

DOI: 10.13376/j.cblls/2018117

文章编号: 1004-0374(2018)09-0980-07



朱健康, 博士, 2010年当选为美国科学院院士, 2011年入选中央“千人计划”顶尖人才, 2016年1月全职回国开展工作, 现任中国科学院分子植物科学卓越创新中心/植物生理生态研究所学术所长、中国科学院上海植物逆境生物学研究中心主任。朱健康院士主要从事植物逆境分子生物学、表观遗传调控、植物基因编辑技术开发及应用等研究, 在相关领域硕果累累, 做出了卓越贡献, 在 *Nature*、*Science*、*Cell* 等世界顶尖学术期刊上发表研究论文 300 多篇, 是世界自然科学领域发表文章引用率最高的科学家之一。

化繁为简——植物多基因编辑体系的优化

王木桂, 朱健康*

(中国科学院上海植物逆境生物学研究中心, 中国科学院分子植物科学卓越创新中心, 上海 201602)

摘要: 多基因编辑技术对于基因家族和通路的研究以及作物改良具有十分重要的意义。利用 CRISPR-Cas 系统在植物中进行多基因编辑需要同时表达 Cas 基因和多个 sgRNA, 组成一个多转录元件系统。为了简化多基因编辑体系, 人们利用 RNA 自剪切工具和植物内源的 RNA 加工机制把多转录元件改造成双转录元件和单转录元件。该文介绍了植物多基因编辑体系由繁至简不断优化过程中的经典案例, 同时对存在的问题及发展的方向进行讨论。

关键词: Cas9; Cpf1; 多基因编辑; RNA 加工剪切; 植物

中图分类号: Q78; Q943 **文献标志码:** A

Simplification of multiplex gene editing systems in plants

WANG Mu-Gui, ZHU Jian-Kang*

(Shanghai Center for Plant Stress Biology and Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai 201602, China)

Abstract: Multiplex gene editing is a powerful tool for targeting plant gene families and is of great importance for crop improvement. Generally, the Cas endonuclease is expressed from an RNA polymerase (pol) II promoter, while sgRNA is expressed from a pol III promoter, such as U6 or U3 for gene editing with the CRISPR-Cas system. For multiplex gene editing, multiple sgRNAs are usually driven by separate pol III promoters. To simplify multiplex gene editing in plants, these sgRNAs can be expressed as a single transcript, which generates functional individual sgRNAs after processing by various RNA processing systems. Further, the Cas endonuclease and its sgRNA array can be co-expressed from a single pol II promoter, without any additional RNA processing machinery. In this review, we discuss some examples and potential problems in the optimization of multiplex gene editing systems in plants.

Key words: Cas9; Cpf1; multiplex gene editing; RNA processing; plant

收稿日期: 2018-05-07; 修回日期: 2018-07-13

基金项目: 中国科学院经费支持

*通信作者: E-mail: jkzhu@sibs.ac.cn

CRISPR (clustered regularly interspaced short palindromic repeats)/Cas (CRISPR-associated gene) 来源于细菌的免疫系统, 主要由 CRISPR 序列元件与 Cas 基因组成。在细菌中, CRISPR 元件是一类独特的 DNA 重复序列簇, 由一些高度保守的重复序列 (repeats) 和间隔序列 (spacer) 相间排列组成, 而 Cas 基因是位于 CRISPR 序列附近一些高度保守的家族基因, 具有核酸酶活性, 可对 DNA 序列进行剪切^[1]。该免疫系统在工作时, CRISPR 序列转录形成 CRISPR RNA (crRNA), 与另一种转录形成的反式作用型 crRNA (trans-activating crRNA, tracrRNA) 部分区域配对形成二元复合体, 然后该二元复合体一起引导具有非特异核酸酶活性的 CAS 蛋白切割与 crRNA 匹配的 DNA 序列, 通常为侵入细菌体内的外来 DNA 序列, 如噬菌体、质粒等等^[2]。为方便在真核生物中应用, 工程化的 CRISPR 系统将 crRNA 与 tracrRNA 融合在一起形成单一的 sgRNA (single guide RNA)^[3-4]。在植物中进行基因编辑通常需要同时构建 sgRNA 表达盒和 Cas 基因表达盒, 其中 sgRNA 表达盒比较小, 一般由长约 300 bp 的 U3 或 U6 等 pol III (RNA polymerase III) 启动子驱动, 以连续 6 个以上的 T 碱基 (poly T, pT) 终止即可, 而 Cas 基因的表达则通常由 pol II (RNA polymerase II) 类型的 35S 或 Ubi 等强启动子驱动, 以 PolyA 或 NOS 等终止子结束^[5-6]。

