

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2022年8月11日 (11.08.2022)



(10) 国际公布号
WO 2022/166572 A1

(51) 国际专利分类号:
C12N 15/70 (2006.01) *C12N 15/66* (2006.01)

(21) 国际申请号: PCT/CN2022/072437

(22) 国际申请日: 2022年1月18日 (18.01.2022)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202110171187.1 2021年2月8日 (08.02.2021) CN

(71) 申请人: 中国科学院分子植物科学卓越创新中心 (CAS CENTER FOR EXCELLENCE IN MOLECULAR PLANT SCIENCES) [CN/CN]; 中国上海市徐汇区枫林路300号, Shanghai 200032 (CN)。

(72) 发明人: 覃重军 (QIN, Zhongjun); 中国上海市徐汇区枫林路300号, Shanghai 200032 (CN)。 鲁宁 (LU, Ning); 中国上海市徐汇区枫林路300号, Shanghai 200032 (CN)。 薛小莉 (XUE, Xiaoli); 中国上海市徐汇区枫林路300号, Shanghai 200032 (CN)。 钟莉 (ZHONG, Li); 中国上海市徐汇区枫林路300号, Shanghai 200032 (CN)。

(74) 代理人: 上海专利商标事务所有限公司 (SHANGHAI PATENT & TRADEMARK LAW OFFICE, LLC); 中国上海市桂平路435号, Shanghai 200233 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,

NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告 (条约第21条(3))。
- 包括说明书序列表部分 (细则5.2(a))。

(54) Title: METHOD AND SYSTEM FOR CONTINUOUS CLONING OF LONG DNA FRAGMENT

(54) 发明名称: 连续克隆长DNA片段的方法和系统

(57) Abstract: Provided are a polynucleotide, a nucleic acid construct, a method, and a system for constructing a long DNA fragment, wherein the polynucleotide contains a double-strand cleavage recognition site, a transfer initiation site ORIT, and an origin of replication.

(57) 摘要: 提供用于构建长DNA片段的多核苷酸、核酸构建物、方法和系统, 所述多核苷酸包含: 双链切割识别位点, 转移起始位点oriT, 复制起点。



WO 2022/166572 A1

连续克隆长 DNA 片段的方法和系统

技术领域

- 5 本发明涉及 DNA 克隆领域，具体涉及长 DNA 片段的连续克隆的方法和系统。

背景技术

全基因组测序信息表明不同生物基因组的大小差异很大，简单和低等的生物基因组相对较小，比如细菌基因组有 1-10 Mb，酵母基因组为 12 Mb，随着生物进化得越来越复杂，生物的基因组也变得越来越复杂，比如植物拟南芥基因组有 120 Mb，昆虫果蝇基因组有 137 Mb，小鼠基因组有 2.6 Gb，而人的基因组有 3.0 Gb。随着人类对生命科学探索的不断深入，人工合成生物基因组工作正在全球多个国家有序展开，我们对于基因组超大 DNA 片段的快速精准组装的需求日益迫切。此外，在复杂高等生物基因组中，功能相同或相关的基因常常聚集在一起，形成超大的功能基因簇，大小可达 1 Mb。而超大的 DNA 片段在操作过程中存在随机断裂、易降解、难以转移、耗时长等缺陷。发展超大 DNA 片段 (≥ 1 Mb) 操作技术对人工生命体的创建以及复杂高等生物基因组功能的研究及应用都至关重要。

20 但是，目前已发展的经典分子生物学常用的克隆载体不能满足克隆基因组超大 DNA 片段的需求。比如来源于小质粒、噬菌体和粘粒的克隆载体通常只能容纳 < 40 kb 的外源 DNA 片段。细菌人工染色体 (BAC) 系统是一种常用的细菌克隆系统，常用于克隆 100-300 kb 的 DNA 片段。BAC 载体来自于大肠杆菌的单拷贝质粒 F 因子，且在宿主体内遗传复制稳定，使细菌人工染色体系统广泛应用于基因组文库构建。但是，传统的 BAC 克隆是一种环型形式的 DNA，较大的环型 DNA 难以与大肠杆菌环型基因组分离、检测及纯化。酵母人工染色体 (YAC) 是酵母中常用来高等生物基因组文库的载体。虽然酵母人工染色体能够容纳大于 1 Mb 的外源 DNA 片段，但长重复序列的 DNA 片段在宿主体内不稳定的，不稳定的 YAC 载体克隆会引发外源片段的重组丢失。另外，由

于共连接或重组，YAC 文库存在 20-60%的嵌合体克隆，即包含多于 2 个非连续片段基因组区段，这极大限制了酵母人工染色体的应用。

大肠杆菌是最常用的遗产操作宿主，它比酵母生长更快，而且遗产操作也更容易。本发明仍需发展在大肠杆菌中连续克隆基因组超大片段 DNA 的新方法。

发明内容

本发明涉及新的克隆超大片段 DNA 的线型载体，并利用新载体建立了超大 DNA 片段的体内连续克隆拼接的新方法。利用本发明实现低成本短时间高效迭代无缝拼接组装出大于 1 Mb 的超大 DNA 片段。

本发明第一方面提供一种多核苷酸，用于构建长 DNA 片段，所述多核苷酸包含：双链切割识别位点，转移起始位点 oriT，复制起点。

在一个或多个实施方案中，所述多核苷酸还包含 5'同源臂和 3'同源臂。所述同源臂用于借助基因编辑系统使所述多核苷酸与目的片段整合(成环形)。

在一个或多个实施方案中，所述多核苷酸还包括借助基因编辑系统通过 5'同源臂和 3'同源臂整合的目的片段。

在一个或多个实施方案中，所述 5'同源臂和 3'同源臂分别识别目的片段的 3'端和 5'端。

在一个或多个实施方案中，所述基因编辑系统选自 CRISPR、ZFN、TALEN。

在一个或多个实施方案中，双链切割识别位点，转移起始位点 oriT 和复制起点位于 5'同源臂和 3'同源臂之间。

在一个或多个实施方案中，5'同源臂长度为 30-1000bp、50-900bp、100-800bp、200-700bp、300-600bp、400-500bp。

在一个或多个实施方案中，3'同源臂长度为 30-1000bp、50-900bp、100-800bp、200-700bp、300-600bp、400-500bp。

在一个或多个实施方案中，双链切割识别位点或其互补序列能被其切割酶切割。

在一个或多个实施方案中，双链切割识别位点是 tos 位点。

在一个或多个实施方案中，*tos* 位点来源于噬菌体 N15。

在一个或多个实施方案中，*tos* 位点序列包含 SEQ ID NO:1 所示的序列或其互补序列。

5 在一个或多个实施方案中，转移起始位点 *oriT* 来源于肠杆菌科细菌，优选埃希氏菌属细菌，例如大肠杆菌。

在一个或多个实施方案中，转移起始位点 *oriT* 来源于大肠杆菌 BAC。

在一个或多个实施方案中，转移起始位点 *oriT* 来源于质粒 pQX17。

在一个或多个实施方案中，转移起始位点 *oriT* 包含 SEQ ID NO:2 所示的序列。

10 在一个或多个实施方案中，复制起点来源于肠杆菌科细菌，优选埃希氏菌属细菌。

在一个或多个实施方案中，复制起点来源于大肠杆菌。

在一个或多个实施方案中，复制起点包含 SEQ ID NO:3 或 4 所示的序列。

在一个或多个实施方案中，转移起始位点 *oriT* 和复制起点的位置可互换。

15 在一个或多个实施方案中，双链切割识别位点位于转移起始位点 *oriT* 和复制起点的 5' 端或 3' 端。

在一个或多个实施方案中，双链切割识别位点位于转移起始位点 *oriT* 和复制起点之间。

20 在一个或多个实施方案中，双链切割识别位点位于复制起点的 5' 端。此时切割和重组效率高于双链切割识别位点位于复制起点的 3' 端的情况。

在一个或多个实施方案中，所述多核苷酸依次包含：

5'同源臂，复制起点，转移起始位点 *oriT*，双链切割识别位点，3'同源臂；

5'同源臂，转移起始位点 *oriT*，复制起点，双链切割识别位点，3'同源臂；

5'同源臂，转移起始位点 *oriT*，双链切割识别位点，复制起点，3'同源臂；

25 5'同源臂，复制起点，双链切割识别位点，转移起始位点 *oriT*，3'同源臂；

5'同源臂，双链切割识别位点，转移起始位点 *oriT*，复制起点，3'同源臂；

5'同源臂，双链切割识别位点，复制起点，转移起始位点 *oriT*，3'同源臂。

在一个或多个实施方案中，所述多核苷酸还包含标记基因，例如抗生素抗性基因。

在一个或多个实施方案中，标记基因选自：壮观霉素筛选标记 *Spc*、阿泊拉霉素筛选标记 *Apr*、氨苄青霉素抗性筛选标记中的一种或多种。

在一个或多个实施方案中，标记基因位于 5' 同源臂和 3' 同源臂之间。

在一个或多个实施方案中，标记基因位于转移起始位点 *oriT* 和复制起点
5 之间。

在一个或多个实施方案中，标记基因位于复制起点的 3' 端。

在一个或多个实施方案中，双链切割识别位点位于标记基因的 5' 端或 3' 端。

本发明第二方面提供一种核酸构建物，包含本文任一实施方案所述的多核
10 苷酸和目的片段，用于构建长 DNA 片段。

在一个或多个实施方案中，所述目的片段是长 DNA 片段的一部分。

在一个或多个实施方案中，所述长 DNA 片段的长度至少 2kb，例如
2kb-1.5Mb。在一个或多个实施方案中，所述长 DNA 片段的长度至少 10kb、
至少 20kb、至少 30kb、至少 40kb、至少 50kb、至少 60kb、至少 70kb、至少
15 80kb、至少 90kb、至少 100kb、至少 200kb、至少 300kb、至少 400kb、至少
500kb、至少 600kb、至少 700kb、至少 800kb、至少 900kb、至少 1Mb、至少
1.1Mb 或上述任意两个数值之间的范围。

在一个或多个实施方案中，所述目的片段的长度至少 1kb，例如为
1kb-30kb，例如 2kb-29kb、3kb-28kb、4kb-27kb、5kb-26kb、6kb-25kb、7kb-24kb、
20 8kb-23kb、9kb-22kb 或上述任意两个数值之间的范围。

在一个或多个实施方案中，所述目的片段位于所述多核苷酸之外。

在一个或多个实施方案中，所述核酸构建物是环形构建物。

在一个或多个实施方案中，所述核酸构建物是线性构建物，其中双链切割
识别位点或其互补序列被切割酶切割。

25 在一个或多个实施方案中，所述双链切割识别位点是 *tos* 位点，所述切割
酶是 *TeIN*。

在一个或多个实施方案中，*tos* 位点来源于噬菌体 N15。

在一个或多个实施方案中，*tos* 位点序列包含 SEQ ID NO:1 所示的序列或
其互补序列。

在一个或多个实施方案中，TelN 的氨基酸序列如 SEQ ID NO:5 所示。

在一个或多个实施方案中，所述核酸构建物包含：

(1) 第一核酸构建物，其具有本文任一实施方案所述的多核苷酸和第一目的片段，

5 (2) 第二核酸构建物，其具有本文任一实施方案所述的多核苷酸和第二目的片段，

其中第一目的片段的 3'端与第二目的片段的 5'端具有 1kb-200kb 的重叠区域。

10 在一个或多个实施方案中，所述重叠区域为 2kb-150kb、3kb-140kb、4kb-130kb、5kb-120kb、5kb-100kb、5kb-80kb、或 5kb-60kb。

在一个或多个实施方案中，第一核酸构建物是线性构建物，其中双链切割识别位点或其互补序列被切割酶切割。

在一个或多个实施方案中，第二核酸构建物是环形构建物。

15 在一个或多个实施方案中，第一核酸构建物与第二核酸构建物分别具有标记基因，例如抗生素抗性基因。

在一个或多个实施方案中，第一核酸构建物具有一种或两种标记基因，且满足以下条件：(1) 至少一种标记基因位于第一目的片段的 5'端，并且任选地(2) 双链切割识别位点位于所述至少一种标记基因的 5'端。优选地，所述两种标记基因不同。在一个或多个实施方案中，两种标记基因分别位于第一目的片段的 5'端和 3'端。

