

利用内源 CRISPR-Cas 系统开展乳酸菌基因编辑的研究进展

朱青, 徐琛, 张书文, 谢宁, 逢晓阳, 吕加平

中国农业科学院 农产品加工研究所, 北京 100193

朱青, 徐琛, 张书文, 谢宁, 逢晓阳, 吕加平. 利用内源 CRISPR-Cas 系统开展乳酸菌基因编辑的研究进展. 生物工程学报, 2022, 38(7): 2447-2458.

ZHU Q, XU C, ZHANG SW, XIE N, PANG XY, LÜ JP. Advances in utilizing the endogenous CRISPR-Cas system for genome editing of lactic acid bacteria. Chin J Biotech, 2022, 38(7): 2447-2458.

摘 要: CRISPR-Cas9 技术是一种高效、精准的基因编辑工具, 该技术的建立推动基因组编辑进入快速发展阶段。目前应用最广泛的 Cas9 蛋白是来源于酿脓链球菌 (*Streptococcus pyogenes*) 的 SpyCas9, 该蛋白作为“基因剪刀”在哺乳动物、植物等真核生物中应用较为广泛且成熟, 但是该蛋白在一些乳酸菌中的应用仍然受到多种因素的限制。乳酸菌基因组上已发现多种类型的 CRISPR 系统, 也蕴含着多种未表征的 Cas 蛋白, 利用乳酸菌内源 CRISPR-Cas 系统, 结合外源导入的向导 RNA 和同源修复模板, 也可实现对乳酸菌基因组的编辑。这种基于内源 CRISPR-Cas 系统实现基因编辑的方式, 具有打靶载体相对较小易转化、无外源 Cas9 蛋白对宿主细胞产生毒性等优势, 相比于 CRISPR-SpyCas9 更适用于乳酸菌基因组进行编辑, 可能是一些乳酸菌未来开展基因组编辑的主要手段, 本文重点对此部分内容进行了综述。

关键词: 自我靶向; 内源 CRISPR-Cas 系统; 乳酸菌; 基因编辑

Received: December 31, 2021; **Accepted:** March 21, 2022; **Published online:** March 25, 2022

Supported by: National Natural Science Foundation of China (32172190, 31871833); Agricultural Science and Technology Innovation Program of Chinese Academy of Agricultural Sciences (CAAS-ASTIP-2020-IFST-04); Central Public-Interest Scientific Institution Basal Research Fund, China (S2020JBKY-18)

Corresponding authors: PANG Xiaoyang. Tel: +86-10-62815542; E-mail: pangxiaoyang@163.com
LÜ Jiaping. Tel: +86-10-62819421; E-mail: lüjiapingcaas@126.com

基金项目: 国家自然科学基金 (32172190, 31871833); 中国农业科学院科技创新工程项目 (CAAS-ASTIP-2020-IFST-04); 中央级公益性科研院所基本科研业务费专项 (S2020JBKY-18)

Advances in utilizing the endogenous CRISPR-Cas system for genome editing of lactic acid bacteria

ZHU Qing, XU Chen, ZHANG Shuwen, XIE Ning, PANG Xiaoyang, LÜ Jiaping

Institute of Food Science and Technology, Chinese Academy of Agricultural Sciences, Beijing 100193, China

Abstract: The CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR-associated proteins) system is an efficient and precise gene editing tool. The development of this technology has promoted genome editing into a rapid development stage. The most widely used Cas9 protein is SpyCas9 from *Streptococcus pyogenes*. As a “gene scissors”, the SpyCas9 protein is widely used in eukaryotes including mammals and plants. However, the application of this protein in some lactic acid bacteria (LAB) is still hampered by many factors. It has been identified that LAB genomes possess many types of CRISPR system and are rich in uncharacterized Cas proteins. Genome editing of LAB is possible by repurposing the endogenous CRISPR-Cas systems in LAB combined with exogenous single guide RNA (sgRNA) and homologous recombination template. This method employs its endogenous CRISPR-Cas system for gene editing, which has the advantages including easiness for transformation due to the relatively small targeting vector, and no concern about the toxicity of heterologous Cas9 to host cells. Compared to CRISPR-SpyCas9, the endogenous CRISPR-Cas system is more suitable for genome editing of LAB, and it may become the main genome editing tool for some LAB in the future. This article summarizes the advances in this field.

Keywords: self-targeting; endogenous CRISPR-Cas system; lactic acid bacteria; genome editing

乳酸菌 (lactic acid bacteria, LAB) 是一种世界公认安全的食品级微生物, 目前被广泛应用于食品发酵领域。近年研究证实, 乳酸菌作为益生菌在促进宿主健康方面发挥着重要作用, 其代谢过程产生的有机酸和分泌在胞外的抗菌肽, 可抑制病原微生物的生长繁殖、改善宿主肠道微生物菌群的结构等^[1-2]。尽管越来越多的研究和临床实践证明, 乳酸菌及其菌剂具有良好促进健康的作用, 在医药和健康领域具有广泛的应用前景^[3], 但是乳酸菌的研究和开发仍然面临两个难题, 一是乳酸菌菌株种类繁多, 常规的乳酸菌筛选是一项耗时耗力的工作; 二是随着越来越多乳酸菌基因组被测序, 大量未知基因需要进行功能解析, 但是传统的乳酸菌

基因操作手段过程繁琐、效率低下, 亟待开发快速高效的遗传操作工具。

传统的乳酸菌基因操作手段多是基于基因同源重组原理开发的, 比如基于染色体和质粒的同源重组方法 (利用乳酸菌内源的同源重组系统)^[4]、Red/RecET 介导的双链 DNA 重组 (利用外源噬菌体辅助重组系统)^[5]、单链 DNA 重组^[6]等。近年来, 以酿脓链球菌 SpyCas9 为核心的 CRISPR-Cas9 技术及其衍生技术已经发展成为基因组精准编辑工具, 这种方法利用 Cas9 的两个核酸酶结构域, 对特定核酸位点进行切割, 使宿主基因组发生双链断裂 (double strand break, DSB), 然后通过对断裂位点进行修复来引入基因突变^[7]。乳酸菌和大多数原核生物一样, 菌体

