



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2013년04월30일
(11) 등록번호 10-1259144
(24) 등록일자 2013년04월23일

(51) 국제특허분류(Int. Cl.)
C12N 15/11 (2006.01) C12Q 1/68 (2006.01)
(21) 출원번호 10-2011-0015756
(22) 출원일자 2011년02월22일
심사청구일자 2011년02월22일
(65) 공개번호 10-2012-0096366
(43) 공개일자 2012년08월30일
(56) 선행기술조사문헌
US20100291578 A1
WO2008060090 A1
Genome Biology, Vol. 8, pp. R143.1- R143.9
(2007.07.20.)

(73) 특허권자
연세대학교 산학협력단
서울특별시 서대문구 연세로 50, 연세대학교 (신촌동)
(72) 발명자
박준홍
서울특별시 종로구 내수동 경희궁의아침2단지 아파트 1401호
이태권
서울특별시 양천구 목동동로 100, 목동아파트 1320동 501호 (신정동)
(74) 대리인
이덕록

전체 청구항 수 : 총 2 항

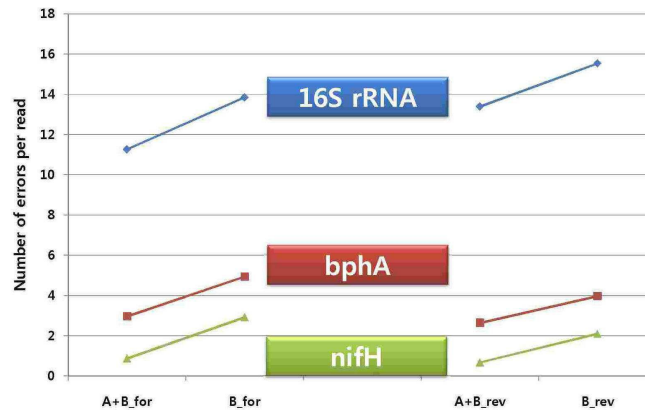
심사관 : 노은주

(54) 발명의 명칭 **파이로시퀀싱 정확도 측정용 인공유전체 및 이를 이용하여 파이로시퀀싱 정확도를 측정하는 방법**

(57) 요약

본 발명은 메타게놈 앰플리콘을 이용하여 FLX 티타늄 파이로시퀀싱하는 경우에 발생하는 오차율을 직접 계산함으로써 파이로시퀀싱 정확도를 측정하는 방법에 관한 것으로, 본 발명에 따르면, 유전자 수 및 미생물 분류수에 대한 FLX 티타늄 파이로시퀀싱의 오차를 확인할 수 있고, 이에 대한 필터링이 가능한 효과가 있다.

대표도 - 도4



이 발명을 지원한 국가연구개발사업

과제고유번호 R33-10076

부처명 교육과학기술부

연구사업명 세계수준의 연구중심대학 육성사업(WCU)

연구과제명 신재생에너지 및 환경오염정화에 활용할 수 있는 미생물자원 탐사를 위한 돌과성 메타지노믹스

주관기관 WCU 그린메타지노믹스 사업단

연구기간 2008.01.01 ~ 2013.12.31

특허청구의 범위

청구항 1

서열목록1의 로도스피릴룸 루브룸(*Rhodospillum rubrum*) ATCC 11170, 서열목록2의 부르크홀데리아 비에타멘시스(*Burkholderia vietamensis*) G4, 서열목록3의 부르크홀데리아 제노보란스(*Burkholderia xenovorans*) LB400, 서열목록4의 디설피토박테리움 하프니엔스(*Desulfitobacterium hafniense*) DCB-2, 서열목록5의 노스톡(Nostoc.) PCC 7120, 서열목록6의 폴라로모나스 나프탈레니보란스(*Polaromonas naphthalenivorans*) CJ2, 서열목록7의 로도코쿠스 속(*Rhodococcus sp.*) RHA 1, 서열목록8의 수도모나스 푸티다(*Pseudomonas putida*) F1, 서열목록9의 네이세리아 시카(*Neisseria sicca*) ATCC 29256, 서열목록10의 오크로박트룸 안트로피(*Ochrobactrum anthropi*) ATCC 49188, 서열목록11의 크로모박테리움 바이오라세움(*Chromobacterium violaceum*) ATCC 12472, 서열목록12의 수도모나스 피케티(*Pseudomonas pickettii*) PK01, 서열목록13의 스팅고비움 야노이쿠야에(*Sphingobium yanoikuyae*) B1, 서열목록14의 에스케리키아 콜라이(*Escherichia Coli*) K-12 서브 W3110, 서열목록15의 바실러스 세루스(*Bacillus cerus*) ATCC 14579, 서열목록16의 코리네박테리움 글루타민(*Corynebacterium glutamin*) ATCC 13032, 서열목록17의 스태필로코쿠스 에피데미디스(*Staphylococcus epidemidis*) ATCC 12228, 서열목록18의 잔토모나스 캄페스트리스(*Xanthomonas campestris*) py. ATCC 33913, 서열목록19의 로소박터 데니트리피칸(*Roseobacter denitrifican*) Och 114 및 서열목록20의 로도박터 스파에로이데스(*Rhodobacter sphaeroides*) KD 131로 이루어진 것을 특징으로 하는 파이로시퀀싱 정확도 측정용 인공유전체(mock community).

청구항 2

제1항의 인공유전체를 파이로시퀀싱한 결과와 제1항의 인공유전체 공지 서열을 비교하여 파이로시퀀싱 정확도를 측정하는 방법.

명세서

기술분야

[0001] 본 발명은 파이로시퀀싱 정확도 측정용 인공유전체 및 이를 이용하여 파이로시퀀싱 정확도를 측정하는 방법에 관한 것으로, 보다 구체적으로는 인공유전체의 애플리콘을 이용하여 FLX 티타늄 파이로시퀀싱을 하는 경우에 발생하는 오차율을 직접 계산하여서 파이로시퀀싱 정확도를 측정하도록 하는 것에 관한 발명이다.

배경기술

[0002] 대량 병렬 파이로시퀀싱(massively parallel pyrosequencing)은 미생물 개체군에서 보다 광범위한 시료 분자량 다양성에 대한 접근방법을 제공한다. 이로 인해, 통상적인 클로닝(cloning)에 의한 서열 플레이트 제조없이도 2~3시간 내에 수십만 개의 DNA 서열(100 내지 200개 뉴클레오타이드)을 생성시킬 수 있다.

[0003] 파이로시퀀서는 454/로슈 게놈 시퀀서(Roche Genome Sequencers, 또는 'GS-FLX'라 함)를 지칭하는 것으로, DNA 합성 동안에 방출되는 파이로포스페이트가 검출되는 시퀀싱 방법에 근거하여 명명되었다. 다수 라이브러리의 분석을 위해 현재 이용가능한 454/로슈 파이로시퀀서는 일정 개수의 독립 시료만을 추적할 수 있으며, 시퀀싱 배지에서 매니폴드(manifold)를 이용한 물리적 분리를 필요로 한다. 이러한 분리 매니폴드는 시퀀싱 플레이트 상의 벽을 차단하여서 비드-결합된 DNA 템플레이트 분자가 추적되지 않도록 하는데, 그 결과 수득되는 서열의 개수가 한정되는 문제를 갖는다.

[0004] 이러한 문제점을 해결하기 위하여, '바코드(Barcode, 이하, 'B '로도 약칭됨)' 또는 독특한 DNA 서열 식별자(identifier)가 고안되었다. 서열 바코드는 각각의 시료를 결합시켜서 시퀀싱할 개별 시료가 풀링(pooling)되어 있도록 하고, 이후에 파이로시퀀서 결과를 분리시키는 것으로, 이종(heterogeneous) 세포-풀(cell-pool) 또는 미생물-풀(organism-pool)에서 식별자 또는 유형 지정자(type specifier)로 작용한다.

[0005] 또한, 454 애플리콘(amplicon) 어댑터(Adapter, 이하, 'A '로도 약칭됨)는 애플리콘 시퀀싱 프라이머 어닐링 부위에 해당하며, 순방향 어댑터(F-어댑터)와 역방향 어댑터(R-어댑터)가 있다.

- [0006] 454를 이용한 파이로시퀀싱에 의해 생성된 메타게놈(metagenome)에는 시스템적 오차가 포함되는 것으로 밝혀졌으며, 이로 인해 유전자 및 미생물 분류군이 과다 측정된다. 즉, 파이로시퀀싱 방법에서 특징적으로 사용되는 인공산물(artifact)은 본래 DNA 시퀀싱 템플레이트보다 15% 이상 많이 증폭된다. 이는 증폭된 DNA가 이멀전(emulsion) PCR동안 공(empty) 비드에 부착될 때 또는 시퀀싱동안 광학 신호가 인접한 공 웰 공간에 흘러들어갈 때 단일 템플레이트를 반복 판독하여 일어나는 것으로 알려졌다.
- [0007] 16S rRNA 분석은 토양, 해수 및 인체와 같은 다양한 서식환경의 미생물 조성을 평가하기 위한 미생물 생태학자의 툴 키트(tool kit)의 필수 구성요소이다. 이는, 다양한 박테리아 중에서 16S 유전자 서열 보존이 높아서 미생물 다양성의 계통발생 분석 및 신규 분류군 동정이 가능하기 때문이다. 이러한 16S rRNA 유전자(16S) 서열은 환경 시료에서 박테리아 다양성(bacterial diversity)을 측정하기 위해 통상적으로 사용된다.
- [0008] 박테리아 다양성은 시료 제조, 프라이머 선택 및 키메라성 16S 증폭 생성물 형성에 의해 영향을 받을 수 있다.
- [0009] 키메라는 다수의 모 서열(parent sequence) 간의 혼성체(hybrid)로서, 신규 미생물로 잘못 해석되어 다양성을 외견상 증가시킬 수 있다. 키메라는 독립적인 증폭 중에서 반복재현가능하게 형성되는 것으로 밝혀졌으며, 시료 다양성을 잘못 인식하고 신규 분류군을 잘못 동정하도록 한다.
- [0010] 파이로시퀀싱과 관련한 선행문헌으로는 US 특허출원 제20100291578호 ‘적하 방법에 의한 파이로시퀀싱(Droplet-based pyrosequencing)’, US 특허출원 제20090325154호 ‘파이로시퀀싱 방법 및 관련 조성물(Pyrosequencing Methods and Related Compositions), WO 특허출원 제2008060090호 ‘파이로시퀀싱을 이용하여 JAK2 V617F를 정량적 검출하기 위한 방법, 프라이머 및 키트(METHODS, PRIMERS AND KITS FOR QUANTITATIVE DETECTION OF JAK2 V617F MUTANTS USING PYROSEQUENCING) ‘ 등이 알려져 있으나, 이들은 파이로시퀀싱 정확도를 측정하기 위한 인공유전체에 대하여 기재하고 있지 않으며, 선행문헌 [Accuracy and quality of massively parallel DNA pyrosequencing, 2007 Huse et al., licensee BioMed Central Ltd.]에는 파이로시퀀싱의 정확도 및 품질을 필터링을 통하여 개선시키는 방법이 기재되어 있으나, 본 발명에 기재된 바와 같은 파이로시퀀싱 정확도 측정용 인공유전체에 대하여는 기재하고 있지 않다.

발명의 내용

해결하려는 과제

- [0011] 본 발명의 목적은 공지 서열의 인공유전체(mock community)를 파이로시퀀싱하여서 그 결과 수득된 서열을 공지 서열과 비교함으로써 파이로시퀀싱의 오차를 직접 계산함으로써 파이로시퀀싱 정확도를 측정하도록 한 인공유전체를 제공하고 이를 이용하여 파이로시퀀싱 정확도를 측정하는 방법을 제공하는데 있다.

과제의 해결 수단

- [0012] 본 발명에 따르면, 메타게놈 앰플리콘을 이용하여 FLX 티타늄 파이로시퀀싱에 대한 오차를 측정하기 위하여, 파이로시퀀싱 그 자체로부터의 오차율을 시험하는 과정, 오차에 영향을 주는 파라미터(프라이머, 바코드, 어댑터)를 밝히는 과정, 인공 서열을 반복하는 과정, 불일치(mismatch), 바코드 또는 어댑터와 관련하여 프라이머 편향오류(primer bias)를 밝히는 과정 및 인공유전체로부터 키메라의 정도 및 범위를 결정하는 과정을 제공한다.
- [0013] 본 발명에 따르면, 서열목록1의 로도스피릴룸 루브룸(*Rhodospillum rubrum*) ATCC 11170, 서열목록2의 부르크홀데리아 비에타멘시스(*Burkholderia vietamensis*) G4, 서열목록3의 부르크홀데리아 제노보란스(*Burkholderia xenovorans*) LB400, 서열목록4의 디설피토박테리움 하프니엔스(*Desulfitobacterium hafniense*) DCB-2, 서열목록5의 노스톡(*Nostoc*.) PCC 7120, 서열목록6의 폴라로모나스 나프탈레니보란스(*Polaromonas naphthalenivorans*) CJ2, 서열목록7의 로도코쿠스 속(*Rhodococcus sp.*) RHA 1, 서열목록8의 수도모나스 푸티다(*Pseudomonas putida*) F1, 서열목록9의 네이세리아 시카(*Neisseria sicca*) ATCC 29256, 서열목록10의 오크로박트륨 안트로피(*Ochrobactrum anthropi*) ATCC 49188, 서열목록11의 크로모박테리움 바이오라세움(*Chromobacterium violaceum*) ATCC 12472, 서열목록12의 수도모나스 피케티(*Pseudomonas pickettii*) PK01, 서열목록13의 스펅고비움 야노이쿠야에(*Sphingobium yanoikuyae*) B1, 서열목록14의 에스케리키아 콜라이(*Escherichia Coli*) K-12 서브 W3110, 서열목록15의 바실러스 세루스(*Bacillus cerus*) ATCC 14579, 서열목록16의 코리네박테리움 글루타민(*Corynebacterium glutamin*) ATCC 13032, 서열목록17의 스타필로코쿠스 에피테미디

스(*Staphylococcus epidemidis*) ATCC 12228, 서열목록18의 잔토모나스 캄페스트리스(*Xanthomonas campestris*) py. ATCC 33913, 서열목록19의 로스오박터 데니트리피칸(*Roseobacter denitrifican*) Och 114 및 서열목록20의 로도박터 스파에로이데스(*Rhodobacter sphaeroides*) KD 131로 이루어진 것을 특징으로 하는 파이로시퀀싱 정확도 측정용 인공유전체(mock community)가 제공된다.

[0014] 또한, 본 발명에 따르면 상기 인공유전체를 파이로시퀀싱한 결과와 상기 인공유전체 공지 서열을 비교하여서 파이로시퀀싱 정확도를 측정하는 방법이 제공된다.

[0015] 삭제

발명의 효과

[0016] 본 발명은 유전자 수 및 미생물 분류수에 대한 FLX 티타늄 파이로시퀀싱의 오차를 직접 계산할 수 있어서 파이로시퀀싱 정확도를 측정할 수 있도록 하는 인공유전체를 제공하는 효과를 갖는다.

도면의 간단한 설명

- [0017] 도 1은 1, 2, 6 내지 8 영역의 순수 시퀀스(raw sequence)를 나타낸 그래프이다.
- 도 2는 RDP (Ribosomal Database Project) 파이로시퀀싱 파이프라인의 초기 과정을 나타낸 그래프이다.
- 도 3은 동적 프로그래밍에 의해 오차 분석을 나타내는 과정을 모식적으로 나타낸 그래프이다.
- 도 4는 1회 read당 16S rRNA, bphA, nifH의 전체 오차율을 나타낸 그래프이다.
- 도 5는 16S rRNA, bphA, nifH의 치환 누적 곡선이다.
- 도 6은 16S rRNA의 오차 분포를 나타낸 그래프이다.
- 도 7는 bphA의 오차 분포를 나타낸 그래프이다.
- 도 8은 nifH의 오차 분포를 나타낸 그래프이다.
- 도 9는 본 발명에서 사용된 인공유전체의 균주, 그 게놈 크기 및 유전자를 나타낸 것이다.
- 도 10은 본 발명의 어댑터를 포함하지 않는 프라이머 서열을 나타낸 것이다.
- 도 11은 본 발명의 어댑터를 포함하는 프라이머 서열을 나타낸 것이다.
- 도 12a 내지 12f는 본 발명에 따른 PCR에서 각 DNA의 농도, 부피 등을 나타낸 것이다.

발명을 실시하기 위한 구체적인 내용

[0018] 본 발명은 이하의 구체적인 실시예를 들어 보다 상세히 설명된다. 그러나, 하기 실시예는 단지 예시를 목적으로 한 것으로, 본 발명의 권리범위가 이에 한정되는 것이 아님이 당업자에게 명백히 이해될 것이다.

[0019] 본 실험에 이용된 재료, 장비 및 문헌은 하기 표 1과 같다.

표 1

재료/장비/문헌	공급처	카탈로그 번호
AccuPrime™ <i>Taq</i> DNA Polymerase High Fidelity	Invitrogen	12346-086
AccuPrime™ <i>Pfx</i> DNA Polymerase	Invitrogen	12344-024
Forward and Reverse Primers premixed	Bioneer (대한민국)	주문 제작
DNase/RNase free water		
Thermo cycler	Biorad	C-1000
Vortex		
Pipettes	Eppendorf	주문 제작
MinElute PCR Purification Kit	Qiagen	28004

QIAquick PCR Purification Kit	Qiagen	28106
Nanodrop Spectrophotometer	Thermo Scientific	ND-1000
PowerSoil DNA Isolation Kit	MoBio	12888
MinElute PCR Purification Kit Manual		

[0020] 삭제

[0021] PCR은 AccuPrime™ Taq DNA 폴리머라아제 하이 피델리티(Polymerase High Fidelity) 또는 AccuPrime™ Pfx DNA 폴리머라아제를 이용하여 실시하였다. 본 실험에 사용된 기능성 유전자(functional gene)는 nifH, bphA 및 16S rRNA 유전자(27F/518R)이었다.

