

冬凌草转录组 SSR 位点分析及多态性初步评价

苏秀红, 李庆磊, 陈随清, 董诚明, 杨晶凡, 尹 雷

(河南中医药大学 药学院, 河南 郑州 450046)

摘 要:以冬凌草转录组为研究对象,使用 Microsatellite(MISA)软件进行 SSR 搜索,分析冬凌草中 SSR 位点信息,并利用 Primer 3 设计 SSR 引物,对其多态性进行初步评价。结果表明:研究共发现 11 114 个 SSR,分布于 8 873 条独立基因(Unigenes)中,SSR 位点出现频率为 24.90%,共有 55 种重复单元,平均每 3 517 bp 含 1 个 SSR 位点。SSR 位点所包含的重复类型中,二核苷酸重复是主要类型,占总 SSR 的 47.71%;其次是单碱基重复类型(35.46%)。SSR 所包含的重复单元中,单核苷酸重复单元 A/T 和二核苷酸重复单元 AG/CT 是优势重复单元,分别占总 SSR 的 34.95%、30.16%。多态性进行评价结果表明冬凌草转录组 SSR 类型丰富,多态性潜能高。

关键词:冬凌草;转录组;SSR

中图分类号:S 567.23⁺9 **文献标识码:**A **文章编号:**1001-0009(2016)04-0093-04

冬凌草(*Rabdosia rubescens* (Hemsl.) Hara)属唇形科香茶菜属多年生植物。以其地上草质部分入药,主要含二萜类化合物,特别是叶片中含量较高。具有清热解毒、消炎止痛、健胃活血及抗肿瘤之功效^[1]。该物种主要分布于我国河南、山西、湖南、湖北等地,作为药材主产于河南。随着冬凌草经济、药用价值不断被发现与认识,以冬凌草为主要原料生产的中成药种类越来越多,对冬凌草的研究报道也逐渐增多,多集中于化学成分的分离鉴定^[1-3]、药理作用^[4-5]、采收加工^[6]、生物技术^[7]等方面。有关冬凌草遗传特性方面的研究报道甚少,目前仅见于有关其 ISSR 方面的报道^[8],且其遗传背景并不清晰,给冬凌草种质资源保存、新种质挖掘等带来了很大困难,因此,亟需开发新的标记方法来研究冬凌草种质资源特性。

简单重复序列(simple sequence repeat, SSR),又称微卫星(microsatellite),通常是由 1~6 个核苷酸组成的串联重复 DNA 序列,其长度在 100~200 bp 左右。高通

量测序技术为研究 SSR 信息位点在基因组中的分布提供了丰富的数据资源。目前药用植物杜仲、丹参、人参等借助测序数据开发了转录组来源的 SSR,且已经证明基于转录组 SSR 开发的分子标记多态性和扩增效果较好,适宜用来进行分子标记的开发^[9]。因此,现通过高通量转录组测序(RNA-seq)技术获得冬凌草的转录组数据,利用 Microsatellite(MISA)软件搜索冬凌草 SSR 位点信息,系统分析了其分布频率、组成特征,并进行初步可用性评价,以期对冬凌草遗传图谱构建、遗传多样性分析提供借鉴和参考。

1 材料与方法

1.1 试验材料

冬凌草(*Rabdosia rubescens* (Hemsl.) Hara),经河南中医学院董诚明教授鉴定为唇形科香茶菜属碎米桠变种。取冬凌草新鲜叶片迅速用液氮冷却后,投入干冰中送北京诺禾致源公司,提取 mRNA 后利用 Illumina HiSeqTM2000/MiseqTM进行转录组测序。共获得 44 626 (39 092 024 bp)条 Unigenes。

