

猕猴桃EST序列的SSR信息分析

姜春芽,徐小彪,廖娇,倪志华,李晶
(江西农业大学农学院,江西南昌 330045)

摘要:【研究目的】通过筛查猕猴桃EST数据库中的SSR重复序列,可为开发出新型的猕猴桃EST-SSR标记和分子生物学研究奠定理论基础;【方法】从NCBI公共数据库中最新公布的猕猴桃表达序列标签(Expressed Sequence Tag, EST)中随机抽取56400条序列,应用SSRHunter软件查找微卫星(Microsatellite, SSR)重复序列;【结果】研究结果表明,从猕猴桃EST序列中获得了7939条SSR,其中包括二核苷酸重复5131条(64.63%),三核苷酸重复1237条(15.58%),四核苷酸重复284条(3.58%),五核苷酸重复397条(5.00%),六核苷酸重复890条(11.21%)。在二核苷酸重复序列中,AG/CT共分布4654条(90.70%)。大约每2.48 kb长度的单一基因序列中即存在1个SSR,即平均7个单一基因中存在1个SSR;【结论】在猕猴桃EST序列中,二核苷酸重复序列是最丰富的重复单元,其次为三核苷酸重复和六核苷酸重复。在所获得的SSR重复单元中,AG/CT为优势重复。

关键词:猕猴桃;表达序列标签;微卫星

中图分类号:S663 **文献标识码:**A

Analysis of SSR Information in EST Resources of Kiwifruit (*Actinidia* spp.)

Jiang Chunya, Xu Xiaobiao, Liao Jiao, Ni Zhihua, Li Jing

(College of Agronomy, Jiangxi Agricultural University, Nanchang Jiangxi 330045)

Abstract: 【OBJECTIVE】We screened and mined the SSR repeats in the EST database of kiwifruit (*Actinidia* spp.) in order to make better use of the EST-SSR resources of kiwifruit for further molecular biological studies and new EST-SSR marker development; 【METHOD】56,400 EST sequences were randomly selected from EST (Expressed Sequence Tag) sequences of kiwifruit in the database of NCBI. EST sequences were analyzed and the SSR (Microsatellite) could be found out by using the SSRHunter software; 【RESULTS】The results showed that the 7939 SSRs were identified from the randomly selected kiwifruit EST resources. Among which there were 5131 (64.63%) dinucleotide, 1237 (15.58%) trinucleotide, 284 (3.58%) tetra nucleotide, 397 (5.00%) pentanucleotide and 890 (11.21%) hexanucleotide SSRs. Among the dinucleotide sequences, AG/CT repeat motif was accounted for 4654 (90.70%). The frequency of SSRs was approximately 1/2.48 kb, which could exist to 1 SSR in 7 unigenes; 【CONCLUSION】The dinucleotide repeats appeared to be the most abundant SSRs, followed by the trinucleotide and hexanucleotide repeats. Among them the repeat motif such as AG/CT was predominant in each type of SSRs.

Key words: Kiwifruit, EST, SSR

0 引言

【本研究的重要意义】微卫星(microsatellite)是指以少数几个核苷酸(一般为1~6个)为基本重复单位的

多次串联重复DNA序列,也称简单序列重复(simple sequence repeats, SSR)。EST(expressed sequence tags, 表达序列标签)是将mRNA在体外反转录成cDNA并

基金项目:国家自然科学基金项目(30660113)。

第一作者简介:姜春芽,男,1980年出生,湖南人,在读硕士,研究方向为果树种质资源研究。

通讯作者:徐小彪,男,1963年出生,江西人,教授,博士,研究方向为果树种质资源与生物技术研究。通信地址:330045江西南昌江西农业大学农学院, Tel: 0791-3813636, E-mail: xiaobiaoxu@hotmail.com。

收稿日期:2009-03-27, 修回日期:2009-04-08。

克隆到载体构建成 cDNA 文库,大规模随机挑取 cDNA 克隆,对其 5' 端或 3' 端进行一步法测序后获得的长约 150-500bp 的表达序列片段^[1-2]。一个 EST 代表生物体某种组织某一时期的一个表达基因,可以说 EST 是基因的“窗口”,它能反映 mRNA 的信息,可代表生物体某种组织或细胞某一时间的一个表达基因^[3]。EST-SSR 标记来源于相对保守的转录区域,较基因组 SSR 标记具有更高的通用性和保守性。EST-SSR 标记具有开发简便及成本低廉等优点,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^[4]。

