

萝卜 EST 资源的 SSR 信息分析及 EST-SSRs 标记开发

崔娜, 邱杨, 李锡香*, 沈镛, 王海平, 宋江萍

(中国农业科学院蔬菜花卉研究所, 农业部蔬菜作物基因资源与种质创制北京科学观测实验站, 北京 100081)

摘要: 从 NCBI 数据库下载了 287 349 条萝卜 EST 序列, 经预处理后得到无冗余 EST 序列 58 105 条, 全长 38 622.476 kb。利用 MISA 搜索 SSR 位点, 得到含有 SSR 位点的 EST 序列 3 523 条, 共 3 718 个 SSR, 平均每 10.39 kb 就出现 1 个 SSR。分析发现萝卜 EST 序列中存在 265 种重复基元类型, 其中二、三、六核苷酸重复是主导的重复基元类型, 二核苷酸中以比例高达 91.34% 的 AG/CT 重复基元为主; 三核苷酸中主导的重复基元类型是 AAG/CTT, 比例为 35.86%; 六核苷酸中, 存在两种主要的重复基元, 分别为 AAGGAG/CCTCTT 和 AAGAGG/CCTTCT, 所占比例为 12.78%。这些结果表明, 萝卜 EST-SSRs 出现频率较高, 类型丰富, 具有良好的多态性潜能和开发利用价值。使用 primer3.0 批量设计, 并以不同主导重复基元类型初步合成 183 对 EST-SSR 引物, 以 12 份典型萝卜种质、1 对亲本及其 F₁ 的基因组 DNA 为模板, 对其有效性进行验证, 筛选出有清晰扩增产物的引物 159 对, 其中具多态性引物 64 对; 在试验亲本及其 F₁ 中表现多态性的引物 30 对, 共显性引物 27 对。

关键词: 萝卜; EST; SSR 特征; 标记开发

中图分类号: S 631.1

文献标识码: A

文章编号: 0513-353X (2012) 07-1303-10

Data Mining for SSRs in EST Resources and EST-SSR Markers Development in Radish

CUI Na, QIU Yang, LI Xi-xiang*, SHEN Di, WANG Hai-ping, and SONG Jiang-ping

(*Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing Research Station of Vegetable Crop Gene Resource and Germplasm Enhancement, Ministry of Agriculture, P.R China, Riculture, Beijing 100081, China*)

Abstract: A total of 287 349 ESTs of radish were downloaded from NCBI database. After preprocessing, 58 105 non-redundant ESTs with the total length of 38 622.476 kb were obtained. By using the MISA program, a total of 3 718 SSRs were observed in 3 523 ESTs and the average distance between SSRs was 10.39 kb. 265 types of motifs were found in radish's ESTs and the dominant motifs types are dinucleotide, trinucleotide and exanucleotide repeats. AG/CT was the main motifs type (91.34%) in dinucleotide repeats; AAG/CTT was the main motifs type (35.86%) in trinucleotide; AAGGAG/CCTCTT and AAGAGG/CCTTCT were two main motifs types (total 12.78%) in exanucleotide. Comprehensively analysis indicated that EST-SSRs of radish with high frequency and rich motifs types were potential for the

收稿日期: 2012-01-09; **修回日期:** 2012-05-24

基金项目: 国家 '863' 计划课题 (2012AA021801); 农业部保种子项 (NB2012-2130135-28); 农业部 '948' 项目 (2011-G1-02); 农业部园艺作物生物学与种质创制重点实验室项目

* 通信作者 Author for correspondence (E-mail: lee0612@sina.com.cn)

development of polymorphic primers. Using the primer 3.0, 183 primers were designed and synthesized based on different dominant motifs types, and verified with 12 typical radish germplasms for polymorphism and one pair of parents and their F₁ for codominance. The results showed that the products of 159 primers are clear and effective, 64 pairs were polymorphic among the germplasms and 27 were codominant between the parents and their hybrids.

Key words: radish; EST; SSR characteristics; markers development

在萝卜 (*Raphanus sativus* L.) 的遗传育种中, 育种者仍以传统育种手段为主, 依据表型和经验判断筛选材料, 育种周期长, 优良性状难以聚合, 育种进展缓慢。此外, 萝卜分子生物学的研究相对滞后, 分子标记的开发尚不能满足遗传育种的需要。

简单重复序列 (simple sequence repeat, SSR) 标记具有在全基因组中随机分布、信息含量高、可靠性强、重复性高、多态性丰富等优点, 同时, SSR 标记属共显性标记, 呈孟德尔遗传, 使得其在遗传关系分析、遗传图谱的构建、品种鉴定等方面得到广泛应用。

近年来, 测序技术的发展突飞猛进, 各种植物的基因组和转录组测序数据迅速增加。就 EST (Expressed Sequence Tags) 资源而言, 在不同物种中 SSR 的出现频率、特点、分布不尽相同, 开发标记的效率也不相同。基于 EST 开发 SSR 标记在花生 (Wei et al., 2011)、油菜 (李小白 等, 2007; Wang et al., 2011)、大豆 (陈相艳 等, 2009)、黄瓜 (胡建斌和李建吾, 2008)、白菜 (忻雅 等, 2006)、棉花 (Qureshi et al., 2004) 和六倍体小麦 (Ju et al., 2004) 等作物中均有报道。尽管白菜和油菜与萝卜同属十字花科, 且基于白菜和油菜 EST 资源开发出的 SSR 标记已广泛应用于芸薹属基因组研究和分子育种 (忻雅 等, 2005; 李丽和郑晓鹰, 2009; Wang et al., 2011), 但在萝卜研究中, 这些标记的通用性不强。关于萝卜 EST-SSR 引物的开发也有报道 (Wang et al., 2007; Nakatsuji et al., 2011; Shirasawa et al., 2011), 但在 EST 资源的利用上很有限, 在 EST 的 SSR 信息分析方面尚不足, 引物开发和应用的效率有待提高。

NCBI (National Center for Biotechnology Information) 网站上公布了来源于不同作者的大量的萝卜 EST 数据, 这为基于 EST 开发适合萝卜的 SSR 标记提供了基础。本研究中拟基于 NCBI 网站中丰富的萝卜 EST 信息资源 (截止至 2011 年 2 月 1 日), 分析 SSR 的分布及其特点, 并通过检测已合成引物的多态性和共显性, 确定开发标记的可行性和有效性。