1.2 试验方法

1.2.1 SSR 检测 采用 MISA 软件对 Unigenes 进行 SSR 检测。

1.2.2 SSR 引物设计 根据找到的 SSR 信息,采用 Primer 3(2.3.5 版,默认参数)进行 SSR 引物设计。引物设计标准如下,扩增产物长度 100~280 bp,引物序列长度 18~27 bp(最适宜长度为 20 bp),退火温度为 57~62℃(最佳退火温度为 60℃),GC 含量为 40%~65%(最适宜为 50%),尽量避免二级结构(如发卡结构)、错配和

第一作者简介:苏秀红(1975-),女,博士,副教授,研究方向为中药资源。E-mail:suxihong80@163.com

责任作者:陈随清(1965-),男,博士,教授,研究方向为中药鉴定。E-mail:suiqingchen@163.com

基金项目:国家自然科学基金资助项目(81173486);河南中医学院创新人才资助项目(2011XCXRC02);河南省高等学校青年骨干教师计划资助项目(2011GGJS-089);河南省教育厅科学技术研究重点资助项目(13A360613);河南中医学院博士科研基金资助项目(BSJ2009-10)。

收稿日期:2015-10-08

二聚体等的出现。

1.2.3 冬凌草 DNA 提取与 PCR 扩增 采用宝生物试剂盒提取不同产地冬凌草 DNA,PCR 反应体系如下,总体积为 50 μ L,DNA 模板 3 μ L,Ex Taq 2 μ L,Master Mix 25 μ L,上下游引物各 1 μ L,RNase-Free H₂O 18 μ L。PCR 反应条件为 94℃ 预变性 4 min;再进 35 个循环,每个循环包括 94℃ 变性 15 s,55℃ 退火 30 s,72℃ 延伸 45 s。最后 72℃ 延伸 7 min。扩增产物用 3.0% 的琼脂糖凝胶进行分离,初步检测引物的可用性和多态性。

2 结果与分析

2.1 冬凌草转录组 SSR 位点的数量

利用软件 MISA (<http://pgrc.ipk-gatersleben.de/misa/misa.html>) 对冬凌草转录组测序获得的 44 626 (39 092 024 bp) 条 Unigenes 的 cDNA 序列进行搜索。由表 1 可知,44 626 条 Unigenes 中共检测到 11 114 个 SSR,这些 SSR 主要分布于 8 873 条 Unigenes 序列中,SSR 发生频率为 19.88%,出现频率为 24.90%。平均每 3 517 bp 就含有 1 个 SSR 位点,其中 1 773 条 (13.22%) Unigenes 中含有 1 个以上 SSR 位点,654 条 Unigenes 含有复合 SSR 位点。

表 1 冬凌草转录组中 SSR 搜索结果

Table 1	Summary of SSR searching results
项目 Item	数量 Number
搜索序列总数 Total number of sequences examined	44 626
搜索序列总长度 Total size of examined sequences/bp	39 092 024
含 SSR 位点的序列数目 Number of SSR containing sequences/条	8 873
SSR 总数 Total number of identified SSR	11 114
SSR 发生频率 Occurring frequency/%	19.88
SSR 出现频率 Frequency/%	24.90
含超过 1 个 SSR 位点的序列数目 Number of sequences containing more than one SSR/条	1 773
含复合型 SSR 的序列数目 Number of SSRs present in compound formation	654

2.2 冬凌草转录组 SSR(EST-SSR)特性

2.2.1 冬凌草 EST-SSR 不同重复基元分布 冬凌草转录组 SSR 种类较为丰富,由表 2 可知,单核苷酸至六核苷酸重复类型的 SSR 均有出现,但是出现的频率不同。出现频率较高的为单核苷酸、二核苷酸、三核苷酸重复,占冬凌草总 SSR 的 99.33%,其中二核苷酸重复类型有 5 302 个,

