

## 刺叶苏铁 EST 序列中 SSR 信息分析

李加敏<sup>1,2</sup> 肖龙骞<sup>2</sup> 李青青<sup>1\*</sup><sup>(1)</sup> 云南师范大学生命科学学院, 云南昆明 650092; <sup>(2)</sup> 中国科学院西双版纳热带植物园热带森林生态重点实验室)

**摘要** 基于 NCBI 数据库中刺叶苏铁的 21 997 条 EST 序列, 统计分析了其 EST-SSR 的组成与分布特点。经过剔除冗余和低质量序列后, 得到长度为 7 926 783 bp 的无冗余 EST 序列 13 640 条。在这些序列中共搜索出了 875 条 EST 序列含有 1 176 个 SSR, 出现频率为 8.6%。这些 SSR 的主要重复基序有 A/T, C/G, AC/GT, AG/CT, AT/AT, CG/CG, AAC/GTT, AAG/CTT, AAT/ATT, ACC/GGT, ACG/CGT, AGC/CTG, AGG/CCT, ATC/ATG, AAAT/ATTT, AAGG/CCCT, AATT/AAATT, ACAT/ATGT 和 AAACCC/GGGTTT。一、二、三核苷酸重复类型是主体, 三者共占总数的 99.1%。A/T, AT/AT, AG/CT, AAG/CTT 和 AAT/ATT 分别是其优势重复基元, 分别占总数的 41.7%、11.5%、11.9%、2.6% 和 2.5%。该研究为刺叶苏铁 EST-SSR 标记的开发与应用奠定了基础。

**关键词** 刺叶苏铁; EST; SSR

**中图分类号** S791.11; Q524 **文献标识码** A **文章编号** 1007-5739(2012)08-0044-02

Analysis of SSR Information in EST Resource of *Cycas rumphii* Miq.LI Jia-min<sup>1,2</sup> XIAO Long-qian<sup>2</sup> LI Qing-qing<sup>1\*</sup>

<sup>(1)</sup> School of Life Sciences, Yunnan Normal University, Kunming Yunnan 650092; <sup>(2)</sup> Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences)

**Abstract** This study analyzed the distribution pattern and compared the characters of EST-SSR, which were derived from a total of 21 997 ESTs of *Cycas rumphii* Miq. downloaded from the NCBI database. After removing redundant and poor quality sequences, 13 640 non-redundant ESTs with 7 926 783 bp length were obtained, and 875 EST sequences, which accounted for 8.6% the total number of non-redundant ESTs, were detected to contain 1 176 SSRs. In these SSRs, the major repeat motifs included A/T, C/G, AC/GT, AG/CT, AT/AT, CG/CG, AAC/GTT, AAG/CTT, AAT/ATT, ACC/GGT, ACG/CGT, AGC/CTG, AGG/CCT, ATC/ATG, AAAT/ATTT, AAGG/CCCT, AATT/AAATT, ACAT/ATGT and AAACCC/GGGTTT, and the major repeat types were mononucleotide, dinucleotide and trinucleotide, which were accounted for 99.1% of the total number of acquired SSRs. Among these repeats, A/T, AT/AT, AG/CT, AAG/CTT and AAT/ATT were the most frequent motifs, accounting for 41.7%, 11.5%, 11.9%, 2.6% and 2.5%, respectively. The present study laid the foundation for the development and further use of EST-SSR markers in *Cycas rumphii* Miq..

