

DOI:10.11937/bfyy.201512021

南药青天葵转录组 SSR 位点信息分析

黄琼林¹, 何 瑞², 詹若挺², 陈蔚文²

(1. 广东医学院, 广东 湛江 524023; 2. 广州中医药大学 中药资源科学与工程研究中心, 教育部岭南中药资源重点实验室, 广东 广州 510006)

摘 要:以青天葵全转录组测序获得的 142 220 条 unigene 为试材, 采用 MicroSATellite (MISA) 软件分析青天葵简单重复序列 (simple sequence repeat, SSR) 的分布频率和重复基元的类型等特征, 利用 Primer 5.0 对符合鉴定要求的 SSR 基元进行引物设计, 并通过 SSRFinder 校验 SSR 引物, 以期为青天葵遗传多样性分析、分子标记辅助育种、遗传图谱绘制等种质资源研究提供参考依据。结果表明:在青天葵转录组共检测到分布在 5 223 条 unigene 上的 5 684 个 SSR 位点, SSR 发生频率为 3.67%, 分布密度为 1/13 kb。二核苷酸重复为青天葵 SSR 主要基元类型, 其中 AG/CT 和 AT/AT 基序出现频率最高, 占 SSR 总数的 52.9%。设计筛选得到 2 486 条 SSR 引物。青天葵转录组 SSR 位点出现频率高、类型丰富、多态性潜能较高。

关键词:青天葵; SSR; 位点信息; 转录组

中图分类号:R 282.71 **文献标识码:**A **文章编号:**1001-0009(2015)12-0076-04

青天葵属兰科芋兰属多年生小草本植物毛唇芋兰 (*Nervilia fordii* (Hance) Schltr.) 的全草或叶片, 是我国岭南地区名贵中药, 具有润肺止咳、清热解毒、散瘀止痛的功效, 对肺部疾病, 尤其是小儿呼吸道疾病有着显著疗效^[1]。由于人类过度开采和自身萌发条件苛刻, 青天葵面临着野生资源枯竭和人工栽培繁殖系数低的困境, 导致药材资源严重短缺。青天葵也早已在 1993 年被列入《中国南部石灰岩稀有濒危植物名录》^[2]。在资源

紧缺的情况下, 青天葵在种植和销售过程中出现了众多的混伪品, 造成其种质混乱, 品质下降, 进一步加剧了青天葵资源破坏程度。因此, 急需加强青天葵的资源保护和利用。但是, 青天葵目前主要开展了人工栽培^[3]、组织培养^[4]、药理药化^[5-6]等方面的研究, 其种质资源遗传多样性和分子标记辅助育种的的研究尚少。

简单重复序列 (simple sequence repeat, SSR) 是指 1~6 个核苷酸串联重复单元, 又叫微卫星 DNA, 是一种广泛存在于原核生物和真核生物基因中的遗传标记, 具有共显性、信息丰富、多态性丰富等特点^[7], 现已广泛应用于植物分子辅助育种、遗传图谱构建、品种鉴定、遗传多样性分析、功能基因定位以及种子纯度鉴定等研究^[8]。目前, 党参^[9]、杜仲^[10]、西洋参^[11]、灯盏花^[12]等药用植物已利用第二代高通量测序 (next generation sequencing, NGS) 技术进行了转录组来源 SSR 位点挖

第一作者简介:黄琼林(1986-), 男, 博士, 讲师, 研究方向为分子生物学。E-mail: perfectql@163.com.

责任作者:何瑞(1972-), 女, 博士, 研究员, 博士生导师, 研究方向为中药资源可持续利用与开发。E-mail: ruihe@gzucm.edu.cn.

基金项目:教育部留学回国人员科研启动资助项目(教外司留[2009]1001号); 广东医学院博士学位人员科研启动资助项目(B2013017)。

收稿日期:2015-01-19

results showed that the optimum chromosome preparation was achieved with collecting root tips at 10:30 AM, root length within 5 mm, pretreated for 2 hours in 0.1% colchicine at room temperature, spreading chromosomes by double-hypotonic treatment, dissociation with 1 mol/L HCl for 50 minutes at room temperature or 20 minutes at 60°C, and stained with 20% modified carbol fuchsin dyeing. The permanent slides could be made by the neutral resin. The karyotype analysis indicated that red pitaya is diploid ($2n=2x=22$), and the ratio of the longest chromosome to the shortest one was 1.63, so that the karyotype formula of red pitaya was $2n=22m$, belonging to 1A karyotype. The karyotype asymmetrical coefficient was 54.29%.

