

# 基因编辑新技术最新进展

张梦娜 柯丽萍 孙玉强\*

(浙江理工大学生命科学学院, 植物基因组学与彩色纤维分子改良实验室, 杭州 310018)

**摘要** 借助基因编辑技术精准编辑植物基因组, 得到性状优良、产量高的农作物种质是目前作物分子育种研究的主要趋势。目前主要的CRISPR/Cas9系统是由产脓链球菌的获得性免疫防御系统改编而来, 该系统以其编辑高效、操作方便、成本低廉等明显优势在基因编辑技术中脱颖而出, 广受青睐。利用CRISPR/Cas技术编辑作物基因组, 能精确引入和改良目标性状, 为作物遗传育种提供新途径。当前, CRISPR/Cas9技术在拟南芥、水稻、土豆、玉米等植物中得到普遍应用。该文简要阐述了锌指核酸酶、转录因子激活样效应物核酸酶以及CRISPR/Cas9系统的结构、作用机制及差异, 重点综述CRISPR/Cas9系统目前在植物中的应用、其改良的CRISPR/Cpf1技术以及该系统相比于其他核酸酶的优势与局限性。

**关键词** 基因编辑; 锌指核酸酶; 转录因子激活样效应物核酸酶; CRISPR/Cas9; CRISPR/Cpf1

## Progress of CRISPR/Cas9 System for Gene Editing

Zhang Mengna, Ke Liping, Sun Yuqiang\*

(Laboratory of Plant Genome and Coloured Fiber Molecular Improvement, School of Life Sciences, Zhejiang Sci & Tech University, Hangzhou 310018, China)

**Abstract** Accurate editing of plant genomes using genetic editing techniques is a very important method for plant molecular breeding to improve crop traits and yields. The CRISPR/Cas9 system is originated from the acquired immune defense system of *Pseudomonas aeruginosa*. Because of its high efficiency, easy operation and low cost, the system stands out from many gene editing technologies and is widely favored by researchers. Using CRISPR/Cas9 technology to edit crop genomes for accurately introducing or improving target traits provides a new approach to crop genetics and breeding. It has been reported that CRISPR/Cas9 technology has been widely used in *Arabidopsis*, rice, potato, corn and other crops. This paper briefly describes the structure, mechanism and differences of zinc finger nuclease (ZFN), transcriptional activation-like effector nuclease (TALEN), CRISPR/Cas9 system, focusing on the progress of the CRISPR/Cas9 system, its improved CRISPR/Cpf1 technology and its advantages as well as limitations compared to other nucleases.

**Keywords** genetic editing; ZFNs; TLENs; CRISPR/Cas9; CRISPR/Cpf1

在人口急剧增长、耕地面积不增反减的社会大背景下, 培育高产、优质、多抗、高效吸收利用等优良性状的农作物来满足人们不断增长的粮食需求

是目前研究的趋势之一, 而传统的育种方式周期长, 耗费大量的人力、物力, 这促使研究者致力于寻找更简单、高效的育种手段。于是, 利用基因编辑技

收稿日期: 2018-08-06 接受日期: 2018-09-25

国家自然科学基金(批准号: 31671738)和浙江省杭州市科技发展计划(批准号: 20160432B06)资助的课题

\*通讯作者。0571-86843335, E-mail: sunyuqiang@zstu.edu.cn

Received: August 6, 2018 Accepted: September 25, 2018

This work was supported by the National Natural Science Foundation of China (Grant No.31671738) and Hangzhou Science and Technology Development Plan of Zhejiang Province (Grant No.20160432B06)

\*Corresponding author. Tel: +86-571-86843335, E-mail: sunyuqiang@zstu.edu.cn

网络出版时间: 2018-11-26 16:58:09 URL: <http://kns.cnki.net/kcms/detail/31.2035.Q.20181126.1658.002.html>

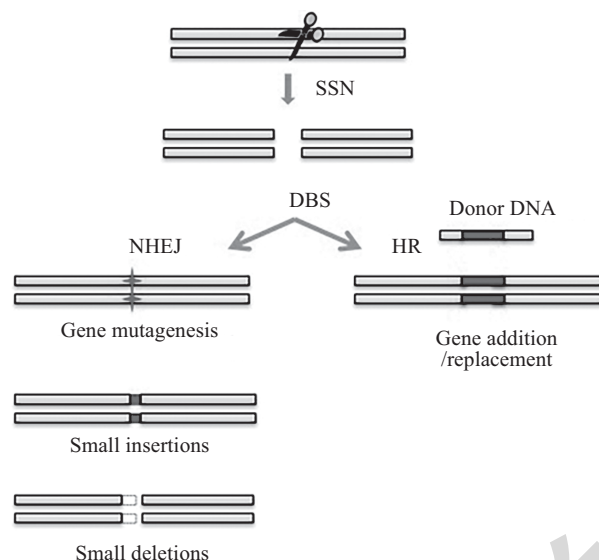


图1 特定核酸酶引起双链断裂激活非同源末端连接或同源重组的DNA修复机制(根据参考文献[2]修改)  
Fig.1 DNA repair mechanism by SSN causing DSB to activate NHEJ or HR (modified from reference [2])

术改良农作物的育种方式应运而生。

基因组编辑技术结合电穿孔、农杆菌介导的遗传转化等方法对植物基因组的特定位点进行插入、替换、碱基删除等编辑,实现跨物种或物种内遗传信息快速高效的编辑和整合,进而快速获得目标性状改良的突出的优良品系,对作物种质资源创新与新品种培育举足轻重。为此,特殊的酶如限制性核酸内切酶和连接酶迅速兴起。然而,使用限制性内切核酸酶和连接酶在编辑较大生物的基因组,包括较大和相对复杂的植物基因组是非常困难的。传统的基因编辑技术如限制性内切核酸酶只能“靶向”相对较短的DNA序列。虽然这种特异性对于短DNA病毒和细菌是足够的,但它很难对具有复杂和较大基因组的植物进行基因编辑,因此,创建简单便捷地能编辑复杂基因组的“人造酶”迫在眉睫<sup>[1]</sup>。

