

(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



(10) 国际公布号
WO 2019/153902 A1

(43) 国际公布日
2019 年 8 月 15 日 (15.08.2019)

(51) 国际专利分类号:
C12N 15/82 (2006.01) *A01H 5/00* (2018.01)

(21) 国际申请号: PCT/CN2018/122014

(22) 国际申请日: 2018 年 12 月 19 日 (19.12.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201810142052.0 2018年2月11日 (11.02.2018) CN

(71) 申请人: 中国科学院上海生命
科学研究院 (SHANGHAI INSTITUTES FOR
BIOLOGICAL SCIENCES, CHINESE ACADEMY OF
SCIENCES) [CN/CN]; 中国上海市徐汇区岳阳
路319号, Shanghai 200031 (CN)。

(72) 发明人: 朱健康 (ZHU, Jiankang); 中国上海市
徐汇区岳阳路 320 号, Shanghai 200031 (CN)。
华凯 (HUA, Kai); 中国上海市徐汇区岳阳路
320号, Shanghai 200031 (CN)。

(74) 代理人: 上海一平知识产权代理有限公司
(XU & PARTNERS, LLC.); 中国上海市普陀
区真北路 958 号天地科技广场 1 号楼 106
室, Shanghai 200333 (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,
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,
SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, 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: PLANT GENOME SITE-DIRECTED SUBSTITUTION METHOD

(54) 发明名称: 植物基因组定点替换的方法

(57) Abstract: The present invention provides a plant genome site-directed substitution method. Specifically, the present invention relates to a nucleic acid construct. The present invention relates to using the nucleic acid construct of a specific structure, and successfully achieves sgRNA-directed base site-directed mutagenesis (such as a mutation from T to C or from A to G) in a plant for the first time.

(57) 摘要: 本发明提供了一种植物基因组定点替换的方法, 具体地, 本发明涉及一种核酸构建物, 本发明采用特定结构的核酸构建物, 首次在植物中成功实现了sgRNA引导的碱基定点突变(如T突变为C或A突变为G)。



WO 2019/153902 A1

植物基因组定点替换的方法

技术领域

本发明涉及生物技术领域，具体地，涉及植物基因组定点替换的方法。

5

背景技术

在真核生物中，用基因组进行单碱基分辨率的精确修饰主要是通过同源重组完成。然而，同源重组的固有的低效率和外源供体 DNA 模板的依赖性极大地限制了其在许多物种中的使用 (Komor et al., 2017a)。目前由 APOBEC/AID 基因家族介导的 C-T 的碱基替换虽然高效，但是突变的形式比较单一，而且突变产物中往往会伴随的 InDel 的产生。由于插入或缺失突变常常会导致移码突变，因此造成基因编辑的风险上升。

10

因此，本领域迫切需要开发一种在植物细胞中可高效且精确地实现 T-C (或者 A-G) 转化同时显著降低插入或缺失突变风险的方法。

15

发明内容

本发明的目的在于提供一种在植物细胞中可高效且精确地实现 T-C 转化的方法。

20

本发明第一方面提供了一种核酸构建物，所述核酸构建物具有 5' -3' (5' 至 3') 的式 I 结构：

$$I1-Z1-Z2-I2 \quad (I)$$

式中，

25

I1 为第一整合元件；

I2 为第二整合元件；

Z1 为第一表达盒；

Z2 为第二表达盒；

并且，Z1 和 Z2 中的一个表达盒具有 Ia 结构，而另一个表达盒具有式 Ib 结构：

30

$$P1-S1-X1-L1-X2-L2-X3 \quad (Ia)$$

$$P2-Y1 \quad (Ib);$$

式中，

P1、S1、X1、L1、X2、L2、X3、P2、Y1 分别为用于构成所述构建物的元件；

P1 为第一启动子，所述第一启动子包括泛素启动子；

S1 为无或信号肽的编码序列；

5 X1 为腺嘌呤脱氨酶(如野生型和/或突变型 TadA)的编码序列；

L1 为无或第一连接肽的编码序列；

X2 为 Cas9 核酸酶的编码序列，所述的 Cas9 核酸酶是无切割活性或单链切割活性的；

L2 为无或第二连接肽的编码序列；

10 X3 为核定位信号的编码序列，所述核定位信号为 VirD2；

P2 为第二启动子；

Y1 为 sgRNA 的编码序列；

并且，各“-”为键或核苷酸连接序列。

在另一优选例中，所述泛素启动子包括玉米泛素启动子。

15 在另一优选例中，所述泛素启动子包括 UBI 启动子。

在另一优选例中，所述第二启动子包括 U6 启动子。

在另一优选例中，所述的“无切割活性或单链切割活性”指 Cas9 核酸酶对于靶位点 T 所在的单链无切割活性。

20 在另一优选例中，本发明的上述核苷酸元件是按阅读框(in-frame)连接的，从而表达氨基酸序列正确的融合蛋白。

在另一优选例中，所述的构建物具有式 IIa 或式 IIb 结构：

II-P1-S1-X1-L1-X2-L2-X3-P2-Y1-I2 (IIa)；

II-P2-Y1-P1-S1-X1-L1-X2-L2-X3-I2 (IIb)；

式中，各元件的定义如上所述。

25 在另一优选例中，所述的 sgRNA 的第 5 位-第 10 位对应于预定发生 T→C 定点突变的位置(即为 T)。

在另一优选例中，所述的 sgRNA 的第 6-14 位对应于预定发生 T→C 定点突变的位置(即为 T)。

30 在另一优选例中，所述的 sgRNA 的第 12 位、第 13 位和/或第 14 位对应于预定发生 T→C 定点突变的位置(即为 T)。

在另一优选例中，所述的 L1 和 L2 的序列长度各自独立地为 3-120nt，较佳地，

3-96nt, 并且优选为 3 的倍数。

在另一优选例中, 所述的核苷酸连接序列长度为 1-300nt, 较佳地 1-100nt。

在另一优选例中, 所述的第一表达盒和第二表达盒均具有终止子。

在另一优选例中, 所述第一整合元件包括 5' 同源臂序列。

5 在另一优选例中, 所述第一整合元件为 RB 序列。

在另一优选例中, 所述 RB 序列如 SEQ ID NO. :1 (TAAACGCTCTTTTCTCTTAGGTTTAC) 所示。

在另一优选例中, 所述信号肽包括 VirD2 的核定位信号肽。

在另一优选例中, 所述 Cas9 核酸酶选自下组: Cas9、Cas9n、或其组合。

10 在另一优选例中, 所述 Cas9 核酸酶为突变的 Cas9 核酸酶。

在另一优选例中, 所述 X2 元件中, 突变位点在 Cas9 核酸酶 (SEQ ID NO. :2) 的 D10A 位。

```
MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR
YTRRNKRICYLQEIFSNEMAKVDDSFHRLVESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKL
15 VDSTDKADRLRIYLALAHMIKFRGHFLIEGDLNPDNSDVKLFIQLVQTYNQLFEENPINASGVDAKAILS
ARLSKSRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDQYDDDLNLLAQIGD
QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQQLPEKYKEIFFDQS
KNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ
EDFYFPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASQSFIERMTN
20 FDKNLPNEKVLPHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDY
FKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTY
AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMLIHDDSLTFKEDIQK
AQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERM
KRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSI
25 DNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVE
TRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG
TALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIET
NGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDS
PTVAYSVLVVAKEVGKSKKLKSVKELLGITIMERSSEKNPIDFLEAKGYKEVKKDLIIKLPKYSLEFELE
30 NGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIIEQISEFSK
RVILADANLDKVL SAYNKHDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIIH
```

QSITGLYETRIDLSQLGGD (SEQ ID NO. :2)

在另一优选例中,所述 X2 元件中,Cas9 核酸酶的氨基酸序列如 SEQ ID NO. :69 所示。

在另一优选例中,所述 X2 元件的来源选自下组:酿脓链球菌(*Streptococcus*
5 *pyogenes*)、葡萄球菌(*Staphylococcus aureus*)、或其组合。

在另一优选例中,所述连接序列包括 XTEN。

在另一优选例中,所述连接序列如 SEQ ID
NO. :3(TCTGGAGGGTCCTCCGGCGGATCGTCCGGCAGCGAGACGCCAGGCACCTCCGAGAGCGCTACGCC
TGAATCCTCCGGGGGATCTTCAGGAGGATCA)所示。

10 在另一优选例中,所述腺嘌呤脱氨酶包括 TadA。

在另一优选例中,所述腺嘌呤脱氨酶包括野生型和突变型。

在另一优选例中,所述腺嘌呤脱氨酶的突变型包括 TadA7-10。

在另一优选例中,所述腺嘌呤脱氨酶为串联型腺嘌呤脱氨酶,所述串联型腺
嘌呤脱氨酶结构如式II所示:

15 Z8-L8-Z9 (II)

其中,

Z8为野生型的腺嘌呤脱氨酶TadA的氨基酸序列;

L8为任意的连接肽序列;

Z9为突变型的腺嘌呤脱氨酶TadA7-10的氨基酸序列。

20 在另一优选例中,所述第一启动子来源于选自下组的一种或多种植物:玉米、
水稻、大豆、拟南芥、番茄。

在另一优选例中,所述第二启动子来源于选自下组的一种或多种植物:水稻、
玉米、大豆、拟南芥、番茄。

在另一优选例中,所述第二整合元件包括 3' 同源臂序列。

25 在另一优选例中,所述第二整合元件为 LB 序列。

在另一优选例中,所述 LB 序列如 SEQ ID NO. :4(TGTTTACACCACAATATATCCTGCCA)
所示。

在另一优选例中,所述核定位信号来源于农杆菌。

在另一优选例中,所述核酸构建物的长度为 5000-10000bp,较佳地,
30 7500-8500bp。

在另一优选例中,在所述的 I1 和 I2 元件之间,还含有额外插入的一个或多个

额外的表达盒。

在另一优选例中，所述的额外表达盒是独立于所述的第一表达盒和第二表达盒的。

在另一优选例中，所述的额外表达盒表达选自下组的物质：

- 5 (a1) 标记基因；
 (a2) 与 Y1 编码的 sgRNA 不同的一种或多种 sgRNA。

在另一优选例中，所述标记基因包括抗性基因(潮霉素基因)、荧光基因、或其组合。

10 本发明第二方面提供了一种载体，所述载体含有本发明第一方面所述的核酸构建物。

在另一优选例中，所述载体为植物表达载体。

在另一优选例中，所述的载体为可转染或转化植物细胞的表达载体。

在另一优选例中，所述的载体为农杆菌 Ti 载体。

在另一优选例中，所述的构建物整合到所述载体的 T-DNA 区。

15 在另一优选例中，所述载体是环状的或线性的。

本发明第三方面提供了一种基因工程细胞，所述细胞含有本发明第一方面所述的核酸构建物，或其基因组整合有一个或多个本发明第一方面所述的核酸构建物。

在另一优选例中，所述的细胞为植物细胞。

20 在另一优选例中，所述的植物选自下组：禾本科植物、豆科植物、十字花科植物、或其组合。

在另一优选例中，所述的植物包括：拟南芥、小麦、大麦、燕麦、玉米、水稻、高粱、粟、大豆、花生、烟草、番茄、或其组合。

25 在另一优选例中，所述的基因工程细胞是用选自下组的方法将本发明第一方面所述的核酸构建物导入细胞的：农杆菌转化法、基因枪法、显微注射法、电击法、超声波法和聚乙二醇(PEG)介导法。

本发明第四方面提供了一种对植物进行基因编辑的方法，包括步骤：

- 30 (i) 提供待编辑植物；和
 (ii) 将本发明第一方面所述的核酸构建物或本发明第二方面所述的载体导入所述待编辑植物的植物细胞，从而在所述植物细胞内进行基因编辑。

在另一优选例中，所述步骤(ii)和步骤(iii)中的植物细胞来自同一部位。

在另一优选例中，所述导入为通过农杆菌导入。

在另一优选例中，所述导入为通过基因枪导入。

在另一优选例中，所述的基因编辑为定点碱基替换(或突变)。

在另一优选例中，所述定点替换(或突变)包括将 T 突变为 C 和/或 A 突变为 G。

- 5 在另一优选例中，所述的植物选自下组：禾本科植物、豆科植物、十字花科植物、或其组合。

在另一优选例中，所述的植物包括：拟南芥、小麦、大麦、燕麦、玉米、水稻、高粱、粟、大豆、花生、烟草、番茄、或其组合。

本发明第五方面提供了一种制备转基因植物细胞的方法，包括步骤：

- 10 (i)将本发明第一方面所述的核酸构建物、或本发明第二方面所述的载体转染植物细胞，使得所述核酸构建物与所述植物细胞中的染色体发生定点替换(或突变)，从而制得所述转基因植物细胞。

