

## 绒毛状烟草基因组中 SSR 位点的信息分析

王卫锋<sup>1</sup>, 晁江涛<sup>1</sup>, 龚达平<sup>1</sup>, 王 鲁<sup>1</sup>, 李凤霞<sup>1</sup>, 陈雅琼<sup>1,2</sup>, 孙玉合<sup>1\*</sup>

(1. 农业部烟草类作物质量控制重点开放实验室, 中国农业科学院烟草研究所, 青岛 266101; 2. 中国农业科学院研究生院, 北京 100081)

**摘要:** 通过 SSRFINDER 软件搜索绒毛状烟草基因组 DNA, 设置条件是重复单元长度 2~5 bp, 重复次数 10~60, 共找到 6 441 个 SSR 位点, 涉及 5 227 个 Scaffold, SSR 序列总长为 303 963 bp, 约占全基因组的 0.12%, 即绒毛状烟草基因组中平均每 400 kb 会出现一个 SSR 位点。在所有 SSR 位点中, 二、三碱基重复基元分别占总数的 45.2% 和 49.6%, 重复单元的重复次数 10~20 之间的 SSR 位点占总数的 74.6%。

**关键词:** 绒毛状烟草; 简单重复序列; 基元

中图分类号: TS413

文章编号: 1007-5119 (2011) 03-0032-04

DOI: 10.3969/j.issn.1007-5119.2011.03.007

## Analysis of SSR Information in *N. tomentosiformis* Goodspeed genome

WANG Weifeng<sup>1</sup>, CHAO Jiangtao<sup>1</sup>, GONG Daping<sup>1</sup>, WANG Lu<sup>1</sup>, LI Fengxia<sup>1</sup>, CHEN Yaqiong<sup>1,2</sup>, SUN Yuhe<sup>1\*</sup>

(1. Key Laboratory of Tobacco Quality Control, Ministry of Agriculture, Tobacco Research Institute of CAAS, Qingdao 266101, China; 2. Graduate School of CAAS, Beijing 100081, China)

**Abstract:** By searching *N. tomentosiformis* Goodspeed genome DNA with SSRFINDER, 6441 SSRs were identified from 5227 scaffold. Total length 303 963bp of SSRs occupied approximately 0.12% of genome DNA. The average distance between SSRs was about 400 kb. Dinucleotide and trinucleotide motifs were most abundant classes, 45.2% and 49.6% respectively. The repeats of motifs from 10-20 accounted for 74.6% of the total SSRs.

**Keywords:** *N. tomentosiformis* Goodspeed; SSR; motif

简单重复序列 (simple sequence repeats, SSR) 又称卫星 DNA、短串联重复序列, 一般为以 1~6 个碱基为核心序列, 首尾相连组成的串联重复序列<sup>[1]</sup>。SSR 位点广泛存在于真核生物基因组中, 不同遗传材料在重复次数上的差异可导致 SSR 位点长度的多态性, 在 SSR 位点两侧有一段相当保守的单拷贝序列, 根据此保守序列设计 PCR 引物能够精确检测出 SSR 位点的长度多态性。以 SSR 位点的长度多态性为基础, 利用 PCR 和电泳技术开发出的 SSR 分子标记具有对基因组覆盖度好、数量丰富、多态性高、共显性标记、实验简单以及重复性好等特点<sup>[2]</sup>。

SSR 分子标记已在植物遗传分析中得到广泛应用, 如遗传作图、分子标记辅助选择、群体遗传分

析以及分类学等方面, 在花生、玉米、大麦<sup>[3-5]</sup>等作物中利用 SSR 标记已构建了遗传连锁图谱; Korzun 等<sup>[6]</sup>利用 SSR 标记对小麦矮秆基因 Rht8 进行了定位; Prevost 等<sup>[7]</sup>对 34 个马铃薯品种构建了 SSR 指纹图谱; SSR 分子标记在小麦、大麦等作物的遗传多样性分析中也得到了成功应用<sup>[8-9]</sup>。另外, 在烟草中也利用 SSR 标记开展了相关研究, 如范静苑等<sup>[10]</sup>成功运用 SSR 分子标记分析了烟草 CMV 抗性基因的遗传连锁距离; Moon 等<sup>[11]</sup>运用 SSR 分子标记分析了 54 份烟草属材料的遗传多样性。