**Key words** *Cycas rumphii* Miq.; EST; SSR

简单重复序列 (Simple Sequence Repeat, SSR) 是由核苷酸重复基元串联组成的高度重复序列, 广泛存在于真核生物基因组的编码区和非编码区。分子标记 SSR 具有共显性、多态性高和重复性好的特点<sup>[1]</sup>, 广泛应用于物种鉴定、遗传连锁图谱的构建和种质资源遗传多样性等研究中。但 SSR 没有物种之间广泛通用的引物, 而基因组 SSR 引物的开发需要构建 DNA 文库, SSR 克隆筛选和测序以弄清楚 SSR 两端的 DNA 序列信息。这一过程费时耗力, 成本也很高。

表达序列标签 (expressed sequence tags, ESTs) 是在生物体特定时空表达基因的一段 cDNA 序列总和, 其长度一般为 150~500 bp。EST 中 7%~10% 的序列含有可用于建立标记的 SSR<sup>[2]</sup>。因此, 日益快速增加的 EST 信息成为开发 SSR 的重要资源。与基因组 SSR 相比, EST-SSR 标记由于源自基因的转录序列, 可直接用于基因作图和基因发掘, 同时还具有信息量大、近缘物种之间通用性好、开发简捷等优点, 是目前一种主要的功能性分子标记。

苏铁是最古老、最原始的现存种子植物。其历经了 2 亿多年的沧桑演化, 蕴涵着丰富的遗传变异信息。这对研究古生物学、古气候学和古地理学以及种子植物的起源与早期演化都具有重要价值<sup>[3-4]</sup>。同时, 苏铁植物是一种人们十分喜爱的观赏园林植物, 其树形优美, 羽叶宛如孔雀开屏, 且四季长青。近几十年来, 因为生境的丧失和过度采挖, 野生苏铁植物急剧减少, 日益变得稀有、濒危, 甚至面临灭绝的危

险。因此, 世界保护联盟 (IUCN) 的《世界野生植物濒危物种红皮书》把苏铁誉为“植物界的大熊猫”, 将其所有物种列为重点保护对象。刺叶苏铁 (*Cycas rumphii* Miq.) 自然分布于巴布亚新几内亚、印度尼西亚与斐济等局部地区, 但在我国华南地区亦能见到这个珍稀物种, 或为植物园的迁地保护对象, 或为家庭庭院的观赏植物。可能是因为引入的历史比较久远, 我国华南地区的人们习惯地称之为“华南苏铁”。

该文对美国国家生物技术信息中心 (NCBI) 数据库中刺叶苏铁的 21 997 条 EST-SSR 信息 (截至 2012 年 2 月) 进行发掘和比较, 对其组成和分布特点进行统计和分析, 并对其多态性潜能进行前瞻性评价, 为刺叶苏铁 EST-SSR 分子标记的开发与应用, 及开展相关分子遗传学研究奠定了基础。

## 1 材料与方法

### 1.1 刺叶苏铁 EST 序列来源与前处理

首先, 以 FASTA 格式从 NCBI 下载刺叶苏铁的 EST 序列。然后利用 EST-trimmer 软件 (<http://www.pgsc.ipk-gatherleben.de/misa/download/est-trimmer.pl>) 先除去 EST 序列 5' 端或 3' 端前 50 bp 内重复次数至少为 5 次的 poly T 或 poly A; 然后筛去长度小于 100 bp 的 EST 序列; 再对长度大于 700 bp 的 EST 序列进行修剪, 仅保留 5' 端前 700 bp 的序列, 以减少测序产生的错误。再利用 CD-HIT 程序 (序列一致度阈值设立为 90%) 除去 EST 中的冗余序列。

### 1.2 刺叶苏铁 EST-SSR 的发掘

利用 MISA (MIcroSATellite) 软件从经过预处理后的 EST 序列中搜索 SSR。重复基元长度设置为 1~6 bp。筛选标准设置为: 一核苷酸重复的次数 ≥ 10 次, 二至六核苷酸重复的次