[0022] (1) PCR 프라이머 준비

[0023] PCR 프라이머 세트는 하기와 같은 2가지 종류로 합성하였으며, 본 실험에서 사용한 프라이머 서열은 도 10과 도 11에 나타내었다.

[0024] 5' -> 3'

[0025] 어댑터-바코드-링커-특이프라이머

[0026] 5' -> 3'

[0027] 바코드-링커-특이프라이머

[0028] 하기 표 2는 본 발명에서 사용한 기능성 유전자의 특이 프라이머 서열을 나타내고 있다.

표 2

타겟 유전자	프라이머 명칭	Sequences (5' -> 3')	문헌	크기
nifH	Poly Forward	TGCGAYCCSAARGCBGACTC	Poly et al. Res. Microbiol. 2001	360 bps
	Poly Reverse	ATSGCCATCATYTCRCCGGA		
bphA	BPHD-f3	AACTGGAARTTYGCIGCVGA	Shoko et al. ISME J, 2009	542 bps
	BPHD-r1	ACCCAGTTYTCICCRTCGTC		
nirK	F1aCu	ATCATGGTSTCTGCCGCG	Michotey et al. Appl. Environ. Microbiol. 2000	472 bps
	R3Cu-GC	GCCTCGATCAGRITGTGGTT		
16S rRNA	27F	GAGTTTGATCMTGGCTCAG		492 bps
	518R	WTTACCGCGGCTGCTGG		

[0029] 삭제

[0030] (2) DNA 템플레이트 제조

[0031] 본 발명에서 사용한 20종의 미생물 인공유전체는 도 9에 나타내었다. 각 미생물의 Genomic DNA를 추출하여 Nanodrop을 이용해서 DNA 농도를 측정하였으며 동일한 DNA 농도를 섞어서 인공유전체를 제조하였다 (표 3).

표 3

Gene	(gram)Phylum	Prepared Strains (20 strains)	#of 16S	GC %	Mismatch		(ng/ml)/(260/280)
					For	Rev	
<i>nifH</i>	(-)Cyano*	Anabaena sp. PCC 9109	?		1	-	157/2.00
	(-)Beta	Burkholderia vietnamsis G4	6	66	0	1	158/1.91
	(-)Beta	Burkholderia xenovorans LB400	6	62	0	0	260/1.86
	(-)Beta	Polaromonas naphthalenivorans CJ2	2	62	0	0	61/1.85
	(+)Firmicute	Desulfitobacterium hafniense DCB-2	5	47	1	2	23/1.88
	(-)Alpha	Rhodospirillum rubrum ATCC 11170	4	65	1	2	128/2.07
<i>bphA</i>	(-)Beta	Burkholderia xenovorans LB400	6	62	0	0	260/1.86
	(-)Beta	Polaromonas naphthalenivorans CJ2	2	62	1	1	61/1.85
	(+)Actino	Rhodococcus sp. RHA1	4	67	0	0	125/1.91
*	(-)Gamma	Pseudomonas putida F1	6	61	0	0	158/2.02
<i>nirK</i>	(-)Beta	Neisseria sicca ATCC 29256	?	50	-	-	51/1.86
	(-)Alpha	Ochrobactrum anthropi ATCC 49188	4	56	0	0	110/1.90
	(-)Beta	Chromobacterium violaceum ATCC 12472	8	64	-	-	58/1.78
	(-)Beta	Polaromonas naphthalenivorans CJ2	2	62	-	-	61/1.85
16S	(-)Gamma	Pseudomonas pickettii PKO1	?				272/1.9
	(-)Alpha	Sphingomonas yanoikuyae B1	?				23/1.95
*	(-)Gamma	Escherichia Coli K-12 sub W3110	7	50			56/1.81
	(+)Firmicute	Bacillus cereus ATCC 14579	13	35			17/1.77
	(+)Actino	Corynebacterium glutamine ATCC 13032	5	53			118/1.9
	(+)Firmicute	Staphylococcus epidermidis ATCC 12228	5	32			77/1.84
	(-)Gamma	Xanthomonas campestris pv. campestris str. A TCC 33913	2	65			42/1.81
	(-)Alpha	Roseobacter denitrificans OCh 114	1	58			85/1.91
	(-)Alpha	Rhodobacter sphaeroides KD131	4	69			100/1.92

? : not Genome sequencing

[0032] 삭제

[0033] 천연 유전체로는 하기 표 4와 같은 물질을 사용하였다.

표 4

시료명	제공자	비고
토양 1	MSU	Performed Illumina V4 16S
강 침전물	연세대학교	강 침전물 (원주천)
조습지	연세대학교	조습지 (강화도)
생물양극	연세대학교	혐기 조건하에 1달 운영

[0034] 삭제

[0035] (3) 파이로시퀀싱할 8개 부분의 제조

[0036] 상기에서 수득한 인공유전체 및 천연 유전체, 어댑터, 바코드, 링커, 특이 프라이머 등을 이용하여 하기 표 5에 기재된 8개 플레이트를 제조하고, 이들을 region 1 내지 8이라 명명하였다. 이 때, 순방향(본원 명세서에서 'for'로 약칭되기도 함) 프라이머 어댑터로는 CGTATGCGCTCCCTCGCGCCATCAG 서열(Roche 사)이 사용되었고, 역방향(본원 명세서에서 'rev'로 약칭되기도 함) 프라이머 어댑터로는 CTATGCGCCTTGCCAGCCGCTCAG 서열(Roche 사)이 사용되었다 (도 11).

표 5

1 플레이트 (region 1)	
DNA 템플레이트	각각의 인공유전체 및 4개 천연 미생물군집으로부터의 gDNA를 동일한 비로 혼합
프라이머 조성	어댑터 + 바코드 + 링커 + 특이 프라이머
타겟 유전자	nifH, bphA, nirK, 16S
2 플레이트 (region 2)	

DNA 템플레이트	각각의 인공유전체 및 4개 천연 미생물군집 공급원으로부터의 gDNA를 동일한 비로 혼합
프라이머 조성	바코드 + 링커 + 특이 프라이머
타겟 유전자	nifH, bphA, nirK, 16S
3 플레이트 (region 3)	
DNA 템플레이트	인공유전체로부터의 gDNA를 5개의 상이한 비로 혼합
프라이머 조성	어댑터 + 바코드 + 링커 + 특이 프라이머
타겟 유전자	nifH, bphA, nirK, 16S
4 플레이트 (region 4)	
DNA 템플레이트	인공유전체로부터의 gDNA를 5개의 상이한 비로 혼합
프라이머 조성	바코드 + 링커 + 특이 프라이머
타겟 유전자	nifH, bphA, nirK, 16S
5 플레이트 (region 5)	
DNA 템플레이트	특정 유전자를 갖는 균주로부터의 개별 gDNA
프라이머 조성	어댑터 + 바코드 + 링커 + 특이 프라이머
타겟 유전자	nifH, bphA, nirK, 16S
6 플레이트 (region 6)	
DNA 템플레이트	인공유전체로부터의 gDNA를 5개의 상이한 비로 혼합
프라이머 조성	어댑터 + 교체된 바코드 + 링커 + 특이 프라이머
타겟 유전자	nifH, bphA
DNA 템플레이트	인공유전체 및 천연 미생물군집 공급원으로부터의 gDNA를 6개의 상이한 비로 혼합
프라이머 조성	어댑터 + 바코드 + 링커 + 특이 프라이머
타겟 유전자	16S rRNA
7 플레이트 (region 7)	
1 플레이트 반복	
8 플레이트 (region 8)	
2 플레이트 반복	

6 플레이트에서 인공유전체: 천연 유전체 공급원의 비는 하기 표 6과 같다.

표 6

천연물	ratio 1	ratio 2	ratio 3
토양 1	0.3:9	1:9	3:9
	ratio 4	ratio 5	ratio 6
생물양극	0.3:9	1:9	3:9

[0037] 삭제

[0038] 삭제

[0039] 삭제

[0040] 삭제

[0041] 삭제

[0042] 삭제

[0043] 삭제

[0044] 삭제

[0045] 삭제

[0046] 삭제

하기 표 7은 1, 2, 7, 8 플레이트 프라이머 세트에 사용된 바코드를 나타낸다.

표 7

타겟 유전자	인공유전체	토양 1	토양 2	조습지	생물양극
nifH	BC1	BC2	BC3	BC4	BC5
bphA	BC6	BC7	BC8	BC9	BC10
nirK	BC11	BC12	BC13	BC14	BC15
16S rRNA	BC16	BC17	BC18	BC19	BC20

[0048] 삭제

3, 4 플레이트의 프라이머 세트에 사용된 바코드는 하기 표 8과 같다.

표 8

타겟 유전자	Ratio 1	Ratio 2	Ratio 3	Ratio 4	Ratio 5
nifH	BC1	BC2	BC3	BC4	BC5
bphA	BC6	BC7	BC8	BC9	BC10
nirK	BC11	BC12	BC13	BC14	BC15
16S rRNA	BC16	BC17	BC18	BC19	BC20

[0050] 삭제

5 플레이트의 프라이머 세트에 사용된 바코드는 하기 표 9와 같다.

표 9

타겟 유전자	균주 1	균주 2	균주 3	균주 4	균주 5
nifH	BC1	BC2	BC3	BC4	BC5
bphA	BC6	BC7	BC8	BC9	BC10
nirK	BC11	BC12	BC13	BC14	BC15
16S rRNA	5개의 대표 균주가 BC16 - BC20에 의해 증폭된다.				

[0052] 삭제

6 플레이트의 프라이머 세트에 사용된 바코드는 하기 표 10과 같다.

표 10

타겟 유전자	Ratio 1	Ratio 2	Ratio 3	Ratio 4	Ratio 5
nifH	BC6	BC7	BC8	BC9	BC10
bphA	BC1	BC2	BC3	BC4	BC5
타겟유전자	Ratio 1	Ratio 2	Ratio 3	Ratio 4	Ratio 5
16S rRNA	BC16	BC17	BC18	BC19	BC20
	Ratio 6	More depth sequencing than others			
	BC21				

[0054] 삭제

상기 표 7 내지 표 10에서 사용된 바코드의 서열은 하기 표 11에 나타내었다.

표 11

Barcode	Sequence	Barcode	Sequence	Barcode	Sequence
BC1	ACACGTCA	BC8	CTAGAGCT	BC15	TGCAGATC
BC2	AGCTACGT	BC9	CTGTCAGA	BC16	ACACGACT
BC3	AGCTGTAC	BC10	CTGAGTCA	BC17	ACAGTCAC
BC4	ATATGCGC	BC11	TAGCTAGC	BC18	AGACGTCT
BC5	ACACACTG	BC12	TCAGACTG	BC19	AGTCACTG
BC6	CACTACAG	BC13	TCGACATG	BC20	ATCGTACG
BC7	CATGACGT	BC14	TGAGTCAC	BC21	CACATGTG

도 1은 표 5에 나타나 있는 1, 2, 6 내지 8 플레이트의 순수 시퀀스 (raw sequences)를 나타낸 그래프이다.

[0056] (4) 마스터믹스(Master mix) 준비

[0057] 2.5ul의 10X AccuPrime PCR 완충액 II, 0.2ul의 Accuprime Taq Hifi, 1.5ul/1.5ul(Forward/Reverse 프라이머 각각의 볼륨)의 혼합 프라이머 템플레이트 DNA(60ng), 및 총 부피가 25ul이 되도록 하는 잔량의 RNase/DNase를 함유하지 않은 물을 준비하였다.

[0058]

[0059] (5) PCR 준비

[0060] 상기에서 획득한 25uL의 PCR 튜브를 원심분리기에서 2000rpm으로 잠깐동안 교반하였다. 써모사이클러에 넣고, 하기 표 12의 조건으로 PCR을 실시하였다.

표 12

nifH	94℃	1분	30사이클	bphA	95℃	3분	30사이클
	94℃	1분			95℃	45초	
	55℃	1분			60℃	45초	
	72℃	2분			72℃	40초	
	72℃	5분			72℃	4분	
	4℃	계속(never)			4℃	계속	
nirK	95℃	1분	30 사이클	16S	94℃	3분	30사이클
	94℃	1분			94℃	30초	
	51℃	1분			55℃	30초	
	72℃	1분			72℃	1분	
	72℃	10분			72℃	5분	
	4℃	계속			4℃	계속	

[0061] 삭제

[0062] QIAquick PCR Purification Kit를 이용하여 PCR 생성물을 정제하였다.

[0063] (6) PCR 겔 분석

[0064] 파라필름 상에서 8uL 탈염증류수 및 1uL 10X 로딩 염료에 1uL PCR 생성물을 첨가한 후 피펫팅으로 혼합하였다. Safeview에 의해 1% 아가로스 1X TAE 겔을 준비하였다. 시료를 로딩하고, 약 1시간동안 100V에서 전기연동을 실시하였다. gel-doc 상에서 겔 이미지를 포착하고, 분석을 위해 보관하였다.

[0065] (7) PCR 생성물 정량화

[0066] 나노드롭 분광광도계를 제조업체 설명서에 따라 이용하여 PCR 생성물을 정량화하였으며, 그 결과 얻어진 농도(concentration)를 도 12a 내지 도 12f에 나타내었다.

[0067] (8) PCR 풀링(pooling)

[0068] 나노드롭 분광광도계로부터 취득한 값을 이용하여, poolingCalculator.xls를 이용하거나 하기 식을 이용하여 풀링 양을 계산하였으며, 그 결과는 도 12a 내지 도 12f의 부피(volume) 항목에 나타내었다.

[0069] 각 시료의 양(uL) = ((부피/2) X (min))/Sampleconc

[0070] 상기에서, '부피'는 각 시료의 전체 부피를 의미하고, 'min'은 가장 낮은 농도를 갖는 시료 농도(ng/ul 단위)를 의미하며, 'Sampleconc'은 타겟 시료의 농도(ng/ul 단위)를 의미한다.

1uL의 최소 이동 부피를 이용하여 시료를 풀링하였다. 1uL 미만이 요청되는 경우에는 희석하여야 한다.

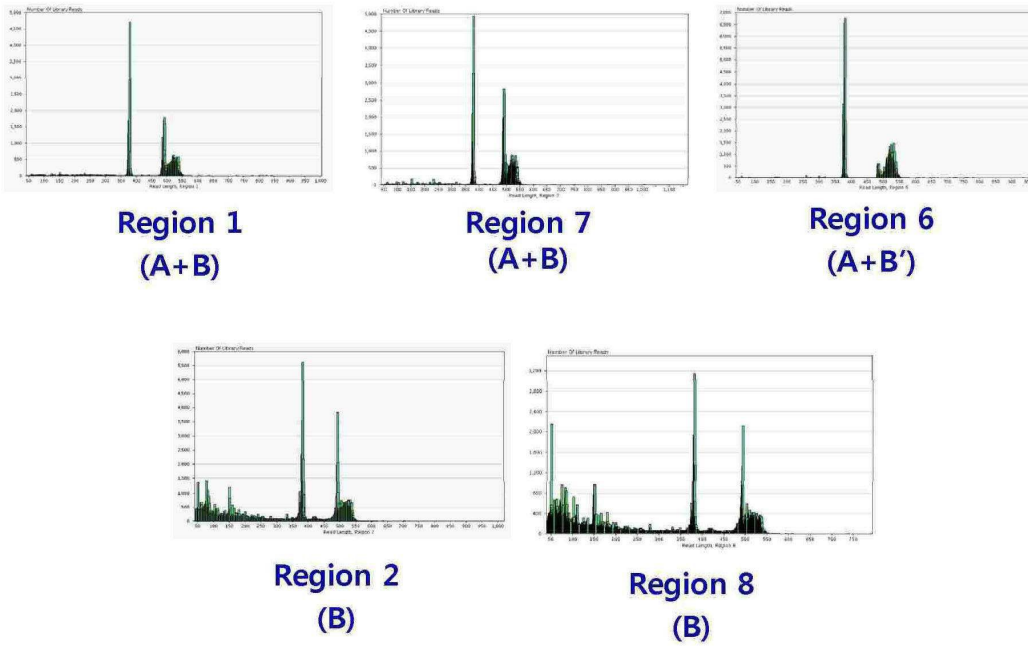
[0071] 삭제

[0072] Qiagen minElute 컬럼을 이용하여 1x low TE, pH 8.0으로 용출하면서 제조업체 설명서에 따라 풀을 정제하였다. 정제율을 높이기 위해서 추가적인 정제를 진행하였으며 정제결과 흡광도 260/280이 모두 1.80 이상으로 나왔다.

