

## 小麦 EST-SSR 标记的开发和遗传作图

潘海涛, 汪俊君, 王盈盈, 齐照良, 李斯深

(山东农业大学农学院/作物生物学国家重点实验室, 山东泰安 271018)

**摘要:** 【目的】利用小麦 EST 序列数据库开发 EST-SSR 标记。【方法】对 GenBank/dbEST 注册的普通小麦 EST 序列 (2006. 4. 18—2007. 2. 4) 进行 SSR 查找, 采用 Primer5.0 软件设计 EST-SSR 引物, 选用 3 个小麦品种进行有效性检测, 利用 RIL 群体和 Mapmaker/Exp3.0 软件进行遗传作图。【结果】在 265 362 条普通小麦 EST 序列中, 共发现 6 314 个 SSR, 占整个 EST 数据库的 2.38%。其中二核苷酸、三核苷酸重复序列最多, 分别为 2 237 (35.43%) 和 2 084 (33.01%) 个。二核苷酸重复中, 以 GA/CT 和 AG/TC 出现频率最高, 分别占 SSR 总数的 17.85% 和 10.37%, 其次是 CA/GT (4.07%) 和 AC/TG (2.53%); 三核苷酸重复中, CAA/GTT (3.93%)、CGG/GCC (3.83%)、CGC/GCG (3.36%)、GGC/CCG (3.14%)、CTT/GAA (2.53%)、TGC/ACG (2.27%) 以较高的频率出现。根据筛选得到的微卫星序列共设计了 596 个 EST-SSR 引物对, 选择其中 95 分以上的 194 个合成。PCR 检测表明, 165 个引物对 (85%) 可以扩增出稳定清晰的带型; 在 RIL 群体中检测到 21 个 EST-SSR 引物 26 个位点有多态性, 将其中的 23 个位点整合到已有的小麦遗传图谱上。【结论】开发了 165 个小麦 EST-SSR 新标记, EST 序列是小麦 SSR 标记的重要来源。

**关键词:** 小麦, EST-SSR, 分子标记, RIL, 遗传图谱

## Development and Mapping of EST-SSR Markers in Wheat

PAN Hai-tao, WANG Jun-jun, WANG Ying-ying, QI Zhao-liang, LI Si-shen

(National Key Laboratory of Crop Biology/College of Agronomy, Shandong Agricultural University, Tai'an 271018, Shandong)

**Abstract:** 【Objective】 The objective of this study was to develop new simple sequence repeat (SSR) markers from EST (Expressed sequence tag) sequences (EST-SSR markers) for wheat. 【Method】 The wheat EST sequences from GenBank/dbEST (from April 18, 2006 to February 4, 2007) were used to search for SSRs, EST-SSR primer pairs were designed using Primer5.0 software, the validity of primer pairs were detected using three wheat varieties, and the marker loci were mapped using RIL population and Mapmaker/Exp3.0 software. 【Result】 From the 265 362 wheat EST sequences released in GenBank/dbEST, 6 314 SSRs were found which amounts to 2.38% of the total number of ESTs. Among those EST-SSRs, the most motifs were dinucleotides and trinucleotides, which were 2 237 (35.43%) and 2 084 (33.01%), respectively. For dinucleotides, GA/CT (17.85%) and AG/TC (10.37%) were the most frequent repeat, followed by CA/GT (4.07%) and AC/TG (2.53%); and for trinucleotides, CAA/GTT (3.93%), CGG/GCC (3.83%), CGC/GCG (3.36%), GGC/CCG (3.14%), CTT/GAA (2.53%) and TGC/ACG (2.27%) appeared high frequency. Based on these EST-SSR sequences, 596 EST-SSR primer pairs were designed, 194 of which were synthesized, and 165 (85%) led to PCR amplification products. Using a RIL population, the polymorphism of 21 EST-SSR primer pairs with 26 loci were detected, and 23 of those loci were integrated on the wheat genetic map. 【Conclusion】 A total of 165 new wheat EST-SSR markers have been developed. EST sequences are important sources for developing SSR markers.

