

鹰嘴豆 EST 资源中 SSR 的信息分析

王显生¹, 贾钰莹¹, 顾汉艳¹, 何小玲¹, 张桦², 张巨松², 石书兵², 李建贵^{2*}, 麻浩^{1*}

(1.南京农业大学 作物遗传与种质创新国家重点实验室, 江苏 南京 210095; 2. 新疆农业大学 农业生物技术重点实验室, 新疆 乌鲁木齐 830052)

摘要: 从 NCBI 公共数据库下载获得 27 904 条鹰嘴豆表达序列标签 EST, 去除其中低质量的和冗余序列后得到全长为 6 198 092 bp 的 11 224 条无冗余 EST. 利用 SSRIT(Simple sequence repeat identification tool)软件共在这些序列中发掘出 982 个 SSR, 它们分布于 865 条 EST 中, 出现频率为 8.69%, 平均长度为 15.58 bp, 平均距离为 6.31 kb. 在鹰嘴豆 EST-SSR 中, 二核苷酸重复是最主要的重复类型(66.90%), 其次是三核苷酸重复类型(30.75%), 出现最多的重复基元类型是 TA/AT(24.95%).

关键词: 鹰嘴豆; 表达序列标签-简单重复序列; 频率; 特性

中图分类号: S529.024 文献标志码: A 文章编号: 1007-1032(2010)02-0137-05

Information analysis of EST resource of chickpea

WANG Xian-sheng¹, JIA Yu-ying¹, GU Han-yan¹, HE Xiao-ling¹, ZHANG Hua²,

ZHANG Ju-song², SHI Shu-bing², LI Jian-gui^{2*}, MA Hao^{1*}

(1.State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, China; 2.Key Laboratory of Agriculture Biotechnology, Xinjiang Agricultural University, Urumqi 830052, China)

Abstract: 27 904 ESTs of chickpea (*Cicer arietinum* L.) were downloaded from the database of NCBI and some redundant or low quality sequences were removed. Finally, 11 224 non-redundant ESTs with total length of 6 198 092 bp were obtained. SSRIT (Simple sequence repeat identification tool), a web tool for SSR hunting, was used to analyze the sequences. The results showed that a total of 982 SSRs distributed in 865 ESTs were detected, with the frequency for the non-redundant ESTs being 8.69%. The average length and distribution distance of the EST-SSRs were about 15.58 bp and 6.31 kb. Among the different SSRs ranging from 2 to 7 bp, dinucleotide repeats were most abundant (66.90%), followed by trinucleotide repeats (30.75%). Among 83 identified motif types, TA/AT had the highest EST frequency (24.95%). The results would contribute to the development of EST-SSR markers in chickpea.

Key words: chickpea; EST-SSR; frequency; characteristics

表达序列标签(expressed sequence tags, EST)是指通过对从 cDNA 文库随机挑选的克隆进行单向测序所获得的一段核酸序列, 长度为 60~500 bp, 其上携带着某个表达基因的部分遗传信息^[1]. 简单重复序列(simple sequence repeats, SSR), 也称卫星序列(microsatellite), 是遗传变异、群体结构、连

锁分析中使用十分广泛的分子标记之一. 但传统基因组 SSR 标记是建立在基因组文库构建的基础上的, 其开发耗时, 劳动强度大且费用高. 与此相比, 基于 EST 序列的 SSR 可以从公用 EST 数据库中获得, 相对简便经济, 而且由于 EST-SSR 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

收稿日期: 2009-09-01

基金项目: 国家科技支撑计划项目(2006BAD09A04, 2006BAD09A08, 2007BAC15B06); 国家自然科学基金项目(30860152); 国家博士后科学基金项目(20080431107); 高等学校学科创新引智计划项目(教育部); 科技支疆项目(200991254)

作者简介: 王显生(1977—), 男, 湖南永兴人, 博士, 主要从事作物种质资源研究与应用; *通讯作者, lijiangui1971@163.com; Lq-ncsi@njau.edu.cn