表 3 冬凌草转录组中 SSR 不同重复基元重复次数分布情况

Table 3	Distribution of identified SSRs using the MISA software							
重复类型 Repeat type	重复次数 Repeat number							
	5	6	7	8	9	10	11	≥12
单核苷酸 Mononucleotide						1 514	689	1 738
二核苷酸 Dinucleotide		1 665	1 102	913	903	557	115	7
三核苷酸 Trinucleotide	1 123	438	209	23	1	1	0	1
四核苷酸 Tetranucleotide	38	12	0	1	1	0	0	0
五核苷酸 Pentanucleotide	7	0	0	0	0	0	1	0
六核苷酸 Hexanucleotide	2	7	2	2	1	0	0	1
总计 Total	1 170	2 122	1 313	939	906	2 072	845	1 747
比例 Proportion/%	10.53	19.10	11.81	8.45	8.15	18.64	7.60	15.73
								100.00

所占比例较大,为 47.71%,发生频率为 11.88%,其次为单核苷酸重复类型(3 941 个),占总 SSR 的 35.46%,发生频率为 8.83%,三核苷酸重复类型有 1 796 个,占总 SSR 的 16.16%,出现频率为 4.02%,四核苷酸至六核苷酸重复较少,仅占总 SSR 的 0.67%,发生频率总计为 0.17%。

表 2 冬凌草转录组 SSR 中不同重复基元分布情况

Table 2	Distribution of different repeat motifs in <i>Rabdosia rubescens</i>		
重复类型 Repeat types	SSR 数 Number of SSR	占总 SSR 比例 The proportion of total SSR /%	出现频率 Frequency /%
单核苷酸 Mononucleotide	3 941	35.46	8.83
二核苷酸 Dinucleotide	5 302	47.71	11.88
三核苷酸 Trinucleotide	1 796	16.16	4.02
四核苷酸 Tetranucleotide	52	0.46	0.12
五核苷酸 Pentanucleotide	8	0.07	0.02
六核苷酸 Hexanucleotide	15	0.14	0.03
总计 Total	11 114	100.00	24.90

2.2.2 冬凌草 EST-SSR 重复基元种类及重复次数 由表 3 可知,冬凌草转录组 SSR 中,重复基元的种类较多。在 11 114 个 SSR 位点中,共发现 55 种重复基元,单核苷酸至六核苷酸重复基元分别有 2、4、10、17、7、15 种。其中以单核苷酸重复基元 A/T 最多,重复个数为 3 884 个,占总 SSR 的 34.95%;其次为二核苷酸重复基元 AG/CT、AC/GT,重复个数分别为 3 352、1 123,分别占总 SSR 的 30.16%和 10.10%;在三核苷酸重复基元中,重复基元 AAG/CTT,CCG/CGG 出现次数最多,分别占总 SSR 的 3.26%、3.01%,四核苷酸至六核苷酸重复基元类型虽然较多,但数量较少,占总 SSR 的比例分别为 0.12%、0.02%、0.03%,出现频率也均较低(图 1)。从重复次数来看,冬凌草转录组中 SSR 位点重复次数以 6 次(2 122)、10 次(2 072)较多,分别占总 SSR 的 19.10%、18.64%,其次为重复 7、5 和 8 次,所占比例分别为 11.81%、10.53%和 8.45%。其中单碱基重复以重复 10 次(1 514)最多,所占比例为 38.42%;二碱基重复以重复 6 次(1 665)最多,所占比例为 31.40%;三碱基重复以重复 5 次(1 123)最多,所占比例为 62.53%;四碱基重复以重复 5 次(38)最多,所占比例为 73.08%;五碱基重复以重复 5 次(7)最多,所占比例为 87.50%;六碱基重复以重复 7 次最多,所占比例为 46.67%。

