

## 大豆 EST 序列长度与 SSR 特性的关系

詹少华<sup>1</sup>, 盛新颖<sup>2</sup>, 樊洪泓<sup>2</sup>, 蔡永萍<sup>2</sup>, 林毅<sup>2</sup>

(<sup>1</sup>皖西学院化学与生命科学系, 安徽六安 237012; <sup>2</sup>安徽农业大学生命科学学院, 安徽合肥 230036)

**摘要:**为了分析大豆 EST-SSR 的特征和比较其不同长度 EST 序列中的 SSR 特性, 从 NCBI 下载了大豆 EST 序列, 应用 SSRIT 搜索 SSR, 分析了 20 183 个无冗余的 EST 序列, 发现了 1747 个 EST 序列 (8.66%) 含有 SSR, 长度超过 400 bp 的 EST 序列含 SSR 的比例为 10.05%, 而长度在 400 bp 以下的 EST 序列 SSR 的比例为 4.74%。二核苷酸和三核苷酸 SSR 是大豆 EST-SSR 的主要类型, 分别的 60.45% 和 37.04%, 最常见的 SSR 基元是: AT、AG、AAT、AC、AAG、ACG。在 1 798 个 EST-SSR 中, 51.45% 的长度大于 12 bp, 13.35% 长度大于 20 bp。推断 NCBI 大豆 EST 数据库中含有 12 3527 个 SSR, 大豆 EST-SSR 可用于大豆分子标记, 具有重要的开发价值, 为有针对性设计 EST-SSR 引物奠定基础。

**关键词:**大豆; 表达序列标签; EST-SSR; 序列长度; 重复基元

**中图分类号:** S565.1

**文献标识码:** A

**文章编号:** 1000-9841(2009)01-0204-06

## Relationship between Length of Soybean ESTs Sequence and Characters of EST-SSR

ZHAN Shao-hua<sup>1</sup>, SHENG Xin-ying<sup>2</sup>, FAN Hong-hong<sup>2</sup>, CAI Yong-ping<sup>2</sup>, LIN Yi<sup>2</sup>

(<sup>1</sup>Chemistry and Biology Department, West Anhui University, Lu'an 237012, Anhui; <sup>2</sup>Life Science School, Anhui Agricultural University, Hefei 230036, Anhui, China)

**Abstract:** The objective of this study was to research the traits of simple sequence repeats (SSR) derived from expressed sequence tags (EST) of soybean and to compare the characters of SSRs in different length of ESTs, so 20183 no redundancy EST sequences were analyzed. The EST sequences were downloaded from NCBI, and SSRs were screened using SSRIT program. The results showed 1747 ESTs (8.66%) contained one or more SSRs. According to EST sequence length over or less than 400bp, the SSRs rates were 10.05% and 4.74%, respectively. Di-nucleotide repeats (60.45%) were the most abundant type of repeats followed by Tri-nucleotide repeats (37.04%). AT, AG, AAT, AC, AAG, ACG were the most common SSR repeat motifs. In the 1798 EST-SSRs, the length of EST-SSR over 12bp accounted for 51.45 percent, and proportion of the length more than 20bp accounted for 13.35 percent. The average distance of distribution was 7.91 (1/kb). 123527 SSRs were deduced in EST database of NCBI, so we claimed exploration value of the EST-SSRs for the soybean molecular marker. The results would contribute to EST-SSR primer efficient design.

**Key words:** Soybean; EST; EST-SSR; Sequence length; Repeat Contig

SSR (Simple Sequence Repeats, 简单重复序列) 具有在基因组中分布广, 多态性好, 表现为复等位共显性的孟德尔式遗传方式, SSR 分子标记由于重复性好、SSR 序列的两侧具有保守序列便于设计引物, 试验成本相对较低<sup>[1]</sup>, 因此 SSR 分子标记成为当前应用最广泛的分子标记之一<sup>[2-4]</sup>, 但是传统的 SSR 引物开发依赖于基因组测序导致开发成本较高。随