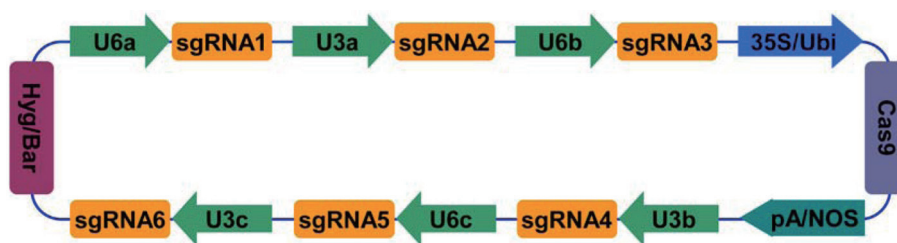
在动物或人类细胞里进行多基因编辑时可以共转染多个含有 sgRNA 表达盒的质粒, 甚至直接混合注射其 RNA, 而在植物体内进行基因编辑往往需要利用农杆菌介导等方法实现稳定的遗传转化, 当进行单次多基因定点编辑时必须把 Cas 基因表达盒、筛选标记表达盒以及多个 sgRNA 表达盒共同装配到一个载体上, 组成一个多转录元件系统 (multi-component transcriptional unit system, MCTU)(图 1)。

1 MCTU——经典的植物多基因编辑系统

2015 年, 刘耀光研究组开发了一套经典的适用于单子叶和双子叶植物的多基因编辑载体系统, 可利用一个载体同时表达多个 sgRNA 单元^[7]。为避免过多重复序列可能导致的片段重组丢失、基因沉默等风险, 该研究在水稻和拟南芥中分别选用 4 个 pol III 类型的启动子: *OsU3*、*OsU6a*、*OsU6b*、*OsU6c* (对水稻), 或 *AtU3b*、*AtU3d*、*AtU6-1*、*AtU6-29* (对拟南芥), 组成若干个不同的 sgRNA 表达盒同时靶向不同的目标位点。为方便组装这一套复杂的载体系统, 研究人员巧妙地设计了一系列的 PCR 扩增引物和酶切黏性末端, 可通过 Gibson 片段装配法或 Golden Gate 连接法将所有 sgRNA 模块一次性克隆到包含有 Cas9 表达盒和筛选标记表达盒的双元载体上。另外, 载体上的克隆位点携带了 *ccdB* 致死基因, 克隆失败的菌株在 LB 板上长不出来, 从而进一步提高了载体构建效率。在这项研究中, 研究人员构建了 20 个靶向水稻单基因和多基因的敲除载体, 其中靶向目标最多的一个质粒搭载了 8 个 sgRNA 表达盒。另外, 还构建了 4 个靶向拟南芥 1~3 个基因的载体。实验结果表明, 水稻所有被编辑的 46 个位点平均突变率高达 85.4%, 并且大多数突变类型为双等位突变或纯合突变, 而在拟南芥的基因敲除效率也达到 35.6%, 充分说明了这套多基因编辑系统的高效性。

2016 年, 朱健康研究组亦报道了类似构架的 MCTU 系统用于拟南芥的多基因编辑^[8]。该研究从拟南芥中克隆了 3 个 pol III 启动子: *AtU6-26*、*AtU3b* 和 *At7SL-2*, 分别使用两次构成一个 6 敲系统靶向拟南芥 PYR/PYL 基因家族的 6 个基因位点, 在 T1 代植株检测发现各靶点的突变率达 13%~93%, 并且能获得 6 基因同时突变的株系。

MCTU 系统作为第一代植物多基因编辑体系已



pA/NOS: poly A或NOS等终止子; Hyg/Bar: 潮霉素或草铵膦等抗性基因。

图1 MCTU系统

经广泛用于植物多基因位点的敲除，由于每个 sgRNA 均由单独的启动子驱动表达，相互之间不干扰，系统内各 sgRNA 单元的基因敲除效率一般与相应的单基因敲除载体无异，但也正是因为这样导致系统构架复杂，通常由一系列中间载体组成，构建起来程序繁琐且费时费力，往往需要有经验的技术人员才能顺利完成。

