

茶树 EST-SSRs 信息分析

程小毛, 王璇, 黄晓霞

(西南林业大学 园林学院, 云南 昆明 650224)

摘要:从 NCBI 公共数据库中公布的茶树表达序列标签中获得 12 378 条茶树 EST, 通过 CAP3 软件进行拼接去冗余后再用 Bathprimer3 v1.0 (<http://wheat.pw.usda.gov/demos/BathPrimer3/>) 在线软件查找微卫星 (Microsatellite, SSR) 重复序列。结果表明: 通过前处理得到全长为 2 307.014 kb 的无冗余 EST 4 779 条, 从中检索出 27 种重复基元类型共 207 个 SSR, 分布于 156 条 EST 中, 出现频率是 3.26%。其中包括二核苷酸重复 72 个 (34.78%), 三核苷酸重复 88 个 (42.51%), 四核苷酸重复 31 个 (14.98%), 五核苷酸重复 16 个 (7.73%)。AG/CT 和 ACC/GGT 分别是二、三核苷酸中的优势重复类型, 分别占二、三核苷酸重复的 86.11% 和 25.00%。

关键词: 茶树; EST; SSR

中图分类号: S 571.1 **文献标识码:** A **文章编号:** 1001-0009(2011)10-0108-04

我国是茶树的原产地, 茶树也是我国目前非常重要的经济作物。近年来飞速发展的分子标记技术为茶树品种改良、遗传多样性分析、种质鉴定等提供了新的手段。但由于对茶树基因组信息知之甚少, 在茶树分子标记及其应用研究中, 目前主要有 RAPD 标记、AFLP 和 ISSR 标记等^[1-4]。但这类标记为未知功能标记, 富含的基因组信息较少, 而 EST-SSR 标记为功能性标记, 且具有共显性、重复性好, 操作简便等优点^[5] 而广泛应用于遗传图谱的构建、QTL 作图和遗传多样性等^[6-8]。近年来, 许多作物在公共数据库上释放了海量的 EST 序列, 使得 EST-SSR 标记具有开发简便及成本低廉等优点, 并且 EST-SSR 标记来自于基因的编码序列, 更容易获得基因表达的信息, 为功能基因的直接鉴定提供了可能性。目前, 油菜^[9]、水稻^[10]、小麦^[11]、辣椒^[6]、棉花^[12] 等作物相继开展了 EST-SSR 标记的开发及其应用方面的工作。

截止到 2010 年 5 月份, 茶树公布的 ESTs 序列已有 12 378 条, 相对于 2006 年茶树 EST 序列数目增加了 5

倍以上^[8]。基于上述茶树 EST 序列信息, 国内从事茶树相关研究工作人员开展了初步研究。金基强^[8] 等利用 1 989 条茶树 EST 序列开发出 19 对 EST-SSR 标记并检测了其在 10 个茶树品种中的多态性。现对已有的茶树 ESTs 中 SSR 信息进行全面的分析, 以期能够更多地了解茶树 ESTs 资源的特性, 为利用茶树 ESTs 建立 EST-SSR 标记和探索其在烟草分子标记辅助育种、基因定位、遗传多样性分析中的应用奠定基础。

1 材料与方法

1.1 茶树 EST 序列来源

茶树 EST 来自 NCBI (美国国家生物技术信息中心) bdEST 数据库 (<http://www.ncbi.nlm.nih.gov/bd-EST/index.html>) (2010 年 5 月份数据), 共计 12 378 条 EST 序列, 以 FASTA 格式保存这些序列。

1.2 EST 前处理

采用 EST-trimmer 软件 (http://pgrc.ipkgatersleben.de/misa/download/est_trimmer.pl) 除去 5' 端或 3' 端的 polyT 或 polyA 以及那些长度小于 100 bp 的 EST 序列。

1.3 EST 去冗余处理

将经前处理后的 EST 序列通过软件 CAP3 进行片段重叠群分析和聚类。拼接时设定的初始装配参数为: 最小匹配碱基数 (Minmatch) 为 40, 最小分值 (Minscore) 为 40, 匹配程度大于 95%。对错误拼接序列设置比较高的装配参数再次进行拼接、判别, 共进行了 3 次。