1 材料与方法

1.1 萝卜 EST 来源

萝卜 EST 序列来源于 NCBI 的 EST 数据库。在 All database 中选择 EST, 使用 “radish” 关键词来检索萝卜 EST 序列, 选择 *Raphanus sativus*, *Raphanus raphanistrum* subsp. *raphanistrum*, *Raphanus raphanistrum* subsp. *landra*, *Raphanus raphanistrum* subsp. *maritimus*, *Raphanus sativus* var. *oleiformis*, *Raphanus sativus* var. *niger* (其中包括野萝卜), 在 “send to” 提示框后选择 “file”, 将检索结果以 FASTA 格式下载并保存, 共计 287349 条 (截止至 2011 年 2 月 1 日)。

1.2 EST 预处理

采用 EST-trimmer (http://pgrc.ipk-gatersleben.de/misa/download/est_trimmer.pl) perl 脚本, 去除 5'端或 3'端长度为 50 bp 的 polyT 或 polyA 以及长度小于 100 bp 的 EST 序列, 对于长度大于 700 bp

的 EST 序列保留其自 5'端起长度为 700 bp 的序列。

1.3 去冗余和拼接

预处理后的 EST 通过软件 CAP3 (<http://seq.cs.iastate.edu/cap3.html>) 进行片段重叠群分析和聚类处理, 拼接时设定的初始装配参数使用默认值。

1.4 SSR 位点搜索和分析

采用 MISA.perl 脚本(<http://pgrc.ipk-gatersleben.de/misa/misa.html>), 于 MISA 系统配置文件(misa.ini) 中修改检索 SSR 位点的最低长度标准, 进行 SSR 位点查找。查找标准设定为: 最低长度 18 bp, 即单核苷酸、二核苷酸、三核苷酸、四核苷酸、五核苷酸、六核苷酸的最少重复次数分别在 18、9、6、5、4、3 次以上, 与此同时, 将间隔小于或等于 10 bp 的被打断的不完全重复 SSR 位点也列为搜索对象。将生成的文本格式文件导入 Excel 表中, 对 EST-SSRs 的基本信息进行统计分析。

EST-SSR 的出现频率 (%) = (检出的 SSR 个数/无冗余 EST 总数) × 100; EST-SSR 平均分布距离 = 无冗余 EST 总长度/SSR 个数。优势基元中 SSR 出现频率 (%) = (优势重复基元 SSR 数量/优势重复基元 SSR 总数量) × 100。

1.5 EST-SSR 引物的设计和合成

使用 primer3.0(<http://sourceforge.net/projects/primer3/files/primer3/1.1.4/primer3-1.1.4-WINXP.zip/download>) 设计引物。以不同主导重复基元类型有选择性地合成 183 对 EST-SSR 引物。引物由上海英骏生物技术有限合成。

1.6 引物有效性检测

引物多态性检测所用的 12 份萝卜种质见表 1。

表 1 12 份萝卜高代自交系
Table 1 The advanced inbred lines of 12 typical radish germplasms

| 序号 Number | 编号 Codes | 品种名称 Name | 类型 Type | 来源 Source |
|--------------|-------------|-------------------------|-----------------------------|----------------------|
| 1 | KB10Q-1 | 夏白玉 Xiabaiyu | 高代自交系 Advanced inbred lines | 韩国 Korea |
| 2 | KB10Q-3 | 秦菜 2 号 Qincai 2 | 高代自交系 Advanced inbred lines | 中国陕西 Shaanxi, China |
| 3 | KB10Q-5 | 洛阳露头青 Luoyang Lutouqing | 高代自交系 Advanced inbred lines | 中国河南 Henan, China |
| 4 | KB10Q-10 | 生粹大根 Shengcui Dagen | 高代自交系 Advanced inbred lines | 日本 Japan |
| 5 | L10Q-113 | 国光 Guoguang | 高代自交系 Advanced inbred lines | 中国山东 Shandong, China |
| 6 | L10Q-184 | 北京大红袍 Beijing Dahongpao | 高代自交系 Advanced inbred lines | 中国北京 Beijing, China |
| 7 | L10Q-433 | 圆红萝卜 Yuanhong Luobo | 高代自交系 Advanced inbred lines | 中国北京 Beijing, China |
| 8 | PP10Q-23 | 耐热白 Nairebai | 高代自交系 Advanced inbred lines | 中国武汉 Wuhan, China |
| 9 | PP10Q-25 | 长莱萝卜 Changjia Luobo | 高代自交系 Advanced inbred lines | 泰国 Thailand |
| 10 | PP10Q-29 | 野生萝卜 Yesheng Radish | 高代自交系 Advanced inbred lines | 美国 America |
| 11 | L10-1-1 | YR | 高代自交系 Advanced inbred lines | 韩国 Korea |
| 12 | Z10Q-xlm | 心里美 Xinlimei | 高代自交系 Advanced inbred lines | 中国北京 Beijing, China |

用于检测共显性的两份亲本分别是本课题组的自交系: YS08Q-5 (中国) 和 YS08Q-9 (中国), F_1 为这两份亲本的杂交一代。

使用 CTAB 改良法提取上述萝卜材料 DNA。经 PCR 扩增后使用 8% 的丙烯酰胺凝胶电泳进行产物检测, 观察其多态性和共显性。有效扩增比例 (%) = (扩增出产物的引物数/所有引物数) × 100; 多态性比例 (%) = (呈多态性引物数/有效扩增引物数) × 100。

多样性指数 $PIC = 1 - \sum_{i=1}^n p_i^2 - 2 \left[\sum_{i=1}^{n-1} \sum_{j=i+1}^n p_i^2 p_j^2 \right]$, 其中 p_i 为等位基因 i 的出现频率, n 为等位基因数。

1.7 PCR 反应体系及程序

15 μL 的反应体系: DNA 模板 ($15 \text{ ng} \cdot \mu\text{L}^{-1}$) 3 μL , $10 \times$ Buffer 1.5 μL , Mg^{2+} ($25 \text{ mmol} \cdot \text{L}^{-1}$) 1.2 μL , dNTPs ($2.5 \text{ mmol} \cdot \text{L}^{-1}$) 0.2 μL , Forward primer ($10 \mu\text{mol} \cdot \text{L}^{-1}$) 0.25 μL , Reverse primer ($10 \mu\text{mol} \cdot \text{L}^{-1}$) 0.25 μL , *Taq* 酶 ($2.5 \text{ U} \cdot \mu\text{L}^{-1}$) 0.2 μL , 用超纯水补足体积。