20 在一个或多个实施方案中，第二核酸构建物的标记基因位于第二目的片段的 3'端，并且双链切割识别位点位于所述标记基因的 3'端。

在一个或多个实施方案中，第一核酸构建物与第二核酸构建物中的标记基因均不相同。

25 在一个或多个实施方案中，标记基因选自：壮观霉素筛选标记 Spc、阿泊拉霉素筛选标记 Apr、氨苄青霉素抗性筛选标记中的一种或多种。

本发明还提供宿主细胞，包含本文第二方面所述的核酸构建物。

在一个或多个实施方案中，所述核酸构建物是环形构建物，并且所述宿主

细胞不表达针对所述双链切割识别位点的切割酶。任选地，所述宿主细胞还包含接合转移辅助质粒或其具有引导接合转移的功能片段，例如 pUZ8002。

在一个或多个实施方案中，所述核酸构建物是线性构建物，并且所述宿主细胞表达针对所述双链切割识别位点的切割酶。

- 5 在一个或多个实施方案中，所述宿主细胞是肠杆菌科细菌，优选埃希氏菌属细菌，更优选大肠杆菌。

本发明第三方面提供一种核酸构建物或包含该核酸构建物的宿主细胞，所述核酸构建物包含与启动子操作性连接的 *TelN* 的编码序列。

- 10 在一个或多个实施方案中，所述启动子是组成型启动子，例如 J23100 启动子。

在一个或多个实施方案中，所述宿主细胞是肠杆菌科细菌，优选埃希氏菌属细菌，更优选大肠杆菌。

- 15 本发明还提供一种构建长 DNA 片段或使目的片段整合成较长片段的方法，包括：

(1) 在使宿主细胞发生接合转移的条件下使受体宿主细胞和供体宿主细胞接触，其中，

- 20 所述受体宿主细胞包含本文第二方面中所述的第一核酸构建物，所述核酸构建物是线性构建物，其中双链切割识别位点或其互补序列被切割酶切割，

所述供体宿主细胞包含本文第二方面中所述的第二核酸构建物，所述核酸构建物是环形构建物；

- 25 (2) 孵育受体宿主细胞，所述第一目的片段和所述第二目的片段发生同源重组，所述同源重组获得具有第一目的片段和第二目的片段的第三核酸构建物。

在一个或多个实施方案中，所述受体宿主细胞表达针对所述双链切割识别位点的切割酶。

在一个或多个实施方案中，所述供体宿主细胞不表达针对所述双链切割识别位点的切割酶。

在一个或多个实施方案中，所述供体宿主细胞还包含接合转移辅助质粒或其具有引导接合转移的功能片段，例如 pUZ8002。

在一个或多个实施方案中，所述第一核酸构建物和/或所述第二核酸构建物的其他特征如本文第二方面中所述。

5 在一个或多个实施方案中，所述方法还包括

(0.1) 利用 5' 同源臂和 3' 同源臂将本文所述的多核苷酸与第一目的片段整合，获得包含所述第一核酸构建物的受体宿主细胞，

(0.2) 利用 5' 同源臂和 3' 同源臂将本文所述的多核苷酸与第二目的片段整合，获得包含所述第二核酸构建物的供体宿主细胞。

10 在一个或多个实施方案中，(0.1) 还包括在所述受体宿主细胞表达针对所述双链切割识别位点的切割酶。例如在所述受体宿主细胞中转入含有所述切割酶的编码序列的核酸构建物，优选表达载体或整合载体。

在一个或多个实施方案中，(0.2) 还包括在所述供体宿主细胞中转入接合转移辅助质粒或其具有引导接合转移的功能片段，例如 pUZ8002。

15 在一个或多个实施方案中，所述 5' 同源臂和 3' 同源臂分别识别目的片段的 3' 端和 5' 端。

在一个或多个实施方案中，所述整合是通过基因编辑系统将所述多核苷酸与目的片段连接成环。

20 在一个或多个实施方案中，所述基因编辑系统选自 CRISPR、ZFN、TALEN。

在一个或多个实施方案中，所述使宿主细胞发生接合转移的条件包括：

供体宿主细胞和/或受体宿主细胞处于指数生长期；

供体宿主细胞与受体宿主细胞的总浓度为 10^8 个/mL；

供体宿主细胞与受体宿主细胞的比例为 10:1-1:10，例如 5:1-1:1，优选

25 4:1；

培养温度 25-40°C，优选 30-34°C；

培养时间至少 0.5 小时。

在一个或多个实施方案中，(2) 所述孵育的温度为 25-40°C，优选 30-34°C。

本发明还提供一种用于构建长 DNA 片段或使目的片段整合成较长片段的系统，包含本文任一实施方案所述的多核苷酸、核酸构建物、或供体和受体宿主细胞。

5

本发明优点：

1、对高等动植物复杂基因组超大功能基因簇功能及人工改造等研究亟需发展超大 DNA (>1 Mb) 的克隆拼接技术。本发明的克隆拼接技术可以避免体外操作 DNA 大片段的困难，快速高效拼接组装出 >1 Mb 的超大 DNA 片段。

10

2、本发明构建一种 *TelN/tos* 系统，可以将传统的环型质粒 DNA 转换成线型质粒 DNA。在不破坏大肠杆菌 *MDS42* 各基因的正常功能前提下，通过 CRISPR/Cas9 系统在大肠杆菌基因组中插入组成型启动子带动的 *TelN* 组件。另外将环型质粒 DNA 上引入 *tos* 位点，再将该环型质粒 DNA 转化进组成型表达 *TelN* 蛋白的宿主体内，端粒酶 *TelN* 将结合在 *tos* 位点，先切开后封住缺口，形成 2 个闭合末端发夹结构。在 *TelN/tos* 系统转换成线型形式存在后，线型质粒 DNA 之间拼接只需要 1 次单交换，而环型质粒 DNA 之间拼接需要交换 2 次，因此，采用线型形式质粒 DNA 的拼接效率更高。

15

3、本发明中我们引入了大肠杆菌的 IV 型分泌系统(T4SS)⁴，巧妙的避免了 DNA 大片段体外操作的困难。即在供体菌的质粒上引入了转移起始位点 *oriT*，在额外转入质粒 pUZ8002 的辅助下，供体菌环型质粒以单链形式接合转移到受体菌中。这种设计一方面避免了从供体菌中提取大质粒和再电击转化受体菌株的复杂操作，另一方面极大的降低了拼接组装成本和时间消耗。

20

4、本发明通过利用 3 个高效元件创建了新型的可容纳 >1 Mb 超大 DNA 的线型克隆载体，并由此发展了可在原核模式菌大肠杆菌体内实现克隆超过 1 Mb 外源 DNA 超大片段的新方法。相比现有方法，本发明避免了大片段 DNA 的复杂低效体外操作，极大降低了拼接组装成本和时间消耗；另一方面，与现有方法需要进行两个位点的重组交换不同，本发明的克隆方法仅需要通过单次重组交换即可，DNA 大片段克隆拼接效率更为高效。我们利用该方法成功克隆拼接了完整的人免疫球蛋白重链可变区 (1.07 Mb) 的超大 DNA 片段。本发

25

明将在高等生物复杂基因组的克隆和拼接组装中有广泛应用。

附图说明

5 **图 1** 是质粒 p1 的构建示意图。将融合 PCR 获得的打靶组件与质粒 S4-gRNA 同时电击转化转入菌株 MDS42 中，再借助于 CRISPR/Cas9 系统替换目标区域，完成环型质粒 p1 的构建。

图 2 是质粒 p2 的构建示意图。将融合 PCR 获得的打靶组件与质粒 SCK-gRNA 同时电击转化转入菌株 MDS42 中，再借助于 CRISPR/Cas9 系统替换目标区域，完成环型质粒 p2 的构建。

10 **图 3** 是质粒 p3 的构建示意图。将融合 PCR 获得的打靶组件与质粒 SC-gRNA 同时电击转化转入菌株 MDS42 中，再借助于 CRISPR/Cas9 系统替换目标区域，完成环型质粒 p3 的构建。

图 4 是一种实施方案的接合转移拼接组装示意图。① 将供体菌 MDS42 的环型质粒 p2 通过接合转移方式转入受体菌 MT 中；② 在不稳定 *oriC* 复制区的压力下，线性化的质粒 p2 与线型质粒 p1 会发生 1 次单交换；③ 在 2 种抗生素的筛选压力下，如果拼接组装成功，会生成线型质粒 pA。其中，字母 T 为 *oriT* 元件，R1 为抗性筛选标记 1，R2 为抗性筛选标记 2。

15 **图 5** 是一种实施方案的接合转移拼接组装示意图。① 将供体菌 MDS42 的环型质粒 p4 通过接合转移方式转入受体菌 MT 中；② 在相同复制区质粒不相容的压力下，线性化的质粒 p4 与线型质粒 pB 会发生 1 次单交换；③ 在 2 种抗生素的筛选压力下，如果拼接组装成功，会生成线型质粒 pC。其中，字母 T 为 *oriT* 元件，R1 为抗性筛选标记 1，R2 为抗性筛选标记 2，R3 为抗性筛选标记 3。

图 6 显示一种实施方案的脉冲场电泳验证拼接组装结果。用 PFGE 琼脂糖制作浓度为 1% 的凝胶，电泳缓冲液为 0.5xTBE，温度为 12°C，转换时间为 10-60 s，角度为 120°，运行时间为 17 h，电压梯度为 6 V/cm。

图 7 显示一种实施方案的脉冲场电泳验证拼接组装结果。用 PFGE 琼脂糖制作浓度为 1% 的凝胶，电泳缓冲液为 0.5xTBE，温度为 12°C，转换时间为 10-60 s，角度为 120°，运行时间为 17 h，电压梯度为 6 V/cm。

图 8 显示脉冲场电泳验证大片段拼接组装结果。图 8A：用 PFGE 琼脂糖制作浓度为 1%的凝胶，电泳缓冲液为 0.5xTBE，温度为 12°C，转换时间为 10-60 s，角度为 120°，运行时间为 17 h，电压梯度为 6 V/cm。图 8B：PFGE 琼脂糖制作浓度为 1%的凝胶，电泳缓冲液为 0.5xTBE，温度为 12°C，角度为 120 电压梯度为 6 V/cm。模块 1 的转换时间为 60 s，22 h；模块 2 的转换时间为 90 s，12 h。

具体实施方式

本发明利用大肠杆菌复制元件（例如大肠杆菌染色体的 *oriC* 复制元件或来源于大肠杆菌 F 因子的 BAC 复制元件）构建了新型克隆载体，该载体可通过切割酶（例如原核端粒酶 TelN 蛋白）识别切割双链切割识别位点（例如 *tos* 位点）使载体 DNA 线性化，再借助于大肠杆菌自身的同源重组能力一步实现两个线型 DNA 大片段的拼接组装。此外，为了使含有两个线型 DNA 大片段同源重组，发明人引入了大肠杆菌的 IV 型分泌系统(T4SS)⁴，通过在供体菌的质粒上引入了转移起始位点 *oriT*，巧妙的避免了 DNA 大片段体外操作的困难。

本发明所用术语“核酸”、“核苷酸”、“多核苷酸”或“核酸分子”可以是 DNA 形式或 RNA 形式。DNA 形式包括 cDNA、基因组 DNA 或人工合成的 DNA。DNA 可以是单链的或是双链的。DNA 可以是编码链或非编码链。提到核酸时，本文所用术语“变体”可以是天然发生的等位变体或非天然发生的变体。这些核苷酸变体包括简并变体、取代变体、缺失变体和插入变体。如本领域所知的，等位变体是一个核酸的替换形式，它可能是一个或多个核苷酸的取代、缺失或插入，但不会从实质上改变其编码的蛋白的功能。本发明核酸可包含与所述核酸序列的序列相同性为至少约 70%、至少约 75%、至少约 80%、至少约 85%、至少约 90%、至少约 95%、至少约 98%、至少约 99%或 100%的核苷酸序列。

本文所述“长 DNA 片段”或“超大 DNA 片段”的长度至少 2kb，例如至少 10kb、至少 20kb、至少 30kb、至少 40kb、至少 50kb、至少 60kb、至少 70kb、至少 80kb、至少 90kb、至少 100kb、至少 200kb、至少 300kb、至少 400kb、至少 500kb、至少 600kb、至少 700kb、至少 800kb、至少 900kb、至少 1Mb、

至少 1.1Mb 或上述任意两个数值之间的范围。

本文中，“双链切割识别位点”表示通过酶促或非酶促方法识别的核酸切割位点。任何双链切割识别位点及其对应的切割酶或同源序列均可用于本发明。有时，切割酶可以识别双链切割识别位点或其互补序列。在具体实施方案中，使用来源于噬菌体 N15 的由端粒酶 TelN 蛋白识别的 *tos* 位点。*tos* 位点序列如 SEQ ID NO:1 或其互补序列所示。TelN 的氨基酸序列如 SEQ ID NO:5 所示。