1.4 茶树 EST-SSR 搜索

用在线软件 Bathprimer3 v1.0 (<http://wheat.pw.usda.gov/demos/BathPrimer3/>)

第一作者简介: 程小毛 (1979-), 女, 安徽安庆人, 博士, 讲师, 现主要从事园林植物分子生物学方面的教学与科研工作。E-mail: xmceng0103@gmail.com。

责任作者: 黄晓霞 (1980-), 女, 博士, 讲师, 研究方向为植物生理生态。

基金项目: “省部级重点学科、省高校重点实验室及校实验室共享平台” 资助项目; 西南林业大学校级重点基金资助项目 (110909)。

收稿日期: 2011-03-09

usda.gov/demos/BatchPrimer3/)对去冗余后的 EST 序列进行 SSR 搜索。筛选标准为:二核苷酸重复的次数在 6 次或 6 次以上,三核苷酸重复的次数在 4 次或 4 次以上,四至五核苷酸重复的次数在 3 次或 3 次以上。同时,也筛选中间被少数碱基(间隔小于 10 或等于 10)打断的(interrupted)不完全重复的 SSR。

2 结果与分析

2.1 茶树 EST 中出现 SSR 的频率

对经过处理后得到的 4 779 条长度总计为 2 307.014 kb 的无冗余 EST 序列进行搜索,共检出含有 SSR 的序列 156 条,发生频率为 3.26%。其中,119 条含单个 SSR,另外 36 条含 2 个或 2 个以上的 SSR。共检出 207 个 SSR,占无冗余 EST 的 4.33%。从分布情况看,茶树 EST 中平均每 1.11 kb 就出现 1 个 SSR(表 1)。

茶树的 EST-SSR 种类较为丰富,二至五核苷酸重复类型都能检测到,但不同重复类型出现的频率不同(表 1)。二核苷酸和三核苷酸是优势重复类型,分别占总 EST-SSR 的 34.78%和 42.51%;四核苷酸和五核苷酸重复所占比例较小,总计仅 22.71%。

表 1 SSR 在茶树无冗余 EST 的出现频率

重复类型	SSR 数目	百分比/%	出现频率/%	平均分布 频率/kb
二核苷酸	72	34.78	1.51	3.20
三核苷酸	88	42.51	1.84	2.62
四核苷酸	31	14.98	0.65	7.44
五核苷酸	16	7.73	0.33	14.42
总 计	207	100	4.33	1.11

2.2 茶树 EST-SSR 的特性

在搜索出的茶树 EST-SSR 中,共观察到 27 种重复基元,二核苷酸重复基元有 3 种,三、四、五核苷酸重复基元分别各有 8 种(表 2)。从出现的频率来看,各种不同重复基元中出现最多的是 AG/CT,占总 SSR 数的 29.95%;其次是 AAG/CTT 和 ACC/GGT 重复,分别占总 SSR 数的 9.18%和 10.63%。在二核苷酸重复基元中,AG/CT 出现的次数最多(62 次),占二核苷酸 SSR 的 86.11%(表 2);在三核苷酸重复基元中,ACC/GGT 出现最多(22 次),占三核苷酸 SSR 的 25.00%(表 2)。在二核苷酸重复类型中未出现 CG/CG 重复基元。