近些年,位点特异性核酸酶(site-specific nuclease, SSNs)在精确编辑模式植物和重要作物基因组方面取得了很大的进展。位点特异性核酸酶是指对基因组的特定核苷酸位点进行切割,在动植物基因组编辑过程中引入双链缺口的工具酶<sup>[2]</sup>。到目前为止,基因编辑创新浪潮主要涉及四种SSNs: 锌指核酸酶(zinc finger nuclease, ZFN)、转录激活样效应物核酸酶(transcription activator-like effector nucleases, TALENs)、RNA引导的规律成簇间隔短回文重复序列(clustered regularly interspaced short palindromic

repeats, CRISPR)/Cas9及通过CRISPR/Cas9改进的CRISPR/Cpf系统。这些SSNs的基因编辑策略分为两类: 随机插入突变与同源重组突变。SSNs在靶位点切割双链形成缺口后,在没有同源模板链的情况下,双链缺口(double stranded break, DSB)引发生物内源性基因修复机制,即非同源末端连接(non-homologous end joining, NHEJ)。NHEJ对DSB处进行不完全修复,发生基因缺失、插入或替换,从而达到编辑基因的目的。然而NHEJ并不是一种忠实的DNA双链断裂修复机制,因为它不需要重组缺口之间有严格的同源性<sup>[3]</sup>。而同源重组修复(homology-directed repair, HR)会将与SSNs共转入细胞的外源同源片段复制到缺口上,进而整合目的基因到靶位点上(图1)<sup>[4]</sup>。在此基础上,人们可以对基因组的特定位点进行各种遗传操作,包括基因打靶、基因定点插入、基因修复等,从而能够方便快捷地对基因组实现靶向遗传修饰。基因编辑技术成为研究基因功能的一条方便、有效的途径,这也进一步对人类疾病治疗、植物种质改良等方面的研究具有重要的生物学意义。在此简要归纳和介绍不同基因组编辑系统及其作物改良的机制,并分析比较这些核酸酶的各自优点与局限性,以便更好地应用基因编辑技术改造生物遗传性状。

## 1 锌指核酸酶

锌指核酸酶作为第一代基因组编辑技术,是

由锌指蛋白结构域与核酸酶(通常为FokI核酸酶)切割结构域两部分组成。锌指蛋白结构域是由3~6个Cys2-His2或Cys4锌指蛋白串联组成, 它可特异性识别靶基因位点<sup>[5]</sup>。每个锌指蛋白由约30个氨基酸构成一个 $\alpha$ -螺旋和两个反平行的 $\beta$ -折叠三个肽段, 锌指N-端的一对半胱氨酸残基与C-端的一对组氨酸残基, 形成恰好容纳一个 $Zn^{2+}$ 的空间结构,  $\alpha$ -螺旋镶嵌于DNA的大沟中, 特异性识别DNA双螺旋中一条单链上的三个连续核苷酸, 这三个连续的核苷酸称为一个三联子<sup>[6]</sup>。ZFN的另一个组分FokI来源于海床黄杆菌的一种非特异性核酸内切酶<sup>[7]</sup>, 能特异性识别双链DNA上5'-GGATG-3'与5'-CATCC-3'序列, 一般在距离两个DNA结合位点间隔区5~7 bp处进行切割, 然后产生双链断裂。由于FokI核酸酶是作为二聚体发挥作用的, 因此, 设计成对的锌指结构域以结合切割位点的上游和下游, 从而将完整ZFN的特异性增加到24~36个核苷酸<sup>[8]</sup>。采用ZFN技术已在多种动植物(如玉米<sup>[9]</sup>、水芹<sup>[10]</sup>、大豆<sup>[11]</sup>、小鼠<sup>[12]</sup>和果蝇<sup>[13]</sup>等)中实现了基因修饰。

然而, 设计成对的锌指结构域以结合切割位点的单一锌指蛋白组合成的ZFN, 并非同样能特异性识别与高效率打靶, 除此之外, ZFN基因编辑技术从设计ZFP到后期实验操作耗时又费力。后基因组时代号召新一代省时便利的基因编辑技术。

## 2 转录激活样效应物核酸酶

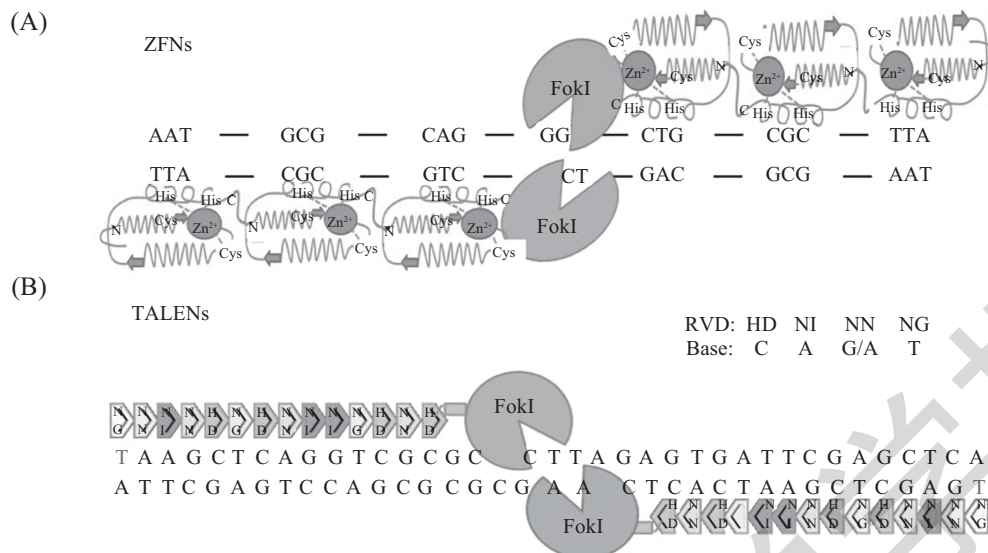
植物病原菌黄单胞菌入侵宿主细胞时释放的一种蛋白效应子, 即TALE(transcriptional activation-like effector), 它能特异性地影响宿主细胞基因的转录。随着TALE特异性结合DNA的规律被破解后, 研究者利用该规律进一步人工改造了TALENs并迅速投入应用, 成为基因编辑的热门工具酶。TALENs是由两条TALE-FokI臂组成, 首先若干个氨基酸重复单元通过尾对尾的方式连接成TALE臂, 在一条TALE臂上连接上一个FokI非特异性核酸酶, 形成一条能特异性结合靶DNA序列的TALE-FokI臂, 即TALEN, 基于FokI是以二聚体的形式发挥作用的, 因此设计两条TALE-FokI臂共同导入细胞内特异性结合并切割靶DNA序列, 从而产生双链断裂(double-strand break, DSB), 且TALEN结合的DNA序列大多5'端为胸腺嘧啶<sup>[14]</sup>。TALE具有特殊的N末端的转运信号, C末端存在核定位信号(nucleus location signal,