反应程序: 预变性 $94 \text{ }^\circ\text{C}$ 4 min, 变性 $94 \text{ }^\circ\text{C}$ 30 s, 退火 $55 \sim 58 \text{ }^\circ\text{C}$ 30 s, 延伸 $72 \text{ }^\circ\text{C}$ 1 min, 变性到延伸一共 35 个循环, 延伸 $72 \text{ }^\circ\text{C}$ 7 min, $4 \text{ }^\circ\text{C}$ 保存。

1.8 种质遗传关系分析

基于 12 份萝卜种质 DNA 扩增产物的电泳图谱, 有带扩增记为 1, 无带记为 0, 使用 NTSYSpc2.0 对统计结果进行聚类分析。

2 结果与分析

2.1 萝卜 EST 资源的 SSR 信息分析

2.1.1 萝卜 EST 的基本信息及 EST-SSR 频率和密度

萝卜 287 349 条 EST 序列经过一系列的预处理后得到 58 105 条无冗余 EST 序列, 全长 38 622.476 kb, 包括 17 850 条重叠群和 40 255 条单一序列。

共检索出含有 SSR 的 Unigene 3 523 条, 其中重叠群 1 363 条, 单一序列 2 160 条。Unigene 占无冗余 EST 序列的 6.06%, 发现 SSR 位点 3 718 个, EST-SSR 出现频率为 6.40%, 萝卜 EST 中平均每 10.39 kb 就出现 1 个 SSR。含有 2 个及 2 个以上 SSR 位点的 Unigene 有 182 条, 完全重复 SSR 有 3 685 个, 不完全重复 SSR 有 33 个。

表 2 表明, 萝卜具有丰富的 EST-SSR 类型, 一、二、三、四、五、六核苷酸重复类型都有出现, 但不同核苷酸重复类型出现的频率相差较多。核苷酸重复类型以二核苷酸重复、三核苷酸重复与六核苷酸重复为主, 三者占总 SSR 比例的 88.17%, 其中三核苷酸重复中的 SSR 数量最多, 占总 SSR 比例的 46.13%, 其次是二核苷酸重复和六核苷酸重复较多, 比例为 22.67% 和 19.37%。

表 2 萝卜 EST-SSR 的数量、出现频率和平均分布距离
Table 2 The number, frequency and average distance of radish EST-SSR

| 重复类型 Type | 数量 Number | 占全部 SSR 的比例/% Proportion in all SSRs | 出现频率/% Frequency | 平均分布距离/kb Average distance |
|----------------------|--------------|---|---------------------|-------------------------------|
| 单核苷酸 Mononucleotide | 200 | 5.38 | 0.34 | 193.11 |
| 二核苷酸 Dinucleotide | 843 | 22.67 | 1.45 | 45.82 |
| 三核苷酸 Trinucleotide | 1715 | 46.13 | 2.95 | 22.52 |
| 四核苷酸 Tetranucleotide | 103 | 2.77 | 0.18 | 374.98 |
| 五核苷酸 Pentanucleotide | 137 | 3.68 | 0.24 | 281.92 |
| 六核苷酸 Hexanucleotide | 720 | 19.37 | 1.24 | 53.64 |
| 总计 Total | 3718 | 100 | 6.40 | 10.39 |

2.1.2 萝卜 EST-SSR 重复基元类型和比例

表 3 显示, 在搜索出的萝卜 EST-SSR 中, 共存在 265 种重复基元类型, 其中单核苷酸 2 种, 二核苷酸 3 种, 三核苷酸 10 种, 四核苷酸 17 种, 五核苷酸 44 种, 六核苷酸的重复基元类型最多, 达到了 189 种。各种不同重复基元中出现最多的是二核苷酸重复类型 AG/CT, 占总 SSR 的比例为 20.71%, 其次是三核苷酸重复类型 AAG/CTT, 比例为 16.54%。

单核苷酸中, A/T 重复基元占绝对优势, 比例达 99%; 二核苷酸中, 以 AG/CT 重复基元占优, 比例同样高达 91.34%, 但 CG/GC 重复基元没有出现; 三核苷酸中, 以 AAG/CTT 重复基元为主, 比例为 35.86%, 此外, CCG/CGG 重复基元出现频率很低 (0.5%); 四核苷酸中, 以 AAAG/CTTT 重复基元为主, 比例为 27.18%; 五核苷酸中, 重复基元主要为 AAGAG/CTCTT, 比例为 18.98%; 六核苷酸中, 以 AAGGAG/CCTCTT 和 AAGAGG/CCTTCT 两种重复基元类型为主, 共占比例 12.78%。

表 3 优势基元重复类型
Table 3 Predominant motifs types

| 重复基元 Motifs types | 重复基元数 Numbers of motifs types | 重复基元 SSR 数 Numbers of motifs types' SSR | 优势重复基元 Predominant motifs | 优势重复基元 SSR 数 Numbers of predominant motifs' SSR | 优势重复基元占本重复基元的比例/% The proportion of predominant motifs | 优势基元中 SSR 出现频率/% Frequency of predominant motifs |
|-------------------------|----------------------------------|--|--------------------------------|--|---|---|
| 单核苷酸 Mononucleotide | 2 | 200 | A/T | 198 | 99 | 11.45 |
| 二核苷酸 Dinucleotide | 3 | 843 | AG/CT | 770 | 91.34 | 44.53 |
| 三核苷酸 Trinucleotide | 10 | 1715 | AAG/CTT | 615 | 35.86 | 35.57 |
| 四核苷酸 Tetranucleotide | 17 | 103 | AAAG/CTTT | 28 | 27.18 | 1.62 |
| 五核苷酸 Pentanucleotide | 44 | 137 | AAGAG/CTCTT | 26 | 18.98 | 1.50 |
| 六核苷酸 Hexanucleotide | 189 | 720 | AAGGAG/CCTCTT AAGAGG/CCTTCT | 47 45 | 6.53 6.25 | 2.72 2.60 |
| 总计 | 265 | 3718 | 6 | 1 729 | | 100 |