**作者简介** 李加敏 (1986-), 女, 云南昭通人, 在读硕士研究生。研究方向: 植物分子生物学。

**\* 通讯作者**

**收稿日期** 2012-03-07

数≥5 次,同时筛选并记录间隔区≤100 bp 的复合 SSR。

1.3 刺叶苏铁 EST-SSR 重复基元的分析统计

刺叶苏铁 EST-SSR 类型、数目和出现频率等采用 Excel 软件的排序、筛选、数据透视表等功能进行统计分析。

2 结果与分析

2.1 刺叶苏铁 EST 基本信息

从 GenBank 中下载到刺叶苏铁 EST 序列共 21 997 条,其中来自于胚珠的 11 562 条(DR062281-DR063343;EX 808571-EX809724;EX920821-EX930168)、来自于新叶片 4 951 条(CB089333-CB094283)、来自于胚珠着生的大孢子叶 5 484 条(CB088332-CB089332;DR061234-DR062280;EX91 7384-EX920820)。经过预处理处理后,获得 13 640 条非冗余且质量高的 EST 序列。这些序列的总长度为约 7 926 783 bp,EST 序列平均长度为 581.1 bp。

2.2 刺叶苏铁 EST 序列中不同 SSR 出现的频率

在非冗余的 13 640 条 EST 序列中,搜索发现共有 875 条 EST 序列存在 SSR,SSR 的发生频率(即含有 SSR 的 EST 数目与 EST 数目的比值)为 6.4%。其中,只含有 1 个 SSR 的 EST 序列有 731 条,含有 2 个或 2 个以上 SSR 的有 144 条。根据预先设定的标准,发现有 257 条 EST 序列存在复合 SSR。总共检测出 1 176 个 SSR。出现频率(即 EST-SSR 个数与 EST 数目之比值)为 8.6%,平均每 11.6 条 EST 序列中存在 1 个 SSR 序列(表 1)。

刺叶苏铁 EST-SSR 的重复类型比较丰富。除了五碱基重复基元的 SSR 没有检测到之外,其余的 5 种碱基重复基元均有不同类型重复,而各种类型出现的频率差异很大(表

表 1 非冗余刺叶苏铁 EST 中各种 SSR 的分布

| 重复类型   | SSR 数目<br>个 | 各类型的比<br>例//% | 频率//% | 平均距离<br>kb |
|--------|-------------|---------------|-------|------------|
| 一核苷酸重复 | 668         | 56.8          | 4.9   | 11.9       |
| 二核苷酸重复 | 360         | 30.6          | 2.6   | 22.0       |
| 三核苷酸重复 | 138         | 11.7          | 1.0   | 57.4       |
| 四核苷酸重  | 9           | 0.8           | 0.1   | 880.8      |
| 六核苷酸重复 | 1           | 0.1           | 0.0   | 7 926.8    |
| 总计     | 1 176       | 100.0         | 8.6   | 6.7        |

1)。重复基元主要集中在一至三核苷酸重复上,占总 EST-SSR 的 99.1%,其中 1 个核苷酸重复基元的 SSR 数量最多,共有 668 个,占 SSR 总数的 56.8%;其次为 2 个核苷酸重复基元的 SSR,有 360 个,占 SSR 总数的 30.6%;3 个核苷酸重复基元的 SSR 有 138 个,占 SSR 总数的 11.7%。而 4 个核苷酸重复基元的 SSR 只检测到为 9 个,占总数的 0.8%;仅发现 1 个由 6 个核苷酸构成重复基元的 SSR,仅占总数的 0.1%。所占比例随着重复基元核苷酸数目的变化而变化。

2.3 不同重复基元 SSR 的分布频率

刺叶苏铁 EST-SSR 中共有 19 种不同的重复基元(表 2)。其中,1 个核苷酸的有 2 种,即 A/T、C/G,所占比例分别为 73.4%和 26.6%(图 1A);2 个核苷酸的有 4 种,即 AC/GT、AG/CT、AT/AT 和 CG/CG,所占比例依次为 AG/CT 重复基元出现频率(38.9%)>AT/AT(37.5%)>AC/GT(23.3%)>CG/CG(0.3%)(图 1B);3 个核苷酸的有 8 种,其中有 5 个主要的重复基元,即 AAG/CTT、AAT/ATT、AGC/CTG、ATC/ATG 和 AGG/CCT,它们所组成的 SSR 占 3 个核苷酸 SSR 的 88%,其余 3 个重复基元分别是 AAC/GTT、ACC/GGT 和 ACG/CGT,它们所组成的 SSR 仅占 3 个核苷酸 SSR 的 12%(图 1C);而 4

