Black枸杞转录组 SSR 信息分析及分子标记开发

尹跃，安巍，赵建华，王亚军，樊云芳，曹有龙

（宁夏林科院 国家枸杞工程技术研究中心，宁夏 银川 750002）

SSR information in transcriptome and development of molecular markers in Lycium ruthenicum

YIN Yue, AN Wei, ZHAO Jianhua, WANG Yajun, FAN Yunfang, CAO Youlong

（National Wolfberry Engineering Research Center, Ningxia Academy of Agriculture and Forestry Sciences, Yinchuan 750002, Ningxia, China）

Abstract: The objective was to analyze Simple Sequence Repeat (SSR) loci information and to develop SSR markers in Lycium ruthenicum (wolfberry) for further analysis with genetic diversity. A total of 73 896 SSRs were identified in L. ruthenicum transcriptome by MISA software, and primer pairs were designed from the SSR loci by BatchPrimer 3 software. Results showed that distribution was in 56 170 unigenes with a frequency of 26.36% and an average density of one SSR per 3.55 kb. Among these SSR loci, the most abundant motifs were mononucleotide (74.33%), dinucleotide (13.30%), and trinucleotide (11.81%) repeat types. Among 73 896 SSR loci, 84 repeat motifs with iteration numbers from 5 to 33 were discovered. The most abundant repeat motifs were A/T, AG/CT, AT/AT, C/G, and AAC/GTT. A total of 12 674 primer pairs were designed from the SSR loci by BatchPrimer 3 software, and 128 primers pairs of randomly selected SSR were validated, of which 74 (57.8%) pairs of primers were able to produce Polymerase Chain Reaction (PCR) products. Among the 74 SSR loci, 28 high polymorphic SSR loci were used to amplify 24 accessions of wolfberry germplasm. A total of 256 alleles were detected for an average of 9.1. The average major allele frequency (0.432), observed heterozygosity (0.439), expected heterozygosity (0.712), and polymorphism information content (PIC) (0.678) were noted. Thus, the development of SSR markers based on the transcriptome of L.
组测序获得的数据为基础。测序技术和生物信息学方法快速发展，也引起人们对重复序列的研究兴趣。虽然来自表达的基因组区域的单核苷酸多态性（SNP）检测已被广泛应用遗传图谱构建及种质资源鉴定和遗传多样性分析等7-9领域。由于缺乏基因组信息，基于基因组测序开发黑果枸杞 SSR 标记成本高，步骤繁琐11-12，而 SSR 标记在黑果枸杞基因组资源研究中的应用受到限制。同时，美国生物技术信息中心 (NCBI) 中公布的黑果枸杞表达序列标签 (EST) 极少，除了 CHEN 等10 研究了基于盐胁迫下黑果枸杞转录组表达序列标签 (EST-SSR) 标记开发之外，基于转录组测序黑果枸杞 EST-SSR 标记开发尚未见报道。EST-SSR 来源于表达的基因组区域，可直接反映相关基因多样性，在不同物种间具有良好的通用性；随着高通量测序技术和生物信息学方法快速发展，基于转录组测序已开发刺梨 Rosa roxburghii，中国樱桃 Cerasus pseudocerasus，马铃薯 Solanum tuberosum 等13-15 多种植物的 SSR 标记。本研究以前期黑果枸杞抗旱转录组测序获得的数据为基础，利用生物信息学方法批量开发 EST-SSR 标记，并分析其分布特点、组成特征，以期为黑果枸杞种质资源遗传多样性分析研究提供参考。

1 材料与方法

1.1 转录组数据来源

取 1 年生黑果枸杞持续干旱胁迫下的叶和根，提取 RNA 后进行转录组测序 (无参，HiSeq TM2500，北京诺禾致源生物信息科技有限公司)，共 12 个样本，计 80 G 数据。利用 Trinity 等软件16 对测序的数据进行拼接和组装，获得 213 058 条单基因簇 (unigene) 作为背景数据进行分析。

1.2 植物材料及 DNA 提取

用于 SSR 引物筛选及评价的 24 份枸杞种质资源（表 1）均来自宁夏农林科学院国家枸杞工程技术研究中心枸杞种质资源圃（38°38′49″N，106°9′10″E）。利用天根试剂盒 (DP5-02，天根) 提取基因组 DNA。

1.3 转录组 SSR 位点鉴别及引物设计

使用 MISA 软件 (http://pgrc.ipk-gatersleben.de/misa/misa.html) 搜索 Unigene 序列的 SSR 位点。搜索标准为：重复单元长度 1-6 bp，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，复合型 SSR 位点碱基间隔 ≤ 100 bp。