内通常缺乏非同源末端连接 (non-homologous end joining, NHEJ) 修复机制, 需导入同源修复模板或借助外源修复途径修复 DSB^[8], 这通常导致构建的打靶质粒较大, 而对于具有较厚细胞壁的乳酸菌来说, 增加了打靶质粒导入宿主细胞的电转难度, 这也是 CRISPR-Cas9 技术在乳酸菌应用时受到的挑战。

最初的 CRISPR-Cas 系统是在细菌基因组上发现的, 它是细菌用来抵抗噬菌体和外源核酸侵染的获得性免疫系统^[9]。乳酸菌基因组上同样含有丰富的 CRISPR-Cas 资源, Crawley 等的研究表明, 1 262 个乳杆菌基因组中, 59.7% 的乳杆菌基因组中可以检测到 CRISPR 阵列, 约 40% 乳杆菌有完整的 CRISPR-Cas 系统, 对这些有完整 CRISPR-Cas 系统的乳杆菌进一步分析, 结果显示 II 型系统丰度最高, I 型次之, III 型系统丰度最低, 大部分乳杆菌基因组上仅含有一种系统^[10]。笔者团队对副干酪乳杆菌 (*Lactobacillus paracasei*) 的研究表明, NCBI 数据库中已经过全基因组测序的 58 株副干酪乳杆菌中, CRISPR 系统的发生率为 43%, 其中 20 株含有 II-A 型系统, 5 株含有 I-E 型系统, 2 株同时含有这两种系统。对 II-A 型系统进行了全面的解析和表征, 该亚型系统重复序列的长度为 36 个核苷酸, 重复序列保守可以形成稳定的二级结构, 不同菌株间隔子数量为 11–25 个不等, 碱基排列顺序非常多样^[11]; tracrRNA 均存在于 *cas9* 和 *cas1* 基因中间, 转录方向与 *cas* 基因相反; 通过对所有间隔子序列和原型间隔子序列进行同源性分析, 初步预测副干酪乳杆菌的 Cas9 蛋白 (*Lactobacillus paracasei* Cas9, lpCas9) 所识别的原型间隔区相邻基序 (protospacer-adjacent motif, PAM), 通过流式细胞术实验和质粒干扰实验进一步解析了 lpCas9 识别的有效 PAM^[12]。本团队已利用此系

统将一株副干酪乳杆菌基因组上的 *terminase* 基因 (与溶源噬菌体包装过程相关的关键蛋白) 成功敲除, 获得一株原噬菌体删除菌株, 解决因噬菌体进入裂解途径导致发酵失败的问题。目前外源 CRISPR-Cas 系统在真核生物中应用广泛, 但应用于乳酸菌仍然受到一些限制。充分利用乳酸菌基因组上已有的 CRISPR-Cas 组件, 构建的打靶质粒只需要包含相应的 crRNA 和待编辑位点两侧翼的修复模板 DNA, 理论上即可成功实现基因编辑。这种策略的优势是: 构建的打靶载体无 Cas 核酸酶基因, 可极大缩减载体大小, 更易转化进乳酸菌, 另外也可以解决异源的 SpyCas9 蛋白对部分乳酸菌的细胞毒性问题。本文重点对利用乳酸菌自身 CRISPR-Cas 组件开展基因编辑的研究进展进行总结。

1 乳酸菌 CRISPR-Cas 系统

1.1 CRISPR 系统

CRISPR-Cas 系统是一种广泛分布于细菌和古菌基因组上, 用来抵御质粒、噬菌体 DNA 等外源遗传物质入侵的获得性免疫系统。此系统主要由 CRISPR 阵列和 Cas 蛋白家族组成, 这两部分在不同的微生物之间存在高度多样性。CRISPR 阵列由一系列高度保守的短回文重复序列 (direct repeat, DR) 和长度相似但高度分化的间隔序列 (spacer) 相间排列而成^[13-14]。CRISPR 阵列侧翼常伴随一系列结构功能保守的基因簇, 被称为 Cas 基因簇。不同类型的 CRISPR 系统含有的 Cas 数目和种类各不相同, 它们在参与不同免疫阶段表现出多种类型的酶活性, 例如核酸酶活性、解旋酶活性和聚合酶活性等。根据干扰阶段发挥作用的 Cas 蛋白的不同将 CRISPR-Cas 系统分为两大类 (Class1 和 Class2)^[15]、6 种类型 (Type I 到 VI), Class1 系统通过多种 Cas 蛋白复合物参与降解外源遗传

物质, 包括 Type I 型、Type III 型、Type IV 型; Class2 系统通过使用单个大 Cas 蛋白执行相同的功能, 包括 Type II 型、Type V 型、Type VI 型^[16-17]。目前利用内源 CRISPR-Cas 系统成功实现基因编辑的乳酸菌多数含有 I-E 型系统 CRISPR-Cas 系统或 II-A 型 CRISPR-Cas 系统, 这两类系统主要区别为系统在表达和干扰阶段发挥作用的 Cas 蛋白种类不同。其中 I-E 型系统带有的 Cas 蛋白基因簇为 *cas3-cse1-cse2-cas7-cas5-cas6-cas1-cas2*^[18], Cas1 和 Cas2 蛋白在适应阶段共同发挥作用^[19], Cas6 蛋白负责将 pre-crRNA 加工成熟, 成熟的 crRNA 招募多个 Cas7 亚基、Cas5 蛋白、Cse2 蛋白、Cse1 蛋白组装形成级联复合物 (cascade); Cas3 蛋白则负责外源 DNA 的切割与降解, 即在干扰阶段 crRNA-cascade 二元复合物与序列相同的外源 DNA 配对时会招募 Cas3 蛋白, 形成的 Cas3-cascade 复合物降解非靶标的单链 DNA^[20]; 而 II-A 型 CRISPR-Cas 系统含有的基因簇为 *cas9-tracrRNA-cas1-cas2-csn2*^[21], pre-crRNA 的成熟和外源遗传物质的降解都由 Cas9 蛋白完成, 干扰阶段 tracrRNA 与 crRNA 的重复序列局部配对并招募 Cas9 蛋白形成三元复合物结构, 当此复合物与外源 DNA 配对时, Cas9 蛋白的两个结构域分别降解外源 DNA 的两条链^[22]。