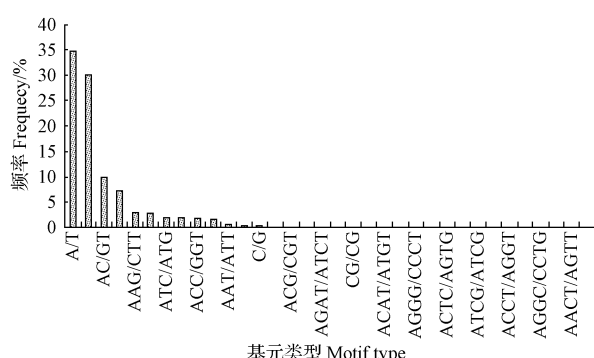


图1 冬凌草转录组中 SSR 基元类型和比例

Fig. 1 Frequency distribution of SSRs based on motif type

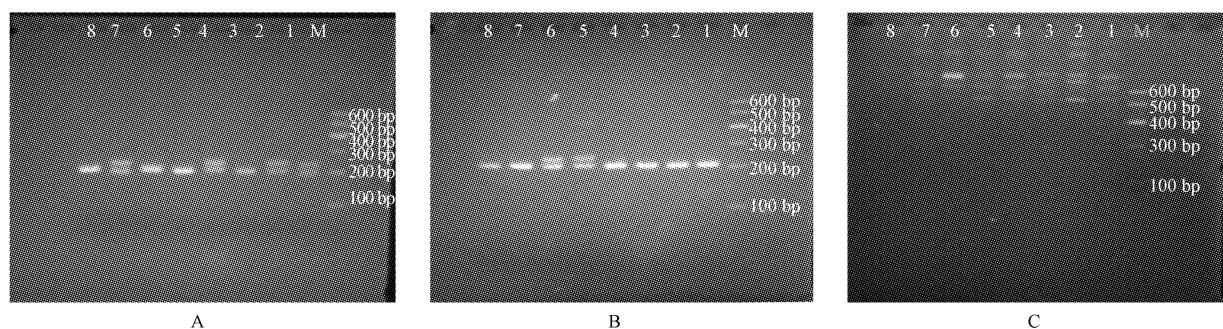
2.3 冬凌草转录组中 SSR 的可用性评价

研究表明当 SSR 长度 ≥ 20 bp 时多态性较高,长度在 12~20 bp 的 SSR 多态性中等,而长度在 12 bp 以下时多态性极低^[10]。冬凌草转录组中 SSR 的长度在 10~232 bp,其中长度在 12~20 bp 的 SSR 有 7 382 条,占

SSR 总数的 66.42%,这类 SSR 位点具有中等多态性;而长度在 20 bp 以上的 SSR 达到 1 131 条,占 SSR 总数的 10.18%,这类 SSR 具有较高多态性。此外,根据 DREISIGACKER 等^[11]的发现,高级基元 SSR 多态性普遍比低级基元的低。经统计发现,长度在 20 bp 以上的冬凌草转录组 SSR 中,包含低级基元一、二、三核苷酸 SSR 共 500 条。可以预计这部分多态性潜能高的 SSR 具有较高的利用价值。

2.4 SSR 引物筛选与多态性分析

随机选择了 26 对引物对 7 种不同来源的冬凌草 DNA 进行 PCR 扩增。扩增结果表明,4 对(15.38%)引物没有扩增产物,5 对(18.23%)扩增产物长度超过预期,1 对引物(3.85%)扩增产物小于预期。其中 16 对(61.54%)PCR 扩增产物与预期大小吻合。利用以上引物,对 8 个产地冬凌草进行了多态性试验,17 对 EST-SSR 引物在 8 个不同产地冬凌草间的多态性扩增。



注:从左到右分别为济源裴村、济源枣庙村、鲁山县辛庄、辉县西莲寺、辉县水磨村、辉县碾平、鲁山县木庙、鲁山宋村、DNA Marker(A,引物 39323;B,引物 34967;C,引物 41764)。

Note: From left to right respectively Pei village, Zaomiao village, Xinzhuang in Lushan county, Xiliansi, Shuimo, Nianping village in Huixian county, Mumiao, Song, Guandimiao village in Lushan county(A, Primer 39323; B, Primer 34967; C, Primer 41764).