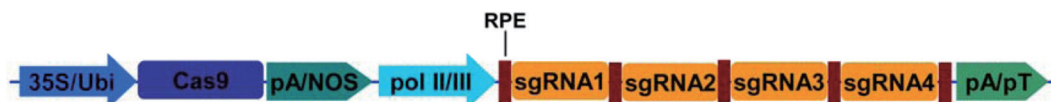
2 TCTU——一个表达盒搭载多个sgRNA单元

为了简化复杂的 MCTU 系统，科研人员尝试把多个 sgRNA 表达盒整合成一个，这样加上 Cas9 基因表达盒可以把多基因编辑体系简化为双转录元件系统 (two component transcriptional unit system, TCTU) (图 2)。当利用一个启动子转录一串 sgRNA 阵列，各 sgRNA 单元需要经过特定的加工剪切机制进行分离释放，于是一系列具有 RNA 加工剪切功能的元件 (RNA processing elements, RPE) 进入了人们的视线。tRNA (transfer RNA) 是生物体内广泛存在的用于携带并转运氨基酸的一类小分子核糖核酸，由一条长 70~90 个核苷酸并折叠成三叶草形的短链组成。在真核细胞核内，tRNA 的前体由 RNase P 和 RNase Z 识别并于特定位点切割，被切除的 5' 和 3' 冗余序列没有序列限制性^[9-10]。根据这些特征，2015 年，杨益农研究组利用植物内源的 tRNA 加工成熟机制开发了多顺反子 tRNA-gRNA 系统 (polycistronic tRNA-gRNA, PTG)，可以利用一个 pol III 启动子同时转录生成多个 sgRNA，在这一过程中不需要添加外源的 RNA 加工剪切工具^[11]。该研究选择在植物体内广泛存在的甘氨酸转运 tRNA (tRNA^{Gly}) 元件 (相应的 DNA 序列长度为 77 bp) 作为间隔，将数量不等的 sgRNA 单元 (最多达 8 个) 组装成一系列的 PTG 系统，通过原生质体转化靶向水稻多个位点，结果表明各 PTG 系统的 DNA 切割效率为 4%~45%，同时，水稻稳定转化实验的结果也表明各靶向位点的突变率最高可达 100%。

与 tRNA 类似的常用于组装多基因编辑器的 RNA 加工剪切工具还有 Csy4、核酶 (ribozyme, RZ) 等。Csy4 是来源于细菌的 RNA 剪切酶，在细菌的

免疫系统里，当 CRISPR 基因簇被转录成前体 (pre-crRNA)，Csy4 就通过序列与结构的特异性把该前体分割成单个成熟的 crRNA^[12]。Csy4 RNA 酶及其附属的仅 20 bp 长度的特异识别切割序列被用于构建人类细胞的多基因编辑器并证实其有效性^[13]。核酶是一类具有催化活性的 RNA 小分子，可通过催化靶位点 RNA 链中磷酸二酯键的断裂，特异性地剪切底物 RNA 分子。两种名为 Hammerhead 和 HDV (hepatitis delta virus) 的短链核酶序列 (其相应 DNA 序列的长度分别为 43 bp 和 68 bp) 曾被用来组装 RGR (ribozyme-gRNA-ribozyme) 系统，分别设置在 sgRNA 序列的 5' 端和 3' 端，用于从 pol II 启动子转录出来的 sgRNA 前体中精确地切割分离成熟的 sgRNA 序列，最先于 2014 年在酵母及人类细胞的基因编辑实验中得到验证^[14-15]。另一种具有 15 bp 识别位点的 Hammerhead 核酶元件也被成功地用于从 pol II 转录物中切割释放 1~2 个 sgRNA 单元，水稻转化实验表明其生成的 sgRNA 引导 Cas9 靶向编辑目标位点的效率最高达 100%^[16]。

2017 年，Voytas 研究组在番茄原生质体细胞中系统地比较了 pol III 构架的 MCTU 体系和 pol II 构架的 TCTU 体系在多基因编辑上的效率^[17]。在 MCTU 体系中，两个 pol III 启动子 *AtU6* 和 *At7SL*，被组装成三套双基因敲除系统。而在 TCTU 体系中，3 种不同类型的 RNA 加工剪切工具 tRNA、Csy4 和核酶元件，各组成一个双基因敲除系统。由于常用的 35S 启动子已被用来驱动 Cas9 基因的表达，为避免在一个 T-DNA 区上重复使用相同启动子造成基因沉默的风险，在 TCTU 体系中选择了来自丁香属黄叶卷曲病毒 (Cestrum Yellow Leaf Curling Virus) 的 pol II 类型启动子 CmYLCV 驱动 sgRNA 阵列的转录，因为该启动子适用于很多类型的作物，并且其驱动基因表达的能力在双子叶植物里不亚于 35S，而在单子叶里不亚于 *ZmUbi*^[18]。研究人员对转化后的原生质体进行高通量测序表明，使用 Csy4 和 tRNA 组装的 TCTU 系统所产生的突变率大概是 MCTU 系统的两倍，3 个 MCTU 系统之间没有显著差异，而利用核酶元件组成的 TCTU 系统进行基因