本文中，“转移起始位点 *oriT*”是细菌 IV 型分泌系统所需的组件。细菌 IV 型分泌系统是与细菌接合机制有关的一类分泌系统。IV 型分泌系统可以转运 DNA。具有含转移起始位点 *oriT* 的质粒的细菌可以在接合转移辅助组件的存在下于 *oriT* 处发生单链缺刻，并通过细菌接触将该质粒的单链拷贝导入邻近接触的细菌中，从而实现 DNA 的转运。任何可引导细菌接合转移的转移起始位点 *oriT* 均可用于本发明。在一些实施方案中，转移起始位点 *oriT* 来源于肠杆菌科细菌，优选埃希氏菌属细菌，例如大肠杆菌。示例性地，转移起始位点 *oriT* 来源于大肠杆菌 BAC（质粒 pQX17）。在具体实施方案中，*oriT* 序列如 SEQ ID NO:2 所示。

本文中，“移辅助组件”可以是接合转移辅助质粒或其具有引导接合转移的功能片段，例如 pUZ8002。

本文中，“复制起点”和“复制元件”可互换使用，是 DNA 复制的起始位置。本领域知晓用于原核或真核细胞的 DNA 复制的复制起点。在一些实施方案中，转移起始位点 *oriT* 来源于肠杆菌科细菌，优选埃希氏菌属细菌，例如大肠杆菌。示例性的复制起点如 SEQ ID NO:3 或 4 所示。

本发明的新型克隆载体包含一段多核苷酸，所述多核苷酸包含：双链切割识别位点，转移起始位点 *oriT*，复制起点，并且任选在两端包含 5'同源臂和 3'同源臂。所述同源臂用于借助基因编辑系统使所述多核苷酸与含有目的片段的载体整合（成环形）。任何适用于原核或真核细胞的基因编辑系统均可用于本发明，例如 CRISPR、ZFN、TALEN。

本文多核苷酸中的各组件可以任意方式排列，只要双链切割识别位点切割后的线性 DNA 可以在细胞中自我复制即可。通常，双链切割识别位点位于复

制起点的 5' 端，转移起始位点位于双链切割识别位点的 5' 端或 3' 端。在优选实施方案中，双链切割识别位点位于转移起始位点 oriT 和复制起点之间。

不希望受任何理论的限制，在具体实施方案中，所述多核苷酸依次包含：5'同源臂，复制起点，转移起始位点 oriT，双链切割识别位点，3'同源臂；5'同源臂，转移起始位点 oriT，复制起点，双链切割识别位点，3'同源臂；5'同源臂，转移起始位点 oriT，双链切割识别位点，复制起点，3'同源臂；5'同源臂，复制起点，双链切割识别位点，转移起始位点 oriT，3'同源臂；5'同源臂，双链切割识别位点，转移起始位点 oriT，复制起点，3'同源臂；或 5'同源臂，双链切割识别位点，复制起点，转移起始位点 oriT，3'同源臂。

10 为了实现对含有所述多核苷酸的细菌的筛选，本文所述多核苷酸还可以包含标记基因，例如抗生素抗性基因。实施例中的示例性标记基因包括：壮观霉素筛选标记 Spc、阿泊拉霉素筛选标记 Apr、氨苄青霉素抗性筛选标记。通常，标记基因位于 5' 同源臂和 3' 同源臂之间。在一个或多个实施方案中，标记基因位于转移起始位点 oriT 和复制起点之间。在一个或多个实施方案中，双链切割识别位点位于标记基因的 3' 端或 5' 端。

15 本发明的核酸构建物（例如克隆载体），包含本文任一实施方案所述的多核苷酸和目的片段，用于构建长 DNA 片段。所述目的片段是本文所述长 DNA 片段的一部分，长度可为 1kb-30kb，例如 2kb-29kb、3kb-28kb、4kb-27kb、5kb-26kb、6kb-25kb、7kb-24kb、8kb-23kb、9kb-22kb 或上述任意两个数值之间的范围。

25 含有目的片段的载体可通过将目的片段克隆到所用载体中获得。例如，在知晓待克隆的长 DNA 片段后，可将其分为两个或多个长度为 1kb-30kb 的片段。然后，可以用 PCR 扩增法、重组法或人工合成的方法获得这些片段。一种可行的方法是用人工合成的方法来合成有关序列，尤其是片段长度较短时。通常，通过先合成多个小片段，然后再进行连接可获得序列很长的片段。一旦获得了有关的序列，就可以用重组法来大批量地获得有关序列。这通常是将其克隆入载体，再转入细胞，然后通过常规方法从增殖后的宿主细胞中分离得到有关序列或包含其的载体。示例性含目的片段的载体包括 pBeloBAC11、pBACe3.6、pBACGK1.1、pUC57-Brick。

通过两个同源臂，所述多核苷酸和目的片段可以顺序存在于核酸构建物中，即目的片段位于核酸构建物中所述多核苷酸之外的位置。示例性地，所述 5' 同源臂和 3' 同源臂分别识别目的片段的 3' 端和 5' 端。借助 CRISPR、ZFN、TALEN 等基因编辑系统，可将多核苷酸与目的片段或含目的片段的载体整合成环形质粒。在使用 CRISPR 的实施方案中，根据含目的片段的载体设计 gRNA，使得所述多核苷酸与目的片段顺序连接，即插入目的片段的某一端。

通常，本文所述核酸构建物是环形构建物。但由于所述核酸构建物中的多核苷酸中含有双链切割识别位点（如 *tos* 位点），因此在存在切割酶（如 *TelN*）的情况下，所述核酸构建物是线性构建物，其中双链切割识别位点或其互补序列被切割酶切割。

为了进行两个目的片段的连接整合操作，本发明提供一种系统，包含本文所述环形和线性核酸构建物和/或包含这些核酸构建物的宿主细胞。所述系统包含：（1）第一核酸构建物（线性构建物），其具有本文任一实施方案所述的多核苷酸和第一目的片段，其中双链切割识别位点或其互补序列被切割酶切割，（2）第二核酸构建物（环形构建物），其具有本文任一实施方案所述的多核苷酸和第二目的片段。第一目的片段的 3' 端与第二目的片段的 5' 端具有 1kb-200kb 的重叠区域，用于同源重组。在一个或多个实施方案中，所述重叠区域为 2kb-150kb、3kb-140kb、4kb-130kb、5kb-120kb、5kb-100kb、5kb-80kb、或 5kb-60kb。

为了对含有核酸构建物的细胞进行筛选，第一核酸构建物与第二核酸构建物分别具有标记基因，例如上述抗生素抗性基因。在一些实施方案中，第一核酸构建物具有一种或两种标记基因，且至少有一种标记基因位于第一目的片段的 5' 端，并且任选地，双链切割识别位点位于所述至少一种标记基因的 5' 端；优选地，所述两种标记基因不同；优选地，两种标记基因分别位于第一目的片段的 5' 端和 3' 端。在一些实施方案中，第二核酸构建物的标记基因位于第二目的片段的 3' 端，并且双链切割识别位点位于所述标记基因的 3' 端。

本文所述宿主细胞包括革兰氏阳性菌酵母，例如肠杆菌科细菌，优选埃希氏菌属细菌，更优选大肠杆菌。用重组 DNA 转化宿主细胞可用本领域技术人员熟知的常规技术进行。当宿主为原核生物如大肠杆菌时，能吸收 DNA 的感

受态细胞可在指数生长期后收获，用 CaCl_2 法处理，所用的步骤在本领域众所周知。另一种方法是使用 MgCl_2 。转化也可用电穿孔的方法进行。当宿主是真核生物，可选用如下的 DNA 转染方法：磷酸钙共沉淀法，常规机械方法如显微注射、电穿孔，脂质体包装等。

5 获得的转化子可以用常规方法培养，表达本发明的基因所编码的多肽。根据所用的宿主细胞，培养中所用的培养基可选自各种常规培养基。在适于宿主细胞生长的条件下进行培养。当宿主细胞生长到适当的细胞密度后，用合适的方法(如温度转换或化学诱导)诱导选择的启动子，将细胞再培养一段时间。

在上面的方法中的多肽可在细胞内、或在细胞膜上表达、或分泌到细胞外。

10 如果需要，可利用其物理的、化学的和其它特性通过各种分离方法分离和纯化重组的蛋白。这些方法是本领域技术人员所熟知的。这些方法的例子包括但不限于：常规的复性处理、用蛋白沉淀剂处理(盐析方法)、离心、渗透破菌、超处理、超离心、分子筛层析(凝胶过滤)、吸附层析、离子交换层析、高效液相层析(HPLC)和其它各种液相层析技术及这些方法的结合。

15 本文宿主细胞包括供体宿主细胞和受体宿主细胞。

供体宿主细胞包含第一核酸构建物(线性构建物)，并且供体宿主细胞不表达针对所述双链切割识别位点的切割酶。任选地，供体宿主细胞还包含接合转移辅助组件(例如转移辅助质粒或其具有引导接合转移的功能片段，例如 pUZ8002)。本领域知晓将转移辅助组件导入宿主细胞的方法，例如通过电转
20 或感受态转化。所述供体宿主细胞不表达针对所述双链切割识别位点的切割酶。

受体宿主细胞包含第二核酸构建物(环形构建物)，并且所述宿主细胞表达针对所述双链切割识别位点的切割酶。本领域知晓使宿主细胞表达所述切割酶的方法，例如通过电转或感受态转化将表达所述切割酶的表达框导入细胞。

25 通常，表达蛋白的表达框包括与编码基因(例如 TeIN 的编码基因)操作性连接的启动子(例如 J23100，序列如 SEQ ID NO:6 所示)、终止子、增强子等。所述表达框可以通过重组载体整合到细胞基因组中或者通过表达载体以质粒形式存在于细胞质中。本领域任何重组载体和表达载体均可用于本发明。载体通常含有用于质粒维系和用于克隆与表达外源性核苷酸序列的序列。所述序列

（在某些实施方案中总称为“侧翼序列”）通常包括一个或多个以下核苷酸序列：启动子、一个或多个增强子序列、复制起点、转录终止序列、含有供体和受体剪接位点的完全内含子序列、编码用于多肽分泌的前导序列的序列、核糖体结合位点、聚腺苷酸化序列、用于插入编码将要表达的抗体的核酸的多连接子区和可选标记元件。

因此，本发明还提供一种构建长 DNA 片段或使目的片段整合成较长片段的方法，包括：（1）在使宿主细胞发生接合转移的条件下使受体宿主细胞和供体宿主细胞接触，其中，所述受体宿主细胞包含本文所述的第一核酸构建物，所述核酸构建物是线性构建物，其中双链切割识别位点或其互补序列被切割酶切割；所述供体宿主细胞包含本文所述的第二核酸构建物，所述核酸构建物是环形构建物；（2）孵育受体宿主细胞，所述第一目的片段和所述第二目的片段发生同源重组，所述同源重组获得具有第一目的片段和第二目的片段的第三核酸构建物。本领域通常用于细菌（尤其是大肠杆菌）接合转移的条件均可用于本发明。在示例性实施方案中，供体宿主细胞和/或受体宿主细胞处于指数生长期；供体宿主细胞与受体宿主细胞的总浓度为 10^8 个/mL；供体宿主细胞与受体宿主细胞的比例为 10:1-1:10，例如 5:1-1:1，优选 4:1；培养温度 25-40°C，优选 30-34°C；培养时间至少 0.5 小时。

举例而言，供体菌细胞含有环形的第二核酸构建物和转移辅助组件；受体菌细胞表达切割酶，其含有线性的第一核酸构建物。当供体菌细胞和受体菌细胞接触时，在转移辅助组件存在下，供体菌细胞的第二核酸构建物接合转移进入受体菌细胞。在受体菌细胞细胞中，在复制起点的压力下，线性化的第二核酸构建物通过重叠区域与第一核酸构建物直接同源重组完成拼接组装。如图 4 和 5 所示。在多种标记基因的筛选下，拼接组装成功可获得两个目的片段重组拼接的线型质粒。

在一个或多个实施方案中，所述方法在步骤（1）之前还包括（a）利用 5' 同源臂和 3' 同源臂将本文所述的多核苷酸与第一目的片段整合，获得包含所述第一核酸构建物的受体宿主细胞，（b）利用 5' 同源臂和 3' 同源臂将本文所述的多核苷酸与第二目的片段整合，获得包含所述第二核酸构建物的供体宿主细胞。

本发明包括如下实施方案：

项目 1、一种多核苷酸，用于构建长 DNA 片段，所述多核苷酸包含：双链切割识别位点，转移起始位点 oriT，复制起点，双链切割识别位点或其互补序列能被切割酶切割，

- 5 优选地，所述多核苷酸还包含用于整合目的片段的 5'同源臂和 3'同源臂，所述目的片段是所述长 DNA 片段的一部分；更优选地，所述双链切割识别位点、转移起始位点 oriT 和复制起点位于 5'同源臂和 3'同源臂之间。