- [0073] (9) 파이로시퀀싱
- [0074] 마크로젠 (회사명)에서 Genome Sequencing FLX titanium pyrosequencing을 이용하여 Roche에서 제공하는 방법대로 시퀀싱을 진행하였다.
- [0075] (10) 표준 시퀀스 선정(Standard sequencing collection)
- [0076] 인공유전체에서 최적의 염기서열을 선별하기 위해 본 발명자들은 2가지 방법으로 표준염기서열을 선정하였다. NCBI 게놈 데이터베이스에서 3개 유전자를 선별하여 특정 프라이머(nifH, bphA, 16S rRNA)를 이용하여 프라이모 유사성 비교(Probe match)를 진행하였다. 동일하지 않은 시퀀스들을 제거하기 위하여 각 유전자의 HMM(Hidden Markov Model) 모델을 사용하여 모든 시퀀스들을 각각의 유전자들과 유효성 평가하였다.
- [0077] (11) 초기의 필터링 관리(Initial Processing)
- [0078] 품질 좋은 시퀀스를 획득하기 위하여 RDP Pyro Initial Process tool [Cole 등, 2009]을 이용하여 순방향 프라이머에 2개 이상의 mismatch 또는 average exponential quality 점수가 0 이상인 sequence들을 걸러냈다. 순방향 프라이머 이전의 염기들은 reads로부터 잘라냈다. Reads가 길기 때문에 역방향 프라이머는 16S와 bphA에 대한 검증을 하지 않았다. nifH reads의 경우 역방향 프라이머와 완벽히 일치해야 하고, 역방향 프라이머는 잘라내었다. 또한, ambiguity 염기 또는 trimming 과정을 거친 후의 길이가 300bps 보다 짧은 read는 버려졌다(도 2).
- [0079] (12) 오염된 시퀀스 검출(Contamination Detection)
- [0080] 초기 품질관리를 통과한 reads는 특별제작된 RDP tool인 ContaminateBot를 이용하여 분석했다. reads들은 RDP Seqmatch tool을 이용하여 높은 질의 RDP public dataset와 인공유전체 서열들과 비교하였다. S_ab 점수차가 0.2 이상이고 인공유전체 보다 RDP public dataset의 sequence 정보가 가까운 reads들은 오염된 sequence로 판단하여 제거하였다.
- [0081] (13) 키메라 검출(Chimera Detection)
- [0082] Potential Chimera를 분별하기 위하여 특별제작된 RDP tool인 ChimeraBot를 이용하여 3% 이상 오차를 갖는 reads를 분석하였다. ChimeraBot는 각각의 read별로 표준 인공유전체 서열 대비 5' 또는 3'에서 시작하는 partial alignments를 만든다. 순방향 및 역방향 alignment들로 가능한 모든 조합의 최대 점수, 해당 좌우 인공유전체 parents, alignment breakpoint 정보를 습득하였다. 종합점수가 최적의 single-parent alignment 점수보다 최소 10% 높고 각 partial alignment가 95% identity일 때 reads는 potential chimera로 가정하여 오차 계산에서 제거하였다.
- [0083] (14) 오차 분석(Error analysis)
- [0084] 초기 품질관리를 통과하고 contaminant가 아닌 reads는 RDP 인공유전체 Analysis tool (<http://pyro.cme.msu.edu/>)를 이용하여 표준 인공유전체 sequence와 비교하였다. 각각의 read는 표준 인공유전체 sequence 간의 alignment를 계산하였고, 최적의 alignment 대비 최고로 근접한 standard sequence를 획득하였다(도 3). 이 최적의 alignment를 토대로 삽입, 결실(indel)와 mismatch error를 계산하였다.
- [0085] 각 유전자(바코드 + 어댑터/바코드/방향성) 별로 총 indel과 mismatch 결과를 파이로시퀀싱 시퀀스 결과로 나누어 전체 오차율을 계산하였고(도 4), 시퀀싱 방향성 별로 차이를 확인하기 위하여 mismatch 누적 곡선을 그려 확인하였다(도 5).
- [0086] 각 유전자별 오차 갯수에 따른 분포를 파악하기 위하여 누적 오차 분포 그래프를 작성하였다(도 6 내지 도 8).

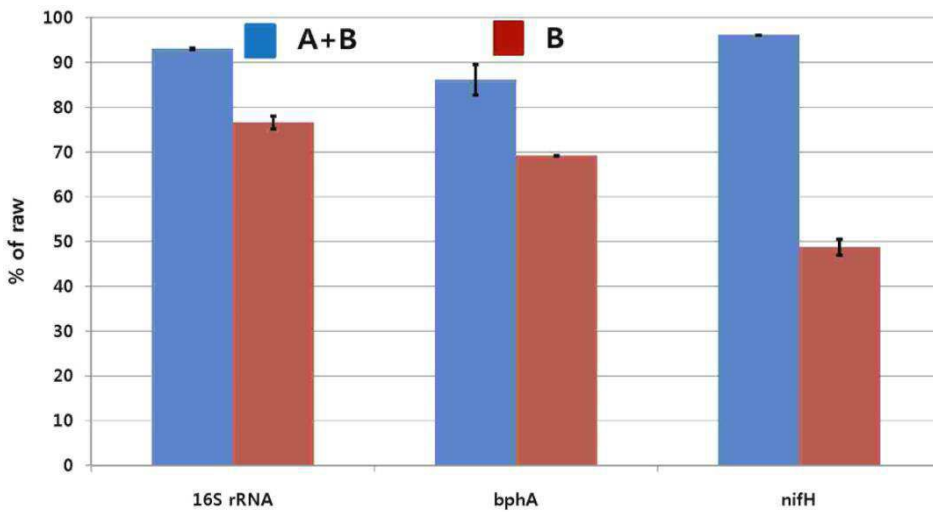
도면

도면1



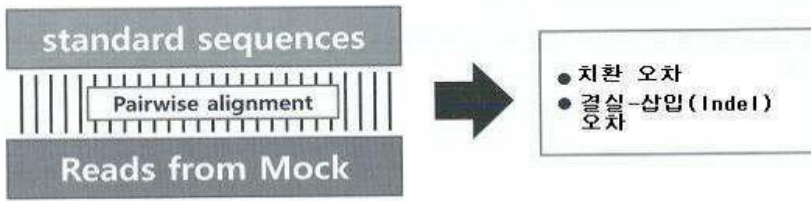
도면2

Initial process in RDP's pyrosequencing pipeline
 Length \geq 300 bps, N = 0, Primer mismatch \leq 2

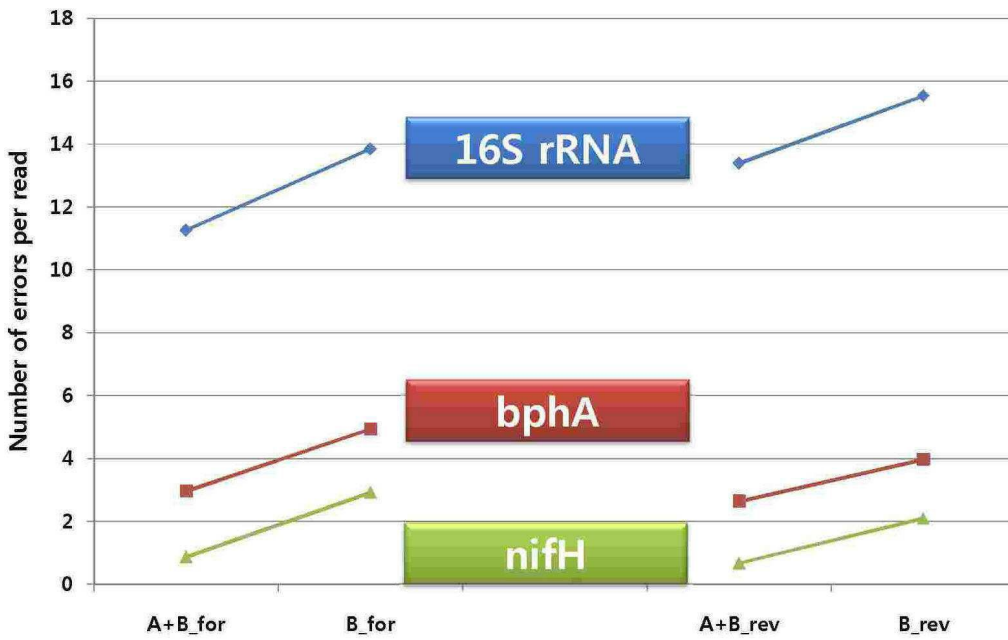


도면3

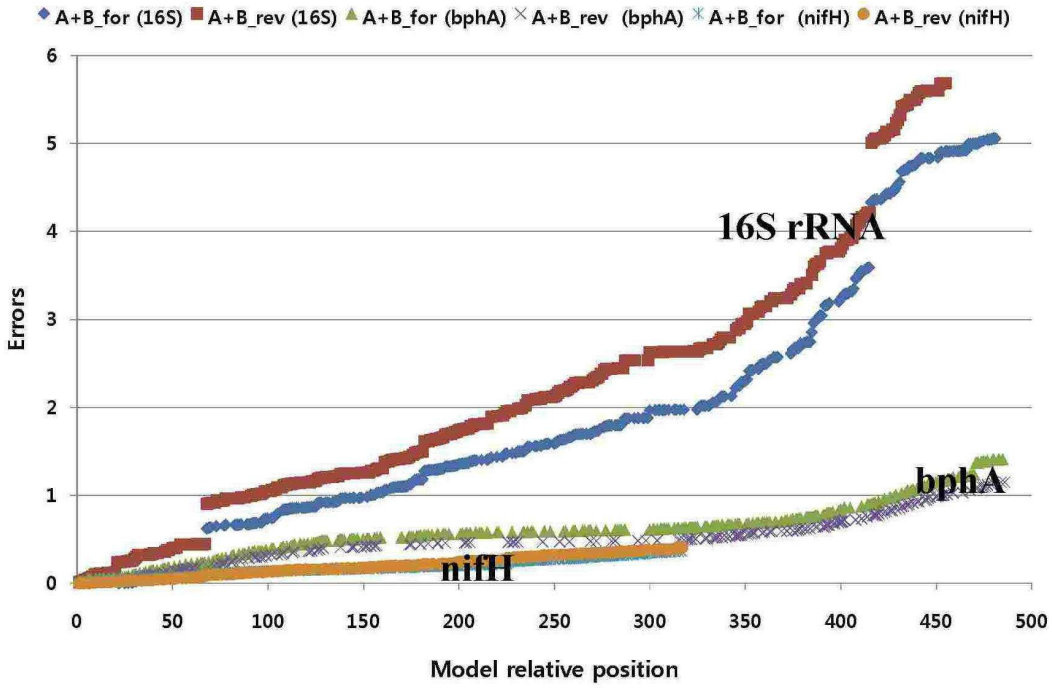
오차 분석 (동적 프로그래밍)



도면4

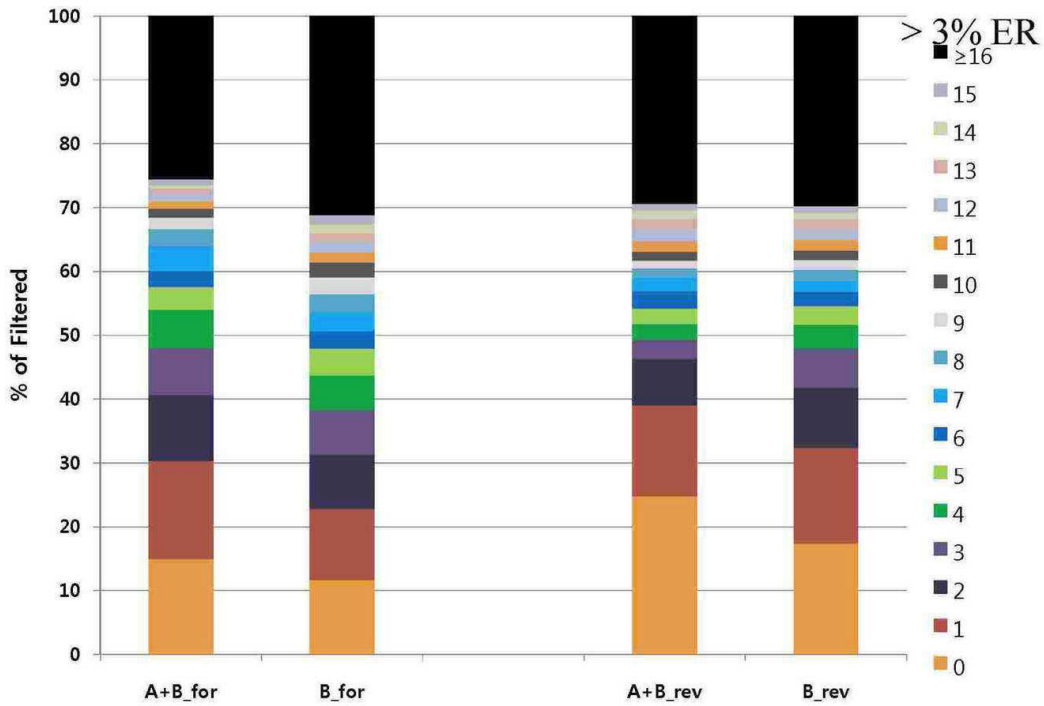


도면5



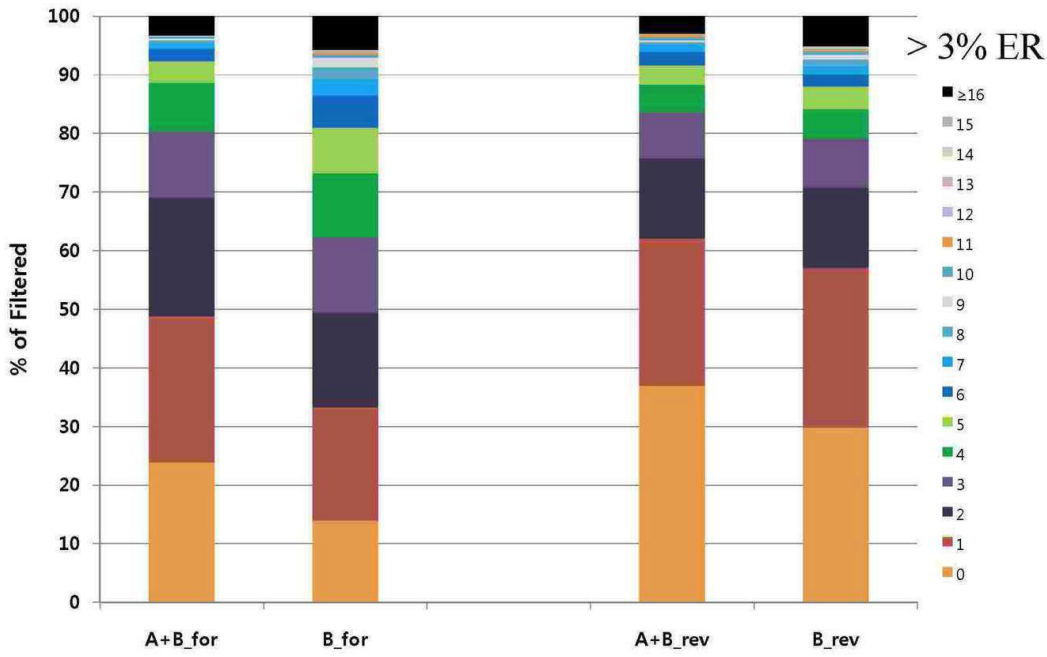
도면6

오차 분포 (16S rRNA)



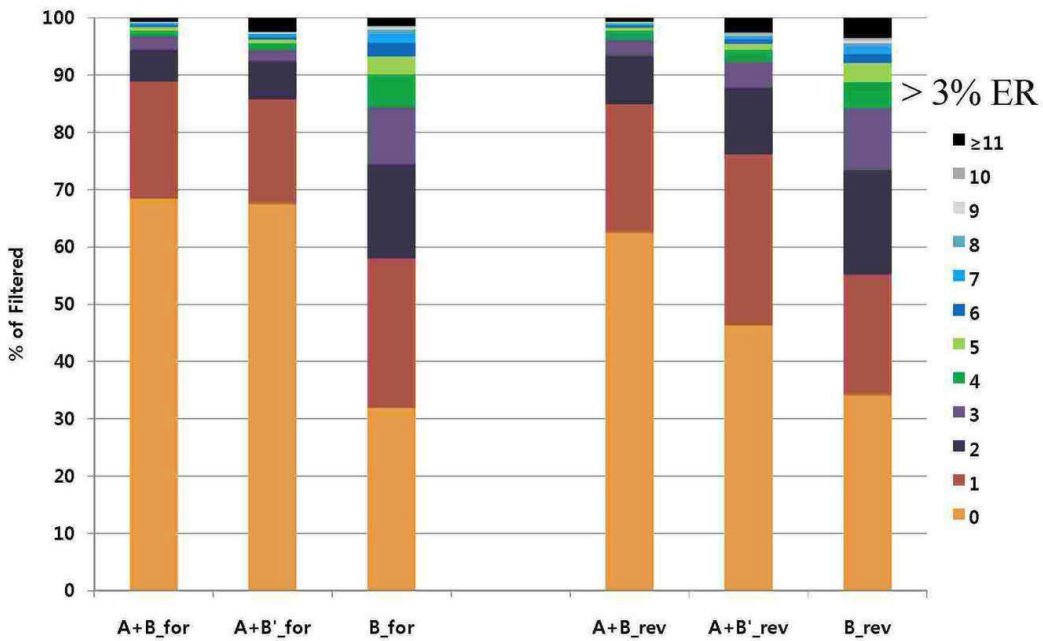
도면7

오차 분포 (bphA)



도면8

오차 분포 (nifH)