NLS)和转录激活结构域(activation domain, AD), 中间是DNA结合结构域(DNA binding domain, DBD)。中央DBD通常由14~20个氨基酸重复单元串联而成, 决定了其特异性<sup>[15]</sup>。每个氨基酸重复单元由33~35个高度同源的氨基酸组成(34-aa重复序列), 第12位与第13位氨基酸是TALEN特异性识别靶序列的关键氨基酸, 被称为重复序列可变的双氨基酸残基(repeat variable diresidues, RVD)<sup>[16]</sup>, 除了识别C碱基需要特定的氨基酸组合外, 第12位氨基酸固定不变, 稳定分子结构, 通过改变34-aa重复序列的第13位氨基酸发挥特异性识别靶序列功能, HD、NI、NN、NG分别识别C、A、G/A、T(图2)<sup>[17-18]</sup>。目前研究者发现, 许多黄单胞菌细菌使用转录激活物样效应物(TALE)蛋白能激活植物疾病易感性(S)基因, 研究其转录激活机制为提高作物对携带TALE的病原体的抗性提供坚实可靠的理论基础, 同时也能较大程度地改良作物, 创造种质资源<sup>[19-20]</sup>。

## 3 CRISPR/Cas9系统

### 3.1 CRISPR/Cas系统的分类和基本结构

CRISPR/Cas9序列在古细菌与细菌中广泛存在, CRISPR/Cas9系统是细菌的一种获得性免疫防御系统, 防止外源DNA或病毒再次入侵。CRISPR最早是1987年在大肠杆菌*K12 iap*基因侧翼序列中被发现的<sup>[21]</sup>; 2002年被命名为串联间隔短回文重复序列(clustered regularly interspaced short palindromic repeats, CRISPR)<sup>[22]</sup>; 至2011年, 其获得性免疫防御机制基本被研究清楚<sup>[23]</sup>并迅速在众多编辑技术中脱颖而出, 被广泛接受与应用。目前, 根据CRISPR/Cas系统作用机制与Cas蛋白的差异性, 分为三个类型: Type I、Type II、Type III<sup>[24]</sup>。I型CRISPR/Cas的显著蛋白是Cas3, 具有核酸酶和解旋酶特性<sup>[25]</sup>, Cas10是III型CRISPR/Cas系统的功能蛋白, 具有RNA酶活性<sup>[26]</sup>。I型与III型CRISPR/Cas系统, 都是由特定Cas蛋白加工出crRNA(CRISPR RNA), crRNA再与多个Cas蛋白复合体识别并剪切外源DNA<sup>[27-28]</sup>, 由于其涉及多个Cas蛋白, 研究难度大。而来自产脓链球菌II型CRISPR/Cas系统切割靶位点只涉及一个核酸内切酶Cas9蛋白, 其含有HNH核酸酶结构域与RuvC-like结构域, 降解外源DNA与病毒核酸, 所以II型CRISPR/Cas9系统应用最多。除Cas9蛋白酶外, CRISPR/Cas9系统还包含5'端的反式激活CRISPR序



A: 锌指核酸酶ZFNs; FokI: 非特异性核酸酶; B: 转录激活样效应物核酸酶TALENs, RVD为HD、NI、NN、NG时能分别识别C、A、G/A、T。  
A: zinc finger nuclease ZFNs; FokI: non-specific nuclease; B: transcriptional activation-like effector nuclease TALENs, RVD for HD, NI, NN, NG can identify C, A, G/A, T, respectively.

图2 两种位点特异性核酸酶的示意图

Fig.2 Schematic representation of two site-specific nucleases

列(trans-acting CRISPR, trasCRISPR)和3'端CRISPR基因座, CRISPR基因座又包括启动转录的300~500 bp前导序列、多个短而高度保守的21~47 bp重复序列以及具有识别外源DNA作用的26~72 bp非重复性的间隔序列三个部分<sup>[29]</sup>。

### 3.2 CRISPR/Cas9系统免疫防御过程

CRISPR/Cas9系统免疫防御外源DNA入侵分为以下三个阶段<sup>[30-31]</sup>。(1)适应阶段, 外源DNA或噬菌体在第一次入侵细菌后, 其核酸序列信息存入到CRISPR基因座前导序列与第一个重复序列之间, 形成间隔序列。(2)表达阶段, CRISPR序列转录形成未成熟的pre-crRNA(pre-CRISPR-derived RNA), 在pre-crRNA转录的同时, 与其重复序列互补的反式激活crRNA(trans-acting CRISPR RNA, tracrRNA)也被转录出来。Pre-crRNA与tracrRNA相互作用形成复合物, 激发RNA内切酶III切割后, 形成成熟的crRNA: tracrRNA: Cas9复合物。(3)干扰阶段, 成熟的crRNA: tracrRNA复合物沉默靶基因, crRNA: tracrRNA复合物通过碱基互补配对识别并结合与crRNA互补的序列, 引导Cas9核酸内切酶切割靶序列。Cas9的HNH结构域切割与crRNA的互补DNA单链, RuvC-like结构域切割非互补链<sup>[2]</sup>。CRISPR/Cas9系统切割靶序列的先决条件是在入侵DNA下游存在一段保守的原间隔序列基序(protospacer adjacent

motif, PAM), crRNA: tracrRNA复合物识别靶位点3'端的PAM通常为5'-NGG-3'序列<sup>[32]</sup>, 三元复合物与靶位点3'端的PAM邻近的20个核苷酸互补结合, 引导Cas9切割靶位点, HNH负责切割crRNA互补链PAM上游3 nt处, 另一条链由RuvC切割, 切割位点为PAM上游3~8 bp处<sup>[33]</sup>, 产生DSB(图3)。