着公用 EST (Expressed Sequence Tag, 表达序列标签) 数据库中序列数不断增多, 利用 EST 开发 EST-SSR 引物成为新的研究热点, 目前已经对甘蔗<sup>[5]</sup>、小麦<sup>[6]</sup>、大麦<sup>[7]</sup>、水稻、高粱<sup>[8]</sup>、油菜<sup>[9]</sup>、茶树<sup>[10]</sup>、棉花<sup>[11]</sup>、烟草<sup>[12]</sup> 等作物进行了 EST-SSR 方面的分析。

EST-SSR 除了基因组 SSR 的优点外, 还反映

收稿日期: 2008-12-17

基金项目: 安徽高校省级自然科学研究重点项目资助 (KJ2008A089)。

作者简介: 詹少华 (1963-), 男, 副教授, 博士, 研究方向为大豆和棉花分子育种。E-mail: zhansh@wxc.edu.cn。

通讯作者: 林毅, 教授, 博士生导师。E-mail: linyiahau@126.com。

了基因的编码部分,这有可能对决定重要表型性状的等位基因进行直接鉴定<sup>[13]</sup>,EST-SSR 物种间通用性高,一旦建立就可以在一系列相关种中应用<sup>[14]</sup>。所以,它在遗传作图、资源多样性、功能基因的发现与定位、物种起源与进化和比较基因组学研究等方面都有重要的利用价值<sup>[15-17]</sup>。

根据排列组合原理,当四种不同的碱基随机组合时,具有  $i$  个碱基的核心基元的种类数为  $4^i$  个,但由于约数重复、单碱基重复、移码重复等,实际上核心基元的种类数远小于  $4^i$ ,由于双链互补关系 AC 与 GT 是同一种核心基元,由于移码关系 AC 与 CA 也是同一种核心基元,这样 AC、GT、CA 和 TG 都是同一个核心基元,所以基本重复基元类型为:2 种单核苷酸、4 种二核苷酸、10 种三核苷酸、33 种四核苷酸、102 种五核苷酸和 350 种六核苷酸<sup>[8,17]</sup>。

2004 年 Song 等<sup>[18]</sup>在构建大豆分子标记的遗传图时采用了 24 个来源于 EST 或基因库的 SSR 引物,证明了大豆 EST-SSR 引物的应用价值,目前大豆的 EST 数据库序列总数已经超过一百万个,大豆 EST-SSR 位点的分布、频率、长度等特征尚缺乏系统研究,其他物种的 EST 序列 SSR 位点分析也缺乏不同长度的 EST 序列的比较研究。试验的目的在于分析大豆 EST-SSR 的总体特征,比较不同长度 EST 序列的 SSR 特性,评价大豆 EST-SSR 的应用价值,以期为设计大豆 EST-SSR 引物提供理论依据。

## 1 材料与方法

### 1.1 大豆 EST 序列来源和处理

从 NCBI(美国国立生物技术信息中心)的 EST 数据库(<http://www.ncbi.nlm.nih.gov/dbEST/>)查询目前(2008 年 12 月 1 日)大豆 EST 序列的总数,按照不同的核苷酸长度(bp)( $\leq 200$ , 201~400, 401~600, 601~800, 801~1 000, >1 000)进行分组,分别查询每个组 EST 序列的数量,核苷酸长度小于 200 bp 的 EST 序列不适合 SSR 位点引物开发,核苷酸长度大于 200 bp 的 EST 序列下载后采用 ClustalX 软件去除冗余的序列,核苷酸长度 201~1 000 bp 的 EST 序列取出 5 000 个无冗余序列用于分析,核苷酸长度大于 1 000 bp 总数少于 5 000,分析全部序列,DNA Counter 软件计算各碱基的比例。

### 1.2 EST-SSR 的发掘

应用 SSRIT(<http://www.gramene.org/db/sea->

[ches/ssrtool](http://www.gramene.org/db/sea-))在线搜索 SSR,检索标准为重复基元 2, 3, 4, 5, 6, 最小重复次数为 5。