基因的直接鉴定提供了可能^[3]。

近年来,随着公用 EST 数据库中序列数的急剧增加,利用 EST 数据库获得 SSR,进而开发 EST-SSR 标记已成为新的研究热点。目前,EST-SSR 的研究和开发已经在水稻^[4]、小麦^[5-6]、棉花^[7]、大豆^[8]、玉米^[9]、油菜^[10]、花生^[11]、白菜^[12]、黄瓜^[13]、辣椒^[14]等作物中有报道。

鹰嘴豆(*Cicer arietinum* L.),属豆科鹰嘴豆属,一年生或多年生草本植物,主要种植于干旱、半干旱地区,是最早被人类驯化利用的豆科植物之一^[15]。作为干旱、半干旱地区的重要农作物,已有研究者采用 RAPD 技术进行了鹰嘴豆的品种鉴定和聚类分析^[16],笔者利用最新公布的鹰嘴豆 EST 数据库,分析鹰嘴豆 EST-SSR 发生频率和特点,以期为进一步建立鹰嘴豆 EST-SSR 标记和探索其在鹰嘴豆遗传育种研究中的应用奠定基础。

1 材料与方法

1.1 鹰嘴豆 EST 来源

鹰嘴豆 EST 序列来自美国国家生物技术信息中心(National Center for Biotechnology Information, NCBI)(<http://www.ncbi.nlm.nih.gov/sites/entrez>)数据库,选取 2009 年 8 月 1 日止公布的 27 904 条 EST 序列,其中 5 098 条由本课题组注册登记。

1.2 鹰嘴豆 EST-SSR 的筛选

利用软件 DNASTAR 中的 SeqMan II 模块对所有 EST 序列进行拼接,结合手工处理,最终获得 11 224 条无冗余 EST。再利用 SSRIT(Simple sequence repeat identification tool)软件(<http://www.gramene.org/db/markers/ssrtool>)在线搜索 EST-SSR。搜索的标准为:二核苷酸、三核苷酸、四核苷酸、五核苷酸、六核苷酸和七核苷酸重复基元(motif)的重复次数大于或等于 5。

2 结果与分析

2.1 鹰嘴豆 EST-SSR 的检出和长度

下载的 27 904 条鹰嘴豆 EST,其总长为 6 640.074 kb,GC 含量为 39.81%,平均长度为 237.96 bp。经利用软件 DNASTAR 中的 SeqMan II 模块处理后,

得到 11 224 条无冗余鹰嘴豆 EST,共计 6 181 998 bp,GC 含量为 40.47%,平均长度为 550.78 bp。

对 11 224 条非冗余 EST 进行 SSR 搜索的结果(表 1)表明,共获得 SSR 982 个,含有 SSR 的 EST 共计 865 条,占全部 EST 总数的 7.71%。在 865 条含有 SSR 的 EST 中,只含 1 个 SSR 的有 771 条,含 2 个 SSR 的有 82 条,含 3 个 SSR 的有 8 条,含 4 个 SSR 的有 2 条,含 6 个和 9 个 SSR 的 EST 均为 1 条。从检出情况来看,平均每 6.31 kb 就出现 1 个 SSR。

对鹰嘴豆 EST-SSR 长度的分析结果(表 1)表明,不同 SSR 类型的长度范围有很大差异,最短为 10 bp,最长达到 205 bp,平均长度为 15.58 bp。由图 1 可知,鹰嘴豆 EST-SSR 主要集中在长度 18 bp 以下,其中又以 10 bp 的最多,为 365 个,其次是 15 bp 的,为 177 个。

表 1 鹰嘴豆 EST-SSR 的数量、比例、频率、平均距离和重复长度

Table 1 The number, percent, frequency, average distance and repeat length of EST-SSR in chickpea

EST-SSR 类型	SSR 数量 /条	所占比例 /%	频率 /%	平均距离 /kb	长度范围 /bp
二核苷酸	657	66.90	5.82	9.43	10~74
三核苷酸	302	30.75	2.67	20.52	15~78
四核苷酸	7	0.71	0.06	885.44	20~36
五核苷酸	8	0.81	0.07	774.76	25~205
六核苷酸	7	0.71	0.06	885.44	30~36
七核苷酸	1	0.10	0.01	6 198.09	35
合计	982	100	8.69	6.31	

