtgctaatcgcaagcttttggttaaaacttggattcgtcatattctgagtaacagagctaaacctgagcatcttgaagcaatctttgatgtagttcttaacgaatttgatatttaa (SEQ ID NO: 457)
47	atggcaaaacaagctaaagcaaagaaagcagttgaaaagaaagttggtgattctaaacgcgctggctacaagcgtgggtcgaactctcgtatcaatcaaactgttgagaagatcatgcgccgagcacgtgcggttcttcgagatgatgcttctcgttttggtaagcagaaagcataa (SEQ ID NO: 458)
48	atgattaaacaattacaacacgctcttgaactgcaacgaaacgcatggaataatggtcacgaaaactatggcgcatctattgatgttgaagccgaagctcttgaaatcctgcgttatttcaaacatctgaatcctgctcaaactgcattagctgccgagcttcaggaaaaagatgaacttaaatatgctaagcctctggcttctgccgcgcgaaaagcagttcgtcactttgtggtaacattgaagtaa (SEQ ID NO: 459)
49	atgtctgaagtacaacagctaccaattcgtgctgtcggtgaatatgttattttagtttctgaacctgcacaagccggcgatgaagaagttacagaatcaggacttattatcggtaaacgtgttcaaggtgaagttcctgaactgtgtgtaattcactctgtcggtcctgatgttcctgaaggttttgtgaagttggtgatttgacttctcttccagttggtcaaattcgaaacgttccgcatccttttgtagctctgggtcttaagcagccaaaagaaattaaacaaaaattcgttacctgtcattataaagctattccgtgtctttataagtga (SEQ ID NO: 460)
50	atgctgctaagtgaaaaaccgattactgttaaagaattccaagaaaaagttaagctatttgcgcaggaattggtaaataaggtttctgaacgatttcctgaaacatcggttcgtgttattaccgaaactcctcgttcagtattagtaattgtgaatccaggtgatggcgatcaaatatcgcatcttaaactggattttgatggattagttgaagcacaaagggtgtatggcgtactatgatgaatttaactgatataattgataattgtcttgaaaatgatactggcgatcatagagcgcttgactctgaaacagcaaagttcattagaataactttaatgaatgatactctggtgaatagtattcatccttctgtgtatgatgctattattgtgacgaagtatccagttgagcttcataaaaagatgactggcgcagtttttattgataagaaaaaccgctttaaagatgggcagaatataattagttctgttattaaaagtataactaaacttcgtcacgaaatttatcgtgttgaaactgctaaatctgcttatctggtgattatgaaatgaaagcgagtacagtacttcaaattgcatatttagtatcgcaggaatcaaaatgttgctcctggaaggtaggagcagtaattgaaaagaatggacgtattatttctactgggtataatggttcacccgcaggggtgtgaactgttgtgattatgctgctgagcaaggttggttgctgaataagcctaaacatactatcattcaaggccataagcctgaatgcgtatcatttggttcaactgatcgttttgtcttggcgaaagaacatcgtagtgctcactctgaatggtcgtctaaaaatgaaattcatgctgagctaaatgcaattttgtttgctgcacgaaatggttcttctattgaaggtgctactatgtatgtaacactttctccttgtccagattgtgcaaaagcgatagctcaatctggtattaaaaagctggtttattgcgaaacatatgataaaaataaacctggctgggatgatattctgcgaaatgcaggtattgaagtgtttaatgttcctaagaaaaacttgaataagttaaactgggaaaatatcaacgaattctgcggtgaataatgaaatttcgtttggtaaagctcacagcaattagttcttattctaatgagaacatctcgtttgctgtagagtataagaaatatttttctctaaatggaaacagtattataagacaaattgggtttgtattgataaaccatatagttggaaatctgatttagaaaaattccaaaaattactttccacccttaaagaacgtggaacaactcatattaaaactgtaataggtaaataaatgaaactgacaactgagcagaaagtagcaattcgtgaaattttgaaaactaaattgtccatgggtgtttcaaacgtagtttttgaaaagtctgatggtactattcgtactatgaaaggtactcgtgatgcagactttatgccaaccatgcaaaccggtaaattgactgaatctactcggaaagaatctacggatatgattccagtatttgatgttgaacttggcgcttggcgaggtttttctattgacaaattgatttctgttaatggtatgaaagttgagcatttgcttcaatttattggtaaataa (SEQ ID NO: 461)

