

张力鹏,滕彦娇,王宏鹏,等. 西藏大花红景天 EST-SSR 开发及通用性分析[J]. 江苏农业科学,2021,49(11):40-48.
doi:10.15889/j.issn.1002-1302.2021.11.007

西藏大花红景天 EST-SSR 开发及通用性分析

张力鹏,滕彦娇,王宏鹏,李天宇,宋文芹,陈成彬

(南开大学生命科学学院,天津 300071)

摘要:以西藏大花红景天为研究对象,用 MISA 软件筛选转录组测序获得的 48 790 条 unigene,开发 EST-SSR(表达序列标签-微卫星序列)并对其通用性进行分析,以期为大花红景天属物种的遗传多样性研究和分子标记辅助育种奠定基础。结果表明,共检测出 10 761 个 SSR 位点,分布于 8 973 条 unigene 中,分布频率为 22.06%,平均分布距离为 4.725 kb。优势重复基序以单核苷酸、二核苷酸、三核苷酸为主,其中单核苷酸重复单元以基序 A/T 为主,共 7 167 个,二核苷酸重复单元以 AG/CT 为主,共 1 008 个。随机合成 280 对引物,其中 248 对表现出有效扩增,54 对具有多态性。此外,SSR 引物在长鞭红景天、狭叶红景天、菊叶红景天、圣地红景天、柴胡红景天、高山红景天中具有较高的通用性,扩增效率分别为 75.0%、73.75%、72.5%、62.5%、70.0%、72.5%。

关键词:大花红景天;转录组;SSR;通用性分析

中图分类号:S567.23+9.01 **文献标志码:**A **文章编号:**1002-1302(2021)11-0040-08

红景天是双子叶蔷薇目(Rosales)景天科(Crassulaceae)红景天属(*Rhodiola* L.)植物,因其根或根茎的浸出液为红色而得名^[1]。全世界红景天属植物共有 90 多种,其中我国占有主要的种质资源有 70 多种,主要分布在海拔(3 500~5 000 m)较高的山区,包括青藏高原及其毗邻带、横断山脉、天山山脉等^[2-4]。其中以西藏占有的种类最多,为 32 种;四川次之,为 22 种;新疆为 14 种。红景天常被称作“高原人参”或“雪山仙草”应用于传统中医和藏医种,已经有数千年的历史。目前主要药用种为蔷薇红景天(*R. rosea*)、大花红景天(*R. crenulata*)、菊叶红景天(*R. sachalinensis*)、圣地红景天(*R. sacra*)、齿叶红景天(*R. serrata*)、高山红景天(*R. cretinii*)和长鞭红景天(*R. fastigiata*)等^[5-7]。现代药理学研究发现,红景天根部含有 40 多种活性物质如红景天苷、没食子酸、苷原酪醇及其衍生物等,具有抗氧化、抗缺氧、抗辐射、抗衰老、抗疲劳、提高机体免疫力等作用。其中以大花红景天的活性物质含量最高、药效最好、使用最为广泛。因此《中华人

民共和国药典》2015 版中唯一指定的人药种为大花红景天^[8]。

基于多聚酶链式反应(PCR)的分子标记技术具有多态性好、遗传性高、稳定性强、简单迅速的特点,已逐渐成为研究植物物种遗传多样性的有力工具^[9]。其中简单重复序列(simple sequence repeat,简称 SSR)也称微卫星,能够均匀、随机地分布于植物基因组 DNA 中,具有数量丰富、稳定性高和简单便捷等优点^[10-12]。2013 年,You 等筛选分离了柴胡红景天的 17 对多态性引物以扩增微卫星位点(CCG)₆、(AAG)₈、(AGG)₆、(CT)₁₃、(AGC)₆、(AC)₁₀和(ATC)₆,并将其成功应用于大花红景天、长鞭红景天和圣地红景天^[13]。雷淑芸等基于高通量测序技术对唐古特红景天全基因组的 SSR 位点进行了分析,为红景天属植物 SSR 标记的开发奠定了一定基础^[14]。然而,目前对红景天基因组 SSR 的研究依然较少,特别是关于主要人药种大花红景天的 SSR 位点尚无报道。

本研究基于大花红景天茎/叶混样转录组测序(RNA-Seq)所获得的数据,利用 MISA 软件对 SSR 标记进行搜索,分析其分布与组成特征,并对其扩增效率和通用性进行了初步评价,以期为大花红景天属植物的分类与鉴定、亲缘关系和遗传多样性研究、种质资源利用和分子标记辅助育种提供理论基础。

收稿日期:2020-08-29

作者简介:张力鹏(1989—),男,河北沧州人,博士,助理研究员,主要从事中药材种质资源品种改良研究。E-mail: nknanhai@163.com。

通信作者:陈成彬,博士,副教授,主要从事中药材种质资源与品种改良研究。E-mail: chenccb@nankai.edu.cn。

1 材料与方法

1.1 材料

本研究用于引物筛选和通用性分析的所有植物材料均为野生种,于 2015 年 7 月 15 日至 7 月 22 日采自中国西藏地区。4 个地理居群的大花红景天 (*Rhodiola crenulata*) 分别采集于拉萨市米拉山 (RcML) 海拔 4 868.4 m 处、山南地区浪卡子县 (RcLK) 海拔 5 237.2 m 处、林芝市色季拉山 (RcSJ) 海拔 4 688 m 处、那曲地区嘉黎县 (RcJL) 海拔 4 499 m 处。其他 6 种红景天属材料分别为长鞭红景天 (*Rhodiola fastigiata*, Rfa)、柴胡红景天 (*Rhodiola bupleuroides*, Rbu)、圣地红景天 (*Rhodiola sacra*, Rsa)、菊叶红景天 (*Rhodiola chrysanthemifolia*, Rch)、狭叶红景天 (*Rhodiola kirilowii*, Rki) 和高山红景天 (*Rhodiola cretinii*, Rcr)。将上述材料的叶片用无菌水洗净并擦干后,再用液氮速冻,于 -80 °C 保存备用。

1.2 方法

1.2.1 DNA 提取和 PCR 扩增 红景天叶片基因组的提取采用改良的十六烷基三甲基溴化铵 (CTAB) 法^[15]。用 1% 琼脂糖凝胶电泳和可见分光光度计 NanoDrop 1000 检测样品 DNA 的质量,保证所有 DNA 样品的浓度在 300 ~ 500 ng/μL 之间, $D_{260\text{ nm}}/D_{230\text{ nm}}$ 和 $D_{260\text{ nm}}/D_{280\text{ nm}}$ 均在 1.8 ~ 2.2 之间,且无 RNA 污染。将各 DNA 样品稀释到浓度为 50 ng/μL 后于 -20 °C 保存备用。