**Key words:** wheat, EST-SSR, molecular marker, RIL, genetic map

## 0 引言

【研究意义】小麦是中国最重要的粮食作物之

一。目前, 分子标记技术已用于小麦遗传图谱构建<sup>[1-3]</sup>、重要农艺性状基因的标记定位<sup>[4-5]</sup>、种质资源的遗传多样性分析<sup>[6-9]</sup>、品种指纹图谱及纯度鉴定<sup>[10-11]</sup>、分子标

收稿日期: 2009-03-04; 接受日期: 2009-08-19

基金项目: 国家自然科学基金 (3057155)

作者简介: 潘海涛, 硕士研究生。通信作者李斯深, 教授, 博士。Tel: 05388242903; E-mail: sslis@sdau.edu.cn

记辅助选择育种<sup>[12-13]</sup>等方面的研究。因此, 开发新的小麦分子标记研究具有重要应用前景。【前人研究进展】众多分子标记中, 基于 PCR 的 SSR (simple sequence repeat) 标记以其丰富且遍布基因组、具有超变异性<sup>[14]</sup>及位点专一性、通常表现共显性和多等位基因<sup>[15]</sup>等优越性得到广泛应用。但传统的基因组 SSR 标记开发需经过文库构建, 包括 SSR 克隆筛选、测序、引物设计和 PCR 扩增与分析等步骤, 要投入大量的人力物力。【本研究切入点】近年来, 许多作物的大规模 cDNA 测序并在网上公开释放, 大大增加了基于 EST (expressed sequence tags) 的 SSR (EST-SSR) 标记开发能力。EST-SSR 来源于表达基因, 与 SSR 相比, 它反映了基因的编码部分, 可以直接获得基因表达的信息, 为功能基因提供“绝对”的标记; 省去了 SSR 引物开发过程中的克隆和测序步骤, 降低了开发成本; 在不同物种间通用性好。因此, 开发 EST-SSR 标记是丰富作物分子标记的简便而有效的途径。已在棉花<sup>[16-17]</sup>、小麦<sup>[18-27]</sup>、大麦<sup>[28-30]</sup>、黑麦<sup>[31-32]</sup>、水稻<sup>[33]</sup>等作物开展了 EST-SSR 标记的开发和利用。【拟解决的关键问题】本研究旨在开发小麦新的 EST-SSR 标记。利用 GenBank 注册的小麦 EST 序列查找 EST-SSR, 分析 SSR 在 EST 中的分布特征, 设计 EST-SSR 引物, 在小麦中进行有效性检测和遗传作图。

## 1 材料与方法

### 1.1 EST 序列来源

从 <http://www.ncbi.nlm.nih.gov/dbEST/> 下载小麦 EST 序列, 时间为 2006.4.18—2007.2.4 (gi|92487338—gi|124110159), 共获得 265 362 条小麦 EST 序列。

### 1.2 EST-SSR 的查找

登陆网站 <http://www.gramene.org/gremene/searches/ssrtool>, 利用 SSRIT (simple sequence repeat identification tool) 软件在线查找二、三、四、五核苷酸 4 种类型的 SSR, 识别标准为: 重复序列长度 $\geq 20$  bp, 即重复次数分别大于或等于 10、7、5、4。

### 1.3 EST-SSR 的引物设计

利用 Primer 5.0 软件根据 SSR 的侧翼区域设计引物。引物设计的主要参数为: EST 序列长度大于 100 bp; SSR 序列的开始和结束位置分别距 5'和 3'端不少于 20 bp; 引物长度 18—22 bp; 退火温度  $T_m$  值 50—65℃, 而且上游和下游引物的  $T_m$  值相差不大于 5℃; PCR 扩增产物长度 100—300 bp; 得分 90 分以上, 尽量避免引物二级结构的出现。引物设计完成以后, 由

上海生物技术有限公司合成引物。引物命名为 SWES 加序号, 如 SWES1。

### 1.4 DNA 提取, SSR 扩增和电泳

EST-SSR 有效性检测材料为中国春、以及笔者所在的作物生物学国家重点实验室构建的 RIL 群体 (recombinant inbred lines) 的亲本山农 483 和川 35050。遗传作图群体为作物生物学国家重点实验室构建 RIL 群体—RIL-8, 其组合为“川 35050 $\times$ 山农 483”,  $F_{15}$  (2007 年), 共 131 个系。