烟草分子标记研究多以 RAPD 和 AFLP 为主, 主要应用于野生种质的抗病基因鉴定<sup>[12-18]</sup>, 限于这些分子标记多态性较低, 用其构建烟草遗传图谱受到了很大限制<sup>[19-21]</sup>。对烟草基因组的测序极大地促

作者简介: 王卫锋, 男, 助理研究员, 从事烟草分子育种工作。E-mail: sdwangweifeng@163.com。\* 通信作者, E-mail: yhsun@163.com

收稿日期: 2010-03-30

进了烟草 SSR 分子标记的开发, Bindler 等<sup>[22]</sup>从 Tobacco Genome Initiative (TGI) 公布的烟草基因组序列信息中搜索到 282 个 SSR 分子标记, 并利用这些标记构建了第一张烟草的遗传连锁图。在本研究中, 通过软件搜索绒毛状烟草基因组序列, 筛选到一批 SSR 位点, 同时对其组成、分布及特征进行了分析, 以期开发新的烟草 SSR 分子标记提供理论依据。

## 1 材料与方法

利用软件 SSRFINDER (<http://www.maizemap.org/bioinformatics/SSRFINDER/>) 搜索绒毛状烟草基因组 DNA 的序列数据, 条件为: 重复单元长度 2 ~ 5 bp, 重复次数为 10 ~ 60。

## 2 结果

### 2.1 SSR 位点的数量与分布

利用软件 SSRFINDER 对绒毛状烟草基因组 DNA 的序列数据进行搜索, 共找到符合条件的 SSR 位点 6 441 个, 涉及 5 227 个 Scaffold, SSR 位点的序列总长 303 963 bp (表 1)。从 SSR 位点的数量上看, 三碱基重复基元的数量最多, 占 49.6%; 其次是二碱基重复基元, 占 45.2%; 四碱基重复基元和五碱基重复基元数量很少, 分别只占 2.8% 和 2.4% (图 1)。二碱基重复基元、三碱基重复基元、四碱基重复基元和五碱基重复基元的 SSR 位点的碱基总长分别是 112 678 bp、167 889 bp、11 616 bp 和 11 780 bp, 位点的平均长度分别是 39 bp、53 bp、66 bp 和 75 bp。

从重复次数看, 10~20 重复的位点有 4 809 个, 占 74.6%; 21~30 重复的位点有 963 个, 占 15.0%; 31~40 重复的位点有 349 个, 占 5.4%; 41~50 重复的位点有 205 个, 占 3.2%; 51~60 重复的位点有 113 个, 占 1.8% (图 2)。

### 2.2 重复单元的类型、数量及特征

在 6 441 个 SSR 位点中, 二碱基重复基元单元共出现 11 种类型, 分别是 CT (417 个)、AG (400 个)、TA (378 个)、TC (373 个)、GA (354 个)

表 1 绒毛状烟草基因组 SSR 位点的数量与分布  
Table1 Distribution of SSR loci in *N.tomentosiformis* Goodspeed genome DNA

重复次数	二碱基	三碱基	四碱基	五碱基
10	0	757	31	27
11	0	355	23	21
12	278	291	18	12
13	346	232	13	20
14	406	201	11	11
15	307	151	14	9
16	146	143	9	10
17	191	109	11	11
18	153	96	6	5
19	122	69	6	4
20	131	46	0	7
21	117	45	2	2
22	97	44	7	3
23	68	33	2	4
24	64	41	3	3
25	75	36	1	1
26	42	30	3	2
27	30	41	1	1
28	33	32	3	1
29	23	22	2	0
30	18	30	0	1
31~40	125	217	7	1
41~50	85	118	1	1
51~60	57	54	3	0
合计	2 914	3 193	177	157
所占比例/%	45.2	49.6	2.8	2.4
碱基总长/bp	112 678	167 889	11 616	11 780
位点平均长度/bp	39	53	66	75