### 3.3 ZFN、TALENs与CRISPR/Cas9系统相比的特异性与局限性

2012年, Jinek等<sup>[34]</sup>在体外证明了tracrRNA: crRNA复合物通过退火产生的嵌合单分子引导RNA(single guide RNA, sgRNA)与Cas9的靶向切割能力, 进一步将CRISPR/Cas9系统组成原件减少到两种, 即gRNA和Cas9基因。这一优化为深入应用CRISPR/Cas9系统进行基因编辑打下了坚实的基础。在后续的发展中, 分别将装载有Cas9与gRNA的两个载体共转导入靶细胞, gRNA与靶DNA序列PAM上游20个碱基互补配对, 待编辑的区域附近需要存在相对保守的PAM序列(NGG), 引导Cas9核酸内切酶到靶位点, 切割形成DSB状态, 引发生物体非同源末端连接基因损伤修复机制。当存在同源序列的情况下, DNA即可能被同源修复(homologous recombination, HR), 该机制可用于精确地修复或插入基因。

对比于ZFN技术与TALENs技术重复单元设计困难、成本高、实验操作难等问题, CRISPR/Cas9技

术具有应用范围广、设计方便、成本低、操作简单、编辑效率高等优势(表1)。目前, CRISPR/Cas9作为新时代的作物基因组编辑技术, 对作物产量和抗性等相关基因的修饰和编辑来提高相应的产量和抗性等特性有着不可替代的作用, 已经成功应用于水稻<sup>[35]</sup>、小麦<sup>[36]</sup>、土豆<sup>[37]</sup>、番茄<sup>[38]</sup>、棉花<sup>[39]</sup>、玉米<sup>[40]</sup>等多种作物中, 其研究最为广泛与深入的作物当数历史悠久的粮食作物——水稻。从2013年至

2017年, 研究者先后针对水稻*OsPDS*、*OsMPK2*、*OsBADH2*、*Os02g23823*、*LOC\_Os10g05490*等几十个功能基因设计sgRNA, 通过农杆菌转化愈伤组织或原生质体或者电穿孔等方法, 对其基因组进行精准编辑。对Cas9与sgRNA分别优化其启动子及转化方法, 其突变率甚至达到80%, 为水稻种质资源创新、遗传性状改良提供了有力、有效的工具<sup>[41]</sup>。在土豆、棉花、玉米等作物中, CRISPR/Cas9系统同

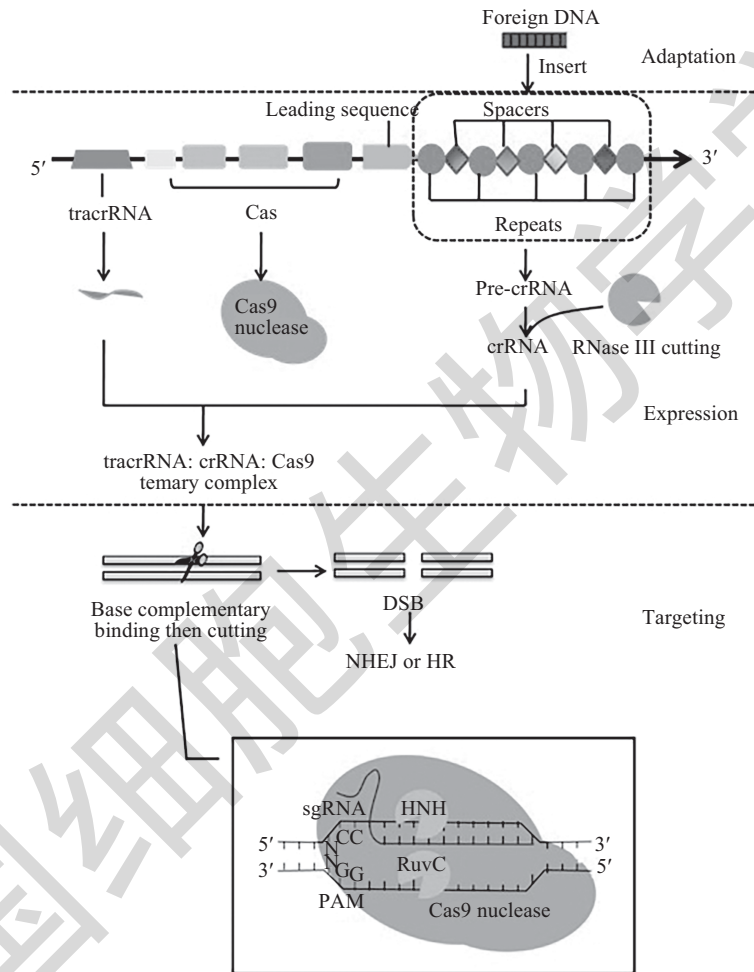


图3 CRISPR/Cas9系统的作用机制模式图

Fig.3 Schematic diagram of the mechanism of action of the CRISPR/Cas9 system

表1 ZFN、TALEN与CRISPR编辑技术比较

Table 1 Comparison of ZFN, TALEN and CRISPR editing techniques

评估项 Evaluation item	锌指核酸酶 ZFN	转录激活样效应物核酸酶 TALEN	CRISPR/Cas9系统 CRISPR/Cas9 system
Core component	ZFA-FokI	TALE-FokI	sgRNA and Cas9
Mode of action	Pair	Pair	Not pair
Design difficulty	Difficult	Easy	Very easy
Cost	High	Moderate	Low
Editing efficiency	Unstable	High	High
Off-target rate	High	Low	High

样表现不俗。2017年, Janga等<sup>[39]</sup>针对棉花中单拷贝的*GFP*基因设计sgRNA, 通过对农杆菌转化幼苗切下的下胚轴、子叶柄和子叶外植体进行研究发现, 靶基因编辑后突变率高达43%。而CRISPR/Cas9技术在土豆中的应用早在2014年已由王晓辉等<sup>[37]</sup>完成。同年, 梁振等<sup>[40]</sup>针对玉米中两个靶向*ZmIPK*基因构建gRNA, 转化原生质体, 靶基因突变频率分别为16.4%和19.1%。