도면9

Mock Community

Strains	Genome size (bp)	Gene (# of gene copy)
Rhodospirillum rubrum ATCC 11170	4,406,557	nifH (5), 16S (1)
Burkholderia vietamensis G4	8,391,070	nifH (1), 16S (2)
Burkholderia xenovorans LB400	9,731,138	nifH (1), bphA (1), 16S (2)
Desulfitobacterium hafniense DCB-2	5,279,134	nifH (4), 16S (5)
Nostoc. PCC 7120	7,211,789	nifH(3), 16S(2)
Polaromonas naphthalenivorans CJ2	5,366,143	nifH (1), bphA (1), 16S (1)
Rhodococcus sp. RHA1	9,702,737	bphA (6), 16S (1)
Pseudomonas putida F1	5,959,964	bphA (1), 16S (3)
Neisseria sicca ATCC 29256	2,830,772	16S (1)
Ochrobactrum anthropi ATCC 49188	5,205,777	16S (1)
Chromobacterium violaceum ATCC 12472	4,751,080	16S (1)
Pseudomonas pickettii PKO1	5,325,729	16S (1)
Sphingobium yanoikuyae B1	5,915,246	16S (1)
Escherichia Coli K-12 sub W3110	4,646,332	16S (6)
Bacillus cerus ATCC 14579	5,427,083	16S (7)
Corynebacterium glutamin ATCC 13032	3,282,708	16S (9)
Staphylococcus epidemidis ATCC 12228	2,564,615	16S (1)
Xanthomonas campestris py. ATCC 33913	5,076,188	16S (1)
Roseobacter denitrifican Och 114	4,331,234	16S (1)
Rhodobacter sphaeroides KD 131	4,711,139	nifH (3), 16S (1)
20 strains		nifH (18), bphA (9), 16S (48)

도면10

Forward primer		Reverse primer	
Primer	Sequence (5'-3')	Primer	Sequence (5'-3')
nifH_f1	ACACGTCA-AC-TGCGAYCCSAARGCBGACTC	nifH_r1	ACACGTCA-AC-ATSGCCATCAITYCRCCGGA
nifH_f2	AGCTACGT-AC-TGCGAYCCSAARGCBGACTC	nifH_r2	AGCTACGT-AC-ATSGCCATCAITYCRCCGGA
nifH_f3	AGCTGTAC-AC-TGCGAYCCSAARGCBGACTC	nifH_r3	AGCTGTAC-AC-ATSGCCATCAITYCRCCGGA
nifH_f4	ATATGCGC-AC-TGCGAYCCSAARGCBGACTC	nifH_r4	ATATGCGC-AC-ATSGCCATCAITYCRCCGGA
nifH_f5	ACACACTG-AC-TGCGAYCCSAARGCBGACTC	nifH_r5	ACACACTG-AC-ATSGCCATCAITYCRCCGGA
bphA_f1	CACTACAG-AC-AACTGGAARTTYGCGCVGA	bphA_r1	CACTACAG-AC-ACCCAGTTYCICCRITCGTC
bphA_f2	CATGACGT-AC-AACTGGAARTTYGCGCVGA	bphA_r2	CATGACGT-AC-ACCCAGTTYCICCRITCGTC
bphA_f3	CTAGAGCT-AC-AACTGGAARTTYGCGCVGA	bphA_r3	CTAGAGCT-AC-ACCCAGTTYCICCRITCGTC
bphA_f4	CTGTGAGA-AC-AACTGGAARTTYGCGCVGA	bphA_r4	CTGTGAGA-AC-ACCCAGTTYCICCRITCGTC
bphA_f5	CTGAGTCA-AC-AACTGGAARTTYGCGCVGA	bphA_r5	CTGAGTCA-AC-ACCCAGTTYCICCRITCGTC
nirK_f1	TAGCTAGC-AC-ATCATGGTISCTGCCGCG	nirK_r1	TAGCTAGC-AC-GCCTCGATCAGRITGTGGTT
nirK_f2	TCAGACTG-AC-ATCATGGTISCTGCCGCG	nirK_r2	TCAGACTG-AC-GCCTCGATCAGRITGTGGTT
nirK_f3	TCGACATG-AC-ATCATGGTISCTGCCGCG	nirK_r3	TCGACATG-AC-GCCTCGATCAGRITGTGGTT
nirK_f4	TGAGTCAC-AC-ATCATGGTISCTGCCGCG	nirK_r4	TGAGTCAC-AC-GCCTCGATCAGRITGTGGTT
nirK_f5	TGCAGATC-AC-ATCATGGTISCTGCCGCG	nirK_r5	TGCAGATC-AC-GCCTCGATCAGRITGTGGTT
Forward primer		Reverse primer	
Primer	Sequence (5'-3')	Primer	Sequence (5'-3')
bphA_iso1	ACACGTCA-AC-AACTGGAARTTYGCGCVGA	bphA_iso1	ACACGTCA-AC-ACCCAGTTYCICCRITCGTC
bphA_iso2	AGCTACGT-AC-AACTGGAARTTYGCGCVGA	bphA_iso2	AGCTACGT-AC-ACCCAGTTYCICCRITCGTC
bphA_iso3	AGCTGTAC-AC-AACTGGAARTTYGCGCVGA	bphA_iso3	AGCTGTAC-AC-ACCCAGTTYCICCRITCGTC
bphA_iso4	ATATGCGC-AC-AACTGGAARTTYGCGCVGA	bphA_iso4	ATATGCGC-AC-ACCCAGTTYCICCRITCGTC
bphA_iso5	ACACACTG-AC-AACTGGAARTTYGCGCVGA	bphA_iso5	ACACACTG-AC-ACCCAGTTYCICCRITCGTC
nifH_iso1	CACTACAG-AC-TGCGAYCCSAARGCBGACTC	nifH_iso1	CACTACAG-AC-ATSGCCATCAITYCRCCGGA
nifH_iso2	CATGACGT-AC-TGCGAYCCSAARGCBGACTC	nifH_iso2	CATGACGT-AC-ATSGCCATCAITYCRCCGGA
nifH_iso3	CTAGAGCT-AC-TGCGAYCCSAARGCBGACTC	nifH_iso3	CTAGAGCT-AC-ATSGCCATCAITYCRCCGGA
nifH_iso4	CTGTGAGA-AC-TGCGAYCCSAARGCBGACTC	nifH_iso4	CTGTGAGA-AC-ATSGCCATCAITYCRCCGGA
nifH_iso5	CTGAGTCA-AC-TGCGAYCCSAARGCBGACTC	nifH_iso5	CTGAGTCA-AC-ATSGCCATCAITYCRCCGGA

도면11

Forward primer		Reverse primer	
Primer	Sequence (5'-3')	Primer	Sequence (5'-3')
nifH_fo1	CGTATGGCCCTCCCTGGGCGCA TDAG-A CACGCTCA-AC-TGCGAYCCSAARGCBGA CTC	nifH_ro1	CTATGGCCCTTSCCAGCCGCGCTCAG-ACA CGTCA-AC-ATSGCCATCATY TDRCCG SA
nifH_fo2	CGTATGGCCCTCCCTGGGCGCA TDAG-AGCTACTGCT-AC-TGCGAYCCSAARGCBGA CTC	nifH_ro2	CTATGGCCCTTSCCAGCCGCGCTCAG-AGCTACTGCT-AC-ATSGCCATCATY TDRCCG SA
nifH_fo3	CGTATGGCCCTCCCTGGGCGCA TDAG-A TA TGGCC-AC-TGCGAYCCSAARGCBGA CTC	nifH_ro3	CTATGGCCCTTSCCAGCCGCGCTCAG-ATA TGGCC-AC-ATSGCCATCATY TDRCCG SA
nifH_fo4	CGTATGGCCCTCCCTGGGCGCA TDAG-A CACTACTG-AC-TGCGAYCCSAARGCBGA CTC	nifH_ro4	CTATGGCCCTTSCCAGCCGCGCTCAG-ACA CACTG-AC-ATSGCCATCATY TDRCCG SA
nifH_fo5	CGTATGGCCCTCCCTGGGCGCA TDAG-CACTACAG-AC-AACTGGAARTY TGCGICV SA	nifH_ro5	CTATGGCCCTTSCCAGCCGCGCTCAG-CACTACAG-AC-CCCACTY TDCORCTGT C
hpaA_fo1	CGTATGGCCCTCCCTGGGCGCA TDAG-CA TBA GST-AC-AACTGGAARTY TGCGICV SA	hpaA_ro1	CTATGGCCCTTSCCAGCCGCGCTCAG-CA TBA GST-AC-ACCCAG TTY TDCORCTGT C
hpaA_fo2	CGTATGGCCCTCCCTGGGCGCA TDAG-CTAGAGCT-AC-AACTGGAARTY TGCGICV SA	hpaA_ro2	CTATGGCCCTTSCCAGCCGCGCTCAG-CTAGAGCT-AC-ACCCAG TTY TDCORCTGT C
hpaA_fo3	CGTATGGCCCTCCCTGGGCGCA TDAG-CTGTGAGA-AC-AACTGGAARTY TGCGICV SA	hpaA_ro3	CTATGGCCCTTSCCAGCCGCGCTCAG-CTGTGAGA-AC-ACCCAG TTY TDCORCTGT C
hpaA_fo4	CGTATGGCCCTCCCTGGGCGCA TDAG-CTGAGTCA-AC-AACTGGAARTY TGCGICV SA	hpaA_ro4	CTATGGCCCTTSCCAGCCGCGCTCAG-CTGAGTCA-AC-ACCCAG TTY TDCORCTGT C
hpaA_fo5	CGTATGGCCCTCCCTGGGCGCA TDAG-TCGADATG-AC-ATCATG GTCTG CCGCG	hpaA_ro5	CTATGGCCCTTSCCAGCCGCGCTCAG-TCGADATG-AC-GCCTGSA TCGAGTTGTGG TT
nifK_fo1	CGTATGGCCCTCCCTGGGCGCA TDAG-TAGTACG-AC-ATCATG GTCTG CCGCG	nifK_ro1	CTATGGCCCTTSCCAGCCGCGCTCAG-TAGTACG-AC-GCCTGSA TCGAGTTGTGG TT
nifK_fo2	CGTATGGCCCTCCCTGGGCGCA TDAG-TCGAGCTG-AC-ATCATG GTCTG CCGCG	nifK_ro2	CTATGGCCCTTSCCAGCCGCGCTCAG-TCGAGCTG-AC-GCCTGSA TCGAGTTGTGG TT
nifK_fo3	CGTATGGCCCTCCCTGGGCGCA TDAG-TCGADATG-AC-ATCATG GTCTG CCGCG	nifK_ro3	CTATGGCCCTTSCCAGCCGCGCTCAG-TCGADATG-AC-GCCTGSA TCGAGTTGTGG TT
nifK_fo4	CGTATGGCCCTCCCTGGGCGCA TDAG-TGAGTCA-AC-ATCATG GTCTG CCGCG	nifK_ro4	CTATGGCCCTTSCCAGCCGCGCTCAG-TGAGTCA-AC-GCCTGSA TCGAGTTGTGG TT
nifK_fo5	CGTATGGCCCTCCCTGGGCGCA TDAG-TCGAGATC-AC-ATCATG GTCTG CCGCG	nifK_ro5	CTATGGCCCTTSCCAGCCGCGCTCAG-TCGAGATC-AC-GCCTGSA TCGAGTTGTGG TT
Forward primer		Reverse primer	
Primer	Sequence (5'-3')	Primer	Sequence (5'-3')
ZFF_o1	CGTATGGCCCTCCCTGGGCGCA TDAG-A CAGTCACT-AC-GAGTTTGATCM TGGCTC AG	538r_o1	CTATGGCCCTTSCCAGCCGCGCTCAG-ACA CGACT-AC-WTTACCGCGGCTGCTGG
ZFF_o2	CGTATGGCCCTCCCTGGGCGCA TDAG-A CAGTCACT-AC-GAGTTTGATCM TGGCTC AG	538r_o2	CTATGGCCCTTSCCAGCCGCGCTCAG-ACA CGTCA-AC-WTTACCGCGGCTGCTGG
ZFF_o3	CGTATGGCCCTCCCTGGGCGCA TDAG-AGA CGTCT-AC-GAGTTTGATCM TGGCTC AG	538r_o3	CTATGGCCCTTSCCAGCCGCGCTCAG-AGA CGTCT-AC-WTTACCGCGGCTGCTGG
ZFF_o4	CGTATGGCCCTCCCTGGGCGCA TDAG-AGTCACTG-AC-GAGTTTGATCM TGGCTC AG	538r_o4	CTATGGCCCTTSCCAGCCGCGCTCAG-AGTCACTG-AC-WTTACCGCGGCTGCTGG
ZFF_o5	CGTATGGCCCTCCCTGGGCGCA TDAG-A TGGTACG-AC-GAGTTTGATCM TGGCTC AG	538r_o5	CTATGGCCCTTSCCAGCCGCGCTCAG-ATGGTACG-AC-WTTACCGCGGCTGCTGG
ZFF_o6	CGTATGGCCCTCCCTGGGCGCA TDAG-CACATGTA-AC-GAGTTTGATCM TGGCTC AG	538r_o6	CTATGGCCCTTSCCAGCCGCGCTCAG-CACATGTA-AC-WTTACCGCGGCTGCTGG
Forward primer		Reverse primer	
Primer	Sequence (5'-3')	Primer	Sequence (5'-3')
hpaA_jso1	CGTATGGCCCTCCCTGGGCGCA TDAG-A CACTGCA-AC-AACTGGAARTY TGCGICV SA	hpaA_jso1	CTATGGCCCTTSCCAGCCGCGCTCAG-ACA CGTCA-AC-CCCACTY TDCORCTGT C
hpaA_jso2	CGTATGGCCCTCCCTGGGCGCA TDAG-AGCTACTGCT-AC-AACTGGAARTY TGCGICV SA	hpaA_jso2	CTATGGCCCTTSCCAGCCGCGCTCAG-AGCTACTGCT-AC-ACCCAG TTY TDCORCTGT C
hpaA_jso3	CGTATGGCCCTCCCTGGGCGCA TDAG-AGCTACTGCT-AC-AACTGGAARTY TGCGICV SA	hpaA_jso3	CTATGGCCCTTSCCAGCCGCGCTCAG-AGCTACTGCT-AC-ACCCAG TTY TDCORCTGT C
hpaA_jso4	CGTATGGCCCTCCCTGGGCGCA TDAG-A TA TGGCC-AC-AACTGGAARTY TGCGICV SA	hpaA_jso4	CTATGGCCCTTSCCAGCCGCGCTCAG-ATA TGGCC-AC-ACCCAG TTY TDCORCTGT C
hpaA_jso5	CGTATGGCCCTCCCTGGGCGCA TDAG-A CACTACTG-AC-AACTGGAARTY TGCGICV SA	hpaA_jso5	CTATGGCCCTTSCCAGCCGCGCTCAG-ACA CACTG-AC-ACCCAG TTY TDCORCTGT C
nifH_jso1	CGTATGGCCCTCCCTGGGCGCA TDAG-CACTACAG-AC-TGCGAYCCSAARGCBGA CTC	nifH_jso1	CTATGGCCCTTSCCAGCCGCGCTCAG-CACTACAG-AC-ATSGCCATCATY TDRCCG SA
nifH_jso2	CGTATGGCCCTCCCTGGGCGCA TDAG-CACTACAG-AC-TGCGAYCCSAARGCBGA CTC	nifH_jso2	CTATGGCCCTTSCCAGCCGCGCTCAG-CACTACAG-AC-ATSGCCATCATY TDRCCG SA
nifH_jso3	CGTATGGCCCTCCCTGGGCGCA TDAG-CTAGAGCT-AC-TGCGAYCCSAARGCBGA CTC	nifH_jso3	CTATGGCCCTTSCCAGCCGCGCTCAG-CTAGAGCT-AC-ATSGCCATCATY TDRCCG SA
nifH_jso4	CGTATGGCCCTCCCTGGGCGCA TDAG-CTGTGAGA-AC-TGCGAYCCSAARGCBGA CTC	nifH_jso4	CTATGGCCCTTSCCAGCCGCGCTCAG-CTGTGAGA-AC-ATSGCCATCATY TDRCCG SA
nifH_jso5	CGTATGGCCCTCCCTGGGCGCA TDAG-CTGAGTCA-AC-TGCGAYCCSAARGCBGA CTC	nifH_jso5	CTATGGCCCTTSCCAGCCGCGCTCAG-CTGAGTCA-AC-ATSGCCATCATY TDRCCG SA
Forward primer		Reverse primer	
Primer	Sequence (5'-3')	Primer	Sequence (5'-3')
ZFF_M1	CGTATGGCCCTCCCTGGGCGCA TDAG-A CACTGAG-GAGTTTGATCM TGGCTC TAG	538R_M1	CTATGGCCCTTSCCAGCCGCGCTCAG-ACA CTGAG-WTTACCGCGGCTGCTGG
ZFF_M2	CGTATGGCCCTCCCTGGGCGCA TDAG-A CACTGAGA-GAGTTTGATCM TGGCTC TAG	538R_M2	CTATGGCCCTTSCCAGCCGCGCTCAG-ACA CGAGA-WTTACCGCGGCTGCTGG
ZFF_M3	CGTATGGCCCTCCCTGGGCGCA TDAG-GCTTGA-GAGTTTGATCM TGGCTC TAG	538R_M3	CTATGGCCCTTSCCAGCCGCGCTCAG-GCTTGA-WTTACCGCGGCTGCTGG
ZFF_M4	CGTATGGCCCTCCCTGGGCGCA TDAG-A CACTGAG-GAGTTTGATCM TGGCTC TAG	538R_M4	CTATGGCCCTTSCCAGCCGCGCTCAG-ACA CACTGAG-WTTACCGCGGCTGCTGG
ZFF_M5	CGTATGGCCCTCCCTGGGCGCA TDAG-A CACTGAG-GAGTTTGATCM TGGCTC TAG	538R_M5	CTATGGCCCTTSCCAGCCGCGCTCAG-ACTGAGT-WTTACCGCGGCTGCTGG
ZFF_M6	CGTATGGCCCTCCCTGGGCGCA TDAG-A CACTGAGT-GAGTTTGATCM TGGCTC TAG	538R_M6	CTATGGCCCTTSCCAGCCGCGCTCAG-ACACTGAGT-WTTACCGCGGCTGCTGG

