

西瓜抗枯萎病相关 EST-SSR 的信息分析

冯建明^{1,2}, 郭绍贵², 吕桂云³, 陈年来¹, 张海英², 宫国义², 许 勇²

(1. 甘肃农业大学 农学院, 甘肃 兰州 730070; 2. 北京市农林科学院 蔬菜研究中心, 北京 100097; 3. 河北农业大学, 河北 保定 071001)

摘要:对西瓜枯萎病菌生理小种 1 诱导的抑制差减杂交 cDNA 文库测序获得的 4 000 条 EST 进行分析, 经过前处理后拼接得到 1 487 条 unigene 序列, 全长为 759 kb。在其中 978 条 unigene 序列中共检索出 2 136 个 EST-SSR, 出现频率为 53.4 %。EST-SSR 的平均分布距离和平均长度分别是 1/0.36 kb、19.59 bp。EST-SSRs 中的重复单元以二核苷酸和三核苷酸重复为主, 二者在总 EST-SSRs 中的出现频率为 98.08 %。GA/TC 和 GAA/TTC 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 25.78 % 和 12.97 %。西瓜抗枯萎病相关 EST 资源的 SSR 信息分析为进一步建立西瓜 EST-SSR 标记和探索其在西瓜基因组学研究中的应用奠定了基础。

关键词:西瓜; 枯萎病; EST; SSR

中图分类号:S651.Q78 **文献标识码:**A **文章编号:**1000-7091(2009)03-0087-05

Analysis of EST-SSR Information Related to Fusarium Wilt Resiatance in Watermelon

FENG Jian-ming^{1,2}, GUO Shao-gui², LU Gui-yun³, CHEN Nian-lai¹,
ZHANG Hai-ying², GONG Guo-yi², XU Yong²

(1. Agronomy College, Gansu Agricultural University, Lanzhou 730070, China; 2. Beijing Vegetable Research Center, Beijing 100097, China; 3. Hebei Agricultural University, Baoding 071001, China)

Abstract: Simple sequence repeats (SSRs) were investigated in the unigene sequences from suppression subtractive hybridization (SSH) cDNA library which constructed with watermelon root induced by *Fusarium oxysporum* f. sp. *niveum* race 1. A total of 1487 unigene sequences corresponding to 795 kb were searched, resulting in the identification of 2 136 SSRs. The estimated frequency of SSRs was approximately 53.4 %, and the overall average length of SSRs was 19.59 bp, the average distance of distribution were 1/0.36 kb. The di- and tri-nucleotide repeats appeared to be the most abundant SSRs, accounting for 98.08 % in all the EST-SSRs. The adenine rich repeats such as GA/TC and GAA/TTC were predominant in each type of SSRs (di-, tri-), accounting for 25.78 % and 12.97 % in di- and tri-nucleotide repeats respectively. These EST-SSRs information could be efficiently exploited for the development of gene-derived SSR markers.

Key words: Watermelon; Fusarium wilt; EST; SSR

目前,我国西瓜的总产量和消费量均居世界第一,现代基因组学研究方法的运用是提高西瓜育种效率的有效途径之一。目前已在西瓜遗传图谱构建、遗传多样性分析和基因定位等方面开展了研究工作^[1]。然而,随着西瓜基因组学研究的不断深入,已有分子标记的种类与数量已经不能满足目前基因组研究的需要,因此,开发更加高效和可靠的新型分子标记势在必行。

与其他分子标记技术相比,SSRs (Simple sequence repeats) 具有多态信息含量高、共显性遗传、技术简单、重复性好、特异性强等特点^[2]。根据 SSR 标记的序列性质不同,SSR 标记可分为基因组 SSR (Genomic SSR, gSSR) 和表达序列标签 SSR (Expressed sequence tag SSR, EST-SSR) 等。gSSR 的开发需要构建基因组文库,技术上的复杂性和高昂的费用制约了其发展和应用。与 gSSR 相比,EST-SSR 可以从

收稿日期:2009-02-11

基金项目:国家“863”计划(2006AA10Z1A8);“十一五”国家科技支撑计划(2006BAD01A7-6-02);国家自然科学基金(30570997; 30571276);北京市自然科学基金(5062008);北京市科委项目(Z0708050150702)