采用 SDS 法提取基因组 DNA。PCR 体系各组分比例及 PCR 反应程序参考 Chen 等的方法<sup>[26]</sup>。扩增产物用 4.5% 脲变性聚丙烯酰胺凝胶电泳, 银染法检测。

### 1.5 遗传图谱绘制

在川 35050 和山农 483 之间有多态性的 EST-SSR 引物被用来扩增 RIL-8 作图群体 131 株系的 DNA。利用软件 Mapmaker/Exp3.0 将其整合到 RIL-8 群体的分子标记遗传连锁图谱<sup>[4]</sup>: 用“near”命令确定其染色体归属; 用“compare”命令确定各染色体上标记的排序; 采用“map”命令作图, Kosambi 图距。

## 2 结果

### 2.1 源于小麦 EST 的 SSR 查找

从 265 362 条 EST 序列中共发现 6 314 个 SSR 序列, 占整个 EST 数据库的 2.38%; 这些 EST 序列全长 214.774 Mb, 平均每 1 Mb 出现 29.40 个 SSR, 即每 34.02 kb 出现 1 个 SSR (表 1)。含有二、三、四和五核苷酸重复基元分别占 SSR 总数 35.43%、33.00%、13.70%和 17.87%。二核苷酸和三核苷酸重复的出现频率接近, 明显高于四、五核苷酸类型。

表 1 基于重复基元大小的 EST-SSR 分布

Table 1 Distribution of EST-SSR based on motif size

基元类型 Motif	数量 Number	百分比 Percentage (%)	频率 Frequency (Mb)
二核苷酸 Dinucleotide	2237	35.43	10.42
三核苷酸 Trinucleotide	2084	33.00	9.70
四核苷酸 Tetranucleotide	865	13.70	4.03
五核苷酸 Pentanucleotide	1128	17.87	5.25
合计 Total	6314	100	29.40

### 2.2 小麦 EST-SSR 的分布特征

6 314 个 SSR 中, 二核苷酸基元 GA/CT 和 AG/TC 出现频率最高, 分别占 SSR 总数的 17.85%和 10.37%, 其次 CA/GT (4.07%) 和 AC/TG (2.53%) 也以较高频率出现 (表 2), AT/TA 和 TA/AT 出现频率较低,

表 2 基于重复基元类型的 EST-SSR 分布

Table 2 Distribution of EST-SSR according to motif sequence type

重复基元 Repeat motif	数量 Number	百分数 Percentage (%)	重复基元 Repeat motif	数量 Number	百分数 Percentage (%)
GA/CT	1127	17.85	CTT/GAA	160	2.53
AG/TC	655	10.37	TGC/ACG	143	2.27
CA/GT	257	4.07	GAG/CTC	117	1.85
AC/TG	160	2.53	GTT/CAA	102	1.62
CAA/GTT	248	3.93	ACA/TGT	97	1.54
CGG/GCC	242	3.83	GGAC/CCTG	214	3.39
CGC/GCG	212	3.36	CGATA/GCTAT	189	2.99
GGC/CCG	198	3.14	GACGA/CTGCT	118	1.87

GC/CG 和 CG/GC 没有出现。三核苷酸基元以 CAA/GTT 出现的频率高, 为 3.93%; 其次是 CGG/GCC、CGC/GCG、GGC/CCG、CTT/GAA、TGC/ACG, 出现频率分别为 3.83%、3.36%、3.14%、2.53%、2.27% (表 2), 其它类型出现频率较低。四核苷酸和五核苷酸以 GGAC/CCTG (3.39%)、CGATA/GCTAT (2.99%)、GACGA/CTGCT (1.87%) 出现频率较高 (表 2), 其余出现频率均较低。

