

DOI:10.3969/j.issn.1672-5565.2017.01.201609002.

靶向番茄 *SlACS2* 基因 CRISPR-Cas9 sgRNA 的设计和分

白云凤¹, 张爱萍², 闫建俊¹, 贺飞燕³, 张维锋^{1*}, 冯瑞云¹, 刘江娜², 张西英²
(1.农业部黄土高原作物基因资源与种质创制重点实验室,作物遗传与分子改良山西省重点实验室
(山西省农业科学院作物科学研究所),太原 030031;
2. 新疆生产建设兵团第 6 师农业科学研究所,新疆 五家渠 831300;
3.山西大学生物工程学院,太原 030006)

摘 要:番茄为呼吸跃变型果实,伴随呼吸跃变产生大量乙烯,即系统 II 乙烯,易使番茄果实过熟,导致腐烂变质。*SlACS2* 是番茄系统 II 乙烯合成的限速酶,通过 CRISPR-Cas9 基因组编辑系统修饰该基因,调控系统 II 乙烯过量表达,将迟滞番茄过熟。本研究基于 RNA-seq 建立了 *SlACS2* 基因的数字表达谱,表明该基因呈果实特异性表达,在植株的根、茎、叶等部位不表达。*SlACS2* 位于番茄 1 号染色体,含 4 个外显子和 3 个内含子。利用在线工具 CRISPRdirect 和 CRISPR-P 发现第 1、2、3 外显子分别具有 18、9 和 11 条 sgRNA。其中,sgRNA1-14 和 sgRNA3-8 及二者的近 PAM 的 12 nt 种子序列在番茄基因组是唯一一序列,GC 含量高于 40%,不存在 TTTT 终止序列。BLAST 结果表明,sgRNA1-14 和 sgRNA3-8 与 GenBank 公布的 8 条 *SlACS2* 同源序列高度一致,位于该基因的保守区,而与 *SlACS4* 和 *SlACS6* 的同源序列存在多个 SNP,预示这 2 条 sgRNA 可用于番茄不同品种 *SlACS2* 基因的靶向编辑,并可规避对 *SlACS* 家族其他同源基因的脱靶效应。
关键词:番茄;ACC 合成酶 2;sgRNA;CRISPR-Cas9;基因组编辑
中图分类号:Q986;S641.2 文献标志码:A 文章编号:1672-5565(2017)01-007-09

Design and evaluation for sgRNAs targeting *SlACS2* gene in CRISPR/Cas9 system

BAI Yunfeng¹, ZHANG Aiping², YAN Jianjun¹, HE Feiyan³, ZHANG Weifeng^{1*},
FENG Ruiyun¹, LIU Jiangna², ZHANG Xiyong²
(1.Key Laboratory of Crop Gene Resources and Germplasm Enhancement on Loess Plateau, Ministry of Agriculture, Shanxi Province Key Laboratory of Crop Genetics and Molecular Improvement (Institute of Crop Science, Shanxi Academy of Agricultural Sciences), Taiyuan 030031, China;
2.Institute of Agricultural Science, Sixth Division, Xinjiang Production and Construction Corps, Wujiaqu 831300, China;
3.College of Bioengineering, Shanxi University, Taiyuan 030006, China)

Abstract: Tomatoes are typical climacteric fruits and produce large amounts of ethylene, i.e. system II ethylene with the climacteric, which makes the tomato fruits overripe and Perishable. ACC synthase 2 in tomato (*SlACS2*) is the key enzymy during the process of biosynthesis of the system II ethylene. Suppressing the expression of the system II ethylene will postpone the fruits overripen by using CRISPR-Cas9 genome editing system to modify the *SlACS2* gene. In this study the digital expression profile of *SlACS2* based on RNA-seq shows that *SlACS2* gene is expressed specifically in fruit. *SlACS2* gene is located in chromosome 1 of tomato, containing four exons and three introns. In the first, second and third exon, 18, 9 and 11 sgRNAs are found by using related online tools CRISPRdirect and CRISPR-P respectively. sgRNA1-14 and sgRNA3-8, and their 12nt seed sequences proximal the protospacer adjacent motif (PAM) all are unique and no completely homologous sequences in other location of tomato genome is found. The GC content is higher than 40% and there is no termination sequence TTTT in sgRNA1-14 and

收稿日期:2016-09-03;修回日期:2016-10-18.
基金项目:国家自然科学基金项目(No.30971838);山西省应用基础研究项目(No.201601D011075)。
作者简介:白云凤,女,博士,研究员,研究方向:植物分子遗传研究;E-mail: byfok@126.com;
张爱萍,女,研究员,研究方向:园艺植物育种;E-mail: 379409419@qq.com.
* 通信作者:张维锋,男,研究员,研究方向:植物遗传育种研究;E-mail: zhwfen@sina.com.

sgRNA3-8. The BLAST results show that sgRNA1-14 or sgRNA3-8 is highly consistent with the sequences of the homologous sites of eight *SIACS2* homologous genes, suggesting that the two sgRNAs are located in the conserved regions of *SIACS2* gene. Multiple single nucleotide polymorphisms (SNPs) are found between sgRNA1-14 or sgRNA3-8 or their respective seed sequence and the sequences of the homologous sites of *SIACS6* or *SIACS4*, indicating that the two sgRNAs can be used to edit the *SIACS2* gene in different varieties with minimal off-target effects for other members in *SIACS* family.