### 1.3 EST-SSR 重复基元分类汇总统计

由于分析的序列数较多,手工汇总难以完成且容易发生错误,将 EST-SSR 查询结果导入到 Excel2003,利用 Excel2003 高级筛选功能获得 EST-SSR 重复基元的种类数,LEN 函数提取重复基元的碱基数,再利用 SUMIF、COUNTIF 函数统计相关指标,MIN/MAX 与 IF 组合取得相关指标的最小值和最大值。

## 2 结果与分析

### 2.1 EST 的序列长度与含 SSR 序列数

2008 年 12 月 1 日从 NCBI 共查询到大豆 EST 序列 1386618 个,其中 201~400 bp 的 EST 序列最多,接近总数的一半(46.9%),其次是小于或等于 200 bp 的 EST 序列,占总数的 25.0%,序列长度大于 800 bp 后,EST 的数目明显减少,801~1 000 bp 的 EST 序列只有总数的 1.29%,大于 1 000 bp 的 EST 序列略高于总数的万分之一(0.0132%)。一般认为 EST 长度多在 700 bp 以下,大豆的 EST 也符合这个特征,但是由于 EST 序列数越来越多,长度大于 700 bp 的 EST 序列的总数也是不可忽视的。共发现 1 747 个 EST 序列含有 SSR 位点,占被分析序列的 8.66%,其中 400 bp 以上序列含 SSR 的比例为 10.05%,400 bp 以下的 EST 序列含 SSR 的比例明显低于 400 bp 以上序列(4.74%)。从 SSR 的分布频率来看,每 7.91 kb 核苷酸含有 1 个 SSR,但是不同长度的 EST 序列 SSR 分布频率不同,800 bp 以下的序列分布频率数值小,说明更少的碱基数就含有 1 个 SSR,801~1 000 bp 序列分布频率的数值是 401~600 bp 序列的 2.27 倍。大豆 EST 碱基比例:AT 为 57.44%,GC 为 42.60%。

801~1 000 bp 的序列中,3 个序列含有 4 个 SSR,7 个序列含有 3 个 SSR,47 个序列含有 2 个 SSR,其他 EST 序列都只有 1 个 SSR。SSR 数目占被分析序列数的 8.91%。

按照重复基元核苷酸数分类,重复基元核苷酸数越小的 SSR 出现的频率越高,二核苷酸重复基元的 SSR 最多,占总数的 60.45%,三核苷酸重复基元出现概率次之,占总数的 37.04%,四核苷酸、五核苷酸和六核苷酸重复基元之和只有总数的 2.50%。33 种四核苷酸重复基元出现了 21 种,分

别是 cttt、gatt、tata、aacc、atta、attg、taaa、aaag、ctca、ttat、cacc、agaa、taat、tttc、cgca、tcat、gggt、acgc、aaag、acca、ttct。102 种五核苷酸重复基元出现了 5 种,

分别是 gtttg、cacaa、tttgg、taaaa、ctatgg, 350 种六核苷酸只出现 6 种, 分别是 atgggt、acgcaa、attttc、ccagca、ccctaa、atttca。

表 1 EST 的序列长度与含 SSR 序列数的关系

Table 1 Relationship between the EST sequence length and number of SSR

序列长度 Sequence length/bp	EST 序列数 No. of EST	被分析的 EST 数 No. of EST analyzed	含 SSR 的序列数 No. of SSR	含 SSR 的序列比例 Rate of SSR/%	SSR 分布频率 Frequency of SSR distribution(1/kb)
≤200	347012				
201 ~ 400	650464	5000	237	4.74	6.33
401 ~ 600	258492	5000	539	10.78	4.64
601 ~ 800	112447	5000	524	10.48	6.68
801 ~ 1000	18020	5000	428	8.56	10.51
> 1000	183	183	19	10.38	11.39
平均 Mean				8.66	7.91
合计 Total	1386618	20183	1747		