项目 2、如项目 1 所述的多核苷酸，其特征在于，所述双链切割识别位点是 tos 位点，所述切割酶是 TelN，

- 10 优选地，tos 位点来源于噬菌体 N15，和/或转移起始位点 oriT 来源于肠杆菌科细菌，优选埃希氏菌属细菌，和/或复制起点来源于肠杆菌科细菌，优选埃希氏菌属细菌，和/或所述多核苷酸还包含标记基因，和/或转移起始位点 oriT 和复制起点的位置可互换；

- 更优选地，tos 位点序列包含 SEQ ID NO:1 所示的序列或其互补序列，和/或转移起始位点 oriT 包含 SEQ ID NO:2 所示的序列，和/或复制起点包含 SEQ ID NO:3 或 4 所示的序列，和/或标记基因位于 5'同源臂和 3'同源臂之间，和/或标记基因位于转移起始位点 oriT 和复制起点之间，和/或标记基因位于复制起点的 3'端，和/或双链切割识别位点位于复制起点的 5'端。
- 15

- 项目 3、如项目 1 所述的多核苷酸，其特征在于，所述多核苷酸依次包含：
- 20 5'同源臂，复制起点，转移起始位点 oriT，双链切割识别位点，3'同源臂；5'同源臂，转移起始位点 oriT，复制起点，双链切割识别位点，3'同源臂；5'同源臂，转移起始位点 oriT，双链切割识别位点，复制起点，3'同源臂；5'同源臂，复制起点，双链切割识别位点，转移起始位点 oriT，3'同源臂；5'同源臂，双链切割识别位点，转移起始位点 oriT，复制起点，3'同源臂；或 5'同源臂，双链切割识别位点，复制起点，转移起始位点 oriT，3'同源臂。
- 25

项目 4、一种核酸构建物，包含项目 1-3 中任一项所述的多核苷酸和目的片段，用于构建长 DNA 片段，

优选地，所述目的片段是长 DNA 片段的一部分，和/或所述长 DNA 片段的长度至少 2kb，和/或所述目的片段的长度为至少 1kb，和/或所述核酸构建物

是环形构建物，或所述核酸构建物是线性构建物并且其中双链切割识别位点或其互补序列被切割酶切割。

项目 5、如项目 4 所述的核酸构建物，其特征在于，所述核酸构建物包含：

- 5 (1) 第一核酸构建物，其具有项目 1-3 中任一项所述的多核苷酸和第一目的片段，(2) 第二核酸构建物，其具有项目 1-3 中任一项所述的多核苷酸和第二目的片段，其中第一目的片段的 3'端与第二目的片段的 5'端具有 1kb-200kb 的重叠区域，

优选地，第一核酸构建物是线性构建物，其中双链切割识别位点或其互补序列被切割酶切割；第二核酸构建物是环形构建物。

- 10 项目 6、如项目 4 所述的核酸构建物，其特征在于，第一核酸构建物与第二核酸构建物分别具有标记基因，其中，

- 第一核酸构建物具有一种或两种标记基因，且满足以下条件：(1) 至少一种标记基因位于第一目的片段的 5'端，并且任选地(2) 双链切割识别位点位于所述至少一种标记基因的 5'端；优选地，两种标记基因不同；更优选地，
15 两种标记基因分别位于第一目的片段的 5'端和 3'端，

第二核酸构建物的标记基因位于第二目的片段的 3'端，并且双链切割识别位点位于所述标记基因的 3'端，

优选地，第一核酸构建物与第二核酸构建物中的标记基因均不相同。

项目 7、一种宿主细胞，包含项目 4-6 中任一项所述的核酸构建物，

- 20 优选地，所述核酸构建物是环形构建物，并且所述宿主细胞不表达针对所述双链切割识别位点的切割酶，所述宿主细胞还任选包含接合转移辅助质粒或其具有引导接合转移的功能片段，例如 pUZ8002；或者，所述核酸构建物是线性构建物，并且所述宿主细胞表达针对所述双链切割识别位点的切割酶，

更优选地，所述宿主细胞是肠杆菌科细菌。

- 25 项目 8、一种构建长 DNA 片段或使目的片段整合成较长片段的方法，包括：

(1) 在使宿主细胞发生接合转移的条件下使受体宿主细胞和供体宿主细胞接触，其中，

所述受体宿主细胞包含项目 5 或 6 中所述的第一核酸构建物，所述核酸构

建物是线性构建物，其中双链切割识别位点或其互补序列被切割酶切割，

所述供体宿主细胞包含项目 5 或 6 中所述的第二核酸构建物，所述核酸构建物是环形构建物；

5 (2) 孵育受体宿主细胞，所述第一目的片段和所述第二目的片段发生同源重组，所述同源重组获得具有第一目的片段和第二目的片段的第三核酸构建物，

10 优选地，所述受体宿主细胞表达针对所述双链切割识别位点的切割酶，和/或所述供体宿主细胞不表达针对所述双链切割识别位点的切割酶，和/或所述供体宿主细胞还包含接合转移辅助质粒或其具有引导接合转移的功能片段，例如 pUZ8002，

15 更优选地，所述方法还包括：(0.1) 利用 5' 同源臂和 3' 同源臂将项目 1-3 中任一项所述的多核苷酸与第一目的片段整合，获得包含所述第一核酸构建物的受体宿主细胞，(0.2) 利用 5' 同源臂和 3' 同源臂将项目 1-3 中任一项所述的多核苷酸与第二目的片段整合，获得包含所述第二核酸构建物的供体宿主细胞。

项目 9、如项目 8 所述的方法，其特征在于，所述使宿主细胞发生接合转移的条件包括：供体宿主细胞与受体宿主细胞的总浓度为 10^8 个/mL；和/或供体宿主细胞与受体宿主细胞的比例为 10:1-1:10；和/或培养温度 25-40°C。

20 项目 10、一种用于构建长 DNA 片段或使目的片段整合成较长片段的系统，包含项目 1-3 中任一项所述的多核苷酸、项目 4-6 中任一项所述的核酸构建物、或项目 7 所述的宿主细胞。

在本文中，浓度、含量、百分数和其它数值均可用范围的形式表示。也应理解，使用这种范围形式只是为了方便和简洁，应该被弹性地解读为包括范围上下限所明确提及的数值，还应包括该范围内包括的所有单个数值或子范围。

25 本发明通过参考以下实验实施例进一步详细地进行描述。这些实施例仅出于说明性的目的提供，并不意欲为限制性的，除非另有规定。因此，本发明决不应被解释为限于以下实施例，而是应被解释为包括由于本文提供的教导变得显而易见的任何和全部的变化。实施例中所用的方法和试剂，除非另有说明，否则为本领域常规的方法和试剂。

实施例

实施例 1, 质粒和菌株构建

质粒 pCas9-Amp 的构建: 以质粒 pSP72 为模板聚合酶扩增氨苄青霉素筛选标记, 诱导 Red 重组系统实现原 pCas 质粒筛选标记的替换。

- 5 质粒 gRNA 的构建: 根据原质粒序列, 选择合适的 20 bp 的切割位点。以质粒 ptargetF⁶ 为模板, 设计具有 20 bp 切割位点重叠区域序列引物进行反向聚合酶链式扩增, 将该核酸片段通过化学转化方式转化入菌株 DH10B 中, 构建环型质粒构建 S0-gRNA、S4-gRNA 和 SC-gRNA。若需更改质粒抗性筛选标记为卡那霉素筛选标记, 以 SC-gRNA 为模板扩增除了筛选标记以外的骨架区域,
- 10 以质粒 pCas9 为模板聚合酶链式扩增卡那霉素筛选标记区域, 将 2 个区域 DNA 片段通过 Gibson 等温拼接方法 (Gibson, D. G.等, *Nature Methods*, 2009) 完成质粒 SCK-gRNA 的克隆构建。

表 1, gRNA 识别序列

名称	20 bp 序列	PAM 位点
S0-gRNA 识别位点	aaaaagccccggcggtcatgcc	ggg
S4-gRNA 识别位点	ccgctgacgcgaaccccttg	cgg
SC-gRNA 识别位点	gaaactgccggaaatcgtcg	tgg
SCK-gRNA 识别位点	gaaactgccggaaatcgtcg	tgg

15

- 菌株 MT 的构建: 打靶融合组件是由 3 个片段融合而成。片段 1: 菌株 MDS42 左侧同源臂区域~500 bp; 片段 2: 由质粒 pZJ431 聚合酶链式扩增的 TelN 组件; 片段 3: 菌株 MDS42 右侧同源臂区域~500 bp。电击转化打靶融合组件和质粒 S0-gRNA, 借助于 CRISPR/Cas9 系统在大肠杆菌 MDS42 基因组上
- 20 引入 DNA 双链断裂, 在菌株 MDS42 基因组位置 3274965-3274969 插入 TelN 组件, TelN 序列选用经酵母表达优化后的 N15 噬菌体基因组序列 NC001901 的 CDS_29 (24995-26890 bp, 共 1896 bp)。启动子选择为组成型启动子 J23100 (<http://parts.igem.org/Promoters/Catalog/Constitutive>)。

- 受体菌质粒 p1 构建: 将 3 个不同区域的片段进行融合 PCR, 从左到右的
- 25 3 个片段依次是片段 1: 质粒的左侧同源臂区域~500 bp (HL); 片段 2: *tos* 位点区域 (来源于金斯瑞公司化学合成 519 bp, 噬菌体 N15 位置 24471-24989

bp)；片段 3：质粒的右侧同源臂区域~500 bp (HR)。每个片段之间的重叠区域为 40 bp。用 Phanta Max DNA 聚合酶一步融合得到目标组件，后将融合组件通过电击转化的方式转入大肠杆菌 MDS42 中，借助 CRISPR/Cas9 系统替换原质粒的目标区域，利用添加了氯霉素的 LB 培养基进行筛选，获得环型质粒改造成功的菌株。示意图如图 1 所示。

供体菌质粒 p2 打靶组件构建：将 6 个不同区域的片段进行融合 PCR，从左到右的 6 个片段依次是片段 1：质粒的左侧同源臂区域~500 bp (HL)；片段 2：*tos* 位点区域（来源于金斯瑞公司化学合成 519 bp，噬菌体 N15 位置 24471–24989 bp）；片段 3：*oriT* 区域（来源于质粒 pQX17）；片段 4：壮观霉素筛选标记 *Spc*（来源于质粒 gRNA）；片段 5：*oriC* 复制区（来源于菌株 MDS42 基因组位置 3352912- 3357362）；片段 6：质粒的右侧同源臂区域~500 bp (HR)。每个片段之间的重叠区域为 40 bp。用 Phanta Max DNA 聚合酶一步融合得到目标组件，后将融合组件通过电击转化的方式转入大肠杆菌 MDS42 中，借助 CRISPR/Cas9 系统替换原质粒的目标区域，利用添加了壮观霉素的 LB 培养基进行筛选，获得环型质粒改造成功的菌株。示意图如图 2 所示。

供体菌质粒 p3 打靶组件构建：将 6 个不同区域的片段进行融合 PCR，从左到右的 6 个片段依次是片段 1：质粒的左侧同源臂区域~500 bp (HL)；片段 2：BAC 复制区（来源于 RP11-413L20）；片段 3：*tos* 位点区域（来源于金斯瑞公司化学合成 519 bp，噬菌体 N15 位置 24471–24989 bp）；片段 4：阿泊拉霉素筛选标记 *Apr*（来源于质粒 pXX22）；片段 5：*oriT* 区域（来源于质粒 pQX17）；片段 6：质粒的右侧同源臂区域~500 bp (HR)。每个片段之间的重叠区域为 40 bp。用 Phanta Max DNA 聚合酶一步融合得到目标组件，后将融合组件通过电击转化的方式转入大肠杆菌 MDS42 中，借助 CRISPR/Cas9 系统替换原质粒的目标区域，利用添加了阿泊拉霉素的 LB 培养基进行筛选，获得环型质粒改造成功的菌株。示意图如图 3 所示。

供体菌质粒 p4 打靶组件构建：与质粒 p3 构建过程相似，区别在于将阿泊拉霉素抗性筛选标记 (*Apr*) 更换成壮观霉素抗性筛选标记 (*Spc*)。

供体菌质粒 p5 构建：与环型质粒 p3 构建过程相似，区别在于将抗性筛选标记和 *oriT* 区域进行了位置互换，其它各元件位置不变。

供体菌质粒 p6 构建：与环型质粒 p4 构建过程相似，区别在于将打靶组件的抗性筛选标记和 *oriT* 区域进行了位置互换，其它各元件位置不变。

供体菌质粒 p7 构建：与环型质粒 p3 构建过程相似，区别在于将打靶组件的抗性筛选标记和 *oriT* 区域进行了位置互换，其它各元件位置不变。

- 5 供体菌质粒 p8-1 构建：与环型质粒 p4 构建过程相似。区别在于将打靶组件的抗性筛选标记和 *oriT* 区域进行了位置互换，其它各元件位置不变。另外，元件中额外添加了氨苄青霉素抗性筛选标记。