作者简介:冯建明(1982-),男,甘肃酒泉人,在读硕士,主要从事蔬菜育种与生理栽培研究。

通讯作者:许 勇(1965-),男,湖北武汉人,研究员,博士,主要从事蔬菜育种与分子生物学研究。

EST 序列中获得,相对简便经济,在近缘物种间具有高度的通用性,而且随着高速自动 DNA 测序技术的发展,EST 的获取也越来越容易。目前 EST-SSR 的开发已经在多种农作物的基因组学研究中广泛应用,在水稻^[3]、小麦^[4]、甘蔗^[5]、葡萄^[6]、番茄^[7]、白菜^[8]等作物中都有成功报道。Joobeur 等^[11]曾利用西瓜 BAC 文库设计了 142 对 SSR 引物,而关于西瓜 EST-SSR 的开发与信息分析尚未见报道。

本实验室构建了西瓜枯萎病菌生理小种 1 诱导的抑制差减杂交 cDNA 文库并进行测序,获得了 4 000 条 EST。本研究对这些 EST 序列的 SSR 信息进行了全面分析,明确西瓜 EST-SSR 发生频率和特点,为进一步建立西瓜 EST-SSR 标记和探索其在西瓜基因组学研究中的应用奠定基础。

1 材料和方法

1.1 西瓜 EST 来源

EST 序列来源于北京市农林科学院蔬菜研究中心构建的西瓜枯萎病菌生理小种 1 诱导的抑制差减杂交 cDNA 文库,共有 4 000 条。

1.2 EST 前处理

采用 Phred 程序(<http://www.phrap.org/consed/consed.html>)去除 EST 中低质量序列。采用 lucy 软件去除载体、接头及小于 100 bp 的序列。采用 SeqClean 软件去除污染序列、5 端或 3 端的 polyT 或 polyA 以及长度小于 100 bp 的序列。

1.3 去冗余—聚类

前处理后的 EST 用 cluster 软件进行聚类,利用

CAP3 软件进行拼接,得到 unigene。用 clview 软件对经 CAP3 拼接过的 EST 进行校正和核对。

1.4 西瓜 EST 序列的 SSR 检索

利用 SSRIT(<http://www.gramene.org/db/searches/ssrtool>)在线进行 SSR 搜索,搜索标准为重复基序长度范围为 2~6 bp,最少 3 次重复。

2 结果与分析

2.1 西瓜 EST-SSR 的重复类型和频率

根据西瓜抗枯萎病抑制差减 cDNA 杂交文库测序信息获得 4 000 条 ESTs,经过前处理后拼接得到 1 487 条 unigene,全长 759 kb。含有 SSR 的 unigene 978 条,占全部 unigene 的 65.8%,其中 226 条 unigene 序列中含有 1 个 SSR,752 条 unigene 中含有 2~15 个 SSR。在 978 条 unigene 中,共包含 2 136 个 SSR,平均分布距离和平均长度分别是 1/0.36 kb 和 19.59 bp。

西瓜 EST-SSR 重复类型丰富,二至六核苷酸的重复类型都能检索到,但各种重复类型在全部 2 136 个 SSR 中所占比例以及出现频率各不相同(表 1)。二核苷酸和三核苷酸的重复类型分别占全部 EST-SSR 总数的 71.72%和 26.36%,四、五、六核苷酸重复类型所占比例较小,总计仅为 1.92%。在西瓜 EST 中,二核苷酸重复类型的出现频率最高,达到 38.3%,其次为三核苷酸重复,占 14.08%;四、五、六核苷酸重复类型的出现频率均不足 1%。根据重复类型和出现频率的数据,说明西瓜基因组中以二、三核苷酸重复为主。

表 1 西瓜 EST-SSRs 的数量和出现频率

Tab.1 Frequency and number occurred in EST-SSRs of watermelon

重复类型 Repeat types	数目 Number	占全部 SSR 比例/ Proportion in all SSRs %	频率/ Frequency %	平均距离/ Average distance kb
二核苷酸 Dinucleotide	1 532	71.72	38.30	0.50
三核苷酸 Trinucleotide	563	26.36	14.08	1.35
四核苷酸 Tetranucleotide	34	1.59	0.85	22.32
五核苷酸 Pentanucleotide	4	0.19	0.10	189.75
六核苷酸 Hexanucleotide	3	0.14	0.08	253.00
总计 Total	2 136	100.00	53.40	0.36

2.2 西瓜 EST-SSR 中主要重复基元特性

在 2 136 个西瓜 EST-SSR 中,共检索到 55 种核苷酸重复基元。其中二核苷酸重复基元有 6 种,三、四、五、六核苷酸重复基元分别有 30、15、2、2 种。二核苷酸重复基元中以 GA/TC、AT/TA、AG/CT 基元为主,三种重复基元所占比例相近,其总数约占二核苷酸重复的 74.15%,三核苷酸重复基元中以 GAA/TTC、AAG/CTT 为主,约占三核苷酸重复的 23%(表 2)。四、五、六核苷酸重复基元共 19 种,各自所占比

例接近,出现数量和频率都很低。

2.3 西瓜 EST-SSR 重复基元的分布特点

EST-SSR 的多态性主要是由于重复基元的重复次数以及 SSR 长度不同造成的。表 3 显示由于西瓜 EST-SSR 重复基元的重复次数不同而获得西瓜 EST-SSR 的数目。进行西瓜 EST 序列的 SSR 检索时,以重复基元的最低重复次数为 3 的标准进行检索,共获得 13 种重复次数,其中重复基元的最高重复次数达到 23 次。重复次数为 3 次时获得的 EST-SSR 数

量最多, 占所有 EST-SSR 总数的 85.86%, 其次为 4 次重复, 占 9.83%。

表 2 西瓜 EST-SSRs 重复类型的出现情况

Tab.2 Repeat type occurred in EST-SSRs of watermelon							
重复类型 Repeat type	重复基元 Repeat Motifs	数量 Number	比例/ % Proportion	重复类型 Repeat type	重复基元 Repeat Motifs	数量 Number	比例/ % Proportion
二核苷酸 Dinucleotide	GA/ TC	395	25.78	四核苷酸 Tetranucleotide	CAA/ TTG	11	1.95
	AT/ TA	371	24.22		CGA/ TCG	8	1.42
	AG/ CT	370	24.15		ACT/ AGT	7	1.24
	CA/ TG	217	14.16		GTA	6	1.07
	AC/ GT	141	9.20		CCG/ CGG	6	1.07
	CG/ GC	38	2.48		GAC	3	0.53
三核苷酸 Trinucleotide	GAA/ TTC	73	12.97		CTA/ TAG	3	0.53
	AAG/ CTT	56	9.95		CGT	2	0.36
	AGA/ TCT	36	6.39		CAAC/ GTTG	4	11.76
	AAT/ ATT	35	6.22		AATT/ TATT	3	8.82
	GGA/ TCC	29	5.15		AAAG/ CTTT	3	8.82
	CCA/ TGG	28	4.97		AAAC/ GTTT	3	8.82
	TCA/ TGA	27	4.80		TCCC/ TTCT	2	5.88
	AGC/ GCT	23	4.09		TAAA/ TTTG	2	5.88
	CAG/ CTG	21	3.73		GCAA/ TTGG	2	5.88
	AGG/ CCT	19	3.37		GATG/ TATC	2	5.88
	AAC/ GTT	18	3.20		CTTG/ ACAT	2	5.88
	TAA/ TTA	17	3.02		ATTA/ TAAT	2	5.88
	GCA/ TGC	17	3.02		ATAG/ CTCA	2	5.88
	ATA/ TAT	17	3.02		ATAA/ TTAT	2	5.88
	CTC/ GAG	14	2.49		AAGA/ TCAT	2	5.88
ATC/ GAT	14	2.49	AAAT/ ATTT		2	5.88	
CCC/ GGC	13	2.31	TCCT		1	2.94	
CAC/ GTG	13	2.31	五核苷酸 Pentanucleotide	AAACG/ CTTTC	2	50.00	
ATG/ CAT	12	2.13	AAACA/ TTTTA	2	50.00		
ACC/ GGT	12	2.13	六核苷酸 Hexanucleotide	ATICCG/ CCTTTC	2	66.67	
ACA/ TGT	12	2.13	TTTTTC	1	33.33		
CGC/ GCG	11	1.95					

表 3 西瓜 EST-SSRs 重复次数的出现情况

Tab.3 Number of repeat occurred in EST-SSRs of watermelon							
重复次数 Number of repeat	重复类型 Repeat type					总计 Total	所占比例/ % Proportion
	二核苷酸 Dinucleotide	三核苷酸 Trinucleotide	四核苷酸 Tetranucleotide	五核苷酸 Pentanucleotide	六核苷酸 Hexanucleotide		
3	1 330	473	28	2	1	1 834	85.86
4	146	54	6	2	2	210	9.83
5	29	15	0	0	0	44	2.06
6	8	12	0	0	0	20	0.94
7	5	3	0	0	0	8	0.37
8	4	5	0	0	0	9	0.42
9	2	0	0	0	0	2	0.09
10	3	0	0	0	0	3	0.14
11	1	0	0	0	0	1	0.05
12	0	1	0	0	0	1	0.05
15	2	0	0	0	0	2	0.09
18	1	0	0	0	0	1	0.05
23	1	0	0	0	0	1	0.05
总计 Total	1 532	563	34	4	3	2 136	100.00