表 2 非冗余刺叶苏铁 EST 中各种 SSR 的分布 (个)

| 重复单元          | 重复次数 |     |     |     |     |      |      |      |      |      |      |       | 合计  |
|---------------|------|-----|-----|-----|-----|------|------|------|------|------|------|-------|-----|
|               | 5 次  | 6 次 | 7 次 | 8 次 | 9 次 | 10 次 | 11 次 | 12 次 | 13 次 | 14 次 | 15 次 | ≥16 次 |     |
| A/T           | -    | -   | -   | -   | -   | 165  | 99   | 51   | 39   | 38   | 31   | 67    | 490 |
| C/G           | -    | -   | -   | -   | -   | 55   | 35   | 27   | 19   | 12   | 10   | 20    | 178 |
| AC/GT         | 44   | 11  | 13  | 2   | 5   | 4    | 1    | 1    | 0    | 2    | 0    | 1     | 84  |
| AG/CT         | 70   | 22  | 7   | 9   | 9   | 7    | 4    | 4    | 0    | 1    | 1    | 6     | 140 |
| AT/AT         | 48   | 26  | 14  | 8   | 8   | 12   | 9    | 2    | 1    | 1    | 0    | 6     | 135 |
| CG/CG         | 1    | 0   | 0   | 0   | 0   | 0    | 0    | 0    | 0    | 0    | 0    | 0     | 1   |
| AAC/GTT       | 6    | 0   | 0   | 0   | 0   | 0    | 0    | 0    | 0    | 0    | 0    | 0     | 6   |
| AAG/CTT       | 18   | 7   | 3   | 0   | 2   | 0    | 0    | 0    | 0    | 0    | 0    | 0     | 30  |
| AAT/ATT       | 14   | 13  | 2   | 0   | 0   | 0    | 0    | 0    | 0    | 0    | 0    | 0     | 29  |
| ACC/GGT       | 2    | 5   | 0   | 0   | 0   | 0    | 0    | 0    | 0    | 0    | 0    | 0     | 7   |
| ACG/CGT       | 2    | 2   | 0   | 0   | 0   | 0    | 0    | 0    | 0    | 0    | 0    | 0     | 4   |
| AGC/CTG       | 18   | 5   | 2   | 0   | 0   | 0    | 0    | 0    | 0    | 0    | 0    | 0     | 25  |
| AGG/CCT       | 17   | 3   | 0   | 1   | 0   | 0    | 0    | 0    | 0    | 0    | 0    | 0     | 21  |
| ATC/ATG       | 13   | 3   | 0   | 0   | 0   | 0    | 0    | 0    | 0    | 0    | 0    | 0     | 16  |
| AAAT/ATTT     | 0    | 3   | 0   | 0   | 0   | 0    | 0    | 0    | 0    | 0    | 0    | 0     | 3   |
| AAGG/CCTT     | 0    | 1   | 0   | 0   | 0   | 0    | 0    | 0    | 0    | 0    | 0    | 0     | 1   |
| AATT/AATT     | 0    | 2   | 0   | 0   | 0   | 0    | 0    | 0    | 0    | 0    | 0    | 0     | 2   |
| ACAT/ATGT     | 0    | 2   | 0   | 0   | 0   | 0    | 0    | 0    | 0    | 0    | 0    | 0     | 3   |
| AAACCC/GGGTTT | 1    | 0   | 0   | 0   | 0   | 0    | 0    | 0    | 0    | 0    | 0    | 0     | 1   |

个和 6 个核苷酸重复基元分别只有 4 种和 1 种。总体而言,刺叶苏铁中,重复基元 A/T 组成的 EST-SSR 所占比例最高(41.7%),而 4 个和 6 个碱基重复基元的 EST-SSR 种类和数量都相对较少。