Keywords: Tomato; ACC synthase 2; sgRNA; CRISPR-Cas9; Genomic editing

乙烯(Ethylene)是植物内源激素,对植物生长发育具有重要调控作用。具有呼吸跃变的果实在呼吸跃变前产生基础水平的少量乙烯,被称为系统 I 乙烯,调控植物正常生长、发育和应对胁迫反应等。系统 I 乙烯具有自抑制机制。随着果实成熟,伴随呼吸跃变产生大量乙烯,被称为系统 II 乙烯。系统 II 乙烯具有自催化机制,促进花的凋亡和果实成熟^[1]。

乙烯的生物合成遵循甲硫氨酸(Methionine, Met) → S-腺苷甲硫氨酸(S-Adenosyl methionine, SAM) → 1-氨基环丙烷 1-羧酸 ACC(1-aminocyclopropane-1-carboxylic acid, ACC) → 乙烯的途径。ACC 是合成乙烯的直接前体,由 ACC 合成酶 ACS(ACC synthase)催化 SAM 形成,这个过程被认为是乙烯合成的限速步骤,ACS 是乙烯合成的限速酶。番茄(*Solanum lycopersicum*)是呼吸跃变型果实,其 ACS(SIACS)由多基因家族编码,*SIACS1A* 和 *SIACS6* 主导系统 I 乙烯的合成,负责番茄果实在呼吸跃变前产生基础水平的少量乙烯,*SIACS2* 和 *SIACS4* 主导系统 II 乙烯的合成^[2],在果实成熟过程中伴随呼吸跃变产生大量乙烯,使得番茄成熟过程难以控制,果实易过熟导致腐烂变质。

通过基因工程手段抑制番茄 *SIACS2* 的表达,主要是把 *SIACS2* 的反义 RNA 基因^[3-4] 或 RNAi 基因^[5-6] 导入番茄,可以减缓番茄果实中的乙烯合成,迟滞果实过熟,延长货架期,提高耐贮运性能。上述利用常规转基因技术,导入的外源基因在植物基因组上的整合位置具有随机性,容易出现一些非预期效应,如转基因整合到宿主基因的“阅读框”,使其不能有效表达,影响基因工程育种效率。民众对于转基因的安全性疑虑,也增加了转基因作物评价和应用成本。

CRISPR-Cas9 是一种新型的基因组定点编辑系统^[7-8],其通过一段 20 nt 的 sgRNA(Single guide RNA)与基因组同源序列互补配对引导 Cas9 蛋白结合于基因组的特异性位点,在 sgRNA 3' 端 PAM(Protospacer adjacent motif, 前间区序列邻近基序)元件上游的第 3~4 bp 处剪切,造成 DNA 双链断裂(Doubled-strand breaks, DSB)。不同 CRISPR-Cas9

的 PAM 序列有所不同,经典酿脓链球菌(*Streptococcus pyogenes*)中的 PAM 元件序列为 5'-NGG-3'。因此,CRISPR-Cas9 系统中靶位点的选择是由 20 nt 的 sgRNA 及 3' 端的 PAM 序列共同决定的。

Cas9 蛋白切割造成的 DNA 双链断裂有两条修复途径^[9],即非同源末端连接(Non-homologous end joining, NHEJ)和同源重组(Homologous recombination, HR)。非同源末端连接是主要修复途径,同源重组只发生在特定的细胞周期和类型中。非同源末端连接容易出错,修复过程中会引入短的缺失或插入,抑制或修饰特定基因功能。与诱发突变相比,CRISPR-Cas9 系统为定点突变,效率高、可控性好。CRISPR-Cas9 系统通过植物基因组原位编辑来抑制或修饰基因功能,无需整合新基因,比转基因技术更为安全可控,易于消除民众安全性疑虑。

CRISPR-Cas9 系统中,20 nt 的 sgRNA 及其 3' 端的 PAM 元件总长度仅 23 nt,易在基因组的非目标区段出现同源序列。另外,Cas9 核酸酶可能切割与 sgRNA 不完全匹配的序列,导致脱靶效应^[10-11]。规避脱靶效应和提高编辑效率是应用 CRISPR-Cas9 系统的关键。本文设计、筛选出了番茄 *SIACS2* 基因的 sgRNA,以期利用 CRISPR-Cas9 技术抑制 *SIACS2* 基因表达、提高番茄的耐贮运性能提供依据。

1 材料与方法

1.1 *SIACS2* 数字表达谱的建立

以番茄功能基因组数据库网站(<http://ted.bti.cornell.edu/cgi-bin/>)中 RNA-seq data 的数据作基础,建立 *SIACS2* 基因的基于 RNA-seq 的数字表达谱。

1.2 *SIACS2* 染色体定位和基因组结构的确定

利用 NCBI(<http://www.ncbi.nlm.nih.gov/>)核酸数据库查询获得 *SIACS2* 的 cDNA,以此作 query,搜索番茄基因组数据库得到相应的 gDNA 序列和所在染色体信息。通过序列比对确定 *SIACS2* 的基因组结构以及外显子所在染色体位置。

1.3 sgRNA 的设计和选择

根据 CRISPR-Cas9 靶点设计原则,利用

CRISPRdirect(<https://crispr.dbcls.jp/>)^[12], 2015) 筛选 *SlACS2* 外显子区域的 sgRNA 序列,即 3' 端有 PAM 元件的 20 个连续的碱基序列,PAM 元件设定为 NGG^[13],sgRNA 的结构为 5'-(N)₂₀NGG-3',N 为任意核苷酸。利用 CRISPR-P(<http://cbi.hzau.edu.cn/crispr/>)^[14]对 sgRNA 的靶向特异性进行综合评分。

2 结果与分析

2.1 基于 RNA-seq 的 *SlACS2* 数字表达谱分析

以番茄功能基因组 RNA-seq 数据库为基础,分析了 *SlACS2* 基因(TC125329)在不同组织以及不同生长时期的的表达模式(见图 1)。结果显示,*SlACS2* 在番茄的种子、胚根、顶端分生组织、不同时期的幼苗、根系、叶片、花蕾、绿果中均无表达,在子房和愈伤组织中有低水平表达,在破色期的果实和