TABLE 18B-continued

DNA sequences of fragments #1-60 in Table 18A

Fragment #	DNA sequence
51	atgtttcctacttattctaaaatcgtagaagtagtgtttagccaaattatcgctaataatatgtttgaaaaacttgataacgcagccgagcttcgaatccatgctcaagtgactcatgtattgaacactttgcttccagaccaggtggattctgttgccattacgctgtatccaggttccgcgcatatcattgttgtattcggtcttgatgctgagctagtcatcaaaggcgatattcgttttgaatcgcagacagcagaattcaaagcaatttaa (SEQ ID NO: 462)
52	atgaaacaataccaagatttaattaaagacattttgaaaatggctatgaaaccgatgatcgaacaggcacaggaacaattgctttgttcggtactaaattacgctgggatttaagtaaaggttttcctgcagtaacaactaaaaagctcgcctggaaagcttgcattgctgagctactttggtttttatcaggaagcacaaatgtcaatgatttacgattaattcagcatgattcattaattcaaggcaaaacagtctgggatgaaaattacgaaaatcaagcaaaagatttaggataccatagcggtgaacttggtccaatttatggaaaacagtggcgtgattttggcggtgtagaccaaattgtagaagttattgatcgtattaaaaaactgccgaatgataggcgacaaattgtttctgcgtggaatccagctgaacttaaatatatggcattaccgccttgtcatatgttctatcagtttaatgtgcgtaatggctatttggatttgcagtggtatcaacgatcagtagatgtttttcttggtcttccatttaatattgcatcatatgctgcgttagttcatattgtagctaagatgtgtaatcttattcctggagatttgatattttctggcggtaatactcatatctatatgaatcacgtagaacaatgtaaagaaattttgcgtcgtgaacctaaagagctttgtgagctggtaataagtggtctaccttataaattccgatatctttctactaaagaacaattaaaatatgttcttaaacttaggcctaaagatttcgttcttaacaactatgtatctcacccgccaattaaaggaaagatggcggtataattttaatttaattgcgaggatatatgattttacgatttaaagatacttctggtgtcgttctttttacacttcctaatccaagcgagttagaagttccaggaccaaatcagcctattatcatttatggcaaaaaatattatactcataaaatgactcgtgagtattttgataataaaatttctacagttaaaacttcttcagattgttactatgatattactgttttaacggaaaaacaatatgacgaattatcgccgcgcgggccgtctatgccaggtagtgaataaatataaatccgactttgatgttaatattcaccgtggtacattttggggaaattacgtcggtaaagatgctggcagccgggaggctgccattgaattattcaaaaaagattttatacgtcgaattaaatccggagaaataactaaagaacatttagagcctttacgtggaatgaggctaggatgcacatgtaaaccaaagccgtgtcatggtgatataatagctcatatagttaaccgattgtttaaagacgattttcaagttgaggacttatgcaattaattaatgttatcaaaagtagtggtgtttctcagagctttgacccgcaaaaaattattaaagttttatcttgggcagctgaaggaacatctgtagatccttatgaattatatgaaaatattaaatcatatctccgtgatggaatgaccactgatgacattcagactattgtcattaaggctgctgcgaattctatttcggttgaagaacctgattatcaatatgtagctgcacgctgtttaatgtttgctcttcgtaaacatgtttatgggcagtatgaaccgcgttcatttattgaccatatttcttactgtgtaaatgaaggtaaatacgaccctgaattgttgtcaaaatattctgcagaagaaattacattttagaatcaaaaattaagcacgagcgggatatggaatttacttattccggggcgatgcaattaaaagaaaaatatctcgttaaagataaaaccactggtcaaatttatgaaactccacagtttgcatttatgactattggaatggcattgcatcaagatgaacctgttgacagattaaaacatgttattcgtttttatgaagcagtatctactcgacagatttcactgccaactcctattatggctggttgccgtactccaactcggcagtttagttcatgcgttgttattgaagctggtgattcattaaagtcaattaataaagcttctgcttcaattgttgaatatatttctaaacgcgctggaattggaattaacgttggtatgattcgtgccgaaggttctaagattggcatgggtgaagtacgccatactggtgttattcctttttggaaacattttcagactgctgttaaatcatgttcacagggtggaattcgtggcggcgctgctactgcttattatcctatttggcatttggaagttgaaaatcttctcgttttgaaaaataacaaaggcgtagaagaaaaccgcatccgtcatatggattatggtgttcaactgaatgatttgatgatggaacgattcggaaagaacgattacattactttgttcagtccgcatgaaatgggtggagagctgtattattcttattttaaagaccaagaccgtttccgtgaattatacgaagcagcagaaaaagaccctaatattcgtaaaaagcgtattaaagcccgtgaactatttgaattgctcatgactgaacgttcaggaacagcaaggatttatgtgcagttcattgataatacgaataactatactccgtttattcgtgaaaaggcacctattcgtcagagtaacttgtgctgtgaaattgctattccaacaaatgatgtgaatagtcctgatgctgaaattggattgtgtactctctctgcattcgtactagataattttgactggcaagaccaagataaaattaatgaattggcagaagttcaagttcgtgctcttgataatctgttggattaccaaggatatccagttcctgaagcagaaaaagctaaaaagcgtcgtaaccttggtgtaggtgttactaactatgcagcttggctggcaagtaactttgcttcttatgaagatgctaacgatttaacacatgaactatttgagagattacagtatggactcattaaagcatccattaagctcgccaaagaaaaaggaccttgtgaatattattcagacactcgttggtctcgaggcgaattacctatcgactggtacaataaaaagattgaccaaatcgcagctccaaaatacgtttgtgactggtcgtcgctgcgggaagaccttaagctctttggcatccgtaatagcacattatcagcacttatgccatgtgagtcatcttcccaagtttctaacagtacaaacggtatcgagcctccacgtggaccagtctctgttaaagaatcaaaagagggttcctttaatcaagtcgtgcccaatattgaacataacatagacctatatgattatacatggaaattagctaagaaaggtaataaaccttatcttacgcaggtagctattatgctgaaatgggtatgtcaatcagcttcagcgaatacatattatgacccgcagattttttccaaaaggaaaggttccaatgtcaataatgattgatgacatgttatacggatggtattatggcattaaaaatttctattatcataatacccgcgatggttctggtactgatgattatgaaatagaaactccaaaagctgaagattgttcatcctgtaaattatga (SEQ ID NO: 463)
53	atgagattacaacgccaaagcatcaaagattcagaagttagaggtaaatggtattttaatatcatcggtaaagattctgaacttgttgaaaaagctgaacatcttttacgtgatatgggatgggaagatgaatgcgatggatgtcctctttatgaagacggagaaagcgcaggattctggatttaccattctgacgtcgagcagtttaaaactgattggaaaattgtgaaaaagtctgtttgaaggagatgatatgattttttgtatttgaatttatgaatgatgaattcgattatgcaatttttaacgcattgcataatcctgatttaaatgaatttaatgaaatgttttctgacgctttgagtatgtcagaagaatactgcggagaatgtcaacgtgtttgtgtgacagtctttgaaaacaaagaaaagacgtatgaagaattattctttgacgctaataaagccactgaatggtttattgaaaggggttttgcgtaatgattaaattggtattcgcttattctccaactaaaacggtcgaaggctttaatgaattagcattcggtttatgtgatggtttaccatggggacgagttaaaaaggacctccagaattttaaagctcgtactgaaggtacaattatgattatgggtgctaaaacgttccagtcattgtctacattacttcctggtcgtagccatattgtagtatgtgacctcgagcgtgattatcctgaaactaaagacggtgatttagcacatttctatattacatgggagcagtacataacttacatttctggcggttcaattcaagtgtcaagtcctaatgcaccattcgaggctatgcttgatcagaattctaatgtaagcgtaattggcggacccgctttgttatatgctgcattaccttgtgcggatgaagtagttgtttctcgcatcgttaaaaggcatcgtgttaattcaacggttcaattagatgcaagttttcttgatgatataagcaagcgtgaaatggttgaaacccattggtataaaatagatgaagtaacaacccttacggaatcagtgtataaatgaaataacgcgtggcggaaaatatgaactttaattattaccctattctattagaaaaagacgcgaaacaaccaaaatggcagggtcctcagtttattaaaggcgtctatcaattagtagttcctaaagacaagatttatagcagttgtttcactgaatccgcttgcagtattttcggtaatagttctccgtattggaattttgatataaaactggatagaaatatcgatatttggttgaaagccatggatattggcaatattacgtttgatgagaataattatcatattattggtcgcttttctaaacgcggtaaagaattatatttcactcctgaaatcgaaagaaaatttgatgctaaaccgtattga (SEQ ID NO: 464)