3 结论与讨论

通过对刺叶苏铁 EST 序列中其 1~6 个核苷酸重复基元 SSR 的分布、组成和频率进行分析,结果表明 EST-SSR 的出现频率为 8.6%,平均每 6.7 kb 核苷酸分布有 1 个 SSR 位

点。刺叶苏铁一、二、三核苷酸重复基元的 EST-SSR 分别占 SSR 总数的 56.8%、30.6%和 11.7%,一核苷酸重复基元主要是(A)<sub>n</sub>,二、三核苷酸基元类型中以(AT)<sub>n</sub>、(AG)<sub>n</sub>、(AAG)<sub>n</sub>与(AAT)<sub>n</sub>占了绝对优势,五核苷酸基元类型没有出现,四、六核苷酸基元类型虽有出现,但类型相对较少、出现频率也很低。这表明同大多数植物一样,刺叶苏铁 EST-SSR 的分布有比较明显的偏倚性。1 176 个刺叶苏铁 EST-SSR 的 (下转第 48 页)

布的效率;收到信息后通过各自渠道快速传递出去。但是在气象预警信号发布时,由于用户量大,传输不够及时,没有真正满足“最后一公里”的需求,而且决策领导职责的变动也较为频繁,影响了信息库的更新。因此,要随时掌握人员变动信息,对“菏泽灾害性天气预警信息库”进行及时的更新,并做到及时发布、及时传输,要求气象信息员第一时间向村民通报,让农民尽早收到预报信息,提前做好防灾减灾准备<sup>[4]</sup>。

### 2.3 短信定制问题

兴农短信平台升级后,移动公司限制了权限,气象部门不能随意添加取消手机号码,存在用户想用但定制不成功的问题,中老年人由于各种原因对定制的方法不了解;同时,由于移动、电信、联通等运营商自身管理的问题,天气预警信息发送的用户手机号码得不到及时更新,至今仍对多年前的号码进行发送,停机号、拆机号和移机号在所难免,既浪费了资源,也削减了灾害性天气预警信息发送的速度。因此,气象部门要充分发挥自身优势,积极主动配合运营商做好推广工作,加强短信服务号码宣传的力度,制订系统全面的宣传方案,有计划、有步骤地长期进行;通过电视、报纸、街区显示屏、网站等多种渠道进行宣传,确保编辑短信代码的短信发送至正确的定制号码(106573101800),即可

(上接第45页)

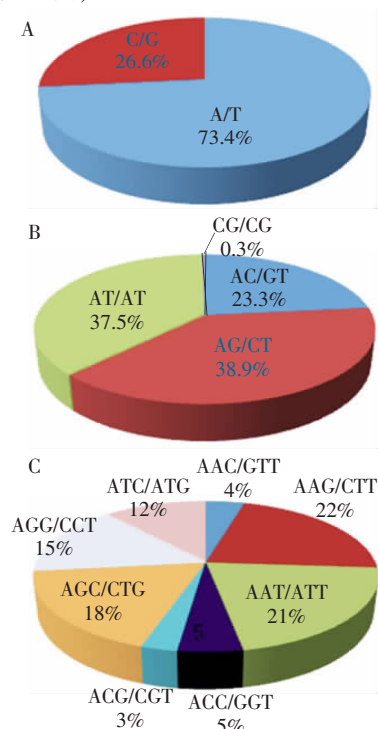


图1 一、二、三核苷酸重复中不同重复基元的比例

注:A:一碱基 SSR;B:二碱基 SSR;C:三碱基 SSR。

分类统计表明:随着重复次数的增加,SSR 数量有明显减少的趋势。较低重复次数(5~10次)的 SSR 有 688 个,占总数的 58.5%,是刺叶苏铁 EST-SSR 的主体部分;较高重复次数(11~30次)的 SSR 有 466 个,占总数的 39.6%;重复次数大于 30 次的 SSR 有 22 个,占总数的 1.9%,主要是在(A/T)重复中出现。SSR 序列越长变异就可能越大,其多态性也就