成熟的红果实中有较高水平的表达。上述基于 RNA-seq 的表达谱分析证实了 *SlACS2* 是与果实成熟相关的基因,利用 CRISPR-Cas9 技术对该基因进行原位编辑,有可能减缓番茄内源乙烯的产生,迟滞番茄过熟,提高贮藏品质。

2.2 *SlACS2* 染色体定位和基因组结构

从 GenBank 搜索到 *SlACS2* 序列(GenBank Accession:NM_001247249.2),mRNA 长度 1 855 nt, cds 位于 181~1 629 nt 之间,长度 1 458 nt。

以 *SlACS2* 的 cds 作 Query,在番茄的基因组数据库进行 Blast。比对结果表明,*SlACS* 的 gDNA 位于番茄的 1 号染色体(GenBank Accession:NC_015438.2),位置在 88 456 011~88 453 490 nt 之间,长度 2 522 nt,由 4 个外显子、3 个内含子组成(见图 2),前 3 个外显子均较短,位于 *SlACS2* 的 5' 端,可在每个外显子区域设计 sgRNA,规避 sgRNA 分布在外显子和内含子连接处。

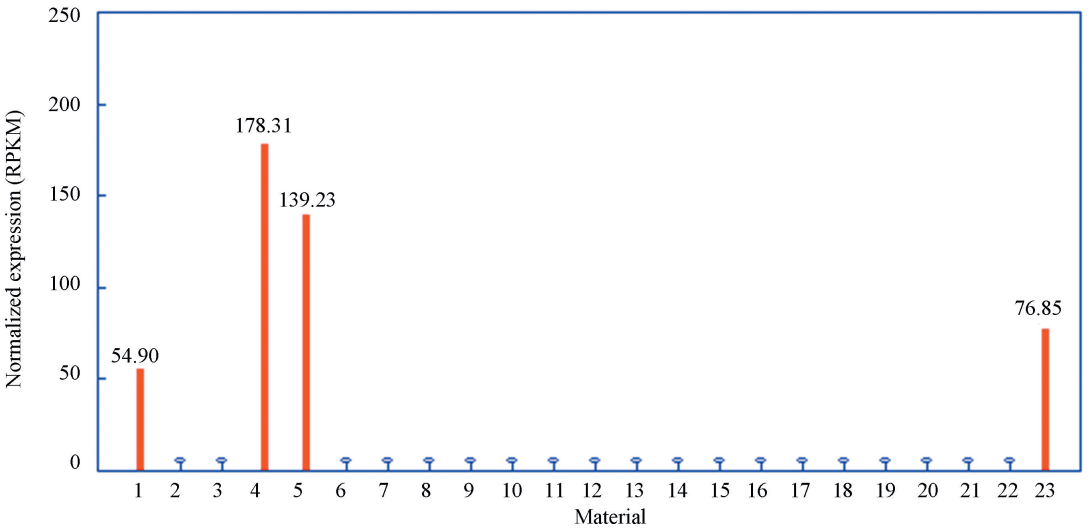


图 1 基于 RNA-seq 的 *SlACS2* 表达谱

Fig.1 The digital expression profile of *SlACS2* based on RNA-sequence

注:1-子房;2-半绿果实;3-成熟绿果实;4-破色期果实;5-成熟红果实;6-花蕾(0~3 mm);7-花蕾(3~8 mm);8-开花前花蕾(8 mm);9-完全展开花;10-发育阶段花的混合物;11-野生番茄花粉;12-感染假单胞菌叶片;13-假单胞菌叶片;14-混合诱导叶片;15-植株的顶端分生组织(4~6 周龄);16-植株顶端分生组织(8 周龄);17-花前根;18-挂果期根;19-营养匮乏的根;20-完全吸涨 5 d 后的胚根;21-完全吸涨 7 d 后的幼苗;22-休眠种子;23-愈伤组织。
PRKM: 每百万 reads 中来自于特定基因每千碱基长度的 reads 数。

Solanum lycopersicum chromosome ch01 (NC 015438.2) *SlACS*



图 2 *SlACS2* 的基因组结构和染色体位置

Fig.2 *SlACS2* enomic costructure and its location on the chromosome

2.3 *SlACS2* 第 1 外显子 sgRNA 的设计和评价

SlACS2 第 1 外显子长度 171 nt,位于染色体

88 456 011~88 455 841 nt 之间。根据 CRISPR-Cas9 靶点设计原则,设定 sgRNA 长度为 20 nt,下游的 PAM

元件为 NGG。CRISPRdirect 在线工具分析表明,第 1 外显子含有 18 条 sgRNA(见表 1)。CRISPR-P 分析表明,这些 sgRNA 的靶向特异性综合得分在 29~49。

在 CRISPR-Cas9 系统中,sgRNA 的转录通常由 U6 或者 U3 启动子驱动,二者驱动转录的活性相对较高。U6 或者 U3 启动子属于 Pol III 型启动子^[15],连续的 TTTT 序列会使转录终止。在第 1 外显子的 18 条 sgRNA 中,有 9 条(sgRNA1-1、-4、-5、-6、-8、-9、-11、-12、-18)含有连续的 TTTT 序列,可排除

选择。另外,sgRNA1-2 的 GC 含量较低,sgRNA1-7 除了在靶位点外,在番茄 5 号染色体上也有与之完全匹配的序列,易产生脱靶效应。在其余 7 条 sgRNA(见图 3)中,位于外显子 132~154 nt 的 sgRNA1-14 特异性最好,该 sgRNA 和它包含的对特异性起重要作用的、邻近 PAM 的 12 nt 种子序列^[16]在番茄基因组上均是唯一序列,其他部位没有与之完全匹配的序列。sgRNA1-14 的 GC 含量 45%,靶向特异性综合得分也最高,可作为优选 sgRNA。