表 2 EST - SSR 数目统计

Table 2 Accounting for number of EST - SSR

序列长度 Sequence Length/ bp	核苷酸数 Number of nucleotide																	合计 Total
	2					3					4					5	6	
	AC	AG	AT	CG	AAC	AAG	AAT	ACC	ACG	AGC	AGG	AGT	ATG	CGG				
201-400	14	87	42	0	6	29	10	14	1	9	6	5	5	4	3	0	2	237
401-600	42	216	62	0	17	54	16	28	7	30	14	20	9	13	9	0	2	539
601-800	41	217	61	0	22	62	19	13	3	24	10	15	20	3	8	4	2	524
801-1 000	37	195	72	1	30	66	10	14	2	17	9	3	26	1	6	5	4	498
合计 Total	134	715	237	1	75	211	55	69	13	80	39	43	60	21	26	9	10	1798

## 1.2 EST - SSR 数目统计

从重复基元个体来看, (AG)<sub>n</sub>、(AT)<sub>n</sub>、(AAG)<sub>n</sub>和 (AC)<sub>n</sub> 是最主要的 SSR, 分别占总数的 39.77%、13.18%、11.74% 和 7.45%, 其他各种 SSR 只有总数的 27.86%。4 种二核苷酸重复基元都能够出现, 但是 (GC)<sub>n</sub> 只出现 1 次, 800 bp 以下的序列中没有 (GC)<sub>n</sub>。10 种三核苷酸重复基元也都能够出现, 其中 (AAG)<sub>n</sub> 的出现频率最高, 最少的是 (ACG)<sub>n</sub>。

从不同长度的 EST 序列来看, 400 bp 以下的序列含 SSR 比例明显低于 400 bp 以上的序列, 而 400 bp 以上的 EST 序列含 SSR 的比例较接近。1 798 个 SSR 中, 201 ~ 400 bp 的 EST 序列占 13.18%, 其他各组平均占有 28.94%。

## 1.3 EST - SSR 长度统计

从表 3 可以看出, 总体来看, 重复基元的碱基越多, EST - SSR 的长度越大, 在重复基元的碱基数相等的情况下, 不同 EST 序列长度的 SSR 长度相接近。

二核苷酸重复基元中最大长度 (AT)<sub>n</sub> 为 98 bp, 三核苷酸重复基元中最大长度 (AAT)<sub>n</sub> 为 48

bp。二核苷酸重复基元中, (AT)<sub>n</sub> 长度变异范围 10 ~ 98 bp 最大, 三核苷酸重复基元中长度变异范围最大是 (AAT)<sub>n</sub>, 达到 15 ~ 48 bp。

在 1 798 个 EST - SSR 中, 有 925 个长度大于 12 bp, 占 51.45%, 有 240 个长度大于 20 bp, 占 13.35%。

## 1.4 EST - SSR 的重复数

基元的重复数是 SSR 的重要特征, 大豆 EST - SSR 重复数总平均为 5.86, 从平均值来看, 不同长度的 EST 序列和不同长度核苷酸基元的重复数都不大于 6.28。

基元重复数最多的 SSR 是: (AT)<sub>n</sub>、(AG)<sub>n</sub>、(AAT)<sub>n</sub>、(AC)<sub>n</sub>、(AAG)<sub>n</sub>、(ACG)<sub>n</sub>。其中 201 ~ 400 bp 的 EST 序列中, 前 5 名 SSR 依次为: (AT)<sub>n</sub>、(AG)<sub>n</sub>、(AAG)<sub>n</sub>、(AAC)<sub>n</sub> 和 (ACC)<sub>n</sub>; 401 ~ 600 的 EST 序列中, 前 5 名 SSR 依次为: (AG)<sub>n</sub>、(AC)<sub>n</sub>、(AT)<sub>n</sub>、(AAG)<sub>n</sub>、(ACG)<sub>n</sub>; 601 ~ 800 的 EST 序列中, 前 5 名 SSR 依次为: (AT)<sub>n</sub>、(AG)<sub>n</sub>、(AAT)<sub>n</sub>、(ACG)<sub>n</sub>、(AC)<sub>n</sub>; 801 ~ 1 000 的 EST 序列中, 前 5 名 SSR 依次为: (AT)<sub>n</sub>、(AG)<sub>n</sub>、(AAG)<sub>n</sub>、(AC)<sub>n</sub>、(ACC)<sub>n</sub>。