在另一优选例中，所述的转染采用农杆菌转化法或基因枪轰击法。

本发明第六方面提供了一种制备转基因植物细胞的方法，包括步骤：

- 15 (i)将本发明第一方面所述的核酸构建物、或本发明第二方面所述的载体转染植物细胞，使得所述植物细胞含有所述核酸构建物，从而制得所述转基因植物细胞。

本发明第七方面提供了一种制备转基因植物的方法，包括步骤：

将本发明第五方面或本发明第六方面所述方法制备的所述转基因植物细胞再生为植物体，从而获得所述转基因植物。

- 20 本发明第八方面提供了一种转基因植物，其特征在于，所述的植物是用本发明第七方面所述的方法制备的。

- 25 应理解，在本发明范围内中，本发明的上述各技术特征和在下文(如实施例)中具体描述的各技术特征之间都可以互相组合，从而构成新的或优选的技术方案。限于篇幅，在此不再一一累述。

附图说明

图 1 显示了水稻中 A.T 到 G.C 的高效碱基编辑。

- 30 其中，(A)两个水稻腺嘌呤碱基编辑载体 pRABEsp-0sU6 和 pRABEsa-0sU6sa 的示意图。(B)OsSPL14 基因中 sgRNA1 靶位点的示意图。OsmiR156 与 OsSPL14 结合的序列以红色突出显示。(C)两个代表性的 OsSPL14 编辑植株 SG1-7 和

SG1-15 的测序峰图。(D)SLR1 基因中 sgRNA2 靶位点的示意图。原间隔序列 (protospacer) 区第 6 位的 T-C 替换可能导致 TVHYNP 基序中 92 位氨基酸 V 变为 A。(E)两个 SLR1 编辑植株 SG2-18 和 SG2-19 的测序峰图。(F)设计的 sgRNA4 同时靶向编辑 OsSPL14 和 OsSPL17 基因。OsSPL14 和 OsSPL17 中的 OsmiR156 结合序列以红色突出显示。(G)OsSPL14 和 OsSPL17 在两个转基因水稻系 SG4-5 和 SG4-25 中都发生碱基编辑。两个株系两个靶位点的测序峰图如图所示。(H)本发明中使用不同 sgRNA 的碱基编辑效率的统计。水稻中的多位点碱基编辑分别由 sgRNAs3-5 引导。

图 2 显示了 sgRNA1 靶向 OsSPL14 基因的两个碱基编辑植株 (SG1-7 和 SG1-15) 的 TA 克隆测序结果。随机选取每行 20 个克隆进行测序, 显示了每个基因型的代表性测序峰图。

图 3 显示了 pRABEsp-0sU6 可用于水稻的多重碱基编辑。

其中, (A)sgRNA3 被设计用于同时靶向编辑水稻基因组中的三个基因。三个基因中的 OsmiR156 结合核苷酸以红色突出显示。(B)提供了三个靶位点的代表性序列峰图。在 SG3-11 和 SG3-12 植株中, 同时编辑 OsSPL16 和 OsSPL18 中的目标站点同时编辑, 但 LOC_0s02g24720 的目标位点在这两个植株中为野生型。

图 4 显示了 pRABEsA-0sU6sa 在 OsSPL16 和 OsSPL18 上同时进行碱基编辑。

其中, (A)sgRNA5 可同时靶向 OsSPL16 和 OsSPL18 两个位点。(B)代表性的三个转基因植株 SG5-7, SG5-18, SG5-44 在两个靶位点的测序峰图。这两个基因都是在三个选定的转基因植株中被同时编辑了。

图 5 显示了 ABE-P1S 对 OsSPL14 靶位点的编辑。

其中, A. ABE-P1S 碱基编辑器的载体示意图. B. OsSPL14 基因中 sgRNA1 靶位点的示意图。OsmiR156 与 OsSPL14 结合的序列以红色突出显示。C. 两个代表性的株系 Line 7 和 Line 12 在 OsSPL14 靶位点的测序峰图, 黑色箭头指示了发生碱基替换的位点。D. ABE-P1S 对 OsSPL17 位点的脱靶编辑, OsmiR156 与 OsSPL17 结合的序列以红色突出显示, 错配碱基字母小写, 黑色箭头指示了发生碱基替换的位点。

图 6 显示了 ABE-P2S 对 SPX2-MFS2 位点的编辑。

其中, A. ABE-P2S 碱基编辑器的载体示意图. B. SPX2-MFS2 基因中 sgRNA8 靶位点的示意图。OsmiR827 与 SPX2-MFS2 结合的序列以红色突出显示。C. 两个

代表性的株系 Line 38 和 Line 40 在 SPX2-MFS2 靶位点的测序峰图，黑色箭头指示了发生碱基替换的位点。

图显示了 7 ABE-P5 和 ABE-P5S 对 OsSPL13 位点的编辑。

其中，A. ABE-P5 和 ABE-P5S 碱基编辑器的载体示意图。B. OsSPL13 基因中 sgRNA11 靶位点的示意图。OsmiR156 与 OsSPL13 结合的序列以蓝色突出显示，PAM 序列由红色字母标出。C. 代表性的株系 Line3, Line 23 和 Line 27 在 OsSPL13 靶位点的测序峰图，黑色箭头指示了发生碱基替换的位点。

具体实施方式

10 本发明人经过广泛而深入地研究，通过大量筛选，筛选出驱动 Cas9 核酸酶与腺嘌呤脱氨酶、核定位信号 VirD2 构成的融合蛋白表达的特定启动子，以及驱动 sgRNA 转录的特定的启动子，并通过采用特定结构的核酸构建物，本发明首次的植物中成功实现了 sgRNA 引导的碱基定点突变(如 T 突变为 C 或 A 突变为 G)，并且突变效率非常高(可高达 $\geq 60\%$ 或更高)，并且申请人意外的发现，简化后的腺嘌呤碱基编辑器 ABE-P1S 具有更高的碱基编辑效率。并在此基础上完成了
15 本发明。

术语

如本文所用，术语“同源臂”指打靶载体上待插入的外源序列两侧的与基因组序列完全一致的侧翼序列，用于识别并发生重组的区域。
20

如本文所用，术语“植物启动子”指能够在植物细胞中启动核酸转录的核酸序列。该植物启动子可以是来源于植物、微生物(如细菌、病毒)或动物等，或者是人工合成或改造过的启动子。

如本文所用，术语“碱基突变”指核苷酸序列的某一位置处发生碱基的替换(substitution)、插入(insertion)和/或缺失(deletion)。
25

如本文所用，术语“碱基替换”指核苷酸序列的某一位置处的碱基突变为另一不同的碱基，比如 T 突变为 C。

如本文所用，术语“A.T 到 G.C”指在双链核酸序列(尤其是基因组序列)中，某一位置上的 A-T 碱基对突变为或替换为 G-C 碱基对。

30 如本文所用，“筛选标记基因”指转基因过程中用来筛选转基因细胞或转基因动物的基因，可用于本申请的筛选标记基因没有特别限制，包括转基因领

域常用的各种筛选标记基因，代表性例子包括(但不限于)：潮霉素抗性基因(Hyg)、卡那霉素抗性基因(NPTII)、新霉素基因、嘌呤霉素抗性基因、潮霉素的抗性基因(HYG)、G418 和卡那霉素的抗性基因(NPTII)、Basta 的抗性基因(BAR)、嘌呤霉素抗性基因(PAC)、和/或新霉素抗性基因(NEO)。

5 如本文所用，术语“Cas 蛋白”指一种核酸酶。一种优选的 Cas 蛋白是 Cas9 蛋白。典型的 Cas9 蛋白包括(但不限于)：来源于酿脓链球菌(*Streptococcus pyogenes*)的 Cas9。在本发明中，Cas9 蛋白为突变的 Cas9 蛋白，具体地，是无切割活性或只具有单链切割活性的突变的 Cas9 蛋白。

10 如本文所用，术语“Cas 蛋白的编码序列”指编码 Cas 蛋白的核苷酸序列。在插入的多聚核苷酸序列被转录和翻译从而产生功能性 Cas 蛋白的情况下，技术人员会认识到，因为密码子的简并性，有大量多聚核苷酸序列可以编码相同的多肽。另外，技术人员也会认识到不同物种对于密码子具有一定的偏好性，可能会根据在不同物种中表达的需要，会对 Cas 蛋白的密码子进行优化，这些变体都被术语“Cas 蛋白的编码序列”所具体涵盖。此外，术语特定地包括
15 了全长的、与 Cas 基因序列基本相同的序列，以及编码出保留 Cas 蛋白功能的蛋白质的序列。

在本发明中，核苷酸序列的描述是从 5' 至 3' 方向，除非特别注明。

腺嘌呤脱氨酶

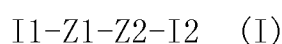
20 如本文所用，术语“腺嘌呤脱氨酶”指 TadA 腺嘌呤脱氨酶，来源于大肠杆菌，原本作用于 tRNA，能够对 tRNA 中的特定腺嘌呤进行脱氨反应。

在本发明中，适用的 TadA 既包含野生型的形式也包含其特定的突变形式 TadA7-10，也可包含野生型的形式和突变形式的组合。TadA7-10 能够以 DNA 作为底物进行脱氨反应。

25 在另一优选例中，本发明的腺嘌呤脱氨酶的编码序列是对密码子进行优化的，从而能够更高效地在植物中表达。

本发明的构建物

30 本发明提供了一种核酸构建物，用于对植物进行基因编辑，所述的核酸构建物具有 5' -3' 的式 I 结构：



其中，

I1 为第一整合元件；

I2 为第二整合元件；

Z1 为第一表达盒；

5 Z2 为第二表达盒；

并且，Z1 和 Z2 中的一个表达盒具有 Ia 结构，而另一个表达盒具有式 Ib 结构：

P1-S1-X1-L1-X2-L2-X3 (Ia)

P2-Y1 (Ib)；

10 I1、P1、S1、X1、L1、X2、L2、X3、L3、P2、Y1、I2 分别为用于构成所述构建物的元件，其定义如本发明第一方面所述；

并且，各“-”为键或核苷酸连接序列。

在上述式 I 结构中，I1 元件(或左侧整合元件)和 I2 元件(或右侧整合元件)可协同作用，从而将位于其间的元件(即从 P1 至 Y1 的核苷酸序列)整合到植物细胞的基因组中。

15 代表性的 I1 和 I2 是来自于农杆菌的 Ti 元件。当然，其他可起到类似整合作用的元件也可用于本发明。

本发明的构建物中所用的各种元件或者是本领域中已知的，或者可用本领域技术人员已知的方法制备。例如，可通过常规方法，如 PCR 方法、全人工化学合成法、酶切方法获得相应的元件，然后通过熟知的 DNA 连接技术连接在一起，就形成了本发明的构建物。

将本发明的构建物插入外源载体(尤其是适合转基因植物操作的载体)，就构成了本发明的载体。

将本发明的载体转化植物细胞从而介导本发明的载体对植物细胞染色体进行整合，制得转基因植物细胞。

25 将本发明的转基因植物细胞再生为植物体，从而获得转基因植物。

将本发明构建好的上述核酸构建物，通过常规的植物重组技术(例如农杆菌转让技术)，可以导入植物细胞，从而获得携带所述核酸构建物(或带有所述核酸构建物的载体)的植物细胞，或获得基因组中整合有所述核酸构建物的植物细胞。

30 载体构建

该载体的主要特征是将腺嘌呤脱氨基酶与 CRISPR/Cas 系统中的 Cas 蛋白

以及核定位信号 VirD2 的编码序列连接在一起,从而形成融合蛋白的编码序列。当该编码序列所编码的融合蛋白在细胞质中表达后,所述的融合蛋白可以非常高效地被转移至细胞核内,并由式 I 构建物所编码的 guide RNA 引导至基因组中的靶点位置,从而在靶点位置进行 A.T 到 G.C 的碱基替换,并基本上避免或消除了发生插入/缺失的风险。

由于腺嘌呤脱氨基酶直接将 T 突变为 C 或 A 突变为 G,并不需要 Cas 蛋白的 DNA 双链切割活性。因此,在本发明中 Cas 蛋白是无切割活性的突变的 Cas 蛋白。在一优选实施方式中,本发明的 Cas 蛋白可以是 SaKKH-Cas9(D10A),其氨基酸序列如 SEQ ID NO.:69 所示。一般的,为了增加融合蛋白的活性,蛋白间一般通过一些柔性短肽连接,即 Linker(连接肽序列)。优选的,该 Linker 可以选用 XTEN。