对于重复次数, 二核苷酸基元的重复次数类型多、跨度大。其中 GA/CT 和 AG/TC 的重复次数类型最多, 跨度最大, 分别为 63 和 60 种, 从重复 10 次到 76 次和 69 次; 其次是 AC/TG 和 CA/GT, 重复次数类型为 19 和 17 种; AT/TA、TA/AT 也有较大的重复次数跨度 (表 3)。三核苷酸基元的重复次数类型和跨度小于二核苷酸。其中, CAA/GTT (12 种)、ACA/TGT (11 种)、GTT/CAA (10 种) 出现重复次数类型最多; 其次为 GAA/CTT、CTT/GAA、TGC/ACG、TCT/AGA (表 3); 其它重复基元多以重复 7、8、9 次 3 种重复次数类型出现。四核苷酸和五核苷酸重复基元多以重复 4、5、6、7 次中 1—2 种重复次数类型出现。

### 2.3 小麦 EST-SSR 引物设计及其有效性

表 3 基元类型的重复次数分布

Table 3 Distribution of repeat motif according to repeat number

重复基元 Repeat motif	跨度 Range	总数 Total	重复基元 Repeat motif	跨度 Range	总数 Total
GA/CT	10—76	63	ACA/TGT	7—24	10
AG/TC	10—69	60	GAA/CTT	7—31	8
AC/TG	10—37	19	CTT/GAA	7—16	7
CA/GT	10—39	17	TGC/ACG	7—16	6
AT/TA	10—23	8	TCT/AGA	7—16	6
TA/AT	10—28	7	GGAC/CCTG	5—6	3
CAA/GTT	7—20	12	CGATA/GCTAT	4—5	2
GTT/CAA	7—18	11	GACGA/CTGCT	4	1

利用 Primer 5.0 软件, 对 6 314 个 EST-SSR 序列进行 EST-SSR 的引物设计, 分值在 90 分以上的有 1 121 (17.75%) 个引物对; 去除相同的引物, 共设计了 596 个 EST-SSR 引物对。选取分值在 95 分以上 266 对引物, 经与笔者所在实验室已合成引物<sup>[26-27]</sup>比较, 最终合成 194 个 EST-SSR 引物对 (引物编号为 SWES966—SWES1159)。经过 PCR 扩增、聚丙烯酰胺凝胶电泳和硝酸银染色, 165 个引物对 (85%) 可以扩增出 PCR 产物 (表 4, 引物的详细信息见笔者课题组的网站: <http://www.sdwgi.com/sdwgie.asp>)。这些引物可以作为小麦的新的 EST-SSR 标记。

### 2.4 遗传作图

利用 RIL-8 的 2 个亲本川 35050 和山农 483 对 EST-SSR 引物进行了多态性分析。有 21 个引物对在川 35050 和山农 483 间检测到多态性, 产生了 26 个多态性位点。一些引物检测到 2 个及以上的位点差异。利用这些 EST-SSR 引物扩增 RIL-8 群体的 131 个系, 进行作图分析, 将其中 20 个引物对的 23 个位点整合到小麦已有的遗传连锁图谱的 8 条染色体上, 1B、1D、4A、5A、5B、5D、6A、6B 染色体的位点数目分别为 4、1、5、4、1、2、1、5 个 (图)。