PCR 扩增反应体系为 25 μL,包括 2 × U Taq PCR Mix (Zoman Biotechnology, Beijing, China) 12.5 μL、上下游引物各 1 μL、模板 DNA 1 μL、ddH₂O 9.5 μL。PCR 扩增程序:94 °C 2 min;94 °C 30 s,56 °C 30 s,72 °C 30 s,33 个循环;72 °C 5 min,4 °C 保存。扩增产物用 1% 琼脂糖凝胶进行电泳检测。

1.2.2 转录组数据来源 以海拔 4 868.4、5 237.2 m 处的大花红景天 (RcML、RcLK) 为材料,用 CTAB 法分别提取叶片、茎的总 RNA,等量混匀后,分别标记为 Rc4800、Rc5200 后备用^[16]。转录组测序由上海欧易生物医学科技有限公司完成,使用 Illumina HiSeq™ 2500 测序仪完成。将测序得到的 reads 用 Trinity (version: trinityrnaseq_r20131110) 软件 paired-end 的拼接方法得到 Transcript 序列,并用 TGICL 软件延伸得到 Unigene^[17-18]。

1.2.3 转录组 SSR 检测和引物设计 用软件 MISA (<http://pgrc.ipk-gatersleben.de/misa/>) 对转录组拼接 unigene 的 SSR 位点进行搜索。搜索标准如下:重复单元长度为 1 ~ 6 bp 时,单核苷酸重复次数 ≥ 10 次,二核苷酸重复次数 ≥ 6 次,三、四、五、六核苷酸重复次数 ≥ 5 次。

用 Primer Premier 3 软件对具有 SSR 位点的 unigene 序列设计引物。设计原则如下:引物序列长度为 18 ~ 25 bp,扩增产物大小为 80 ~ 300 bp,GC 含量为 40% ~ 65%,退火温度为 55 ~ 65 °C 且上、下游引物的退火温度相差不大于 2 °C,避免出现错配、引物二聚体和发卡结构。从设计好的引物中随机选择 280 对送至生工生物工程(上海)股份有限公司合成。

2 结果与分析

2.1 大花红景天转录组测序、unigene 组装与注释

通过 Illumine 测序平台共得到 81 105 580 raw reads (原始读序),通过对 raw reads 进行质量监控,过滤低质量的 reads 后,Rc5200 获得 40 584 060 个 clean reads (有效读序),Rc4800 获得 40 521 520 个 clean reads。用 de novo 拼接技术将有 overlap (重叠区)的 clean reads 连接成 1 个更长的序列,使用 Trinity (version: trinityrnaseq_r20131110) 软件的 paired-end 方法不断延伸然后拼接成 transcript (转录本序列),通过 TGICL 软件聚类、去掉冗余序列,最终得到 1 套 unigene (单基因序列) (图 1)。本研究共得到 48 790 条 unigene,总长度为 50 851 045 bp,平均大小为 1 042.24 bp,其中最长的为 15 505 bp,最短的为 301 bp。

基于 BLAST 对 de novo 拼接的 unigene 进行功能注释和分类,通过 BLASTx 将 unigene 序列分别与 NR、SWISSPROT 和 KOG 库进行比对,取 $e < 1 \times 10^{-5}$ 的注释,选择与 unigene 相似性最高的蛋白,将其功能作为该 unigene 蛋白功能的注释信息。5 个数据库注释统计结果如下:NR 注释到 31 024 条 (63.59%), SWISSPROT 注释到 23 770 条 (48.72%), KOG 注释到 19 792 条 (40.57%), KEGG 注释到 7 773 条 (15.93%), GO 注释到 20 660 条 (42.34%)。

根据 KOG 功能分类的结果,其中大部分 unigene 注释到 R (即一般性功能预测),其数量为 7 228 个,3 665 条 unigene 注释到 O (即次生代谢产物的合成、转运和代谢),2 942 条 unigene 注释到 T

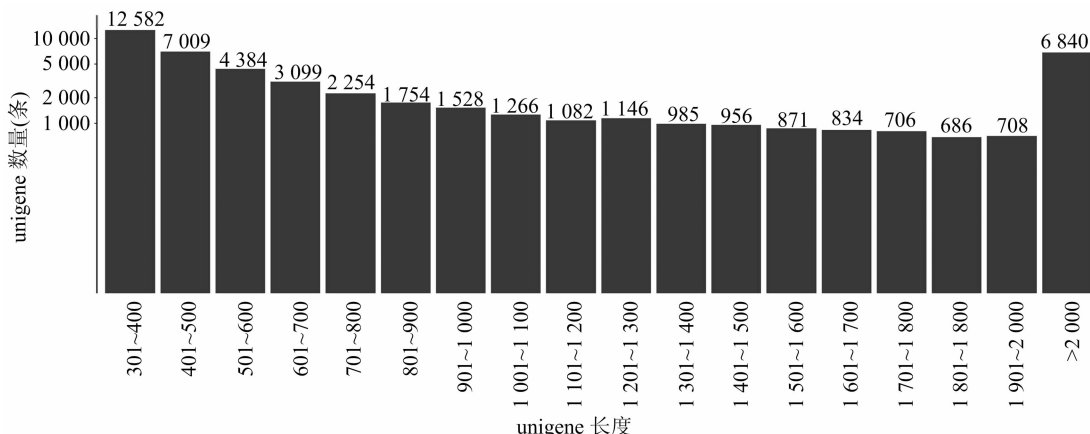
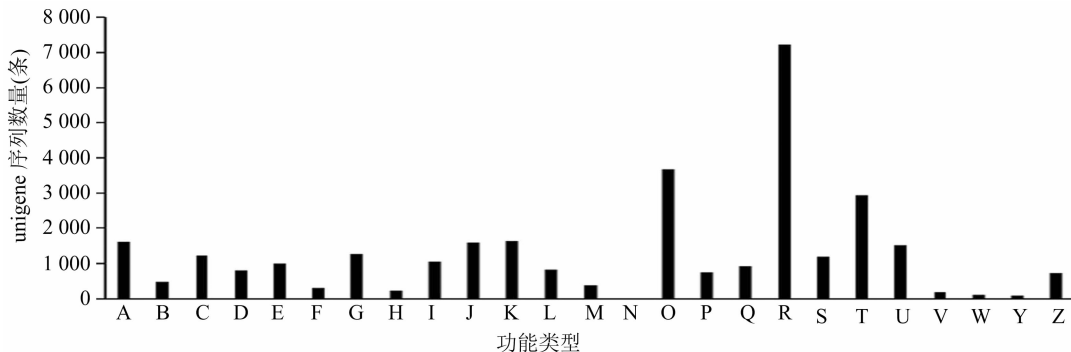