2.2 鹰嘴豆 EST-SSR 的类型、频率及次数分布

鹰嘴豆的 EST-SSR 类型较为丰富,二至七核苷酸重复类型都能看到,但频率各不相同(表 1)。其中,二核苷酸重复类型的 SSR 最多,共有 657 个,占全部 SSR 的 66.90%,出现频率为 5.82%;其次是三核苷酸重复类型,共有 302 个,占全部 SSR 的 30.75%,出现频率为 2.67%;四、五、六和七核苷酸重复类型在鹰嘴豆 EST 中的出现频率较低,共计 0.20%,占 SSR 总数的比例仅为 2.31%。可见,在鹰嘴豆 EST-SSR 中,二、三核苷酸重复类型占主导地位。

就鹰嘴豆 EST-SSR 重复类型的重复次数而言,

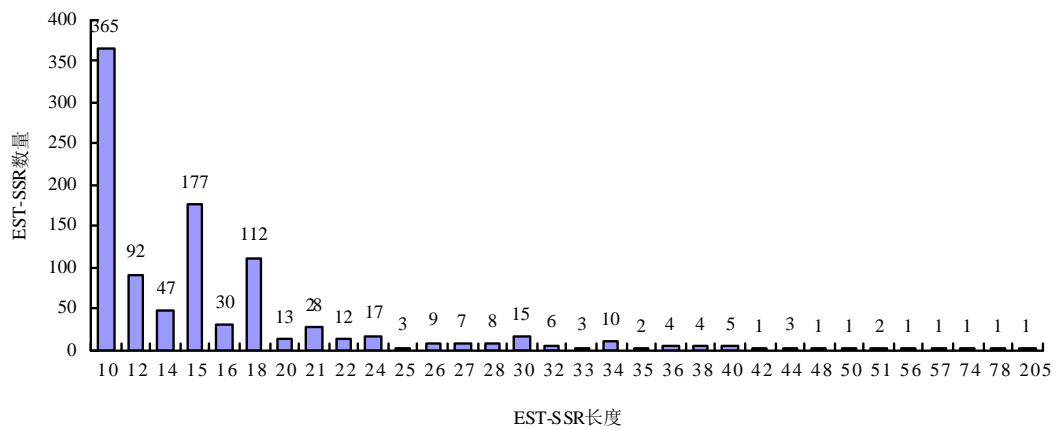


图 1 鹰嘴豆 EST-SSR 长度分布

Fig.1 The length distribution of EST-SSR in chickpea

在 5~22 次的重复范围内都有连续分布，其中以 5 次重复最多(556 个)，占 SSR 总数的 56.62%；其次为 6 次重复(172 个)；7 次以上重复则少得多，均未超过 10%。鹰嘴豆 EST-SSR 重复类型的重复次数最多的达到 41 次，但仅有一例，即 AAAAT 重复类型。

表 2 鹰嘴豆 EST-SSR 重复类型的次数分布

Table 2 Frequency of repeat type of EST-SSR in chickpea

重复次数	重复类型						合计	占总 SSR 比率/%
	二核苷酸	三核苷酸	四核苷酸	五核苷酸	六核苷酸	七核苷酸		
5	365	177	4	3	6	1	556	56.62
6	92	74	2	3	1	0	172	17.52
7	47	28	0	1	0	0	76	7.74
8	30	7	0	0	0	0	37	3.77
9	38	7	1	0	0	0	46	4.68
10	9	2	0	0	0	0	11	1.12
11	12	3	0	0	0	0	15	1.53
12	8	0	0	0	0	0	8	0.81
13	9	0	0	0	0	0	9	0.92
14	8	0	0	0	0	0	8	0.81
15	4	0	0	0	0	0	4	0.41
16	6	0	0	0	0	0	6	0.61
17	10	2	0	0	0	0	12	1.22
18	2	0	0	0	0	0	2	0.20
19	4	1	0	0	0	0	5	0.51
20	5	0	0	0	0	0	5	0.51
21	1	0	0	0	0	0	1	0.10
22	3	0	0	0	0	0	3	0.31
24	1	0	0	0	0	0	1	0.10
25	1	0	0	0	0	0	1	0.10
26	0	1	0	0	0	0	1	0.10
28	1	0	0	0	0	0	1	0.10
37	1	0	0	0	0	0	1	0.10
41	0	0	0	1	0	0	1	0.10
合计	657	302	7	8	7	1	982	100