表 1 *SlCAS2* 基因第 1 外显子含有的 sgRNAs
Table 1 sgRNAs of exon 1 in *SlCAS2* gene

No.	Position		Target sequence 20mer+ PAM (total 23mer)	Sequence information			Number of target sites		score
	start	+		GC% of 20mer	Tm(℃) of 20mer	TTTT in 20mer	20mer +PAM	12mer +PAM	
	- end	-							
1-1	23 - 45	-	CCA ACTCAATCTTATCAAAATTG	25.00	57.14	+	1	14	36
1-2	24 - 46	+	CAACTCAATCTTATCAAAAT TGG	25.00	57.14	-	1	27	37
1-3	43 - 65	+	TTGGCTACTAATGAAGAGCA TGG	40.00	68.91	-	1	6	48
1-4	67 - 89	+	GAAAACTCGCCATATTTTGA TGG	35.00	63.59	+	1	3	35
1-5	68 - 90	+	AAAACTCGCCATATTTTGAT GGG	30.00	62.44	+	1	13	31
1-6	71 - 93	+	ACTCGCCATATTTTGATGGG TGG	45.00	70.11	+	1	11	32
1-7	76 - 98	-	CCA TATTTTGATGGGTGGAAGC	40.00	67.75	-	2	5	29
1-8	112 - 134	-	CCT TTCCAGCCTCTAAAAAACCC	45.00	70.92	+	1	2	45
1-9	117 - 139	-	CCA CCCTCTAAAAAACCCCAACG	50.00	71.96	+	1	6	47
1-10	118 - 140	+	CACCCTCTAAAAAACCCCAA CGG	45.00	70.87	-	1	5	46
1-11	120 - 142	-	CCC TCTAAAAAACCCCAACGGAG	45.00	69.93	+	1	10	48
1-12	121 - 143	-	CCT CTAAAAAACCCCAACGGAGT	45.00	70.44	+	1	3	48
1-13	132 - 154	+	CCCCAACGGAGTTATCCAAA TGG	50.00	72.85	-	1	4	49
1-14	132 - 154	-	CCC CAACGGAGTTATCCAAATGG	45.00	68.06	-	1	1	49
1-15	133 - 155	+	CCCAACGGAGTTATCCAAAT GGG	45.00	70.12	-	1	5	49
1-16	133 - 155	-	CCC AACGGAGTTATCCAAATGGG	45.00	70.12	-	1	2	49
1-17	134 - 156	-	CCA ACGGAGTTATCCAAATGGGT	45.00	71.70	-	1	2	48
1-18	147 - 169	-	CCA AATGGGTCTTGCTGAAAATC	40.00	67.24	+	1	3	45

注:字体加粗表示该 sgRNA 具有高特异性;含有连续 TTTT 的靶向序列用灰色表示,含有 polIII 启动子的载体应避免 sgRNA 中含连续的 TTTT 序列。

Solanum lycopersicum (SL2.5), Position:SL2.50ch01:86456010-86455840,Lenth:171 nt,SIACS,exon 1

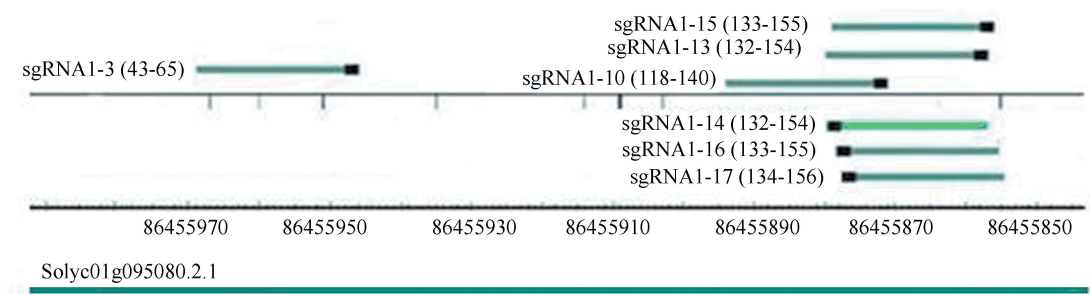


图 3 *SlACS2* 第 1 外显子主要 sgRNAs 的分布。

Fig.3 The position of some important sgRNAs of exon 1 in *SlACS2*

注:括号内数值为该 sgRNA 所在外显子位置。

Cas9 蛋白包含氨基端的 RuvC-like 结构域及位于蛋白中间位置的 HNH 核酸酶结构域。HNH 核酸酶结构域切割与 sgRNA 互补配对的模板链, RuvC-like 结构域切割另 1 条链。突变 Cas9 核酸酶的一个功能结构域时,则只能切割 DNA 双链中的一条链,形成 Cas9 单切口酶(Nickase)。体内 DNA 单链断裂修复采用碱基切除修复途径,不能产生突变。当 1 对分别位于染色体正负链上、方向相反且距离较近的两个 sgRNA (即“Paired-sgRNA”)同时引导 Cas9 单切口酶对基因组编辑时,两个近距离的单链断裂会组成双链断裂,通过非同源末端连接修复,造成短的缺失或插入,导致移码突变;而潜在的脱靶位点处只有单链断裂,修复过程不易发生突变。因此,采用

“paired-sgRNA”和 Cas9 单切口酶编辑策略时,酶切识别所涉及的核苷酸数就由 20 nt 增加到 40 nt,从而提高靶位点的专一性,降低脱靶风险^[17-18]。据此,采用“paired-sgRNA”和 Cas9 单切口酶编辑策略,对第 1 外显子 sgRNA1-3 及其互补链的 sgRNA1-14 进行切割,间隔 66 nt,有较大概率产生专一性移码突变。