图1 unigene 长度的分布

(即信号转导机制),此外 K(转录)、A(RNA 的加工和修饰)、J(翻译、核糖体结构和生物合成)、U(胞质运输、分泌和囊泡运输)、G(碳水化合物转运与代谢)、C(能源的产生与转化)、I(脂质转运与代谢)、E(氨基酸转运与代谢)也均在 1 000 条 unigene 以上。GO 功能分类可以在基因的分子功能、细胞成分、参与的生物过程 3 个方面对其进行分类。对所

获得的 unigene 基因进行 GO 分类,结果显示在分子功能中, binding、catalytic activity(结合、催化活性)类最多;在细胞成分中, cell、cell part(细胞、细胞成分)最多,分别为 15 830、15 792 条;在生物过程中, cellular process、metabolic process(细胞过程、代谢过程)最多,分别为 14 116、12 057 条(图 2)。



A—RNA 的加工和修饰; B—染色体结构和动态变化; C—能源的产生与转化; D—细胞周期调控, 细胞分离, 染色体分配; E—氨基酸转运与代谢; F—核苷酸转运与代谢; G—碳水化合物转运与代谢; H—辅酶转运与代谢; I—脂质转运与代谢; J—翻译、核糖体结构和生物合成; K—转录; L—复制、重组和修复; M—细胞壁/细胞膜的生物发生; N—细胞运动; O—次生代谢产物的合成、转运和代谢; P—无机离子转运和代谢; Q—次生代谢产物的合成、转运和代谢; R—一般性功能预测; S—未知功能; T—信号转导机制; U—胞质运输、分泌和囊泡运输; V—防御机制; W—细胞外结构; Y—核结构; Z—细胞骨架

图2 unigene 序列的 KOG 注释

2.2 大花红景天转录组 SSR 位点的分布

用软件 MISA 对 48 790 条 unigene 的 SSR 位点进行搜索,结果显示,符合条件的 SSR 共有 10 761 个,分布频率(总 SSR 位点数/总 unigene 数)为 22.06%,平均分布距离(总 unigene 长度/总 SSR 位点数)为 4.725 kb。含有 SSR 位点的 unigene 为 8 973 条,发生频率(含有 SSR 位点的 unigene/总 unigene 数)为 18.39%。其中有 1 512 条 unigene 含有 2 个及以上 SSR 位点,449 条 unigene 含有复合 SSR 位点。SSR 位点的序列总长度为 16 590 bp。

由表 1 可以看出,大花红景天转录组 SSR 的类型较为丰富,单核苷酸、多核苷酸重复类型均有分布,但是不同类型的比例差异显著。主要为单核苷酸重复类型,为 7 207 个,占总 SSR 位点数的 66.97%,平均分布距离为 5.622 kb。其次为二核苷酸、三核苷酸重复类型,分别含有 1 468、1 960 个位点,分别占 SSR 位点数的 13.64%、18.21%,平均分布距离分别为 27.603、20.674 kb;四、五、六核苷酸重复类型较少,均不多于 80 个位点,分别为 80、29、17 个。

表 1 SSR 基序类型分布

SSR 类型	基序类型数(个)	基序类型	出现次数(次)	占总 SSR 重复基序的比例(%)
单核苷酸	2	A/T	7 167	66.60
		C/G	40	0.37
二核苷酸	4	AC/GT	193	1.79
		AG/CT	1 008	9.37
		AT/AT	266	2.47
		CG/CG	1	0.01
		其他	1	0.01
三核苷酸	10	AAC/GTT	225	2.09
		AAG/CTT	427	3.97
		ACC/GGT	194	1.80
		AGC/CTG	315	2.93
		AGG/CCT	271	2.52
		ATC/ATG	323	3.00
		其他	205	1.91
		其他	205	1.91
四核苷酸	23	AAAG/CTTT	12	0.11
		AAAT/ATTT	10	0.09
		AATC/ATTG	7	0.07
		ACTC/AGTG	7	0.07
		ATCC/ATGG	8	0.07
		其他	36	0.33
五核苷酸		AAACT/AGTTT	2	0.02
		AGAGG/CCTCT	2	0.02
		其他	25	0.23
六核苷酸		ACAGCC/CTGTGG	2	0.02
		其他	15	0.14

2.3 大花红景天转录组 SSR 基序重复类型和频率特征

对大花红景天转录组 SSR 基序类型进行分析显示,共有 82 种重复基序,单、二、三、四、五、六核苷酸重复各有 2、4、10、23、27、16 种;重复次数主要为 5~24 次,不同重复次数的 SSR 位点数目不同,主要为 5、10、11 次重复类型(图 3)。

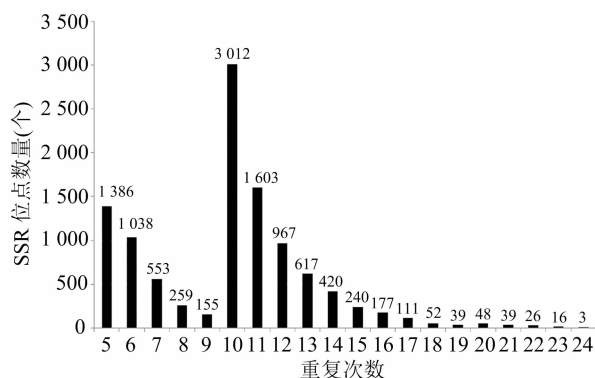


图 3 SSR 重复次数分布

由表 1 可以看出,大花红景天转录组不同 SSR 基序出现的频率相差较大,以单、二、三核苷酸重复基序为主要类型,占总 SSR 位点的 98.83%。其中以基序 A/T 最多,为 7 167 个,其次为 AG/CT(1 008 个)。单核苷酸重复基序主要为 A/T,重复次数集中在 10~14 次,而 G/C 较少,仅有 40 个,且重复次数主要为 10~13 次。二核苷酸重复基序主要为 AG/CT,其次为 AT/AT(266 个)和 AC/GT(193 个),CG/CG 仅含有 1 个,主要重复次数为 5~11 次。三核苷酸重复基序中 AAG/CTT 最多,为 427 个,其次为 ATC/ATG(323 个)、AGC/CTG(315 个)、AGG/CCT(271 个)、AAC/GTT(225 个)、ACC/GGT(194 个),其余 4 种均未超过 100 个,主要重复次数为 5~7 次。四、五、六核苷酸重复类型虽然最多,但是数量较少,除 AAAG/CTTT、AAAT/ATTT 类型分别为 12、10 个外,其他 66 种类型均低于 10 个。