Keywords: red pitaya; aerial root; chromosome; preparation technique; karyotype analysis

掘。在前期青天葵转录组 RNA-seq 高通量测序结果的基础上,利用 MicroSATellite(MISA)^[13] 软件检测青天葵 SSR 位点,分析其分布、组成特征,并初步评价其可用性,以为青天葵的遗传多样性分析、功能基因定位和遗传图谱构建等研究中进一步开发利用 SSR 标记提供序列数据。

1 材料与方法

1.1 试验材料

青天葵转录组数据来源:在前期研究^[14]中,课题组利用 Illumina HiSeq™ 2000 高通量测序平台对青天葵叶片和球茎进行全转录组测序,共组装获得 142 220 条 unigene。

1.2 试验方法

1.2.1 青天葵 SSR 的筛选 采用软件 MISA (<http://pgrc.ipk-gatersleben.de/misa/Misa.html>)对青天葵转录组 unigene 进行序列分析,搜索 SSR 位点。所检测 SSR 位点包括二核苷酸重复、三核苷酸重复、四核苷酸重复、五核苷酸重复和六核苷酸重复 5 类。判断标准为二核苷酸重复不少于 6 次;三核苷酸重复不少于 5 次;四核苷酸、五核苷酸重复和六核苷酸重复不少于 4 次。单核苷酸重复容易发生错配而测序失败,所以没有列入分析。

1.2.2 青天葵 SSR 引物设计 使用 Primer 5 软件对含有 SSR 位点且 SSR 上下游序列均不小于 150 bp 的 unigene 进行批量设计引物,每条 SSR 序列产生 3 条引物。主要的引物参数设置为:退火温度(T_m)在 57~63℃,上、下游引物 T_m 值相差小于 5℃;扩增产物大小为 100~280 bp,引物长度为 18~27 bp;GC 含量为 40%~65%。将设计出的引物通过以下方式筛选:引物不含 SSR 序列;与 unigene 序列相比,引物的 5'端不得有超过 3 个碱基的错配,3'端只允许有 1 个碱基的错配;去除同时比对到多个 unigene 上的引物;使用 SSRFinder 检验出与 MISA 搜索结果相同的 SSR,并据此筛选出相同的 SSR 产物。

2 结果与分析

2.1 青天葵 SSR 数量与分布

通过 MISA 工具对青天葵转录组的 142 220 条 unigene 序列进行 SSR 检索,表 1 表明,在 5 223 条 unigene 中鉴定出 5 684 个符合搜索标准的 SSR 位点,SSR 发生频率(包含 SSR 的 unigene 数与总 unigene 数之比)为 3.67%,出现频率(SSR 数目与总 unigene 数目之比)为 4.00%。408 条 unigene 含有一个以上 SSR 位点,235 条 unigene 含有复合 SSRs。平均每 10 kb 有 0.77 个 SSR。

表 1 青天葵转录组中 SSR 搜索结果

项目	数量
搜索序列总数	142 220
搜索序列总长度/bp	73 629 571
SSR 总数	5 684
SSR 发生频率/%	3.67
SSR 出现频率/%	4.00
10 kb 含 SSR 数目	0.77
含 SSR 位点的序列数目	5 223
含超过 1 个 SSR 位点的序列数目	408
含复合型 SSR 的序列数目	235

由表 2 可知,青天葵转录组 SSR 种类比较丰富,从重复次数来看,重复序列以重复 6 次出现的频率最高,有 1 548 个,占 SSR 总量的 27.23%。其次为 5、7、4、8 次,分别占 SSR 总量的 19.25%、14.23%、10.80% 和 9.02%。

由图 1 可知,青天葵 SSR 以二核苷酸 AG/CT 和 AT/AT 重复基序为主要类型,占总 SSR 的 51.9%。其次是三、四核苷酸重复基序,出现频率分别是 26.2% 和 7.2%。其中,三核苷酸重复序列 AAT/ATT 最多, AAG/CTT 和 ATC/ATG 次之,ACC/GGT 和 AGC/CTG 最少;四核苷酸重复序列以 AAAT/ATTT 最多, ACAT/ATGT 最少。另外,青天葵转录组中还含有约 0.7% 的五核苷酸重复序列 AAAAT/ATTTT。