RPE: RNA加工剪切元件; pol II/III: RNA聚合酶II或III相关的启动子; pA/NOS: poly A或NOS等终止子; pT: poly T终止序列。

图2 TCTU系统

编辑的效率最低。随后, 上述3个TCTU系统被扩大到同时编辑8个靶位点, 结果表明, Csy4 构架的效率最高, tRNA 次之, 而核酶系统的表现最差^[17]。核酶元件有很多种, 由于在上述研究中所使用的核酶的识别位点仅为15 bp, 可能会对细胞体内相同或类似的转录序列也进行识别切割从而产生毒性, 最终影响基因编辑的效率。尽管不同的研究对核酶元件的评价不一, 但 Voytas 的上述研究充分证明了TCTU体系在多基因编辑中的效率可以优于经典的MCTU体系, 而且结构简单, 构建更加方便。另外, 在TCTU系统里由于RNA加工剪切工具的使用, sgRNA的表达不必受限于pol III启动子, 众多各具特色的pol II启动子为其提供了更加灵活的选择。

3 Cpf1——天然的TCTU系统

Cpf1 (CRISPR from *Prevotella* and *Francisella* 1, 又名为 Cas12a) 是一种新型的 class 2-type V CRISPR 系统, 与属于 class 2-type II 的 Cas9 相比具有诸多独特的特征: 它识别富含 A/T 碱基的 PAM 序列 (TTN 或 TTTN), DNA 切割位点位于 PAM 的远端并产生黏性末端。另外, 在 CRISPR/Cpf1 系统中, 成熟的 crRNA 仅由一条很短的直接重复序列 (direct repeat, DR, 20~21 bp) 和引导序列 (guide, 22~23 bp) 组成, 不需要反式作用型 tracrRNA 的辅助; 更重要的是, Cpf1 蛋白兼具 DNA 剪切酶和 RNA 剪切酶的功能, 可以把转录出来的 crRNA 前体加工剪切成成熟型 crRNA^[19-20]。这样, Cpf1 蛋白自身就具备了类似 TCTU 系统中 RNA 加工剪切元件的功能。利用 Cpf1 的这种特性可以把多个 crRNA 单元直接串联在一起表达, 转录后 Cpf1 蛋白自身可以把单个 sgRNA 切割分离并分别靶向不同的位点, 这一过程不需要像 Cas9 多基因编辑系统那样添加额外的 RNA 处理元件。这一构思率先通过动物细胞实验和小鼠体内实验取得成功。在体外和体内实验中, 研究人员利用 AsCpf1 (Cpf1 from *Acidaminococcus*) 分别构建了由4个和3个DR-guide单元组成的crRNA阵列进行多基因编辑, 发现各靶点的突变效率与单基因编辑系统相当, 阵列内不同guide的排

列次序不存在位置效应; 另外, 在研究中还发现由成熟型 DR 和 guide 序列组成的阵列比序列更长的非成熟型阵列具有更高的效率^[21]。

2017年, 朱健康研究组利用 FnCpf1 (Cpf1 from *Francisella novicida*) 和 LbCpf1 (Cpf1 from *Lachnospiraceae bacterium* ND2006) 对水稻进行单基因和多基因敲除测试, 发现上述两类 Cpf1 只需一条非常短的 DR 序列 (20~21bp) 加上 22~24 bp 的 guide 序列即可在植物体内实现单基因敲除; 更重要的是, 把多个成熟型 DR-guide 单元直接串联, 只需要一个 pol III 启动子驱动即可简单高效地实现多基因敲除。实验中测试了由4个DR-guide单元组成的crRNA短阵列分别对水稻RLK和CYP81A家族的四个基因进行编辑, 结果表明, 各靶点的敲除效率达到40%~75%^[22]。Cpf1的开发利用拓展了CRISPR系统在动植物中的应用, 利用Cpf1可以组装成一个天然的TCTU系统, 为多基因编辑提供了一个新利器(图3)。但是相对于Cas9, 目前Cpf1在动植物中尚未得到广泛的使用, 在一些双子叶植物(如拟南芥)中基因编辑效果不佳, 仍然需要进一步优化改良。