供体菌质粒 p8-2 构建：与环型质粒 p3 构建过程相似。区别在于将打靶组件的抗性筛选标记和 *oriT* 区域进行了位置互换，其它各元件位置不变。

- 10 供体菌质粒 p8-3 构建：与环型质粒 p4 构建过程相似。区别在于将打靶组件的抗性筛选标记和 *oriT* 区域进行了位置互换，其它各元件位置不变。另外，该质粒是通过 Gibson 等温组装拼接获得。

表 2，原质粒改造情况汇总

原质粒名称	原质粒载体	质粒包含的人免疫球蛋白区域	改造后的质粒名称	所用打靶 gRNA 的名称	改造后的质粒大小
CTD-2572O2	pBeloBAC11	chr14:105815179-106008117	p1	S4-gRNA	200576 bp
RP11-659B19	pBACe3.6	chr14:105975326-106144835	p2	SCK-gRNA	176639 bp
RP11-413L20	pBACe3.6	chr14:106085460-106270391	p3	SC-gRNA	192494 bp
RP11-72N10	pBACe3.6	chr14:106234071-106427650	p4	SCK-gRNA	201291 bp
CTD-3074B5	pBeloBAC11	chr14:106375022-106531377	p5	S4-gRNA	163666 bp
CTD-2195P5	pBeloBAC11	chr14:106509911-106656138	p6	SCK-gRNA	153751 bp
CTD-2366K3	pBeloBAC11	chr14:106621115-106766495	p7	S4-gRNA	152649 bp
CH17-314I7	pBACGK1.1	chr14:106648191-106860132	p8-1	SCK-gRNA	220967 bp
CTD-3087C18	pBeloBAC11	chr14:106849834-106875815	p8-2	S4-gRNA	33331 bp
pUC57-1	pUC57-Brick	chr14:106869845-106879844	p8-3	无	17170 bp

15 实施例 2，目的片段的拼接组装

- 拼接组装方式 1（BAC 复制区与 *oriC* 复制区的线型质粒拼接）：**供体菌株是在只含有环型质粒 p2（*oriC* 复制区）的菌株 MDS42 的基础上，电击转化辅助质粒 pUZ8002 后待用。受体菌株是基因组上携带组成型表达 TeIN 蛋白的 MDS42 菌株，同时受体菌株体内只含有已经线性化的待组装线型质粒 p1（BAC 复制区）。然后将供体菌株的 p2 接合转移进入受体菌株，线性化质粒 p2 可通过 32 kb 的同源区域与线型质粒 p1 直接拼接组装。示意图如图 4 所示。具体的接合转移步骤参照文献（Ma, N. J.等 *Nat Protoc.* 2014）如下操作：
- 20

1. 在 5 mL 的 LB 试管中独立培养供体菌和受体菌,等到菌株生长到指数后期;
2. 按浓度 $OD_{600}=1.0$, 体积 1 mL 来收集相应的细胞量, 约含有大肠杆菌 10^9 个;
3. 13500 g 室温离心 1 min, 去除上清液, 再用 1 mL 新鲜的 LB 培养基洗涤细胞 3 次;
4. 取 100 μ L LB 重悬细胞, 再取 10 μ L 细胞悬液加入 990 μ L LB, 约含有大肠杆菌 10^8 个;
5. 按供受体比例 4:1 混匀细胞, 取 80 μ L 的供体细胞和 20 μ L 受体细胞混匀, 约含有大肠杆菌 10^7 个;
6. 在 LB 培养基平板上滴 2 个 20 μ L 的点和 6 个 10 μ L 的点, 放置在 30-34°C 培养 1 h, 如果转移质粒过大, 可以考虑延长接合转移时间到 2 h;
7. 用 750 μ L LB 洗涤平板上的细胞 2 次, 转移细胞到 1.5 mL EP 管中;
8. 将细胞悬液涂布在添加了氯霉素和壮观霉素的培养基平板上。

待平板上长出可见克隆后, 在供受体比例为 80:10 时, 1/2 体积涂布培养基平板, 统计克隆数目为 12 个。另外在添加了氯霉素和壮观霉素的培养基平板上划线, 准备制作菌株胶块, 通过脉冲场电泳验证线型质粒拼接组装的阳性率。具体操作步骤如下所示:

1. 将每个单克隆接种入 5 mL 液体 LB 培养基, 37°C 过夜培养;
2. 吸取 1 mL 菌液转入 1.5 mL 的 EP 管中, 离心 10000 g, 1 min, 去除上清液 (体积 1 mL 菌量可制作 5 个胶块);
3. 吸取 1 mL 无菌水洗涤细胞 1 次, 10000 g, 1 min;
4. 吸取 1 mL 50 mM EDTA pH8.0 洗涤细胞 1 次, 10000 g, 1 min;
5. 吸取 750 μ L 细胞重悬缓冲液(10 mM Tris·HCl pH7.2)洗涤细胞 1 次, 10000 g, 1 min, 去除上清液;
6. 向细胞沉淀再加入 250 μ L 细胞重悬缓冲液, 取用 100 μ L (因为后面需要制作 2 个胶块), 放置 50 °C 水浴锅内平衡;
7. 另外准备低熔点琼脂糖浓度 2% 的 TE25S (TE25S 配方: 25 mM Tris·HCl pH8.0, 25 mM EDTA pH 8.0, 10.3%蔗糖; 配置方法: 加热溶解时注意放水中加热, 防止爆沸, 后放在 50 °C 水浴锅中待取用);

8. 吸取 100 μ L 的琼脂糖溶液，加入到 100 μ L 预热的细胞悬液中，缓慢吹打均匀（此时琼脂糖胶浓度约为 1%），尽量减少气泡产生，缓慢注入洗净晾干的模具中，放冰箱 4°C 静置 30 min，让胶块凝固；
9. 向 2 mL EP 管中加入现配置好的蛋白酶 K 反应液（每 1 mL 胶块加入 5 mL
5 反应液，蛋白酶 K 反应液配置为：100 mM EDTA pH8.0，0.2% 脱氧胆酸钠盐，1% 十二烷基基肌氨酸钠，1 mg/mL 蛋白酶 K，配置前可以提前用无菌水配置 10 mg/mL 蛋白酶 K 的母液），50 °C 水浴锅中消化胶块 36 h，消化时间可以根据样品不同进行适度调整；
10. 待胶块消化结束，胶块成透明状态，去除消化液，加入 2 mL Wash buffer
10 (50 mM EDTA pH8.0，20 mM Tris·HCl pH8.0) 洗涤胶块 4 次，每次洗涤时间为 30-60 min；
11. 洗涤结束后，上样 1/3 体积胶块进行脉冲场凝胶电泳验证；
12. 用 PFGE 琼脂糖制作浓度为 1% 的凝胶，电泳缓冲液为 0.5xTBE，温度为
12°C，转换时间为 10-60 s，角度为 120°，运行时间为 17 h，电压梯度为 6
15 V/cm。

脉冲场电泳结果分析：受体菌 MT 中线型质粒 p1 大小为 200576 bp，供体菌 MDS42 中环型质粒 p2 大小为 176639 bp。若环型质粒 p2 通过接合转移方式进入受体菌株内，先经 *TelN/tos* 系统线性化后再与线型质粒 p1 拼接组装，若拼接组装成功，理论上会得到大小为 343373 bp 的线型质粒 pA，电泳条带位置
20 符合我们的预期大小，2 个线型质粒拼接成功的阳性率分别为 11/11。结果如图 6 所示。

拼接组装方式 2 (BAC 复制区与 BAC 复制区的线型质粒拼接)：供体菌株是在只含有环型质粒 p4 的菌株 MDS42 的基础上，电击转化辅助质粒 pUZ8002 后待用。受体菌株是基因组上携带组成型表达 *TelN* 蛋白的 MDS42
25 菌株，同时受体菌株体内只含有已经线性化的待组装线型质粒 pB。然后将供体菌株的 p4 接合转移进入受体菌株，线性化质粒 p4 可通过 36 kb 的同源区域与线型质粒 pB 直接拼接组装。示意图如图 5 所示。实验步骤参考“拼接组装方式 1”的步骤。

待平板上长出可见克隆后，在供受体比例为 4:1 时，1/2 体积涂布培养基

平板，统计克隆数目为 212 个。另外在添加了氯霉素和壮观霉素的培养基平板上划线，准备制作菌株胶块，通过脉冲场电泳验证线型质粒拼接组装的阳性率。根据脉冲场电泳结果分析，受体菌 MT 中线型质粒 pB 大小为 464352 bp，供体菌 MDS42 中环型质粒 p4 大小为 201291 bp。若环型质粒 p4 通过接合转移方式

5 进入受体菌株内，先经 *TelN/tos* 系统线性化后再与线型质粒 pB 拼接组装，若拼接组装成功，理论上会得到大小为 621741 bp 的线型质粒 pC，电泳条带位置符合我们的预期大小，2 个线型质粒拼接成功的阳性率分别为 7/7。结果如图 7 所示。

两种拼接组装方式比较：拼接组装方式 1（BAC 复制区与 *oriC* 复制区的线型质粒拼接）是利用复制区稳定性差异实现 2 个 DNA 片段的拼接组装，而

10 拼接组装方式 2（BAC 复制区与 BAC 复制区的线型质粒拼接）是利用相同 BAC 复制区的排斥性压力完成 2 个 DNA 片段的拼接组装。拼接组装方式 1 中 *oriC* 复制区带动的质粒由于不含分配系统元件，所以该质粒是不稳定的。其宿主在不添加抗生素的培养基中培养传代过程中容易造成质粒丢失，且不易抽提检测

15 菌株内环型质粒的大小。另外，*oriC* 复制区带动的质粒接合组装效率较 BAC 复制区带动的质粒接合组装效率低，所以接下来的 DNA 大片段连续克隆拼接采用拼接组装方式 2 进行。

实施例 3，大片段的连续拼接组装

20 **人免疫球蛋白重链区域的连续拼接组装：**从赛默飞公司购买 BAC 克隆文库 CTD-2572O2、RP11-659B19、RP11-413L20、RP11-72N10、CTD-3074B5、CTD-2195P5、CTD-2366K3 和 CTD-3087C18；从睿铂赛公司购买 BAC 克隆文库 CH17-314I7；从金斯瑞公司人工合成质粒 pUC57-1。借助 CRISPR/Cas9 系统替换原质粒的目标区域，将 BAC 克隆文库各质粒依次在菌株 MDS42 中改造

25 成环型质粒 p1、p2、p3、p4、p5、p6、p6、p7、p8-2、p8-1 和 p8-3。通过连续的接合转移拼接组装方法依次获得 DNA 大片段组装成功的线型质粒 pA、pB、pC、pD、pE、pF、p8 和 pG，其中 p8 是由质粒 p8-1、p8-2 和 p8-3 连续拼接组装得到，pG 是由 pF 和 p8 拼接组转得到。

表 3 连续拼接组装数据统计

受体菌质粒	供体菌质粒	重叠区域	接合子数目	拼接组装	拼接组装
-------	-------	------	-------	------	------

名称	名称	大小	(1/2 体积)	质粒大小	阳性率
p1 (200576 bp)	p2 (176639 bp)	32792 bp	12 个 (80:10)	pA (343373 bp)	11/11
pA	p3 (192494 bp)	59376 bp	230 个 (4:1)	pB (464352 bp)	8/8
pB	p4 (201291 bp)	36321 bp	212 个 (4:1)	pC (621741 bp)	7/7
pC	p5 (163666 bp)	52629 bp	130 个 (40:10)	pD (725323 bp)	5/6
pD	p6 (153751 bp)	21467 bp	91 个 (40:10)	pE (850112 bp)	8/8
pE	p7 (152649 bp)	35024 bp	118 个 (40:10)	pF (960384 bp)	3/5
p8-1 (220967 bp)	p8-2 (33331 bp)	10299 bp	54 个 (80:20)	p8-A (236366 bp)	3/3
p8-A	p8-3 (17170 bp)	5971 bp	30 个 (80:20)	p8 (240271 bp)	9/9
pF	p8 (240271 bp)	118305 bp	450 个 (80:20)	pG (1073650 bp)	3/5