表 2 4 779 条茶树非冗余序列中不同 SSR 的出现情况

重复基元	重复次数													合计	不同基元 比率/%
	3	4	5	6	7	8	9	10	11	12	13	14	>15		
AC/GT	0	0	0	1	1	1	2	0	0	0	0	0	0	5	2.42
AG/CT	0	0	0	12	5	8	12	6	5	4	2	3	5	62	29.95
AT/AT	0	0	0	0	4	1	0	0	0	0	0	0	0	5	2.42
AAC/GTT	0	2	1	2	0	0	0	0	0	0	0	0	0	5	2.42
AAG/CTT	0	12	7	0	0	0	0	0	0	0	0	0	0	19	9.18
AAT/ATT	0	6	2	0	0	0	0	0	0	0	0	0	0	8	3.86
ACC/GGT	0	12	7	1	1	1	0	0	0	0	0	0	0	22	10.63
ACG/CGT	0	7	0	1	0	0	0	0	0	0	0	0	0	8	3.86
ACT/AGT	0	8	0	1	0	0	0	0	0	1	0	0	0	10	4.83
AGG/CCT	0	8	0	1	0	0	0	0	0	0	0	0	0	9	4.35
CCG/CGG	0	4	1	2	0	0	0	0	0	0	0	0	0	7	3.38
AAAC/GTTT	1	0	1	0	0	0	0	0	0	0	0	0	0	2	0.97
AAAG/CTTT	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0.48
AAAT/ATTT	12	1	0	0	0	0	0	0	0	0	0	0	0	13	6.28
AACT/AGTT	4	0	0	0	0	0	0	0	0	0	0	0	0	4	1.93
AAGG/CCTT	2	1	0	0	0	0	0	0	0	0	0	0	0	3	1.45
AAGT/ACTT	5	0	0	0	0	0	0	0	0	0	0	0	0	5	2.42
AATT/AAAT	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0.48
ACCT/AGGT	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0.97
AAAAC/GTTTT	3	0	0	0	0	0	0	0	0	0	0	0	0	3	1.45
AAAAT/ATTTT	3	1	0	0	0	0	0	0	0	0	0	0	0	4	1.93
AGAGA/TCTCT	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0.97
AGGAG/CTCCT	1	1	0	0	0	0	0	0	0	0	0	0	0	2	0.97
ACAAC/GTTGT	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0.48
ATTCG/CGAAT	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0.48
ATACC /GGTAT	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0.97
ATAAT /ATTAT	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0.48
合计 Total	42	63	19	21	11	11	14	6	5	5	2	3	5	207	100.00

3 讨论

茶树 EST-SSR 出现频率较高,平均出现频率为 1/1.11kb,高于茶树(1/2.61 kb)^[8]、油菜(1/4.34)^[9]、烟草(1/14.21 kb)^[13],表明茶树的 EST-SSR 非常丰富。同一物种 SSR 出现频率与搜索标准(如 SSR 重复类型、重复次数)、分析方法、分析数据多少不同,而出现不同的结果。Cardle 等^[14]认为,拟南芥 EST 序列中,SSR 出现频率是 1/13.8 kb,水稻中出现频率是 1/3.4 kb,大豆是 1/7.4 kb,玉米为 1/8.1 kb。而根据 Morgante^[15]的统计标准,拟南芥中每 2.1 kb 就出现 1 个 SSR。Cao 等^[16]认为水稻中每 11.81 kb 出现 1 个 SSR。结果表明,发掘 EST-SSRs 的标准不一样时,得出的研究结果可能就存在差异。此外,EST-SSRs 的出现频率除与发掘标准有关外,还与 EST 序列来源的组织、器官不同相关。该试验方法与金基强等^[8]的统计标准相比,搜索二、三、四、五核苷酸重复基元重复次数的标准降低,减少了对单核苷酸和六核苷酸重复基元的搜索,用于搜索 SSR 的 EST 数目大大增加,结果表明在茶树中每 1.11 kb 的 EST 就出现 1 个 SSR。

前人研究认为,不同作物 SSR 的分布特征会存在一定差异,但存在一定的共性,即三核苷酸重复基元出现频率相对其它基元较高^[14,16],茶树与这些作物的 EST-SSRs 分布规律存在一致。在二核苷酸重复基元中,(AG/CT)_n 出现频率最高,这与小麦、玉米、水稻和大豆的分布规律类似^[17],但与拟南芥^[18]、棉花^[12]、甘蓝型油菜^[19]分布迥异,在这些物种中,(AT)_n 是最丰富的基元。在茶树二核苷酸重复类型中没有出现(GC)_n 基元,主要是因为真核基因组中(GC)_n 含量非常少^[18]。茶树 EST 序列中,三核苷酸重复基元的分布与甘蓝型油菜^[19]、棉花^[12]、水稻^[17]和拟南芥基因组^[18]三核苷酸重复基元分布一致,均是(AAG)_n 相对较丰富。