采用 BatchPrimer 3 软件 (https://probes.pw.usda.gov/batchprimer3/) 对获得的 SSR 位点批量设计引物。引物设计参数为：引物长度 18-27 bp（最佳 22 bp），GC 为 40%-70%（最适 50%），退火温度为 50-60 ℃（最适 55 ℃），PCR 扩增产物长度为 100-300 bp，其他参数为默认设置。

1.4 EST-SSR 引物筛选及验证

随机挑选 128 对引物，在上游引物 5‘端添加 18 bp 的 M13 通用引物序列 (TGTTAACGACGCAGCTT)，3’端保持不变，用 FAM 荧光基团修饰后，由英潍捷基基贸易有限公司合成。

PCR 扩增体系 (15.0 μL): DNA 模板 1.0 μL，M13 通用荧光引物 0.1 μL，上游引物和下游引物各 0.4 μL，2X Taq PCR Master Mix 7.5 μL，双蒸水 5.6 μL。PCR 扩增在 ABI-2720（应用生物系统公司，美国）上进行。扩增程序为：95 ℃ 5 分；95 ℃ 30 s，60-45 ℃ 30 s，72 ℃ 30 s，15 个循环；95 ℃ 30 s，50 ℃ 30 s，72 ℃ 30 s，20 个循环；72 ℃ 7 min。扩增产物经过 ABI3730XL DNA (应用生物系统公司，美国) 检测，用 LIZ500 作为分子量内标。
表 1 供试种质材料信息

<table>
<thead>
<tr>
<th>序号</th>
<th>种质名称</th>
<th>果实颜色</th>
<th>序号</th>
<th>种质名称</th>
<th>果实颜色</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>‘宁杞 1 号’ Lycium bararum ‘Ningqi 1’</td>
<td>红色</td>
<td>13</td>
<td>黑果枸杞 L. ruthenicun</td>
<td>黑色</td>
</tr>
<tr>
<td>2</td>
<td>‘宁杞 2 号’ L. bararum ‘Ningqi 2’</td>
<td>红色</td>
<td>14</td>
<td>中国枸杞 L. chinsense</td>
<td>红色</td>
</tr>
<tr>
<td>3</td>
<td>‘宁杞 5 号’ L. bararum ‘Ningqi 5’</td>
<td>红色</td>
<td>15</td>
<td>昌吉枸杞 L. changxicun</td>
<td>粉红色</td>
</tr>
<tr>
<td>4</td>
<td>‘宁杞 7 号’ L. bararum ‘Ningqi 7’</td>
<td>红色</td>
<td>16</td>
<td>新疆枸杞 L. dasysystem</td>
<td>红色</td>
</tr>
<tr>
<td>5</td>
<td>‘宁杞菜 1 号’ L. bararum ‘Ningqicai 1’</td>
<td>红色</td>
<td>17</td>
<td>‘HGB’ L. bararum ‘HGB’</td>
<td>黄色</td>
</tr>
<tr>
<td>6</td>
<td>‘宁杞菜 3 号’ L. bararum ‘Ningqicai 3’</td>
<td>红色</td>
<td>18</td>
<td>SC-09 (拉丁名未知)</td>
<td>红色</td>
</tr>
<tr>
<td>7</td>
<td>‘宁杞菜 2 号’ L. bararum ‘Ningqi 2’</td>
<td>红色</td>
<td>19</td>
<td>HB-09 (拉丁名未知)</td>
<td>红色</td>
</tr>
<tr>
<td>8</td>
<td>‘宁杞菜 5 号’ L. bararum ‘Ningqi 5’</td>
<td>红色</td>
<td>20</td>
<td>DHG (拉丁名未知)</td>
<td>红色</td>
</tr>
<tr>
<td>9</td>
<td>‘宁杞菜 9 号’ L. bararum ‘Ningqi 9’</td>
<td>红色</td>
<td>21</td>
<td>AN-YN-01 (拉丁名未知)</td>
<td>红色</td>
</tr>
<tr>
<td>10</td>
<td>‘蒙杞 1 号’ L. bararum ‘Mengqi 1’</td>
<td>红色</td>
<td>22</td>
<td>W-12-30 (拉丁名未知)</td>
<td>黄色</td>
</tr>
<tr>
<td>11</td>
<td>‘蒙杞 2 号’ L. bararum ‘Mengqi 2’</td>
<td>红色</td>
<td>23</td>
<td>W-12-27 (拉丁名未知)</td>
<td>黄色</td>
</tr>
<tr>
<td>12</td>
<td>‘天精 3 号’ (拉丁名未知)</td>
<td>红色</td>
<td>24</td>
<td>W-11-15 (拉丁名未知)</td>
<td>黄色</td>
</tr>
</tbody>
</table>