4 STU——只有一个表达盒的多基因编辑器

TCTU系统借助外源或内源的RNA加工工具已经实现了把多个sgRNA单元组装到一个表达盒里, 但仍然独立于Cas9或Cpf1等核酸酶的表达盒之外, 能否进一步把这两个基本的表达盒也整合到一起, 这一设想首先在人类细胞的基因编辑中实现。2015年, Naito研究组构建了由上文提到的Hammerhead和HDV核酶元件组装而成的包含sgRNA的RGR模块, 并把RGR模块内嵌于Cas9表达盒的5' UTR (untranslated region) 区, 由一个pol II启动子同时驱动sgRNA和Cas9, 转录后sgRNA可由核酶元件分离, 而Cas9的mRNA则进入下一步的翻译阶段, 这样就构成了一种由单个转录元件组成的基因编辑系统 (single transcriptional unit system, STU)^[23]。该研究利用包含1~2个RGR模块的STU系统对HEK 293细胞系进行了转化实验, 结果表明STU系统能有效地切割靶位点; 但是在上述研究中所构建STU



DR: 直接重复序列; pol II/III: RNA聚合酶II或III相关的启动子; pA/NOS: poly A或NOS等终止子; pT: poly T终止序列。

图3 利用Cpf1构建的TCTU系统

系统的缺陷在于，由于 RGR 模块被设置在 Cas9 的 5' 端，有可能由于 ATG 碱基的存在而导致 mRNA 被提前翻译，所以，在这个系统中还有必要在 Cas9 的 5' 端添加一段内部核糖体进入位点序列 (internal ribosome entry site, IRES) 以保证 Cas9 准确地起始翻译。2016 年，张勇研究组设计的 STU 系统在植物的基因编辑中也取得了成功^[24]。研究人员先利用 eYFP 报告基因在水稻原生质体中测试了多种 linker 序列对 STU 系统内基因转录和翻译的影响，发现当设置有核酶元件时，以一段人工合成的 polyA 序列 (pA) 作 linker 能保证 eYFP 报告基因的正常表达而不受核酶剪切的影响。随后，研究人员构建了以下构架的 STU 系统：pol II-Cas9-pA-RZsite-sgRNA1-RZsite-sgRNA2-RZ，在水稻中进行单基因和双基因位点的敲除实验，结果表明，靶位点的突变率达到 53.8%~100%。在这个构架的设计上，由于 sgRNA 模块被设置在 Cas9 完整表达盒的 3' 端，所以不需要添加额外的 IRES 序列去保证 Cas9 的起始翻译。

谢卡斌研究组于 2018 年开发了分别基于 Cas9 和 Cpf1 的 STU 系统^[25]，研究人员发现在 UBI10 启动子的 5' UTR 区域包含一段非常保守的内含子序列，该序列的边界上具有在 mRNA 成熟过程中内含子被切除时识别的基序 5'-GU-A-AG-3'，理论上可以把其中的内含子序列换成 sgRNA 模块，这样在用 UBI10 启动子驱动 Cas9 时，位于 5' UTR 区域 GU-A-AG 基序内的 sgRNA 被识别为内含子而切割分离，而作为外显子的 Cas9 则正常表达。在装载 Cas9 的 STU 系统里，当需要串联多个 sgRNA 时，各单元之间可用 tRNA 元件连接，而在装载 Cpf1 的 STU 系统里，由于 Cpf1 蛋白自身兼具 RNA 酶活性，各 guide 单元之间可直接用其 DR 序列作间隔。水稻转化实验表明，这两套 STU 系统均能有效地进行单基因和多基因位点敲除。

以上论述的三套 STU 系统在设计上各有特色：Naito 研究组把 sgRNA 模块设置在 Cas9 基因的 5' UTR 区域，并且利用 Hammerhead 和 HDV 核酶元件进行分离，但需要添加额外的 IRES 元件保证 Cas9 的起始翻译；张勇研究组充分利用了 polyA 终