表 2 青天葵转录组中不同基序长度和重复的 SSR 数量分布

基序长度/bp	4	5	6	7	8	9	10	11	12	13	14	≥15	总计
2	0	0	1 188	664	495	411	405	230	10	0	0	0	3 403
3	0	966	343	179	16	3	0	0	0	0	0	2	1 509
4	448	117	15	0	0	0	1	0	0	0	0	0	581
5	82	7	2	0	0	0	0	1	0	0	0	0	92
6	84	10	0	2	2	0	1	0	0	0	0	0	99
总计	614	1 100	1 548	845	513	414	407	231	10	0	0	2	5 684

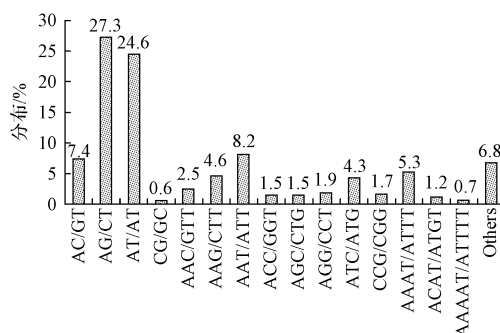


图 1 青天葵转录组中 SSR 基序类型分布

2.2 青天葵 SSR 的可用性评价

SSR 分子标记的多态性是评判其可用性的重要标准和依据,而 SSR 长度是影响其多态性高低的关键因

素^[15]。SSR 长度大于 20 bp 时多态性较好,在 12~20 bp 之间时多态性次之,而低于 12 bp 时其多态性极低,因此该研究去除了 12 bp 以下的 SSR。青天葵 SSR 长度在 12~20 bp 的有 4 679 条,占 SSR 总数的 82.32%,而长度在 20 bp 以上的达到 681 条,占 SSR 总数的 11.98%,这些 SSR 具有较高的多态性。另外,低级基序 SSR 的多态性被认为普遍优于高级基序 SSR^[15]。在长度大于 20 bp 的 SSR 中,含有低级基序(二、三核苷酸重复)SSR 共 396 个,推测这部分多态性潜能高的 SSR 在青天葵上

具有较高的利用价值。

2.3 青天葵 SSR 引物设计

为了筛选出可在试验中应用的青天葵 SSR,该课题组使用 Primer 5 软件对含有 SSR 位点且 SSR 上下游序列均不小于 150 bp 的 unigene 进行引物设计。结果表明,共有 829 条青天葵 SSR 序列成功设计出 2 486 对引物(部分引物见表 3),占到青天葵 SSR 总数的 14.6%。其中 20 bp 以上 SSR 序列且包含低级基序(二、三核苷酸重复)共设计出 155 对引物。

表 3 青天葵 SSR 引物序列

SSR 类型	SSR 序列	上游引物	下游引物	扩增产物长度/bp
二核苷酸重复	(AC) ₁₀	GATCACATTGAGCTGCAGGA	GTAGTACCCAAAGCCCATGT	274
	(AG) ₁₀	CGAACCCAAACATCTCGAACT	ATCAGTCCATTCGGGTGGTA	262
	(AT) ₁₀	CTGAAGATGCTGGGACATA	TAATGCAAAACGCTCAAGACG	199
	(CA) ₁₀	CGGCCAATTTCTCATAAAG	CAACTTCTTTTCTTATGTTTGTAGG	225
	(CT) ₁₁	CGATCAAAGCTCTCTCTCCC	GGAGTTGAGTGCGGTGAAAT	126
	(GA) ₁₀	TGTGATTAAAGCATTTGTAACATGG	TGACAAAAGCTGTGCTGTGC	251
	(TA) ₁₀	CAAAATGCTGCTGCTTCAGA	CCATTGTTATTCAGGTAGGTGATG	258
	(TC) ₁₀	GCCTGTGGAAGGAGGAGAG	AAAAGCGATCAAAACATTGC	266
	(AAT) ₇	CCACAAAAATCCATGCCTTC	AAACAATGAGGTGCGGAGAT	227
	(ATT) ₇	GGGCGATCACAATCCAGTTT	CAATTGATACAGCTCAGCG	176
三核苷酸重复	(CTT) ₁₈	TTTGCTTTTCTTTTCATATCA	GGGTGAAGATAAAGAGCTGGAA	226
	(ATC) ₈	GAGAACCCCTCATTTGAGCCA	GAGGAGGACATCGATGAGGA	229
	(AAC) ₇	TGGTAGGCATCCCAAGAAAG	GATCCTCGGCTGTGTGTAT	129
	(AGG) ₇	TGCTGCAAGACATGATCACA	CACTGCCATCCTCATCACAG	106
	(CCT) ₇	ATGATGAAAAAGCCAGCAOC	AGATGCAGCAAGAAGGATCG	182
	(ACC) ₈	CTCGCTCTCAAACCAACCTC	GTCGTGGTGGTCACATCTTG	273
	(CCT) ₈	GATGATGGGTGCAATGACTG	GTATTTGGTGGCCCTTGCTA	153
	(AAAT) ₅	GAGGTTGGAAATGCGGATAA	GCGTGTGTTTCTTTTGGGT	272
	(ATTT) ₅	GAACCTCTCTGCGCTCATA	TGGGCATTCAAACATGAAC	172
	(ACAT) ₅	CAAAAATTTTCAATCGACCAAA	TTCAGTGCCCCAAATTAGC	240
四核苷酸重复	(ATGT) ₈	TGAGGGTGTATCATAAAGGC	TCCCCAGTTACCCAGTCATC	248
	(AAAAT) ₄	CGGAAGATCCAGATCATCTAAA	ATGATGGCACATGAAGGTCA	205
	(TTTTA) ₄	GGGTACAGTAACCAACCCCT	GTGTGGAGGTTGGCATATT	270