도면12a

Description	Concentration	1 μg to ng	1 mol bp DNA to g DNA	mol bp to bp size (bp)	Added mass (ng)	volume	gene copies
#1 Adapter_Same_16S	115.7	1E+09	660	6.023E+23	492	500	4.32 9.27414E+11
#1 Adapter_MSU_16S	74.75	1E+09	660	6.023E+23	492	500	6.69 9.27414E+11
#1 Adapter_River_16S	34.88	1E+09	660	6.023E+23	492	500	14.33 9.27414E+11
#1 Adapter_Tidal_16S	101.08	1E+09	660	6.023E+23	492	500	4.95 9.27414E+11
#1 Adapter_Biocathode_16S	93.08	1E+09	660	6.023E+23	492	500	5.37 9.27414E+11
#1 Adapter_Same_nifH	65.49	1E+09	660	6.023E+23	360	380	5.59 9.27414E+11
#1 Adapter_MSU_nifH	10.56	1E+09	660	6.023E+23	360	380	34.65 9.27414E+11
#1 Adapter_River_nifH	72.45	1E+09	660	6.023E+23	360	380	5.05 9.27414E+11
#1 Adapter_Tidal_nifH	46.1	1E+09	660	6.023E+23	360	380	7.94 9.27414E+11
#1 Adapter_Biocathode_nifH	40.6	1E+09	660	6.023E+23	360	380	9.01 9.27414E+11
#1 Adapter_Same_nirK	47.25	1E+09	660	6.023E+23	472	380	10.15 9.27414E+11
#1 Adapter_MSU_nirK	24.29	1E+09	660	6.023E+23	472	380	19.75 9.27414E+11
#1 Adapter_River_nirK	34.59	1E+09	660	6.023E+23	472	380	13.87 9.27414E+11
#1 Adapter_Tidal_nirK	25.16	1E+09	660	6.023E+23	472	380	19.06 9.27414E+11
#1 Adapter_Biocathode_nirK	39.42	1E+09	660	6.023E+23	472	380	12.17 9.27414E+11
#1 Adapter_Same_bpha	45.34	1E+09	660	6.023E+23	542	380	12.15 9.27414E+11
							6600 185.0411926

도면12b

Description	Concentration 1g to ng	1 mol bp DNA to g DNA	mol bp to bp size (bp)	Added mass (ng)	volume	gene copies		
#2 Barcode_Same_16S	121.86	1E+09	660	6.023E+23	492	60C	4.92	1.1129E+12
#2 Barcode_MSU_16S	100.33	1E+09	660	6.023E+23	492	60C	5.98	1.1129E+12
#2 Barcode_River_16S	101.23	1E+09	660	6.023E+23	492	60C	5.93	1.1129E+12
#2 Barcode_Tidal_16S	99.83	1E+09	660	6.023E+23	492	60C	6.01	1.1129E+12
#2 Barcode_Biocathode_16S	126.94	1E+09	660	6.023E+23	492	60C	4.73	1.1129E+12
#2 Barcode_Same_nifH	48.38	1E+09	660	6.023E+23	360	40C	9.07	1.1129E+12
#2 Barcode_MSU_nifH	12.05	1E+09	660	6.023E+23	360	40C	36.43	1.1129E+12
#2 Barcode_River_nifH	15.72	1E+09	660	6.023E+23	360	40C	27.93	1.1129E+12
#2 Barcode_Tidal_nifH	44.5	1E+09	660	6.023E+23	360	40C	9.87	1.1129E+12
#2 Barcode_Biocathode_nifH	32.17	1E+09	660	6.023E+23	360	40C	13.65	1.1129E+12
#2 Barcode_Same_nirK	34.67	1E+09	660	6.023E+23	472	40C	16.60	1.1129E+12
#2 Barcode_MSU_nirK	29.35	1E+09	660	6.023E+23	472	40C	19.61	1.1129E+12
#2 Barcode_River_nirK	46.09	1E+09	660	6.023E+23	472	40C	12.49	1.1129E+12
#2 Barcode_Tidal_nirK	33.74	1E+09	660	6.023E+23	472	40C	17.06	1.1129E+12
#2 Barcode_Biocathode_nirK	51.25	1E+09	660	6.023E+23	472	40C	11.23	1.1129E+12
#2 Barcode_Same_bphA	65.63	1E+09	660	6.023E+23	542	40C	10.07	1.1129E+12
							7400	211.5825348

도면12c

Description	Concentration 1g to ng	1 mol bp DNA to g DNA	mol bp to bp size (bp)	Added mass (ng)	volume	gene copies		
#3 Adapter_Ratio1_16S	92.75	1E+09	660	6.023E+23	492	100C	10.78	1.85483E+12
#3 Adapter_Ratio2_16S	93.28	1E+09	660	6.023E+23	492	100C	10.72	1.85483E+12
#3 Adapter_Ratio3_16S	77.14	1E+09	660	6.023E+23	492	100C	12.96	1.85483E+12
#3 Adapter_Ratio4_16S	110.81	1E+09	660	6.023E+23	492	100C	9.02	1.85483E+12
#3 Adapter_Ratio5_16S	120.14	1E+09	660	6.023E+23	492	100C	8.32	1.85483E+12
#3 Adapter_Ratio1_nifH	74.66	1E+09	660	6.023E+23	360	40C	9.80	1.85483E+12
#3 Adapter_Ratio2_nifH	92.69	1E+09	660	6.023E+23	360	40C	7.89	1.85483E+12
#3 Adapter_Ratio3_nifH	75.00	1E+09	660	6.023E+23	360	40C	9.76	1.85483E+12
#3 Adapter_Ratio4_nifH	72.56	1E+09	660	6.023E+23	360	40C	10.08	1.85483E+12
#3 Adapter_Ratio5_nifH	64.50	1E+09	660	6.023E+23	360	40C	11.35	1.85483E+12
#3 Adapter_Ratio1_nirK	56.36	1E+09	660	6.023E+23	472	40C	17.02	1.85483E+12
#3 Adapter_Ratio2_nirK	68.52	1E+09	660	6.023E+23	472	40C	14.00	1.85483E+12
#3 Adapter_Ratio3_nirK	51.79	1E+09	660	6.023E+23	472	40C	18.53	1.85483E+12
#3 Adapter_Ratio4_nirK	54.41	1E+09	660	6.023E+23	472	40C	17.63	1.85483E+12
#3 Adapter_Ratio5_nirK	48.04	1E+09	660	6.023E+23	472	40C	19.97	1.85483E+12
#3 Adapter_Ratio1_bphA	66.84	1E+09	660	6.023E+23	542	40C	16.48	1.85483E+12
#3 Adapter_Ratio2_bphA	57.11	1E+09	660	6.023E+23	542	40C	19.29	1.85483E+12
#3 Adapter_Ratio3_bphA	62.13	1E+09	660	6.023E+23	542	40C	17.73	1.85483E+12
#3 Adapter_Ratio4_bphA	77.53	1E+09	660	6.023E+23	542	40C	14.21	1.85483E+12
#3 Adapter_Ratio5_bphA	46.23	1E+09	660	6.023E+23	542	40C	23.83	1.85483E+12
							11000	279.3968131

도면12d

Description	Concentration 1g to ng	1 mol bp DNA to g DNA	mol bp to bp size (bp)	Added mass (ng)	volume	gene copies		
#4 Barcode_Ratio1_16S	119.725	1E+09	660	6.023E+23	492	100C	8.35	1.85483E+12
#4 Barcode_Ratio2_16S	109.01	1E+09	660	6.023E+23	492	100C	9.17	1.85483E+12
#4 Barcode_Ratio3_16S	129.315	1E+09	660	6.023E+23	492	100C	7.73	1.85483E+12
#4 Barcode_Ratio4_16S	97.525	1E+09	660	6.023E+23	492	100C	10.25	1.85483E+12
#4 Barcode_Ratio5_16S	117.92	1E+09	660	6.023E+23	492	100C	8.48	1.85483E+12
#4 Barcode_Ratio1_nifH	48.095	1E+09	660	6.023E+23	360	40C	15.21	1.85483E+12
#4 Barcode_Ratio2_nifH	54.19	1E+09	660	6.023E+23	360	40C	13.50	1.85483E+12
#4 Barcode_Ratio3_nifH	32.93	1E+09	660	6.023E+23	360	40C	22.22	1.85483E+12
#4 Barcode_Ratio4_nifH	41.555	1E+09	660	6.023E+23	360	40C	17.61	1.85483E+12
#4 Barcode_Ratio5_nifH	33.8	1E+09	660	6.023E+23	360	40C	21.65	1.85483E+12
#4 Barcode_Ratio1_nirK	44.26	1E+09	660	6.023E+23	472	40C	21.68	1.85483E+12
#4 Barcode_Ratio2_nirK	44.835	1E+09	660	6.023E+23	472	40C	21.40	1.85483E+12
#4 Barcode_Ratio3_nirK	49.45	1E+09	660	6.023E+23	472	40C	19.40	1.85483E+12
#4 Barcode_Ratio4_nirK	31.025	1E+09	660	6.023E+23	472	40C	30.92	1.85483E+12
#4 Barcode_Ratio5_nirK	67.68	1E+09	660	6.023E+23	472	40C	14.17	1.85483E+12
#4 Barcode_Ratio1_bphA	35.255	1E+09	660	6.023E+23	542	40C	31.25	1.85483E+12
#4 Barcode_Ratio2_bphA	71.75	1E+09	660	6.023E+23	542	40C	15.35	1.85483E+12
#4 Barcode_Ratio3_bphA	43.39	1E+09	660	6.023E+23	542	40C	25.39	1.85483E+12
#4 Barcode_Ratio4_bphA	17.35	1E+09	660	6.023E+23	542	40C	63.48	1.85483E+12
#4 Barcode_Ratio5_bphA	43.14	1E+09	660	6.023E+23	542	40C	25.54	1.85483E+12
							11000	402.7759248

도면12e

Description	Concentration	1g to ng	1 mol bp DNA to g DNA	mol bp to bp size (bp)	Added mass (ng)	volume	gene copies
#5 G4_16S	131.46	1E+09	660	6.023E+23	492	65C	4.94 1.20564E+12
#5 RHA1_16S	104.93	1E+09	660	6.023E+23	492	65C	6.19 1.20564E+12
#5 B1_16S	112.24	1E+09	660	6.023E+23	492	65C	5.79 1.20564E+12
#5 Sataphyl_16S	141.92	1E+09	660	6.023E+23	492	65C	4.58 1.20564E+12
#5 Roseobac_16S	103.57	1E+09	660	6.023E+23	492	65C	6.28 1.20564E+12
#5 DCB_nifH	157.745	1E+09	660	6.023E+23	360	40C	3.02 1.20564E+12
#5 Rhodo_nifH	119.44	1E+09	660	6.023E+23	360	40C	3.98 1.20564E+12
#5 CJ2_nifH	115.32	1E+09	660	6.023E+23	360	40C	4.12 1.20564E+12
#5 G4_nifH	89.615	1E+09	660	6.023E+23	360	40C	5.31 1.20564E+12
#5 PCC_nifH	115.915	1E+09	660	6.023E+23	360	40C	4.10 1.20564E+12
#5 Ochro_nirK	73.955	1E+09	660	6.023E+23	472	40C	8.43 1.20564E+12
#5 CJ2_nirK	42.22	1E+09	660	6.023E+23	472	40C	14.77 1.20564E+12
#5 F1_bphA	19.075	1E+09	660	6.023E+23	542	40C	37.54 1.20564E+12
#5 RHA1_bphA	67.665	1E+09	660	6.023E+23	542	40C	10.58 1.20564E+12
#5 CJ2_bphA	53.325	1E+09	660	6.023E+23	542	40C	13.43 1.20564E+12
							7250 133.0690225

도면12f

Description	Concentration	1g to ng	1 mol bp DNA to g DNA	mol bp to bp size (bp)	Added mass (ng)	volume	gene copies
#6 MSU R1_16S	54.46	1E+09	660	6.023E+23	492	55C	10.10 1.02016E+12
#6 MSU R2_16S	80.955	1E+09	660	6.023E+23	492	55C	6.79 1.02016E+12
#6 MSU R3_16S	96.72	1E+09	660	6.023E+23	492	55C	5.69 1.02016E+12
#6 Bio R1_16S	70.63	1E+09	660	6.023E+23	492	55C	7.79 1.02016E+12
#6 Bio R2_16S	95.76	1E+09	660	6.023E+23	492	55C	5.74 1.02016E+12
#6 Bio R3_16S	93.1	1E+09	660	6.023E+23	492	55C	5.91 1.02016E+12
#6 Same_jsonifH	48.53	1E+09	660	6.023E+23	360	40C	8.29 1.02016E+12
#6 MSU_jsonifH	31.645	1E+09	660	6.023E+23	360	40C	12.72 1.02016E+12
#6 River_jsonifH	83.535	1E+09	660	6.023E+23	360	40C	4.82 1.02016E+12
#6 Tidal_jsonifH	29.785	1E+09	660	6.023E+23	360	40C	13.51 1.02016E+12
#6 Bio_jsonifH	65.09	1E+09	660	6.023E+23	360	40C	6.18 1.02016E+12
#6 Same_isobphA	117.32	1E+09	660	6.023E+23	542	40C	5.16 1.02016E+12
#6 NL1	18.995	1E+09	660	6.023E+23	492	40C	28.95 1.02016E+12
#6 NL2	18.33	1E+09	660	6.023E+23	492	40C	30.01 1.02016E+12
#6 NL3	14.7	1E+09	660	6.023E+23	492	40C	37.41 1.02016E+12
#6 NL4	19.245	1E+09	660	6.023E+23	492	40C	28.58 1.02016E+12
#6 NL5	14.705	1E+09	660	6.023E+23	492	40C	37.40 1.02016E+12
#6 NL6	15.95	1E+09	660	6.023E+23	492	40C	34.48 1.02016E+12
							8100 289.5432742

$$copies = (DNA\ concentration [ng/\mu l]) \left(\frac{1g}{1000^3 ng} \right) \left(\frac{1\ mol\ bp\ DNA}{660\ g\ DNA} \right) \left(\frac{6.023 \times 10^{23}\ bp}{mol\ bp} \right) \left(\frac{1\ copy}{\geq\ name\ or\ plasmid\ size\ [bp]} \right) (volume\ of\ template\ [\mu l])$$

서열 목록

- <110> Yonsei University
- <120> Mock community for measuring pyrosequencing accuracy and a method of measuring pyrosequencing accuracy using the same
- <130> p4885
- <140> 10-2011-0015756
- <141> 2011-02-22
- <160> 20
- <170> Kopatent In 2.0
- <210> 1
- <211> 1477
- <212> DNA
- <213> Rhodospirillum rubrum ATCC 11170
- <400> 1

agagtttgat cctggctcag gacgaacgct ggcggcaggc ctaacacatg caagtcgaac 60
 gcatccttcg ggatgagtgg cgcacgggtg agtaacacgt gggaacgtac cttggagtgc 120

ggaataatct ttggaaacga ggactaatac gcatacgc cttaggggga aagatttatc 180
 gctccaagat cggcccgcgt ccgattagct agttggcggg gtaatggccc accaaggcga 240
 cgatcggtag ctggtctgag aggatggcca gccacactgg gactgagaca cggcccagac 300
 tcctacggga ggcagcagtg gggaaatattg cgcaatgggg gcaacctga cgcagccatg 360
 ccgcgtgagt gaagaaggcc ttcgggttgt aaagctcttt cgggttgtaa gatgatgacg 420
 gtaacaccag aagaagcccc ggctaacttc gtgccagcag ccgcggtaat acgaaggggg 480
 caagcgttgt tcggaattac tgggcgtaaa gagcgcgtag gcggtctgat tagtcagagg 540

 tgaaatccca gagctcaact ttggaactgc ctttgatact gttagactag aatccgtgag 600
 aggggtggtg aattcccagt gtagaggta aattcgtaga tattgggagg aacaccagtg 660
 gcgaaggcgg ccacctggcg cggtattgac gctgaggcgc gaaagcgtgg ggagcaaaca 720
 ggattagata ccttgtagt ccacgccga aacgatgagt gctagatgac ggggtacatg 780
 tacctcggtg tcgcagctaa cgcattaagc actccgctg gggagtacgg ccgcaaggtt 840
 aaaactcaaa ggaattgacg ggggcccgca caagcgggtg agcatgtggt ttaattcgaa 900
 gcaacgcgca gaaccttacc agcccttgac atcccgtgac acttccagag atggaaggtt 960

 cccttcgggg acacggtgac aggtgctgca tggctgtcgt cagctcgtgt cgtgagatgt 1020
 tgggttaagt cccgcaacga gcgcaacct catcttcagt tgccagcaag taacgttggg 1080
 cactctgaag agactgccg tgacaagccg gaggaaggtg gggatgacgt caagtcctca 1140
 tggcccttac gggctgggct acacacgtgc tacaatggcg cctacaatgg gcagcgacct 1200
 cgcgagggga agctaacttc caaaaggcgt ctcagttcgg attgcactct gcaactcggg 1260
 tgcataaagt cggaaatcgt agtaatcgtg gatcagcatg ccacggtgaa tacgttcccg 1320
 ggcttctgac acaccgccc tcacaccatg ggagttggtt ctaccggaag acggtacgct 1380

 aaccgcaagg aggcagccgg ccacggtagg gtcagcgact ggggtgaagt cgtaacaagg 1440
 tagccgtagg ggaacctgcg gctggatcac ctctttt 1477