2.3 鹰嘴豆 EST-SSR 重复基元的特性

在鹰嘴豆 EST-SSR 中,共观察到 83 种重复基元(表 3),其中以三核苷酸的 SSR 重复基元最多,为 50 种,但在所有可能的三核苷酸 SSR 重复基元中没有检出以下 8 种:AGG、CCG、CGC、CGT、GGC、GGT、GTC、TAC.其次为二核苷酸,有 12 种,四、五、六核苷酸的 SSR 重复基元分别为 7、7、6 种,七核苷酸的 SSR 重复基元最少,仅为 1 种,即 TCTTCTC.二核苷酸重复基元中以 TA/AT、TC/AG 和 CT/GA 基元为主,合计占二核苷酸型 SSR

总数的 84.32%, 占有检出 SSR 总数的 56.42%. 三核苷酸重复基元中以 TTC/AAG、TCT/AGA、TAT/ATA 和 AAT/TTA 较多,合计占三核苷酸型 SSR 总数的 31.46%, 占有检出 SSR 总数的 9.67%.

从表 3 中所列的不同重复基元可以看出,鹰嘴豆 EST-SSR 中没有三核苷酸 CCG/GGC 重复基元,而且,二核苷酸重复中 CG/GC 重复基元的数目也较少(均为 5 个),说明鹰嘴豆 EST 中缺乏连续的 GC 碱基区域;相比而言,AT 连续区域则较多.

表 3 鹰嘴豆 EST-SSR 重复基元的数量及比例
Table 3 The number and percent of repeat motif of EST-SSR in chickpea

EST-SSR 类型	重复基元	数量	占总 SSR 比率/%	EST-SSR 类型	重复基元	数量	占总 SSR 比率/%
二核苷酸	TA/AT	127/118	24.95	三核苷酸	TCC/AGG	4/0	0.41
	TC/AG	122/57	18.23		TGC/ACG	3/1	0.41
	CT/GA	103/27	13.24		GCA/CGT	3/0	0.31
	AC/TG	42/20	6.31		CGG/GCC	2/1	0.31
	CA/GT	23/8	3.16		CAG/GTC	2/0	0.20
	CG/GC	5/5	1.02		CGA/GCT	1/1	0.20
三核苷酸	TTC/AAG	20/11	3.16	四核苷酸	AGC/TCG	1/0	0.10
	TCT/AGA	12/10	2.24		AACA	1	0.10
	TAT/ATA	12/9	2.14		AATC	1	0.10
	AAT/TTA	11/10	2.14		AATT/TTAA	1/1	0.20
	CAA/GTT	14/3	1.73	ATAA	1	0.10	
	CAC/GTG	9/8	1.73	GAAA	1	0.10	
	TGA/ACT	15/2	1.73	TTCT	1	0.10	
	TAA/ATT	11/5	1.63	五核苷酸	GATTC	1	0.10
	CTT/GAA	8/8	1.63		GGAAA	1	0.10
	TGG/ACC	9/6	1.53		AAAAT/TTTAA	2/1	0.31
	AAC/TTG	6/6	1.22		AACAG	1	0.10
	GAT/CTA	11/1	1.22	AAGTA	1	0.10	
	TGT/ACA	7/5	1.22	ACCAC	1	0.10	
	TCA/AGT	9/1	1.02	六核苷酸	ATGTGA	1	0.10
	ATG/TAC	8/0	0.81		AAATGA	2	0.20
	ATC/TAG	6/1	0.71		GAAAGT	1	0.10
	CAT/GTA	6/1	0.71		GCTCCT	1	0.10
	CCT/GGA	6/1	0.71	GGAGCA	1	0.10	
	CCA/GGT	6/0	0.61	GTCATT	1	0.10	
	CTG/GAC	3/2	0.51	七核苷酸	TCTTCTC	1	0.10
CTC/GAG	2/2	0.41					