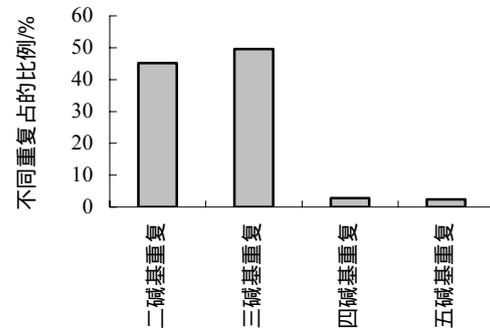


图 1 4 种重复单元的分布比例图

Fig.1 Distribution of four repeat motifs

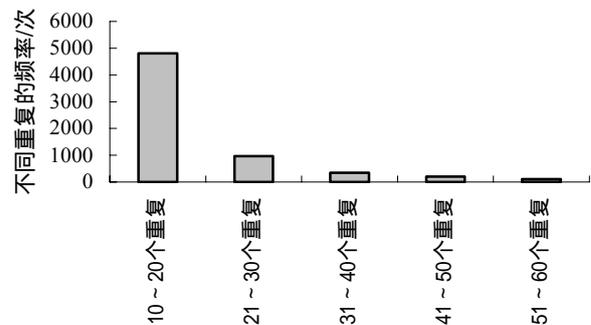


图 2 重复次数分布图

Fig.2 Distribution of the number of repeats

AT(312个)、TG(270个)、AC(229个)、CA(109个)、GT(71个)、CG(1个);三碱基重复基元单元共出现51种类型,数目最多的前10种依次是GAA(304个)、TTC(295个)、AAG(224个)、TCT(218个)、AGA(215个)、TTG(198个)、ACA(184个)、CTT(177个)、AAC(163个)、GTT(156个);四碱基重复基元单元共出现44种,数目最多的前10种依次是AAAG(16个)、ATAG(16个)、TATG(13个)、TTTA(12个)、ACAT(9个)、ATAC(7个)、TTTC(7个)、AAGA(6个)、AGAG(6个)、ATCT(6个);五碱基重复基元单元共出现45种类型,数目最多的前10种依次是CGGGA(23个)、TTTTTC(20个)、AAAAG(14个)、GTCCC(14个)、CGTCC(10个)、ACAAT(5个)、AGAAA(4个)、GACGG(4个)、TCCCG(4个)、TTTTTA(4个)(表2)。A、T、C、G在所有SSR位点中所占的比例分别为33.1%、33.4%、16.6%和16.9%(图3),GC含量为33.5%。

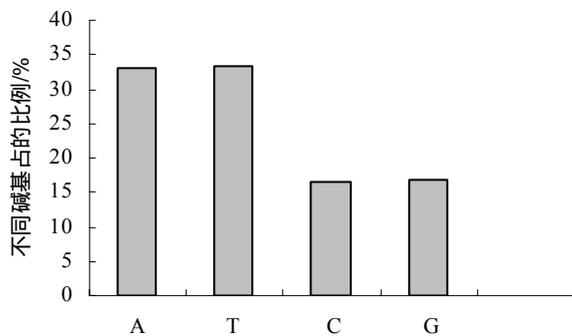


图3 SSR位点中四种碱基的比例

Fig.3 Distribution of A, T, C and G in SSR loci

### 3 讨论

根据测序结果,绒毛状烟草基因组大小为2.6Gb(待发表),本研究从中筛选SSR总长约为300kb,约占基因组的万分之1.2,换句话说绒毛状烟草基因组平均每400kb出现一个SSR。而本研究并未计算单碱基重复基元和六碱基重复基元的SSR,所以SSR实际出现的频率还会略高一些。不同物种间,SSR的数量、重复次数、重复单元以及在染色体上的分布均存在很大的差异<sup>[9]</sup>。在植物中,