毋庸置疑, CRISPR/Cas9系统的发现掀起了基因编辑技术的新浪潮。ZFN技术、TLNAs技术目前尚无法实现对任意一段DNA序列进行基因编辑, 并且脱靶效应引起的细胞毒性等缺点使ZFN与TLNAs的应用受到很大限制。而应用CRISPR/Cas9技术对基因组进行编辑时, 研究者可通过设计与靶序列同源互补的gRNA, 达到对靶序列准确定位, 进而Cas9对靶序列准确切割的目的。然而, CRISPR/Cas9系统并非完美无缺, 它同样存在诸多亟待解决的问题: 其一, 由于CRISPR/Cas9系统通过gRNA与PAM上游的前18~22个碱基互补配对特异性识别靶序列, 起主要作用的是PAM上游的8~10 bp, 而远离PAM的10~12 bp与gRNA不完全配对则对靶向切割影响不大, 由此出现碱基错配<sup>[42]</sup>, 引起基因序列被非特异性切割, 因此产生细胞毒性, 需进一步提升CRISPR/Cas9系统对基因组的精确编辑能力, 降低脱靶效应带来的细胞毒性; 其二, 该系统要求靶序列具有特定的间隔序列(PAM), 虽然生物的基因组内存在大量的间隔序列, CRISPR/Cas9系统能得到广泛的应用, 但是若遇到研究的基因没有富含胞嘧啶的间隔序列便束手无策了; 其三, CRISPR/Cas9介导的DNA双链断裂后修复引发靶位点远端DNA大片段丢失、插入等事件值得深入的思考和广泛的关注<sup>[43]</sup>。另一方面, CRISPR/Cas9仍存在的缺点与不足也促使研究者不懈努力, 为改进与研发更简易、高效、脱靶效率低的基因编辑技术指引前进方向。

#### 4 CRISPR/Cpf1系统的结构与分子作用机理

2015年, 张锋教授<sup>[44]</sup>在改进CRISPR/Cas9系统方面取得了突破性的成果。在几百个来自细菌的蛋白酶中, 张锋团队发现其中一个能防御病毒入侵并对核酸序列具有编辑能力的蛋白酶Cpf1, 与旧的CRISPR/Cas9系统相比较, CRISPR/Cpf1编辑靶基因

序列更简单、高效且脱靶效应几乎没有。CRISPR/Cpf1系统属于两类五型, 由42~44 nt gRNA和Cpf1两部分组成, Cpf1仅有RuvC-like结构域和推定的某核酸酶结构域, 没有HNH结构域, 具有5'端DNA与RNA切割能力。其作用机制与CRISPR/Cas9的作用机制大同小异, 效率却有突破性的提高。在CRISPR转录成crRNA前体(pre-crRNA)后, 由Cpf1切割pre-crRNA形成成熟的crRNA, crRNA通过识别靶序列5'端富含胸腺嘧啶的PAM序列, 引导Cpf1在PAM序列下游靶DNA链的第23位核苷酸和非靶DNA链的18位核苷酸处进行切割, 产生5 bp凸出的黏性末端<sup>[44]</sup>, 引发NHEJ修复机制。

#### 4.1 CRISPR/Cas9与CRISPR/Cpf1系统两者的异同

新改进的CRISPR/Cpf1与CRISPR/Cas9仍有很大差异(表2)。首先在组分数目上, 前者由crRNA与Cpf1蛋白酶两部分构成, 不需要tracrRNA序列, 而CRISPR/Cas9由tracrRNA、crRNA与Cas9蛋白酶三部分构成, 需要tracrRNA序列与crRNA结合成复合物发挥识别作用。在各组分性质方面, CRISPR/Cas9系统中的gRNA是由tracrRNA与crRNA融合而成, 长度大于100 nt, 而CRISPR/Cpf1系统中的gRNA是一个小片段序列, 仅42~44 nt。gRNA序列长度的减少大大减轻了针对靶序列的gRNA的设计工作, 同时在一定程度上提高了CRISPR/Cpf1系统的转染效率。其次在分子量上, Cpf1酶分子量比Cas9小, 进入细胞更容易, 进一步提高了编辑成功率。值得注意的是, Cas9具有HNH结构域与RuvC-like结构域, 而Cpf1没有HNH结构域, 只有RuvC-like结构域和一个目前功能不明确的推定的核酸酶结构域; Cpf1具有双重切割活性, 不仅具有Cas9一样的5'端DNA酶活性, 而且还有5'端RNA酶活性。

Cpf1的结构决定其功能, 在两个系统的作用机制方面也有很大不同。总体上都是蛋白酶在gRNA的引导下, 靶向切割DNA序列, 然而在CRISPR转录成pre-crRNA后, CRISPR/Cas9系统需要RNA内切酶III对pre-crRNA进行切割生成成熟的crRNA, CRISPR/Cpf1系统中pre-crRNA的切割直接由Cpf1执行; 然后在gRNA识别靶序列阶段, CRISPR/Cpf1系统中gRNA识别靶序列5'富含胸腺嘧啶(T)的PAM序列, 而CRISPR/Cas9中gRNA识别靶序列3'富含胞嘧啶(C)PAM序列, 两者识别序列的特性, 大大拓宽

表2 CRISPR/Cpf1与CRISPR/Cas9系统的比较

Table 2 Comparisons of CRISPR/Cpf1 and CRISPR/Cas9 Systems

评估项	CRISPR/Cpf1 系统	CRISPR/Cas9系统
Evaluation item	CRISPR/Cpf1 system	CRISPR/Cas9 system
Core component	sgRNA and Cpf1	sgRNA and Cas9
sgRNA length	42-44 nt	>100 nt
Protein size	Consisting of 1 200-1 300 amino acids with a low molecular weight	Consists of 1 368 amino acids with a large molecular weight
Protein structure	Containing a RuvC domain and a putative nucleic acid domain	Containing RuvC domain and HNH domain
Protease activity	DNA and RNA endonuclease activity	DNA endonuclease activity
Factor for processing crRNA	Matured by Cpf1	Matured by RNase III
DNA recognition site	T-rich PAM sequence	C-rich PAM sequence
Cleavage site	The 23rd nucleotide downstream of the PAM and the 14-18 nucleotides of the complementary DNA strand	The 3rd nucleotide upstream of the PAM
End type	Sticky end	Flat end
Off-target rate	Almost none	High