2.1.3 萝卜 EST-SSR 重复基元长度

在本研究中, 只有 22.8% 的萝卜 EST-SSR 为高重复次数基元, 即重复基元的重复次数在 10 次以上, 主要以低重复次数基元 (3~10 次重复) 为主。

基元重复长度集中在 18~20 bp 之间的萝卜 EST-SSR 位点有 2 711 个, 占总数的 72.9%; 在 21~30 bp 长度范围的重复基元有 825 个, 比例为 22.2%, 其余 182 个重复基元长度大于 30 bp。

2.2 萝卜 EST-SSRs 引物有效性验证

2.2.1 EST-SSRs 标记在 12 份萝卜种质中的多态性

根据上述信息分析结果, 选择了不同类型和重复次数的 EST-SSRs 设计合成了一批引物。其中具单核苷酸重复的引物 10 对, 二核苷酸重复引物 43 对, 三核苷酸重复引物 95 对, 四核苷酸重复引物 6 对, 五核苷酸重复引物 7 对, 六核苷酸重复引物 20 对, 不完全重复的引物 2 对。

以 12 份萝卜种质的基因组 DNA 为模板对合成的 183 对引物进行多态性检验, 得到有清晰扩增产物的引物 159 对, 扩增有效率为 87%; 64 对 EST-SSR 引物具有多态性, 共扩增出 172 个多态性位点, 平均每个引物产生 2.5 个多态性位点。

由表 4 可以初步看出, 不同核苷酸重复引物的扩增有效率和多态性比例不同, 以二核苷酸重复开发出的引物, 有效扩增比例和多态性比例最低, 而基于单核苷酸重复、三核苷酸重复与五核苷酸重复开发的引物多态性比例较高, 分别为 40%, 41% 和 57%。总体而言, 在基元重复数量最多 (表 3) 的三核苷酸、六核苷酸和二核苷酸中, 基于三核苷酸重复基元开发引物的效率较高, 但是单核苷酸、四核苷酸、五核苷酸和六核苷酸也不可忽视。

表 4 不同核苷酸重复引物的扩增有效率和多态性比例

Table 4 The proportion of effective primers and polymorphic primers in different motifs types of primers

| 重复基元 Motifs types | 合成引物数 Numbers of Synthesized primers | 有效扩增数 Numbers of effective primers | 有效扩增比例/% The proportion of effective primers | 具有多态性的引物数 Numbers of polymorphic primers | 多态性比例/% The proportion of polymorphic primers |
|-------------------------|--|--|--|--|---|
| 单核苷酸 Mononucleotide | 10 | 8 | 80.0 | 4 | 40 |
| 二核苷酸 Dinucleotide | 43 | 28 | 65.1 | 9 | 21 |
| 三核苷酸 Trinucleotide | 95 | 89 | 93.7 | 39 | 41 |
| 四核苷酸 Tetranucleotide | 6 | 5 | 83.3 | 2 | 33 |
| 五核苷酸 Pentanucleotide | 7 | 7 | 100 | 4 | 57 |
| 六核苷酸 Hexanucleotide | 20 | 20 | 100 | 5 | 25 |
| 不完全重复 Others | 2 | 2 | 100 | 1 | 50 |
| 总计 Total | 183 | 159 | — | 64 | — |

不同引物在 12 份萝卜种质中扩增出的多态性位点数量不同, 其揭示的不同种质间的多样性指数 (PIC 值) 的分布范围在 0.141 ~ 0.633 之间。图 1 显示的是部分引物对 12 份萝卜种质基因组 DNA 的扩增结果, 可以看到多态性的显示程度比较丰富。

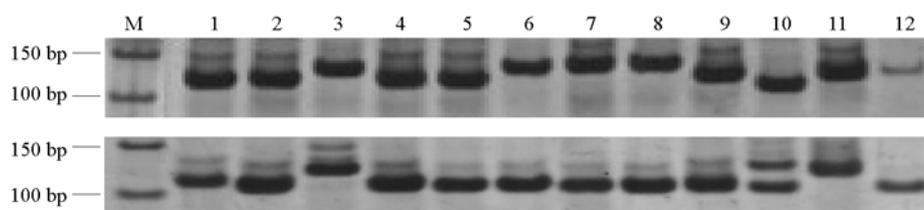


图 1 引物 PR006 (上) 和 PR007 (下) 对 12 份萝卜种质基因组 DNA 的扩增结果

M: 第 1 泳道为 50 bp marker; 1 ~ 12: 萝卜种质。

Fig. 1 The products amplified by primer PR006 and PR007

M: 50 bp marker; 1 to 12: radish germplasm.

2.2.2 EST-SSRs 标记在特定材料中的共显性

用开发的具有多态性的 64 对引物, 以亲本 YS08Q-5 和 YS08Q-9 及其杂种 F_1 的基因组 DNA 为模板进行扩增, 其中 30 对引物在这一特定亲本间仍具有多态性, 在亲本及其 F_1 中呈共显性的标记共有 27 对 (表 5), 图 2 中为部分共显性标记的扩增图。

从多态性和共显性来看, 基于 EST 序列开发萝卜 SSR 引物是可行的和有效的, 这些标记可为萝卜遗传多样性分析、指纹图谱、遗传图谱构建、基因定位以及分子标记辅助育种等所用。

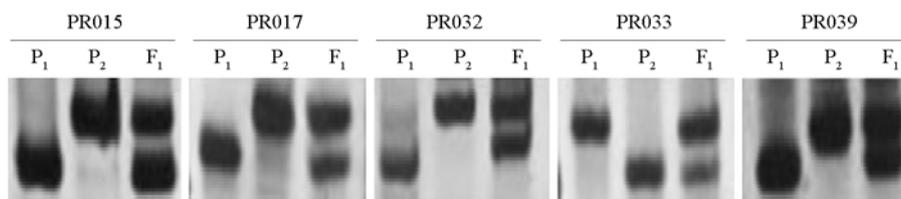
图 2 引物 PR015、PR017、PR032、PR033 和 PR039 在 1 对亲本及其 F_1 中的扩增结果