TABLE 18B-continued

DNA sequences of fragments #1-60 in Table 18A

Fragment #	DNA sequence
54	atgtatattggcaaaaagtatgaacttgttccaagacttattgatacatttattaattatcgcccacgttctaattcatcaatagttaaaattattgaagaaaatggcgggtggtttgaagttaaagaaactttctttgttgatggatttagagcaataaaacacattgaatgcgcaaatggaaagcatttttactttaacatttgtgaagatgaatttcattgttttcgtgagtataaagaacagacttctgaagaagatgaaatcgaagacaaggtttctggcgtaacaaaaattcactgcattgtagacgaaaacaatgtagatgaaatcattgaacttttgcgaaaaactttcaaaaagtag (SEQ ID NO: 465)
55	atggctaaagttgatattgacatcgttgattttgaatatattgaagaaattattcgtaatcgttatcctgaacttagtatcacaagcgtgcaagattctaagttttggagtattcaaatcgttattgaaggtcctcttgaagacctcacccgctttatggctaatgaatattgcgatggtatggattctgaagacgcagaattttacatgggactgattgaacaataa (SEQ ID NO: 466)
56	atgtttaaacgtaaatctactgctgaactcgctgcacaaatggctaaactggctggaaataaaggtggtttttcttctgaagataaaggcgagtggaaactgaaactcgataatgcgggtaacggtcaagcagtaattcgttttcttccgtctaaaaatgatgaacaagcaccatttgcaattcttgtaaatcacggtttcaagaaaaacggtaaatggtatatcgaaaattgctcatctacccacggtgattacgattcttgtccagtatgtcagtacatcagtaaaaatgatttgtacaacactgacaataaagagtacggtcttgttaaacgtaaaacttcttactgggctaacattcttgtagtaaaagatccagctgctccagaaaacgaaggtaaagtatttaaataccgtttcggtaagaaaatctgggataaaatcaatgcaatgattgcagttgatgttgaaatgggtgaaactccggttgatgtaacttgtccgtgggaaggtgctaactttgtactgaaagttaaacaagtttccggatttagtaactacgacgaatctaaattcctgaatcaatctgcgattccaaacattgacgatgaatctttccagaaagaactgttcgaacaaatggttgacctttctgaaatgacttctaaagataaattcaaatcgttcgaagaactgagcactaagtttagtcaagttatgggaactgctgctatgggtggtgccgcagcgactgctgctaagaaagctgataaagttgctgatgatttggatgcattcaatgttgatgacttcaatacaaaaactgaagatgattttatgagctcaagctctggcagttcatctagtgctgatgacacggacctggatgaccttttgaatgacctttaa (SEQ ID NO: 467)
57	atggatttagaaatgatgctggatgaagattacaaagagggaatttgctttattgactttagtcaaattgcgctttcaactgctttagtaaacttcccagataaagaaaaaattaatttatcaatggttcgtcatttgatattgaactcaattaagtttaatgtcaaaaaagcaaaaacgcttggatacactaaaatcgtgttgtgtattgataacgcgaaatctggatattggcgtcgtgattttgcttattattataagaaaaaccgtggaaaagcacgagaagaatctacttgggactgggaaggttattttgaatccagccataaagttatagatgaattgaaagcttatatgccatacattgttatggatattgataagtatgaagcggatgaccatattgctgttcttgttaaaaagttctctttagaaggacataagattttaatcatttcgtcggatggtgactttacacagcttcacaaatatccaaatgttaagcaatggtctccaatgcataagaaatgggttaaaattaaaagcggttctgctgaaattgactgtatgactaaaatccttaaaggcgacaaaaaggataacgttgcttcagttaaagtacgatctgacttctggtttaccagagttgaaggtgaacgaactccttcaatgaaaacttcaatcgttgaagccattgctaatgaccgtgagcaagctaaggtgcttctcacagaatctgaatataatcgttataaagaaaatttagttctaattgattttgattatattcctgataatattgcttcaaacattgtgaattactataattcatataaattaccaccgcgtggcaaaatttattcatattttgtaaaagcgggtctttctaaattaactaatagcattaatgaattttgaggtgaataatggctaaaaaagaaatggttgaatttgatgaagctatccatggcgaagacttggctaaatttattaaagaagcatctgatcataaactgaaaatttccggttataatgaactgattaaagatattcgaattcgtgctaaagatgaacttggcgttgatggtaagatgtttaatcgtctattagctttgtatcataaagataaccgtgatgtgtttgaagctgaaactgaagaggtagttgaactttatgacacagttttctctaaatgatattcgtccggtcgatgagaccggtctttcagaaaaagaactttcaatcaagaaagaaaaggatgaaatagcaaagcttcttgatcgtcaagaaaatggatttattattgaaaaaatggtagaagagtttggaatgagttatcttgaagctacaacagcattcttagaagaaaattctattcctgaaactcaatttgctaaatttattccttcgggtataattgaaaaaattcagtcagaagctattgacgaaaatcttttacgtccttctgttgttcgctgtgaaaaaactaatacattagattttctactatgattaaattccgcatgcctgctggtggtgaaagatacattgatggtaaatcagtttataaattatacttaatgataaaacagcatatgaatggaaagtatgatgttattaagtataattggtgcatgcgggtgtctgatgccgcttatcaaaagcgaagggataagtatttttccagaagttatcagaaaaatataaattaaaggaacttgctttaatttttataagtaatttggttgctaaccaagatgcttggattggtgacatctctgacgctgatgcacttgtgttttatcgtgaatatatcggacgcttaaagcaaattaaatttaagtttgaagaagatattcgcaacatttattattttagtaaaaaagttgaagtttctgcttttaaagaaatctttgaatataatccaaaggttcaatcaagttatatttttaaactgcttcagtcgaatataatttcgtttgaaacgtttatcttgcttgattcgtttttaaatataattgataaacacgatgaacagactgataatttagtctggaataattattctataaagttaaaggcttatagaaaaattttaaatattgattcacagaaagctaaaaatgttttcattgaaactgtgaaatcttgcaagtattaa (SEQ ID NO: 468)
58	atggccgagattaaaagaaagttcagagcagaagatggtctggacgcaggtggtgataaaataatcaacgtagctttagctgatcgtgccgtaggaactgacggtgttaacgttgattacttaattcaagaaaatacagttcaacaatatgatccaactcgtggatatttaaaagatttgtaatcatttatgataaccgcttttgggctgctataaatgatattccaaaaccagcaggagcttttaatagcggacgctggagagcattacgtaccgatgcaaactggattacggtttcatccggttcatatcaattaaaatccggtgaagcaatttcggttaatactgcagctggaaatgacatcacgtttactttaccatcttctccaattgatggtgatactatcgttctccaagatattggaggaaaacccggagttaaccaagttttaattgtagctccagtgcaaagtattgtaaactttagaggtgaacaagtacgttcagtactaatgactcatccaaagtcacagctagttttaatttttagtaatcgtctgtggcaaatgtatgttgctgattatagtagagaagctgtaattgtaacaccagcgaatacttatcaagcacaatcaaacgattttatcgtgcatagatttacttctgccgcaccgataaatattaaacttccgagatttgctaatcacggagatattattaatttcgttgatttagataaactaaatccactttatcatacaattgttactacatacgatgaaactacttcaatacaagaagatggaactcattctattgaagaccgtacatcaatcgacggtttcttgatgtttgatgataatgagaaattgtggagattgtttgacggggacagtaaagcacgtttacgtatcataacgactaattcaaacattcttccaaatgaagaagttatggtatttggtgcgaataacggaacaactcaaacaattgagcttcagcttccaactaatatttctgttggtgatactgttaaaatttccatgaattacatgagaaaaggacaaacagttaaaatcaaagctgctgatgaagataaaattgcttcttcagttcaattactgcaattcccaaaacgctcagaatatccgcctgaagctgaatgggtaactgtccaagaattagtttttaacggtgaaactaattatgttccagttttggagcttgcttatattgaagattctgatggaaaatactgggttgtacagcaaaacgttccaaccgtagaaagagtagattctttaaatgattctactagagcaagattaggcgtaattgattagctacacaagctcaagctaacgtcgatttagaaaattctccacaaaaagaattagcaattactccagaaacgttagctaatcgcactgctactgaaactcgcagaggtattgcaagaatagcaactactgctcaagtgaatcagaacaccacattctcttttgctgacgatattat