2.4 大花红景天 SSR 引物的开发和通用性检测

基于含有 SSR 位点的 unigene 序列,利用 Primer Premier 3 软件设计引物,每条 unigene 产生 3 对引物。共得到 18 696 对符合条件的引物,分别位于 6 232 条 unigene 上。随机挑选 280 对引物,包括单核苷酸(125 对)、二核苷酸(32 对)、三核苷酸(48 对)、四核苷酸(51 对)、五核苷酸(14 对)、六核苷酸(9 对)。以 4 个地理居群的大花红景天 DNA 为模板对上述引物进行扩增,共 246 对引物能够扩增出与预期产物大小一致的特异性条带,有效扩增率为 87.86%。不同 SSR 重复类型引物的扩增成功率分别为 89.6%、90.63%、79.17%、94.12%、78.57%、88.89%。此外 54/246 对引物在不同产地大花红景天中的扩增产物具有多态性,占有效引物的 21.95%。共扩增得到 167 条电泳条带,其中多态性片段为 95 条,平均每对引物产生的多态性片段为 1.76 条(图 4)。

为了进一步验证大花红景天转录组 SSR 引物是否对其他种具有通用性,在 246 对引物中随机选择 80 对引物,以 6 种红景天(*Rfa*、*Rbu*、*Rki*、*Rcr*、*Rsa*、*Rch*)的 DNA 为模板进行扩增。结果显示,大花红景天转录组 SSR 引物对长鞭红景天、狭叶红景天、菊叶红景天、圣地红景天、柴胡红景天和高山红景天的 PCR 扩增成功率较高,分别为 75%、73.75%、72.5%、62.5%、70.0%、72.5%。用筛选到的 54 对大花红景天 SSR 多态性引物(表 2)对 20 个样品,包括 4 个产地(RcML、ReLK、ReJL、ReLZ)

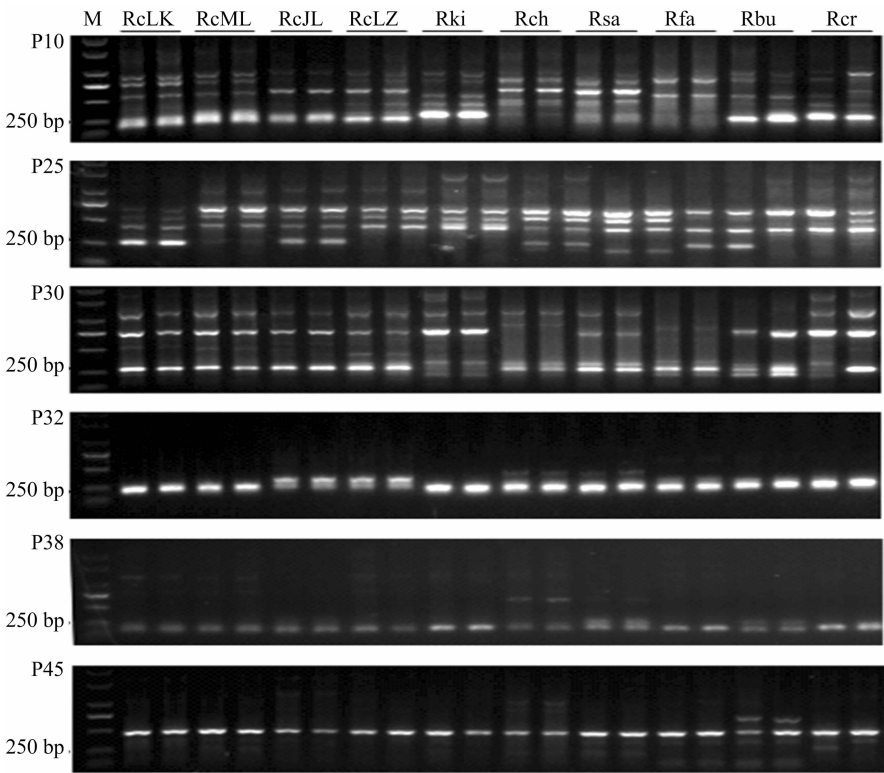


图4 部分 SSR 引物在 20 个红景天样品中的 PCR 扩增结果

的大花红景天和 6 种红景天进行 PCR 扩增,每个样品包含 2 个重复。结果显示,45/54 对 SSR 引物在 10 个样本中均能够扩增出清晰明亮的多态性产物条带。基于扩增产物多态性信息,利用 NtSYS 2.1 软件构建上述样品的遗传进化树,结果显示,红景

天属植物的种间遗传多态性显著高于种内。由此可见,不同地理居群的大花红景天明显聚为 1 支,狭叶红景天、菊叶红景天和圣地红景天明显聚为 1 支,柴胡红景天、长鞭红景天和高山聚为 1 支(图 5)。

表 2 54 对多态性 SSR 引物的序列信息

基因序列	SSR 类型	上游引物		下游引物		产物大小 (bp)
		序列(5'→3')	<i>T_m</i> 值(℃)	序列(5'→3')	<i>T_m</i> 值(℃)	
comp10589	(A) ₁₁	GTTCTCCCTCTCCATCTCC	60.0	TCATCCTCCTCCACGTCTTC	60.2	261
comp16785	(T) ₁₀ + (TG) ₆	TGCGATGAAAATGGAGATTG	59.6	AAGCCCTTAACAGAGCACGA	60.0	246
comp19385	(T) ₁₁	AGTGTTTTGCTGGCAGTTGA	59.5	ACTAAATTCCCAAGGCCCAT	60.0	156
comp20636	(A) ₁₀	AACAACAAGCCAGGTATCG	60.0	TGTTTCGCAACAACCTCGATG	60.8	252
comp21715	(TCA) ₅	TTGCTCATCCAACCTCATCA	60.2	GAGAAGAAGCGTCATCAGCC	60.1	154
comp21842	(TTAT) ₅	CAGCATTTTTGTGGCACATC	60.1	CTACAGAGTGGGACTGCAA	60.2	143
comp2272	(GT) ₆	TTGCTCCTTTTGGATCTTGG	60.2	AACCCTGTTTGGATGTTGGA	60.2	241
comp23189	(T) ₁₀	GTTTCGTGTCGTGAGATGGA	59.7	CATCACAATCAACCGACCAG	60.0	245
comp24119	(T) ₁₁	TGACCAGCGCATCTACTGTT	59.5	GCATCTGATACTCGACGCAA	60.0	253
comp24754	(T) ₁₃	CACTGTTGCCGTGTGAGTCT	60.0	CAAACCCAACCACATCTCCT	60.0	204
comp25385	(GAGT) ₅	ACAAGAGCATCATCAACCCC	60.0	CCAAATCAAATTAACTCCCCA	58.8	205
comp261	(CAT) ₅	CCCTACAGAACTTTCCACGG	59.6	TGAAAGAGATGAGGGACGCT	60.0	149
comp26828	(TCA) ₆	GGATATGGCTACAGTGGCGT	60.0	CGAAAACCTTTCGCAAAGGAG	60.0	248
comp27803	(T) ₁₄	GAAAAAGATGGAGCGACAGC	60.0	ACAGACATCCCCACCACAAT	60.1	246
comp29175	(A) ₁₀	AAGAAAGAGGAGCAGGAGGC	60.1	AGTTGCTGCGTCACTCTCAA	60.0	248
comp29420	(AGA) ₅	TGCACCACCACAAGAAGAAG	60.0	CCCTCTCCGTTCTCGTTTTT	60.6	178