线型质粒的连续拼接组装：供体菌株是在只含有环型质粒的菌株 MDS42 的基础上，电击转化辅助质粒 pUZ8002 后待用。受体菌株是基因组上携带组成型表达 TeIN 蛋白的 MDS42 菌株，同时受体菌株体内只含有已经线性化的待组装线型质粒。然后将供体菌株的环型质粒接合转移进入受体菌株，线性化质粒可通过同源区域与受体菌线型质粒直接拼接组装获得。

待平板上长出可见克隆后，准备制作菌株胶块，通过两种不同的脉冲场电泳条件验证线型质粒拼接组装的阳性率。根据脉冲场电泳结果分析，质粒 p1 和 p2 拼接得到大小为 343373 bp 的线型质粒 pA，这 2 个线型质粒拼接成功的阳性率分别为 11/11；质粒 pA 和 p3 拼接得到大小为 464352 bp 的线型质粒 pB，拼接成功的阳性率分别为 8/8；质粒 pB 和 p4 拼接得到大小为 621741 bp 的线型质粒 pC，2 个线型质粒拼接成功的阳性率分别为 7/7；质粒 pC 和 p5 拼接得到大小为 725323 bp 的线型质粒 pD，2 个线型质粒拼接成功的阳性率分别为 5/6；质粒 pD 和 p6 拼接得到大小为 850112 bp 的线型质粒 pE，2 个线型质粒拼接成功的阳性率分别为 8/8；质粒 pE 和 p7 拼接得到大小为 960384 bp 的线型质粒 pF，2 个线型质粒拼接成功的阳性率分别为 3/5；质粒 pF 和 p8 拼接得到大小为 1073650 bp 的线型质粒 pG，2 个线型质粒拼接成功的阳性率分别为 3/5。结果如图 8A 和 8B 所示。

权 利 要 求 书

- 1、一种多核苷酸，用于构建长 DNA 片段，所述多核苷酸包含：双链切割识别位点，转移起始位点 oriT，复制起点，双链切割识别位点或其互补序列
5 能被切割酶切割，
优选地，所述多核苷酸还包含用于整合目的片段的 5'同源臂和 3'同源臂，所述目的片段是所述长 DNA 片段的一部分；更优选地，所述双链切割识别位点、转移起始位点 oriT 和复制起点位于 5'同源臂和 3'同源臂之间。
- 10 2、如项目 1 所述的多核苷酸，其特征在于，所述双链切割识别位点是 tos 位点，所述切割酶是 TelN，
优选地，
tos 位点来源于噬菌体 N15，和/或
转移起始位点 oriT 来源于肠杆菌科细菌，优选埃希氏菌属细菌，和/或
15 复制起点来源于肠杆菌科细菌，优选埃希氏菌属细菌，和/或
所述多核苷酸还包含标记基因，和/或
转移起始位点 oriT 和复制起点的位置可互换；
更优选地，
tos 位点序列包含 SEQ ID NO:1 所示的序列或其互补序列，和/或
20 转移起始位点 oriT 包含 SEQ ID NO:2 所示的序列，和/或
复制起点包含 SEQ ID NO:3 或 4 所示的序列，和/或
标记基因位于 5'同源臂和 3'同源臂之间，和/或
标记基因位于转移起始位点 oriT 和复制起点之间，和/或
标记基因位于复制起点的 3'端，和/或
25 双链切割识别位点位于复制起点的 5'端。
- 3、如项目 1 所述的多核苷酸，其特征在于，所述多核苷酸依次包含：
5'同源臂，复制起点，转移起始位点 oriT，双链切割识别位点，3'同源臂；
5'同源臂，转移起始位点 oriT，复制起点，双链切割识别位点，3'同源臂；

5'同源臂，转移起始位点 oriT，双链切割识别位点，复制起点，3'同源臂；
 5'同源臂，复制起点，双链切割识别位点，转移起始位点 oriT，3'同源臂；
 5'同源臂，双链切割识别位点，转移起始位点 oriT，复制起点，3'同源臂；

或

5 5'同源臂，双链切割识别位点，复制起点，转移起始位点 oriT，3'同源臂。

4、一种核酸构建物，包含项目 1-3 中任一项所述的多核苷酸和目的片段，
 用于构建长 DNA 片段，

优选地，

10 所述目的片段是长 DNA 片段的一部分，和/或

所述长 DNA 片段的长度至少 2kb，和/或

所述目的片段的长度为至少 1kb，和/或

所述核酸构建物是环形构建物，或所述核酸构建物是线性构建物并且其中
 双链切割识别位点或其互补序列被切割酶切割，

15

5、如项目 4 所述的核酸构建物，其特征在于，所述核酸构建物包含：（1）
 第一核酸构建物，其具有项目 1-3 中任一项所述的多核苷酸和第一目的片段，

（2）第二核酸构建物，其具有项目 1-3 中任一项所述的多核苷酸和第二目的
 片段，其中第一目的片段的 3'端与第二目的片段的 5'端具有 1kb-200kb 的重叠
 20 区域，

优选地，第一核酸构建物是线性构建物，其中双链切割识别位点或其互补
 序列被切割酶切割；第二核酸构建物是环形构建物。

6、如项目 4 所述的核酸构建物，其特征在于，第一核酸构建物与第二核
 25 酸构建物分别具有标记基因，其中，

第一核酸构建物具有一种或两种标记基因，且满足以下条件：（1）至少
 一种标记基因位于第一目的片段的 5'端，并且任选地（2）双链切割识别位点
 位于所述至少一种标记基因的 5'端；优选地，两种标记基因不同；更优选地，
 两种标记基因分别位于第一目的片段的 5'端和 3'端，

第二核酸构建物的标记基因位于第二目的片段的 3'端,并且双链切割识别位点位于所述标记基因的 3'端,

优选地,第一核酸构建物与第二核酸构建物中的标记基因均不相同。

5 7、一种宿主细胞,包含项目 4-6 中任一项所述的核酸构建物,

优选地,所述核酸构建物是环形构建物,并且所述宿主细胞不表达针对所述双链切割识别位点的切割酶,所述宿主细胞还任选包含接合转移辅助质粒或其具有引导接合转移的功能片段,例如 pUZ8002;或者,所述核酸构建物是线性构建物,并且所述宿主细胞表达针对所述双链切割识别位点的切割酶,

10 更优选地,所述宿主细胞是肠杆菌科细菌。

8、一种构建长 DNA 片段或使目的片段整合成较长片段的方法,包括:

(1) 在使宿主细胞发生接合转移的条件下使受体宿主细胞和供体宿主细胞接触,其中,

15 所述受体宿主细胞包含项目 5 或 6 中所述的第一核酸构建物,所述核酸构建物是线性构建物,其中双链切割识别位点或其互补序列被切割酶切割,

所述供体宿主细胞包含项目 5 或 6 中所述的第二核酸构建物,所述核酸构建物是环形构建物;

20 (2) 孵育受体宿主细胞,所述第一目的片段和所述第二目的片段发生同源重组,所述同源重组获得具有第一目的片段和第二目的片段的第三核酸构建物,

优选地,

所述受体宿主细胞表达针对所述双链切割识别位点的切割酶,和/或

所述供体宿主细胞不表达针对所述双链切割识别位点的切割酶,和/或

25 所述供体宿主细胞还包含接合转移辅助质粒或其具有引导接合转移的功能片段,例如 pUZ8002,

更优选地,所述方法还包括

(0.1) 利用 5' 同源臂和 3' 同源臂将项目 1-3 中任一项所述的多核苷酸与第一目的片段整合,获得包含所述第一核酸构建物的受体宿主细胞,

(0.2) 利用 5' 同源臂和 3' 同源臂将项目 1-3 中任一项所述的多核苷酸与第二目的片段整合，获得包含所述第二核酸构建物的供体宿主细胞。

- 9、如项目 8 所述的方法，其特征在于，所述使宿主细胞发生接合转移的条件包括：
- 供体宿主细胞与受体宿主细胞的总浓度为 10^8 个/mL；和/或
供体宿主细胞与受体宿主细胞的比例为 10:1-1:10；和/或
培养温度 25-40℃。
- 10 10、一种用于构建长 DNA 片段或使目的片段整合成较长片的系统，包含项目 1-3 中任一项所述的多核苷酸、项目 4-6 中任一项所述的核酸构建物、或项目 7 所述的宿主细胞。