图2 代表性引物对不同种群冬凌草的多态性扩增

Fig. 2 Polymorphic amplification of representative primers

3 讨论

分析发现,冬凌草转录组中 SSR 出现频率高,类型丰富、密度大。通过搜索转录组序列,从冬凌草转录组 44 626 条 Unigenes 中共搜索到 11 114 个 SSR,平均出现频率为 24.90%。与其它药用植物比较,冬凌草 SSR 出现的频率高于野三七(16.86%)^[12]、党参(16.10%)^[13]、半夏(16.24%)^[14]、灯盏花(6.26%)^[15]、杜仲^[16](2.90%),出现这种频率差异的原因可能是与搜索序列标准、数据库大小和物种的特异性有关,同时从某种程度上也表明冬凌草转录组中 SSR 数量很丰富,多态性潜能较高。

许多研究表明,植物转录组中的 SSR 以二、三核苷酸重复类型为主,但主要的重复基元类型有所差异。研究发现党参 EST 中二核苷酸重复占总 SSR 58.7%,三核苷酸重复占 21.63%,半夏转录组中 SSR 的主要重复

类型是二核苷酸和三核苷酸,二者共占 SSR 总数的 92.65%,白菜 EST 中二核苷酸、三核苷酸重复占主导类型^[17]。该研究表明冬凌草转录组中 SSR 以二核苷酸重复类型最多,占总 SSR 的 47.71%,其次为单核苷酸类型,这与以往的报道不尽相同,可能与冬凌草 EST-SSR 自身的特点有关。

GA 在大多数植物二核苷酸重复中占主导类型,该研究表明 AG 是冬凌草转录组中出现频率最高的基元类型,这与白菜、番茄、普通小麦等植物转录组结果有所不同,但与党参、半夏的转录组分析结果相一致,GC 重复基元在多数植物中很难见到或者出现频率较低,如拟南芥、玉米等植物中均未发现该重复基元,但在党参、半夏、杜仲等的转录组分析中发现了出现频率极低的 GC 重复基元,以水稻为代表的单子叶植物基因组则倾向出

现高 G+C 含量的重复类型^[18], 该研究表明冬凌草转录组中发现了 57 个 GC 重复, 出现频率为 0.51%, 远远高于党参的 5 个 GC 重复。

冬凌草三核苷酸重复中 AAG/CTT 占主导地位, 这与党参、马蹄香^[19]的报道结果相一致, 而与半夏(AAT/ATT)、白菜(GAA/TTC)、杜仲(AGA/TCT)等的类型不同。这种分布差异可能与物种基因组的组成结构有关, 但也有可能与物种 EST 来源及 EST 数目不同所致。如马蹄香、白菜数据来源于 NCBI 的数据库, EST 数目也仅为数千条, 而其它植物如杜仲、党参、野三七夏等的数据均来源于高通量测序的结果, 且 EST 数目较多, 均为数万条。

总之, 从该研究结果可知, 冬凌草转录组 SSR 类型丰富且出现频率高, 从多态性的潜能来看, 冬凌草转录组中长度在 20 bp 以上的 SSR 达 1 132 条, 占 SSR 总数的 10.18%, 其中有 44.17% 是多态性潜能高的以低级基元一、二、三核苷酸重复类型的 SSR, 这部分 SSR 可能具有较高的利用价值。这表明冬凌草转录组 SSR 进行分子标记具有很大的潜力。另外, 该试验中多态性的检测仅仅采用了琼脂糖凝胶电泳, 如采用非变性聚丙烯酰胺凝胶, 呈现多态性引物的比例将会更高, 这是今后需要进一步开展的工作。

因此, 该研究的结果对于丰富冬凌草分子标记类型、遗传资源评价及加速冬凌草功能因资源的开发利用具有重要的意义。同时也为进一步开发新的冬凌草功能基因、SSR 标记奠定了基础。

参考文献

- [1] 孙汉董, 韩全斌. 冬凌草的植物资源、化学和抗癌活性成分的研究 [C]. 中国植物学会七十周年年会论文摘要汇编 (1933-2003).
- [2] 刘净, 谢韬, 魏秀丽, 等. 冬凌草化学成分的研究 [J]. 中国天然药物, 2004, 2(5): 276-279.