Fig. 2 The codominant products of primer PR015, PR017, PR032, PR033 and PR039

表 5 27 对共显性引物信息
Table 5 The information of 27 codominant primer pairs

| 引物编号 Number | 核苷酸重复类型 Type | 序列 (5'-3') Sequence(5'-3') | T _m /°C | 扩增产物片段/bp Production's length |
|----------------|---------------------|---|--------------------|----------------------------------|
| PR015 | (AGA)9 | TCAATGTGTTTCTCACACCCAGATTCGGGTAGCTTCAGCAG | 55 | 234 |
| PR017 | (CTC)7 | TCTTTAACAGCCCCAGCAGTCGAGAACCAGCATCAACAGA | 58 | 142 |
| PR032 | (ATC)10 | TGAGAACTTGATTGCTTGCGTCACTGCTTGTCTTTCCAGA | 55 | 226 |
| PR033 | (TTGTG)4 | TTGGTCTCTTGTGCTGTGGCAATCATAAAGTGTGCTTTT | 55 | 194 |
| PR039 | (TCTCA)3 | CCACCCAGTTTGTGGAGAATACTGCCCTGCTGAGACTGTT | 55 | 131 |
| PR052 | (AG)1 | AACACAAACCACGAGAAGGGGAACAAGAAGCAGGGATTGC | 57 | 268 |
| PR053 | (CTC)6 | TCGTTTCATAGGCTCAATCCCTGTTACCAACTCCACGTCCA | 55 | 237 |
| PR054 | (GGA)6 | ATGTGAACGATCCAGCATCAAGTCCACAGCTGGAGATTGC | 55 | 201 |
| PR070 | (T)18 | GTTACTGGGATGGGGATCTTCCTCAAAAAGGGAAATGATGA | 55 | 150 |
| PR071 | (AG)12 | TTCCAGCGTCAAACAGACAGTTGGAGTCCCATCTTTGCT | 55 | 246 |
| PR074 | (GAA)6 attgga(GAT)7 | TCTGTCATTAGTGGCGCATCGTGCCCTGGTTTTTCAATGT | 55 | 211 |
| PR077 | (AGC)8 | ATTAACGTGGGTGATGGGAAAATTTAGGCCCGAGAAAACG | 55 | 198 |
| PR085 | (TCT)8 | TCAAGGGATCTCTTAGCCGACGATTTTGAGGCCATTAGGA | 55 | 249 |
| PR088 | (AGC)6 | TTCTGTGCTTCCACGATCAGTTCTTTCACACGCCATTG | 55 | 193 |
| PR100 | (CTT)7 | CTTTTTAGCGCTTTTGGGTGGCAAATTGAAGCTCCAGAGG | 55 | 259 |
| PR103 | (AG)9 | TCTTCTTTCAGTGCACACCTCGCTTTCCTCCTTCCTT | 55 | 136 |
| PR104 | (TCA)6 | GGGTCCAATGCTAGTTGTGCAATGCCAGAAGGACAAGGT | 55 | 161 |
| PR107 | (T)20 | TTCTTCCAATCTCCTCCCTCTCATCAATCATATCGGCT | 55 | 116 |
| PR109 | (GCTCCA)3 | AACAAGTGGCGAGAAGGAGAGTCTGCTCAGCTGGTACCGT | 55 | 198 |
| PR123 | (GAG)6 | GAACTGAACGAAAACCAATTCCGACTTCAGCTTGGTCT | 55 | 256 |
| PR124 | (CTT)6 | CTTCCCTTCTCCGTCATCAGTCTGAATCGAGAGAGGCCAT | 55 | 128 |
| RS009 | (AGA)6 | AGAACTCGGCAGATAAAGAAGAGGAATACGCAAACC | 56 | 156 |
| RS027 | (TTGTG)4 | TTCTCGTCTTGTGAGAGCAATCATAAAGTGTGCG | 54 | 170 |
| RS030 | (ATG)7 | AACTCCGTCCTGATGCTTTTCTCATATTACCCGATG | 54 | 133 |
| RS032 | (CTG)6 | AAACACCATTTTCGCTCAGTTCAGGGACGGCTTCTAC | 56 | 135 |
| RS040 | (AGA)13 | GAGGGACGATAGGAAGGGATTCTGCTCCGCCAACTC | 57 | 192 |
| RS046 | (CTC)6 | GAGGCGGAAAGCCCAATACTTGAACCTCGCCGTGT | 57 | 210 |

2.3 已开发的 EST-SSR 标记在萝卜种质遗传关系分析中的应用

以 64 对多态性引物对 12 份萝卜种质进行聚类分类 (图 3), 在相似系数 0.5260 处可将其分为 3 大类。KB10Q-1, KB10Q-3, L10Q-113, KB10Q-10, L10Q-184, L10Q-433, Z10Q-xlm 和 KB10Q-5 聚为一类, 其中来源于北京的 L10Q-184, L10Q-433 和 Z10Q-xlm 亲缘关系相对较近, 均为圆形红皮和红心萝卜; 来源于韩国的 KB10Q-1 和中国的 KB10Q-3, L10Q-113 也显出相对较近的关系, 均为白皮和绿皮长圆柱大根萝卜; KB10Q-10 和 KB10Q-5 也为大根萝卜。PP10Q-3, PP10Q-25, L10-1-1 被聚为一类, 其关系很远: PP10Q-23 生育期较短, 不耐抽薹, 耐热; PP10Q-25 来自泰国, 短生育期, 长莢; L10-1-1 特耐抽薹, 大根。以上这两类均为萝卜的栽培种。PP10Q-29 为野生萝卜, 单独聚为一类。