CRISPR-Cas 系统介导的免疫包括适应、表达、干扰 3 个阶段^[14]。适应阶段: 即新闻隔子的获取过程, 当外源遗传物质入侵宿主时, 在相关 Cas 蛋白的辅助下将其捕获形成特定长度的片段并整合到最靠近前导序列的位置, 形成新闻隔子^[23]; 表达阶段: 在前导序列的驱动下, CRISPR 阵列转录形成前体 pre-crRNA, 在单一 Cas 蛋白或 Cas 蛋白复合体作用下, pre-crRNA 被切割加工成成熟的 crRNA^[24]; 干扰阶段: 破坏外源遗传物质的过程, 当外源遗传物质再次

入侵宿主细胞时, 成熟的 crRNA 与 Cas 蛋白形成核糖蛋白复合物识别相同序列的外源遗传物质并将其降解, 该复合物可识别入侵核酸 PAM 的方式区分“自身”和“非自身”逃避自我攻击^[25]。

1.2 微生物利用内源 CRISPR-Cas 系统产生自我靶向现象

如果宿主在适应阶段错误获取了针对宿主 DNA 的免疫记忆, 即把宿主基因组片段当作原始间隔序列整合进 CRISPR 系统的间隔序列区形成新的 Spacer 序列, 宿主的 CRISPR 系统会把自身基因组的片段错误识别为外源核酸, 从而发挥切割核酸功能导致宿主染色体裂解, 存在自身免疫的风险^[26], 这种现象在乳酸菌宿主中是普遍存在的^[27]。

这种自然发生的微生物利用内源 CRISPR-Cas 系统自我切割现象主要由两种原因导致, 即质粒的水平转移和溶源噬菌体进入裂解循环时对宿主基因组的错误包装^[28-29]。尽管自我靶向现象普遍发生, 且发生自我靶向的宿主基因组被降解, 受到强烈的负选择, 但是宿主可以通过发生突变或编码抗 CRISPR 蛋白逃避 CRISPR-Cas 系统对自身的攻击^[30-31]。乳酸菌基因组上也存在这种靶向自身基因组的间隔子, 例如嗜热乳杆菌、嗜酸乳杆菌均报道过自我靶向和靶向逃逸事件^[26]。这类在乳酸菌菌体细胞内发生的自然事件给予了研究者利用该内源系统实现菌体基因组的精准编辑一个重要启示。

1.3 利用内源 CRISPR-Cas 系统进行基因编辑

II 型 CRISPR-Cas 系统结构简单紧凑、遗传背景被研究得较为透彻, 已经开发出了成熟的基因编辑工具并在多种生物中成功应用, 例如酵母^[32]、小鼠^[33]、斑马鱼^[34]、植物^[35]和人类细胞^[36]等。然而 CRISPR-Cas9/Cpf1 技术在乳酸菌中的应用受到了一定的限制, 主要原因是:

(1) 构建一个异源表达载体需要表达一个较大的 Cas9 蛋白基因 (约 4.2 kb)、0.4 kb 的 sgRNA 以及约 2 kb 的同源重组模板 DNA, 大尺寸的编辑载体不仅难以在微生物中稳定存在, 且对涉及多个 sgRNA 和编辑模板的多基因编辑也有很大限制; (2) 分子遗传工具的限制, 乳酸菌是一类细胞壁较厚的革兰氏阳性菌, 较大的编辑质粒通常导致较低的转化效率, 影响了 CRISPR-Cas9 技术在乳酸菌中的应用; (3) 目前应用较广泛的 SpyCas9 蛋白对部分乳酸菌宿主有细胞毒性。乳酸菌基因组上含有丰富的 CRISPR-Cas 资源, 在乳酸菌菌体开展基因编辑, 充分利用乳酸菌基因组上的内源 CRISPR-Cas 组件, 可以克服现有编辑系统在乳酸菌上的限制, 提高编辑效率。

利用内源 CRISPR-Cas 系统建立基因编辑工具, 首要需要解决的问题是解析该物种 CRISPR 系统效应蛋白所识别的 PAM, 只有明确该 Cas 蛋白识别的 PAM, 才能够有效激活该蛋白的“基因剪刀”功能。目前已经有多数细菌利用内源性 CRISPR-Cas 系统实现基因编辑的实例: 如利用酪丁酸梭菌 (*Clostridium tyrobutyricum*) 基因组上的 I-B 型系统实现了多重基因编辑, 提高菌株的丁醇生产水平^[37]; 利用铜绿假单胞菌 (*Pseudomonas aeruginosa*) 基因组上的 I-F 型系统删除耐药性相关基因, 从临床分离物中分离耐药病原体^[38]。这些利用细菌内源 CRISPR 系统成功开展基因组编辑的研究为在乳酸菌上应用提供了宝贵的经验。

2 乳酸菌利用内源 CRISPR-Cas 系统基因编辑研究进展

虽然乳酸菌基因组上含有丰富的 CRISPR-Cas 系统资源, 但是 CRISPR-Cas 含量和类型因

乳酸菌菌株水平而异, 鲜少有乳酸菌的内源性系统被表征或被开发成新的基因编辑工具。目前为止, 已经被表征的乳酸菌种类有嗜热链球菌 (*Streptococcus thermophilus*)^[39]、清酒乳杆菌 (*Lactobacillus sakei*)^[40]、加氏乳杆菌 (*Lactobacillus gasseri*)^[41]、双歧杆菌 (*Bifidobacterium*)^[42]、卷曲乳杆菌 (*Lactobacillus crispatus*)^[43]、乳酸片球菌 (*Pediococcus acidilactici*)^[44]等 (表 1)。

利用乳酸菌基因组上的内源 CRISPR-Cas 系统实现基因编辑, 首先要表征菌种基因组上的 CRISPR 系统, 并解析 Cas 蛋白识别的 PAM, 即可利用这些信息针对靶标基因构建打靶质粒^[52], 主要包括以下步骤 (图 1)。