SSR的出现频率与基因组大小、重复DNA所占比例呈负相关,而与转录区所占的比例、低拷贝序列的出现频率成显著正相关<sup>[23]</sup>。

绒毛状烟草基因组的GC含量为38.9%,其中SSR序列的GC含量为33.5%,明显低于全基因组

表2 重复单元的类型及数量

Table2 Type and number of repeat motifs

二碱基重复基元		三碱基重复基元		四碱基重复基元		五碱基重复基元	
重复单元类型	数量/个	重复单元类型	数量/个	重复单元类型	数量/个	重复单元类型	数量/个
CT	417	GAA	304	AAAG	16	CGGGA	23
AG	400	TTC	295	ATAG	16	TTTTTC	20
TA	378	AAG	224	TATG	13	AAAAAG	14
TC	373	TCT	218	TTTA	12	GTCCC	14
GA	354	AGA	215	ACAT	9	CGTCC	10
AT	312	TTG	198	ATAC	7	ACAAT	5
TG	270	ACA	184	TTTC	7	AGAAA	4
AC	229	CTT	177	AAGA	6	GACGG	4
CA	109	AAC	163	AGAG	6	TCCCG	4
GT	71	GTT	156	ATCT	6	TTTTTA	4
CG	1	AAT	111	GAGA	6	AATAC	3
		TTA	109	TATC	6	ATACA	3
		TGT	88	ATGT	5	TTTCT	3
		CAA	86	CATA	4	ACGGG	2
		ATA	84	CTCT	4	AGAAG	2
		TAT	75	TCTA	4	ATTTT	2
		ATT	57	TGTA	4	CAAGC	2
		TAA	53	AGAT	3	CTTGA	2
		GAT	48	ATAA	3	CTTTT	2
		CAT	33	ATAT	3	GCTTG	2
		TCA	28	CTTT	3	GGGAC	2
		GAG	27	TATT	3	TATTG	2
		TCC	25	TTAT	3	TCTTT	2
		AGG	24	AAAT	2	TGTAT	2
		CTC	21	AGAA	2	TTTTG	2
		TGA	20	GTGT	2	TTCTC	2
		CCT	19	TCTC	2	TTGGC	2
		ATC	18	TCTT	2	AAAAT	1
		ATG	16	TTCT	2	AAAGA	1
		GTA	11	TTTG	2	AATGG	1
		AGT	11	AATT	1	ACGAT	1
		TGG	10	AGGG	1	ATACT	1
		TAC	10	AGTT	1	ATTGT	1
		GGA	10	CATG	1	CCAAG	1
		ACT	10	CTAA	1	CCGTC	1
		TAG	9	CTAT	1	CTTCT	1
		CTG	6	GAGG	1	GAATG	1
		CCA	6	GGGT	1	GAGAA	1
		ACC	6	TAAA	1	GATAC	1
		CTA	4	TAAT	1	TAAAA	1
		TGC	3	TAGA	1	TCTTC	1
		GGT	3	TAGT	1	TGCCT	1
		CAG	3	TATA	1	TGTGG	1
		CAC	3	TCAC	1	TTCTT	1
		AGC	3			TTGGT	1
		ACG	3				
		TCG	2				
		GTG	1				
		GTC	1				
		GCA	1				
		CGC	1				

的 GC 含量。SSR 序列显示出显著的碱基偏好性, A/T 的频率显著高于 G/C。二、三碱基重复基元的 SSR 共 6107 个, 占 SSR 总体的 94.8%。而 (CG)<sub>n</sub>、(GTG)<sub>n</sub>、(GTC)<sub>n</sub>、(GCA)<sub>n</sub>、(CGC)<sub>n</sub> 分别只出现了 1 次, 说明这些重复单元的 SSR 在绒毛状烟草基因组中非常少见。从总体来看, 各种重复类型的 SSR 位点都存在着长度越长出现的频次越少的趋势。目前, 对于植物 SSR 序列信息的分析, 大多来源于 EST 数据<sup>[24-30]</sup>, 由于 EST-SSR 来自于转录区域, 一般保守性较高, 通用性好。Cho<sup>[31]</sup>等人对水稻的研究结果表明, 来源于基因组的 SSR 比来源于 EST 的 SSR 具有更丰富的多态性。研究发现, 当 SSR 长度大于等于 20 bp 时, 不同品种间显示出较高的多态性<sup>[32]</sup>, 而且高级基元的 SSR 多态性比低级基元的低<sup>[33]</sup>。在这 6 441 个 SSR 位点中, 二、三、四、五重复基元 SSR 位点的平均长度分别为 39 bp、53 bp、66 bp 和 75 bp, 其中二、三碱基重复基元的总和占总体的 94.8%, 表明这些 SSR 具有较高的利用价值。用 PRIMER5.0 对这 6 441 个 SSR 设计引物, 其中 6 020 可获得满足条件的引物。至于哪些引物可用于开发新的烟草 SSR 分子标记还有待进一步研究。

### 参考文献

[1] Beckmann J S, Soller M. Toward a unified approach to genetic mapping of eukaryotes based on sequence tagged microsatellite sites[J]. *Bio Technology*, 1990, 8: 930-932.

[2] Powell W, Machray G C, Provan J. Polymorphism revealed by simple repeats[J]. *Trends Plan Sci*, 1996, 1: 215-222.