 <210> 2
 <211> 1526
 <212> DNA
 <213> Burkholderia vietnamiensis G4
 <400> 2
 agagtttgat cctggetcag attgaacgct ggcggcatgc cttacacatg caagtcgaac 60
 ggcagcacgg gtgcttgac ctggtggcga gtggcgaac ggtgagtaat acatcggaa 120
 atgtcctgta gtgggggata gcccgcgaa agccggatta ataccgata cgatctatgg 180

atgaaagcgg gggaccttcg ggcctcgcgc tatagggttg gccgatggct gattagctag	240
ttggtggggg aaaggcctac caaggcgacg atcagtagct ggtctgagag gacgaccagc	300
cacactggga ctgagacacg gccagactc ctacgggagg cagcagtggg gaattttgga	360
caatgggcga aagcctgac cagcaatgcc gcgtgtgtga agaaggcctt cgggttghtaa	420
agcacttttg tccggaaga aatccttggc tctaatacag tcgggggatg acggtaccgg	480
aagaataagc accggctaac tacgtgccag cagcccggt aatacgtagg gtgcaagcgt	540
taatcggaat tactgggct aaagcgtgcg caggcggttt gctaagaccg atgtgaaatc	600
cccgggctca acctgggaac tgcattggtg actggcaggc tagagtatgg cagagggggg	660
tagaattcca cgtgtagcag tgaatgcgt agagatgtgg aggaataccg atggcgaagg	720
cagccccctg ggccaatact gacgctcatg cacgaaagcg tggggagcaa acaggattag	780
ataccctggt agtccacgcc ctaaacgatg tcaactagt gttggggatt catttctta	840
gtaacgtagc taacgcgtga agttgaccgc ctggggagta cgttcgcaag attaaaactc	900
aaaggaattg acggggacc gcacaagcgg tggatgatgt ggattaattc gatgcaacgc	960
gaaaaactt acctaccctt gacatggtcg gaatcctgaa gagattcggg agtgctcga	1020
agagaaccgg cgcacaggtg ctgcatggct gtcgtcagct cgtgtcgtga gatgtgggt	1080
taagtccgc aacgagcga acccttgtcc ttagttgcta cgcaagagca ctctaaggag	1140
actgccggtg acaaaccgga ggaaggtggg gatgacgtca agtcctcatg gcccttatgg	1200
gtagggcttc acacgcata caatggtcgg aacagagggt tgccaaccg cgagggggag	1260
ctaataccag aaaaccgatc gtagtcgga ttgcaactctg caactcgagt gcatgaagct	1320
ggaatceta gtaatcgcg atcagatgc cgcggtgaat acgttcccg gtcttgtaca	1380
caccgccgt cacacatgg gagtgggttt taccagaagt ggctagtcta accgcaagga	1440
ggacggtcac cacgtagga ttcatgactg ggggaagtc gtaacaaggt agccgtatcg	1500
gaaggtgcgg ctggatcacc tccttt	1526
<210> 3	
<211> 1490	
<212> DNA	
<213> Burkholderia xenovorans LB400	
<400> 3	
cttacacatg caagtcaac ggcagcacgg gggcaaccct ggtggcgagt ggccaacggg	60
tgagtaatac atcggaactg gtctcttagt ggggatagc cggcgaaag ccgattaat	120
accgcatacg ctctcggag gaaagcggg gatcttcgga cctcgcgcta cagggcggc	180

cgatggcaga ttagctgggt ggtggggtaa aggcctacca aggcgacgat ctgtagctgg 240
tctgagagga cgaccagcca cactgggact gagacacggc ccagactcct acgggaggca 300

gcagtgggga attttggaca atgggcgcaa gcctgatcca gcaatgccgc gtgtgtgaag 360
aaggccttcg gtttgaag cacttttctc cggaagaaa acctctgcc taatacgggtg 420
gggggatgac ggtaccgaa gaataagcac cggctaacta cgtgccagca gccgcggtaa 480
tacgtagggt gcaagcgtta atcgaatta ctgggcgtaa agcgtgcgca ggcggttcgc 540
taagacagat gtgaaatccc cgggcttaac ctgggaactg cttttgtgac tggcgggcta 600
gagtatggca gaggggggtg gaattccacg tgtagcagtg aaatgcgtag agatgtggag 660
gaataccgat ggcaaggca gccccctggg ccaatactga cgctcatgca cgaaagcgtg 720

gggagcaaac aggattagat accctggtag tccacgccct aaacgatgc aactagtgt 780
cgggtcttca ttgacttggg aacttagcta acgcgtgaag ttgaccgct ggggagtacg 840
gtcgaagat taaaactcaa aggaattgac ggggaccgc acaagcggtg gatgatgtgg 900
attaattcga tgcaacgca aaaaccttac ctaccctga catgtatgga agtctgctga 960
gaggtggatg tgcccgaag ggagccataa cacagtgct gcatggctgt cgtcagctcg 1020
tgtcgtgaga tgttgggtta agtcccga caagcgcaac ccttgtccct agttgctacg 1080
caagagcact ccaggagac tgccggtgac aaaccggagg aaggtgggga tgacgtcaag 1140

tectcatggc cttatgggt agggcttcac acgtcataca atggtcggaa cagaggtcgc 1200
ccaacccgcg agggggagcc aatcccagaa aaccgatcgt agtccggtc gactctgca 1260
actcgggtgc gtgaagctgg aatcgctagt aatcgcggat cagcatgccg cggtgaatac 1320
gttcccgggt ctgtacaca ccgccctca caccaatggga gtgggtttta ccagaagtgg 1380
ctagtctaac cgcaaggagg acggtcacca cggtaggatt catgactggg gtgaagtcgt 1440
aacaaggtag ccgtatcgga aggtgcggt ggatcacctc ctttcccag 1490

<210> 4
<211> 1671
<212> DNA

<213> Desulfitobacterium hafniense DCB-2
<400> 4

agagtttgat cctggctcag gacgaacgt ggccgctgc ctaacacatg caagtcaac 60
ggactaacgt ttagcactga gtgttcagtg aagttggta aagcgagagc tcgaacgtag 120
tgaagcgaag agagactcc aggtcttata aagtccaac acagtgtgct aaaggaaac 180
agtaagaaca ctgagtgcta gacgttagtt agtggcggac gggtagtaaa cgcgtggata 240

acctacctag tagaccggga caacccttgg aaacgagggc taataccgga ttagcttaat 300
tagtggcatc actgattaag gaaagatggc ctctgaaaat gctatcgta gtagatggat 360

ccgcgtctga ttagctagtt ggtggggtaa aggcctacca aggcgacgat cagtagccgg 420
cctgagaggg tgaacggcca cactgggact gagacacggc ccagactcct acgggaggca 480
gcagtgggga atcttccgca atggacgaaa gtctgacgga gcaacgccgc gtgtacgacg 540
aaggccttcg ggttgtaaag tactgtcttc agggacgaac ggtaaagtatg taaataatgt 600
acttacatga cggtaactga ggaggaagcc ccggctaact acgtgccagc agcccggtta 660
atacttaggg ggcaagcgtt gtccggaatc attgggcgta aagggcgcgt agggcgatac 720
ttaagtctgg tgtgaaaacc tagggctcaa ccctgggact gcatcggaaa ctgggtatct 780

tgaggacagg agaggaaagt ggaattccac gtgtagcggg gaaatgcgta gatatgtgga 840
ggaacaccag tggcgaagcg gactttctgg actgtaactg acgctgaggc gcgaaagcgt 900
ggggagcaaa caggattaga taccctggtg gtccacgccg taaacgatga gtgctagggtg 960
tagagggtat cgacccttc tgtgcccgag ttaacacact aagcactccg cctggggagt 1020
acggccgcaa ggttgaaact caaaggaatt gacgggggcc cgcacaagcg gtggagcatg 1080
tggtttaatt cgacgcaacg cgaagaacct taccaaggct tgacatccat agaatcctgt 1140
ggaaacatgg gagtgcctt cggggagcta tgagacaggt ggtgcatggt tgtcgtcagc 1200

tctgtctgtg agatgttggg ttaagtcccg caacgagcgc aaccctatg tttagttgct 1260
aacgcgtaat ggtgagcact ctagacagac tgccggtgac aaaccggagg aaggtgggga 1320
tgacgtcaaa tcatcatgcc cttatgtct tgggctacac acgtgctaca atggccagta 1380
cagacggaag cgaagccgtg aggtgaagcc aatccgagaa agctggcttc agttcggatt 1440
gttctctgca actcgagaac atgaagtggc aatcgctagt aatcgaggt cagcactatg 1500
cggatgaatac gttcccgggc cttgtacaca ccgccgtca caccacgaaa gtctgcaaca 1560
cccgaagccg tgagagtaac ccgaaaggga gctagccgtc gaaggtgggg ccgatgattg 1620

gggtgaagtc gtaacaaggt agccgtatcg gaaggtgagg ctggatcacc t 1671

<210> 5
<211> 1489
<212> DNA
<213> Nostoc sp. PCC 7120
<400> 5

aaaacggaga gtttgatcct ggctcaggat gaacgctggc ggtatgctta acacatgcaa 60
gtcgaacggt ctcttcggag atagtggcgg acgggtgagt aacgcgtgag aatctagctt 120

caggtcgggg acaaccactg gaaacggtgg ctaataccgg atgtgccgaa aggtgaaaga 180
 tttattgctt gaagatgagc tcgctctga ttagctagtt ggtgtggtaa gagcgacca 240
 aggcgacgat cagtagctgg tctgagagga tgatcagcca cactgggact gagacacggc 300

 ccagactcct acgggaggca gcagtgggga attttccgca atgggcgaaa gcctgacgga 360
 gcaataccgc gtgaggagg aaggctcttg ggttgtaaac ctcttttctc agggaataaa 420
 aaaatgaagg tacctgagga ataagcatcg gctaactccg tgccagcagc cgcgtaata 480
 cggaggatgc aagcgttacc cggaatgatt gggcgtaaag cgtccgcagg tggcactgta 540
 agtctgctgt taaagagcaa ggctcaacct tghtaaaggca gtggaaacta cagagctaga 600
 gtacgttcgg ggcagaggga attcctggtg tagcggtgaa atgcgtagag atcaggaaga 660
 acaccggtgg cgaaagcgtc ctgctaggcc gtaactgaca ctgagggacg aaagctaggg 720

 gagcgaatgg gattagatac cccagtagtc ctaccgtaa acgatggata ctaggcgtgg 780
 cttgtatcga cccgagccgt gccggagcca acgcgttaag tatccgcct ggggagtacg 840
 cacgcaagtg tgaactcaa aggaattgac gggggcccgc acaagcggtg gagtatgtgg 900
 tttaatcga tgcaacgca agaacctac caagacttga catgtcgcga atcttcttga 960
 aagggaagag tgccttaggg agcgcgaaca cagggtgtgc atggctgtcg tcagctcgtg 1020
 tcgtgagatg ttgggttaag tcccgaacg agcgcgaacc tcgttttag ttgccagcat 1080
 taagttgggc actctagaga gactgccgtt gacaaaccgg aggaaggtgg ggatgacgtc 1140

 aagtcagcat gcccttacg tcttgggcta cacacgtact acaatgctac ggacagaggg 1200
 cagcaagcta gcgatagcaa gcaaatccc taaaccgtag ctcaagttcag atcgcaggct 1260
 gcaactcgcc tgcgtgaagg aggaatcgct agtaattgca ggtcagcata ctgcagtgaa 1320
 ttcgttccc ggccttgtac acaccgccc tcaacatg gaagctggca acgcccgaag 1380
 tcattactcc aacttttagg agaggaggat gcctaaggca gtgctggtga ctggggtgaa 1440
 gtcgtaacaa ggtagecgta ccggaaggtg tggtggatc acctcctt 1489

 <210> 6
 <211> 1523
 <212> DNA

 <213> Polaromonas naphthalenivorans CJ2
 <400> 6
 agagtttgat cctggctcag attgaacgct ggcggcatgc cttacacatg caagtcgaac 60
 ggtaacgggt taagccgac agtggcgaac gggtagtaa tatatcgaa cgtgccagc 120
 cgtgggggat aacgtagaga aatttacgct aataccgat acgatctaag gatgaaagc 180

ggggaccttc gggcctcgcg cgattggagc ggctgatac agattaggtt gttggtgagg 240
 taaaagctca ccaagccgac gatctgtagc tggtttgaga gaacgaccag ccacactggg 300
 actgagacac ggcccagact cctacgggag gcagcagtgg ggaatthttgg acaatgggcg 360

 aaagcctgat ccagcaatgc cgcgtgcagg aagaaggcct tcgggttgta aactgctttt 420
 gtacggaacg aaaaggccag ccctaatacg gcaggcccat gacggtaccg taagaataag 480
 caccggctaa ctacgtgcca gcagccgagg taatacgtag ggtgcgagcg ttaatcggaa 540
 ttactgggcg taaagcgtgc gcagggcgtg atgtaagaca gttgtgaaat ccccgggctc 600
 aacctgggaa ttgcatctgt gactgcatcg cttagtagc gtagaggggg atggaattcc 660
 gcgtgtagca gtgaaatgcg tagatatgcg gaggaacacc gatggcgaag gcaatccct 720
 ggacctgtac tgacgctcat gcacgaaagc gtggggagca aacaggatta gataccctgg 780

 tagtccacgc cctaaacgat gtcaactggt tgttgggtgc attagtactc agtaacgaag 840
 ctaacgcgtg aagttgaccg cctggggagt acggccgcaa ggttgaaact caaaggaatt 900
 gacggggacc cgcacaagcg gtggatgatg tggtttaatt cgatgcaacg cgaaaaacct 960
 tacctacctt tgacatgtac ggaactcgcc agagatggct tggtgctcga aagagagccg 1020
 taacacaggt gctgcatggc tgtcgtcagc tcgtgtcgtg agatgthggg ttaagtcccg 1080
 caacgagcgc aaccctgtc attagthgt acatttagtt gggcactcta atgagactgc 1140
 cggtgacaaa ccggaggaag gtggggatga cgtcaagtcc tcatggccct tatagtagg 1200

 gctacacacg tcatacaatg gatggtacaa agggtcgcca acccgcgagg gggagccaat 1260
 cccataaagc cattcgtagt ccggatcgea gtctgcaact cgactgctg aagtcggaat 1320
 cgctagtaat cgtggatcag aatgtcacgg tgaatagtt cccgggtctt gtacacaccg 1380
 cccgtcacac catgggagcg gtttctgcca gaagtagtta gcctaactg caaaggaggg 1440
 cgattaccac ggcagggttc gtgactggg tgaagtcgta acaaggtagc cgtatcgaa 1500
 ggtgcggtg gatcacctcc ttt 1523

 <210> 7
 <211> 1519
 <212> DNA

 <213> Rhodococcus sp. RHA1
 <400> 7
 cttcaacgga gattttgatc ctggctcagg acgaacgctg gcggcgtgct taacacatgc 60
 aagtcgagcg gtaagccct tcgggttaca cgagcggcga acgggtgagt aacacgtggg 120
 tgatctgccc tgcacttcgg gataagcctg ggaaactggg tctaataccg gatatgacct 180

tcggtgcat ggctgagggt gaaaggttt actggtgcag gatgggcccg cggcctatca 240
gcttgttggt ggggtaatgg cctaccaagg cgacgacggg tagccgacct gagagggtga 300
ccggccacac tgggactgag acacggccca gactcctacg ggaggcagca gtggggaata 360

ttgcacaatg ggcgaaagcc tgatgcagcg acgccgctg agggatgacg gccttcgggt 420
tgtaaacctc tttcagcagg gacgaagcga aagtgacggt acctgcagaa gaagcaccgg 480
ctaactacgt gccagcagcc gcgtaatac gtagggtgca agcgttgtcc ggaattactg 540
ggcgtaaaga gttcgtaggc ggtttgtcgc gtcgtttgtg aaaactcaca gctcaactgt 600
gagcctgcag gcgatacggg cagacttgag tactgcaggg gagactggaa ttcttggtgt 660
agcggtgaaa tgcgcagata tcaggaggaa caccggtggc gaaggcgggt ctctgggcag 720
taactgacgc tgaggaacga aagcgtgggt agcaaacagg attagatacc ctggtagtcc 780

acgccgtaaa cgggtggcgc tagtgtggg ttccttccac gggatctgtg ccgtagctaa 840
cgcatlaagc gccccctg gggagtacgg ccgcaaggt aaaaactcaaa ggaattgacg 900
ggggcccga caagcggcgg agcatgtgga ttaattcgat gcaacgcgaa gaaccttacc 960
tgggtttgac atataccgga aagccgtaga gatacggccc cccttgtggt cggatacag 1020
gtggtgcatg gctgtcgtca gctcgtgicg tgagatgttg ggttaagtcc cgcaacgagc 1080
gcaacccttg tcttatgttg ccagcacgta atgggtggga ctcgtaagag actgccgggg 1140
tcaactcgga ggaaggtggg gacgacgta agtcatcatg ccccttatgt ccagggttc 1200

acacatgcta caatggccag tacagagggc tgcgagaccg tgagggtgag cgaatccctt 1260
aaagctggtc tcagttcgga tcgggtctg caactcgacc ccgtgaagtc ggagtcgcta 1320
gtaatcgag atcagcaacg ctgcggtgaa tacgttcccg ggccttgtac acaccgcccg 1380
tcacgtcatg aaagtcggtg acaccgaag ccggtggcct aacccttgt gggagggagc 1440
cgtcgaaggt gggatcggcg attgggacga agtcgtaaca aggtagccgt accggaaggt 1500
gcggtggat cacctcctt 1519