2.4 *SlACS2* 第 2、3 外显子 sgRNA 的设计和分析

SlACS2 第 2 外显子位于染色体 86 455 742 ~ 86 455 614 nt 之间,长 129 nt,与第 1 外显子相隔 98 nt 的内含子。CRISPRdirect 在线分析表明,第 2 外显子有 9 条 sgRNA 序列,其 GC 含量均低于 35%,邻近 PAM 12 nt 的种子序列的脱靶位点在 2 个以上(见表 2),没有合适的 sgRNA 可供选择。

表 2 *SlACS2* 基因第 2 外显子含有 sgRNAs
Table 2 sgRNAs of exon 2 in *SlACS2* gene

No.	Position		Target sequence 20mer+ PAM (total 23mer)	Sequence information			Number of target sites		score
	start	+		GC% of	Tm(°C) of	TTTT in	20mer	12mer	
	- end	-		20mer	20mer	20mer	+PAM	+PAM	
2-1	5 - 27	+	GTTTAGACTTGATAGAAGAT TGG	30.00	60.81	-	1	8	46
2-2	25 - 47	+	TGGATTAAGAGAAACCCAA AGG	35.00	65.58	-	1	3	43
2-3	39 - 61	-	CCC AAAAGGTTCAATTGTCTG	30.00	60.44	+	1	18	41
2-4	40 - 62	-	CCA AAAGGTTCAATTGTCTGA	30.00	61.91	-	1	13	42
2-5	43 - 65	+	AAAGGTTCAATTGTCTGA AGG	30.00	61.91	-	1	10	38
2-6	60 - 82	+	TGAAGGAATCAAATCATTCA AGG	30.00	61.46	-	1	14	39
2-7	83 - 105	-	CCA TTGCCAACTTTCAAGATTAT	30.00	62.73	-	1	4	43
2-8	88 - 110	+	GCCAACTTTCAAGATTATCA TGG	35.00	64.09	-	2	20	30
2-9	89 - 111	-	CCA ACTTTCAAGATTATCATGGC	35.00	64.34	-	1	29	46

SlACS2 第 3 外显子位于染色体 86 455 528 ~ 86 455 368 nt 之间,长 161 nt,与第 2 外显子相隔 85 nt 的内含子,与第 4 外显子相隔 881 nt 的内含子。CRISPRdirect 在线分析表明,第 3 外显子有 11 条 sgRNAs。CRISPR-P 分析表明,sgRNAs 的靶向特异性综合得分在 22~48,差异较大(见表 3)。在这些 sgRNAs 中,sgRNA3-9 和 sgRNA3-10 含有连续的 TTTT 序列,易使转录提前终止;sgRNA3-1、-2、-3 的 GC 含量较低;sgRNA3-6 的靶向特异性综合得分最低。在其余 5 条 sgRNAs(见图 4)中,sgRNA3-8 特异性最好,其邻近 PAM 12 nt 的种子序列在番茄全基因组上是唯一序列,其他部位没有与之完全匹配的序列,靶向特异性综合得分也较高。该 sgRNA 3’端第 20 位的碱基为鸟嘌呤 G,意味着基因组编辑效率较高^[19],可作为优选序列。sgRNA3-11 特异性也较好,靶向特异性综合得分也较高,其 12 nt 的种子序列除了靶位点外,在 5 号染色体的 1 个位点上有与其完全匹配的序列,其脱靶效应可根据脱靶

位点两侧的序列(见图 5)设计引物,对 PCR 产物测序予以验证。另外该 sgRNA 的第一碱基为 G,也适合 U6 启动子转录,是另一优选序列。sgRNA3-8 与位于互补链的 sgRNA3-4 间隔 5 nt,sgRNA3-11 与位于互补链的 sgRNA3-7 之间相隔 34 nt,这两对 sgRNA 均可设计为“paired-sgRNA”,用 Cas9 单切口酶进行编辑。

2.5 优选 sgRNA 的 SNP 分析

以 *SlACS2* 的第 1 和第 3 外显子作 Query 在 GenBank 中进行 BLAST,搜索同源序列,进行多序列比对(见图 6,图 7)。比对结果表明,sgRNA1-14、sgRNA3-11 在搜索到的 8 条序列上均完全一致,sgRNA3-8 也只有 1 个 SNP,表明本研究优选的 3 条 sgRNA 均具有较好的保守性,可用于不同品种 *SlACS2* 基因的靶向编辑。而与 GenBank 公布的 *SlACS4* 和 *SlACS6* 比对,sgRNA1-14 均有 5 个 SNP,其中 3 个位于种子序列区;sgRNA3-8 分别有 3 个和 6 个 SNP,且有 1 个 SNP 位于 PAM 的第 2 个核苷酸,

TGG 突变为 TAG;sgRNA3-11 有 3 个 SNP,其中 2 个可规避对 *SIACS* 家族其他成员的脱靶效应。位于种子序列区;预示本研究优选的 3 条 sgRNA 均

表 3 *SIACS2* 基因第 3 外显子含有 sgRNAs
Table 3 sgRNAs of exon 3 in *SIACS2* gene

No.	Position		Target sequence 20mer+ PAM (total 23mer)	Sequence information			Number of target sites		score
	start	+		GC% of 20mer	Tm(°C) of 20mer	TTTT in 20mer	20mer +PAM	12mer +PAM	
3-1	10 - 32	+	AAATTTATGGAGAAAACAAG AGG	25.00	57.51	-	2	20	22
3-2	13 - 35	+	TTTATGGAGAAAACAAGAGG AGG	35.00	63.79	-	2	8	28
3-3	48 - 70	+	TGATCCAGAAAGACTTGTTA TGG	35.00	64.71	-	1	14	44
3-4	52 - 74	+	CCAGAAAGAGTTGTTATGGC TGG	45.00	68.83	-	1	2	42
3-5	52 - 74	-	CCA GAAAGAGTTGTTATGGCTGG	45.00	68.83	-	1	7	43
3-6	55 - 77	+	GAAAGAGTTGTTATGGCTGG TGG	45.00	68.83	-	1	2	14
3-7	64 - 86	+	GTTATGGCTGCTGGTGCCAC TGG	60.00	77.87	-	1	49	48
3-8	80 - 102	-	CCA CTGGAGCTAATGAGACAATT	40.00	66.97	-	1	1	47
3-9	93 - 115	+	TGAGACAATTATATTTTGT TGG	20.00	55.94	+	1	54	25
3-10	103 - 125	+	ATATTTTGTTTGGCTGATCC TGG	35.00	64.83	+	1	5	30
3-11	121 - 143	-	CCT GGCGATGCATTTTATGACC	45.00	69.51	-	1	2	47