表 3 大豆 EST – SSR 的长度  
Table 3 Length of soybean EST – SSR sequence

		序列长度 Sequence length/bp									
		201-400		401-600		601-800		801-1000		201-1000	
		范围	平均	范围	平均	范围	平均	范围	平均	范围	平均
		Range	Mean	Range	Mean	Range	Mean	Range	Mean	Range	Mean
二核苷酸 Di – nucleotide	AC	10-28	11.9	10-28	12.5	10-20	12.1	10-22	11.8	10-28	12.0
	AG	10-44	13.1	10-40	13.5	10-46	13.7	10-32	13.6	10-46	13.4
	AT	10-70	15.1	10-22	12.8	10-54	14.3	10-98	16.9	10-98	14.8
	CG							10-10	10	10-10	10.0
平均 Mean			13.37		12.93		13.37		13.07		12.55
三核苷酸 Tri – nucleotide	AAC	15-21	17.0	15-24	18.0	15-21	17.3	15-24	17.7	15-24	17.5
	AAG	15-39	18.0	15-33	18.5	15-33	17.6	15-42	18.7	15-42	18.2
	AAT	15-24	17.4	15-21	16.9	15-48	20.4	15-21	17.4	15-48	18.0
	ACC	15-21	15.9	15-27	17.4	15-24	16.8	15-24	16.1	15-27	16.5
	ACG	18-18	18.0	15-27	17.6	15-18	16.0	15-21	18.0	15-27	17.4
	AGC	15-27	17.3	15-24	16.1	15-21	16.3	15-18	15.5	15-27	16.3
	AGG	15-24	17.5	15-21	16.3	15-21	16.8	15-21	17.0	15-24	16.9
	AGT	15-18	16.8	15-24	16.7	15-18	15.6	15-21	18.0	15-24	16.8
	ATG	15-21	16.2	15-24	18.0	15-21	16.7	15-21	17.2	15-24	17.0
	CGG	15-15	15.0	15-21	16.2	15-15	15.0	15-15	15.0	15-15	15.3
平均 Mean			16.91		17.17		16.85		17.06		16.99
四核苷酸 Tetra – nucleotide		24-64	37.3	20-28	20.7	20-24	21.8	20-24	20.7	24-64	25.1
五核苷酸 Hexa – nucleotide						25-25	25	30-35	32.5	25-35	28.8
六核苷酸 Hept – nucleotide		30-30	30	30-30	30	30-30	30	30-30	30	30-30	30.0

表 4 EST – SSR 的重复数统计  
Table 4 Accounting for number of EST – SSR repeats unit

		序列长度 Sequence length/bp									
		201-400		401-600		601-800		801-1000		合计 Total	
		范围	平均	范围	平均	范围	平均	范围	平均	范围	平均
		Range	Mean	Range	Mean	Range	Mean	Range	Mean	Range	Mean
二核苷酸 Di – nucleotide	AC	5-14	5.9	5-14	6.2	5-10	6.0	5-11	5.9	5-14	6.0
	AG	5-22	6.6	5-20	6.8	5-23	6.7	5-16	6.7	5-23	6.7
	AT	5-35	7.6	5-11	6.4	5-27	7.2	5-49	8.4	5-35	7.4
	CG							5-5	5	5-5	5
三核苷酸 Tri – nucleotide	AAC	5-7	5.7	5-8	6	5-7	5.8	5-8	5.9	5-8	5.8
	AAG	5-13	6	5-11	6.2	5-11	5.9	5-14	6.2	5-14	6.0
	AAT	5-8	5.8	5-7	5.6	5-16	6.8	5-7	5.8	5-16	6.0
	ACC	5-7	5.3	5-9	5.8	5-8	5.6	5-8	5.4	5-9	5.5
	ACG	6-6	6	5-9	5.6	5-6	5.3	5-7	6.0	5-9	5.8
	AGC	5-9	5.8	5-8	5.4	5-7	5.4	5-6	5.2	5-9	5.4
	AGG	5-8	5.8	5-7	5.4	5-7	5.6	5-7	5.7	5-8	5.6
	AGT	5-6	5.6	5-8	5.6	5-6	5.2	5-7	6	5-8	5.6
	ATG	5-7	5.4	5-8	6	5-7	5.6	5-7	5.7	5-8	5.7
	CGG	5-5	5	5-7	5.4	5-5	5	5-5	5	5-7	5.1
四核苷酸 Tetra – nucleotide		6-16	9.3	5-7	5.5	5-6	5.4	5-5	5	5-16	6.3
五核苷酸 Hexa – nucleotide						5-5	5	7-7	7	5-7	6
六核苷酸 Hept – nucleotide		5-5	5	5-5	5	5-5	5	5-5	5	5-5	5