[3] 洪彦彬, 梁炫强, 陈小平, 等. 花生栽培种 SSR 遗传图谱的构建[J]. *作物学报*, 2009, 35 (3): 395-402.

[4] 向道权, 曹海河, 曹永国, 等. 玉米 SSR 遗传图谱的构建及产量性状基因定位[J]. *遗传学报*, 2001, 28 (8): 778-784.

[5] Thiel T, Michalek W, Varshney R K, et al. Exploiting EST database for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare*L.)[J]. *Theor Appl Genet*, 2003, 106 (3): 11-22.

[6] Korzun V, Roder M, Wendehake K, et al. Integration of dinucleotide microsatellites from hexaploid bread wheat into a genetic linkage map of durum wheat[J]. *Theor Appl Genet*, 1999, 98 (1): 1202-1207.

[7] Prevost A, Wilkinson M. A new system of comparing

PCR primers applied to SSR fingerprinting of potato cultivars[J]. *Theor Appl Genet*, 1998, 98 (1): 107-112.

[8] Prasad M, Varshney R, Roy J, et al. The use of microsatellites for detecting DNA polymorphism, genotype identification and genetic diversity in wheat[J]. *Theor Appl Genet*, 2000, 100 (4): 584-592.

[9] 朱宏波, 方宣钧, 杨仁崔. 利用水稻基因组序列数据开发 SSR 标记的方法[J]. *分子植物育种*, 2003, 1 (2): 273-276.

[10] 范静苑, 王元英, 蒋彩虹, 等. 烟草 CMA 抗性鉴定及抗性基因的 SSR 标记研究[J]. *分子植物育种*, 2009, 7 (2): 355-359.

[11] H S Moon, J S Nicholson, R S Lewis. Use of transferable *Nicotiana tabacum* L. microsatellite markers for investigating genetic diversity in the genus *Nicotiana*[J]. *Genome*, 2008, 51: 547-559.

[12] Bai D, Reeleder, R, Brandle, J E. Identification of two RAPD markers tightly linked with the *Nicotiana debneyi* gene for resistance to black root rot of tobacco[J]. *Theor Appl Genet*, 1995, 91: 1184-1189.

[13] Yi Y H, Ruffy R C, Wernsman E A. Identification of RAPD markers linked the wildfire resistance gene of tobacco using bulked segrement analysis[J]. *Tob Sci*, 1998, 42: 52-57.

[14] Johnson E S, Wolff M F, Wernsman E A. Marker assisted selection for resistance to black shank disease in tobacco[J]. *Plant Dis*, 2002, 86: 1303-1309.

[15] Lewis R S, Milla S R, Kernodle S P. Analysis of an introgressed *Nicotiana tomentosa* genomic region affected leaf number and correlated traits in *Nicotiana tabacum*[J]. *Theor Appl Genet*, 2007, 114: 841-854.

[16] Milla, S R., Levin, J.S., Lewis, R.S., and Rufty, R.C. RAPD and SCAR markers linked to an introgressed gene conditioning resistance to *Peronospora tabacina* D.B. Adam. in tobacco[J]. *Crop Sci*, 2005, 45: 2346-2354.

[17] Lewis R S. Transfer of resistance to potato virus Y(PVY) from *Nicotiana africana* to *Nicotiana tabacum*: possible influence of tissue culture on the rate of introgression[J]. *Theor Appl Genet*, 2005, 110: 678-687.

[18] Lewis R S, Milla S R, Levin J S. Molecular and genetic characterization of *Nicotiana glutinosa* L. chromosome segments in tobacco mosaic virus-resistant tobacco accessions[J]. *Crop Sci*, 2005, 45: 2355-2362.

[19] Moon H S, Nicholson J S. Identification of AFLP markers linked to Tomato Spotted Wilt Virus resistance in tobacco[J]. *Crop Sci*, 2007, 47: 1887-1894.

[20] Ren N, Timko M P. AFLP analysis of genetic polymorphism and evolutionary relationships among cultivated and wild *Nicotiana* species[J]. *Genome*, 2001, 44: 559-571.

[21] Rossi L, Bindler G, Pijenburg H, et al. Potential of molecular marker analysis for variety identification in processed tobacco[J]. *Plant Var. Seed*, 2001, 14: 89-101.