表 2(续)

基因序列	SSR 类型	上游引物		下游引物		产物大小 (bp)
		序列(5'→3')	T _m 值(℃)	序列(5'→3')	T _m 值(℃)	
comp30949	(T) ₁₀	TTTTCGACCTTTCATTGGCT	60.0	CAATCACATTTTCATTGGCG	60.0	144
comp31499	(GGA) ₅	AGGGTTTTGGAACAATGGC	61.1	GGGGTGGAATATCCGATTCT	60.0	127
comp33265	(A) ₁₁	TTGCTACACATACACGGGA	60.0	TCAGGGCAACAAGTGTTTTCT	60.0	249
comp33612	(TA) ₇	TGACGAATGTGCATAACCAGA	60.1	TGAGAGAAAACAAGCAGGCA	60.0	169
comp34529	(CCT) ₅	TCTGCAACAATGGCGTTTTA	60.3	GTCGGATCATTACGTCCTT	59.9	149
comp34926	(T) ₁₆	CGCACACACTTACGAGAGGA	60.0	GCATTGAATAAAAAATGGGCA	58.5	179
comp36037	(TCG) ₅	CTGTCTCACACAATGGCTCG	60.5	TTCAGGCAACTTTGGAGCTT	60.0	279
comp36320	(G) ₁₂	TCAGGTAGGCAAGAAGGGAA	59.8	CCGCCTTTGTCTCCATTAAA	60.1	224
comp37508	(T) ₁₀	ATTGGACTGGACTTGTGGCC	60.0	CGGGATACCCATTACATTTC	60.0	270
comp38308	(TC) ₆	ATCTACTCCGCGTCACATCC	60.1	CAAAGCCAACTCCTCTACCG	59.9	241
comp39907	(GAA) ₆	AGGATGACAAGAATCGGTGC	60.1	CCCCTTTAGTCCGACCTTTT	59.5	195
comp40116	(TCA) ₅	ATCGCAGGAATGCCATACTC	60.1	TCCCTAACACCCCATGAAA	60.2	217
comp40475	(T) ₁₃	CTGACTCAAGGAGTGAGCCC	60.0	ATGGTGGCAATCTATCGAGC	60.1	269
comp40629	(A) ₁₂	TGACGTGATGGGATGACCTA	60.0	TTGGCTCTCACGCTCAAATA	59.6	277
comp40991	(CT) ₆	CTCCCAAAACCAAAGTCCAA	59.9	GTAGAGAGGCGGTGGAGATG	59.8	175
comp41352	(C) ₁₂	AAGGGCCTCGATTTCATCTTT	60.0	TACAGTGGGTGATGCGTGAT	60.0	218
comp41458	(CCAA) ₅	AGGTGGGCTTGACAACAGAC	60.2	AATTATGCGTCAATCCCGAG	59.9	187
comp41736	(A) ₁₁	TGGATATTGCAACACCTGA	60.0	CCATCGTGTCACTTTTCCT	60.0	276
comp42671	(TTG) ₅	CTTCTGGATAGGCGTCTTCG	60.0	CAGGAACTGTCCAATCCGTT	60.0	222
comp42837	(T) ₁₀	CATCACTGCTTCAAAGCCAA	60.0	TTGGCTATCACAATCCTCCC	59.9	276
comp43011	(CT) ₆	TCCGATTGCTTTGATGTTCA	60.0	CTCAATCTCCATCCCCTCAA	60.0	207
comp43248	(GT) ₁₀	TGCAACAACGTGTTGAAGAG	60.0	TTTCACAAGGGACGCTTTTT	59.7	206
comp43722	(TG) ₆	GTTTTGAAATTGGCGATGCT	60.1	AGATGGATGTCTGTTTCGG	59.9	240
comp43989	(T) ₁₄	ACGCGTTTGCTTACTTCGAT	60.0	ATCACCTTCTGGGGGACTCT	59.9	271
comp44763	(A) ₁₆	ATCTGAACCAACCGCAAAAC	60.0	CTGAATTCAAAATCCGCCAT	59.9	201
comp45833	(ATCCCA) ₆	CCGAAGACGAGTTAGCGAAG	60.1	ACCAGCGTGTGAGAGGACTT	59.9	206
comp45954	(ATCC) ₅	TCTGGGAAAGGGAGCTTAT	60.0	GCCACTAGACATCTCTCCGC	60.0	256
comp46022	(C) ₁₁	CCTTTTCGGGTCGTACAGTAA	60.1	TTGACTTTAGCCGCCATTTT	59.7	251
comp46437	(TCT) ₅	TTCTTGGATGTTACCTGGCT	59.6	GAGAGAAGAGATGTTGGCGG	60.0	208
comp47839	(GTG) ₆	TCACCCAGCTTCTGTGATA	59.5	AGAGAGCCCATCTGCCTACA	60.0	268
comp48084	(CTC) ₅	TGTCTCACTCCACCATTCCA	60.1	CACGCAATCGAAATCCTTTT	60.1	248
comp56889	(A) ₁₀	GGGTGCAATAGGGTAAGCAA	60.0	CCCCCAGAAAATGAAGTCAA	59.9	245
comp57101	(A) ₁₀	TCTATAACCAGCAGGGTCGG	60.1	CCGATGTGCTGAATGACTTG	60.1	202
comp57316	(G) ₁₁	AGCAAAAGGAGAAGAAGGGG	59.8	TTCGATAAGCAACTCGGGAC	60.2	279
comp58091	(GA) ₇	TACAGAATTTGGGTGTGGCA	60.0	CGAAATCCAGTGGTCAGGTT	60.0	192
comp60530	(A) ₁₄	GCCCTGGGTGACAAGATAAA	59.9	TTTATATGGAGGTGAGGCGG	59.9	265
comp6718	(T) ₁₀	AACACTGTCACTGGCCTTCC	60.2	TGGTGTGAGCACCAGACTTG	60.9	124
comp8191	(TC) ₈	GACCAGCGCTACCTGATCTC	60.0	CGATTCTCACCCACCTTCC	60.5	239