### 3 讨论

通过分析可知,大豆含 SSR 的 EST 序列占 8.66%,据此推测大豆 EST 数据库中共有 122 937 余个 EST 序列含有 SSR,SSR 数目占被分析序列数的 8.91%,大豆 EST 数据库中含有 123 527 个 SSR。Temnykh 等<sup>[19]</sup>报道,长度在 12 ~ 20 bp 之间的 SSR 可以应用于多态性分析,SSR 长度在 20 bp 以上时多态性表现更加丰富,大豆 EST - SSR 长度大于 20 bp 的占 13.35%,大于 12 bp 占 51.45%,这样可以推算出大豆 EST 数据库中长度大于 20 bp 的 SSR 有 18 961 个,大于 12 bp 的 SSR 有 63 555 个,所以大豆 EST - SSR 开发前景广阔,必然在大豆的分子标记连锁图的构建、种质资源鉴定、进化关系分析等方面发挥重要作用。

二核苷酸和三核苷酸 SSR 是大豆 EST - SSR 的主要类型,占 97.49%,其中二核苷酸重复基元的 SSR 最多,占总数的 60.45%,二核苷酸 SSR 中,出现数目最多的是 (AG)<sub>n</sub>、(AT)<sub>n</sub>,同时重复的次数也多,但是 (GC)<sub>n</sub> 只占二核苷酸 SSR 的 0.092%。Dreisigacker 等<sup>[20]</sup>发现低级基元的 SSR 多态性较高,大豆 EST - SSR 中,二核苷酸和三核苷酸 SSR 比例特别高,这也为开发大豆 EST - SSR 创造了很好的条件。

大豆 EST 序列长度不同,EST - SSR 的特性具有差别,400 bp 以上 EST 序列含 SSR 的比例 (10.05%) 明显高于 400 bp 以下的 EST 序列 (4.74%),400 bp 以上各组 EST 序列的 SSR 数目比例平均值 (28.94%) 也明显高于 201 ~ 400bp 以下的 EST 序列 (13.18%),所以开发 SSR 引物时,下载 400 bp 以上的 EST 序列会显著提高试验效率。

首次统计了 SSR 基元重复数变异范围,推测 SSR 基元重复数变异范围越大,分子标记的多态性可能越强,因为 SSR 来源于复制过程中 DNA 复制酶的滑动或同源染色体不等价交换,而 SSR 基元重复数发生变异也可能是这些原因造成的。这一点如果被将来的试验证实,有助于提高设计 SSR 引物针对性。

### 参考文献

[1] Powell W, Machray G C, Provan J. Polymorphism revealed by simple sequence repeats [J]. Trends Plant Science, 1996, 1: 215 - 222.