在本发明中,合适的启动子包括组成型和/或诱导型启动子。优选地,为了增加效率,可以选择适用于植物细胞的强启动子,代表性的例子包括(但不限于):CaMV 35S 启动子或者 UBI 启动子或 Actin 启动子等。

选择适用于植物细胞的 guide RNA 的表达框,并将其与上述融合蛋白的开放表达框(ORF)构建在同一载体。

靶点设计

在本发明中,当腺嘌呤脱氨基酶通过 CRISPR/Cas9 系统引导至靶点位置后,脱氨基酶的作用区域就被固定的。

例如,将腺嘌呤脱氨基酶 TadA 或突变型的腺嘌呤脱氨基酶 TadA7-10 蛋白通过 32 个氨基酸的 XTEN Linker 连接至 Cas9 的 N 端后,通常,在人细胞系里面其脱氨基的作用区域为原间隔序列(protospacer)区的第 4-7 个碱基区域,本发明得到的实验结果表明,如果用 SpCas9,碱基编辑窗口为原间隔序列(protospacer)区第 5-10 位,如果用 SaCas9,碱基编辑窗口为原间隔序列(protospacer)区第 6-14 位。根据这一原则,在设计靶点时,需将待编辑的腺嘌呤处于“脱氨基窗口区”。

遗传转化

在本发明中,对于将本发明的式 I 构建物导入细胞或整合到基因组的方法,没有特别限制。可以用常规的方法进行,例如将式 I 构建物或相应的载体通过

合适的方法导入到植物细胞中。代表性的导入方法包括但不限于：农杆菌转染法、基因枪法、显微注射法、电击法、超声波法、和聚乙二醇 (PEG) 介导法等。

在本发明中，对于受体植物没有特别限制，其中包括各种不同的农作物植物(如禾本科植物)、林业植物、园艺植物(如花卉植物)等。代表性的例子包括但不限于：水稻、大豆、番茄、玉米、烟草、小麦、高粱等。

上述 DNA 载体或片段导入植物细胞后，使转化的植物细胞中的 DNA 表达该融合蛋白和 gRNA。融合腺嘌呤脱氨基酶的 Cas 蛋白在相应 gRNA 的引导下，将靶点位置的 T 突变为 C(进而使得互补链的 A 突变为 G)。

对于用本发明方法进行植物基因组定点替换后的植物细胞或组织或器官，可以用常规方法再生获得相应的转基因植株。例如，通过组织培养，再生获得碱基替换后的植株。

应用

本发明可以用于植物基因工程领域，用于植物研究和育种，尤其是具有经济价值的农作物、林业作物或园艺植物的遗传改良。

本发明的主要优点包括：

(1) 本发明首次提供了一种在植物中实现定点碱基突变(T→C 或 A→G)的高效方法，可以广泛的用于植物研究和育种。

(2) 本发明的定点碱基突变的方法可高效地在植物细胞的特定位置进行碱基突变(如 T→C 或 A→G)。

(3) 用本发明的方法在植物细胞中进行碱基替换的效率非常高，可高达 60% 或更高。

(4) 本发明能够同时对植物基因组中的多个位点同时进行编辑。

(5) 本发明可以通过使用不同形式的 Cas9，扩大植物基因组中可被定点编辑的位点。

(6) 本发明可显著降低或基本上消除在靶位点发生插入和/或缺失的风险。

(7) 本发明发现，sgRNA 的第 12 位、第 13 位和/或第 14 位对应于预定发生 T→C 定点突变的位置(即为 T)(从 5' 至 3')的碱基替换效率更高。

下面结合具体实施例，进一步阐述本发明。应理解，这些实施例仅用于说明本发明而并不用于限制本发明的范围。下列实施例中未注明具体条件的实验方法，通常按照常规条件如 Sambrook 等人，分子克隆：实验室手册 (New York: Cold Spring Harbor Laboratory Press, 1989) 中所述的条件，或按照制造厂商所建议的条件。除非另外说明，否则百分比和份数按重量计算。本发明中所涉及的实验材料和试剂如无特殊说明均可从市售渠道获得。

通用方法

1. 载体构建

用常规方法合成大肠杆菌野生型 tRNA 腺嘌呤脱氨基酶基因 TadA(ecTadA) 和其突变体形式 TadA7-10(ecTadA * 7.10)。TadA 和 TadA7-10 3' 端还各加有一段编码 32 个氨基酸残基的连接肽序列的 96 bp 的接头编码序列。TadA-linker 和 TadA7-10-linker 两端分别加有 AarI 酶切位点。

通过对 pCas9(0sU6) 载体的改造，我们获得水稻腺嘌呤碱基编辑载体。

pCas9(0sU6) 载体中 Cas9 基因由玉米泛素启动子驱动表达，sgRNA 骨架由水稻 U6 启动子驱动表达。Cas9 (D10A) 切口酶和 SaCas9(D10A) 切口酶通过 PCR 以 pCas9(0sU6) 和 pX600 (Addgene, #61592) 为模板扩增获得。上游引物 5' 端含有两个 AarI 酶切位点，下游引物 3' 端加有 VirD2 核定位信号，扩增产物回收后用于替换 pCas9(0sU6) 中的 Cas9 基因。分别得到中间载体 pRSp-0sU6 and pRSa-0sU6。。TadA-linker 和 TadA7-10-linker 片段通过 Golden Gate 的方式连入 pRSp-0sU6 和 pRSa-0sU6，分别获得 pRABESp-0sU6 和 pRBESa-0sU6 载体。

通过 overlap PCR，将水稻 U6 启动子，两个 BsaI 酶切位点，和 SaCas9 的 sgRNA 骨架连接在一起用于替换 pRBESa-0sU6 中 0sU6-sgRNA 片段，获得载体 pRABESa-0sU6Sa。

简化版腺嘌呤碱基编辑载体构建

为了构建简化版的腺嘌呤碱基编辑载体，我们首先对合成的 ecTadA* (7.10) 两端的 AarI 酶切位点接头进行了改造，然后通过金门克隆的方式将单个 ecTadA*(7.10) 连接到之前构建的 pRSp-0sU6 and pRSa-0sU6 中间载体中，得到载体 pRSABESp-0sU6 和 pRSABESa-0sU6。pRSABESp-0sU6 即对应腺嘌呤碱基编辑器 ABE-P1S。用 HindIII 和 XmaI 从之前构建的 pRABESa-0sU6Sa

载体上切下 OsU6-SasgRNA 表达盒来替换掉 pRSABESa-OsU6 中 OsU6-SpsgRNA 表达盒，获得 pRSABESa-OsU6Sa 载体，它对应腺嘌呤碱基编辑器 ABE-P2S。我们用 PCR 从 pRABESa-SaKKH 载体上扩增得到 nSaKKH-Cas9 (D10A) 缺口酶，用它替换 pRSABESa-OsU6Sa 载体中的 nSaCas9 (D10A)，
5 得到 pRSABESa-SaKKH 载体，它对应腺嘌呤碱基 编辑器 ABE-P5S。

pRABESp-OsU6 和 pRABESa-OsU6Sa 载体骨架上含有潮霉素基因用于转基因植株的筛选。

人工合成 sgRNA1 至 sgRNA5 的序列。引物在 PCR 仪上退火后形成短的寡聚核苷酸接头，寡聚核苷酸接头插入到由 BsaI 酶切的 pRABESp-OsU6，
10 pRABESa-OsU6Sa 载体中。

所有的载体都通过 Sanger 测序确认其准确性。

2. 水稻转化

所有的二元载体通过冻融法转入常规的农杆菌 EHA105 之中。水稻转化受体为日本晴。具体的实验步骤按(Nishimura et al., 2007)等报道的进行。愈伤组织在农杆菌侵染两天以后转到含 50mg/L 潮霉素的筛选培养基上进行筛选。筛选 2 周以后，抗性愈伤组织直接转到分化培养基之中进行分化。待小苗长到 4-5cm 长时，转到生根培养基中诱导生根。10 天过后，打开生根盒炼苗 3 天后转到营养土中，放入温室生长，温室条件为 28° C，光照 12 小时，22° C 黑暗
20 12 小时。

3. 目标位点基因型检测

以 CTAB 法提取所有转基因水稻植株的叶片 DNA。在靶标位点上下游各 250bp 左右设计引物，通过 PCR 扩增出目标位点。PCR 产物回收后直接送公司
25 测序。测序结果有 Sequencher 软件分析。对于有些在目标位点发生编辑的植株，我们通过 TA 克隆进行了进一步的验证。具体为目标位点通过 PCR 扩增，纯化后连入 p-EASY blunt Zero 载体转化大肠杆菌。随机挑选 20 个克隆进行测序。碱基编辑的效率的计算由产生碱基编辑的植株数目除以总的转基因数目。

30

4. 脱靶位点的检测

对于 sgRNA1 和 sgRNA4 的脱靶位点进行预测分析。水稻基因组中与 sgRNA1 和 sgRNA4 序列错配低于 5 个碱基以下的位点被认为是潜在的脱靶位点。在所有潜在脱靶位点上下游各 250bp 左右设计引物, PCR 扩增出这些潜在脱靶位点, PCR 产物纯化后直接送公司进行测序。测序结果有 Sequencher 软件分析。

5

实施例 1 测试 ABE7-10 在水稻中的编辑效果

在本实施例中, 根据通用方法, 本发明合成了野生型 ecTadA 及其突变体形式 ecTadA * 7.10 的编码序列。然后通过使用编码 32 个氨基酸残基的接头编码序列将它们连接在一起。接下来, 将该重组蛋白的编码序列与具有相同接头的 Cas9(D10A)切口酶编码序列的 N-末端融合。最后将 VirD2 核定位信号的编码序列连接到 Cas9(D10A)切口酶的 C 端, 以形成 ABE7-10。然后将 ABE7-10 克隆到玉米泛素启动子控制下的双元载体中, 并且通过水稻 U6 启动子驱动 sgRNA, 形成载体 pRABEsp-0sU6(图 1A)。

为了测试 ABE7-10 在水稻中的编辑效果, 我们选择调节水稻理想株型的 *IPA1(0sSPL14)* 作为靶基因(图 1B)。

本发明设计了一个 sgRNA (sgRNA1) 靶向 *0sSPL14* 的 0smiR156 结合位点序列(图 1B), sgRNA1 载体构建所对应的引物为表 3 中的 sgRNA1F 和 sgRNA1R。该双元载体以冻融法转化农杆菌, 然后以农杆菌侵染水稻日本晴的愈伤组织, 获得了 23 个独立的转基因系。然后, 通过 PCR 扩增目标区域, 并通过 Sanger 测序进行基因型鉴定, 扩增引物为表 3 中 SPL14-seq-F 和 SPL14-seq-R。

表3 本发明用到的引物

引物名字	SEQ ID NO.:	引物序列 5'-3'	目的
sgRNA1-F	5	TGTGAGAGAGAGCACAGCT CGAGT	sgRNA1载体构建
sgRNA1-R	6	AAACACTCGAGCTGTGCTCT CTCT	
sgRNA2-F	7	TGTGAGTGCACGGTGTCCGT GGCC	sgRNA2载体构建
sgRNA2-R	8	AAACGGCCACGGACACCGT GCACT	
sgRNA3-F	9	TGTGCAGAAGAGAGAGAGC	sgRNA3载体构建

		ACAAT	
sgRNA3-R	10	AAACATTGTGCTCTCTCTCT TCTG	
sgRNA4-F	11	TGTGTGACAGAAGAGAGAG AGCACAGC	sgRNA4载体构建
sgRNA4-R	12	AAACGCTGTGCTCTCTCTCT TCTGTCA	
sgRNA5-F	13	TGTGTGACAGAAGAGAGAG AGCACAAT	sgRNA5载体构建
sgRNA5-R	14	AAACATTGTGCTCTCTCTCT TCTGTCA	
SPL14-seq-F	15	AGGGTTCCAAGCAGCGTAA GGA	在 <i>OsSPL14</i> 中对目标位点进行基因分型
SPL14-seq-R	16	TGGTGCTGGGGCTGGACCGT TC	
SLR-seq-F	17	GCGCAATTATTACTAGCTAT AGC	在 <i>SLR1</i> 中对目标位点进行基因分型
SLR-seq-R	18	AGCCGTCGCCACCACCGTA AGG	
SPL16-seq-F	19	CAGCTGCTGCTGCTGCTTCA GTGTG	在 <i>OsSPL16</i> 中对目标位点进行基因分型
SPL16-seq-R	20	ACATCCCATTGTAGTTCATC TCATTG	
02G-seq-F	21	GCCTGCAGGCGGAGGAGTG G	在 <i>LOC_Os02g24720</i> 中对目标位点进行基因分型
02G-seq-R	22	AGTTCATCTCATTGTCATTG GA	
SPL17-seq-F	23	GGACCTCGTTCAGCACAACC C	在 <i>OsSPL17</i> 中对目标位点进行基因分型
SPL17-seq-R	24	GGGTCCAAGCAGTGTGAGG GA	
SPL18-seq-	25	TGGGATCATCAAATCCGAGG	在 <i>OsSPL18</i> 中对目标位点