TABLE 18B-continued

DNA sequences of fragments #1-60 in Table 18A

Frag-ment #	DNA sequence
	catcactcctaaaaagctgaatgaaagaactgctactgaaactcgcagaggtgttgctgaaattgctacgcagcaagaaactaatacaggtactgatgatactacaatcatcactcctaaaaagcttcaagcccgtcaaggttctgaatcattatctggtattgtaacttttgtatctactgcaggtgctactccagcttctagccgtgaattaaatggtacgaatgtttataataaaaacactaataatttagttgtttcacctaaagctttggatcagtataaagctactccaacgcagcaaggtgcagtaattttagcagttgaaagtgaagtaattgctggaaaaagtcaggaaggatgggcgaatgctgttgtaacgccagaaacgttacataaaaagacatcaactgatggaagaattggtttaattgaaattgctacgcaaagtgaagttaatacaggaactgattatactcgtgcagtcactcctaaaactttaaatgaccgtagagcaactgaaagtttaagtggtatagctgaaattgctacacaagttgaattcgacgcaggcgtcgacgatactcgtatctctacaccattaaaaattaaaaccagatttaatagtactgatcgtacttctgttgttgctctatctggattaattgaatcaggaactctctgggaccattatacccttaatattcttgaagcaaatgagacacaacgtggtacacttcgtgtagctacacaagttgaagctgctgcaggaaaattagataatgttttaataactcctaaaaagcttttaggtactaaatctaccgaatcgcaagagggtgttattaaagttgcaactcagtctgaagctgtggctggaacgtcagcaaatactgctatatctccaaaaaatttaaaatggattgtgcagagtgaaccttcttggagagcaactactacggtaagagggtttgttaaaacttcgtctggttcaattacattcgttggtaatgatacagtcggttctacccaagatttagaactttatgagaaaaataattatgcagtatcaccatatgaattaaaccgtgtattagcaaattatttgccgttaaaagcaaaagctgtagatagtaatttattggatggtctagattcatcccagttcattcgtagggatattgcacagacggttaatggttcactaaccttaacccaacaaacgaatctgagtgcccctcttgtatcatctagtactgctacgtttggtggttcagtttcggcaaatagtacattaactatttctaatactggtacgacttcttctcgatttacatttgagaaaggtcctgcttctggtagtaatgctgattctgcattgtatgttcgtgtatggggtaataagtacagcggcggttctgatgtaactcgtgcaacgattatagaattctctgatgctaccggctctcatttctattctcaaagagatacgtcaaataatgtgttgttcaacatttcaggtacgatgcaatcagtcaacgctagcgttcgtggtgttctgaacgttacaggtgtctcaacgtttaatagttcagttacagccaatggtgaattcatcagtaaatcaccaaatgcttttagagcaataaatggaaattacggattctttattcgtaatgctggtaatgacacctattttatgctcactgcagcaggtgatcagagcggtggatttaatggattacgtccattatcaattaataatcaatccggtcaggttacgattggtgaaagcttaatcattgccaaaggtgctactataaattcaggtggtttgactgttaactcgagaattcgttctcagggtactaaaacatctgatttatatacccgtgcgccaacatctgatactgtaggattctggtcaatcgatattaatgattcagccacttataaccagttcccgggttattttaaaatggttgaaaaaactaatgaagtgactgggcttccatacttagaacgtggcgaagaagttaaatctcctggtacactgactcagtttggtaatacacttgattcgctttaccaagattggattacttatccaacgacgccagaagcgcgtaccactcgctggacacgtacatggcagaaaaccaaaaactcttggtcaagttttgttcaggtatttgatggtggaaaccctcctcaaccttctgatattggtgctttaccttctgataatgcaacaatcggaaacttgacaataagggatttcttaaggattggtaatgtccgcattattccagaccctgtgaataaatctgttaaattcgagtggattgaataagaggtattatggaaaaatttatggctaagtttggacaaggatacgtccaaacgccatttttatcggaaagcaattcagtacgatttaaattaagcatagcgggatcttgcccgctttctacagcaggaccatacgttaaatttcaagataatcctgtaggaagtcaaacatttagcgcaggtcttcatttaagagtttttgacccttccaccggagcattagttgatagtaagtcatatgcttttttcgacttcaaatgatactacatcagctgcttttgttagcttcatgaattctttgacaaataatagaattgttgctatattaactaacggaaaggttaattttcctcctgaagtagtatcttggttaagaactgcaggaacgtctgcttttccatctgattctatattgtcaagattcgacgtatcatatgctgctttttatacttcttctaaaagagctattgcattagagcatgttaaactgagtaatagaaaaagcacagatgattatcaaactattttagatgtcgtatttgacagtttagaagatgttggagctaccgggtttccaagaagaacgtatgaaagcgttgagcaatttatgtcagcggttggtggaactaataacgaaatcgcgcgtttaccaacttcggctgctataagtaaattatctgattataatttaattcctggtgatgttctttatcttaaaacacagctatacgccgatgccgatttacttgctcttggaactacgaatatatccattcgattttataatgcatcaaatggatatatttcctcgacacaagctgaatttaccgggcaagctggtgtttgggaattaaaagaagattatgtagttgttccagaaaatgcagtaggatttacgatatatgcacaaagaactgcacaagctggtcaaggtggaatgaggaacttaagcttttctgaggtatcaagaaatggtagtatttcgaaacccgctgaatttggtgtcaatggtattcgagttaattatgtctgtgaatctgcttcacctccggatataatggtacttcctacacaagcatcgtctaaaactggtaaagtgtttgggcaagaatttagagaagtataa (SEQ ID NO: 469)
59	atgtttactacagctgaactaaaacgagcaaaagctaagaaagggcaaggaaaatataaagctgaattagttaaagaacttcagtttgctgaggctgaattgaattcaatgattattcaaaatgctccagaaactgaaattgctcttaaacgtattgcgaataagtgtcttcgtgatgcaatcgtcgatcttttagcggattattgagtaaaatgaaaatcgttgagattgaactatgagttcattatggtggtgttttgtttggttaattagtattccattaatttgtttaacatttacttttgtgatgaggttattatgaaaatttttaattctgtacttattgcttgtgcgtggtgggttgcacaagtttcggcagtagtgattggtattcacatttattacgaatatttttaa (SEQ ID NO: 470)
60	atgtacaatattaaatgcctgaccaaaaacgaacaagctgaaattgttaaactgtattcaagtggtaattacacccaacaggaattggctgattggcaaggtgtatcggttgacacaatccgtcgtgttttgaaaaatgctgaagaagctaaacgccctaaagttactattagcggtgatattacagttaaagttaatagcgatgcagttattgctccagttgctaaatctgacattatttggaatgcatctaaaaaattcatttcaattactgttgacggtgtaacttataacgcaactcctaatactcattcaaactttcaggaaattcttaatctgcttgtagcggataagctggaagaagctgcgcaaaaaattaatgttcgtcgcgctgttgaaaaatatatttccggcgatgttcgaattgaaggtggaagcttgttctatcaaaatattgaattgcggtctggtttggttgatcgtattcttgactcgacggaaaaaggcgaaaactttgaattttattttccgttcttggaaaatctgctggaaaacccaagccaaaaagcggtatctcgactctttgatttcttggtagcaaacgatattgaaatcaccgaagatggttacttctatgcttggaaagtagttcgtgacaactactttgactgtcactcaaacacctttgataacagtccgggtaaagtagttaaaatgccacgtactcgtgtgaatgacgatgatacacaaacttgttctcgtggtctgcatgtgtgttctaaatcttatattcgtcactttggcagttcaaccagtcgagttgtaaaagttaaagtacatccgcgtgatgtagtatcaattccgattgattacaacgatgctaaaatgcgtacctgccaatacgaagtagttgaagacgttactgaacaatttaaataagggcttcggcccttatcatattaaggaaaattatgttaggttatcaagcacgagtaaaagaagaatacgatcaattaatgctcaaaattaatgcactgagtaaattttagaaagcacaaagtttctaacggttagtgcagttgagcaagaactgctactttcgcagtttatctcaatgaaatcttatgctgagtgtctagagaaaagaattgcgcaattcaaataa (SEQ ID NO: 471)