SSR 基元类型、重复长度与多态性间的关系对于开发 SSR 标记非常重要。一些研究结果表明,SSRs 的突变频率随着重复长度的增加而提高,重复长度长的 SSRs 一般多态性高^[6,20-21]。但在芸薹属物种中 SSR 重复长度在 11~20 bp 的标记多态性相对 SSR 重复长度长的标记多态性较高^[22]。因此,该试验在筛选 SSR 时,将长度标准定位 12 bp 以上,以期在今后的标记开发中能够获得更多的多态性标记,为茶树的分子育种、遗传多样性、种质鉴定等工作做出贡献。

参考文献

[1] Ratnaparkhe M B, Tekeoglu M, Muehlbauer F J. Inter-simple sequence repeat (ISSR) polymorphisms are useful for finding markers associated with disease resistance gene clusters [J]. Theor Appl Genet, 1998, 97: 515-519.
[2] Wachira F N, Powell W, Waugh R. An assessment of genetic diversity among *Camellia sinensis* L. (cultivated tea) and its wild relatives based on

randomly amplified polymorphic DNA and organelle-specific STS [J]. Heredity, 1997, 78(6): 603-611.

[3] 梁慧玲, 梁月荣. 植物分子标记技术原理及其在出书育种中的应用 [J]. 茶叶, 2003, 29(4): 191-194.
[4] 黄建安, 李家贤, 黄意欢, 等. 茶树 AFLP 分子连锁图谱的构建 [J]. 茶叶科学, 2005, 25(1): 7-15.
[5] Powell W, Machray G C, Provan J. Polymorphism revealed by simple sequence repeats [J]. Trends Plant Sci, 1996(7): 215-222.
[6] Yi G, Lee J M, Lee S, et al. Exploitation of pepper EST-SSRs and an SSR-based linkage map [J]. Theor Appl Genet, 2006, 14(1): 113-130.
[7] Tanaka J. QTL analysis for Branching Angle using F₁ segregating population derived from a cross between Tea Cultivar "Yabukita" and Tea Strain "Shizu-Inzatsul31" (in Japanese) [J]. Japan J. Breed, 1998, 48: 116-121.
[8] 金基强, 崔海瑞, 陈文岳, 等. 茶树 EST-SSR 的信息分析与标记建立 [J]. 茶叶科学, 2006(1): 17-23.
[9] 李小白, 张明龙. 油菜 EST-SSR 标记的建立 [J]. 分子细胞生物学报, 2007, 40(2): 137-144.
[10] 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]. Theor Appl Genet, 2000, 100: 713-722.
[11] Peng J H, Nore L, Lapitan V. Characterization of EST-Derived microsatellites in the wheat genome and development of eSSR markers [J]. Funct Integr Genomics, 2005(5): 80-96.
[12] 王长彪, 郭旺珍, 蔡彩平, 等. 雷蒙德氏棉 EST-SSRs 分布特征及开发利用 [J]. 科学通报, 2006, 51(3): 316-320.
[13] 胡重怡, 蔡刘体, 陈兴江. 烟草 ESTs 资源的 SSR 信息分析 [J]. 生物技术通报, 2009(7): 82-85, 93.
[14] Cardle L, Ratnsay L, Milbourne, et al. Computational and experimental characterization of physically clustered simple sequence repeats in plants [J]. Genetics, 2000, 156: 847-854.
[15] Morgante M, Hanafey M, Powell W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes [J]. Nat Genet, 2002, 30: 194-200.
[16] Gao L F, Tang J F, Li H W, et al. Analysis of microsatellites in major crops assessed by computational and experimental approaches [J]. Mol Breed, 2003(12): 245-261.
[17] Kantety R V, La R M, 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.
[18] Katti M V, Ranjekar P K, Gupta V S. Differential distribution of simple sequence repeats in eukaryotic genome sequences [J]. Mol Biol Evol, 2001, 18: 1161-1167.
[19] Cheng X M, Xu J S, Xia S, et al. Development and genetic mapping of microsatellite markers from genome survey sequences in *Brassica napus* [J]. Theor Appl Genet, 2009, 118: 1121-1131.
[20] Sharopova N, Mc Mullen M D, Schultz L, et al. Development and mapping of SSR markers for maize [J]. Plant Mol Biol, 2002, 48: 463-481.
[21] 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.
[22] Iniguez-Luy F, Voort A, Osborn T. Development of a set of public SSR markers derived from genomic sequence of a rapid cycling *Brassica oleracea* L. genotype [J]. Theor Appl Genet, 2008, 117: 977-985.