了CRISPR技术的应用范围。在gRNA的引导下,蛋白酶特异性结合并切割靶位点,Cpf1的切割位点距离靶序列PAM较远,在靶DNA链的PAM序列下游第23位核苷酸和非靶DNA链的第14位到18位核苷酸范围内进行切割,切割后产生5 bp突出的黏性末端,而Cas9在靶序列PAM上游第3个核苷酸处切割,产生平末端。相比较而言,Cas9切割位点距PAM序列太近,在第一次切割后,由于小片段的缺失、插入、替换可能造成PAM改变,影响Cas9再次识别,从而无法对靶位点进行多次编辑;而Cpf1切割后产生的黏性末端比起Cas9切割后产生的平末端,在基因损伤修复方面更有优势。除此之外,Cpf1的切割位点与crRNA spacer序列长度相关,当spacer序列长度小于20 nt时,Cpf1在非互补链的14位切割,而spacer序列大于20 nt时,则在非互补链的18位切割。利用Cpf1的这个特性,在较短spacer长度的crRNA介导下,可以特异性切割靶DNA的22位核苷酸与非靶DNA链的14位核苷酸形成8 nt的长黏性末端,研究人员利用此特性联合Taq DNA连接酶的高精度连接开发了大DNA片段体外无缝编辑的新工具并得到了成功的应用<sup>[45]</sup>。靶基因编辑产生的末端,若是平末端,则需要在设计插入基因时添加较长的同源臂来提高同源重组的效率,根据Cpf1切割后产生黏性末端的特性,只需在插入基因上添加与黏性末端突出的5 bp互补序列,即可通过NHEJ修复途径进行对靶序列的可控编辑,无需依赖发生频率低的同源重组方式<sup>[46]</sup>。Li

等<sup>[47]</sup>于2018年发现,只需左侧同源臂的供体模板足以用于精确靶向等位基因替换,极大地简化了作物改良中精确基因组编辑的供体模板设计,减轻了研究者的工作量与成本。研究发现,通过预组装重组Cpf1能有效消除之前对CRISPR/Cas9脱靶效应的顾虑<sup>[49]</sup>。十分重要的是,相比较CRISPR/Cas9系统一般在靶位点产生20 bp以内的缺失与插入<sup>[48]</sup>,CRISPR/Cpf1切割靶位点产生的黏性末端,引起较大基因片段的缺失与插入,这在一定程度上增加了编辑靶序列的可塑性<sup>[50]</sup>。

## 5 CRISPR/Cas系统的最新研究进展

CRISPR/Cas9系统中的核酸酶是来自化产脓链球菌的spCas9<sup>[51]</sup>,此系统只能识别切割5'-NGG-3'靶序列,人类基因组中每8 bp出现一个“NGG”序列可供研究者进行定位研究,该系统的局限性之一也在于此<sup>[52]</sup>。针对这个问题,近期Liu及其科研团队<sup>[53]</sup>对spCas9进行改进,迅速扩增、积累spCas9,最终筛选出一个spCas9的变体,称xCas9。xCas9蛋白能在多种PAM附近切割DNA序列,能广泛地应用于基因组的编辑。同时,研究还发现,xCas9不仅能广泛识别多种PAM,还大大增加了靶序列的识别能力,降低了脱靶风险<sup>[53]</sup>。CRISPR/Cpf1技术实现了对水稻多个基因进行有效编辑<sup>[54]</sup>。此外,无催化活性的Cas9(dCas9)和Cpf1(dCpf1)蛋白允许在不引入DSB的情况下对DNA进行可编程靶向。在真核生物中,dCas9和dCpf1通常与特定的效应区域

融合进行转录抑制(CRISPR干扰, 也称CRISPRi)或转录激活(CRISPR激活, 也称CRISPRa), 诱导生物表现遗传发生变化<sup>[55]</sup>。

2018年5月, 麻省理工学院和哈佛大学更是推出了一款快速、廉价、高度敏感的CRISPR诊断工具——夏洛克纸片测试条(SHERLOCK)。此工具无需仪器, 可直观看到阳性基因条带, 扩展了诊断信息的广度和灵敏度, 并能一次检测多个目标, 大大推广了该编辑技术的应用<sup>[56]</sup>。

## 6 小结与展望

目前, 以识别靶序列的5'-NGG-3' PAM为前提进行基因剪切CRISPR/Cas9技术已在植物、动物、微生物基因功能研究领域得到广泛应用, 包括后续由CRISPR/Cas9改良发展来的CRISPR/Cpf1与CRISPR/xCas9, 都能在多种PAM附近切割DNA序列, 大大拓宽了可编辑的靶基因覆盖面, 为研究者研究基因功能打下坚实的基础。应用基因编辑技术进行作物育种, 不仅有效规避了传统作物育种手段周期长、效率低的弊端, 而且就该编辑技术不涉及导入外源基因这方面而言, 安全性更高, 更容易为人们所接受。

同时, CRISPR/Cas9系统也存在诸多亟待解决的问题, 如脱靶效应等。再者, 后续研究者发现了Cpf1与xCas9蛋白能对多种PAM序列进行切割, 使基因编辑技术更广泛地得以运用。然而归根究底这类基因编辑技术均依赖于靶基因的序列特异性, 那么是否存在一种蛋白酶能摆脱特异性序列的限制, 可以对任意序列进行编辑? 徐澍等<sup>[57]</sup>设计了一种基于结构的核酸酶(structure-guided endonuclease, SGN), 它是由一个能识别3' flap结构的FEN-1识别区域和一个能切割DNA链的FokI切割区域两个部件组成。通过实验验证, 发现SGN能识别一条单链导向DNA与基因组DNA一条链杂交形成的3' flap结构, 并能切割DNA单链, 在距离3'末端9~10个碱基处形成缺口; SGN识别缺口结构并进一步切割掉一段长片段, 从而达到任意编辑生物基因组的目的。然而经实验发现, SGN编辑效率很低且编辑后造成大片段DNA序列缺失或插入, 大大提高了无效等位基因突变概率, 目前尚未见到有关SGN应用的相关报道。其次, CRISPR/Cas9在不同种植物中的基因编辑效率不同。不同的启动子引导与不同的转化方式对植物基因编辑效率产生很大的影响<sup>[58]</sup>, 这就需要针