F		AG	进行基因分型
SPL18-seq-R	26	CTGTCCATGCTCGGGCAGGC G	
sgRNA1M 1-F	27	GGGACCTCGTTCAGCACAAC CCC	用于检测sgRNA1的脱靶 位点1
sgRNA1M 1-R	28	GCAGGTCCAGAAGCTTTGTG GTA	
sgRNA1M 2-F	29	TTGCTCCCTAAACAACTCCC AG	用于检测sgRNA1的脱靶 位点2(Off target site 2 detection for sgRNA1)
sgRNA1M 2-R	30	CCTGGTGCAGAGAGTACAA ATG	
sgRNA1M 3-F	31	TGCAATGCATCTACTCCCTC AG	用于检测sgRNA1的脱靶 位点3 (Off target site 3 detection for sgRNA1)
sgRNA1M 3-R	32	GATCACACTAGCGACAGCG AGC	
sgRNA1M 4-F	33	ACGAGAGCTTCGACTGAACG CA	用于检测sgRNA1的脱靶 位点4(Off target site 4 detection for sgRNA1)
sgRNA1M 4-R	34	AATCTGCCGCGTGTGTCAGCAG AG	
sgRNA1M 5-F	35	TGCAGGTGGTCGGCGATCGC G	用于检测sgRNA1的脱靶 位点5 (Off target site 5 detection for sgRNA1)
sgRNA1M 5-R	36	AGCGCGCGCAGCTTGGCCT	
sgRNA1M 6-F	37	GATCACAACTGCTCGTAAGC TA	用于检测sgRNA1的脱靶 位点6(Off target site 6 detection for sgRNA1)
sgRNA1M 6-R	38	ATATGTCCTTATCGGACACG AC	
sgRNA1M 7-F	39	CCAAACTCTATCTTGAGCCT TC	用于检测sgRNA1的脱靶 位点7 (Off target site 7 detection for sgRNA1)
sgRNA1M 7-R	40	GAGTTCGACGAGGGAGAGA GA	
sgRNA1M	41	ATATTCATAATCCCCCTAGA	用于检测sgRNA1的脱靶

8-F		A	位点8(Off target site 8 detection for sgRNA1)
sgRNA1M 8-R	42	CCCCAACGCTTCGCGAATCG CA	
sgRNA1M 9-F	43	CAGAGCCTGGACGGGTTCTA C	用于检测sgRNA1的脱靶 位点9 (Off target site 9 detection for sgRNA1)
sgRNA1M 9-R	44	ACAGTACAGCAAATCGCAC GT	
sgRNA4M 1-F	45	GCAACAAGATGTTCTCCTCC G	用于检测sgRNA4的脱靶 位点1 (Off target site 1 detection for sgRNA4)
sgRNA4M 1-R	46	CTGCCGCCGAAGCTGCTGCTG C	
sgRNA4M 2-F	47	ACACCTGCCAAGAGAATGG CA	用于检测sgRNA4的脱靶 位点2 (Off target site 2 detection for sgRNA4)
sgRNA4M 2-R	48	CCGATAGCTGATCAGAACGC	
sgRNA4M 3-F	49	TTGGCATCAGCAGCAGGCAG CA	用于检测sgRNA4的脱靶 位点3 (Off target site 3 detection for sgRNA4)
sgRNA4M 3-R	50	GAACCAGGCTGAGCTGCCGC C	
sgRNA4M 4-F	51	TATTATTCGTGCAATGCATC TA	用于检测sgRNA4的脱靶 位点4(Off target site 4 detection for sgRNA4)
sgRNA4M 4-R	52	ATGCGAGCGGTCTAACCGAC GA	

根据测序结果，在 23 株转基因植株中有 6 株在目标位点显示了预期的 T-C 替换，碱基编辑效率高达 26%(6/23) (图 1H)。在这 6 个株系中，两个株系 (SG1-11, SG1-21) 在 protospacer 的第 5 位具有 T-C 替换(将 PAM 位点定为 21-23)，两个株系 (SG1-10, SG1-15) protospacer 的 5 和 7 位具有 T-C 替换，而其余两株 (SG1-7, SG1-23) 在 protospacer 的 5 和 10 具有 T-C 置换(图 1C)。

本发明发现两个转基因植株 (SG1-7 和 SG1-10) 在 protospacer 的第 10 位发生编辑(图 1C)。为了进一步确认 Sanger 测序结果，选择了 SG1-7 和 SG1-15 进行 TA 克隆，随机挑选了 20 个克隆进行测序。有趣的是，本发明发现来自 SG1-7

的 11 个克隆在 protospacer 第 5 位具有 T-C 替换, 9 个克隆在 protospacer 的第 10 位具有 T-C 取代, 表明这株转基因是 bi-allele(图 2)。然而, SG1-15 株系只有 15%的克隆 (3/20) 在 protospacer 的 5 和 7 位有 T-C 替换, 表明该转基因植株是嵌合型的(图 2)。

- 5 为了评估 ABE7-10 在水稻中的特异性, 通过在线工具 CRISPR-GE 预测了 sgRNA1 在水稻基因组中的潜在脱靶位点。对九个潜在脱靶位点的测序未发现任何碱基编辑事件。九个脱靶位点的信息见表 1, 对九个潜在脱靶位点扩增所用的引物见表 3。

表1. sgRNA1潜在脱靶位点的编辑效率

位点	染色体	位置	引导-PAM 序列	SEQ ID NO.:	错配数	编辑效率
On target	8	2527516	AGAGAGAGCACAGCTCGAG TCGG	53	0	26%
Off-target 1	8	1891891	AGAGAGAGCACAGCTgGAGT CGG	54	1	0
Off-target 2	1	2417964	tGAtcGgGCACAGCTCGcGT CG G	55	5	0
Off-target 3	3	3613708	AGAaAGAGCAtgGgTCGAGTC GG	56	4	0
Off-target 4	3	8596704	AGAGtGAGCACAGCggGAGaC GG	57	4	0
Off-target 5	4	2302582	gGAGAGcGCgCgGCTCGAGgC GG	58	5	0
Off-target 6	5	1028490	AGAGtGtGCAgAGtTCGAGTC GG	59	4	0
Off-target 7	7	1903042	AGAGAGAGCtCgGCTCGgcTC GG	60	4	0
Off-target 8	10	1012640	AGAGAGAtCtCAGaTCGAGgC GG	61	4	0
Off-target	7	2080424	AGAcAGAGCACAGCaaGAaTC	62	4	0

9

0

GG

注：PAM序列的核苷酸以粗体书写，潜在脱靶的错配碱基以小写字母显示。

测序结果表明，腺嘌呤碱基编辑在水稻中是高度特异的。

实施例 2 测试水稻中腺嘌呤碱基编辑系统的编辑能力

5 为了进一步测试水稻中腺嘌呤碱基编辑系统的编辑能力，对 *SLR1* 基因座做进一步实验。*SLR1* 编码水稻中的 DELLA 蛋白，其作为 GA 信号传导途径中的阻遏物。

本发明设计了针对 *SLR1* 的 TVHYNP 结构域的 sgRNA(sgRNA2)。sgRNA2 载体构建所对应的引物为表 3 中的 sgRNA2F 和 sgRNA2R。Protospacer 第 6 位的 T-C
10 碱基替换可导致 TVHYNP 基序中的 V92A 替换(图 1D)。

从获得的 40 个转基因株系中，5 个株系在 protospacer 的 6 位具有预期的 T-C 替换(图 1E, 1H)。目标区域扩增引物为表 3 中 SLR-seq-F 和 SLR-seq-R 由于在 *SLR1* 的碱基编辑窗口中只有一个可编辑的 T，因此在目标基因座中没有发现任何其他突变形式。这些结果证实了水稻中腺嘌呤碱基编辑系统的高特异性。
15 性。

实施例 3 检测本发明的碱基编辑系统是否可以同时编辑水稻基因组中的两个或多个位点

为了检测本发明的碱基编辑系统是否可以同时编辑水稻基因组中的两个
20 或多个位点，同时设计了第三个 sgRNA(sgRNAs3)，同时靶向 *OsSPL16* 和 *OsSPL18* 的 *OsmiR156* 结合位点(图 3, A 和 B)。sgRNA3 载体构建所对应的引物为表 3 中的 sgRNA3F 和 sgRNA3R。另外，在 *LOC_Os02g24720* 基因的内含子中发现了一个与 sgRNA3 序列 100%匹配的脱靶位点(图 3, A 和 B)。因此，sgRNA3 可以同时靶向水稻基因组中的三个位点。本发明对 21 个转基因株系中的这三个靶位点进行
25 了基因型鉴定。各有四个转基因株系在 *OsSPL16* 和 *OsSPL18* 靶向区域中显示 T-C 替换。然而，在 *LOC_Os02g24720* 中只有一个株系含有 T-C 取代(图 1H 和图 3)。三个目标位点的扩增引物为表 3 中 SPL16-seq-F/SPL16-seq-R, SPL18-seq-F/SPL18-seq-R, 02G-seq-F/02G-seq-R。有趣的是，在 SG3-11, SG3-12 两个转基因株系中，*OsSPL16* 和 *OsSPL18* 能够
30 被同时编辑，这表明本发明的碱基编辑系统可以在水稻中同时靶向多个基因

(图 3, A 和 B)。

利用 pRABEsp-0sU6 载体进行碱基编辑需要在 protospacer 下游含有 NGG 的 PAM 序列。这一要求显著限制了水稻基因组中可被 pRABEsp-0sU6 编辑的位点数目。为了进一步增加水稻腺嘌呤碱基编辑器在水稻基因组中可编辑的目标, 我们用 SaCas9 (D10A) 切口酶和 SaCas9 的骨架 sgRNA 替换了 pRABEsp-0sU6 载体中的 Cas9 (D10A) 切口酶及其 sgRNA 骨架。最终构建成的载体 pRABEsa-0sU6sa 可以识别不同的 PAM 序列 NNGRRT (图 1A)。

实施例 4 测试该载体的可行性和效率

为了测试该载体的可行性和效率, 我们设计了第四个 sgRNA (sgRNA4), 同时靶向 *OsSPL14* 和 *OsSPL17* 的 0smiR156 结合位点 (图 1F)。sgRNA4 载体构建所对应的引物为表 3 中的 sgRNA4F 和 sgRNA4R。值得注意的是, 尽管 PAM 序列不同, 但 sgRNA4 识别的 protospacer 序列与 sgRNA2 识别的 protospacer 序列重叠 (图 1F 和 1B)。从我们基因鉴定的 31 个转基因株系中, 我们发现 14 个株系在 *OsSPL14* 目标位点具有 T-C 替换, 而在 *OsSPL17* 目标位点有 19 个株系发生了 T-C 的替换。目标位点的扩增引物为表 3 中 SPL14-seq-F/SPL14-seq-R, SPL17-seq-F/SPL17-seq-R, 因此, 在 *OsSPL14* 和 *OsSPL17* 靶位点处 pRABEsa-0sU6sa 的碱基编辑效率分别为 45.2% 和 61.3%, 远高于 sgRNA2 靶向的 pRABEsp-0sU6 (图 1H)。更重要的是, 在这两个靶位点有 13 个株系 (41.9%) 被同时编辑。

此外, SaCas9 (D10A) 切口酶的碱基编辑窗口比 Cas9 (D10A) 切口酶更宽, 这可能是由于在由 SaCas9 (D10A) 在诱导形成 R-环复合物形成期间有更多的单链暴露于腺嘌呤脱氨酶。出乎意料的是, 甚至在 protospacer 第 12 和 14 位的 T 在两个靶位点都被编辑了, 具体的 ABE-P1 的碱基编辑窗口口如表 4 和表 5 所示。

表4. ABE-P1的碱基编辑窗口

sgRN	靶点	编号	碱基编辑位点	编辑形式
A				
sgRN	<i>OsSPL14</i>	SG1-11, SG1-21	5	T-C的转换
A1				
		SG1-10, SG1-15	5,7	T-C的转换