(1) CRISPR-Cas 系统类型的鉴定: 在 GenBank 上获取某一菌种的所有基因组数据, 利用 CRISPRCasFinder 和 CRISPRdb 鉴定这些菌体基因组包含的 CRISPR 系统类型 (cas 基因)、间隔子序列和重复序列。

(2) 重复序列二级结构的预测: 利用 RNA Fold Web Sever 软件对重复序列进行二级结构的预测, 包括重复序列的碱基配对及茎环结构等。

(3) 间隔序列来源分析: 间隔子对应的原间隔序列一般来源于噬菌体、质粒等, 使用 CRISPR Target (http://crispr.otago.ac.nz/CRISPRTarget/crispr_analysis.html) 针对数据库中所有已经公开的质粒和噬菌体基因组进行搜索比对。

(4) 预测 tracrRNA 的位置、序列: tracrRNA 典型特征是具有一个反向重复序列结构, 该序列可以和 pre-crRNA 中的重复序列部分互补配对形成复合物, 同时 tracrRNA 经常出现在 4 个位置: cas9 上游、cas9 和 cas1 之间、CRISPR 阵列的前后, 根据上述特征可以比较容易地确定 tracrRNA 的全长序列。

(5) PAM 的预测及筛选: 使用 WebLogo 程序对原间隔序列进行分析, 预测可能的 PAM。

表 1 乳酸菌中已经被表征的 CRISPR-Cas 系统汇总

Table 1 Summary of characterized CRISPR-Cas systems in lactic acid bacteria

Species	CRISPR types	Predicted PAM 5'→3'	Application	Efficiency	References
<i>Streptococcus thermophilus</i>	I -E	I -E AGG	Type II -A system used for self-targeting, resulting in genomic island excision (37.4 kb)	None	[39,45]
	II -A	II -A AGAAW			
	II -C	II -C GGNG			
<i>Streptococcus mutans</i>	II -A	NGG	Endogenous system utilized for deletion in <i>gtfs</i> gene, resulting in decrease of EPS synthesis and thinner biofilms	None	[46]
<i>Lactobacillus crispatus</i>	II -A	I -E AA	Type I -E system reprogrammed for deletion in <i>p-gtf</i> gene, insertion stop codons in <i>p-gtf</i> gene, and single base substitution in <i>p-gtf</i> gene	Deletion: 100%; insertion: 36%; single base substitution: 19%	[43]
	I -B				
	I -E				
<i>Pediococcus acidilactici</i>	II -A	NGG	Endogenous system used for deletion in <i>pyre</i> gene, insertion in <i>l-ldh</i> gene, point mutation in <i>mpi</i> gene	Deletion: average 68.75%; insertion: close to 100%; point mutation: 90%	[44]
<i>Lentilactobacillus buchneri</i>	II -A	NAAAA	None	None	[47]
<i>Bifidobacterium longum</i>	I -C	I -C TTC	None	None	[48]
	I -E	I -E NAAG			
	I -U	I -U TAN			
	II -C	II -C GCN			
<i>Lactobacillus gasseri</i>	II -A	cTAACc	None	None	[10,49-50]
<i>Lactobacillus sakei</i>	II -A	II -A RAAA	None	None	[40]
	II -C	II -C RYAC			
<i>Lactobacillus casei</i>	II -A	tGAAAA	None	None	[10]
<i>Lacticaseibacillus rhamnosus</i>	II -A	aGAAA	None	None	[10]
<i>Lactobacillus jensenii</i>	II -A	tGGc	None	None	[10]
<i>Lactobacillus pentosus</i>	II -A	TTAAT	None	None	[10]
<i>Lactobacillus fermentum</i>	II -A	YAAA	None	None	[51]

N: A/T/C/G; W: A/T; R: A/G; Y: C/T.

进行细胞实验验证 PAM 的有效性,即构建含有候选 PAM 文库的干扰质粒转化宿主细胞,通过宿主细胞的表型来判断宿主细胞内是否发生了预期的基因编辑事件,从而筛选有效的 PAM。

(6) 利用质粒干扰实验检测内源 CRISPR 系统活性:将 spacer 序列作为靶标序列,构建

一系列带有 spacer 和 PAM 的抗性质粒,将这些质粒电转进宿主细胞,利用抗性平板筛选,若与阳性对照相比转化效率显著下降,则说明步骤 5 筛选的 PAM 是有效的,并且内源系统活性较好,在宿主基因组上的 spacer 位点处发生了切割,导致宿主死亡。

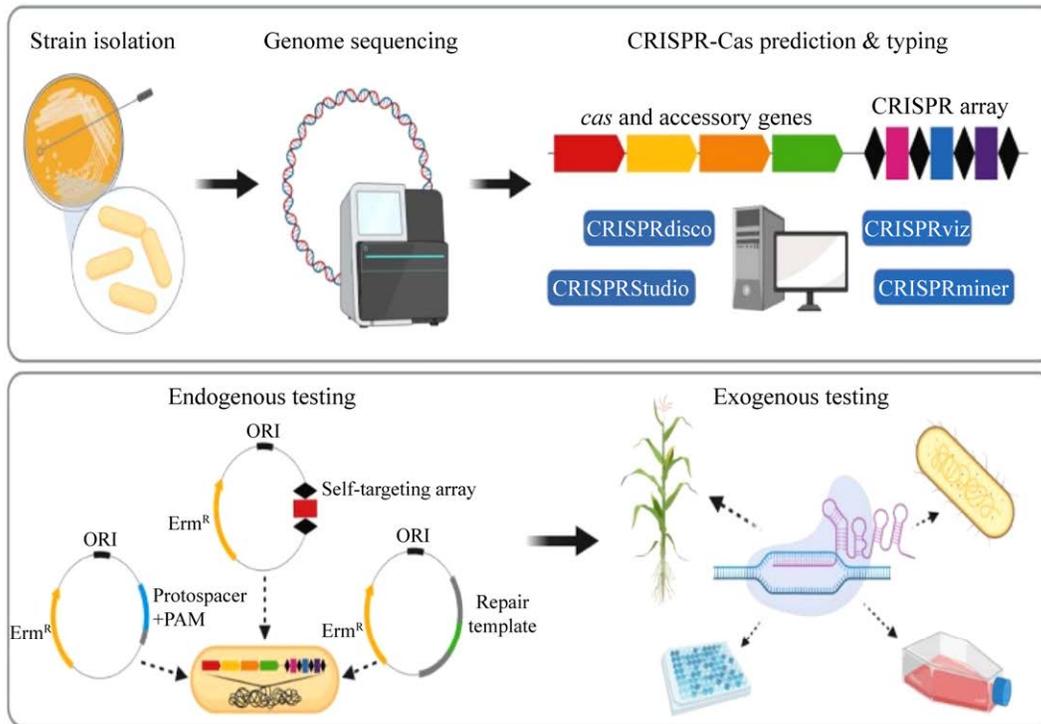


图 1 CRISPR-Cas 系统预测和表征工作流程图^[52]

Figure 1 CRISPR-Cas prediction and characterization workflow^[52].

(7) 体内实验实现基因编辑:即设计打靶质粒验证是否可以在宿主体内发生基因编辑。打靶质粒的构建,模拟 CRISPR 系统的结构,在骨架质粒上插入 mini-CRISPR 基因盒和同源修复模板,mini-CRISPR 基因盒是 leader-repeat-spacer-repeat 结构;对于 II 型系统,也可以直接将成熟的 crRNA 序列和 tracrRNA scaffold 结构合二为一设计成一个独立的 sgRNA 序列,如图 2 所示,图中的 *Bsa* I 属于 Type II S 型限制性内切酶,切割位点在识别位点 (5'-GGTCTCN-3') 下游,因此两个 *Bsa* I 酶切位点间可以随意更换不同的 crRNA,将打靶质粒电转进宿主细胞,用红霉素抗性平板筛选阳性克隆并进一步测序验证。

2.1 利用卷曲乳杆菌内源 I-E 型系统实现多种类型基因突变

卷曲乳杆菌是一类存在于人类阴道和家禽

肠道的重要共生益生菌,对人类和动物健康有重要影响^[53-54],但是由于缺乏有效的分子遗传学工具、细胞壁较厚导致转化效率有限、遗传背景不清晰等原因,对卷曲乳杆菌的生理生化特性研究进展缓慢^[55]。Hidalgo-Cantabrana 等研究发现,卷曲乳杆菌基因组包含有完整的 I-E 型 CRISPR 系统,通过分析其间隔子序列与噬菌体、质粒和细菌染色体的同源性,预测该菌 Cas 蛋白识别的 PAM 是 5'-AA-3'。通过设计一个仅带 sgRNA 不带修复模板的靶向自身基因组的自我靶向质粒,电转至卷曲乳杆菌中,结果发现细胞自我靶向死亡概率是 99%,证实该菌的 CRISPR 系统比较活跃。随后研究者设计了 3 种带不同修复模板的打靶质粒分别验证基因编辑系统在基因缺失、基因插入和单碱基替换的编辑效率,结果显示 *p-gtf* 基因缺失效率是 100%,终止密码子插入效率是 36%,单碱基替

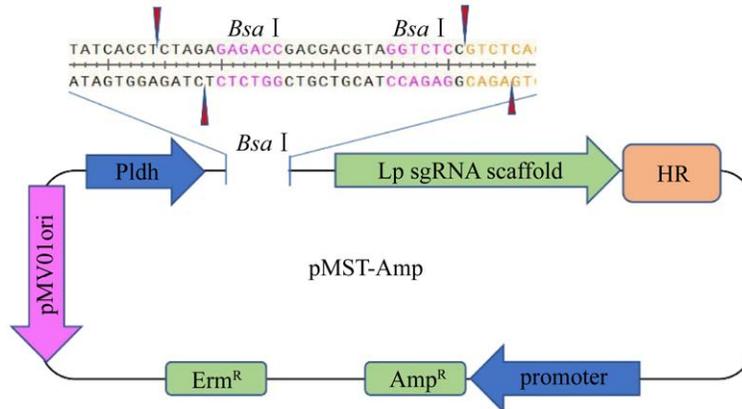


图2 打靶质粒图谱

Figure 2 The map of targeting plasmid. Targeting plasmid is usually a shuttle plasmid, containing the following basic elements: Pldh: promoter for sgRNA expression; *Bsa* I site: crRNA insertion site; Lp sgRNA scaffold: tracrRNA sequence portion of *Lactobacillus paracasei* sgRNA; HR: homologous recombination template DNA; promoter- Amp^R : ampicillin resistance sequence operon; Erm^R : erythromycin resistance; pMV01ori: origin of replication.

换效率是 19%。利用 *L. crispatus* 的内源 I-E 型 CRISPR-Cas 系统在该菌基因组上成功引入了多个突变,为进一步理解 *L. crispatus* 的生理生化特征奠定了基础^[43]。

2.2 利用变形链球菌 II-A 型系统敲除 *gtfs* 基因

变形链球菌是引起人类龋齿的主要病原菌^[56],它不仅可以利用口腔内残留的糖类发酵产生大量酸腐蚀牙齿的釉层;还可以合成胞外多糖 (exopolysaccharides, EPS) 为其他细菌的定殖和聚集提供黏附位点,在牙齿上形成生物膜。为了更好地研究这种生物膜的形成和这类产 EPS 口腔病原体的功能基因组学,研究者开发了各种靶向目标基因的分子生物学工具,如克隆独立无标记诱变系统、反选择诱变系统^[57]。虽然这些方法可以产生变形链球菌突变体,但无标记突变需两次转化,效率低,程序繁杂。李雨庆等对变形链球菌 UA159 的基因组进行分析,确定该菌株含有完整的 II-A 型 CRISPR-Cas