3 讨论

利用分子标记对青天葵遗传学背景进行研究,可为青天葵资源保护提供科学依据。目前,青天葵基因组和遗传信息研究仍属基本空白,遗传多样性研究和分子标记研究也只有杜勤等^[16]利用 RAPD 分析青天葵遗传多样性;黄琼林等^[17-18]使用 DNA 条形码技术成功鉴别青天葵及其混伪品、同属植物。因此,青天葵仍缺乏能够大量应用于青天葵遗传图谱构建、功能基因定位等研究的简便、高效、稳定且具有种属特异性的分子标记体系。

该研究大规模地从青天葵转录组中检测出 5 684 个 SSR 位点,分布于 5 223 条 unigene。青天葵 SSR 出现频率为 4.00%,与单子叶植物如水稻(4.7%)、大麦(3.4%)、小麦(3.2%)和高粱(3.6%)^[19]等基本一致,低于双子叶植物如党参(16.10%)^[9]、野三七(16.86%)^[20]等,说明 SSR 含量具有物种特异性,亲缘关系相近的物种的 SSR 含量比较相似。

二核苷酸和三核苷酸重复是大多数植物 SSR 的主要类型,但主要重复基元因物种而异,这种占优势的重复基元可能与其编码相应蛋白质的使用频率较高有关^[21]。青天葵转录组 SSR 重复基元以二核苷酸为主,并以 AG/CT 和 AT/AT 的出现频率最高,共占总 SSR 的 51.9%,其次是三核苷酸重复基元。GC 重复基元在很多植物中出现频率均最低^[22],在青天葵转录组中也仅找到 9 个 GC 重复。此外,该研究还对所获得的青天葵 SSR 利用软件设计引物,方法快速、通量大、成本低,并通过筛选共获得 2 486 对 SSR 引物,为青天葵的 SSR 分析奠定数据基础。

总的来说,青天葵 SSR 出现频率较高,而且类型丰富。从多态性潜能和引物设计方面来考虑,鉴定获得的 SSR 也具有较高的可用性。因此,该研究结果为进一步开发新的青天葵功能基因 SSR 标记奠定了基础,将有力地推动青天葵功能基因资源的开发利用,丰富其分子标