说明：1-12 为栽培品种，13-24 为野生资源

1.5 数据分析

以 SSR 出现频率和 SSR 平均分布距离描述 EST-SSR。公式如下：①SSR 出现频率 f = n/100，其中 n 为所搜索到的 SSR 次数(个)，N 为所有 EST 数量(个)。②SSR 平均分布距离 f = N/n，其中 N 为所有 EST 数量的总和(个)。由 GeneMapper 4.0 软件获得扩增产物片段大小，用 DataFormatter 2.7 软件将数据转化为 PowerMarker v3.25 软件的输入格式，计算等位基因数(number of alleles，N)，主效等位基因频率(major allele frequency，f_M)，期望杂合度(expected heterozygosity，H_e)、观察杂合度(observed heterozygosity，H_o)和多态信息量(polymorphism information content，C_P)

2 结果与分析

2.1 黑果枸杞转录组中 SSR 位点的分布特点

利用 MISA 软件对黑果枸杞转录组测序获得的 213 058 条 Unigene(序列总长度为 262 643 598 bp)序列进行搜索，发现其中 56 170 条 Unigene 序列中含有 73 896 个 SSR 位点，其中 13 611 条 Unigene 含有 2 个或 2 个以上 EST-SSR 位点。总体上，SSR 发生频率为 26.36%，平均 3.55 kb 位置出现 1 个 SSR 位点。SSR 类型丰富，单核苷酸至六核苷酸重复类型均存在；单核苷酸、二核苷酸和三核苷酸重复出现频率占优势，分别占总 SSR 的 74.33%，13.30% 和 11.81%；四核苷酸、五核苷酸和六核苷酸重复类型数量较少，分别占总 SSR 的 0.49%，0.03% 和 0.04%(表 2)。

表 2 黑果枸杞转录组 SSR 数量、类型和频率

<table>
<thead>
<tr>
<th>重复类型</th>
<th>不同重复次数下 SSR 数量/条</th>
<th>总数/条</th>
<th>百分比/%</th>
</tr>
</thead>
<tbody>
<tr>
<td>单核苷酸</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>二核苷酸</td>
<td>3 661</td>
<td>2 116</td>
<td>1 502</td>
</tr>
<tr>
<td>三核苷酸</td>
<td>5 285</td>
<td>2 414</td>
<td>982</td>
</tr>
<tr>
<td>四核苷酸</td>
<td>5</td>
<td>317</td>
<td>32</td>
</tr>
<tr>
<td>五核苷酸</td>
<td>16</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>六核苷酸</td>
<td>16 631</td>
<td>6 117</td>
<td>3 110</td>
</tr>
</tbody>
</table>

2.2 转录组 SSR 基序重复类型和频率特征

从黑果枸杞转录组 SSR 核苷酸基序重复类型来看，73 896 个 SSR 位点共有 84 种重复基元，单核苷酸至六核苷酸重复数目比较为 2, 4, 10, 26, 18 和 24 种。从出现的频率来看(表 3)：占优势的前 3 种重复
单元类型是单核苷酸、二核苷酸和三核苷酸重复基元，单核苷酸重复基元以 A/T 为主，占该类型 SSR 位点总数的 95.64%；二核苷酸主要以 AG/CT 为主，占二核苷酸总数的 40.42%，其次是 AT/AT，AC/GT 和 CG/GG，分别所占比例为 35.37%，23.74% 和 4.06%；三核苷酸重复单元以 AAC/GTT，AAG/CTT，ATT/ATT，ATC/ATG、ACC/CTG 为主，占所有三核苷酸重复单元的 75.88%；四核苷酸、五核苷酸和六核苷酸重复单元类型分布较为分散，出现频率相对较低，仅为 0.56%。