<210> 8
<211> 1549
<212> DNA

<213> Pseudomonas putida F1
<400> 8

gaactgaaga gtttgatcat ggctcagatt gaacgctggc ggcaggccta acacatgcaa 60
gtcagcggga tgacgggagc ttgtccttg attcagcggc ggacgggtga gtaatgccta 120
ggaatctgcc tggtagtggg ggacaacgtt tcgaaaggaa cgtaataacc gcatacgtcc 180

tacgggagaa agcagggggac cttcgggect tgcgctatta gatgagccta ggtcggatta 240
gctagttagt ggggtaatgg ctaccaagg cgacgatccg taactggtct gagaggatga 300
tcagtcacac tggaaactgag acacggtcca gactcctacg ggaggcagca gtggggaata 360

ttggacaatg ggcgaaagcc tgatccagcc atgcccgctg tgtgaagaag gtcttcggat 420
tgtaaagcac ttttaagtgg gaggaagggc agtaagctaa taccttgctg ttttgacgtt 480
accgacagaa taagcaccgg ctaactctgt gccagcagcc gcgtaatac agagggtgca 540
agcgttaatc ggaattactg ggcgtaaagc gcgctaggt ggtttgttaa gttggatgtg 600
aaagccccgg gctcaacctg ggaactgcat ccaaaactgg caagctagag tacggtagag 660
ggtggtggaa tttcctgtgt agcgggtgaaa tgcgtagata taggaaggaa caccagtggc 720
gaaggcgacc acctggactg atactgacac tgagggtcga aagcgtgggg agcaaacagg 780

attagatacc ctggtagtcc acgccgtaaa cgatgtcaac tagccgttgg aatccttgag 840
atthtagtgg cgcagetaac gcattaagtt gaccgcctgg ggagtacggc cgcaaggtta 900
aaactcaaat gaattgacgg gggccccgac aagcgggtga gcatgtggtt taattcgaag 960
caacgcgaag aaccttacca ggccttgaca tgcagagaac tttccagaga tggattggtg 1020
ccttcgggaa ctctgacaca ggtgctgcat ggctgtcgtc agctcgtcgtc gtgagatgtt 1080
gggttaagtc ccgtaacgag cgcaaccctt gtccttagtt accagcacgt tatggtgggc 1140
actctaagga gactgccggt gacaaccgg aggaaggtgg ggatgacgtc aagtcatcat 1200

ggcccttacg gcctgggcta cacacgtgct acaatggtcg gtacagaggg ttgccaagcc 1260
gcgaggtgga gctaactca caaaaccgat cgtagtcgag atcgcagtct gcaactcgac 1320
tgcgtgaagt cggaaactgt agtaatcgcg aatcagaatg tgcgggtgaa tacgttcccg 1380
ggccttgiac acacccccc tcacaccatg ggagtggtt gcaccagaag tagctagtct 1440
aaccttcggg gggacggtta ccacggtgtg attcatgact ggggtgaagt cgtaacaagg 1500
tagccgtagg ggaacctgcg gctggatcac ctcttaate gacgacatc 1549

<210> 9
<211> 457
<212> DNA

<213> Neisseria sicca
<400> 9

catgcaagtc ggacggcagc acagagaagc ttgcttcttg ggtggcgagt ggccaacggg 60
tgagtaacat atcggaactg accgagcagt gggggataac taatcgaag attagctaat 120
accgcatatt ttctgaggaa gaaagcaggg gaccatttgg ccttgcgctg tttgagcggc 180

cgatatctga ttagctggtt ggtggggtaa aggcctacca aggcgacgat cagtagcggg 240
tctgagagga tgatcccca cactgggact gagacacggc ccagactcct acgggaggca 300
gcagtgggga attttgaca atgggcgcaa gcctgatcca gccatgccgc gtgtctgaag 360

aaggccttcg ggttgtaaag gacttttgtc agggaagaaa aagatagggt taataccct 420
gtctgatgac ggtacctgaa gaataagcac cggctaa 457

<210> 10
<211> 1476
<212> DNA
<213> Ochrobactrum anthropi ATCC 49188
<400> 10

agagtttgat cctggctcag aacgaacgt ggcggcaggc ttaacacatg caagtcgagc 60
gccccgcaag gggagcggca gacgggtgag taacgcgtgg gaacgtacct tttgctacgg 120
aataactcag ggaaacttgt gctaataacc tatgtgccct tcgggggaaa gatttatcgg 180
caaaggatcg gcccgcgttg gattagctag ttggtgaggt aaaggctcac caaggcgacg 240

atccatagct ggtctgagag gatgatcagc cacactggga ctgagacacg gcccagactc 300
ctacgggagg cagcagtggg gaatattgga caatgggcgc aagcctgac cagccatgcc 360
gcgtgagtga tgaagccct agggttgtaa agctctttca ccggtgaaga taatgacggt 420
aaccggagaa gaagccccg ctaacttctg gccagcagcc gcgtaatac gaagggggct 480
agcgttgctc ggatttactg ggcgtaaac gcacgtaggc ggacttttaa gtcaggggtg 540
aatccccgg gctcaacccc ggaactgctt ttgatactgg aagtcttgag tatggtagag 600
gtgagtggaa ttccgagtgt agaggtgaaa ttcgtagata ttcggaggaa caccagtggc 660

gaaggcggct cactggacca ttactgacgc tgaggcgcga aagcgtgggg agcaaacagg 720
attagatacc ctggtagtec acgccgtaaa cgatgaatgt tagccgttgg ggagtttact 780
cttcggtggc gcagctaac cattaaacat tccgcctggg gactacggtc gcaagattaa 840
aactcaaagg aattgacggg ggccccaca agcgtgggag catgtggttt aattcgaagc 900
aacgcgaga accttaccag cccttgacat accggtcgcg gacacagaga tgtgtctttc 960
agttcggctg gaccgatac aggtgctgca tggctgtcgt cagctcgtgt cgtgagatgt 1020
tgggttaagt cccgcaacga gcgcaacct cgcccttagt tgccagcatt tagttgggca 1080

ctctaagggg actgccgggtg ataagccgag aggaaggtgg ggatgacgtc aagtcctcat 1140
ggcccttac ggctgggcta cacactgct acaatggtgg tgacagtggg cagcgagcac 1200
gcgagtgtga getaatctcc aaaagccatc tcagttcgga ttgcactctg caactcgagt 1260

gcatgaagtt ggaatcgcta gtaatcgcg atcagcatgc cgcggtgaat acgttcccgg 1320
gccttgtaga caccgcccgt cacaccatgg gagttggttt tacccgaagg cgctgtgcta 1380
accgcaagga ggcagcgac cacggtaggg tcagcgactg ggggaagtc gtaacaaggt 1440
agccgtaggg gaacctcgcg ctggatcacc tccttt 1476

<210> 11

<211> 1474

<212> DNA

<213> Chromobacterium violaceum ATCC 12472

<400> 11

aactgaagag ttgatcctg gctcagattg aacgctggcg gcatgcitta cacatgcaag 60
tcgaacggta acagggtgct tgcaccgctg acgagtgccg aacgggtgag taatgcgtcg 120
gaatgtaccg tgtaatgggg gatagctcgg cgaaagccgg attaataccg catacgcct 180
gaggggaaa gcgggggac gaaagacctc gcgttatacg agcagccgac gtctgattag 240
ctagttgggt agtaagagc tcaccaagc gacgatcagt agcgggtctg agaggatgat 300
ccgccacact gggactgaga cacggcccag actcctacgg gaggcagcag tggggaattt 360

tggacaatgg gggcaacct gatccagcca tcccgcgtgt ctgaagaagg cttcgggtt 420
gtaaaggact ttgtcaggg aggaaatccc gctggttaat acccggcggg gatgacagta 480
cctgaagaat aagcaccggc taactacgtg ccagcagccg cgtaaatcag tagggtgca 540
gcgttaatcg gaattactgg gcgtaaagcg tgcgcaggcg gttgtgcaag tctgatgtga 600
aagccccggg cttaacctgg gaacggcatt ggagactgca cagctagagt gcgtcagagg 660
ggggtagaat tccacgtgta gcagtgaat gcgtagagat gtggaggaat accgatggcg 720
aagcagccc cctgggatga cactgacgct catgcacgaa agcgtgggga gcaaacagga 780

ttagataccc tggtagtcca cgcctaacc gatgtcaact agctgttggg ggtttgaatc 840
cttggtagcg tagctaaccg gtgaagtga cgcctgggg agtacggccg caaggttaa 900
actcaaagga attgacgggg acccgcaaa gcggtggatg atgtgatta attcagatca 960
acgcgaaaaa cttactcgc tcttgacatg tacggaactt gccagagatg gcttgggtcc 1020
cgaaaggag ccgtaacaca ggtgctgcat ggctgtcgtc agctcgtgtc gtgagatgtt 1080
gggttaagtc ccgcaacgag cgcaacctt gtcattagtt gccatcattc agttgggcac 1140
tctaagaga ctgccgtga caaacggag gaaggtgggg atgacgtcaa gtcctcatgg 1200

cccttatgag cagggttca cacgtcaca aatggtcggg acagagggtt gccaagccgc 1260
gaggtggagc taatctcaga aaaccgatcg tagtccggat cgcactctgc aactcgatg 1320

cgtgaagtcg gaatcgctag taatcgcaga tcagcatgct gcggtgaata cgttcccggg 1380
 tcttgtacac accgcccgtc acaccatggg agtgagtttc accagaagtg ggtaggctaa 1440
 ccgcaaggag gccgcttacc acggtgggat tcat 1474
 <210> 12
 <211> 1359
 <212> DNA
 <213> Pseudomonas pickettii PK01
 <400> 12
 ggctcagatt gaacgtggc ggcattgcctt acacatgcaa gtcgagcggc agcatgatct 60

 agcttgctag attgatggcg agtggcgaac gggtagtaaa tacatcggaa cgtgccctgt 120
 agtgggggat aactagtca aagattagct aataccgcat acgacctgag ggtgaaagt 180
 ggggaccgca aggcctcatg ctataggagc ggccgatgct tgattagcta gttggtgggg 240
 taaaggccca ccaaggcgac gatcagtagc tggctgaga ggacgatcag ccacactggg 300
 actgagacac ggcccagact cctacgggag gcagcagtgg ggaattttgg acaatgggcg 360
 aaagcctgat ccagcaatgc cgcgtgtgtg aagaaggcct tcgggttgta aagcactttt 420
 gtccgaaag aatggctct ggttaatacc tggggtcgat gacggtaccg gaagaataag 480

 gaccggctaa ctacgtgcca gcagccgagg taatacgtag ggtccaagcg ttaatcggaa 540
 ttactgggcg taaagcgtgc gcagggggtt gtgcaagacc gatgtgaaat ccccagctt 600
 aacttgggaa ttgcatggg gactgcacgg ctagagtgtg tcagaggggg gtagaattcc 660
 acgtgtagca gtgaaatgcg tagagatgtg gaggaatacc gatggcgaag gcacccccct 720
 gggataaac tgacgtcat gcacgaaagc gtggggagca aacaggatta gataccctgg 780
 tagtccacgc cctaaacgat gtcaactagt tgttggggat tcatttcctt agtaacgtag 840
 ctaacgcgtg aagttgaccg cctggggagt acggtcgcaa gattaaaact caaaggaatt 900

 gtcggggacc cgcacaagcg gtgatgatg tggattaatt cgatgcaac cgaaaaacct 960
 tacctacct tgacatgcca ctaacgaagc agagatgcat tagtgctcg aaagagaaag 1020
 tggacacagg tgctgatgg ctgtcgtcag ctctgtcgt gagatgttg gctaagtccc 1080
 gcaacgagcg caaccctgt ctctagttgc tacgaaaggg cactctagag agactgccgg 1140
 tgacaaccg gaggaaggtg gggatgacgt caagtcctca tggcccttat gggtagggct 1200
 tcacagtc tacaatgggt catacagagg gttgccaagc cgcgaggtgg agctaatccc 1260
 agaaaatgca tcgtagtccg gatcgtatgc tgcaactcga ctactgaag ctggaatcgc 1320