[0431] Various modifications and variations of the described methods, pharmaceutical compositions, and kits of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific embodiments, it will be understood that it is capable of further modifications and that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in the art are intended to be within the scope of the invention. This application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure come within known customary practice within the art to which the invention pertains and may be applied to the essential features herein before set forth.

SEQUENCE LISTING

The patent application contains a lengthy "Sequence Listing" section. A copy of the "Sequence Listing" is available in electronic form from the USPTO web site (https://seqdata.uspto.gov/?pageRequest=docDetail&DocID=US20210130833A1). An electronic copy of the "Sequence Listing" will also be available from the USPTO upon request and payment of the fee set forth in 37 CFR 1.19(b)(3).

---

1. An engineered system comprising an ATPase and an adenosine deaminase wherein the ATPase and the adenosine deaminase are derived from same or different prokaryotes.

2. The engineered system of claim **1**, wherein the ATPase comprises a sequence of WP_012906049.1 or WP_155731552.1, and the adenosine deaminase comprises a sequence of WP_012906048.1 or WP_064360593.1.

3. The engineered system of claim **1**, wherein the ATPase comprises 1100 or less amino acid residues.

4. The engineered system of claim **1**, wherein the adenosine deaminase comprises 1100 or less amino acid residues.

5. The engineered system of claim **1**, further comprising a membrane protein.

6. The engineered system of claim **5**, wherein the membrane protein comprises a SLATT domain or Csx27.

7. The engineered system of claim **1**, wherein the system is configured to modify a target nucleic acid.

8. The engineered system of claim **7**, wherein the target nucleic acid is RNA.

9. The engineered system of claim **7**, wherein modification of the target nucleic acid comprises causing an A to G mutation in the target nucleic acid.

10. The engineered system of claim **1**, further comprising one or more phage proteins.

11. The engineered system of claim **10**, wherein the one or more phage proteins are in Tables 18A-18B.

12. An engineered system comprising one or more reverse transcriptases comprising one or more UG1, UG2, UG3, UG8, UG15, or UG16 reverse transcriptase.

13. The engineered system of claim **12**, comprising a first and a second reverse transcriptase.

14. The engineered system of claim **13**, wherein the first and the second reverse transcriptases are comprised in a protein.