表5 羊肚菌发酵多糖在卷烟中的试用效果

Table 5 The evaluation of cased cigarettes and flue-cured tobacco of the polysaccharide extract

样品	用量/ %	光泽 (5.0)	香气 (32)	协调性 (6.0)	杂气 (12)	刺激性 (20)	余味 (25)	总计 (100)
1#	0.0000	5.00	29.43	5.07	10.50	17.50	22.07	89.6
	0.0001	5.00	29.57	5.00	10.57	17.71	21.86	89.7
	0.0002	5.00	29.86	5.07	10.57	17.57	22.29	90.4
	0.0003	5.00	29.5	5.07	10.71	17.64	22.29	90.2
2#	0.0000	5.00	28.93	5.00	10.21	17.43	21.93	88.5
	0.0001	5.00	29.36	5.00	10.21	17.57	22.14	89.3
	0.0002	5.00	29.43	5.07	10.21	17.64	22.14	89.5
	0.0003	5.00	29.36	5.00	10.21	17.50	21.93	89.0

乳糖和甘露糖等残基为重复单元组成的杂多糖<sup>[7]</sup>，根据真菌多糖不溶于乙醇的特点，对真菌多糖进行粗提，其中真菌多糖含量达到 42.6%，香气分析表明，致香成分含量较少，仅检测出 13 种，大部分为长链碳酸及其酯类，但通过热裂解表明，羊肚菌发酵多糖可在抽吸时形成糠醛类物质外，还有多种呋喃类和醛酮类香气成分，对丰富烟香、改善吸味具有较好效果。

### 3.3 羊肚菌发酵多糖的应用效果

羊肚菌发酵多糖在卷烟中的试用结果表明，其在醇和烟香、提高舒适性、去除杂气等方面具有较

好的作用，且用量远远低于普通香精香料的用量，可适用于卷烟加料及烟叶醇化。

### 参考文献

- [1] 兰进,曹文苓,徐锦堂.中国羊肚菌属真菌资源[J].资源科学,1999,21(2):56-61.
- [2] Duncan C J, Pugh N, Pasco D S, et al. Isolation of a galactomannan that enhances macrophage activation from the edible fungus *Morchella esculenta*[J]. *J Agric Food Chem*, 2002, 50(20): 5683-5685.
- [3] 许周善,周晓燕.冬虫夏草多糖的研究进展[J].工业微生物,2000,30(1):56-57.
- [4] 李林辉.野外分离大型真菌菌种技术[J].中国食用菌,2002,21(5):11-12.
- [5] 武秋立,安家彦.羊肚菌的液体深层培养[J].食品与发酵工业,2004,30(2):72-76.
- [6] 谭晓虹,王治宝,李如章.北五味子中多糖含量的测定[J].河北北方学院学报:医学版,2005,22(6):45-46.
- [7] 魏芸,张天佑.羊肚菌多糖的分离纯化及组成结构分析[J].植物资源与环境学报,2000,9(2):14-17.

(上接第 35 页)

- [22] Bindler G, Rutger van der Hoeven, et al. A microsatellite marker based linkage map of tobacco[J]. *Theor Appl Genet*, 2007, 114: 341-349.
- [23] Morgante M, Hanafey M, Powell W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes [J]. *Nature Genetics*, 2002, 30: 194-200.
- [24] 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.
- [25] 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 Mol Biol*, 2002, 48: 501-510.
- [26] Thiel T, Michalek W, Varshney R K, et al. Exploiting EST database for the development and characterization of gendered SSR-markers in barley (*Hordeum vulgare* L.) [J]. *Theor Appl Genet*, 2003, 106: 411-422.
- [27] 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):49-554.
- [28] 林范学,程水明,李安政,等.香菇基因组中 EST-SSR 的构成和分布[J].微生物学通报,2007,34(3):438-442.
- [29] 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(1):17-23.
- [30] Jiang Dong, Zhong Guangyan, Hong Qibing. Analysis of Microsatellites in *Citrus Unigenes* [J]. *Theor Appl Genet*, 2003, 106: 411-422.
- [31] 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.
- [32] Temnykh S, DeClerck G, Lukashova A, et al. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.) [J]. *Genome Res*, 2001, 11: 1441-1452.
- [33] Dreisigacker S, Zhang P, Warburton M L, et al. SSR and pedigree analysis of genetic diversity among CIM-MYT wheat lines targeted to different mega environments [J]. *Crop Science*, 2004, 44: 381-388.