【前人研究进展】随着 EST 计划在不同物种间的扩展和研究内容的深入,来源于不同物种,不同组织、不同细胞类型和不同发育阶段的表达基因序列的数目急剧上升。公共数据库中的 EST 数据已逐渐发展成为寻找新型分子标记 EST-SSR 的主要数据来源。1991 年各个公共数据库中的 EST 数目不足 2000 条,而目前小麦 (*Triticum Aestivum*) 已有 106 万条,玉米 (*Zea Mays*) 有 117 万条,水稻 (*Oryza sativa*) 有 112 万条 EST。快速增长的 EST 数据为 SSR 标记的开发提供了丰富的来源,在大麦、黑麦、甘蔗、小麦、水稻、高羊茅、葡萄、棉花、甘蔗、黑麦草、云杉、杏树等多种植物中都建立了 EST-SSR 标记。目前,EST-SSR 标记已成为重要农艺性状定位、基因作图、遗传多样性、比较基因组学研究的新型重要工具^[4]。【本研究切入点】前人的研究为植物 EST-SSR 标记的开发打下了良好的基础,但有关猕猴桃 EST 数据库中 SSR 标记的建立尚未见报道。【拟解决的关键问题】据此,该研究旨在利用最新公布的猕猴桃 EST 数据库,筛查 SSR 重复序列,为开发出新型的猕猴桃 EST-SSR 标记和分子生物学研究奠定理论基础。

1 材料与方法

1.1 猕猴桃 EST 来源

截至 2008 年 8 月,GenBank 已公布了 13 多万条的猕猴桃 EST 序列,从 NCBI bdEST 数据库 (www.ncbi.nlm.nih.gov/bdEST/index.html) 下载所有的猕猴桃 EST 序列。用记事本保存至本地电脑备用。

1.2 EST-SSR 的发掘

利用 SSRHunter1.3 程序,结合人工搜索,从所有下载的 EST 序列中随机抽取其中 56,400 条序列进行 SSR 查找。搜索的标准是:二核苷酸、三核苷酸、四核苷酸、五核苷酸和六核苷酸最少重复次数分别为 7 次、5 次、4 次、4 次、3 次,包括复合型微卫星。微卫星的重复次数越多,相应的检验出等位基因数目就越多。

2 结果与分析

对来源于猕猴桃 EST 数据库 (NCBI bdEST) 的 56400 条单一基因序列使用 SSRHunter 进行了简单序列重复 SSRs 搜索。结果表明,从中共获得含 SSR 的 EST 序列 7939 条,其中包括二核苷酸重复 5131 条,三核苷酸重复 1237 条,四核苷酸重复 284 条,五核苷酸重复 397 条,六核苷酸重复 890 条,分别占总 SSR 的 64.63%、15.58%、3.58%、5.00% 和 11.21%。大约每 2.48kb 长度的单一基因序列中即存在 1 个 SSR,即平均 7 个单一基因中存在 1 个 SSR。其中,二核苷酸重复序列是最丰富的重复单元,其次为三核苷酸重复和六核苷酸重复(表 1)。在所获得的 SSR 重复单元中,AG/CT 为优势重复,共分布 4654 条,占二核苷酸重复中各重复基元的 90.70% 其次是 AT/AT 重复,共分布 286 条,占二核苷酸重复中各重复基元的 5.57%(表 2)。

表 1 猕猴桃 SSR 在 EST 中出现的频率

类型	数目	各类型的比例
二核苷酸 dinucleotide	5131	64.63%
三核苷酸 trinucleotide	1237	15.58%
四核苷酸 tetranucleotide	284	3.58%
五核苷酸 pentanucleotide	397	5.00%
六核苷酸 hexanucleotide	890	11.21%
总计 Total	7939	100%

表 2 二核苷酸重复中各重复基元的数量及比例

类型	数目	各类型的比例
AG/CT	4654	90.70%
AT/AT	286	5.57%
AC/GT	190	3.70%
CG/CG	1	0.02%
总计 Total	5131	100%

在三核苷酸中,最为丰富的三种重复基元是 ACC/GGT, AAG/CTT 和 CGC/GCG, 共分布 817 条,它们占三核苷酸重复中各重复基元的 66.04%(表 3)。在其它基元中,还包括 AGG/CCT, AGC/GCT, AAC/GTT, ATC/GAT, AGT/ACT, CGA/TCG, AAT/ATT 等 7 种重复基元,共 420 条 SSR 序列,占 33.96%。AGT 与 CGT 重复基元未见分布。

3 小结

猕猴桃的 EST 序列中,二核苷酸比例最高,其次是三核苷酸重复,它们分别占总 SSR 的 64.63% 和 15.58%,而在拟南芥基因组的研究中发现,单核苷酸、二核苷酸和三核苷酸重复的 SSR 所占比例分别为