止子不完全终止转录的特征把 sgRNA 模块设置在 Cas9 表达盒的 3' 端，同样利用核酶元件进行分离；而谢卡斌研究组则利用了一段保守的内含子剪切基序用于内嵌 sgRNA 序列，利用 tRNA 元件实现各个 sgRNA 模块的分离。这些 STU 系统尽管各有各的特点，共同的特征是仅需要一个 pol II 启动子就能同时实现 Cas9 (或者 Cpf1) 和 sgRNA 的转录，在结构上比 TCTU 系统更简单，但其串联的单个和多个 sgRNA 模块之间还必需借助额外添加的各种 RNA 加工剪切元件才能有效地分离释放 (图 4)。

5 SSTU——把多基因编辑体系简化到极致

TRBO (*Tobacco mosaic virus* RNA-based overexpression) 是一个来源于烟草花叶病毒的过表达系统，常用于植物体内大量表达目的基因。在一项研究中，Cody 等^[26]本计划利用 TRBO 系统在烟草中大量表达 sgRNA，却发现在植物体内 sgRNA 的 5' 和 3' 端均能容忍一定的多余序列并保持活性，这与体外实验的结果并不一样。研究人员随即尝试把两个靶向不同位点的 sgRNA 直接串联在一起，发现两者均能有效地在目标位点产生突变，并且这两个 sgRNA 的排列顺序不存在位置效应。在这个 TRBO 表达系统中，甚至可以把 GFP 的编码序列设置在 sgRNA 的 5' 或 3' 端，同时执行 GFP 表达和 sgRNA 介导靶位点突变的功能。在这项研究中，尽管在植物体内 sgRNA 的 5' 和 3' 端能容忍多余序列这一现象已经证实，但其具体机制并未明确，作者推测了 3 个可能的原因：(1) 在植物体内 Cas9 蛋白结合 sgRNA 时可以容忍其两端的冗余序列；(2) 在植物体内 Cas9 蛋白可以切除 sgRNA 两端的多余序列；(3) 烟草自身的一些蛋白能够帮助切除 sgRNA 两端的多余序列。

2017 年，来自日本的 Endo 研究组通过一系列的体内和体外实验进一步证实 Cas9 蛋白和植物内源的 RNA 酶，如 RNA 酶 III 或 RNA 酶 T1，对于带有多余序列的 sgRNA 的成熟是必需的^[27]。在这项研究中，作者提出了以下假设的模型：在植物体内，Cas9 蛋白可以结合包含 sgRNA 的转录产物，然后



RPE: RNA加工剪切元件；pA/NOS: poly A或NOS等终止子。

图4 STU系统

植物内源的 RNA 酶可以帮助切除 sgRNA 两端的多余序列进而形成成熟的 Cas9-sgRNA 复合体。当然, 具体的作用流程和机理仍有待进一步研究。利用这一机制, 研究人员简化了多基因敲除的载体模型: 把 Cas9 的编码序列和一到多个 sgRNA 直接串连在一起通过一个 pol II 启动子驱动, sgRNA 之间只需一个短的 linker 序列 (12 bp) 作间隔。通过对水稻两个内源基因进行双基因敲除实验表明, 该系统的基因敲除效率与包含核酶元件的多敲系统相当。

朱健康研究组进一步提出了植物多基因编辑简化单转录元件系统 (simplified single transcriptional unit, SSTU) 的概念^[28](图 5)。在该系统的设计中, 不论是利用 Cas9 还是 Cpf1 做多基因敲除, 均可以把 Cas9 或 Cpf1 的编码序列和一个或多个 sgRNA 直接串连在一起, 由一个 pol II 启动子驱动形成一个表达盒。Cas9 系统里的各个 sgRNA 之间由一个很短的 linker 序列 (可以只有几个碱基) 隔开, 而 Cpf1 系统里的各个 guide 之间则由其独有的 DR 序列作间隔。为验证这一概念, 研究人员分别构建了 FnCpf1 和 LbCpf1 的 SSTU 系统, 同时靶向水稻内源的 8 个和 9 个基因位点, 同时, 还构建了基于 Cas9 的 SSTU 系统对 3 个水稻内源基因进行多位点敲除。水稻稳定转化结果表明, SSTU 系统的多基因编辑效率总体上与经典的 MCTU 或 TCTU 系统的效率相当。