		SG1-7, SG1-23	5,10	T-C的转换
sgRN	<i>SLR1</i>	SG2-4, SG2-18, SG2-19,	6	T-C的转换
A2		SG2-26, SG2-36		
sgRN	<i>OsSPL16</i>	SG3-3, SG3-11, SG3-12, SG3-13	7	T-C的转换
A3				
	<i>OsSPL18</i>	SG3-11, SG3-12	7	T-C的转换
		SG3-15	7,9	T-C的转换
		SG3-19	5,7	T-C的转换
	<i>LOC_Os02g24</i>	SG3-6	5	T-C的转换

注意：从PAM远端计算碱基编辑位置，将PAM记录为位置21-23(Note: Base editing position was counted from the PAM-distal end, scoring the PAM as position 21-23.)。

表5. ABE-P2的碱基编辑窗口

sgRN	靶点	编号	碱基编辑位置	编辑形式
A				
sgRN	<i>OsSPL14</i>	SG4-6, SG4-11, SG4-16	8	T-C的转换
A4		SG4-2, SG4-5, SG4-17, SG4-19	10	T-C的转换
		SG4-4	6, 10	T-C的转换
		SG4-1, SG4-24, SG4-27	8, 10	T-C的转换
		SG4-7, SG4-10	6, 8, 10	T-C的转换
		SG4-25	10, 12	T-C的转换
	<i>OsSPL17</i>	SG4-2, SG4-7	8	T-C的转换
		SG4-8	10	T-C的转换
		SG4-12, SG4-20	12	T-C的转换
		SG4-1, SG4-5, SG4-6, SG4-10,	8, 10	T-C的转换
		SG4-16, SG4-18, SG4-24,		
		SG4-26		
		SG4-11, SG4-19	8, 10, 12	T-C的转换
		SG4-27	8, 10, 14	T-C的转换
		SG-3	6, 10, 12, 14	T-C的转换

sgRN A5	<i>OsSPL16</i>	SG17, SG4-25	8, 10, 12, 14	T-C的转换
		SG5-18, SG5-30, SG5-45	12	T-C的转换
		SG5-7, SG5-23, SG5-32, SG5-33, SG5-44	14	T-C的转换
	<i>OsSPL18</i>	SG5-18, SG5-21, SG5-29, SG5-37, SG5-38, SG5-40, SG5-44, SG5-45	10	T-C的转换
		SG5-7	10, 12	T-C的转换
		SG5-23	8, 10, 12	T-C的转换
		SG5-12	10, 12, 14	T-C的转换

注意：从PAM远端计算碱基编辑位置，将PAM记录为位置22-27(Note: Base editing position was counted from the PAM-distal end, scoring the PAM as position 22-27.)。

- 5 为了评估 pRABEsA-0sU6sa 的脱靶编辑，用在线工具 CRISPR-GE 预测了 sgRNA4 在水稻基因组中的潜在脱靶位点。对 sgRNA4 的潜在脱靶位点进行了测序，发现这些位点没有发生任何形式的突变，表明 pRABEsA-0sU6sa 在水稻中也是高度特异的。sgRNA4 的潜在脱靶位点信息见表 2，扩增 sgRNA4 潜在脱靶位点的引物见表 3。

10

表2 sgRNA4潜在脱靶位点的编辑效率

位点	染色体	位置	引导-PAM 序列	SEQ ID NO.:	错配数	编辑 效率
On target	8	2527515	ACAGAAGAGAGAGAGCAC	63	0	45.2
1		6	AGCTCGAGT			%
On target	9	1891891	ACAGAAGAGAGAGAGCAC	64	0	61.3
2		1	AGCTGGAGT			%
Off-target	9	1964783	ACAGAAGAGAGAGAGCAC	65	2	0
1		9	Aat CCGAGT			
Off-target	11	1763182	ACAGAAGAGAGAGAGCAC	66	2	0
2		7	Act CCGGGT			
Off-target	8	2650555	ACAGAAGAGAGAGAGCAC	67	2	0

3	5	Act CCGGGT				
Off-target	3	3613708	ACgGAAGAGAAAGAGCAtg	68	5	0
5	3	Gg TCGAGT				

注：PAM序列的核苷酸以粗体书写，潜在脱靶的错配碱基以小写字母显示。

实施例 5 测试 pRABesa-0sU6sa 是否可以用于水稻的多位点碱基编辑

本发明还设计了另一个同时靶向 *OsSPL16* 和 *OsSPL18* 的 0smiR156 结合位
5 点的 sgRNA (sgRNA5) (图 4)。sgRNA5 载体构建所对应的引物为表 3 中的 sgRNA5F 和 sgRNA5R。然而，sgRNA5 的编辑效率远远低于 sgRNA4。在 *OsSPL16* 目标位点只有 17% (8/47) 的株系具有 T-C 替换，在 *OsSPL18* 位点有 23.4% (11/47) 株系发生了编辑 (图 4，A 和 B)。目标位点的扩增引物为表 3 中 SPL16-seq-F/SPL16-seq-R, SPL18-seq-F/SPL18-seq-R。值得注意的是，有
10 14.6% (6/47) 的株系在这两个位点同时发生了 T-C 的替换，进一步证实 pRABesa-0sU6sa 也可以用于水稻的多位点碱基编辑。

实施例 6 测试简化后的腺嘌呤碱基载体 pRSABESp-0sU6 在水稻中的编辑效果

本发明为进一步提高水稻中腺嘌呤碱基编辑的效率，对原有的碱基编辑载体
15 pRABESp-0sU6 (称为腺嘌呤碱基编辑器 ABE-P1) 进行了简化，得到新的腺嘌呤碱基载体 pRSABESp-0sU6 (称为腺嘌呤碱基编辑器 ABE-P1S) (图 5A)。与 pRABESp-0sU6 相比较，在载体 pRSABESp-0sU6 中，我们只把 ecTadA * 7.10 通过编码 32 个氨基酸残基的接头连接到 SpCas9 (D10A) 切口酶编码序列的 N-末端，其它序列没有任何改变。

20 为了测试新的腺嘌呤碱基器 ABE-P1S 在水稻中的编辑效果，我们同样选择 sgRNA1 靶向 *OsSPL14* 的 0smiR156 结合位点序列 (图 5B)，sgRNA1 载体构建所对应的引物为表 6 中的 sgRNA1F 和 sgRNA1R。该二元载体以冻融法转化农杆菌，然后以农杆菌侵染水稻日本晴的愈伤组织，获得了 17 个独立的转基因系。然后，通过 PCR 扩增目标区域，并通过 Sanger 测序进行基因型鉴定，扩增引物为表 10 中
25 SPL14-seq-F 和 SPL14-seq-R。

表10 扩增靶位点及测序所用的引物

引物名称	引物序列 5'-3'	SEQ ID NO.:	目的
------	------------	-------------	----

OsSPL14-F	AGGGTTCCAAGCAGCGTAAGGA	70	扩增所有的 <i>OsSPL14</i> 靶位
OsSPL14-R	TGGTGCTGGGGCTGGACCGTTC	71	点
OsSPL14seq-F	TCTCCGGTGGTATCCAGTGGCAC	72	Sanger 测序引物
SLR1-F	GCGCAATTATTACTAGCTATAGC	73	
SLR1-R	AGCCGTCGCCACCACCGGTAAGG	74	扩增 <i>SLR1</i> 靶位点
SLR1seq-F	CGTCGACGAGCTGCTGGCG	75	Sanger 测序引物
OsSERK2-F	CCATCTGCACGCATGACTT	76	
OsSERK2-R	AGAGTTGTACGTACAATACAGCAG	77	扩增 <i>OsSERK2</i> 靶位点
OsSERK2seq-F	GCTTCGTGGAGCATGAGGT	78	Sanger 测序引物
Tms9-1-F	TGCCGTCGTACCGGCAGTCG	79	
Tms9-1-R	CACGAGGCCTCCGACATGATGC	80	扩增 <i>Tms9-1</i> 靶位点
Tms9-1seq-F	CTCCATGCCGCTCTGCGTGCTG	81	Sanger 测序引物
OsNRT1.1B-F	GGGAGTTCATGATTGGAGGCACG	82	
OsNRT1.1B-R	GTGCTACCTCGCCCAGTGCTCATC	83	扩增 <i>OsNRT1.1B</i> 靶位点
OsNRT1.1Bseq-F	CTGTGACATTTGCAACCACAC	84	Sanger 测序引物
-F			
OsACC1-F	GCAATCTGTTCTCGTGCTGGAC	85	
OsACC1-R	CTTGAAGCGATTTTGTGTCAGCAC	86	扩增 <i>OsACC1</i> 靶位点
OsACC1seq-F	GGCATTGCTGGACTTCAACCG	87	Sanger 测序引物
OsDEP1-F	CTCTTTCACACTCGCTCGCTCG	88	
OsDEP1-R	AGCACAGCACAGTGTCAAGAAC	89	扩增 <i>OsDEP1</i> 靶位点
OsDEP1seq-F	AGGAGATCCGATTCGCCGCG	90	Sanger测序引物
SPX-MSF2-F	AGCACACACACAGACAGTGTC	91	
SPX-MSF2-R	GGCAACTTAGGTTGGCATTGCC	92	扩增 <i>SPX-MSF2</i> 靶位点
SPX-MSF2seq-F	CGCTGCATGCATGCAGGCGCGT	93	Sanger测序引物
F			
OsSPL17-F	GGGTTCCAAGCAGTGTGAGGGA	94	
OsSPL17-R	GGACCTCGTTCAGCACAACCC	95	扩增 <i>OsSPL17</i> 靶位点
OsSPL17seq-F	ATGAACTCCATCCTCACC GCAG	96	Sanger测序引物
OsSPL16-F	TTGGCATCAGCAGCAGGCAGCAG	97	
OsSPL16-R	AGAGACGACGACGTGCCATCCGATG	98	扩增 <i>OsSPL16</i> 靶位点
OsSPL16seq-F	CAGCTGCTGCTGCTGCTTCAGTGTG	99	Sanger测序引物

OsSPL18-F	TGGGATCATCAAATCCGAGGAG	100	扩增 <i>OsSPL18</i> 靶位点
OsSPL18-R	CTGTCCATGCTCGGGCAGGCG	101	
OsSPL18seq-F	CCTCCATGAGGGCGATCAGATA	102	Sanger测序引物
OsSPL13-F	ATTTGACAAGAGTGCATTATGCA	103	扩增 <i>OsSPL13</i> 靶位点
OsSPL13-R	AGCACTGCATATATGTAGCAGCAG	104	
OsSPL13seq-F	GCTTCATCATCATCTTGTC	105	Sanger测序引物
SNB-F	GGTTAGAATCATCCATTTCATG	106	扩增 <i>SNB</i> 靶位点
SNB-R	CCAGCTTTAGTGCCCACTGTTC	107	
SNBseq-F	CTGGTAGCACGGTCACGATTTCG	108	Sanger测序引物

根据测序结果，在 17 株转基因植株中有 12 株在目标位点显示了预期的 T-C 替换，碱基编辑效率高达 70.6% (12/17)。是 pRABEsp-0sU6 在此位点编辑效率的 2.7 倍（对比表 7 和图 1H）。对所有发生碱基替换的位置进行统计后我们发现

5 pRSABEsp-0sU6 可以对 *OsSPL14* protospacer 区第 1, 3, 5, 7, 10, 12 位的腺嘌呤进行编辑，碱基编辑窗口较 pRABEsp-0sU6 有了一定的扩展（图 5C 显示出代表性结果）。

此外，我们还设计了其它的 sgRNAs 靶向水稻基因组不同位点，同时选择了不同的水稻品种（日本晴和 kittake）作为转基因受体材料。结果表明除了 *OsDEPI*

10 靶位点以外，简化后的腺嘌呤碱基编辑器 ABE-P1S 比原来的腺嘌呤碱基编辑器 ABE-P1 有更高的碱基编辑效率（表 7）。

表7 ABE-P1和ABE-P1S在不同靶位点编辑效率的统计

sgRNA	靶基因	碱基编辑器	水稻品种	基因分型的株系数	碱基编辑的株系数	编辑效率
sgRNA1	<i>OsSPL14</i>	ABE-P1	日本晴 (Nipponbare)	48	18	37.5%
		ABE-P1S	日本晴 (Nipponbare)	17	12	70.6%
sgRNA2	<i>SLR1</i>	ABE-P1	日本晴 (Nipponbare)	40	7	17.5%