大花萱草组培苗的练苗技术

史淑一, 彭广霖, 丛惠芳, 潘仕梅, 张志芬

(中国农业大学(烟台), 山东 烟台 264670)

摘要:大花萱草幼嫩花梗切片接种于 MS+6-BA 3.0~5.0 mg/L+NAA 0.4~0.8 mg/L 的培养基上诱导丛生芽, 将丛生芽分割进行增殖培养和生根培养, 根系长度 3~5 cm 时在不同的练苗基质上进行驯化练苗。结果表明: 基质配比①草炭土: 珍珠岩=2:1, ②草炭土: 珍珠岩: 沙土=1:1:1, ③草炭土: 沙土=1:1 的 3 个基质组合均发根良好, 成活率在 95% 以上, 特别是③组合, 沙土代替珍珠岩, 既免去了冲洗珍珠岩的麻烦, 减少珍珠岩粉尘的污染, 又降低了成本。同时介绍了组培苗的常规练苗、冬季室内练苗、簇生苗练苗技术, 以期为大花萱草组培苗的大面积推广提供理论依据。

关键词:大花萱草; 组培苗; 练苗技术

中图分类号:S 682.1⁺9 **文献标识码:**B **文章编号:**1001-0009(2011)10-0111-03

大花萱草 (*Hemerocallis middendorffi*) 为百合科萱草属的多年生耐寒宿根草本花卉。现在世界各地广泛栽培。大花萱草花色多样, 花期 5~10 月中、下旬, 喜气候温暖, 抗旱, 抗病虫能力强, 适应性广, 对盐碱土壤有特别的耐性。华北地区可露地越冬, 栽培管理简单, 是优良的园林绿地花卉, 目前多采用分株、播种繁殖, 或扦插繁殖。大花萱草做为地被绿化花卉正在被开发, 种苗

缺口很大, 因此通过离体培养是快速繁殖大花萱草的有效途径, 而练苗技术是降低生产成本的另一重要技术。

1 组培苗的诱导与生根

1.1 丛生芽的培育

摘取没有开花的幼嫩花梗, 消毒后将花梗, 斜切成 0.2~0.3 mm 厚的椭圆形的薄片, 接入 MS+6-BA 3.0~5.0 mg/L+NAA 0.4~0.8 mg/L 培养基上进行培养, 7~8 周由愈伤组织直接分化形成丛生芽。将诱导形成的丛生芽分割后转移到 MS 增殖培养基上进行培养, 切分的小芽又会形成丛生芽。如此每 4 周左右转接 1 次, 进行快速增殖。该研究培养基均添加 3% 蔗糖和 0.75% 琼脂 (pH 5.8), 培养条件为 (23±1.0) °C, 每日 12~14 h 光照, 光照度 2 000~3 000 lx。

第一作者简介:史淑一(1964-), 女, 在读硕士, 副教授, 现主要从事应用气象研究工作。E-mail: ssy19640723@yahoo.com.cn。

责任作者:张志芬(1959-), 女, 本科, 教授, 现主要从事生物技术和农业气象教学工作。

基金项目:中国农业大学(烟台)校级资助项目(Yt200609)。

收稿日期:2011-02-28

Analysis of SSR Information in EST Resource of Tea Plants

CHENG Xiao-mao, WANG Xuan, HUANG Xiao-xia

(College of Landscape Architecture, Southwest Forestry University, Kunming, Yunnan 650224)

Abstract: 12 378 ESTs of tea plants in the public database of NCBI were downloaded and some of them redundant were removed using software CAP3, then resulting in 4 779 non-redundant ESTs with total length of 2 307.014 kp. Total of 207 SSRs distributed in 156 ESTs were detected, accounting for 3.26% of the non-redundant ESTs. The distribution distance of EST-SSRs was 1/1.11 kb. The dinucleotide and trinucleotide repeats were the main types, accounting for 34.78% and 42.51% of all the SSRs, respectively. AG/CT and ACC/GGT were the dominant motif, accounting for 86.11% and 25.00% in dinucleotide and trinucleotide repeats, respectively. The result in this paper provided a base for the development and future application of EST-SSR markers in tea plants.

Key words: tea plant; EST; SSR