西瓜 EST-SSR 的长度差异较大, 最长的 SSR 达到了 46 bp, 平均长度为 19.59 bp (图 1)。大部分 EST-SSR 长度都集中在 12 bp 以下, 这些 SSR 多以低级基元的二、三核苷酸重复为主。SSR 长度为 6 bp

的 EST-SSR 共有 1 330 个, 全部为二核苷酸 3 次重复, 占总 SSR 的 62.3%; SSR 长度为 9 bp 的共有 473 个, 全部为三核苷酸 3 次重复, 占总 SSR 的 22.1%。

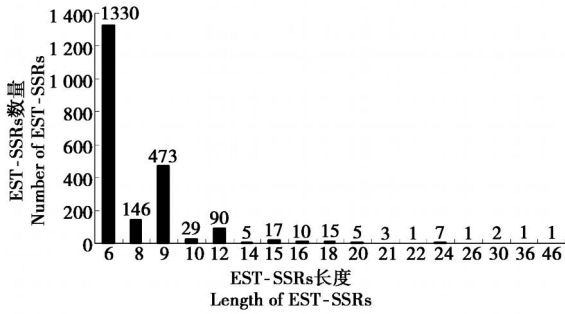


图1 西瓜 EST-SSR 长度的分布

Fig.1 Length of EST-SSRs distributed on watermelon

3 讨论

本研究对西瓜枯萎病菌生理小种 1 诱导的抑制差减杂交 cDNA 文库测序获得 4 000 条 ESTs 进行分析,共检索出 2 136 个 EST-SSR,其 EST-SSR 重复类型与大多数植物有一定的相似性,即主要属于二、三核苷酸重复类型。已有研究表明,茶树^[9]、猕猴桃^[10]、杏树、桃树^[11]和欧洲白桦^[12]中以二核苷酸重复为主;而在甘蔗^[5]、葡萄^[6]、大豆^[13]、柑橘^[14]和百合^[15]中,则是三核苷酸重复为主;在白菜^[8]和油菜^[16]中,二、三核苷酸重复类型相当。在本研究中西瓜的二、三核苷酸重复共占全部 EST-SSR 总数的 98.08 %。同时发现,GA/TC 是大多数植物二核苷酸重复基元的主导类型^[3,17],而 GA/TC 在西瓜二核苷酸重复中也最多。出现这种情况可能的原因之一是二核苷酸重复单元 AG/TC 在 mRNA 中根据不同的阅读起始位点可以分别读为:GAG、AGA、UCU 和 CUC,从而分别翻译为 Arg、Glu、Ala 和 Leu 氨基酸,而 Ala 和 Leu 氨基酸在蛋白质中出现的频率分别高达 8 %和 10 %。这也说明了为什么 GA/CT 重复在 EST 序列中高频率存在的原因^[18]。

核苷酸重复基元的偏倚性在多数植物中普遍存在,且出现与否和物种无关。本研究表明西瓜 EST-SSR 也存在明显的偏倚性。GA/TC 是大多数植物二核苷酸重复基元的主导类型,而 CG/GC 含量极少,白菜^[8]中仅为其 EST 的 0.21 %,本研究为 0.95 %;在拟南芥^[13]、杏树、桃树^[11]、水稻、玉米、大豆^[19]等植物中甚至没有发现 GC 重复基元;小麦的研究报道中陈军方等^[20]和 Nicot 等^[21]均未发现 GC 重复基元,而 Gao 等^[19]发现小麦 EST-SSR 中 GC 重复只以很低的频率(0.09/100 kb)出现。此外,我们还发现在西瓜中四、五、六核苷酸重复占全部 EST-SSR 的比例很少,分别为 1.59 %、0.19 %和 0.14 %,这些高级基元的核苷酸重复都不是西瓜 EST-SSR 重复基元的主导类型,表明西瓜 EST-SSR 确实存在明显的偏倚性。