		ABE-P1S	日本晴	20	7	35%
			(Nipponbare)			
sgRNA3	<i>OsSERK2</i>	ABE-P1	日本晴	39	26	66.7%
			(Nipponbare)			
		ABE-P1S	日本晴	27	21	77.8%
			(Nipponbare)			
sgRNA4	<i>Tms9-1</i>	ABE-P1	Kitaake	22	1	4.5%
		ABE-P1S	Kitaake	36	4	11.1%
sgRNA5	<i>OsNRT1.1</i>	ABE-P1	Kitaake	37	27	72.9%
	<i>B</i>					
		ABE-P1S	Kitaake	54	52	96.3%
sgRNA6	<i>OsACCI</i>	ABE-P1	Kitaake	36	25	69.4%
		ABE-P1S	Kitaake	42	31	73.8%
sgRNA7	<i>OsDEP1</i>	ABE-P1	Kitaake	50	40	80%
		ABE-P1S	Kitaake	35	25	71.4%

为了评估 pRSABESp-0sU6 在水稻中的特异性, 我们对 sgRNA1 的九个潜在脱靶位点进行了测序, 测序结果表明有一个株系在 OsSPL17 基因内的脱靶位点发生了碱基替换, 该脱靶位点和 sgRNA1 在 PAM 上游第 5 位存在 1 个碱基的错配 (图 5D)。

- 5 在其它潜在脱靶位点我们没有检测到任何形式的突变。九个脱靶位点的信息见表 1, 对九个潜在脱靶位点扩增所用的引物见表 3。

实施例 7 测试简化后的腺嘌呤碱基载体 pRSABESa-0sU6Sa 在水稻中的编辑效果

- 10 腺嘌呤碱基编辑载体 pRABESa-0sU6sa (称为腺嘌呤碱基编辑器 ABE-P2) 使用了 SaCas9 (D10A) 切口酶, 它可以识别不同的 PAM 序列 NNGRRT, 这可以扩展水稻腺嘌呤碱基编辑器在水稻基因组中可编辑的目标。

本发明为进一步提高腺嘌呤碱基编辑器 ABE-P2 的编辑效率, 对 pRSABESa-0sU6Sa 载体也进行了简化, 得到新的腺嘌呤碱基载体 pRSABESa-0sU6Sa

(称为腺嘌呤碱基编辑器 ABE-P2S) (图 6A)。与 pRABESa-0sU6Sa 相比较, 在载体 pRSABESa-0sU6Sa 中, 我们只把 ecTadA * 7.10 通过编码 32 个氨基酸残基的接头连接到 SaCas9 (D10A) 切口酶编码序列的 N-末端, 其它序列没有任何改变。

为了比较新的腺嘌呤碱基器 ABE-P2S 与 ABE-P2 在水稻中的编辑效果, 我们同样选择 sgRNA8 靶向水稻 SPX-MFS2 的 OsmiR827 结合位点(图 6B), sgRNA8 所对应的引物为表 6 中的 sgRNA8F 和 sgRNA8R。将 sgRNA8 分别装入 pRABESa-0sU6Sa 和 pRSABESa-0sU6Sa, 将双元载体以冻融法转化农杆菌, 然后以农杆菌侵染水稻日本晴的愈伤组织。对于 pRABESa-0sU6Sa, 我们得到了 41 株转基因阳性苗。通过 PCR 扩增目标区域, 并通过 Sanger 测序进行基因型鉴定, 扩增引物为表 10 中 SPX-MFS2-F 和 SPX-MFS2-R。经鉴定发现有 4 株在靶位点发生了单个碱基 A-G 的替换, 编辑效率为 9.8%(表 8)。发生碱基替换的位点为 protospacer 区第 1, 9 或 15 位腺嘌呤, 但没有株系表现为纯合替换。对于 pRSABESa-0sU6Sa, 我们获得 47 株转基因阳性苗, 其中 7 株在靶位点发生了 A-G 的替换, 编辑效率为 14.9%(表 8), 是 pRABESa-0sU6Sa 在此位点编辑效率的 1.5 倍。除株系 Line 40 外, 其它株系均在 protospacer 区的 3, 6, 9, 15 位发生了单个碱基 A-G 替换, 其中株系 Line 38 在 protospacer 第 6 位表现为纯合替换(图 6C)。

表 8 ABE-P2和ABE-P2S在不同靶位点碱基编辑效率的统计

sgRNA	靶基因	碱基编辑器	背景	基因分型 的株系数	碱基编辑的株 系数	编辑效率
sgRNA8	SPX-MSF2	ABE-P2	日本晴 (Nipponbare)	41	4	9.8%
		ABE-P2S	日本晴 (Nipponbare)	47	7	15.9%
sgRNA9	OsSPL14	ABE-P2S	日本晴 (Nipponbare)	36	20	55.6%
	OsSPL17	ABE-P2S	日本晴 (Nipponbare)	36	22	61.1%

sgRNA1	<i>OsSPL16</i>	ABE-P2S	日本晴	33	10	30.3%
0			(Nipponbare)			
	<i>OsSPL18</i>	ABE-P2S	Nipponbare	33	13	39.4%

此外，我们还设计了其它的 sgRNAs 靶向水稻基因组不同位点。结果表明除了 *OsSPL17* 靶位点以外，简化后的腺嘌呤碱基编辑器 ABE-P2S 比原来的腺嘌呤碱基编辑器 ABE-P2 有更高的碱基编辑效率（对比表 8 与图 1H）。

5

实施例 8 对含 Cas9 蛋白变体的腺嘌呤碱基器 ABE-P5 进行简化后观测其在水稻中的编辑效果

腺嘌呤碱基编辑器 ABE-P5 使用了 SaKKH-Cas9(D10A) 切口酶（图 7A），SaKKH-Cas9(D10A) 是在 SaCas9 (D10A) 中引入了 E782K/N968K/R1015H 三个突变，它可以识别不同的 PAM 序列 NNNRRT。

10

本发明为进一步提高腺嘌呤碱基编辑器 ABE-P5 的编辑效率，对其进行了简化，得到新的腺嘌呤碱基编辑器 ABE-P5S（图 7A）。与 ABE-P5 相比较，在载体 ABE-P5S 中，我们只把 ecTadA * 7.10 通过编码 32 个氨基酸残基的接头连接到 SaKKH-Cas9(D10A) 切口酶编码序列的 N-末端，其它序列没有任何改变。

15

为了比较 ABE-P5 和 ABE-P5S 的编辑效率，我们设计了 sgRNA11 靶向 *OsSPL13* 的 OsmiR156 结合位点（图 7B）。sgRNA11 所对应的引物为表 6 中的 sgRNA11-F 和 sgRNA11-R。将 sgRNA11 分别装入 ABE-P5 和 ABE-P5S，将二元载体以冻融法转化农杆菌，然后以农杆菌侵染水稻日本晴的愈伤组织。对于 ABE-P5，我们共获得 46 株转基因苗。通过扩增引物 0sSPL13-F 和 0sSPL13-R（表 10）对靶位点扩增测序后只发现 1 株转基因苗 Line 3 在 protospacer 第 11 位发生了 A-G 的替换，编辑效率仅为 2.2%（表 9）。而且从 Sanger 测序峰图可以看出这个株系可能只有极少部分细胞在 protospacer 第 11 位发生了 A-G 的替换（图 7C）。对于 ABE-P5S，我们共获得了 33 株转基因苗，其中有 2 株在靶位点发生了 A-G 替换，编辑效率为 6.1%（表 9）。Line 23 在 protospacer 第 7 位发生了 A-G 的替换，而 Line 27 在 protospacer 第 9 位发生了 A-G 的替换（图 7C）。因此在 sgRNA11 靶位点，ABE-P5S 的效率是 ABE-P5 的 2.8 倍。

20

25

此外，我们还设计了其它的 sgRNA12 靶向水稻基因组的 *SNB* 基因。结果表明

简化后的腺嘌呤碱基编辑器 ABE-P5S (33.9%) 的编辑效率是 ABE-P5 (6.5%) 的 5.2 倍 (表 9)。

表9 ABE-P5和ABE-P5S在不同靶位点碱基编辑效率的统计

sgRNA	靶基因	碱基编辑器	背景	基因分型 的株系数	碱基编辑的株 系数	编辑效率
sgRNA1 1	OsSPL13	ABE-P5	日本晴 (Nipponbare)	46	1	2.2%
		ABE-P5S	日本晴 (Nipponbare)	33	2	6.1%
sgRNA1 2	SNB	ABE-P5	日本晴 (Nipponbare)	46	3	6.5%
		ABE-P5S	日本晴 (Nipponbare)	62	21	33.9%

5 对比例 1

方法同实施例 1，区别在于，用水稻 Actin 启动子驱动 Cas9 核酸酶、核定位信号 VirD2 和腺嘌呤脱氨酶的表达，用依赖于 II 型 RNA 聚合酶的启动子或者 U3 启动子驱动 sgRNA 的转录。

结果表明，在水稻中的编辑效率（即定点替换效率）仅为实施例 1 的 50%。

10

对比例 2

方法同实施例 1，区别在于，用 SV40 核定位信号替换 VirD2。

结果表明，在水稻中的编辑效率（即定点替换效率）仅为实施例 1 的约 50%-70%。

15

参考文献

1. Komor, A. C., Badran, A. H., and Liu, D. R. (2017a). Editing the Genome

Without Double-Stranded DNA Breaks. ACS Chem Biol. Advance Access published September 28, 2017, doi: 10.1021/acscchembio.7b00710.

2. Nishimura, A., Aichi, I., and Matsuoka, M. (2007). A protocol for Agrobacterium-mediated transformation in rice. Nat. Protoc. 1:2796-2802.

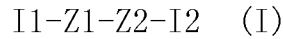
5

在本发明提及的所有文献都在本申请中引用作为参考，就如同每一篇文献被单独引用作为参考那样。此外应理解，在阅读了本发明的上述讲授内容之后，本领域技术人员可以对本发明作各种改动或修改，这些等价形式同样落于本申请所附权利要求书所限定的范围。

权 利 要 求

1. 一种核酸构建物，其特征在于，所述核酸构建物具有 5' -3' (5' 至 3') 的式 I 结构：

5



式中，

I1 为第一整合元件；

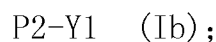
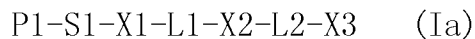
I2 为第二整合元件；

10

Z1 为第一表达盒；

Z2 为第二表达盒；

并且，Z1 和 Z2 中的一个表达盒具有 Ia 结构，而另一个表达盒具有式 Ib 结构：



15

式中，

P1、S1、X1、L1、X2、L2、X3、P2、Y1 分别为用于构成所述构建物的元件；

P1 为第一启动子，所述第一启动子包括泛素启动子；

S1 为无或信号肽的编码序列；

X1 为腺嘌呤脱氨酶 (如野生型和/或突变型 TadA) 的编码序列；

20

L1 为无或第一连接肽的编码序列；

X2 为 Cas9 核酸酶的编码序列，所述的 Cas9 核酸酶是无切割活性或单链切割活性的；

L2 为无或第二连接肽的编码序列；

X3 为核定位信号的编码序列，所述核定位信号为 VirD2；

25

P2 为第二启动子；

Y1 为 sgRNA 的编码序列；

并且，各“-”为键或核苷酸连接序列。

2. 如权利要求 1 所述的核酸构建物，其特征在于，所述泛素启动子包括玉米泛素启动子。

30

3. 如权利要求 1 所述的核酸构建物，其特征在于，所述第二启动子包括 U6 启动子。

4. 如权利要求 1 所述的核酸构建物, 其特征在于, 所述的 sgRNA 的第 5 位-第 10 位对应于预定发生 T→C 定点突变的位置(即为 T)。

5. 如权利要求 1 所述的核酸构建物, 其特征在于, 所述的 sgRNA 的第 6-14 位对应于预定发生 T→C 定点突变的位置(即为 T)。

5 6. 如权利要求 1 所述的核酸构建物, 其特征在于, 所述信号肽包括 VirD2 的核定位信号肽。

7. 如权利要求 1 所述的核酸构建物, 其特征在于, 所述第一启动子来源于选自下组的一种或多种植物: 玉米、水稻、大豆、拟南芥、番茄。

10 8. 如权利要求 1 所述的核酸构建物, 其特征在于, 所述第二启动子来源于选自下组的一种或多种植物: 水稻、玉米、大豆、拟南芥、番茄。

9. 一种载体, 其特征在于, 所述载体含有权利要求 1 所述的核酸构建物。

10. 一种基因工程细胞, 其特征在于, 所述细胞含有权利要求 1 所述的核酸构建物, 或其基因组整合有一个或多个权利要求 1 所述的核酸构建物。

11. 一种对植物进行基因编辑的方法, 其特征在于, 包括步骤:

15 (i) 提供待编辑植物; 和

(ii) 将权利要求 1 所述的核酸构建物或权利要求 9 所述的载体导入所述待编辑植物的植物细胞, 从而在所述植物细胞内进行基因编辑。

12. 一种制备转基因植物细胞的方法, 其特征在于, 包括步骤:

20 (i) 将权利要求 1 所述的核酸构建物、或权利要求 9 所述的载体转染植物细胞, 使得所述核酸构建物与所述植物细胞中的染色体发生定点替换, 从而制得所述转基因植物细胞。

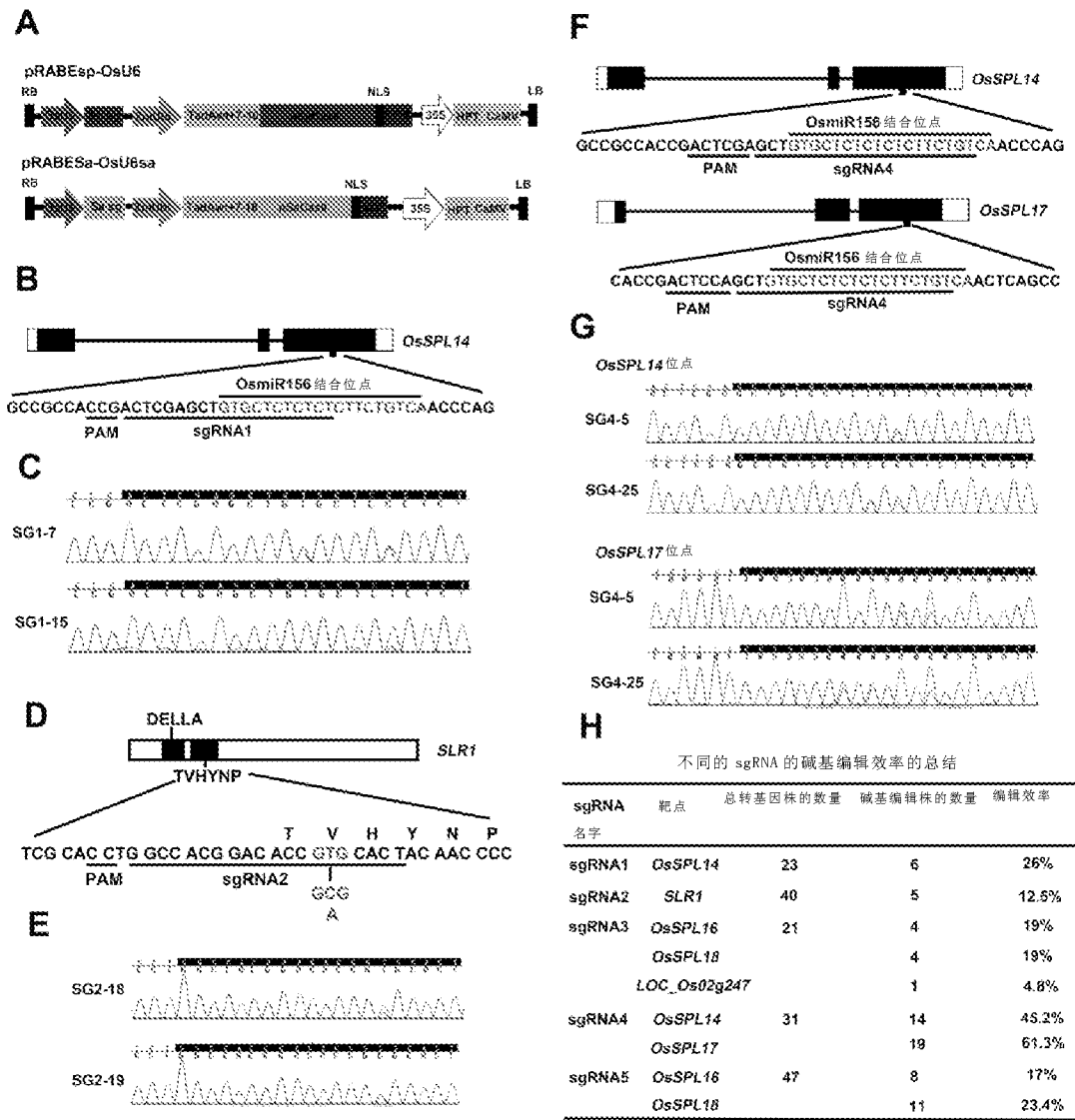
13. 一种制备转基因植物细胞的方法, 其特征在于, 包括步骤:

(i) 将权利要求 1 所述的核酸构建物、或权利要求 9 所述的载体转染植物细胞, 使得所述植物细胞含有所述核酸构建物, 从而制得所述转基因植物细胞。

25 14. 一种制备转基因植物的方法, 其特征在于, 包括步骤:

将权利要求 12 或权利要求 13 所述方法制备的所述

转基因植物细胞再生为植物体, 从而获得所述转基因植物。



图

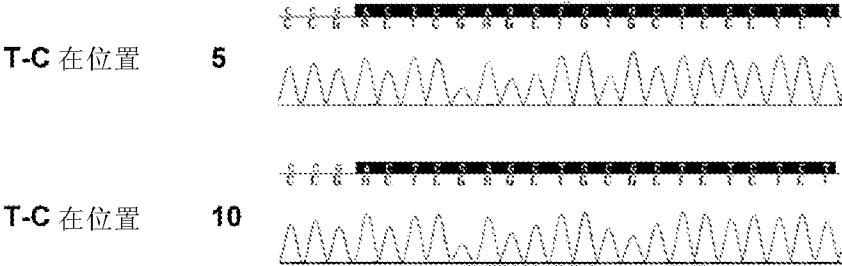
1

OsSPL14 SG1-7

参考： 5'-GCCGCCACCGACTCGAGCTGTGCTCTCTCTCTTCTGT-3'

T-C at position 5: GCCGCCACCGACTCGAGCTGTGCTCCCTCTCTTCTGT 9/20

T-C at position 10: GCCGCCACCGACTCGAGCTGCGCTCTCTCTCTTCTGT 11/20



OsSPL14 SG1-15

参考： 5'-GCCGCCACCGACTCGAGCTGTGCTCTCTCTCTTCTGT-3'