3 讨论与结论

近年来,随着红景天药用价值和保健价值的逐渐开发,人们对其需求量也与日俱增。然而由于缺少人工栽培技术,对红景天种质资源的掠夺和破坏

十分严重,从而引起多个物种资源数量急剧下降,极大破坏了冻原植被带的高寒草甸生态系统^[19]。此外,红景天属植物多分布于海拔较高的山区,因生长环境恶劣,常伴有低温、低氧、强紫外等特点,使得对其采集和研究均十分困难^[20-22]。本研究对

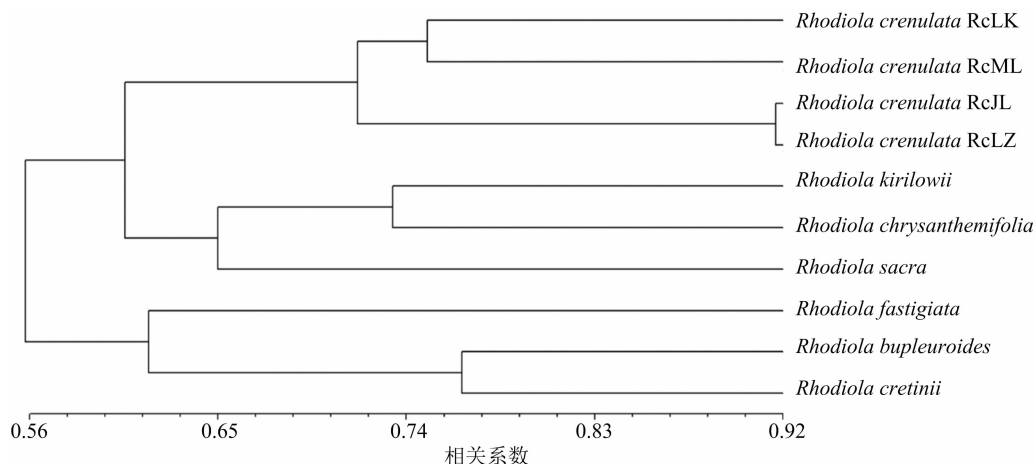


图5 基于 45 对多态性 SSR 引物对 10 种红景天的 20 份材料的聚类结果

2 种海拔地区的大花红景天进行转录组测序,基于测序结果分析 SSR 分子标记的类型与分布情况,并对 SSR 引物的有效性和通用性进行评估,为进一步开展红景天属植物遗传多样性分析、物种鉴定、遗传图谱构建、分子标记辅助育种和功能基因的发掘提供坚实的理论基础。

根据表达序列标签 (EST) 开发 SSR 分子标记具有简单、便捷等特点,特别对于尚无参考基因组序列的非模式植物而言,利用转录组测序技术可以获得大量基因序列,为开发该物种的 SSR 分子标记提供有力的基础^[23]。目前已对多个物种的 SSR 位点进行了比较详尽的研究,如黑绿豆 (*Vigna mungo*)^[24]、花生 (*Arachis hypogaea*)^[25]、绿豆 (*Vigna radiate*)^[26]、赤小豆 (*Vigna umbellata*)^[27]、豇豆 (*Vigna unguiculata*)^[28]、洋麻 (*Hibiscus cannabinus*)^[29]、开心果 (*Pistacia vera*)^[30]、白桦 (*Betula platyphylla*)^[31]、苜蓿 (*Medicago sativa*)^[32] 和刺梨 (*Rosa roxburghii*)^[33] 等。本研究通过对转录组组装的 unigene 序列 SSR 位点进行搜索,共得到 10 761 个 SSR,分布频率为 22.06%,平均分布距离为 4.725 kb,与已报道的万寿菊 (*Tagetes erecta*)^[34]、柑橘 (*Citrus sinensis osbeck*)^[35]、萝卜 (*Raphanus sativus*)^[36] 等相似。不同物种的 SSR 重复类型十分不同,如大花红景天 SSR 重复基序类型以单核苷酸、二核苷酸和三核苷酸重复类型最多,占总位点数的 98.83%,且 A/T 重复 10 次的最多,为 2 899 个,占总位点的 26.93%,这一结果与绿豆的 SSR 分布相似^[26]。ACG/AGT、CG/CG 重复基序比例最少,与刺梨的结果一致^[33]。

在随机合成的 280 对引物中,有 249 对引物可

以在 4 个地理居群的大花红景天中成功扩增,比例较高,为 87.86%,且筛选到 54 对多态性引物。说明大花红景天的种内遗传变异较小。此外,随机挑选的 80 对引物对 6 种红景天植物扩增成功率均在 60% 以上,说明利用转录组数据开发的大花红景天 SSR 具有较高的通用性,为其他红景天种材料提供了较为丰富的 SSR 信息来源。通过对多态性扩增产物构建进化树可以观察到红景天属植物的种内遗传变异明显小于种间,且圣地红景天与菊叶红景天明显聚为一支,长鞭红景天与柴胡红景天聚类关系较近,这与笔者所在实验室前期基于 DNA 条形码 (ITS) 所获得红景天属植物聚类结果^[37] 一致。说明 SSR 分子标记技术可以作为区分红景天属植物类群的重要手段。

参考文献:

- [1] 中国科学院中国植物志委员会,等. 中国植物志:景天科[M]. 北京:科学出版社,1984.
- [2] Liu Z L, Liu Y Y, Liu C S, et al. The chemotaxonomic classification of *Rhodiola* plants and its correlation with morphological characteristics and genetic taxonomy[J]. Chemistry Central Journal, 2013, 7(1): 118–126.
- [3] Zhang J Q, Meng S Y, Wen J, et al. DNA barcoding of *Rhodiola* (Crassulaceae): a case study on a group of recently diversified medicinal plants from the Qinghai–Tibetan Plateau[J]. PLoS One, 2015, 10(3): 1–15.
- [4] Zhang J Q, Meng S Y, Wen J, et al. Phylogenetic relationships and character evolution of *Rhodiola* (Crassulaceae) based on nuclear ribosomal ITS and plastid trnL–F and psbA–trnH sequences[J]. Systematic Botany, 2014, 39(2): 441–451.
- [5] Kelly G S. *Rhodiola rosea*: a possible plant adaptogen [J]. Alternative Medicine Review A Journal of Clinical Therapeutic, 2001, 6(3): 293–302.