表3 三核苷酸重复中各不同重复基元的数量及比例

类型	数目	各类型的比例
ACC/GGT	500	40.42%
AAG/CTT	194	15.68%
CGC/GCG	123	9.94%
AGG/CCT	101	8.16%
AGC/GCT	92	7.44%
AAC/GTT	70	5.66%
ATC/GAT	49	3.96%
AGT/ACT	34	2.75%
CGA/TCG	32	2.59%
AAT/ATT	42	3.40%
总计 Total	1237	100%

33%、30%、30%，且分布较为均匀^[5]。在大麦、玉米、大豆、黑麦、甘蔗和小麦中，三核苷酸重复的SSR最多，在54%~78%之间^[6-7]。葡萄^[8]、柑橘^[9]也以三核苷酸占主导。茶树^[10]、杏树和桃树^[11]是以二核苷酸为主，此项研究证实猕猴桃也是以二核苷酸重复为主导，它占到总SSR的64.63%。值得注意的是，猕猴桃中的六核苷酸重复所占比例达到11.21%，这在所报道的植物种类中还未见到过，大麦、玉米、大豆、黑麦、甘蔗、桃树、柑橘和茶树等一般都在3%以下。

猕猴桃的二核苷酸重复以AG/CT为主，它占二核苷酸重复中各重复基元的90.70%，而AT/AT、AC/GT与CG/CG三种重复类型所占比例还不到10%，这与Huang^[12]的研究结果基本一致。在很多中脊椎动物中，CA重复是最丰富的种类，但是在酵母菌和植物中，AT是最常见的重复^[13]。在对拟南芥、杏树、桃树^[11]、水稻、玉米以及大豆^[14]的研究中，均未见GC重复的报道，而在油菜^[15]的研究中发现GC出现的频率很低(0.016/100KB)，对咖啡的研究也只发现2条GC重复^[16]，此次研究也筛查到了一次GC重复，这可能反映植物对CG/GC重复有明显的偏倚性，但也不缺乏这种重复序列。猕猴桃的三核苷酸中最为丰富的三种基元是ACC/GGT，AAG/CTT和CGC/GCG，这三种类型共占总数的66.04%，这与拟南芥、柑橘、大豆等双子叶植物中最为丰富的类型的AAG/CTT有所差别。该研究通过筛查猕猴桃EST数据库中的SSR重复序列，可为开发出新型的猕猴桃EST-SSR标记和分子生物学研究奠定理论基础。

参考文献

- [1] 骆蒙,贾继增.植物基因组表达序列标签(EST)计划研究进展[J].生物化学与生物物进展,2001,(28):494-497.
- [2] Ilatey F,Tosser Klopp G, Clouscard Martinato C, et al. Expressed sequenced tags for genes:a review[J]. Genet Sel Evol, 1998, (30): 521-541.
- [3] 于凤池. EST技术及其应用综述[J].中国农学通报,2005,(2):54-58.
- [4] Varshney R K, Grnaer A, Sorrells M E. Genetic microsatellite markers in plants:features and applications[J]. Trends in Biotechnology, 2005, (23):48-55.
- [5] Kantety R V, La Rota M, Matthews D E, et al. Data mining of simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat[J]. Plant Mol Bio, 2002, 148 (5-6):501-510.
- [6] Thiel T, Michalek W, Varshney K, et al. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.) [J].Theor Appl Genet, 2003, 106(3):411-422.
- [7] Cardle L,Rmasay L,Milboume D,et al. Computational and experimental characterization of physically clustered simple sequence repeats in plants [J]. Genetics, 2000, 156:847-85.
- [8] Scott K D, Eggler P,Seaton G,et al. Analysis of SSRs derived from grape ESTs[J]. Theor Appl Genet, 2000, (1):723-726.
- [9] Jiang D, ZHong G Y, Hong Q B. Analysis of Microsatellites in Citrus Unigenes[J]. Acta Genetica Sinica, 2006, 33(4):345 - 353.
- [10] 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,(26):17-23.
- [11] Jung S, Abbott A, Jesudurai C, et al. Frequency type distribution and annotation of simple sequence repeats in Rosaceae ESTs[J]. Funct Integr Genomics, 2005, (5):136-143.
- [12] Huang W G, Morange M, Testolin R. Microsatellite DNA in *Actinidia chinensis*: isolation, characterization, and homology in related species [J]. Theor Appl Genet, 1998, (97):1269-1278.
- [13] Tóth G, Gáspári Z, Jurka J. Microsatellites in different eukaryotic genomes: survey and analysis [J]. Genome Research, 2000, 967-981.
- [14] 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.
- [15] 李小白,张明龙,崔海瑞,等.油菜 EST 资源的 SSR 信息分析[J].中国油料作物学报,2007,29(1):20-25.
- [16] Ramesh K, Aggarwal Prasad S, Hendre, et al. Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species [J]. Theor Appl Genet, 2007, (114):359 - 372.