对不同的植物设计对其最优的策略。关于CRISPR/Cas9技术特异性编辑序列后是否能在后代稳定遗传的问题, 沈春修等<sup>[59]</sup>通过CRISPR/Cas9技术编辑水稻LOC\_Os05g31750基因, 获得T<sub>0</sub>代阳性植株, 其后对突变植株的T<sub>1</sub>、T<sub>2</sub>代群体的靶位点检测证实了CRISPR/Cas9编辑水稻基因位点能在代间稳定遗传。但目前关于这方面的研究尚少, 并且能否稳定遗传与编辑位点与PAM之间的距离远近相关, 其稳定遗传性有待进一步研究证实。2018年, 英国惠康桑格研究所Bradley教授的研究发现在学术界掀起了轩然大波。他认为, CRISPR/Cas9介导的DNA双链断裂后修复存在诸多致病隐患, 例如CRISPR/Cas9会随机引发靶位点远端DNA大片段丢失、重排等基因突变, 这种高达数千bp的片段丢失很可能影响细胞功能, 且大片段丢失处在切割位点远端, 因此不容易被检测到<sup>[43]</sup>。CRISPR/Cas9编辑后, 基因序列变化多样、不可控<sup>[42]</sup>, 因此, 关于引发靶位点远端大片段丢失的分子机制有待进一步研究。Bradley选用小鼠胚胎干细胞为实验材料, 目前并没有其他的报道进一步实证CRISPR/Cas9技术编辑基因会引起靶位点远端出现大片段缺失的问题。CRISPR/Cas9编辑靶基因引发切割位点远端大片段缺失与细胞自身状态有无直接联系? 是否增殖活跃的细胞内基因互作机制与静态期细胞存在显著差异并影响基因编辑效果? 这些问题还有待进一步探究。

## 参考文献 (References)

- 1 Kamburova VS, Nikitina EV, Shermatov SE, Buriev ZT. Genome editing in plants: an overview of tools and applications. *Int J Agron* 2017; 7: 1-15.
- 2 Liang Z, Zhang K, Chen K, Gao C. Targeted mutagenesis in *Zea mays* using TALENs and the CRISPR/Cas system. *J Genet Genomics* 2014; 41(2): 63-8.
- 3 Van Gent DC, Van Der Burg M. Non-homologous end-joining, a sticky affair. *Oncogene* 2007; 26(56): 7731-40.
- 4 Mahfouz MM, Li L, Shamimuzzaman M, Wibowo A, Fang X, Zhu JK. *De novo*-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *Proc Natl Acad Sci* 2011; 108(6): 2623-8.
- 5 Bibikova M, Beumer K, Trautman JK, Carroll D. Enhancing gene targeting with designed zinc finger nucleases. *Science* 2003; 300(5620): 764.
- 6 Pavletich NP, Pabo CO. Zinc finger-DNA recognition: Crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* 1991; 252(5007): 809-17.
- 7 Wah DA, Bitinaite J, Schildkraut L, Aggarwal AK. Structure of



- FokI has implications for DNA cleavage. *Proc Natl Acad Sci USA* 1998; 95(18): 10564-9.
- 8 Davies JP, Kumar S, Sastry-Dent L. Use of zinc-finger nucleases for crop improvement. *Prog Mol Biol Transl Sci* 2017; 149: 47-63.
- 9 Shukla VK, Doyon Y, Miller JC, DeKolver RC, Moehle EA, Worden SE, *et al.* Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. *Nature* 2009; 459(7245): 437-41.
- 10 Carbery ID, Ji D, Harrington A, Brown V, Weinstein EJ, *et al.* Targeted genome modification in mice using zinc-finger nucleases. *Genetics* 2010; 186: 451-9.
- 11 Curtin SJ, Zhang F, Sander JD, Haun WJ, Starker C, Baltes NJ, *et al.* Targeted mutagenesis of duplicated genes in soybean with zinc-finger nucleases. *Plant Physiol* 2011; 156(20): 466-73.
- 12 Cui X, Ji D, Fisher DA, Wu Y, Briner DM, Weinstein EJ. Targeted integration in rat and mouse embryos with zinc-finger nucleases. *Nat Biotechnol* 2011; 29(1): 64-8.
- 13 Carroll D, Beumer KJ, Trautman JK. High-efficiency gene targeting in *Drosophila* with zinc finger nucleases. *Humana Press* 2010; 649(4): 271-80.
- 14 Lamb BM, Mercer AC, Barbas CF. Directed evolution of the TALE N-terminal domain for recognition of all 5' bases. *Nucleic Acids Res* 2013; 41(21): 9779-85.
- 15 Boch J, Bonas U. Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu Rev Phytopathol* 2010; 48(1): 419-36.
- 16 Yu Y, Streubel J, Balzergue S, Champion A, Boch J, Koebnik R, *et al.* Colonization of rice leaf blades by an African strain of *Xanthomonas oryzae* pv. *oryzae* depends on a new TAL effector that induces the rice nodulin-3 Os11N3 gene. *Mol Plant Microbe Interact* 2011; 24(9): 1102-13.
- 17 Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, *et al.* Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* 2009; 326(5959): 1509-12.
- 18 Moscou MJ, Bogdanove AJ. A simple cipher governs DNA recognition by TAL effectors. *Science* 2009; 326(5959): 1501.
- 19 Yuan M, Ke Y, Huang R, Ma L, Yang Z, Chu Z, *et al.* A host basal transcription factor is a key component for infection of rice by TALE-carrying bacteria. *Elife* 2016; 29: 5.
- 20 Renyan Huang, Shugang Hui, Meng Zhang, Pei Li, Jinghua Xiao, Xianghua Li, *et al.* A conserved basal transcription factor is required for the function of diverse TAL effectors in multiple plant hosts. *Front Plant Sci* 2017; 8: 1919.
- 21 Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol* 1987; 169(12): 5429-33.
- 22 Jansen R, van Embden JDA, Gastra W, Schouls LM. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* 2002; 43(6): 1565-75.
- 23 Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 2014; 157(6): 1262-78.
- 24 Makarova K, Wolf Y, Koonin E. The basic building blocks and evolution of CRISPR-Cas systems. *Biochem Soc Trans* 2013; 41(6): 1392-400.
- 25 Wang R, Preamplume G, Terns MP, Terns RM, Li H. Interaction of the Cas6 ribonuclease with CRISPR RNAs: Recognition and cleavage. *Structure* 2011; 19(2): 257-64.
- 26 Anantharaman V, Iyer LM, Aravind L. Presence of a classical RRM-fold palm domain in Thg1-type 3'-5' nucleic acid polymerases and the origin of the GGDEF and CRISPR polymerase domains. *Biol Direct* 2010 30; 5: 43.
- 27 Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 2008; 321(5891): 960-4.
- 28 Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 2011; 471(7340): 602-7.
- 29 Godde JS, Bickerton A. The repetitive DNA elements called CRISPRs and their associated genes: Evidence of horizontal transfer among prokaryotes. *J Mol Evol* 2006; 62(6): 718-29.
- 30 Barrangou R, Marraffini LA. CRISPR-cas systems: Prokaryotes upgrade to adaptive immunity. *Mol Cell* 2014; 54(2): 234-44.
- 31 Mohanraju P, Makarova KS, Zetsche B, Zhang F, Koonin EV, van der Oost J. Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* 2016; 353(6299): aad5147.
- 32 Gasjunas G, Barrangou R, Horvath P, Siksnys V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci USA* 2012; 109(39): E2579-86.
- 33 Luo M, Gilbert B, Ayliffe M. Applications of CRISPR/Cas9 technology for targeted mutagenesis, gene replacement and stacking of genes in higher plants. *Plant Cell Rep* 2016; 35(7): 1439-50.
- 34 Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012; 337(6096): 816-21.
- 35 Shan Q, Wang Y, Li J, Gao C. Genome editing in rice and wheat using the CRISPR/Cas system. *Nat Protoc* 2014; 9(10): 2395-410.
- 36 Xing HL, Dong L, Wang ZP, Zhang HY, Han CY, Liu B, *et al.* A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biol* 2014; 14: 327.
- 37 Wang S, Zhang S, Wang W, Xiong X, Meng F, Cui X. Efficient targeted mutagenesis in potato by the CRISPR/Cas9 system. *Plant Cell Rep* 2015; 34(9): 1473-6.
- 38 Brooks C, Nekrasov V, Lippman ZB, Van Eck J. Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated 9 system. *Plant Physiol* 2014; 166(3): 1292-7.
- 39 Janga MR, Campbell LM, Rathore KS. CRISPR/Cas9-mediated targeted mutagenesis in upland cotton (*Gossypium hirsutum* L.). *Plant Mol Biol* 2017; 94(4/5): 349-60.
- 40 Liang Zhen, Kang Zhang, Kunling Chen, Caixia Gao. Targeted mutagenesis in *zea mays* using TALENs and the CRISPR/Cas system. *J Genet Genomics* 2014; 41(2): 63-8.
- 41 时欢, 林玉玲, 赖钟雄, 杜宜殷, 黄鹏林. CRISPR/Cas9介导的植物基因编辑技术研究进展. *应用与环境生物学报*(Shi Huan, Lin Yuling, Lai Zhongxiong, Du Yiyin, Huang Penlin. *Progress*