注:字体加粗显示该 sgRNA 具有高特异性;含有连续 TTTT 的靶向序列用灰色表示,含有 polIII 启动子的载体应避免 sgRNA 中含连续的 TTTT 序列。

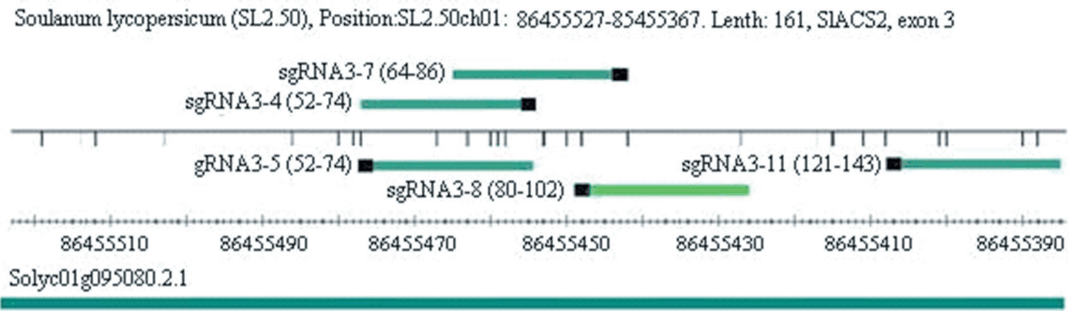


图 4 *SIACS2* 第 3 外显子主要 sgRNAs 的分布
Fig.4 The position of some important sgRNAs of exon 3 in *SIACS 2*

注:括号内数值为该 sgRNA 所在外显子位置。

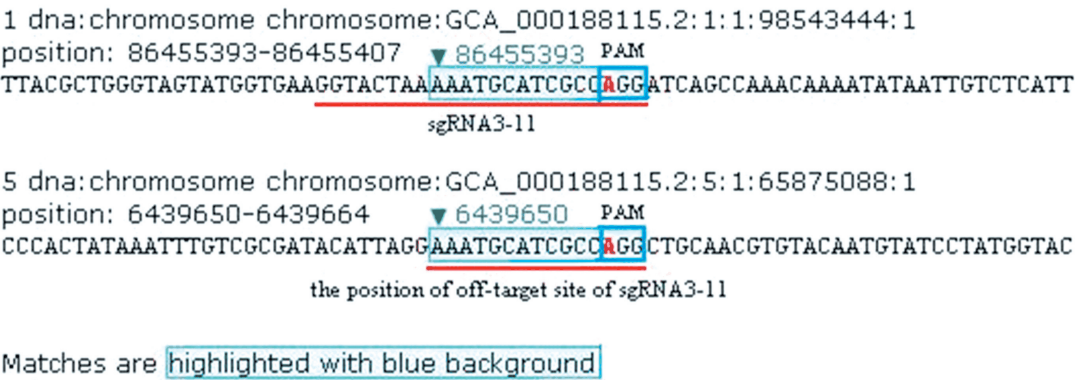


图 5 第 3 外显子 sgRNA3-11 的脱靶位点和染色体位置
Fig.5 The position of off-target site of sgRNA3-11 in exon 3