[2] 杨喆,刘丽君,高明杰,等. 大豆高蛋白基因分子标记及其在大豆育种中的应用[J]. 大豆科学,2008,27(2):186 - 189. (Yang Z, Liu L J, Gao M J, et al. QTL Tagging for high protein gene and using molecular marker assistant selection in soybean breeding [J]. Soybean Science, 2008, 27(2):186 - 189. )

[3] 王俊,王林林,刘章雄,等. 大豆地方品种遗传结构及其保存研究[J]. 大豆科学,2008,27(3):361 - 365. (Wang J, Wang L L, Liu Z X, et al. Genetic structure and conservation of soybean landraces [J]. Soybean Science, 2008, 27(3):361 - 365. )

[4] 汤复跃,周立人,程潇,等. 大豆 M 型细胞质雄性不育恢复基因 SSR 标记初步定位[J]. 大豆科学,2008,27(3):383 - 386. (Tang F Y, Zhou L R, Cheng X, et al. SSR marker location for fertility restorer gene of M - cytoplasmic male sterility in soybean [J]. Soybean Science, 2008, 27(3):383 - 386. )

[5] Cordeiro G M, Casu R, McIntyre C L, et al. Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum [J]. Plant Science, 2001, 160:1115 - 1123.

[6] Eujayl I, Sorrells M E, Baum M, et al. Assessment of genotypic variation among cultivated durum wheat based on EST - SSRs and genomic SSRs [J]. Euphytica, 2001, 119:39 - 43.

[7] Thiel T, Michalek W, Varshney R K, et al. Exploiting EST databases for the development and characterization of gene - derived SSR - markers in barley (*Hordeum vulgare* L.) [J]. Theoretical and Applied Genetics, 2003, 106:411 - 422.

[8] 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.

[9] 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. (Li X B, Zhang M L, Cui H R. Analysis of SSR information in EST resource of oilseed rape [J]. Chinese Journal of Oil Crop Sciences, 2007, 29(1):20 - 25. )

[10] 金基强,崔海瑞,龚晓春,等. 用 EST - SSR 标记对茶树种质资源的研究[J]. 遗传,2007,29(1):103 - 108. (Jin J Q, Cui H R, Gong X C, et al. Studies on tea plants (*Camellia sinensis*) germplasms using EST - SSR marker [J]. Hereditas, 2007, 29(1):103 - 108. )

[11] 林忠旭,王锦峰,张献龙. 瑟伯氏棉和异常棉的陆地棉导入系的 EST - SSR 和 gSSR 分析[J]. 棉花学报,2008,20(4):243 - 248. (Lin Z X, Wang J F, Zhang X L. Characteristics of *Gossypium thurberi* and *G. anomalum* Introgression Lines of *G. hirsutum* Revealed by EST - SSR and gSSR [J]. Cotton Science, 2008, 20(4):243 - 248. )

[12] 张俊娥,李芬,孙蒙祥. 烟草 EST - SSR 位点分析[J]. 武汉植物学研究,2007,25(5):427 - 431. (Zhang J E, Li F, Sun M X. Analysis of EST - SSR loci in tobacco (*Nicotiana tabacum* L.) [J]. Journal of Wuhan Botanical Research, 2007, 25(5):427 - 431. )

[13] Varshney R K, Graner A, Sorrells M E. Genic microsatellite markers in plants: features and applications [J]. Trends Biotechnology, 2005, 23, 48 - 55.