- [3] 郑晓珂, 李钦, 冯卫生. 冬凌草中酚酸类化学成分研究 [J]. 中国药理学杂志, 2004, 39(5): 335-336.
- [4] 李钦, 冯卫生. 冬凌草化学成分、药理作用及开发研究进展 [J]. 河南中医学院学报, 2003, 18(6): 31-33.
- [5] 宋芳, 冯一中, 蒋小岗, 等. 冬凌草甲素对人胰腺癌 SW1990 细胞系生长抑制作用 [J]. 中国药理学通报, 2010, 26(2): 240-243.
- [6] 陈随清, 冯卫生, 王丽, 等. 冬凌草化学成分积累动态的研究 [J]. 世界科学技术-中医药现代化, 2007, 9(3): 67-70.
- [7] 董诚明, 苏秀红, 王伟丽. 氮碳源对冬凌草再生植株生长及次生代谢产物的影响 [J]. 西北植物学报, 2009, 29(3): 494-498.
- [8] 李国民, 李忠超, 许凯扬, 等. 抗癌植物冬凌草种质资源遗传多样性分析 [J]. 热带亚热带植物学报, 2008, 16(2): 116-122.
- [9] 李小白, 向林, 罗洁, 等. 转录组测序 (RNA-seq) 策略及其数据在分子标记开发上的应用 [J]. 中国细胞生物学学报, 2013, 35(5): 720-726, 740.
- [10] 杨华, 陈琪, 韦朝领, 等. 茶树转录组中 SSR 位点的信息分析 [J]. 安徽农业大学学报, 2011, 38(6): 882-886.
- [11] DREISIGACKER S, ZHANG P, WARBURTON M L, et al. SSR and pedigree analyses of genetic diversity among CIMMYT wheat lines targeted to different mega environments [J]. Crop Science, 2004, 44(2): 381-388.
- [12] 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.
- [13] 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.
- [14] 王森, 张震, 姜倪皓, 等. 半夏转录组中的 SSR 位点信息分析 [J]. 中药材, 2014, 37(9): 1566-1569.
- [15] 陈茵, 李翠婷, 姜倪皓, 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.
- [16] 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.
- [17] 忻雅, 崔海瑞, 卢美贞, 等. 白菜 EST-SSR 信息分析与标记的建立 [J]. 园艺学报, 2006, 33(3): 549-554.
- [18] 李珊珊, 孙春玉, 蒋世翠, 等. SSR 分子标记及其在植物遗传育种中的应用 [J]. 吉林蔬菜, 2014(5): 33-38.
- [19] 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.

Analysis of SSR Loci Information in Transcriptome of *Rabdosia rubescens* (Henmsl.) Hara and Its Polymorphism

SU Xiuhong, LI Qinglei, CHEN Suiqing, DONG Chengming, YANG Jingfan, YIN Lei
(College of Medicine, Henan College of Traditional Chinese Medicine, Zhengzhou, Henan 450046)

Abstract: Taking transcriptome of *Rabdosia rubescens* as research object, using Microsatellite (MISA) software, SSR loci information of *Rabdosia rubescens* was analyzed, and SSR primers were designed to evaluate the polymorphism. The results showed that 11 114 SSR loci were found, which distributed in 8 873 Unigenes, SSR loci frequency was 24.90%. There were a total of 55 kinds of repeating units, average per 3 517 bp containing a SSR loci. The dinucleotide repeating was the most abundant type, accounting for 47.71% of the total number of SSRs, followed by a single base repeat type (35.46%). A/T and AG/CT were the main repeat motifs, accounting for 34.95% and 30.16%, respectively. Conclusion was SSR loci of *Rabdosia rubescens* had a high frequency, high polymorphism.

Keywords: *Rabdosia rubescens*; transcriptome; simple sequence repeat