表 4 165 个小麦 EST-SSR 引物对情况

Table 4 Description of 165 wheat EST-SSR primer pairs

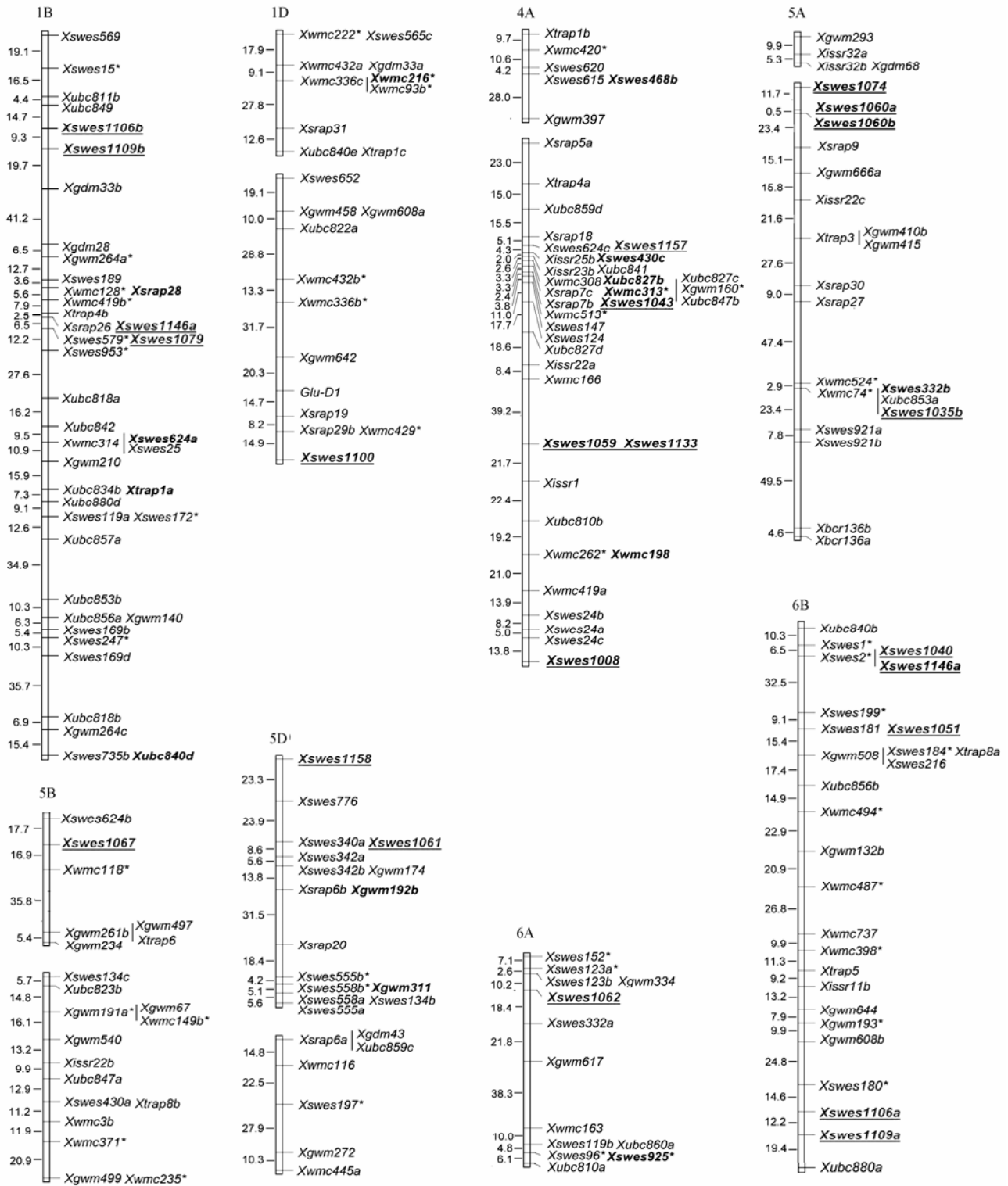
引物 Primer	基元 Motif	Tm (°C)	左端引物 (5'-3') Forward primer	右端引物 (5'-3') Reverse primer	引物 Primer	基元 Motif	Tm (°C)	左端引物 (5'-3') Forward primer	右端引物 (5'-3') Reverse primer
SWES966	gcg	61	AGCAGAGGAAGACAGGCAGGAC	AAGCAGCAGGAGGCGAAGC	SWES1002	gcg	57	CTTCCACAGCCACAACCG	CCCGCCGTCAACAATCTA
SWES967	ccg	61	GAAGCAGCAGGAGGCGAAGC	GCAGATGACGCACGAGTTGG	SWES1004	cgc	57	AATCACAAACGCCGAAGAAG	CGAACCACGACGAGAACG
SWES968	cgg	60	ACTTCCCTAACCTTACCCAACC	GACATCGCAGGACCATAACCG	SWES1005	cccct	57	AAAGAAGAGGTCTGGGAAGGA	ACGGGAATCAAATCAAGCAA
SWES969	ac	59	CTTGACCCGCCCGCTTCT	TCGCCCTTCGCCGCCATCT	SWES1006	gga	57	AGAGCGAAGAGGCGAAAC	GGTGGGAAGTCCGAACAG
SWES970	cgc	59	CCGTTGAAGGACGAAGCC	AGGGAGGAGGTGGTGTGG	SWES1007	gctc	57	CAGAGCGGCAGCAGTTAT	ACCAGCCTTGACCATTCC
SWES971	gag	59	CCACCTTGACTCTCGCTTTC	TCTCCTGCCAGACCTTC	SWES1008	ca	57	ACCCTGCCAGTCTCATCAC	CGCCGACATCAACACCTT
SWES972	gca	59	CGGCGTTCTACCATTTCC	TCCTCAGCGAGCCACAAG	SWES1009	aggc	57	GCGGGACAGTAAACAAAT	CCTGGCGAACCTGACCTT
SWES973	cgc	59	AGCCTTCCAACCTCATCTCCT	TAGCCACCGTCCGCATCT	SWES1010	gac	56	CAGCGGCGAGGAAGATGA	CACGGAATAGGCGGCAGT
SWES975	cgg	58	CCGTCGCTCGTTAGATTAGA	TGATGTGCTCCACCCTGA	SWES1011	ggc	56	GAACCACGACGAGAACGC	CACAAACGCCGAAGAAGA
SWES976	gca	58	TCCGCTCGTCTCCGATTT	ACGGCGTGATAGGGCATC	SWES1012	gagag	56	GTCATCAAATCAAATGCTCGTA	TCCACCACCTTCTGTCTG