- [14] Guo W, Cai C, Wang, C, et al. A microsatellite based gene – rich linkage map reveals genome structure, function and evolution in *Gossypium* [J]. *Genetics*, 2007, 176: 527 – 541.
- [15] 李永强, 李宏伟, 高丽锋, 等. 基于表达序列标签的微卫星标记 (EST – SSRs) 研究进展 [J]. *植物遗传资源学报*, 2004, 5 (1): 91 – 95. ( Li Y Q, Li H W, Gao L F, et al. Progress of simple sequence repeats derived from expressed sequence tags [J]. *Journal of Plant Genetic Resources*, 2004, 5 (1): 91 – 95. )
- [16] 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.
- [17] Rota L R, Kantety R V, Yu J K. Nonrandom distribution and frequencies of genomic and EST – derived microsatellite markers in rice, wheat and barley [J]. *BMC Genomics*, 2005, 6: 23.
- [18] Song Q J, Marek L F, Shoemaker R C, et al. A new integrated genetic linkage map of the soybean [J]. *Theoretical and Applied Genetics*, 2004, 109: 122 – 128.
- [19] Temnykh S, De Clerck G, Lukashova A, et al. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.) [J]. *Genome Research*, 2001, 11: 1441 – 1452.
- [20] Dreisigacker S, Zhang P, Warburton M L, et al. SSR and pedigree analyses of genetic diversity among CIM – MYT wheat lines targeted to different megaenvironments [J]. *Crop Science*, 2004, 44: 381 – 388.
- 
- (上接第 203 页)
- [3] Chu Y W, Sun C T. Regularity of secondary protein structures: A genetic algorithm approach [C]. *Proceedings of the 5<sup>th</sup> World Congress on Intelligent Control and Automation*, 2004: 2104–2108.
- [4] 邹先霞, 陈孝卫, 许龙飞. 基于关联规则与遗传算法的蛋白质二级结构预测 [J]. *计算机工程与应用*, 2006, 42 (15): 152–156. (Zou X X, Chen X W, Xu L F. Protein secondary structure prediction based on association rules and genetic algorithm [J]. *Computer Engineering and Applications*, 2006, 42 (15): 152–156. )
- [5] 张宁, 张涛. 蛋白质二级结构预测样本集数据库的设计与实现 [J]. *生物信息学*, 2006, 4 (4): 163–166. (Zhang N, Zhang T. Design and implementation of the database of data sets for prediction of protein secondary structures [J]. *China Journal of Bioinformatics*, 2006, 4 (4): 163–166. )
- [6] Chu Y W, Sun C T. A hybrid genetic algorithm approach for protein secondary structure [C]. *Proceedings of the 6<sup>th</sup> World Congress on Intelligent Control and Automation*, 2006: 3320–3324.
- [7] 杜耀华, 王正志, 倪青山. 基于滑动窗口的原核转录起始点计算定位方法 [J]. *生物物理学报*, 2006, 22 (5): 360–366. (Du Y H, Wang Z Z, Ni Q S. Computational location of transcription start sites in prokaryotic genome based on sliding window [J]. *Acta Biophysica Sinica*, 2006, 22 (5): 360–366. )
- [8] 黄敏, 沈辉, 肖奕. 不同类蛋白质氨基酸疏水序列周期性 [J]. *生物物理学报*, 2000, 16 (4): 755–760. (Huang M, Shen H, Xiao Y. Correlation properties of protein sequences of different types [J]. *Acta Biophysica Sinica*, 2000, 16 (4): 755–760. )
- 

## 立足黑龙江 辐射全中国 聚焦大农业 促进快发展

### 欢迎订阅 2009 年《黑龙江农业科学》

《黑龙江农业科学》是黑龙江省农业科学院主办的综合性科技期刊。内容丰富、栏目新颖、信息量大、可读性强,读者群大、发行面广,是全国优秀期刊、黑龙江省优秀期刊。现已被《中国科学引文数据库》《中国核心期刊(遴选)数据库》、CNKI 系列数据库、万方数据库、重庆维普中文科技期刊数据库和华艺电子出版事业群等多家权威数据库收录。

《黑龙江农业科学》为双月刊,单月 10 日出版,国内外公开发行。国内邮发代号 14 – 61,每期定价 8.00 元,全年 48.00 元;国外由中国国际图书贸易总公司发行,发行代号 BM8321,每期定价 8.00 美元,全年 48.00 美元。

热忱欢迎广大农业科研工作者、农业院校师生、国营农场及农业技术推广人员、管理干部和广大农民群众踊跃订阅。全国各地邮局均可订阅。漏订者可汇款至本刊编辑部补订。汇款写明订购份数,收件人姓名、详细邮寄地址及邮编。

另外,编辑部现有少量 2007 年合订本珍藏版。每册 80.00 元,邮费 10.00 元,共计 90.00 元,售完为止。

地址:哈尔滨市南岗区学府路 368 号《黑龙江农业科学》编辑部

邮编:150086 电话:0451 – 86668373 电子函件:nykx13579@sina.com