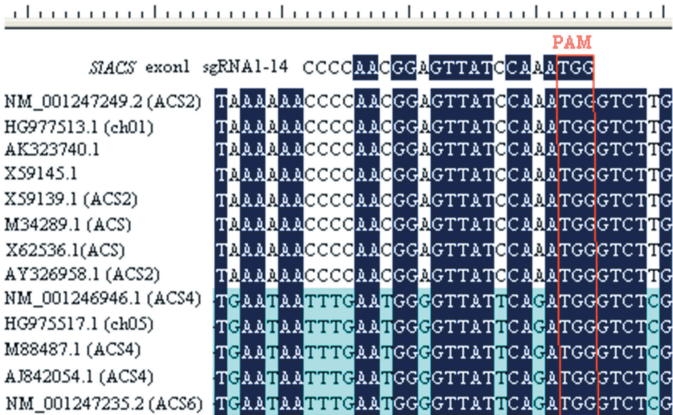


图 6 *SIACS2* 第 1 外显子 sgRNA1-14 的 SNP 位点

Fig.6 SNP loci on sgRNA1-14 of exon 1 in *SIACS2*

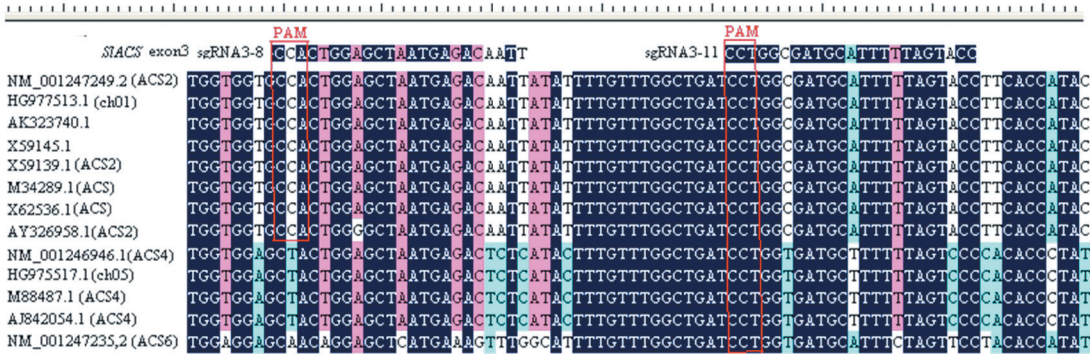


图 7 *SIACS2* 第 3 外显子 sgRNA3-8 和 sgRNA3-11 的 SNP

Fig.7 SNP loci on sgRNA3-8 and sgRNA3-11 of exon 3 in *SIACS2*

3 讨论

科学设计 sgRNA,提高基因组编辑效率和最大限度降低脱靶风险是 CRISPR/Cas9 系统面临的最大挑战。番茄是模式植物之一,已完成了全基因组测序,sgRNA 的特异性及其脱靶位点可通过全基因组层面扫描获得相关信息。sgRNA 的设计已有多个工具软件和在线分析平台,如 CRISPRdirect 综合考虑了 sgRNA 全长和 12 nt 及 8 nt 种子序列的特异性,结果输出中会标出每条 sgRNA 在外显子上的位置和 GC 含量,还将特异性高、脱靶位点少的 sgRNA 以及含有转录终止序列“TTTT”的 sgRNA 分别用亮绿色和灰色作出标示;华中农业大学作物遗传改良国家重点实验室开发的 CRISPR-P 则根据特定公式对每个 sgRNA 序列的靶向特异性进行综合评分和排序,并标示出 sgRNA 靶标位点中是否含限制性内切酶识别序列,以方便采用酶切法检测基因组切割效率。本研究综合运用 CRISPRdirect 和 CRISPR-P 两个在线分析平台,对 *SIACS2* 基因 5' 端的第 1、2、3 个外显

子上所分布的 sgRNAs 进行了综合评价,筛选出 3 条 sgRNAs 优选序列,即第 1 外显子 sgRNA1-14、第 3 外显子 sgRNA3-8 和 sgRNA3-11。其中 sgRNA1-14 和 sgRNA3-8 的全序列或 12 nt 的种子序列在番茄基因组上均为唯一序列,但它们的 5' 端第 1 个碱基为 A 或 C。双子叶植物 sgRNA 的转录一般选用 U6 启动子驱动,其转录活性相对较高,有明确的转录起始位点,即以“G”碱基开头。精确起始转录能消除无关 DNA 序列的转录,从而减少脱靶效应的产生^[20-21]。因此,如果选用 U6 启动子,需要在这 2 条 sgRNA 的 5' 端添加鸟嘌呤核苷酸。有研究表明,在 sgRNA 的 5' 端额外增加两个鸟嘌呤核苷酸后能够显著提高 CRISPR/Cas9 系统的特异性^[17],为非 G 起始的 sgRNA 应用提供了依据。

乙烯除了促进果实成熟,在植物种子萌发、叶片扩展、根毛伸长、侧根生长、开花、植株衰老等阶段均起重要作用。抑制番茄系统 I 或过度抑制系统 II 乙烯的生成,将影响番茄的正常生长或致使果实不能成熟。本研究从 *SIACS2* 中筛选出的 3 条 sgRNAs 与 GenBank 中公布的 *SIACS4* 和 *SIACS6* 的同源序列存

在 3~8 个 SNP, 其中 2~3 个 SNP 分布在 12 nt 的种子序列区。这些 SNP 的存在规避了 sgRNA 对 *SIACS4* 和 *SIACS6* 的脱靶效应, 可望达到预期目的: (1) 既抑制系统 II 乙烯合成限速酶基因 *SIACS2* 的表达, 又不影响 *SIACS4* 的正常表达, 使系统 II 乙烯保持在一定水平而不致过量, 以迟滞番茄过熟; (2) 不干扰系统 I 乙烯形成的 *SIACS6* 基因的正常表达, 保证番茄的正常生长。生物信息学可以提供许多有用信息, 但不能代替实验验证。在应用这 3 条优选的 sgRNA 改良番茄的耐贮运性状前, 可先用原生质体系统验证其靶向能力和编辑效率。

CRISPR-Cas9 系统通过 sgRNA 引导 Cas9 蛋白在 PAM 元件上游第 3~4 bp 处剪切, 造成 DNA 双链断裂, 在非同源末端连接修复过程中引入短的缺失或插入, 使基因失活, 但失活基因仍能复制和转录, 消耗植物营养和能量, 如能删除失活基因的完整序列或大片断, 将降低植物营养和能量的消耗。近年来, 已通过双 sgRNA 引导 Cas9 系统在植物^[22-23]和动物^[24]上删除了一段 DNA 序列, 最大缺失片段达 105 kb。本研究筛选的 sgRNA1-14 和 sgRNA3-8 特异性较好, 在染色体上相隔 409 nt, 利用双 sgRNA 引导 Cas9 将其删除, 造成 DNA 大片段缺失, 将更彻底的抑制 *SIACS2* 基因功能。

参考文献(References)