- [6] Xin T Y, Li X J, Yao H, et al. Survey of commercial *Rhodiola* products revealed species diversity and potential safety issues[J]. *Sci Rep*, 2015, 5: 8337 – 8342.
- [7] Baran M G, Baranek K B, Pietrosiuk A. Biotechnological approaches to enhance salidroside, rosin and its derivatives production in selected *Rhodiola* spp. *in vitro* cultures[J]. *Phytochem Rev*, 2015, 14(4): 657 – 674.
- [8] 国家药典委员会. 中华人民共和国药典: 一部[M]. 北京: 中国医药科技出版社, 2015: 154.
- [9] Hou Y, Lou A. Population genetic diversity and structure of a naturally isolated plant species, *Rhodiola dumulosa* (Crassulaceae) [J]. *PLoS One*, 2011, 6(9): 1 – 10.
- [10] Chen J F, Li R H, Xia Y S, et al. Development of EST – SSR markers in flowering Chinese cabbage (*Brassica campestris* L. ssp. *chinensis* var. *utilis* Tsen et Lee) based on *de novo* transcriptomic assemblies[J]. *PLoS One*, 2017, 12(9): 1 – 14.
- [11] Luo H Y, Xu Z J, Li Z D, et al. Development of SSR markers and identification of major quantitative trait loci controlling shelling percentage in cultivated peanut (*Arachis hypogaea* L.) [J]. *Theoretical and Applied Genetics*, 2017, 130(8): 1635 – 1648.
- [12] Kempk K, Mora – irtiz M, Ssith L M, et al. Characterization of novel SSR markers in diverse sainfoin (*Onobrychis viciifolia*) germplasm [J]. *BMC Genet*, 2016, 17(1): 124 – 138.
- [13] You J, Liu W, Zhao Y, et al. Microsatellite markers in *Rhodiola* (Crassulaceae), a medicinal herb genus widely used in traditional Chinese medicine[J]. *Applications in Plant Sciences*, 2013, 1(3): 1 – 4.
- [14] 雷淑芸, 高庆波, 付鹏程, 等. 基于 Solexa 高通量测序的唐古特红景天(*Rhodiola algida*) 微卫星信息分析[J]. *植物研究*, 2014, 34(6): 829 – 834.
- [15] 赵 为, 邓科君, 杨足君, 等. 景天科植物基因组 DNA 的高效提取方法[J]. *安徽农业科学*, 2006, 34(22): 5804 – 5805.
- [16] 张力鹏, 张银兴, 宋文芹, 等. 红景天属植物叶片 RNA 高效提取的方法[J]. *南开大学学报(自然科学版)*, 2017, 50(6): 48 – 53.
- [17] Perte G, Huang X Q, Liang F, et al. TIGR gene indices clustering tools (TGICL): a software system for fast clustering of large EST datasets[J]. *Bioinformatics*, 2003, 19(5): 651 – 652.
- [18] Grabherr M G, Haas B J, Yassour M, et al. Trinity: reconstructing a full – length transcriptome without a genome from RNA – Seq data [J]. *Nature Biotechnology*, 2011, 29(7): 644 – 652.
- [19] 孟庆文, 崔卫东, 白光红, 等. 红景天种内遗传多样性分析 AFLP 方法建立[J]. *新疆农业科学*, 2008, 45(1): 88 – 92.
- [20] Fu Y, Li L, Hao S, et al. Draft genome sequence of the Tibetan medicinal herb *Rhodiola crenulata* [J]. *Gigascience*, 2017, 6(6): 1 – 5.
- [21] Ahemed S, Zhan C S, Yang Y Y, et al. The transcript profile of a traditional Chinese medicine, *Atractylodes lancea*, revealing its sesquiterpenoid biosynthesis of the major active components [J]. *PLoS One*, 2016, 11(3): 1 – 19.
- [22] Zhang L P, Wu M, Yu D S, et al. Identification of glutathione peroxidase (GPX) gene family in *Rhodiola crenulata* and gene expression analysis under stress conditions[J]. *International Journal of Molecular Sciences*, 2018, 19(11): 3329 – 3346.
- [23] 刘超, 张力鹏, 王春国, 等. 日本落叶松 EST – SSR 标记挖掘及特征分析[J]. *林业科学研究*, 2013, 26(f10): 60 – 68.
- [24] Souframanien J, Reddy K S, Denovo A. *De novo* assembly, characterization of immature seed transcriptome and development of genic – SSR markers in black gram *Vigna mungo* L. Hepper [J]. *PLoS One*, 2015, 10(6): 1 – 18.
- [25] Bosamia T C, Mishra G P, Thankappan R, et al. Novel and stress relevant EST derived SSR markers developed and validated in peanut [J]. *PLoS One*, 2015, 10(6): 1 – 19.
- [26] Chen H L, Wang L X, Wang S H, et al. Transcriptome sequencing of mung bean (*Vigna radiate* L.) genes and the identification of EST – SSR markers[J]. *PLoS One*, 2015, 10(4): 1 – 15.
- [27] Chen H L, Chen X, Tian J, et al. Development of gene – based SSR markers in rice bean (*Vigna umbellata* L.) based on transcriptome data[J]. *PLoS One*, 2016, 11(3): 1 – 13.
- [28] Chen H L, Wang L X, Liu X Y, et al. *De novo* transcriptomic analysis of cowpea (*Vigna unguiculata* L. Walp.) for genic SSR marker development[J]. *BMC Genet*, 2017, 18(1): 65 – 77.
- [29] Li H, Li D, Chen A, et al. Characterization of the kenaf (*Hibiscus cannabinus*) global transcriptome using Illumina paired – end sequencing and development of EST – SSR markers[J]. *PLoS One*, 2016, 11(3): 1 – 8.
- [30] Ziya M E, Kafkas S, Khodaeiaminjan M, et al. Genome survey of pistachio (*Pistacia vera* L.) by next generation sequencing: Development of novel SSR markers and genetic diversity in *Pistacia* species[J]. *BMC Genomics*, 2016, 17(1): 998 – 1012.
- [31] Hao W, Wang S J, Liu H J, et al. Development of SSR markers and genetic diversity in white birch (*Betula platyphylla*) [J]. *PLoS One*, 2015, 10(4): 1 – 14.
- [32] Wang Z, Yu G H, Shi B B, et al. Development and characterization of simple sequence repeat (SSR) markers based on RNA – sequencing of *Medicago sativa* and in silico mapping onto the *M. truncatula* genome[J]. *PLoS One*, 2014, 9(3): 1 – 7.
- [33] Yan X Q, Zang X, Lu M, et al. *De novo* sequencing analysis of the *Rosa roxburghii* fruit transcriptome reveals putative ascorbate biosynthetic genes and EST – SSR markers [J]. *Gene*, 2015, 561(1): 54 – 62.
- [34] 张华丽, 丛日晨, 王茂良, 等. 基于万寿菊转录组测序的 SSR 标记开发[J]. *园艺学报*, 2018, 45(1): 159 – 167.
- [35] Jiang D, Zhong G Y, Hong Q B. Analysis of microsatellites in *Citrus unigenes* [J]. *Acta Genetica Sinica*, 2006, 33(4): 345 – 353.
- [36] Wang S F, Wang X X, He Q W, et al. Transcriptome analysis of the roots at early and late seedling stages using Illumina paired – end sequencing and development of EST – SSR markers in radish [J]. *Plant Cell Reports*, 2012, 31(8): 1437 – 1447.
- [37] 张力鹏, 滕艳娇, 于得水, 等. 西藏地区红景天属植物 ITS, rbcL, trnS – G 序列多态性分析[J]. *南开大学学报(自然科学版)*, 2019, 52(5): 93 – 101.