记类型、遗传资源评价、绘制遗传图谱,实现特定性状的辅助选择和比较基因组学研究的开展。

参考文献

- [1] 陈蔚文. 岭南本草(二)[M]. 广州:广东科技出版社,2010:318-334.
- [2] 文和群,许兆然, Villa-Lobos J,等. 中国南部石灰岩稀有濒危植物名录[J]. 广西植物,1993,13(2):110-127.
- [3] 杜勤. 青天葵药材规范化种植(GAP)研究[D]. 广州:广州中医药大学,2005.
- [4] 李林轩,唐美琼,韦坤华,等. 青天葵组织培养条件优化[J]. 江苏农业科学,2014,42(7):66-68.
- [5] 邱少玲,焦杨,谢集照,等. 青天葵提取物的抗炎作用研究[J]. 时珍国医国药,2014,24(5):1033-1034.
- [6] 韦柳斌,陈金嫚,周光雄,等. 青天葵极性化学成分研究[J]. 天然产物研究与开发,2014(26):43-46,59.
- [7] Liu T, Zhu S, Fu L, et al. Development and characterization of 1 827 expressed Sequence tag-derived simple sequence repeat markers for ramie (*Boehmeria nivea* L. Gaud)[J]. PLoS One,2013,8(4):e60346.
- [8] 闫秋良. 基于生物信息学方法从牛和绵羊表达序列标签中筛选 SSR 标记的初步研究[D]. 西安:西北农林科技大学,2007.
- [9] 王东,曹玲亚,高建平. 党参转录组中 SSR 位点信息分析[J]. 中草药,2014,45(16):2390-2394.
- [10] 黄海燕,杜红岩,乌云塔娜,等. 基于杜仲转录组序列的 SSR 分子标记的开发[J]. 林业科学,2013,49(5):176-181.
- [11] 杨维泽,金航,赵振玲,等. 西洋参 EST 资源的 SSR 信息分析[J]. 西南农业学报,2011,24(1):275-278.
- [12] 陈茵,李翠婷,姜倪浩,等. 灯盏花转录组中 SSR 位点信息分析及其多态性研究[J]. 中国中药杂志,2014,39(7):1220-1224.
- [13] Lu X, Wang H, Liu B, et al. Three EST-SSR markers associated with QTL for the growth of the clam *Meretrix meretrix* revealed by selective genotyping[J]. Mar Biotechnol,2013,15(1):16-25.
- [14] 黄琼林. 青天葵转录组特征研究[D]. 广州:广州中医药大学,2013.
- [15] Temnykh S, De Clerck G, Lukashova A, et al. Computational and experimental analysis of microsatellites in Rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential[J]. Genome Res,2001,11:1441-1452.
- [16] 杜勤,魏志强,田军. 基于 RAPD 的青天葵遗传多样性及鉴别研究[J]. 中药新药与临床药理,2009,26(6):554-557.
- [17] Huang Q L, Liang L L, He R, et al. Applying DNA barcodes to identify *Nervilia fordii* and six congeneric species[J]. Plant Omics,2013,6(5):325-332.
- [18] 黄琼林,梁凌玲,何瑞,等. 应用 matK 基因鉴别青天葵及其常见混伪品[J]. 广西植物,2014,34(3):299-303.
- [19] Kantety R V, Rota M L, Matthews D E, et al. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat [J]. Plant Mol Biol,2002,48:501-510.
- [20] 李翠婷,张广辉,马春花,等. 野三七转录组中 SSR 位点信息分析及其多态性研究[J]. 中草药,2014,45(10):1468-1472.
- [21] 范三红,郭蔼光,单丽伟,等. 拟南芥基因密码子偏爱性分析[J]. 生物化学与生物物理进展,2003(30):221-225.
- [22] 李珊,周天华,赵桂仿,等. 马蹄香表达序列标签资源的 SSR 信息分析[J]. 中草药,2010,41(3):464-468.

SSR Loci Information Analysis in *Nervilia fordii* Transcriptome

HUANG Qiong-lin¹, HE Rui², ZHAN Ruo-ting², CHEN Wei-wen²

(1. Guangdong Medical University, Zhanjiang, Guangdong 524023; 2. Research Center of Chinese Medicinal Resource Science and Engineering, Guangzhou University of Chinese Medicine; Key Laboratory of Chinese Medicinal Resource from Lingnan, Ministry of Education, Guangzhou, Guangdong 510006)

Abstract: Taking 142 220 unigenes in *Nervilia fordii* transcriptome as materials to identify functional simple sequence repeat (SSR), which would promote the researches on resource protection including genetic diversity, molecular breeding and genetic mapping of *N. fordii*. The distribution frequency of high-flux transcriptome SSR and the basic characteristics of repeat motifs were analyzed using MicroSatellite software, and then SSR primers were designed by Primer 5.0 software and validated by SSR Finder software. The results showed that 5 684 SSR loci were acquired from 5 223 unigenes with an occurrence frequency of 3.67% and a density of 1/13 kb. Dinucleotide repeat was the main SSR type of *N. fordii*, and AG/CT and AT/AT exhibited the highest occurrence frequency (52.9%) among all SSR motifs. A total of 2 486 primer pairs were designed for marker development.

Keywords: *Nervilia fordii*; SSR; loci information; transcriptome