WT GCCGCCACCGACTCGAGCTGTGCTCTCTCTCTTCTGT 17/20

T-C 在位置 5,7 GCCGCCACCGACTCGAGCTGTGCCCCCTCTCTTCTGT 3/20

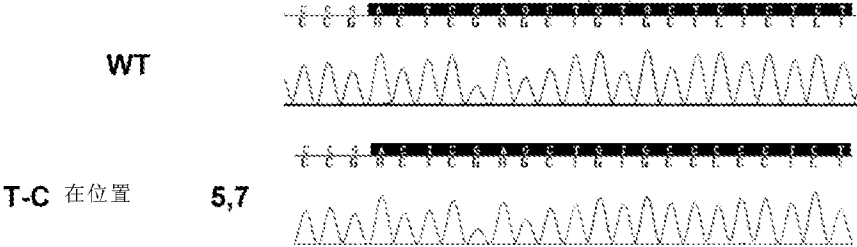


图 2

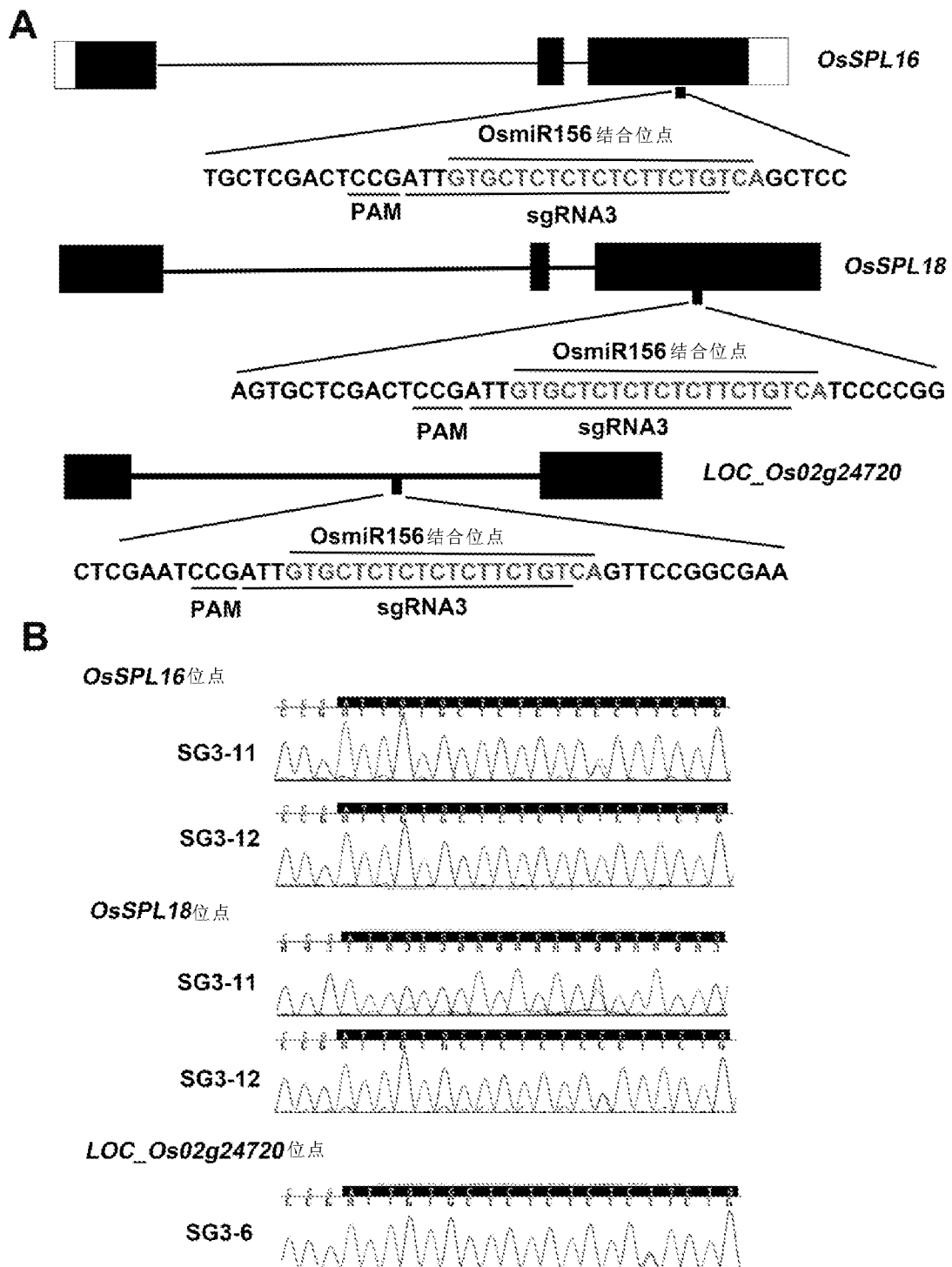


图 3

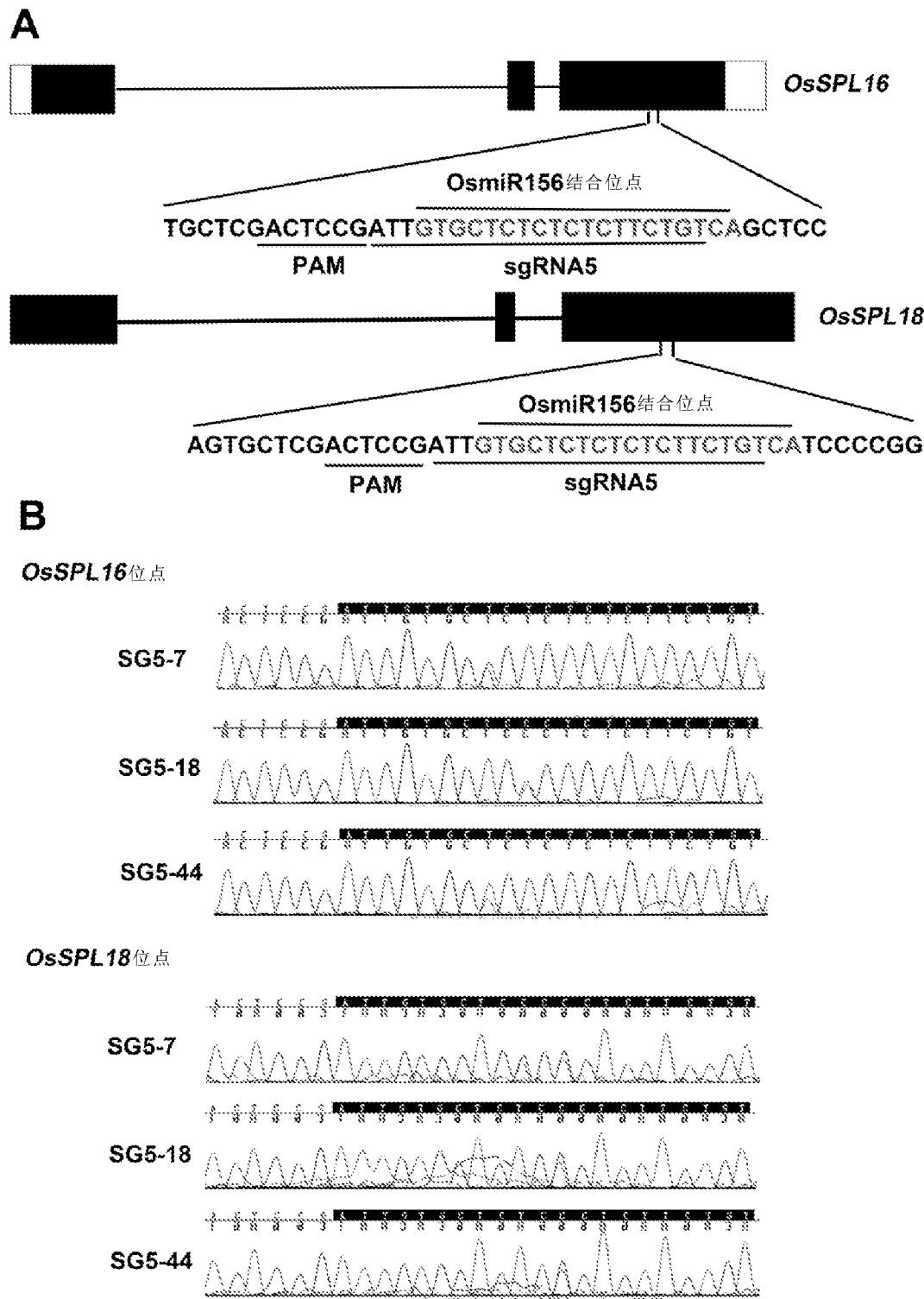


图 4

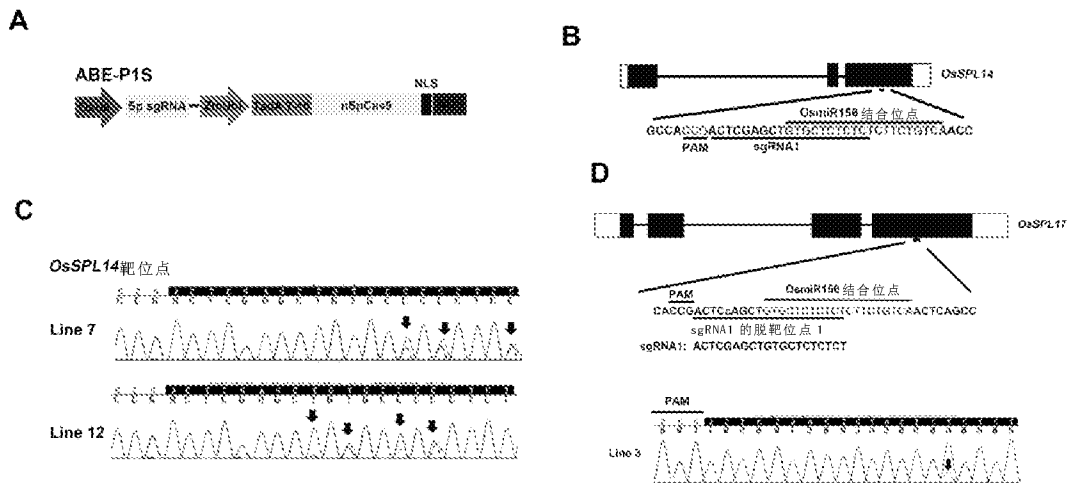


图 5

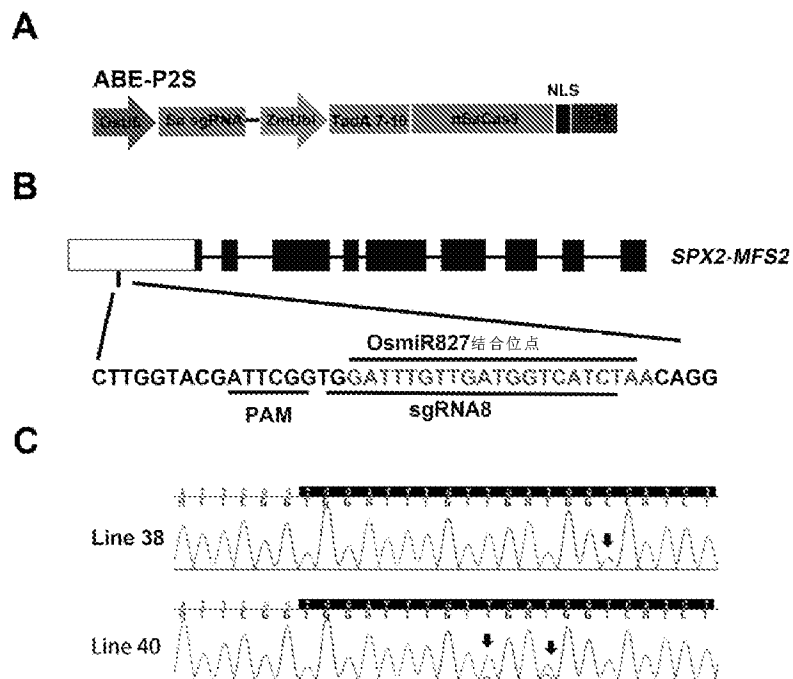


图 6

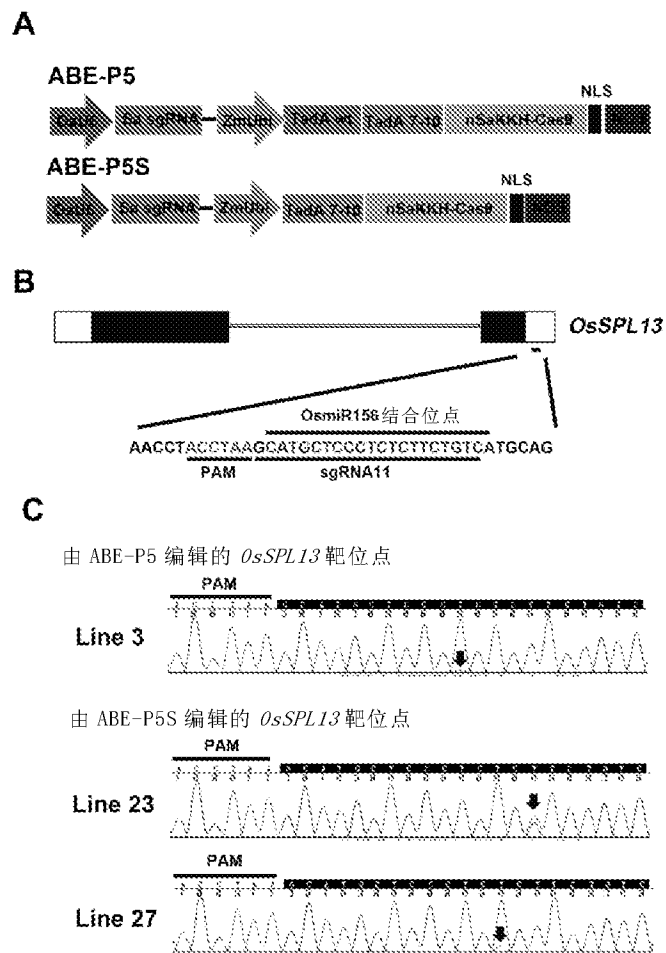


图 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/122014

A. CLASSIFICATION OF SUBJECT MATTER

C12N 15/82(2006.01)i; A01H 5/00(2018.01)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 A01H

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; CNTXT; VEN; WOTXT; USTXT; EPTXT; CNKI; 万方, WANFANG; 读秀学术, DUXIU ACADEMIC; Springer; ISI Web of Science: 朱健康, 华凯, 核酸, 腺嘌呤脱氨酶, 腺苷, 脱氨, 酶, ZHU Jiankang, HUA Kai, nucleic acid, adenine deaminase, adenine, enzyme, TadA, cas9, gRNA

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 106609282 A (SHANGHAI INSTITUTES FOR BIOLOGICAL SCIENCES, CHINESE ACADEMY OF SCIENCES) 03 May 2017 (2017-05-03) description, paragraphs [0006]-[0041], [0058], [0066], [0071]-[0073], [0080] and [0082]-[0085]	1-14
Y	GAUDELLI, N.M. et al. "Programmable Base Editing of A•T to G•C in Genomic DNA without DNA Cleavage" <i>Nature.</i> , Vol. 551, No. (7681), 23 November 2017 (2017-11-23), abstract, and page 14, paragraph 1	1-14



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"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)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"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

"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

"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

"&" document member of the same patent family

Date of the actual completion of the international search

05 March 2019

Date of mailing of the international search report

18 March 2019

Name and mailing address of the ISA/CN

State Intellectual Property Office of the P. R. China
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Authorized officer

Facsimile No. (86-10)62019451

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/122014

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 13*ter*.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 13*ter*.1(a)).
 - ☐ on paper or in the form of an image file (Rule 13*ter*.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/CN2018/122014

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	106609282	A	03 May 2017	WO	2018098935	A1	07 June 2018

国际检索报告

国际申请号

PCT/CN2018/122014

A. 主题的分类 C12N 15/82(2006.01)i; A01H 5/00(2018.01)i 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											
B. 检索领域 检索的最低限度文献(标明分类系统和分类号) C12N A01H 包含在检索领域中的除最低限度文献以外的检索文献 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)) CNABS;CNTXT;VEN;WOTXT;USTXT;EPTXT;CNKI;万方;读秀学术;Springer;ISI Web of Science; 朱健康, 华凯, 核酸, 腺嘌呤脱氨酶, 腺苷, 脱氨, 酶, ZHU Jiankang, HUA Kai, nucleic acid, adenine deaminase, adenine, enzyme, TadA, cas9, gRNA											
C. 相关文件 <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>CN 106609282 A (中国科学院上海生命科学研究院) 2017年 5月 3日 (2017 - 05 - 03) 说明书第[0006]-[0041]、[0058]、[0066]、[0071]-[0073]、[0080]、[0082]-[0085]段</td> <td>1-14</td> </tr> <tr> <td>Y</td> <td>Nicole M. Gaudelli 等. "Programmable Base Editing of A•T to G•C in Genomic DNA without DNA Cleavage" Nature., 第551卷, 第7681期, 2017年 11月 23日 (2017 - 11 - 23), 摘要, 第14页第1段</td> <td>1-14</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	CN 106609282 A (中国科学院上海生命科学研究院) 2017年 5月 3日 (2017 - 05 - 03) 说明书第[0006]-[0041]、[0058]、[0066]、[0071]-[0073]、[0080]、[0082]-[0085]段	1-14	Y	Nicole M. Gaudelli 等. "Programmable Base Editing of A•T to G•C in Genomic DNA without DNA Cleavage" Nature., 第551卷, 第7681期, 2017年 11月 23日 (2017 - 11 - 23), 摘要, 第14页第1段	1-14
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求									
Y	CN 106609282 A (中国科学院上海生命科学研究院) 2017年 5月 3日 (2017 - 05 - 03) 说明书第[0006]-[0041]、[0058]、[0066]、[0071]-[0073]、[0080]、[0082]-[0085]段	1-14									
Y	Nicole M. Gaudelli 等. "Programmable Base Editing of A•T to G•C in Genomic DNA without DNA Cleavage" Nature., 第551卷, 第7681期, 2017年 11月 23日 (2017 - 11 - 23), 摘要, 第14页第1段	1-14									
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。											
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “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) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “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 “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 “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 “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 “&” 同族专利的文件											
国际检索实际完成的日期 2019年 3月 5日		国际检索报告邮寄日期 2019年 3月 18日									
ISA/CN的名称和邮寄地址 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 传真号 (86-10)62019451		受权官员 李小晶 电话号码 (86-512)88996463									

国际申请号
PCT/CN2018/122014

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 106609282 A	2017年 5月 3日	W0 2018098935 A1	2018年 6月 7日