系统^[58],利用质粒干扰实验证实了变形链球菌 Cas9 识别的有效 PAM 是 5'-NGG-3',与应用广泛的 SpyCas9 识别的 PAM 一致,最后选取 5'-TGG-3'为后续实验所用的 PAM^[46]。研究团队以 *gtfs* 为靶标基因,利用变形链球菌基因组上的 II-A 型 CRISPR-Cas 系统元件,实现了 *gtfs* 基因的缺失。模仿宿主体内天然的 II-A 型 CRISPR 阵列的排布,设计出一个带有 mini-CRISPR 阵列的自我靶向质粒表达盒,在 *gtfs* 基因 3 个不同位点设计相应的 crRNA,分别将这 3 个自我靶向质粒和修复模板转化进宿主细胞,结果显示成功编辑的变形链球菌 EPS 合成量明显减少,生物膜的形成能力显著下降。这项工作为在变形链球菌中实现基因编辑提供了理论依据,为研究变形链球菌 EPS 合成和生物膜的形成机制奠定了基础^[46]。

2.3 利用乳酸片球菌 II-A 型系统实现多种类型基因突变

乳酸片球菌常用于片球菌素的生产 and 乳酸

发酵,但是除了传统的同源重组方法外,该物种目前没有高效的基因组编辑工具。Liu 等^[44]对最近分离出的乳酸片球菌 LA412 菌株基因组进行解析,开发了一种基于内源性 II-A 型 CRISPR-Cas 系统的基因组编辑工具,实现了乳酸片球菌的无标记基因缺失、基因插入和点突变。利用质粒干扰实验证实了当 PAM 是 5'-NGG-3'时,乳酸片球菌的 Cas9 蛋白有干扰活性。以 *pyrE* 基因(编码核糖基转移酶)为靶标,设计了 4 种识别不同 PAM 的编辑质粒进行电转,全部成功实现基因的缺失,识别 PAM 为 AGG、CGG、GGG、TGG 的质粒缺失效率分别为 50%、50%、75%和 100%;以 *mpi* 基因(编码甘露糖磷酸异构酶)为靶标,将突变的核苷酸引入编辑质粒的同源臂,在基因组靶位点成功实现了 T→C 单核苷酸突变,效率为 90%;设计编辑质粒,将 *l-ldh* 基因(编码乳酸脱氢酶)插入在乳酸片球菌的非编码区域(*GE01508* 和 *GE01509* 基因位点之间)编辑效率接近 100%,电穿孔后,几乎所有的菌落都被鉴定为混合基因型。这项工作为后续乳酸片球菌益生特性的改良奠定了重要基础,例如利用乳酸片球菌的内源 CRISPR-Cas 系统删除有毒基因、加强乳酸生产途径、提高片球菌素分泌量等。

2.4 利用嗜热链球菌 II-A 型系统实现基因组大片段缺失

与上述利用内源 CRISPR-Cas 系统实现靶标基因编辑的例子不同,研究者在研究嗜热链球菌自我靶向逃逸机制时观察到基因组大片段缺失^[45]。嗜热链球菌基因组上存在 II-A 型和 I-E 型 CRISPR-Cas 系统^[59],为了比较两种系统的自我靶向效率,选择 *lacZ* 基因为靶标,依据两类系统的 crRNA 结构分别设计了两类自我靶向质粒,将这两类自我靶向质粒电转进嗜热链

球菌中,与阴性对照相比,电转了自我靶向质粒的细胞转化效率大大降低,但仍有少部分存活,且电转了 I-E 型自我靶向质粒与电转了 II-A 型自我靶向质粒的细菌存活率比例为 9:1^[39]。

对存活菌株基因组测序分析显示,所有 I-E 型幸存细胞含有完整的 *lacZ* 基因,但将幸存细胞的质粒回收并测序发现, I-E 型质粒是有缺陷的,它缺少间隔子和一个重复序列,所以 I-E 型系统自我靶向逃逸的主要机制是间隔子序列的缺失;所有 II-A 型幸存细胞不再含有 *lacZ* 基因,在靶标位点缺失了 37.4 kb,占基因组总长的 2%,进一步分析发现 *lacZ* 基因的上下游存在两个独立的 *galE* 基因,且这两个基因即为 *lacZ* 基因缺失边界,这表明嗜热链球菌通过两个保守的基因间发生同源重组,修复了 II-A 型系统靶向基因组造成的损伤,这种大片段基因缺失在野生型菌群里以较低的概率发生,基于 CRISPR-Cas 系统的自我靶向可以筛选出这种罕见的基因组重塑事件。

3 总结与展望

尽管基于 SpyCas9 蛋白为核心的 II 型 CRISPR-Cas 基因编辑系统是目前机制研究最清晰、应用最广泛的基因组编辑系统,但是仍然无法满足种类繁多的乳酸菌多种应用场景的需求。相较于不断优化 CRISPR-Cas9 技术,充分利用乳酸菌自身基因组的 CRISPR 相关组件,构建内源性 CRISPR-Cas 系统对乳酸菌进行自我靶向基因编辑,可以克服目前 CRISPR-Cas9 技术在乳酸菌应用上的缺陷和不足,未来可能更有潜力。利用内源 CRISPR/Cas 基因编辑工具箱对乳酸菌进行基因修饰,以改善益生菌性状、增强其发酵性能、定殖宿主肠道将具有广泛的应用前景。例如利用内源 CRISPR 系统,结合

sgRNA 和带有突变位点的模板, 将单核苷酸突变引入嗜热链球菌的 *epsC* 基因上, 可以增加 EPS 的合成、改善酸奶的流变特性、增强其感官特性; 利用已经表征的双歧杆菌 CRISPR-Cas 系统, 将外源 β -半乳糖苷酶插在双歧杆菌基因组上, 可以提高双歧杆菌代谢寡糖的能力, 从而增强该菌株在肠道的定殖能力。但是这种自我靶向基因编辑技术仍然面临一些问题需要解决: (1) 特异性 PAM 的确定。同一物种存在多种类型的 CRISPR 阵列, 甚至同一类型的 CRISPR-Cas 系统有多种潜在的 PAM 位点, 但是目前生物信息预测 PAM 的手段有限, 预测的 PAM 可能与 Cas 蛋白所识别的最优 PAM 有偏差, 导致基因编辑效率不高。现阶段解决这个问题只能依赖于构建数量庞大的 PAM 库, 通过实验来筛选有效的 PAM, 这大大增加了实验的工作量。(2) 无法生成适合多种微生物的通用工具箱, 因为 PAM 位点具有物种特异性, 不同生物识别的 PAM 位点的长度、简单性、位置都不同, 针对不同的研究对象都要重新确定 PAM 位点。后期还需要对每种类型的 CRISPR-Cas 系统进行深入研究确定常见的 PAM 的位点, 通过评估高度普遍的 PAM 或经过验证的同源物种的 PAM 来确定目的菌株的 PAM 位点。