- of CRISPR/Cas9 mediated plant gene editing technology. Chinese Journal of Applied And Environmental Biology) 2018; 1006(687): 498-513.
- 42 Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol 2013; 31(3): 233-9.
- 43 Kosicki M, Tomberg K, Bradley A. Repair of CRISPR-Cas9-induced double-stranded breaks leads to large deletions and complex rearrangements. Nat Biotechnol 2018; doi: 10.1038/nbt.4192.
- 44 Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, *et al.* Cpf1 is a single RNA-guided endonuclease of a Class 2 CRISPR-Cas system. Cell 2015; 163(3): 759-71.
- 45 Zou X, Jiang X, Xu L, Lei T, Peng A, He Y, *et al.* Transgenic citrus expressing synthesized cecropin B genes in the phloem exhibits decreased susceptibility to Huanglongbing. Plant Mol Biol 2017; 93(4/5): 341-53.
- 46 杨帆, 李寅. 新一代基因组编辑系统CRISPR/Cpf1. 生物工程学报(Yang Fan, Li Yan. Next generation genome editing system CRISPR/Cpf1. Chinese Journal of Biotechnology) 2017; 33(3): 361-71
- 47 Lei C, Li YS, Liu JK, Zheng X, Zhao GP, Jin W. The CCTL (Cpf1-assisted cutting and Taq DNA ligase-assisted ligation) method for efficient editing of large DNA constructs *in vitro*. Nucleic Acids Res 2017; 45(9): e74.
- 48 Kim D, Kim J, Hur JK, Been KW, Yoon SH, Kim JS. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. Nat Biotechnol 2016; 34(8): 863-8.
- 49 van Overbeek M, Capurso D, Carter MM, Thompson MS, Frias E, Russ C, *et al.* DNA repair profiling reveals nonrandom outcomes at Cas9-mediated breaks. Mol Cell 2016; 63(4): 633-46.
- 50 Hu X, Wang C, Liu Q, Fu Y, Wang K. Targeted mutagenesis in rice using CRISPR-Cpf1 system. J Genet Genomics 2017; 44(1): 71-3.
- 51 Horvath P, Barrangou R. CRISPR/Cas, the immune system of *Bacteria* and *Archaea*. Science 2010; 327(5962): 167-70.
- 52 Paul JW, Qi Y. CRISPR/Cas9 for plant genome editing: accomplishments, problems and prospects. Plant Cell Rep 2016; 35(7): 1417-27.
- 53 Ledford H. Powerful enzyme could make CRISPR gene-editing more versatile. Nature 2018; 556: 57-63
- 54 Wang M, Mao Y, Lu Y, Tao X, kang Zhu J. Multiplex gene editing in rice using the CRISPR-Cpf1 system. Mol Plant 2017; 10(7): 1011-3.
- 55 Kim SK, Kim H, Ahn W, Park K, Lee D, Lee S. Efficient transcriptional gene repression by Type V-A CRISPR-Cpf1 from eubacterium eligens. ACS Synth Biol 2017; 6(7): 1273-82.
- 56 Gootenberg JS, Abudayyeh OO, Kellner MJ, Collins JJ, Zhang F. Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. Science 2018; 360(6387): 439-44.
- 57 Xu S, Cao S, Zou B, Yue Y, Gu C, Chen X, *et al.* An alternative novel tool for DNA editing without target sequence limitation: The structure-guided nuclease. Genome Biol 2016; 17(1): 186.
- 58 黄娟, 邓国富, 高利军, 高菊, 卿冬进, 朱昌兰. CRISPR/Cas9 基因编辑系统在作物育种研究中的应用进展. 南方农业学报 (Huang Juan, Deng Guofu, Gao Lijun, Gao Ju, Qin Dongjin, Zhu Changlan. Progress in the application of CRISPR/Cas9 gene editing system in crop breeding. Journal of Southern Agriculture) 2018; 49(1): 14-21.
- 59 沈春修, 廖锦风, 却志群, 刘莹. CRISPR/Cas9 定点编辑水稻基因位点及靶修饰位点遗传稳定性研究. 核农学报 (Shen Chunxiu, Miao Jinfeng, Que Zhiqun, Liu Ying. Genetic stability of rice gene locus and target modification locus by CRISPR/Cas9. Journal of Nuclear Agriculture Science) 2018; 32(6): 1041-9.