定制成功;在稳定现有气象短信用户的基础上,扩大宣传气象信息服务产品,扩大公共气象短信服务的覆盖面;并千方百计使通信公司形成“上层重视、中层支持、具体操作人员积极配合”的局面,使其对手机短信定制用户进行及时的更新,提高信息的有效性。

### 3 结语

近年来,各地极端天气气候事件频繁发生,气象灾害及气象次生灾害造成的损失愈来愈严重。气象部门作为基础性服务部门,担负着向群众提供天气资讯和灾害性天气预报的责任,要不断提升气象信息服务能力,为手机用户提供稳定、实用和高质量的气象服务信息,提高天气预报的准确率,加强气象短信的时效性,研究更多更好的气象服务产品,开展多样性、精细化的气象服务产品,以期更有力地吸引用户的注意力,最大限度地减轻气象灾害对人民生命财产的影响。

### 4 参考文献

- [1] 闫英,达成荣,袁成鑫,等.论气象手机短信在防灾减灾信息服务中的优势[J].青海气象,2004(1):52-54.
- [2] 梁毅进.气象短信的特色及注意的问题[J].广西气象,2004(3):59-61.
- [3] 保广裕,赵应章,达成荣,等.青海气象手机短信及编辑技巧[J].青海气象,2006(4):52-43.
- [4] 李芬,杨世刚,王小青.山西气象手机短信发展前景展望[J].山西气象,2004(2):27-28.

可能越丰富,也就越有利于 SSR 标记的开发与利用<sup>[5-6]</sup>。该研究得到的刺叶苏铁 SSR 重复次数在 10 次以上(含 10 次)的占总数的 62.2%。这暗示刺叶苏铁 EST-SSR 将具有较高的多态性与良好的实用性。有研究者基于刺叶苏铁 EST 序列信息开发设计了一些 EST-SSR 的引物,分别在我国特有的海南苏铁(*C.hainanensis*)和攀枝花苏铁(*C.panzhihuaensis*)中进行引物通用和其性多样性检测,结果支持了该文的分析预期,不仅引物通用性好,而且遗传多态性也较高<sup>[7-8]</sup>。

该研究结果将有助于刺叶苏铁及其近缘物种的 SSR 标记的开发与利用,也将有利于开展苏铁属植物基因组学研究如遗传图谱的绘制、遗传多样性评价、种质鉴定和系统进化及物种保护等研究。

### 4 参考文献

- [1] POWELL W, MACHRAY G C, PROVAN J. Polymorphism revealed by simple sequence repeats[J]. Trends Plant Science, 1996, 1(7): 215-222.
- [2] 陈军方,任正隆,高丽峰,等.从小麦 EST 序列中开发新的 SSR 引物[J].作物学报,2005,31(2):154-158.
- [3] 管仲天,周林.中国苏铁植物[M].成都:四川科学技术出版社,1996:128-131.
- [4] 王定跃.一个被错误鉴定的苏铁名称——华南苏铁[J].南京林业大学学报:自然科学版,2000(4):35-38.
- [5] CHO Y G, ISHII T, TEMNYKH S, et al. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.)[J]. Theoretical and Applied Genetics, 2000, 100(5): 713-722.
- [6] TEMNYKH S, PARK W D, AYERSN, et al. Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.)[J]. Theoretical and Applied Genetics, 2000, 100(5): 697-712.
- [7] WANG ZHENG FENG, YE WAN-HUI, CAO HONG-LIN. Identification and characterization of EST-SSRs and cpSSRs in endangered *Cycas hainanensis* [J]. Springer, 2008, 9(16): 1079-1081.
- [8] ZHANG FANG MING, SU TAO, YANG YING. Development of seven novel EST-SSR markers from *cycas panzhihuaensis* (cycadaceae)[J]. American Journal of Botany, 2010, 97(12): 159-161.