 tagtaatcgc ggatcagcat gcccggtga attcgttcc 1359

<210> 13

<211> 1430

<212> DNA

<213> Sphingomonas yanoikuyae B1

<400> 13

agagtttgat cctggctcag aacgaacgct ggcggcatgc ctaatacatg caagtccaac 60

gagatcttcg gatctagtgg cgcacgggtg cgtaacgcgt gggaatctgc ccttgggttc 120

ggaataactt ctggaaacgg aagctaatac cggatgatga cgtaagcca aagatttacc 180

gccaaggat gagcccgcgt aggattagct agttggtggg gtaaaggccc accaaggcga 240

cgatccttag ctggtctgag aggatgatca gccacactgg gactgagaca cggcccagac 300

tcctacggga ggcagcagta gggaatatg gacaatgggc gaaagcctga tccagcaatg 360

ccgcgtgagt gatgaaggcc ttagggttgt aaagctcttt taccgggat gataatgaca 420

gtaccgggag aataagctcc ggctaactcc gtgccagcag ccgcggtaat acggagggag 480

ctagcgttgt tcggaattac tggcgttaa gcgcacgtag gcggctattc aagtcagagg 540

tgaaagcccg gggctcaacc ccggaactgc ctttgaact agatagcttg aatccaggag 600

aggigagtgg aattccgagt gtagaggta aattcgtaga tattcggaag aacaccagtg 660

gcgaaggcgg ctcaactggac tggattgac gctgaggtgc gaaagcgtgg ggagcaaaca 720

ggattagata ccctgtagt ccacgccgta aacgatgata actagctgtc agggcacatg 780

gtgttttggg ggcgcagcta acgcattaag ttatccgcct ggggagtacg gtcgcaagat 840

taaaactcaa aggaattgac gggggcctgc acaagcgtg gagcatgtgg ttttaattcga 900

agcaacgcgc agaaccttac caacgttga catccctatc gcggatcgtg gagacacttt 960

ccttcagttc ggctggatag gtgacaggtg ctgcatggct gtcgtcagct cgtgtcgtga 1020

gatgttgggt taagtcccgc aacgagcga accctgcct ttagttgcca gcatttagtt 1080

gggiactcta aaggaaccgc cggtgataag ccggaggaag gtggggatga cgtcaagtcc 1140

tcatggccct tacgcgttgg gctacacacg tgctacaatg gcgactacag tgggcagcca 1200

cctcgcgaga gggagctaat ctccaaaagt cgtctcagtt cggatcgttc tctgcaactc 1260

gagagcgtga aggcggaatc gctagtaatc gcggatcagc atcccgcgtg gaatacgttc 1320

ccaggccttg tacacaccgc ccgtcacacc atgggagttg gattcactcg aaggcgttga 1380

gctaaccgta aggaggcagg cgaccacagt gggtttagcg actggggtga 1430

<210> 14

<211> 1542

<212> DNA

<213> Escherichia coli str. K12 substr. W3110

<400> 14

taaggagggtg atccaaccgc aggttccctt acggttacct tgttacgact tcaccccagt 60

catgaatcac aaagtggtaa ggcctctccc gaaggttaag ctacctactt cttttgcaac 120

ccactcccat ggtgtgacgg gcggtgtgta caaggcccgg gaacgtattc accgtggcat 180

tctgatccac gattactagc gattccgact tcatggagtc gagttgcaga ctccaatccg 240

gactacgacg cactttatga ggtccgcttg ctctcgcgag gtcgcttctc tttgtatgcg 300

ccattgtagc acgtgtgtag ccctggctgt aaggccatg atgacttgac gtcatcccca 360

ccttctcca gtttatcact ggcagtctcc tttgagttcc cggccggacc gctggcaaca 420

aaggataagg gttgcgctcg ttgcgggact taaccaaca tttcacaaca cgagctgacg 480

acagccatgc agcacctgtc tcacggttcc cgaaggcaca ttctcatctc tgaaaacttc 540

cgtggatgtc aagaccaggt aaggttcttc gcggttgcatc gaattaaacc acatgctcca 600

ccgcttgtgc gggccccctg caattcattt gagttttaac ctgcccggcc tactccccag 660

gcggtcgcact taacgcgtta gtcccggaag ccacgcctca agggcacaac ctccaagtgc 720

acatcgttta cggcgtggac taccagggta tctaactctg tttgctcccc acgctttcgc 780

acctgagcgt cagtcttctg ccagggggcc gccttcgcca ccggtattcc tccagatctc 840

tacgcatttc accgctacac ctggaattct acccccctct acgagactca agcttgccag 900

tatcagatgc agttcccagg ttgagcccgg ggatttcaca tctgacttaa caaacgcct 960

gcgtgcgctt tacgcccagt aattccgatt aacgcttgca cctccglat taccgcgct 1020

gctggcacgg agttagccgg tgccttctct cggggtaacg tcaatgagca aaggtattaa 1080

ctttactccc ttctccccg ctgaaagtac tttacaacc gaaggccttc ttcatacacg 1140

cggcatggct gcatcaggct tgcgcccatt gtgcaatatt ccccaactgct gcctcccgta 1200

ggagtctgga ccgtgtctca gttccagtggt ggctgggtcat cctctcagac cagctaggga 1260

tcgtcgccta ggtgagccgt taccccact actagctaat cccatctggg cacatccgat 1320

ggcaagaggc ccgaaggtag cctcttttgg tcttgcgacg ttatgaggta ttagctaccg 1380

tttccagtag ttatccccct ccatcaggca gtttcccaga cattactcac ccgtccgcca 1440

ctcgtcagca aagaagcaag ctcttctctg ttaccgttcg acttgcatgt gttaggcctg 1500

ccgccagcgt tcaatctgag ccatgatcaa actcttcaat tt 1542

<210> 15

<211> 1556

<212> DNA

<213> Bacillus cereus ATCC 14579

<400> 15

```

ctttattgga gagtttgatc ctggctcagg atgaacgctg gcggcgtgcc taatacatgc      60
aagtcgagcg aatggattaa gagcttgctc ttatgaagtt agcggcggac gggtgagtaa      120

cacgtgggta acctgcccat aagactggga taactccggg aaaccggggc taataccgga      180
taacattttg aaccgcatgg ttcgaaatg aaaggcggct tcggctgtca cttatggatg      240
gaccgcgctc gcattagcta gttggtgagg taacggctca ccaaggcaac gatgcgtagc      300
cgacctgaga gggatgatcg ccacactggg actgagacac ggcccagact cctacgggag      360
gcagcagtag ggaatcttcc gcaatggacg aaagtctgac ggagcaacgc cgcgtgagtg      420
atgaaggctt tcgggtcgta aaactctggt gttaggggaag aacaagtgct agttgaataa      480
gctggcacct tgacgttacc taaccagaaa gccacggcta actacgtgcc agcagccgcg      540

gtaatacgtg ggtggcaagc gttatccgga attattgggc gtaaagcgcg cgcaggtggt      600
ttcttaagtc tgatgtgaaa gcccacggct caaccgtgga gggtcattgg aaactgggag      660
acttgagtgc agaagaggaa agtggaaatc catgtgtagc ggtgaaatgc gtagagatat      720
ggaggaacac cagtggcgaa ggcgactttc tggctctgta ctgacactga ggcgcgaaag      780
cgtgggggagc aaacaggatt agataccctg gtagtccacg ccgtaaacga tgagtgctaa      840
gtgttagagg gtttccgccc tttagtctg aagttaacgc attaagcact ccgcctgggg      900
agtacggccg caaggtgaa actcaaagga attgacgggg gcccgcaaa gcggtggagc      960

atgtggttta attcgaagca acgcaagaa ccttaccagg tcttgacatc ctctgaaaac      1020
cctagagata gggcttctcc ttcgggagca gagtgacagg tggatgatgg ttgtcgtcag      1080
ctcgtgtcgt gagatgttgg gtttaagtccc gcaacgagcg caaccctga tcttagttgc      1140
catcattaag ttgggcactc taaggtgact gccggtgaca aaccggagga aggtggggat      1200
gacgtcaaat catcatgccc cttatgacct gggctacaca cgtgctacaa tggacggtac      1260
aaagagctgc aagaccgca ggtggagcta atctcataaa accgttctca gttcggattg      1320
taggctgcaa ctgcctaca tgaagctgga atcgctagta atcgcgatc agcatgccgc      1380

ggtgaatacg ttccccggcc ttgtacacac cgcccgtcac accacgagag tttgtaacac      1440
ccgaagtcgg tgggtaacc tttttgagc cagccgcta aggtgggaca gatgattggg      1500
gtgaagtcgt aacaaggtag ccgtatcgga aggtgcggtt ggatcacctc ctttct      1556

```

<210> 16

<211> 1538

<212> DNA

<213> Corynebacterium glutamicum ATCC 13032

<400> 16

ttttttgtgg agagtttgat cctggctcag gacgaacgct ggcggcgtgc ttaacacatg 60
caagtgaac gctgaaaccg gagcttgctt tggatgatga gtggcgaacg ggtgagtaac 120
acgtgggtga tctgccttac actttgggat aagcctggga aactgggtct aataccgaat 180

attcacacca ccgtaggggt ggtgtggaaa gccttgtgcg gtgtgggatg agcctgcggc 240
ctatcagctt gttggtgggg taatggccta ccaaggcgtc gacgggtagc cggcctgaga 300
gggtgtacgg ccacattggg actgagacac ggcccagact cctacgggag gcagcagtgg 360
ggaatattgc acaatgggcg caagcctgat gcagcgacgc cgcgtggggg atgaaggcct 420
tcgggttga aactccttc gctagggacg aagcctttta ggtgacgta cctggagaag 480
aagcaccggc taactacgtg ccagcagccg cggtaatagc tagggtgcga gcgttgcg 540
gaattactgg gcgtaaagag ctctaggtg gtttgtcgcg tcgtctgtga aatcccgggg 600

cttaacttcg ggcgtgcagg cgatacgggc ataactttag tgctgtaggg gagactggaa 660
ttcctggtgt agcggtgaaa tgcgcagata tcaggaggaa caccaatggc gaaggcaggt 720
ctctgggcag taactgacgc tgaggagcga aagcatgggt agcgaacagg attagatacc 780
ctggtagtcc atgccgtaaa cggtagggcgc taggtgtagg ggtcttcac gacttctgtg 840
ccgcagctaa cgcattaage gcccgcctg gggagtacgg ccgcaaggct aaaactcaa 900
ggaattgacg ggggcccga caagcggcgg agcatgtgga ttaattgat gcaacgcgaa 960
gaaccttacc tgggcttgac atggaccgga tcggcgtaga gatacgtttt cccttgtggt 1020

cggttcacag gtggtgcatg gttgtcgtca gctcgtgctg tgagatgttg ggttaagtcc 1080
cgcaacgagc gcaacccttg tcttatgttg ccagcacatt gtggtgggta ctcatgagag 1140
actgccgggg ttaactcgga ggaaggtggg gatgacgtca aatcatcatg ccccttatgt 1200
ccagggttc acacatgcta caatggtcgg tacagcgagt tgccacaccg tgaggtggag 1260
ctaactcttt aaagccggcc tcagttcgga ttggggtctg caactcgacc ccatgaagtc 1320
ggagtcgcta gtaatcgag atcagcaacg ctgcggtgaa tacgttcccg ggccttgtac 1380
acaccgccg tcacgtcatg aaagtggta acaccgaag ccagtggccc aaccttttag 1440

gggggagctg tcgaaggtgg gatcggcgat tgggacgaag tcgtaacaag gtagccgtac 1500
cggaaggtgc ggctggatca cctcctttct aaggagct 1538

<210> 17

<211> 1554

<212> DNA

<213> Staphylococcus epidermidis ATCC 12228

<400> 17

ttttatggag agtttgatcc tggctcagga tgaacgctgg cggcgtgcct aatacatgca 60
 agtcgagcga acagatgagg agcttgctcc tctgacgtta gcggcggacg ggtgagtaac 120
 acgtggataa cctacctata agactgggat aacttcggga aaccggagct aataccggat 180
 aatatattga accgcatggt tcaatagtga aagacggttt tgctgtcact tatagatgga 240

tccgcgccgc attagetagt tggttaaggta acggcttacc aaggcaacga tgcgtagccg 300
 acctgagagg gtgatcggcc aacttggaac tgagacacgg tccagactcc tacgggaggc 360
 agcagtaggg aatcttccgc aatgggcgaa agcctgacgg agcaacgccg cgtgagtgat 420
 gaaggtcttc ggatcgtaaa actctgttat tagggaagaa caaatgtgta agtaactatg 480
 cacgtcttga cggtaacaa tcagaaagcc acggctaact acgtgccagc agcccggtta 540
 atacgtaggt ggcaagcgtt atccggaatt attgggcgta aagcgcgcgt aggcggtttt 600
 ttaagtctga tgtgaaagcc cacggctcaa ccgtggaggg tcattggaaa ctggaaaact 660

tgagtgcaga agaggaaagt ggaattccat gtgtagcggg gaaatgcgca gagatatgga 720
 ggaacaccag tggcgaaggc gactttctgg tctgtaactg acgctgatgt gcgaaagcgt 780
 ggggatcaaa caggattaga taccttggtg gtccacgccg taaacgatga gtgctaagtg 840
 ttagggggtt tccgcccctt agtctgcag ctaacgcatt aagcactccg cctggggagt 900
 acgaccgcaa ggttgaacct caaaggaatt gacggggacc cgcacaagcg gtggagcatg 960
 tggtttaatt cgaagcaacg cgaagaacct taccaaatct tgacatcctc tgaccctct 1020
 agagatagag ttttcccctt cgggggacag agtgacaggt ggtgcatggt tgtcgtcagc 1080

tcgtgtcgtg agatgttggg ttaagtcccg caacgagcgc aacccttaag cttagtgtcc 1140
 atcattaagt tgggcactct aagttgactg ccggtgacaa accggaggaa ggtggggatg 1200
 acgtcaaatc atcatgcccc ttatgatttg ggctacacac gtgctacaat ggacaataca 1260
 aagggtagcg aaaccgcgag gtcaagcaaa tcccataaag ttgttctcag ttcggattgt 1320
 agtctgcaac tcgactatat gaagctggaa tcgctagtaa tcgtagatca gcatgctacg 1380
 gtgaatacgt tcccgggtct tgtacacacc gccctcaca ccacgagagt ttgtaacacc 1440
 cgaagccggt ggagtaacca tttggagcta gccgtcgaag gtgggacaaa tgattggggt 1500

gaagtcgtaa caagtagcc gtatcggaag gtgcggctgg atcacctcct ttct 1554

<210> 18

<211> 1547

<212> DNA

<213> Xanthomonas campestris ATCC 33913

<400> 18

taagtgaaga gtttgatcct ggctcagagt gaacgctggc ggcaggccta acacatgcaa	60
gtcgaacggc agcacagtaa gagcttgctc ttatgggtgg cgagtggcgg acgggtgagg	120
aatacatcgg aatctactct ttcgtggggg ataacgtagg gaaacttacg ctaataccgc	180
atacgaccta cgggtgaaag cggaggacct tcgggcttcg cgcgattgaa tgagccgatg	240
tcggattagc tagttggcgg ggtaaaggcc caccaaggcg acgatccgta gctggtctga	300
gaggatgac agccacactg gaactgagac acggtccaga ctctacggg aggcagcagt	360
ggggaatatt ggacaatggg cgcaagcctg atccagccat gccgcgtggg tgaagaaggc	420
cttcgggttg taaagccctt ttgttgggaa agaaaagcag tcggttaata cccgattgtt	480
ctgacggtac ccaaagaata agcaccgct aacttcgtgc cagcagccgc ggtaatacga	540
agggtgcaag cgttactcgg aattactggg cgtaaagcgt gcgtaggtgg tggtttaagt	600
ctgttgtgaa agccctgggc tcaacctggg aattgcagtg gatactgggt cactagagtg	660
tggtagaggg tagcgaatt cccggttag cagtgaaatg cgtagagatc gggaggaaca	720
tccgtggcga aggcggctac ctggaccaac actgacactg aggcacgaaa gcgtggggag	780
caaacaggat tagataccct ggtagtccac gccctaaacg atgcgaactg gatgttgggt	840
gcaatttggc acgcagtatc gaagctaacg cgtaaagttc gccgcctggg gagtacggtc	900
gcaagactga aactcaaagg aattgacggg ggccccaca agcgggtggag tatgtggttt	960
aattcgatgc aacgcgaaga accttacctg gtcttgacat ccacggaact ttccagagat	1020
ggattggtgc cttcgggaac cgtgagacag gtgctgcatg gctgtcgtca gctcgtgctg	1080
tgagatgttg ggtaaagtcc cgcaacgagc gcaacccttg tccttagttg ccagcacgta	1140
atggtgggaa ctctaaggag accgccgtg acaaaccgga ggaaggtggg gatgacgtca	1200
agtcatcatg gcccttacga ccagggttac acacgtacta caatggtagg gacagagggc	1260
tgcaaaccg cgagggtaa ccaatcccag aaacctatc tcagtccgga ttggagtctg	1320
caactcgact ccatgaagtc ggaatcgta gtaatcgag atcagcattg ctgcbgtgaa	1380
tacgttccc ggcttgtac acaccgccg tcacacatg ggagtttgtt gcaccagaag	1440
caggtagctt aacctcggg agggcgcttg ccacggtgtg gccgatgact ggggtgaagt	1500
cgtaacaagg tagccgtatc ggaaggtgcg gctggatcac ctctttt	1547

<210> 19

<211> 1410

<212> DNA

<213> *Roseobacter denitrificans* OCh 114
 <400> 19
 tcaacttgag agtttgatcc tggctcagaa cgaacgctgg cggcaggcct aacacatgca 60
 agtcgagcgc tcacttcggt gggagcggcg gacgggttag taacgcgtgg gaacataccc 120
 ttctctacgg aatagccttt ggaaacgaag agtaataaccg tatacgcctc tcgggggaaa 180
 gatttatcgg agatggattg gcccgcgtta gattagatag ttggtggggt aatgcctac 240
 caagtctacg atctatagct ggttttagag gacgatcagc aacctggga ctgagacacg 300
 gcccagactc ctacgggagg cagcagtggg gaatcttaga caatgggcga aagcctgatc 360

 tagccatgcc gcgtgagtga tgaaggccct agggctgtaa agctctttcg ccagggatga 420
 taatgacagt acctggtaaa gaaaccccg gtaactccgt gccagcagcc gcggtaatc 480
 ggagggggtt agcgtttgtc ggaattactg ggcgtaaagc gcacgtaggc ggatcagaaa 540
 gttaggggtg aatcccag gctcaacctc ggaactgcct ctaaaactcc tggctttgag 600
 ttcgagagag gtgagtgaa ttccaagtgt agaggtgaaa ttcgtagata tttggaggaa 660
 caccagtggc gaaggcggct cactggctcg atactgacgc tgaggtcga aagtgtgggg 720
 agcaaacagg attagatacc ctggtagtcc acaccgtaa cgatgaatgc cagtctcgg 780

 gcagtatact gttcggtgac acacctaacg gattaagcat tccgcctggg gactacggtc 840
 gcaagattaa aactcaaagg aattgacggg ggccccaca agcggtgagg catgtggttt 900
 aattcgaagc aacgcgcaga acctaccaa cccttgacat cctgtgctaa cccgagagat 960
 cgggcgttct cgcaagagac gcagtgcagc gtgctgcatg gctgtcgtca gctcgtgtcg 1020
 tgagatgttc ggttaagtcc ggcaacgagc gcaaccaca tctttagtgt ccagcagttc 1080
 ggctgggcac tctaaagaaa ctgcccgtga taagcgggag gaagggtgtg atgacgtcaa 1140
 gtcctcatgg cccttacggg ttgggtaca cacgtgtac aatggtagtg acaatgggtt 1200

 aatcccaaaa agctatctca gttcggattg gggtctgcaa ctcgaccca tgaagtcgga 1260
 atcgctagta atcgctaac agcatgacgc ggtgaatacg ttcccgggcc ttgtacacac 1320
 cccccgtcac accatgggag ttggttctac ccgacgacgc tgcgctaacc cttcggggag 1380
 gcaggcggcc acggtaggat cagcagactgg 1410

 <210> 20
 <211> 1455
 <212> DNA
 <213> *Rhodobacter sphaeroides* KD131
 <400> 20
 agagtttgat cctggctcag aatgaacgct ggcggcaggc ctaacacatg caagtcgagc 60

gaagtcttcg gacttagcgg cggacgggtg agtaacgcgt gggaacgtgc cctttgcttc 120

 ggaatagccc cgggaaactg ggagtaatac cgaatgtgcc ctttggggga aagatttatc 180
 ggcaaaggat cggccccgct tggattaggt agttgggtgg gtaatggcct accaagccga 240
 cgatccatag ctggtttgag aggatgatca gccacactgg gactgagaca cggcccagac 300
 tcctacggga ggacagcagt gggaaatcta gacaatgggc gcaagcctga tctagccatg 360
 ccgctgatc gatgaaggcc ttagggttgt aaagatcttt caggtgggaa gataatgacg 420
 gtaccaccag aagaagcccc ggctaactcc gtgccagcag ccgcggtaat acggaggggg 480
 ctagcgttat tcggaattac tgggcgtaaa gcgcacgtag gcggatcgga aagtcagagg 540

 tgaaatecca gggctcaacc ctggaactgc ctttgaaact cccgatcttg aggtcgagag 600
 aggtgagtgg aattccgagt gtagaggta aattcgtaga tattcggagg aacaccagtg 660
 gcgaaggcgg ctcaactggct cgatactgac gctgaggtgc gaaagcgtgg ggagcaaaca 720
 ggattagata cctggttagt ccacgccgta aacgatgaat gccagtcgtc gggcagcatg 780
 ctgttcgggt acacacctaa cggattaagc attccgcctg gggagtacgg ccgcaaggtt 840
 aaaactcaa ggaattgacg gggccccga caagcgggtg agcatgtggt ttaattcгаа 900
 gcaacgcgca gaaccttacc aaccttgac atggcgatcg cggttcaga gatggttct 960

 tcagttcggc tggatcgac acaggtgctg catggctgtc gtcagctcgt gtcgtgagat 1020
 gttcggttaa gtcggcaac gagcgcaacc cacgtcctta gttgccagca ttcagttggg 1080
 cactctaggg aaactgccgg tgataagccg gaggaaggtg tggatgacgt caagtcctca 1140
 tggcccttac gggttgggct acacacgtgc tacaatggca gtgacaatgg gttaatcca 1200
 aaaagctgtc tcagttcgga ttggggtctg caactcgacc ccatgaagtc ggaatcgcta 1260
 gtaatcgct aacagcatga cgcggtgaat acgttcccgg gccttgtaca caccgccct 1320
 cacacatgg gaattggtc taccggaagg cggtcgcca acctcgcaag aggaggcagc 1380

 cgaccacggt aggatcagt actggggtga agtcgtaaca aggtagccgt aggggaacct 1440
 gcggctggat cacct 1455