- [1] MCMURCHIE E, MCGLASSON W, EAKS I. Treatment of fruit with propylene gives information about the biogenesis of ethylene[J]. *Nature*, 1972, 237(5352): 235-236. DOI: 10.1038/237235a0.
- [2] BARRY C, LOP-TOUS M, GRIERSON D. The regulation of 1-aminocyclopropane-1-carboxylic acid synthase gene expression during the transition from system-1 to system-2 ethylene synthesis in tomato [J]. *Plant Physiology*, 2000, 123(3): 979-986. DOI: 10.1104/pp.123.3.979.
- [3] OELLER P, LU M, TAYLOR L, et al. Reversible inhibition of tomato fruit senescence by antisense RNA[J]. *Science*, 1991, 254(5030): 437-439. DOI: 10.1126/science.1925603.
- [4] 熊爱生, 姚泉洪, 李贤, 等. ACC 氧化酶和 ACC 合成酶反义 RNA 融合基因导入番茄和乙烯合成的抑制[J]. *实验生物学报*, 2003, 36(6): 35-41.
XIONG Aisheng, YAO Quanhong, LI Xian, et al. Double antisense ACC oxidase and ACC synthase fusion gene introduced into tomato by agrobacterium-mediated transformation and analysis the ethylene production of transgenic plants[J]. *Acta Biologiae Experimentalis Sinica*, 2003, 36(6): 35-41.
- [5] 陈银华, 李汉霞, 叶志彪. 不同结构的外源 ACO 基因导入番茄对乙烯生成速率的影响[J]. *园艺学报*, 2007, 34(3): 644-648.
CHEN Yinhan, LI Hanxia, YE Zhibiao. Effects of introduction of different T-DNA structures of ACO gene to tomato genome on ethylene production rate [J]. *Acta Horticulturae Sinica*, 2007, 34(3): 644-648.
- [6] GUPTA A, PAL R, RAJAM M. Delayed ripening and improved fruit processing quality in tomato by RNAi-mediated silencing of three homologs of 1-aminopropane-1-carboxylate synthase gene [J]. *Journal of Plant Physiology*, 2013, 170(11): 987-995. DOI: 10.1016/j.jplph.2013.02.003.
- [7] CONG L, RAN F, COX D, et al. Multiplex genome engineering using CRISPR/Cas systems [J]. *Science*, 2013, 339(6121): 819-823. DOI: 10.1126/science.1231143.
- [8] Doudna J A, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9 [J]. *Science*, 2014, 346(6213): 1258096-1258096. DOI: 10.1126/science.1258096.
- [9] ILIAKIS G, WANG H, PERRAULT A, et al. Mechanisms of DNA double strand break repair and chromosome aberration formation [J]. *Cytogenetic & Genome Research*, 2004, 104(1-4): 14-20. DOI: 10.1159/000077461.
- [10] FU Y, FODEN J, KHAYTER C, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells [J]. *Nature Biotechnology*, 2013, 31(9): 822-826. DOI: 10.1038/nbt.2623.
- [11] HSU P, SCOTT D, WEINSTEIN J, et al. DNA targeting specificity of RNA-guided Cas9 nucleases [J]. *Nature Biotechnology*, 2013, 31(9): 827-832. DOI: 10.1038/nbt.2647.
- [12] NATIO Y, HINO K, BONO H, et al. CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites [J]. *Bioinformatics*, 2015, 31(7): 1120-1123. DOI: 10.1093/bioinformatics/btu743.
- [13] ZHANG Y, GE X, YANG F, et al. Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells [J]. *Scientific Reports*, 2014(4): 5405. DOI: 10.1038/srep05405.
- [14] LEI Y, LU L, LIU H, et al. CRISPR-P: a web tool for synthetic single-guide RNA design of CRISPR-system in plants [J]. *Molecular Plant*, 2014, 7(9): 1494-1496. DOI: 10.1093/mp/ssu044.
- [15] SHAN Q, WANG Y, LI J, et al. Targeted genome modification of crop plants using a CRISPR-Cas system [J]. *Nature Biotechnology*, 2013, 31(8): 686-688. DOI: 10.1038/nbt.2650.
- [16] PATTANAYAK V, LIN S, GUILINGER J, et al. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity [J]. *Nature Biotechnology*, 2013, 31(9): 839-843. DOI: 10.1038/nbt.2673.

[17] CHO S W, KIM S, KIM Y, et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases [J]. *Genome Research*, 2014, 24(1): 132–141. DOI: 10.1101/gr.162339.113.

[18] RAN F, HSU P, LIU C, et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity[J]. *Cell*, 2013, 154(6): 1 380–1 389. DOI: 10.1016/j.cell.2013.08.021.

[19] DOENCH J, HARTENIAN E, GRAHAM D, et al. Rational design of highly active sgRNAs for CRISPRCas9-mediated gene inactivation [J]. *Nature Biotechnology*, 2014, 32(12):1262–1267.

[20] BELHAJ K, CHAPARRO-GARCIA A, KAMOUN S, et al. Editing plant genomes with CRISPR/Cas9 [J]. *Current Opinion in Biotechnology*, 2015 (33): 76 – 84. DOI: 10.1016/j.copbio.2014.11.007.

[21] LI X, JIANG D H, YONG K L, et al. Varied transcriptional efficiencies of multiple *Arabidopsis* U6 small nuclear RNA genes[J]. *Journal of Integrative Plant Biology*, 2007, 49(2): 222 – 229. DOI: 10.1111/j.1744 – 7909.2007.00393.x.

[22] LI J, NORVILLE J, ACH J, et al. Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9[J]. *Nature Biotechnology*, 2013, 31(8): 688–691. DOI:10.1038/nbt.2654.

[23] BROOKS C, NEKRASOV V, LIPPMAN Z, et al. Efficient gene editing in tomato in the first generation using the CRISPR/Cas9 system [J]. *Plant Physiology*, 2014(166): 1 292–1 297. DOI: 10.1104/pp.114.

[24] SONG Y, YUAN L, WANG Y, et al. Efficient dual sgRNA-directed large gene deletion in rabbit with CRISPR/Cas9 system[J]. *Cellular & Molecular Life Sciences Cmls*, 2016(73):2959. DOI: 10.1007/s00018–016–2143–z.