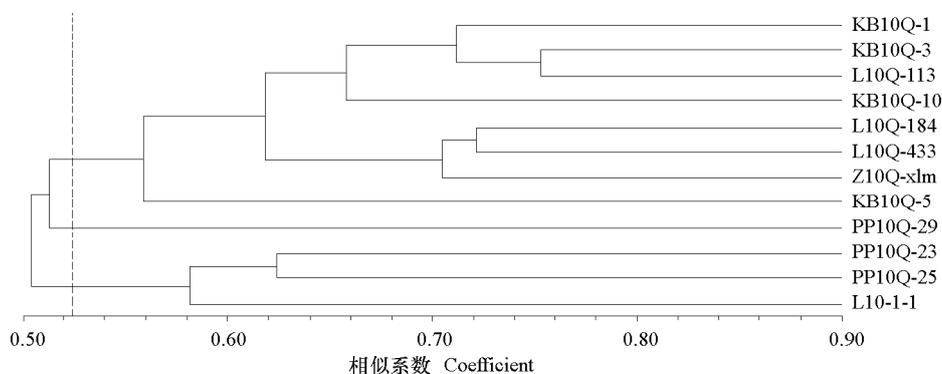


图 3 12 份萝卜种质的聚类图
Fig. 3 Dendrogram of 12 radish germplasms

3 讨论

本研究中基于 287 349 条萝卜 EST 的预处理和 SSR 位点搜索, 得到 3 718 个 SSR, 出现频率 6.06%, 平均分布距离为 10.39 kb。在甘蔗(闫学兵 等, 2010), 百合(杨素丽 等, 2008), 棉花(Qureshi et al., 2004), 玉米(赵美琼 等, 2008 年), 欧洲白桦(王艳敏 等, 2008) 和油菜(李小白 等, 2007) 中的 EST-SSR 位点的检出比例分别为 4.7%、5.98%、6.00%、7.25%、12.01% 和 15.58%。可见萝卜 EST-SSR 的丰富度处于较高水平。不同作物检出 SSR 比例的范围值相差较大, 一方面可能与不同物种 EST-SSR 的分布有关, 另一方面可能与搜索的核苷酸重复类型的长度标准设定不同有关。现行的 SSR 搜索标准中有 12 bp、15 bp 以及 18 bp, 3 种重复基元总长度。不同的研究者采用不同的搜索标准, 导致结果不同, 大多数的研究人员采用了 15 bp 或 18 bp 的标准。在本研究中, 采用 15 bp 标准可搜索到 7 944 个 SSR 位点, 而采用 18 bp 的标准, 搜索的 SSR 位点数减少到了 3 718 个。根据 Schlötterer 等(2000) 的研究, SSR 的变异频率与基元重复数呈正相关, 即重复基元的重复次数越多则 SSR 的多态性潜能越大。在本研究中采用了 18 bp 的标准进行 SSR 位点的筛选。此外, 是否进行序列的预处理, 也会导致 EST-SSR 检出的频率不同。对下载的 EST 序列经过了预处理, 在去冗余、聚类 and 拼接后, 得到 58 105 条单一序列, 尔后搜索得到了 3 718 个 SSR 位点。若不进行预处理, 采用相同的搜索标准, 竟能搜索到 11 055 个位点。这说明, 如果不进行预处理的拼接工作, 会产生大量的 EST-SSR 位点重复开发的弊端, 孔秋生(2006) 的研究也证实了预处理的重要性。

据报道, 大多数植物的 EST-SSR 序列以二、三核苷酸重复基元类型为主, 不同作物中的主导重复基元类型不大相同。茶树(金基强 等, 2006) 二、三核苷酸重复为主的 EST-SSR 发生频率分别为 11.5% 和 3.08%。白菜(忻雅 等, 2006) 分别为 4.36% 和 4.19%, 油菜(李小白 等, 2007) 分别为 7.24% 和 6.63%, 木薯(彭丁文 等, 2008) 分别为 4.61% 和 4.07%。大麦(Thiel et al., 2003), 葡萄(蔡斌 等, 2009 年), 拟南芥、小麦、水稻、玉米、大豆(Cardle et al., 2000) 等均均以三核苷酸重复为主。而萝卜的 EST-SSR 序列主要为二、三、六核苷酸重复基元类型, 其发生频率分别为 1.45%、2.95% 和 1.24%, 均低于其他作物。这既可能与物种基因组或转录组序列的特异性有关, 也可能是由于作物间已测序的 EST 数量不同且不同研究者搜索 SSR 的标准不同造成的。在这些作物中, 二核苷酸重复基元中占有绝对优势的均为 AG/GT, 三核苷酸重复基元中则以 AAG/CCT 为主, 这与本研究中萝卜的二核苷酸重复与三核苷酸重复的优势基元类型相同。另外, 萝卜中出现的大量六核苷酸重复基元中以 AAGGAG/CCTCTT 和 AAGAGG/CCTTCT 两种为主, 占六核苷酸重复基元总比例的 12.78%。在以往研究的大多作物中, 六核苷酸重复基元类型均不常见, 仅在甘薯(黄立飞 等, 2009) 和玉米(赵美琼 等, 2008) 中见到报道, 且与萝卜的重复基元类型不同。这些可能与物种本身的基因组或转录组序列差异性有关。

SSR 位点的多态性是由重复基元数量和基元碱基数不同所产生的简单序列长度多态性 (simple sequence length polymorphism, SSLP) 和随机扩增微卫星多态性 (random amplified microsatellites polymorphism, RAMP) 共同决定的。Temnykh 等(2001) 认为, 当 SSR 长度在 20 bp 及 20 bp 以上时, 在不同品种间显示出较高的多态性, 长度低于 20 bp, 则多态性就会降低。此外, Dreisigacker 等(2007) 的研究又发现高重复次数基元的微卫星标记多态性比低重复次数基元低。在本研究中, 只有 22.8% 的 EST-SSR 为高重复次数基元 (重复次数在 10 次以上), 以低重复次数基元 (3~10 次重复) 为主。共有 1 776 个萝卜 EST-SSR 序列长度都在 20 bp 及 20 bp 以上, 因此, 据前者的观点, 本研究中的 48% 的 EST-SSR 位点具有良好的多态性潜能; 据后者的观点, 本研究中低重复次数基元占优势 (77.2%), 具有很高的多态性潜能。