当前我国人均 GDP 超过 1 万美元, 已昂首迈入营养健康时代, 人民群众对于营养健康的追求比以往任何时候都迫切。在这个时代, 促进人体健康方面具有巨大潜力的乳酸菌遇到飞速发展的基因编辑技术, 我们完全可以利用安全的乳酸菌作为底盘细胞按照需要去定向改良乳酸菌菌株的发酵性能、益生性能, 甚至我们可以畅想在不久的未来, 按照合成生物学的理念去设计、开发和定制新型乳酸菌, 达到促进人类健康和治疗疾病的目的, 这将对整个社会生产生活方式产生巨大的影响。

REFERENCES

- [1] Huang H, Song X, Yang S. Development of a RecE/T-assisted CRISPR-Cas9 toolbox for *Lactobacillus*. *Biotechnol J*, 2019, 14(7): e1800690.
- [2] Ganji-Arjenaki M, Rafieian-Kopaei M. Probiotics are a good choice in remission of inflammatory bowel diseases: a meta analysis and systematic review. *J Cell Physiol*, 2018, 233(3): 2091-2103.
- [3] Zitvogel L, Daillère R, Roberti MP, et al. Anticancer effects of the microbiome and its products. *Nat Rev Microbiol*, 2017, 15(8): 465-478.
- [4] Singleton MR, Dillingham MS, Gaudier M, et al. Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. *Nature*, 2004, 432(7014): 187-193.
- [5] Zhang Y, Buchholz F, Muyrers JP, et al. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat Genet*, 1998, 20(2): 123-128.
- [6] Van Pijkeren JP, Britton RA. High efficiency recombineering in lactic acid bacteria. *Nucleic Acids Res*, 2012, 40(10): e76.
- [7] Paix A, Folkmann A, Goldman DH, et al. Precision genome editing using synthesis-dependent repair of Cas9-induced DNA breaks. *PNAS*, 2017, 114(50): e10745-e10754.
- [8] Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, 2014, 157(6): 1262-1278.
- [9] Deveau H, Garneau JE, Moineau S. CRISPR/Cas system and its role in phage-bacteria interactions. *Annu Rev Microbiol*, 2010, 64: 475-493.
- [10] Crawley AB, Henriksen ED, Stout E, et al. Characterizing the activity of abundant, diverse and active CRISPR-Cas systems in lactobacilli. *Sci Rep*, 2018, 8(1): 11544.
- [11] Song X, Zhang XY, Xiong ZQ, et al. CRISPR-Cas-mediated gene editing in lactic acid bacteria. *Mol Biol Rep*, 2020, 47(10): 8133-8144.
- [12] Yang L, Li WX, Ujiroghene OJ, et al. Occurrence and diversity of CRISPR loci in *Lactobacillus casei* group. *Front Microbiol*, 2020, 11: 624.
- [13] Barrangou R, Marraffini LA. CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. *Mol Cell*, 2014, 54(2): 234-244.
- [14] Barrangou R, Fremaux C, Deveau H, et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 2007, 315(5819): 1709-1712.

- [15] Makarova KS, Wolf YI, Koonin EV. Classification and nomenclature of CRISPR-Cas systems: where from here? *CRISPR J*, 2018, 1(5): 325-336.
- [16] Westra ER, Van Houte S, Gandon S, et al. The ecology and evolution of microbial CRISPR-Cas adaptive immune systems. *Philos Trans R Soc Lond B Biol Sci*, 2019, 374(1772): 20190101.
- [17] Makarova KS, Wolf YI, Alkhnbashi OS, et al. An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol*, 2015, 13(11): 722-736.
- [18] Makarova KS, Zhang F, Koonin EV. SnapShot: class 1 CRISPR-Cas systems. *Cell*, 2017, 168(5): 946-946.e1.
- [19] Wang JY, Li JZ, Zhao HT, et al. Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR-Cas systems. *Cell*, 2015, 163(4): 840-853.
- [20] Xue CY, Sashital DG. Mechanisms of type I-E and I-F CRISPR-Cas systems in *Enterobacteriaceae*. *Ecosal Plus*, 2019, 8(2).
- [21] Makarova KS, Zhang F, Koonin EV. SnapShot: class 2 CRISPR-Cas systems. *Cell*, 2017, 168(1/2): 328-328.e1.
- [22] Deltcheva E, Chylinski K, Sharma CM, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 2011, 471(7340): 602-607.
- [23] Yosef I, Goren MG, Qimron U. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res*, 2012, 40(12): 5569-5576.
- [24] Liu T, Li YJ, Wang XD, et al. Transcriptional regulator-mediated activation of adaptation genes triggers CRISPR *de novo* spacer acquisition. *Nucleic Acids Res*, 2015, 43(2): 1044-1055.
- [25] Van Houte S, Ekroth AKE, Broniewski JM, et al. The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature*, 2016, 532(7599): 385-388.
- [26] Stern A, Keren L, Wurtzel O, et al. Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet*, 2010, 26(8): 335-340.
- [27] Horvath P, Coûté-Monvoisin AC, Romero DA, et al. Comparative analysis of CRISPR loci in lactic acid bacteria genomes. *Int J Food Microbiol*, 2009, 131(1): 62-70.
- [28] Stachler AE, Turgeman-Grott I, Shtifman-Segal E, et al. High tolerance to self-targeting of the genome by the endogenous CRISPR-Cas system in an archaeon. *Nucleic Acids Res*, 2017, 45(9): 5208-5216.
- [29] Rollie C, Chevallereau A, Watson BNJ, et al. Targeting of temperate phages drives loss of type I CRISPR-Cas systems. *Nature*, 2020, 578(7793): 149-153.
- [30] Maxwell KL. The anti-CRISPR story: a battle for survival. *Mol Cell*, 2017, 68(1): 8-14.
- [31] Fischer S, Maier LK, Stoll B, et al. An archaeal immune system can detect multiple protospacer adjacent motifs (PAMs) to target invader DNA. *J Biol Chem*, 2012, 287(40): 33351-33363.
- [32] Horwitz AA, Walter JM, Schubert MG, et al. Efficient multiplexed integration of synergistic alleles and metabolic pathways in yeasts via CRISPR-Cas. *Cell Syst*, 2015, 1(1): 88-96.
- [33] Wang HY, Yang H, Shivalila CS, et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell*, 2013, 153(4): 910-918.
- [34] Hwang WY, Fu YF, Reyon D, et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol*, 2013, 31(3): 227-229.
- [35] Shan QW, Wang YP, Li J, et al. Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat Biotechnol*, 2013, 31(8): 686-688.
- [36] Cong L, Ran FA, Cox D, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science*, 2013, 339(6121): 819-823.
- [37] Zhang J, Zong WM, Hong W, et al. Exploiting endogenous CRISPR-Cas system for multiplex genome editing in *Clostridium tyrobutyricum* and engineer the strain for high-level butanol production. *Metab Eng*, 2018, 47: 49-59.
- [38] Xu ZL, Li M, Li YR, et al. Native CRISPR-Cas-mediated genome editing enables dissecting and sensitizing clinical multidrug-resistant *P. aeruginosa*. *Cell Rep*, 2019, 29(6): 1707-1717.e3.
- [39] Cañez C, Selle K, Goh YJ, et al. Outcomes and characterization of chromosomal self-targeting by native CRISPR-Cas systems in *Streptococcus thermophilus*. *FEMS Microbiol Lett*, 2019, 366(9): fnz105.
- [40] Schuster JA, Vogel RF, Ehrmann MA. Characterization and distribution of CRISPR-Cas systems in *Lactobacillus sakei*. *Arch Microbiol*, 2019, 201(3): 337-347.
- [41] Stout EA, Sanozky-Dawes R, Goh YJ, et al. Deletion-based escape of CRISPR-Cas9 targeting in *Lactobacillus gasseri*. *Microbiology (Reading)*, 2018, 164(9): 1098-1111.
- [42] Pan MC, Nethery MA, Hidalgo-Cantabrana C, et al. Comprehensive mining and characterization of CRISPR-Cas systems in *Bifidobacterium*.