在目前主要利用 Cas9 和 Cpf1 作为基因编辑工具的技术条件下, SSTU 系统已经把植物多基因编辑体系简化到了极致。目前该体系, 特别是基于 Cas9 的 SSTU 系统的主要问题在于, 尽管 Endo 研究组的实验已经证明 Cas9 蛋白和植物内源的 RNA 酶对于系统里 sgRNA 的成熟是必需的, 但是具体的切割分离机理仍不明确。与 Cpf1 对其 DR 序列的特异识别切割机制不同, Cas9 和植物内源的 RNA 酶对包括 linker 在内的 sgRNA 前体序列的处理可能是一种随机的非特异性切割。截至目前, 基于 Cas9 的 SSTU 系统仅在拟南芥、水稻和烟草中对 3 个以内的靶位点进行多基因敲除实验, 对更多靶点的测试及在其他物种上的优化应用有待进一步深入。

6 问题及展望

在短短 3~4 年时间里, 植物多基因编辑载体系统就伴随着 Cas9 和 Cpf1 等基因编辑工具的迅猛发展而经历了四代的进化, 多个 sgRNA 模块从连续组装到利用 RNA 自剪切元件进行串联再到单转录系统的一体化应用, 总体上由繁至简不断优化, 载体构建越来越简单方便, 必将加速植物多基因突变体的获得, 打破功能冗余, 进一步促进植物基因家族或通路的机理机制研究以及推动作物的改良。

目前植物多基因编辑系统仍然存在以下几个问题: (1) 尽管已经在水稻、拟南芥、烟草、番茄等多种植物上进行过多基因编辑, 但主要应用的还是第一代的 MCTU 体系, 简化版多基因编辑系统的使用还少有报道, 还需要在更多植物上进行适用性及效率的评价。(2) 目前的多基因编辑体系主要依赖 SpCas9 和 Cpf1 构建, 靶位点的设计受到这些基因编辑工具酶 PAM 序列 (SpCas9 为 NGG, Cpf1 为 TTTV, V = A/C/G) 的限制, 所以, 需要更多的基因编辑工具酶来做多基因编辑以提供更灵活的靶位点选择。2018 年报道的 XCas9 就是一个很好的候选工具, 它从 SpCas9 人工进化而来, 可以识别 NG、GAA 和 GAT 的 PAM 序列, 并且特异性更高^[29]。而 LbCpf1 的变体 LbCpf1 (RR) 可识别 TYCV (Y = C/T, V = A/C/G) 的 PAM 序列, 并且已经在水稻的基因敲除上得到验证^[30], 也可以用来建立多基因编辑系统。(3) 面向多个靶位点的系统设计及靶向效率有待进一步提高。在进行多基因编辑时, 尽管多数位点靶向效率较高, 但还是会经常出现个别靶点突变率低下, 甚至没有突变的情况, 导致一次性多基因全部敲除的努力前功尽弃。在多基因编辑系统里, sgRNA 阵列内是否存在 RNA 结构上的相互拮抗或干扰而导致其中一些 sgRNA 单元不能完全分离或被过度切割, 最终导致个别靶点效率低下, 目前尚没有明确的研究。在植物基因编辑领域, 针对单个位点的 sgRNA 设计预测软件已经不断完善, 但是面向多个靶位点进行系统设计及模拟分析的程序还鲜有报道。(4) 除了 sgRNA 本身可能会造成脱靶之外, 多基因编辑系统还可能会带来额外的脱靶



pA/NOS: poly A或NOS等终止子。

图5 SSTU系统

效应和细胞毒性风险。为了系统的简化而添加的各种 RNA 加工剪切工具所识别切割的序列一般都很短，例如其中一些 Hammerhead 核酶的识别位点仅有 15 bp，而以 Cpf1 构建的多基因编辑系统里，能被 Cpf1 蛋白自身识别切割的 DR 序列也仅有 20 bp 左右。尽管 Hammerhead 核酶和 Cpf1 均能通过序列及结构的双重特异性对其识别位点进行切割，但仍有可能对细胞体内相同或类似的转录序列进行误切从而产生额外的脱靶或毒性效应，需要在实际应用中深入观察分析。总体而言，在经历过代简化之后，植物多基因编辑系统未来的发展方向将是更广谱、更高效和更安全。

[参 考 文 献]