在萝卜 EST-SSR 标记的开发方面, Wang 等 (2007) 仅开发了 8 对 EST-SSR 引物。Nakatsuji 等 (2011) 基于部分基因组和转录组序列, 设计了 417 对萝卜 SSR 引物, 256 对引物有扩增产物, 扩增有效率 61%, 其中 130 对引物在两份自交系间有多态性, 比例 50%。Shirasawa 等 (2011) 基于自建的栽培萝卜 cDNA 文库测序获得的 26 606 条 ESTs, 分析仅获得 10 381 无冗余 ESTs, 通过鉴定 ≥ 15 bp 的二、三和四核苷酸基元序列 (未涉及其他基元类型), 搜索出 3 800 个 SSR 位点, 开发出 642 对多态性 (16.9%) EST-SSR 引物用于构建连锁图谱, 82% 的标记呈共显性。本研究利用 NCBI 中已公开的包括萝卜栽培和野生种的 287 349 条萝卜 EST 序列, 得到无冗余 EST 序列 58 105 条, 采用 18 bp 的标准, 搜索的 SSR 位点, 发现 SSR 位点 3 718 个, 一至六核苷酸重复类型都存在, 主要以二核苷酸重复、三核苷酸重复与六核苷酸重复为主。这使我们对萝卜 EST-SSR 出现的频率、特点、分布有了较前人更详尽的了解。本研究按照不同的重复基元类型, 合成了 183 对引物, 使用 12 份不同的萝卜种质对其有效性进行了验证, 得到有清晰扩增产物的标记 159 对, 扩增有效率 (87%) 和多态性比例 (40%) 均较高。而且, 由于 EST 序列的来源和基于的重复基元不同, 所获的有效引物与前人获得的有效引物不同。另外, 还发现在基元重复数量最多的三核苷酸、六核苷酸和二核苷酸中, 基于三核苷酸重复基元开发引物的有效扩增率和多态率均较高, 但是单核苷酸、四核苷酸、五核苷酸和六核苷酸也显示较高的开发价值, 值得进一步开发验证。使用开发的多态性引物不仅能有效鉴定 12 份萝卜种质, 而且在特定组合中的共显性标记比例达 90%。故本研究基于萝卜 EST 资源整合进行的 SSR 信息分析和标记开发是有效的, 可为萝卜遗传多样性分析、指纹图谱、遗传图谱构建、基因定位以及分子标记辅助育种提供更有力的技术支持。

References

- Cai Bin, Li Cheng-hui, Yao Quan-hong, Zhou Jun, Tao Jian-min, Zhang Zhen. 2009. Analysis of SSRs in grape genome and development of SSR database. *Journal of Nanjing Agricultural University*, 32 (4): 28 - 32. (in Chinese)
- 蔡 斌, 李成慧, 姚泉洪, 周 军, 陶建敏, 章 镇. 2009. 葡萄全基因组 SSR 分析和数据库构建. *南京农业大学学报*, 32 (4): 28 - 32.
- Cardle L, Ramsay L, Milbourne D, Milbourne D, Macaulay M, Marshall D, Waugh R. 2000. Computational and experimental characterization of physically clustered simple sequence repeats in plants. *Genetics*, 156: 847 - 854.
- Chen Xiang-yan, Li Wei, Dai Hai-ying, Zhang Li-feng. 2009. Analysis of SSR information in EST resource of soybean (*Glycinemax*). *Soybean Science*, 28 (3): 394 - 399. (in Chinese)
- 陈相艳, 李 伟, 戴海英, 张礼凤. 2009. 大豆 EST 资源的 SSR 信息分析. *大豆科学*, 28 (3): 394 - 399.
- Dreisigacker S, Taba S, Bedoya C, Franco J, Charcosset A, Mir C, Zhang Shi-huang, Chaun Xiao X, Prasanna B M, Hearne S, Muthamia Z, Yunus M, Jampatong C, Mahn Cuong B, Warburton M. 2007. Molecular characterization of maize landraces and various subspecies of teosinte from Mexico and Central America. *Proceedings of Crop Science Society of America 2007 Annual Meetings. International maize and wheat improvement center*.
- Hu Jian-bin, Li Jian-wu. 2008. Information of EST - SSR loci in cucumber. *Acta Agriculturae Jiangxi*, 20 (11): 58 - 60. (in Chinese)
- 胡建斌, 李建吾. 2008. 黄瓜 EST-SSR 位点信息. *江西农业学报*, 20 (11): 58 - 60.
- Huang Li-fei, Fang Bo-ping, Chen Jing-yi, Zhang Xiong-jian, Luo Zhong-xia. 2009. Analysis of SSR information in EST resource of sweet potato [*Ipomoea batatas* (L.) Lam]. *Plant Physiology Communications*, 45 (1): 23 - 27. (in Chinese)
- 黄立飞, 房伯平, 陈景益, 张雄坚, 罗忠霞. 2009. 甘薯 EST 资源的 SSR 信息分析. *植物生理学通讯*, 45 (1): 23 - 27.
- Jin Ji-qiang, Cui Hai-rui, Chen Wen-yue, Lu Mei-zhen, Yao Yan-ling, Xin Ya, Gong Xiao-chun. 2006. Data mining for sssrs in ests and development of est-ssr marker in tea plant (*Camellia sinensis*). *Journal of Tea Science*, 26 (1): 17 - 23. (in Chinese)
- 金基强, 崔海瑞, 陈文岳, 卢美贞, 姚艳玲, 忻 雅, 龚晓春. 2006. 茶树 EST-SSR 的信息分析与标记建立. *茶叶科学*, 26 (1): 17 - 23.
- Ju-Kyung Yu, Trevor M. Dake, Sukhwinder Singh, Benschel D, Wanlong L, Gill B, Mark E. Sorrells. 2004. Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat. *Genome*, 47: 805 - 818.
- Kong Qiu-sheng. 2006. Characterization, development and application of EST-SSR markers in cucumis genus based on public sequence database [Ph.D. Dissertation]. Wuhan: Huazhong Agricultural University. (in Chinese)