表 3 黑果枸杞 EST-SSR 中重复基元的类型、数量及频率

<table>
<thead>
<tr>
<th>重复基序类型</th>
<th>数量/条</th>
<th>占该类型 SSR 比例%</th>
<th>占总 SSR 比例%</th>
<th>重复基序类型</th>
<th>数量/条</th>
<th>占该类型 SSR 比例%</th>
<th>占总 SSR 比例%</th>
</tr>
</thead>
<tbody>
<tr>
<td>A/T</td>
<td>52531</td>
<td>95.64</td>
<td>71.09</td>
<td>ACG/GGT</td>
<td>635</td>
<td>7.28</td>
<td>0.86</td>
</tr>
<tr>
<td>CG/CT</td>
<td>2074</td>
<td>3.46</td>
<td>3.24</td>
<td>ACG/CTG</td>
<td>577</td>
<td>6.84</td>
<td>0.81</td>
</tr>
<tr>
<td>AG/CT</td>
<td>2333</td>
<td>23.74</td>
<td>3.16</td>
<td>ACT/AGT</td>
<td>319</td>
<td>3.66</td>
<td>0.43</td>
</tr>
<tr>
<td>AT/AT</td>
<td>2973</td>
<td>40.42</td>
<td>5.38</td>
<td>AGG/CTG</td>
<td>597</td>
<td>6.84</td>
<td>0.81</td>
</tr>
<tr>
<td>CG/GT</td>
<td>347</td>
<td>35.37</td>
<td>4.71</td>
<td>AGG/CTG</td>
<td>424</td>
<td>4.86</td>
<td>0.57</td>
</tr>
<tr>
<td>AAG/GTT</td>
<td>2108</td>
<td>24.16</td>
<td>2.85</td>
<td>ATC/ATG</td>
<td>978</td>
<td>11.21</td>
<td>1.32</td>
</tr>
<tr>
<td>AAG/CTT</td>
<td>1865</td>
<td>21.37</td>
<td>2.52</td>
<td>CCG/CGG</td>
<td>591</td>
<td>6.77</td>
<td>0.80</td>
</tr>
<tr>
<td>AAT/ATT</td>
<td>1305</td>
<td>11.86</td>
<td>1.40</td>
<td>其他类型</td>
<td>414</td>
<td>4.82</td>
<td>0.56</td>
</tr>
</tbody>
</table>

2.3 SSR 引物有效性及多态性检测

为检测所开发的 EST-SSR 标记的可用性，对 56170 条 Unigene 序列的 73896 个 EST-SSR 位点设计引物，共得到引物 12674 对。随机挑选 128 对引物扩增 20 个不同物种转录组，以检测到的等位基因为有效引物，而位点搜索时并未将单核苷酸重复单元设置为搜索对象。

表中数据表明，共得到引物 12674 对，有效引物 12640 对，有效引物占总引物的 99.33%，引物扩增有效率为 99.33%。引物扩增的有效性可用多态信息量、多态性引物占有效引物的百分比、引物扩增的多态位点数、引物扩增的等位基因数等来评价。

2.4 SSR 位点遗传多样性分析

以筛选的 28 对多态性引物对 24 份枸杞植物种质的遗传多样性进行分析，共检测到等位基因 256 个，引物扩增出的等位基因为 4~19 个，平均为 9.1 个；共检测到引物型数 303 个，有效等位基因频率变化范围为 0.188~0.729，平均为 0.432；观察杂合度为 0.167~0.833，平均为 0.439；期望杂合度为 0.433~0.904，平均为 0.712，多态信息量为 0.395~0.897，平均为 0.678。由 Botstein 理论预测出的等位基因数 256 个。

3 讨论

本研究搜索了黑果枸杞的 213058 条 Unigene 序列，发现其中的 56170 条中含有 73896 个 SSR 位点，发生频率为 26.36%；高于中国樱桃 Cerasus pseudocerasus (15.62%) [14]，马铃薯 Solanum tuberosum (3.43%) [15]。夏橙 Hexinoclyanthus chinensis (21.25%) [16]，马尾松 Pinus massoniana (4.69%) [17] 和中间锦鸡儿 Caragana intermedia (14.78%) [18]，低于蓝莓果忍冬 Lonicera caerulea var. edulis (32.51%) [19]，山桐子 Ilesia polycarpa (35.00%) [20]，普通油茶 Camellia oleifera (39.67%) [21] 和芙蓉李 Prunus salicina (54.51%) [22]。可见，不同物种转录组 SSR 位点的出现频率不同，出现差异的原因除了物种本身差异，还可能与分析数据库大小、SSR 挖掘工具及搜索条件有关。

从重复单元频率来看，黑果枸杞转录组的 SSR 主要类型为单核苷酸重复基序 (74.33%)，其次为二核苷酸重复 (13.30%) 和三核苷酸重复 (11.81%)。研究发现：多数植物中 EST-SSR 以二核苷酸和三核苷酸为主，但重复基序类型不同 [23]。推测原因可能与 SSR 搜索参数设置有关，多数植物在进行 SSR 位点搜索时并未将单核苷酸重复基序设置为搜索对象 [12,14]。

本研究从 213058 条 Unigene 序列中鉴定出 73896 个 SSR 位点，丰富了黑果枸杞 EST-SSR 标记的数量。为了进一步评估这些 EST-SSR 引物的质量，从设计到的 12674 对 EST-SSR 引物中随机挑选出 128 对引物，54 对引物未能扩增出产物，原因可能是扩增产物中存在大量内含子或引物设计不合理 [24,25]。
有 74 对引物成功扩增出产物，引物扩增有效性达 57.8%，表明开发的引物质量较高，对后续黑果枸杞种质资源遗传多样性分析、品种鉴定、遗传图谱构建和分子标记辅助育种等均有应用价值。
4 参考文献