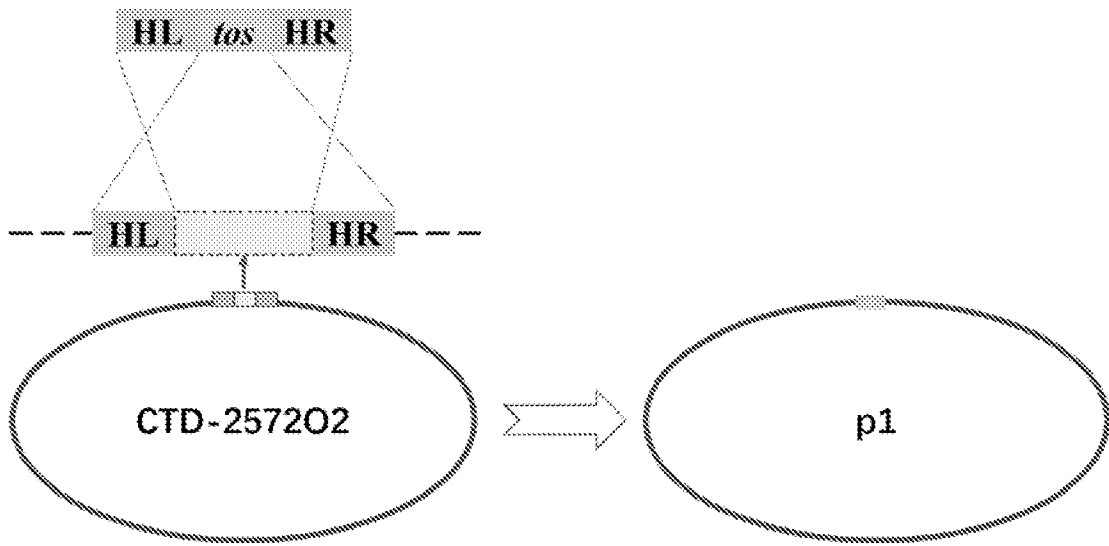


图 1

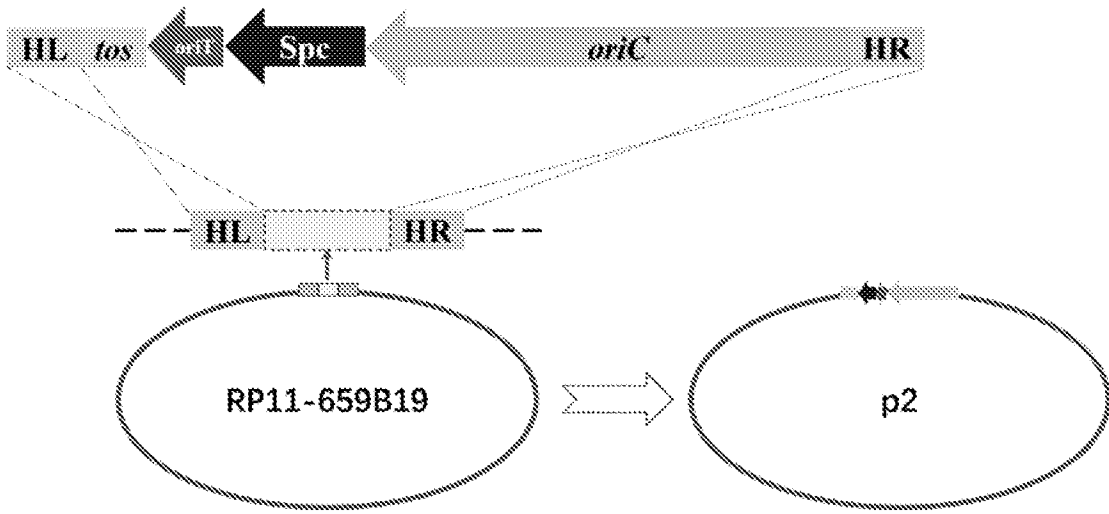


图 2

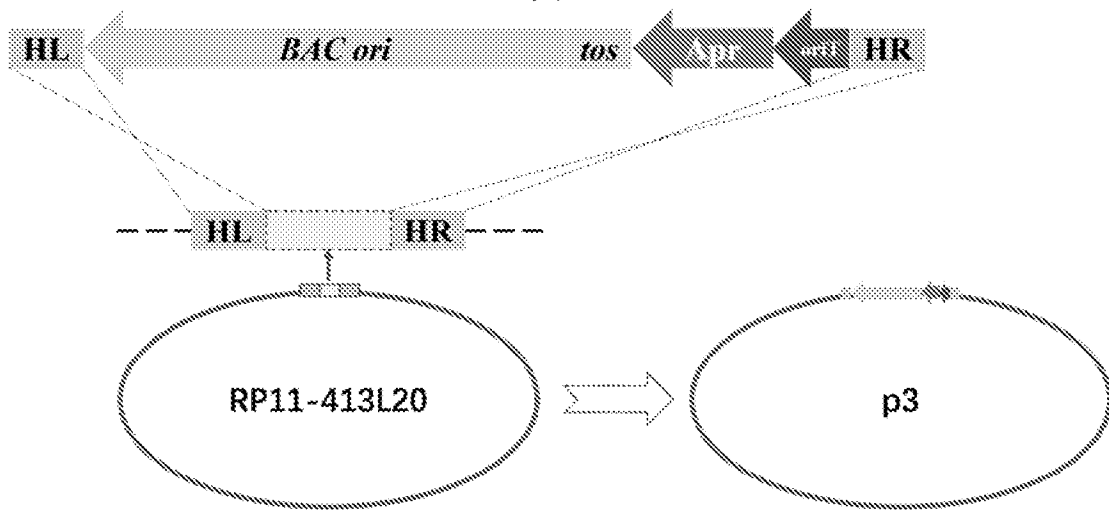


图 3

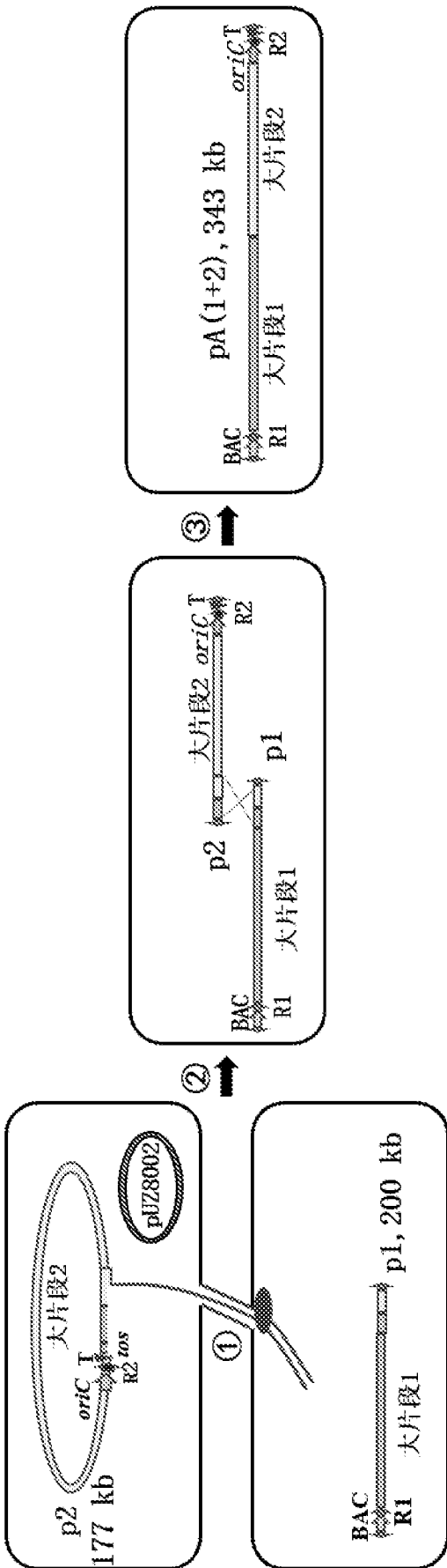


图 4

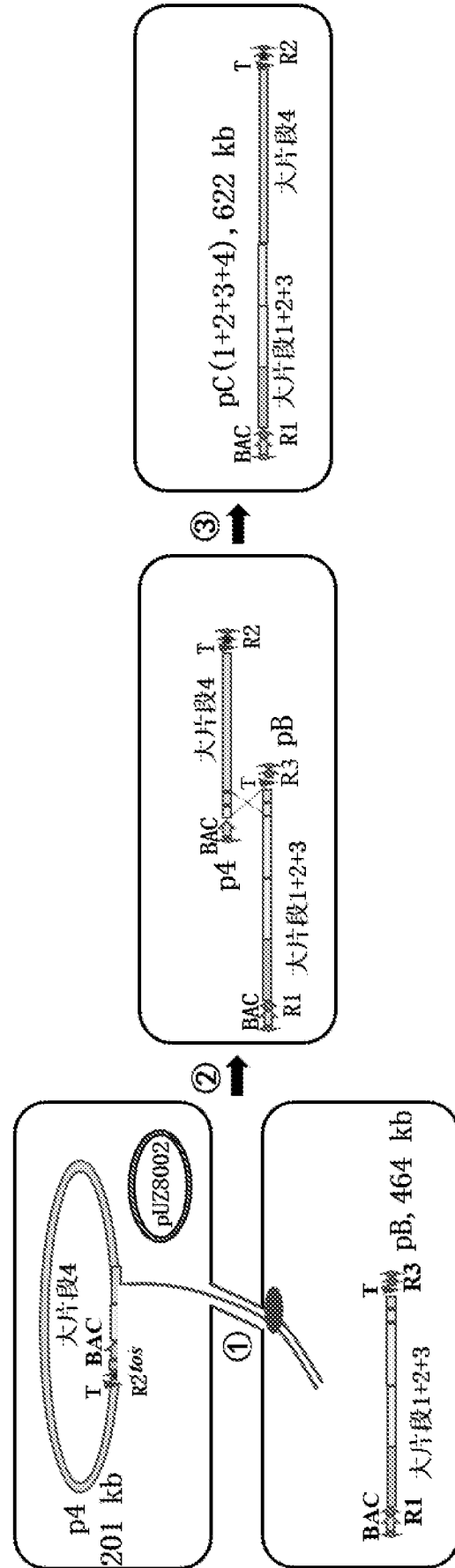


图 5

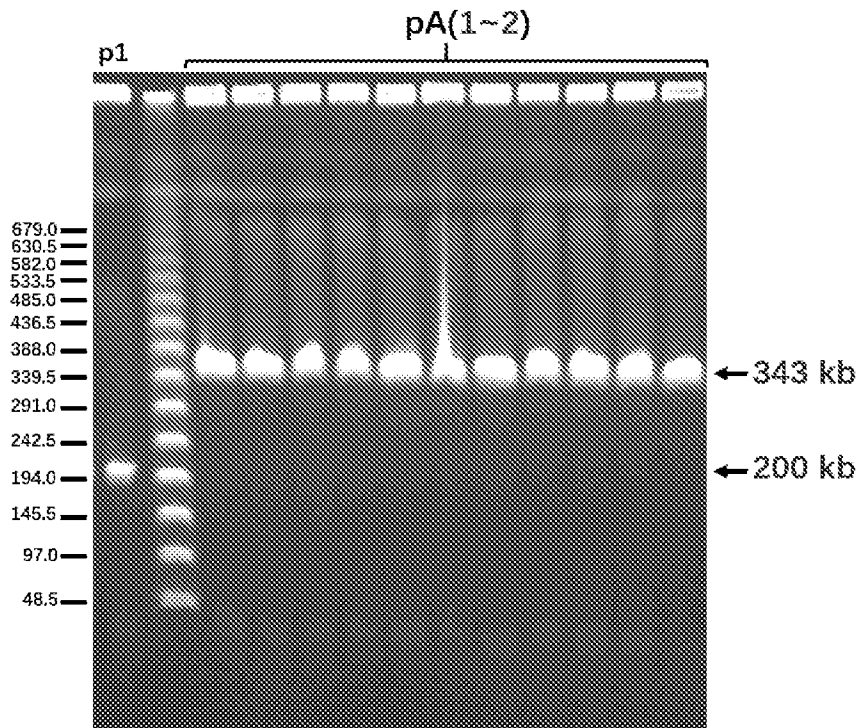


图 6

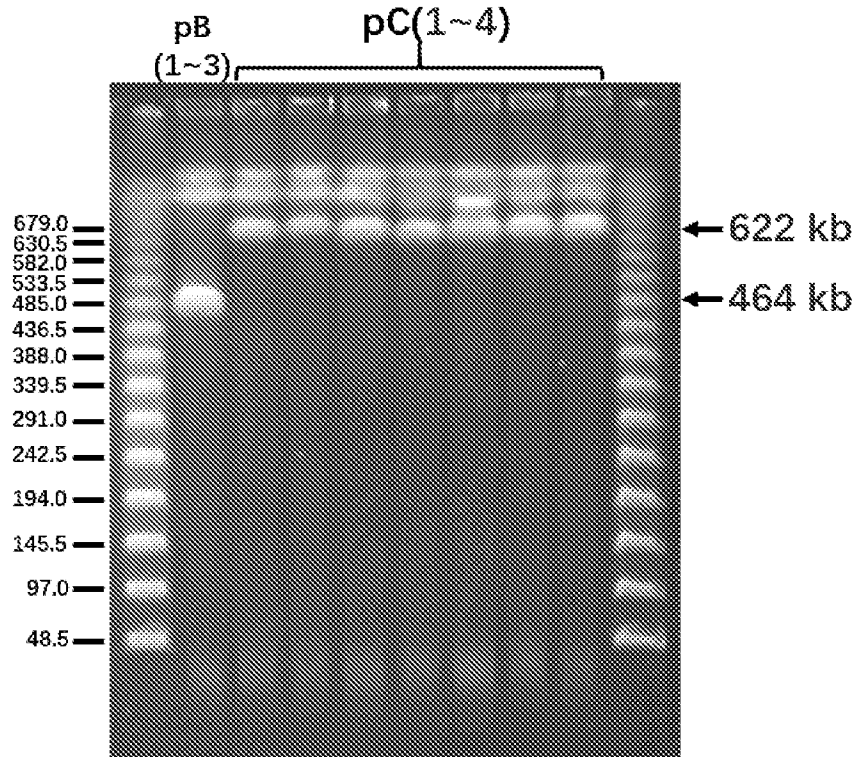


图 7

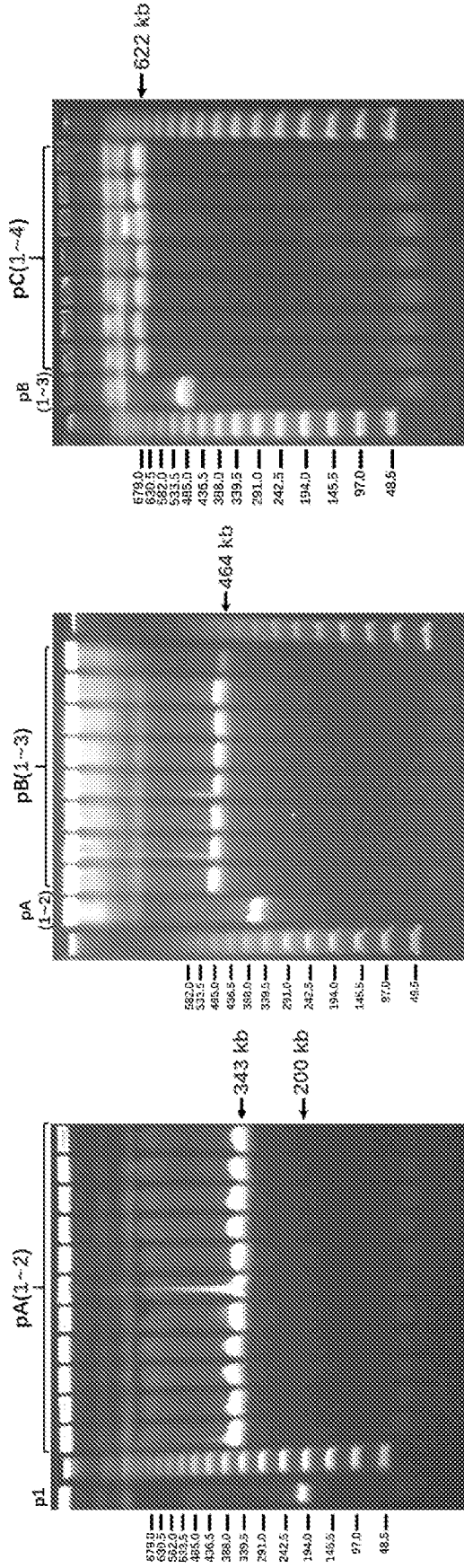


图 8A

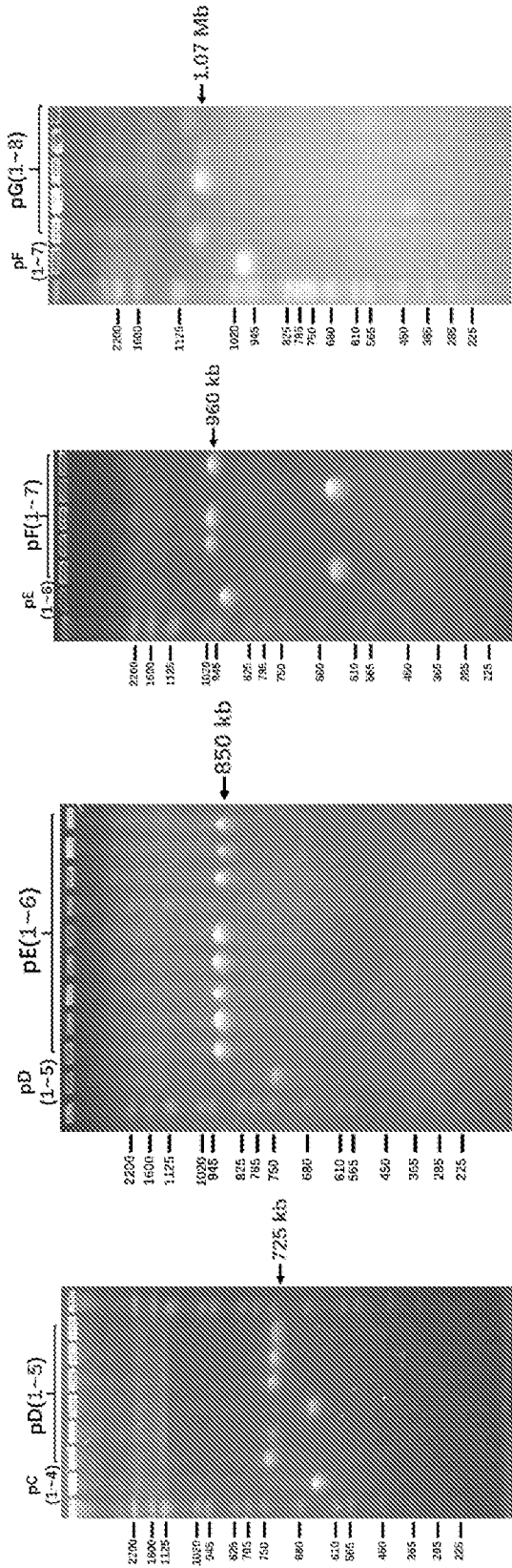


图 8B

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/072437

A. CLASSIFICATION OF SUBJECT MATTER		
C12N 15/70(2006.01)i; C12N 15/66(2006.01)i		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols)		
C12N		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
CNABS, SIPOABS, DWPI, CNKI, baidu, patentics, NCBI, 中国科学院分子植物科学卓越创新中心, 覃重军, 鲁宁, 薛小莉, 钟莉, TelN/tos系统, CRISPR/Cas9, TelN组件, 端粒酶TelN, tos位点, T4SS, 大肠杆菌IV行分泌系统, oriT, MDS42菌株, pUZ8002质粒, 同源臂, 双链切割识别位点, 切割酶, SEQ ID NO: 1-6.		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	LIEW, P.S. et al. "Phage N15 protelomerase resolves its tos recognition site into hairpin telomeres within mammalian cells." <i>Analytical Biochemistry.</i> , Vol. 583, 12 July 2019 (2019-07-12), chapter 113361, pages 1-10	1-10
Y	CN 101016551 A (NANJING NORMAL UNIVERSITY) 15 August 2007 (2007-08-15) claim 1, description figures 1-2	1-10
Y	VOSTROV, A.A. et al. "Bacteriophage N15, complete genome, GenBank: AF064539.1, 46375bp DNA linear." <i>NCBI GenBank.</i> , 25 July 2016 (2016-07-25), pp. 1-21	1-10
Y	NA, G. et al. "Cloning vector pBAC-DE, complete sequence, GenBank: HQ670404.1, 8821bp DNA circular." <i>NCBI GenBank.</i> , 06 May 2013 (2013-05-06), pp. 1-3	1-10
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
<p>* Special categories of cited documents:</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier application or patent but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>"&" document member of the same patent family</p>		
Date of the actual completion of the international search		Date of mailing of the international search report
17 March 2022		14 April 2022
Name and mailing address of the ISA/CN		Authorized officer
China National Intellectual Property Administration (ISA/CN) No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088, China		
Facsimile No. (86-10)62019451		Telephone No.

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	DILA, D. et al. "Escherichia coli strain NEB_Turbo chromosome, complete genome, GenBank: CP053605.1, 4527032bp DNA circular." <i>NCBI GenBank.</i> , 26 May 2020 (2020-05-26), pp. 1-42	1-10
A	DENEKE, J. et al. "Phage N15 Telomere Resolution." <i>The Journal of Biological Chemistry.</i> , Vol. 277, No. 12, 22 March 2002 (2002-03-22), pp. 10410-10419	1-10
A	JAIN, M. et al. "Nanopore sequencing and assembly of a human genome with ultra-long reads." <i>nature biotechnology.</i> , 29 January 2018 (2018-01-29), pp. 1-13	1-10
A	NA, G. et al. "Cloning vector pBAC-FC, complete sequence, GenBank: HQ670403.1, 9260bp DNA circular." <i>NCBI GenBank.</i> , 06 May 2013 (2013-05-06), pp. 1-3	1-10
A	CN 102876702 A (THE INSTITUTE OF MICROBIOLOGY, CHINESE ACADEMY OF SCIENCES) 16 January 2013 (2013-01-16) entire document	1-10

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
 - a. forming part of the international application as filed:
 - in the form of an Annex C/ST.25 text file.
 - on paper or in the form of an image file.
 - b. furnished together with the international application under PCT Rule 13ter.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
 - c. furnished subsequent to the international filing date for the purposes of international search only:
 - in the form of an Annex C/ST.25 text file (Rule 13ter.1(a)).
 - on paper or in the form of an image file (Rule 13ter.1(b) and Administrative Instructions, Section 713).
2. In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
3. Additional comments:

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2022/072437

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	101016551	A	15 August 2007	CN	101016551	B	19 May 2010
CN	102876702	A	16 January 2013	CN	102876702	B	20 August 2014

<p>A. 主题的分类</p> <p>C12N 15/70(2006.01)i; C12N 15/66(2006.01)i</p> <p>按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类</p>																	
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>C12N</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))</p> <p>CNABS, SIPOABS, DWPI, CNKI, baidu, patentics, NCBI, 中国科学院分子植物科学卓越创新中心, 覃重军, 鲁宁, 薛小莉, 钟莉, Te1N/tos系统, CRISPR/Cas9, Te1N组件, 端粒酶Te1N, tos位点, T4SS, 大肠杆菌IV行分泌系统, oriT, MDS42菌株, pUZ8002质粒, 同源臂, 双链切割识别位点, 切割酶, SEQ ID NO:1-6.</p>																	