15. The engineered system of claim **12**, further comprising:

a SLATT domain;
a DNA polymerase;
a family A DNA polymerase;
a serine protease domain linked to or associated with the one or more reverse transcriptases;
an MBL domain;
a nitrilase;
a nitrilase, wherein the nitrilase and the one or more reverse transcriptases are comprised in a protein, and the nitrilase is at a C-terminus of the protein; or
a protease.

16. (canceled)
17. (canceled)
18. (canceled)
19. (canceled)
20. (canceled)
21. (canceled)
22. (canceled)

23. The engineered system of claim **12**, wherein the one or more reverse transcriptase comprises (Y/F)XDD (SEQ ID NOS: 1-2), wherein X is any amino acid.

24. An engineered system comprising a retron or one or more molecules encoded by the retron.

25. The engineered system of claim **24**, wherein the retron is an Ec67 retron, Ec86 retron, or Ec78 retron.

26. (canceled)
27. (canceled)

28. The engineered system of claim **24**, wherein the retron is a Tol/interleukin 1 (TIR) domain-associated retron.

29. The engineered system of claim **28**, wherein the TIR domain has NAD+ hydrolase activity.

30. The engineered system of claim **24**, wherein the retron is a topoisomerase-primase (TOPRIM) domain-associated retron.

31. The engineered system of claim **30**, wherein the TOPRIM domain has nuclease activity.

32. An engineered system comprising:

an NTPase of a STAND (signal transduction ATPases with numerous associated domains) superfamily;
an NTPase of a STAND superfamily, DUF4297, Mrr-like nuclease, SIR2, a trypsin-like serine protease, and/or a helical domain;
von Willebrand factor (VWF), a PP2C-like serine/threonine protein phosphatase, and a serine/threonine kinase;
SIR2;
transmembrane ATPase;
ATPase, QueC synthase n, and TatD endonuclease;

S8 peptidase;
DUF4011, a helicase, and a Vsr endonuclease;
a silent information regulator (SIR)2-DUF4020;
SIR2-STAND-TPR;
a Polymerase and Histidinol Phosphatase (PHP)-ATPase;
SIR2 and HerA;
DUF1887;
DUF499, DUF3780, and DUF1156 methyltransferase and a helicase;
a Type I-E CRISPR-associated ATPase; or
ApeA.

33. (canceled)

34. (canceled)

35. (canceled)

36. (canceled)

37. (canceled)

38. (canceled)

39. (canceled)

40. (canceled)

41. (canceled)

42. (canceled)

43. (canceled)

44. (canceled)

45. (canceled)

46. (canceled)

47. (canceled)

48. (canceled)

49. The system of claim **1**, wherein the system comprises two proteins fused together.

50. The system of claim **1**, comprising one or more components in a retrotransposon system.

51. A polynucleotide comprising coding sequences for one or more proteins in the system of claim **1**.

52. A vector comprising a polynucleotide of claim **51**.

53. A cell comprising the polynucleotide of claim **51**.

54. A method of identifying a defense system in a microorganism, the method comprising:
identifying genes of known defense systems in a plurality of genomes of the microorganism;
recording candidate genes located within 10 kb or 10 open reading frames from the identified genes of known defense systems in the genomes;
identifying homologs of each candidate gene in the genomes; and
selecting candidate genes wherein at least 10% of homologs of the candidate genes are within 5000 nucleotides or 5 genes from one or more known defense systems on the genomes.

55. The method of claim **54**, wherein identifying genes of known defense systems comprises identifying known defense genes and filtering false positive hits among the identified known defense genes.

56. The method of claim **54**, further comprising validating the selected candidate genes.

57. The method of claim **54**, wherein the homologs of the candidate genes share at least 70% sequence identity with the candidate genes and/or the homologs have an E-value of 10^{-5} or lower.

58. The method of claim **54**, wherein the recorded candidate genes are within 10 kb from the identified genes of known defense systems on the genomes.

59. The method of claim **54**, wherein at least 15% of homologs of the selected candidate genes are within 5000 nucleotides or 5 genes from one or more known defense systems on the genomes.

60. The method of claim **54**, wherein the plurality of genomes comprises at least 100,000 genomes.

61. The method of claim **54**, wherein the known defense systems comprise one or more of a CRISPR system, Type I RM and McrBC system, BREX-associated system, Zorya system, Wadjet system, Druantia-associated system, Hachiman system, Lamassu system, Thoeris-like system, Gabija system, Septu system, pAgo system, Shedu system, Kiwa system, DUF499-DUF1156 system, and Toxin/antitoxin system.

62. The method of claim **54**, wherein the microorganism is *E. coli.*

* * * * *