- 孔秋生. 2006. 基于公共序列数据库的 Cucumis 属 EST-SSR 标记的鉴定、开发和利用[博士论文]. 武汉: 华中农业大学.
- LI Li, Zheng Xiao-ying. 2009. The development of multiplex EST-SSR markers to identification chinese cabbage [*Brassica campestris* L. *chinensis* (L.) makino and *Brassica campestris* L. *pekinensis* (Lour.) Olsson] Cultivars. *Acta Horticulturae Sinica*, 36 (11) : 1627 - 1634. (in chinese)
- 李 丽, 郑晓鹰. 2009. 用于白菜和大白菜品种鉴定的 EST-SSR 复合标记的建立. *园艺学报*, 36 (11) : 1627 - 1634
- Li Xiao-bai, Zhang Ming-long, Cui Hai-rui. 2007. Data mining for SSRs in ESTs and development of EST-SSR marker in oilseed rape. *Journal of Molecular Cell Biology*, 40 (2): 137 - 144. (in Chinese)
- 李小白, 张明龙, 崔海瑞. 2007. 油菜 EST-SSR 标记的建立. *分子细胞生物学报*, 40 (2): 137 - 144.
- Nakatsuji R, Hashida T, Matsumoto N, Tsuru M, Kubo N, Hirai M. 2011. Development of genomic and EST-SSR markers in radish (*Raphanus sativus* L.). *Breeding Science*, 61: 413 - 419.
- Peng Ding-wen, Zheng Liu-cheng, Zhu Hong-bo. 2008. Analysis of SSR information in EST resource of cassava. *Chinese Agricultural Science Bulletin*, 24 (2): 433 - 436. (in Chinese)
- 彭丁文, 郑柳城, 朱宏波. 2008. 木薯 EST 资源的 SSR 信息分析. *中国农学通报*, 24 (2): 433 - 436.
- Qureshi S N, Saha S, Kantety R V, Jenkins J N. 2004. EST-SSR: A new class of genetic markers in cotton. *Journal of Cotton Science*, 8 (2): 112 - 123.
- Schlötterer C. 2000. Evolutionary dynamics of microsatellite DNA. *Chromosoma*, 109: 365 - 371.
- Shirasawa K, Oyama M, Hirakawa H, Sato S, Tabata S, Fujioka T, Kimizuka-Takagi C, Sasamoto S, Watanabe A, Kato M, Kishida Y, Kohara M, Takashi C, Tsuruoka H, Wada T, Sakai T, Isobe S. 2011. An EST-SSR linkage map of raphanus sativus and comparative genomics of the brassicaceae. *DNA Research*, 6: 1 - 12.
- Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Res*, 11: 1441 - 1452.
- Thiel T, Michalek W, Varshney R, Graner A. 2003. Exploiting EST database for the development and characterization of gendered SSR markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet*, 106: 411 - 422.
- Wang Fang, Wang Xiao-feng, Chen Xun, Xiao Ying-jie, Li Hai-tao, Zhang Shun-chang, Xu Jin-song, Fu Jie, Huang Lei, Liu Chao, Wu Jiang-sheng, Liu Ke-de. 2011. Abundance, marker development and genetic mapping of microsatellites unigenes in *Brassica napus*. *Molecular Breeding*, DOI: 10.1007/s11032-011-9658-7.
- Wang Ning, Hu Jing-jie, Ohsawa R, Ohta M, Fujimura T. 2007. Identification and characterization of microsatellite markers derived from expressed sequence tags (ESTs) of radish (*Raphanus sativus* L.). *Molecular Ecology*, 7 (3): 503 - 506.
- Wang Yan-min, Wei Zhi-gang, Yang Chuan-ping. 2008. Data mining for SSRs in ESTs and EST-SSR marker development in *Betula platyphylla*. *Scientia Silvae Sinicae*, 44 (2): 78 - 84. (in Chinese)
- 王艳敏, 魏志刚, 杨传平. 2008. 白桦 EST-SSR 信息分析与标记的开发. *林业科学*, 44 (2): 78 - 84.
- Wei Xin-yan, Liu Li-feng, Cui Shun-li, Chen Huan-ying, Zhang Jing-jing. 2011. Development of EST-SSR markers in peanut (*Arachis hypogaea* L.). *Front. Agric. China*, 5 (3): 268 - 273.
- Xin Ya, Cui Hai-rui, Zhang Ming-long, Lim Yong-pyo, Choi Su-ryun. 2005. Development of EST (expressed sequence tags) marker in chinese cabbage and its transferability to rapeseed. *Hereditas (Beijing)*, 27 (3): 410 - 416 (in Chinese)
- 忻 雅, 崔海瑞, 张明龙, 林容杓, 崔水莲. 2005. 白菜的 EST 标记及其对油菜的通用性. *遗传*, 27 (3): 410 - 416
- Xin Ya, Cui Hai-rui, Lu Mei-zhen, Yao Yan-ling, Jin Ji-qiang, Lim Yong-Pyo, Choi Su-Ryun. 2006. Data mining for SSRs in ESTs and EST-SSR marker development in chinese cabbage. *Acta Horticulturae Sinica*, 33 (3): 549 - 554. (in Chinese)
- 忻 雅, 崔海瑞, 卢美贞, 姚艳玲, 金基强, 林容杓, 崔水莲. 2006. 白菜 EST-SSR 信息分析与标记的建立. *园艺学报*, 33 (3): 549 - 554.
- Yan Xue-bing, Que You-xiong, Xu Li-ping, Guo Jin-long, Chen Ru-kai, Pan Yong-bao. 2010. Analysis of SSR information in EST resource of sugarcane. *Chinese journal of tropical crops*, 31 (9): 1497 - 1500. (in Chinese)
- 闫学兵, 阙友雄, 许莉萍, 郭晋隆, 陈如凯, 潘永保. 2010. 甘蔗 EST 序列的 SSR 信息分析. *热带作物学报*, 31 (9): 1497 - 1500.
- Yang Su-li, Ming Jun, Liu Chun, Mu Ding, Li Ming-yang. 2008. Data mining for simple sequence repeats marker development in expressed sequence tags from *Lilium* L. *Acta Horticulturae Sinica*, 2008, 35 (7): 1069 - 1074. (in Chinese)
- 杨素丽, 明 军, 刘 春, 穆 鼎, 李名扬. 2008. 基于 EST 信息的百合 SSR 标记的建立. *园艺学报*, 35 (7): 1069 - 1074.
- Zhao Mei-qiong, Song Yan, Fan Jing-hua, Ji Yun-zuo, Luo Qiong, Zhu You-yong, Li Cheng-yun. 2008. Frequency and distribution of different simple sequence repeats (SSRs) in the maize ESTs. *Journal of Yunnan Agricultural University*, 23 (5): 585 - 589. (in Chinese)
- 赵美琼, 宋 燕, 范静华, 纪韵祚, 罗 琼, 朱有勇, 李成云. 2008. 玉米 EST 序列中微卫星的频率和分布. *云南大学学报*, 23 (5): 585 - 589.