张二豪, 张 杰. CRISPR/Cas9 基因编辑技术应用于绿僵菌[J]. 江苏农业科学, 2021, 49(11): 48–53.
doi:10.15889/j.issn.1002-1302.2021.11.008

CRISPR/Cas9 基因编辑技术应用于绿僵菌

张二豪¹, 张 杰²

(1. 西藏农牧学院食品科学学院, 西藏林芝 860000; 2. 周口师范学院生命科学与农学院, 河南周口 466001)

摘要:近年来的研究显示, CRISPR/Cas9 系统是强有力的基因编辑新技术。以蝗绿僵菌为试验对象, 以同源重组敲除系统为对照, 研究 CRISPR/Cas9 系统敲除蝗绿僵菌的基因 *isp4* 核苷酸序列。阐明了 CRISPR/Cas9 载体构建的方法, 比较了 CRISPR/Cas9 和 Recombinase 敲除技术的异同, 最后通过 PCR 和突变菌株的表型验证了 CRISPR/Cas9 系统能够应用于绿僵菌。结果表明, CRISPR/Cas9 在昆虫病原真菌绿僵菌中是有效的基因编辑技术。

关键词: CRISPR/Cas9; 重组酶; 基因编辑; 绿僵菌

中图分类号: S188 **文献标志码:** A **文章编号:** 1002-1302(2021)11-0048-06

分子遗传修饰和基因工程技术需要精确地改变基因组的核苷酸序列。随着基因修饰新技术的不断发展, 某些基因可被敲除或被降低表达水平^[1], 这些技术为基因功能的研究提供了便捷的途径。基因编辑就是对细胞基因组中目的基因的核苷酸序列甚至是单个核苷酸进行替换、切除, 增加或者是插入外源的 DNA 序列, 使之产生可遗传的改变, 进而研究其功能的手段^[2]。绿僵菌的研究已有 135 年的历史^[3], 它作为一个昆虫病原真菌的模式丝状真菌, 常常用来研究真核细胞的生物学过程、基因的相关表型和侵染昆虫致病的过程。

为了更好地研究病原真菌的基因特征, 多种基因操作技术已经在绿僵菌研究中应用^[4]。以位点为靶向的 DNA 内切酶是基因编辑技术中强有力的方法之一^[5]。这些内切酶能在基因组范围内与靶向序列直接结合, 进而使 DNA 双链断裂, 断裂的 DNA 被修复过程中产生 DNA 序列的修饰。开发这些工具酶的起始方向集中在同源内切酶、锌指内切酶和转录激活内切酶类^[6-7]。与传统突变体材料的获得方法相比, 基因编辑技术能定向改变基因的组成和结构, 具有高效、可控和定向操作的优点^[8-9]。

在目标基因 DNA 产生双链断裂^[10]的基础上进行基因编辑是传统基因编辑技术的共同之处, 但是各种基因编辑技术的原理及作用方式并不相同。

同源内切酶利用单一结构域识别和切割双链 DNA^[11] 的基因编辑技术具有一定的局限性。CRISPR/Cas9 是新一代的识别和绑定特异 DNA 序列为导向的基因编辑技术^[12]。Generoso 等首次报道, CRISPR/Cas 作为特异的短间隔序列, 成簇存在于大肠杆菌 (*Escherichia coli*) 基因组内的特殊短重复序列之中^[13]。随后的研究表明, CRISPR 位点在细菌和古细菌的基因组中分别占 40% 和 90%, 这些位点具有适应性的免疫功能^[14]。在 2012 年, CRISPR/Cas9 在链球菌 (*Streptococcus pyogenes*) 中, 通过锚定 crRNA 5' 端的 20 nt 核苷酸序列行使功能^[15]。经过优化的 CRISPR/Cas9 系统编码序列在哺乳动物细胞中可以高度激活^[16-17]。在实际操作时, 在 sgRNA 上 5' 端设计的 20 nt 核苷酸序列能够与目标基因完全互补, 即可达到基因编辑的作用^[18]。基因编辑技术 CRISPR/Cas9 在高等真核生物上已经成功应用^[19-20]。CRISPR/Cas9 介导的基因编辑过程实质上是非同源末端重组 (non-homologous end joining, NHEJ) 修复和同源末端重组修复 (homology-directed repair, HDR) 过程^[21]。与传统基因编辑技术相比, CRISPR/Cas9 系统最大的优点是易形成基因序列倍增, 这些倍增区域在 sgRNA 的锚定下很容易突变^[22]。如果 2 个 sgRNA 位于基因组侧的两翼, 它们中间区域的基因组会被删除或发生颠倒^[23-24]。通过双 sgRNA 可以进行基因组序列置换^[25]。

收稿日期: 2020-10-09

基金项目: 周口师范学院高层次人才科研启动经费研究项目 (编号: 2018B180061); 西藏自治区自然科学基金 (编号: XZ2018ZRG-19)。

作者简介: 张二豪 (1989—), 男, 河南平顶山人, 硕士, 讲师, 主要研究方向为分子生物学。E-mail: 1158496424@qq.com。

通信作者: 张 杰, 博士, 讲师, 主要研究方向为分子生物学与生物工程。E-mail: Zhangjiezk@qq.com。