- Microorganisms, 2020, 8(5): 720.
- [43] Hidalgo-Cantabrana C, Goh YJ, Pan MC, et al. Genome editing using the endogenous type I CRISPR-Cas system in *Lactobacillus crispatus*. Proc Natl Acad Sci USA, 2019, 116(32): 15774-15783.
- [44] Liu L, Yang DL, Zhang ZY, et al. High-efficiency genome editing based on endogenous CRISPR-Cas system enhances cell growth and lactic acid production in *Pediococcus acidilactici*. Appl Environ Microbiol, 2021, 87(20): e0094821.
- [45] Selle K, Klaenhammer TR, Barrangou R. CRISPR-based screening of genomic island excision events in bacteria. PNAS, 2015, 112(26): 8076-8081.
- [46] Gong T, Tang BY, Zhou XD, et al. Genome editing in *Streptococcus mutans* through self-targeting CRISPR arrays. Mol Oral Microbiol, 2018, 33(6): 440-449.
- [47] Briner AE, Barrangou R. *Lactobacillus buchneri* genotyping on the basis of clustered regularly interspaced short palindromic repeat (CRISPR) locus diversity. Appl Environ Microbiol, 2014, 80(3): 994-1001.
- [48] Hidalgo-Cantabrana C, Crawley AB, Sanchez B, et al. Characterization and exploitation of CRISPR loci in *Bifidobacterium longum*. Front Microbiol, 2017, 8: 1851.
- [49] Sanozky-Dawes R, Selle K, O'Flaherty S, et al. Occurrence and activity of a type II CRISPR-Cas system in *Lactobacillus gasseri*. Microbiology (Reading), 2015, 161(9): 1752-1761.
- [50] Anderson EM, McClelland S, Maksimova E, et al. *Lactobacillus gasseri* CRISPR-Cas9 characterization *in vitro* reveals a flexible mode of protospacer-adjacent motif recognition. PLoS One, 2018, 13(2): e0192181.
- [51] Brandt K, Nethery MA, O'Flaherty S, et al. Genomic characterization of *Lactobacillus fermentum* DSM 20052. BMC Genomics, 2020, 21(1): 328.
- [52] Roberts A, Barrangou R. Applications of CRISPR-Cas systems in lactic acid bacteria. FEMS Microbiol Rev, 2020, 44(5): 523-537.
- [53] Wei S, Morrison M, Yu Z. Bacterial census of poultry intestinal microbiome. Poult Sci, 2013, 92(3): 671-683.
- [54] Dec M, Nowaczek A, Stępień-Pyśniak D, et al. Identification and antibiotic susceptibility of lactobacilli isolated from turkeys. BMC Microbiol, 2018, 18(1): 168.
- [55] Beasley SS, Takala TM, Reunanen J, et al. Characterization and electrotransformation of *Lactobacillus crispatus* isolated from chicken crop and intestine. Poult Sci, 2004, 83(1): 45-48.
- [56] Gao L, Xu TS, Huang G, et al. Oral microbiomes: more and more importance in oral cavity and whole body. Protein Cell, 2018, 9(5): 488-500.
- [57] Zhang S, Zou ZZ, Kreth J, et al. Recombineering in *Streptococcus mutans* using direct repeat-mediated cloning-independent markerless mutagenesis (DR-CIMM). Front Cell Infect Microbiol, 2017, 7: 202.
- [58] Chen J, Li TC, Zhou XD, et al. Characterization of the clustered regularly interspaced short palindromic repeats sites in *Streptococcus mutans* isolated from early childhood caries patients. Arch Oral Biol, 2017, 83: 174-180.
- [59] Carte J, Christopher RT, Smith JT, et al. The three major types of CRISPR-Cas systems function independently in CRISPR RNA biogenesis in *Streptococcus thermophilus*. Mol Microbiol, 2014, 93(1): 98-112.

(本文责编 郝丽芳)