- [1] Makarova KS, Haft DH, Barrangou R, et al. Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol*, 2011, 9: 467-77
- [2] van der Oost J, Westra ER, Jackson RN, et al. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat Rev Microbiol*, 2014, 12: 479-92
- [3] Jinek M, Chylinski K, Fonfara I, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 2012, 337: 816-21
- [4] Cong L, Ran FA, Cox D, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science*, 2013, 339: 819-23
- [5] Shan Q, Wang Y, Li J, et al. Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat Biotechnol*, 2013, 31: 686-8
- [6] Mao Y, Zhang H, Xu N, et al. Application of the CRISPR-Cas system for efficient genome engineering in plants. *Mol Plant*, 2013, 6: 2008-11
- [7] Ma X, Zhang Q, Zhu Q, et al. A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. *Mol Plant*, 2015, 8: 1274-84
- [8] Zhang Z, Mao Y, Ha S, et al. A multiplex CRISPR/Cas9 platform for fast and efficient editing of multiple genes in *Arabidopsis*. *Plant Cell Rep*, 2016, 35: 1519-33
- [9] Barbezier N, Canino G, Rodor J, et al. Processing of a dicistronic tRNA-snoRNA precursor: combined analysis *in vitro* and *in vivo* reveals alternate pathways and coupling to assembly of snoRNP. *Plant Physiol*, 2009, 150: 1598-610
- [10] Canino G, Bocian E, Barbezier N, et al. *Arabidopsis* encodes four tRNase Z enzymes. *Plant Physiol*, 2009, 150: 1494-502
- [11] Xie K, Minkenberg B, Yang Y. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc Natl Acad Sci USA*, 2015, 112: 3570-5
- [12] Haurwitz RE, Jinek M, Wiedenheft B, et al. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science*, 2010, 329: 1355-8
- [13] Tsai SQ, Wyvekens N, Khayter C, et al. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol*, 2014, 32: 569-76
- [14] Gao Y, Zhao Y. Self-processing of ribozyme-flanked RNAs into guide RNAs *in vitro* and *in vivo* for CRISPR-mediated genome editing. *J Integr Plant Biol*, 2014, 56: 343-9
- [15] Nissim L, Perli SD, Fridkin A, et al. Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. *Mol Cell*, 2014, 54: 698-710
- [16] Tang X, Zheng X, Qi Y, et al. A single transcript CRISPR-Cas9 system for efficient genome editing in plants. *Mol Plant*, 2016, 9: 1088-91
- [17] Cermak T, Curtin SJ, Gil-Humanes J, et al. A multipurpose toolkit to enable advanced genome engineering in plants. *Plant Cell*, 2017, 29: 1196-217
- [18] Stavolone L, Kononova M, Pauli S, et al. Cestrum yellow leaf curling virus (CmYLCV) promoter: a new strong constitutive promoter for heterologous gene expression in a wide variety of crops. *Plant Mol Biol*, 2003, 53: 663-73
- [19] Zetsche B, Gootenberg JS, Abudayeh OO, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, 2015, 163: 759-71
- [20] Fonfara I, Richter H, Bratovic M, et al. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature*, 2016, 532: 517-21
- [21] Zetsche B, Heidenreich M, Mohanraju P, et al. Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat Biotechnol*, 2017, 35: 31-4
- [22] Wang M, Mao Y, Lu Y, et al. Multiplex gene editing in rice using the CRISPR-Cpf1 system. *Mol Plant*, 2017, 10: 1011-3
- [23] Yoshioka S, Fujii W, Ogawa T, et al. Development of a mono-promoter-driven CRISPR/Cas9 system in mammalian cells. *Sci Rep*, 2015, 5: 18341
- [24] Tang X, Zheng X, Qi Y, et al. A single transcript CRISPR-Cas9 system for efficient genome editing in plants. *Mol Plant*, 2016, 9: 1088-91
- [25] Ding D, Chen K, Chen Y, et al. Engineering introns to express RNA guides for Cas9- and Cpf1-mediated multiplex genome editing. *Mol Plant*, 2018, 11: 542-52
- [26] Cody WB, Scholthof HB, Mirkov TE. Multiplexed gene editing and protein overexpression using a tobacco mosaic virus viral vector. *Plant Physiol*, 2017, 175: 23-35
- [27] Mikami M, Toki S, Endo M. In planta processing of the SpCas9-gRNA complex. *Plant Cell Physiol*, 2017, 58: 1857-67
- [28] Wang M, Mao Y, Lu Y, et al. Multiplex gene editing in rice with simplified CRISPR-Cpf1 and CRISPR-Cas9 systems. *J Integr Plant Biol*, 2018, 60: 626-31
- [29] Hu JH, Miller SM, Geurts MH, et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature*, 2018, 556: 57-63
- [30] Li S, Zhang X, Wang W, et al. Expanding the scope of CRISPR/Cpf1-mediated genome editing in rice. *Mol Plant*, 2018, 11: 995-8