本研究 EST-SSR 的出现频率为 53.4 %,而其他物种从 EST 数据库中检出的 SSR 均低于 20 %,如水稻 4.7 %^[3]、甘蔗 2.9 %^[5]、白菜 10.34 %^[8]、欧洲白桦 12.05 %^[12]。这种差异可能是物种间的真实 SSR 信息差异或搜寻 SSR 时所用长度最低标准不同造成的^[8]。因此,不同研究者采取鉴定标准的不同会对 SSR 在 EST 上的分布特征及不同作物研究结果的比较产生较大的影响。本试验西瓜 EST-SSR 出现频率高可能与拼接聚类后高质量的 unigene 中有多个 SSR 有关,也可能是由于西瓜本身基因组特性决定的。另外,为了使尽可能多的差减杂交获得的 EST 序列转化有价值的遗传标记进行验证,在搜索 SSR 时采用了相对宽松的搜索标准也是原因之一。目前正在根据 SSR 位点检索结果设计 EST-SSR 标记,以对差减获得的枯萎病抗性相关 unigene 进行连锁验证和遗传图谱定位研究。

参考文献:

- [1] Joobeur T, Gusmini G, Zhang X, *et al.* Construction of a watermelon BAC library and identification of SSRs anchored to melon or Arabidopsis genomes[J]. *Theor Appl Genet*, 2006, 112:1553 - 1562.
- [2] Powell W, Machray G C, Provan J. Polymorphism revealed by simple sequence repeats[J]. *Trends Plant Science*, 1996, 1: 215 - 222.
- [3] Kantety R V, Rota M L, Matthews D E, *et al.* Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat[J]. *Plant Molecular Biology*, 2002, 48:501 - 510.
- [4] Eujayl I, Sorrells M E, Baum M, *et al.* Isolation of EST derived microsatellite markers for genotyping the A and B genomes of wheat [J]. *Theoretical and Applied Genetics*, 2002, 104(23):399 - 407.
- [5] Cordeiro G M, Casu R, McIntyre C L, *et al.* Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to helianthus and sorghum[J]. *Plant Science*, 2001, 160(6):1115 - 1123.
- [6] Scott K D, Eggler P, Seaton G, *et al.* Analysis of SSRs derived from grape ESTs[J]. *Theoretical and Applied Genetics*, 2000, 100(5):723 - 726.
- [7] Areshchenkova T, Ganal M W. Comparative analysis of polymorphism and chromosomal location of tomato microsatellite markers isolated from different sources [J]. *Theoretical and Applied Genetics*, 2002, 104:229 - 235.
- [8] 忻雅, 崔海瑞, 卢美贞, 等. 白菜 EST-SSR 信息分析与标记的建立[J]. *园艺学报*, 2006, 33(3):549-554.
- [9] 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.

- [10] Fraser L G, Harvey C F, Crowhurst R N, *et al.* EST-derived microsatellites from *Actinidia* species and their potential for mapping [J]. *Theor Appl Genet*, 2004, 108: 1010 - 1016.
- [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] 王艳敏, 魏志刚, 杨传平. 白桦 EST-SSR 信息分析与标记的开发[J]. *林业科学*, 2008, 44(2): 78 - 84.
- [13] Cardle L, Ramsay L, Milbourne D, *et al.* Computational and experimental characterization of physically clustered simple sequence repeats in plants [J]. *Genetics*, 2000, 156: 847 - 854.
- [14] Chen C X, Zhou P, Choi Y A, *et al.* Mining and characterizing microsatellites from citrus ESTs [J]. *Theor Appl Genet*, 2006, 112: 1248 - 1257.
- [15] 杨素丽, 明 军, 刘 春, 等. 基于 EST 信息的百合 SSR 标记的建立[J]. *园艺学报*, 2008, 35(7): 1069 - 1074.
- [16] 李小白, 张明龙, 崔海瑞. 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.
- [17] Morgante M, Hanafey M, Powell W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes[J]. *Nat Genet*, 2002, 30: 194 - 200.
- [18] 孔秋生. 基于公共序列数据库的 *Cucumis* 属 EST-SSR 标记的鉴定、开发和利用[D]. 武汉: 华中农业大学, 2006.
- [19] 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.
- [20] 陈军方, 任正隆, 高丽锋, 等. 从小麦 EST 序列中开发新的 SSR 引物[J]. *作物学报*, 2005, 31: 154 - 158.
- [21] Nicot N, Chiquet V, Gandon B, *et al.* Study of simple sequence repeat (SSRs) markers from wheat expressed sequence tags (ESTs) [J]. *Theor Appl Genet*, 2004, 109: 800 - 805.