<p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>Y</td> <td>LIEW, P.S.等. "Phage N15 protelomerase resolves its tos recognition site into hairpin telomeres within mammalian cells." Analytical Biochemistry., 第583卷, 2019年7月12日 (2019 - 07 - 12), 第113361篇, 第1-10页</td> <td>1-10</td> </tr> <tr> <td>Y</td> <td>CN 101016551 A (南京师范大学) 2007年8月15日 (2007 - 08 - 15) 权利要求1, 说明书图1-2</td> <td>1-10</td> </tr> <tr> <td>Y</td> <td>VOSTROV, A.A.等. "Bacteriophage N15, complete genome, GenBank: AF064539.1, 46375bp DNA linear." NCBI genbank., 2016年7月25日 (2016 - 07 - 25), 第1-21页</td> <td>1-10</td> </tr> <tr> <td>Y</td> <td>NA, G.等. "Cloning vector pBAC-DE, complete sequence, GenBank: HQ670404.1, 8821bp DNA circular." NCBI genbank., 2013年5月6日 (2013 - 05 - 06), 第1-3页</td> <td>1-10</td> </tr> </tbody> </table>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	Y	LIEW, P.S.等. "Phage N15 protelomerase resolves its tos recognition site into hairpin telomeres within mammalian cells." Analytical Biochemistry., 第583卷, 2019年7月12日 (2019 - 07 - 12), 第113361篇, 第1-10页	1-10	Y	CN 101016551 A (南京师范大学) 2007年8月15日 (2007 - 08 - 15) 权利要求1, 说明书图1-2	1-10	Y	VOSTROV, A.A.等. "Bacteriophage N15, complete genome, GenBank: AF064539.1, 46375bp DNA linear." NCBI genbank., 2016年7月25日 (2016 - 07 - 25), 第1-21页	1-10	Y	NA, G.等. "Cloning vector pBAC-DE, complete sequence, GenBank: HQ670404.1, 8821bp DNA circular." NCBI genbank., 2013年5月6日 (2013 - 05 - 06), 第1-3页	1-10
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求															
Y	LIEW, P.S.等. "Phage N15 protelomerase resolves its tos recognition site into hairpin telomeres within mammalian cells." Analytical Biochemistry., 第583卷, 2019年7月12日 (2019 - 07 - 12), 第113361篇, 第1-10页	1-10															
Y	CN 101016551 A (南京师范大学) 2007年8月15日 (2007 - 08 - 15) 权利要求1, 说明书图1-2	1-10															
Y	VOSTROV, A.A.等. "Bacteriophage N15, complete genome, GenBank: AF064539.1, 46375bp DNA linear." NCBI genbank., 2016年7月25日 (2016 - 07 - 25), 第1-21页	1-10															
Y	NA, G.等. "Cloning vector pBAC-DE, complete sequence, GenBank: HQ670404.1, 8821bp DNA circular." NCBI genbank., 2013年5月6日 (2013 - 05 - 06), 第1-3页	1-10															
<p><input checked="" type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。</p>																	
<p>* 引用文件的具体类型:</p> <p>"A" 认为不特别相关的表示了现有技术一般状态的文件</p> <p>"E" 在国际申请日的当天或之后公布的在先申请或专利</p> <p>"L" 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)</p> <p>"O" 涉及口头公开、使用、展览或其他方式公开的文件</p> <p>"P" 公布日先于国际申请日但迟于所要求的优先权日的文件</p> <p>"T" 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件</p> <p>"X" 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性</p> <p>"Y" 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性</p> <p>"&" 同族专利的文件</p>																	
<p>国际检索实际完成的日期</p> <p>2022年3月17日</p>		<p>国际检索报告邮寄日期</p> <p>2022年4月14日</p>															
<p>ISA/CN的名称和邮寄地址</p> <p>中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088</p> <p>传真号 (86-10)62019451</p>		<p>授权官员</p> <p>朱宁</p> <p>电话号码 86-(10)-53961956</p>															

C. 相关文件		
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求
Y	DILA, D. 等. "Escherichia coli strain NEB_Turbo chromosome, complete genome, GenBank: CP053605.1, 4527032bp DNA circular." NCBI genbank., 2020年5月26日 (2020 - 05 - 26), 第1-42页	1-10
A	DENEKE, J. 等. "Phage N15 Telomere Resolution." THE JOURNAL OF BIOLOGICAL CHEMISTRY., 第277卷, 第12期, 2002年3月22日 (2002 - 03 - 22), 第10410-10419页	1-10
A	JAIN, M. 等. "Nanopore sequencing and assembly of a human genome with ultra-long reads." nature biotechnology., 2018年1月29日 (2018 - 01 - 29), 第1-13页	1-10
A	NA, G. 等. "Cloning vector pBAC-FC, complete sequence, GenBank: HQ670403.1, 9260bp DNA circular." NCBI genbank., 2013年5月6日 (2013 - 05 - 06), 第1-3页	1-10
A	CN 102876702 A (中国科学院微生物研究所) 2013年1月16日 (2013 - 01 - 16) 全文	1-10

第1栏 核苷酸和/或氨基酸序列(续第1页第1. c项)

1. 关于国际申请中所公开的任何核苷酸和/或氨基酸序列,国际检索是基于下列序列列表进行的:
- a. 作为国际申请的一部分提交的:
- 附件C/ST. 25文本文件形式
 - 纸件或图形文件形式
- b. 根据细则13之三. 1(a) 仅为国际检索目的以附件C/ST. 25文本文件形式与国际申请同时提交的:
- c. 仅为国际检索目的在国际申请日之后提交的:
- 附件C/ST. 25文本文件形式(细则13之三. 1(a))
 - 纸件或图形文件形式(细则13之三. 1(b)和行政规程第713段)
2. 另外,在提交/提供了多个版本或副本的序列列表的情况下,提供了关于随后提交的或附加的副本中的信息与申请时提交的作为申请一部分的序列列表的信息相同或未超出申请时提交的申请中的信息范围(如适用)的所需声明。
3. 补充意见:

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2022/072437

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	101016551	A	2007年8月15日	CN	101016551	B	2010年5月19日
CN	102876702	A	2013年1月16日	CN	102876702	B	2014年8月20日