3 小 结

利用 11 224 条无冗余 EST 序列发掘出 982 个 EST-SSR,其出现频率为 8.69%,平均长度为 15.58

bp,平均距离为 6.31 kb.与其他作物 EST-SSR 的报道相比,鹰嘴豆 EST-SSR 的出现频率高于小麦(4.1%)^[5]、甘薯(4.4%)^[17]、棉花(4.46%)^[18]、玉米(7.25%)^[9]、花生(7.8%)^[11]、大豆(8.66%)^[8],表明鹰

嘴豆 EST 中含有较丰富的 SSR。

研究表明,多数作物的 EST-SSR 以三核苷酸和二核苷酸重复类型为主^[5, 10-14]。本研究结果表明,二核苷酸和三核苷酸重复类型是鹰嘴豆 EST-SSR 的主要类型,它们分别占 SSR 总数的 66.90% 和 30.75%,在观察到的 83 个重复基元中又以二核苷酸重复类型中的 TA/AT 重复基元出现次数最多。

参考文献:

- [1] 胡松年.基因表达序列标签(EST)数据分析手册[K].杭州:浙江大学出版社,2005:1.
- [2] Powell W, Machray G C, Provan J. Polymorphism revealed by simple sequence repeats[J]. Trends in Plant Science, 1996, 1(7): 215-222.
- [3] Varshney R K, Graner A, Sorrells M E. Genic microsatellite markers in plants: Features and applications [J]. Trends in Biotechnology, 2005, 23(1): 48-55.
- [4] Yu J K, Rota M L, Kantety R V, et al. EST derived SSR markers for comparative mapping in wheat and rice [J]. Molecular Genetics and Genomics, 2004, 271(6): 742-751.
- [5] 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].麦类作物学报,2005,31(2): 154-158.
- [6] 王盈盈,刘玉新,汪俊君,等.62 个小麦品种基于 EST-SSR 标记的遗传多样性分析[J].麦类作物学报,2008,28(5): 749-754.
- [7] 董薇,杜雄明,赖童飞.雷蒙德氏棉和亚洲棉 SSR 重复序列类型和丰度差异比较[J].棉花学报,2008,20(6): 418-424.
- [8] 詹少华,盛新颖,樊泓泓,等.大豆 EST 序列长度与 SSR 特性的关系[J].大豆科学,2009,28(2): 204-209.
- [9] 赵美琼,宋燕,范静华,等.玉米 EST 序列中微卫星的频率和分布[J].云南农业大学学报,2008,23(5): 585-589.
- [10] 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.
- [11] 梁炫强,洪彦彬,陈小平,等.花生栽培种 EST-SSRs 分布特征及应用研究[J].作物学报,2009,35(2): 246-254.
- [12] 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.
- [13] 胡建斌,李建吾.黄瓜基因组 EST-SSRs 的分布规律及 EST-SSR 标记开发[J].西北植物学报,2008,28(12): 2429-2435.
- [14] 李晶晶,王述彬,刘金兵,等.辣椒 EST-SSR 标记的开发[J].分子植物育种,2008,6(6): 1219-1222.
- [15] 王显生,高文瑞,石庆华,等.鹰嘴豆籽粒贮藏蛋白双向电泳的研究[J].湖南农业大学学报:自然科学版,2008,34(6): 635-639, 672.
- [16] 许磊,王希东,张桦,等.鹰嘴豆的 RAPD 品种鉴定和聚类分析[J].新疆农业大学学报,2008,31(1): 51-56.
- [17] 黄立飞,房伯平,陈景益,等.甘薯 EST 资源的 SSR 信息分析[J].植物生理学通讯,2009,45(1): 23-27.
- [18] 李华盛,范术丽,沈法富.从棉花 ESTs 数据库中筛选微卫星标记的初步研究[J].棉花学报,2005,17(4): 211-216.

责任编辑:罗慧敏

英文编辑:胡东平