SWES977	cgg	58	GCCTCCGAAACCCTAAC	AGCCCTCTGGTACAACCC	SWES1013	cag	56	GCAAGCACAACAGCCACA	GAGCCGAACCTGAAATCC
SWES978	gca	58	CCTTCGCAGTAGCCACCT	GCCGACATTCACCACCAG	SWES1014	ggc	56	CGCCAATGTATCCGTGCTT	CTTCTGCTTGCCGCTCT
SWES979	ccgc	58	GCAGACCTTCCCTTTCCC	CCATCTCCAATGCCCTTC	SWES1016	ggc	56	CAAACCTCGCCAATGTATCCG	CTTCTGCTTGCCGCTCT
SWES980	gac	58	AGCCCACGCTGCCACTCTT	CCTCTGTGTGCCATCTCCC	SWES1018	cgc	56	TTGATTAGACGACCAGACCG	CAACAACAGCCACCAGACA
SWES981	gcg	58	TGTTTCCCGTCCATCAGTT	CTCCTCACAAGCAGCGTCTA	SWES1065	gag	54	GGAGGACAAGGGAGGTGA	CTTTGGAGGTGGGCAGTT
SWES982	gctgt	58	TTTACATTGCCATCCTCG	GCGGTCCGTCATCGTATT	SWES1066	gct	54	AGCCCTTACTTGCTACCT	AGCCCAACACTGCTTACC
SWES983	gt	58	CGGACCAGCAGAGCAAGA	CAAGCAAGGGCGTCGTAG	SWES1067	cag	54	ACCGATGGCAATCAACAT	TGGCAAAGCCCTAAGAAG
SWES984	ccgat	58	ATTCACCCTCCACCTC	ACCTGCCACTCAAACACG	SWES1068	cgc	54	CATCTTCGCCTCCTTCAA	TCCTCAATGCGTCCAATCT
SWES985	tcc	57	GCGGAAGTTGCCGATGAA	TGGTGCCTGCCTTTGTGG	SWES1069	gag	54	GCCAGGTGGAGGGATAGT	TGCTTCGGTCTTCTGTCTTT
SWES991	gag	57	CCTCCCGACCTACTACTATCC	TGTACCCTGTTCGGCTCAT	SWES1070	ag	54	ACGAAGCCAGCCCTATCA	TGCCGACACCAAGGTCAT
SWES992	ct	57	CTTGCTCCGTGGTCTTCTT	TGTAGGACAGCCTCATCTGC	SWES1071	tg	54	TCTTCGTCGCTTTCATCG	AGTTTAGCAAGGCGTCCC
SWES993	cag	57	CCACCAGCAATGCCAGGAT	TGCCGATGACGATGAGCC	SWES1072	gac	54	CGATGACTACGGTGATAAGAA	TGGGTGAATGAGGCAGGT
SWES994	cgg	57	GCAAGTACCCACTCCCACG	CGCTTCCGCTGTCCATTT	SWES1073	ggaa	54	TTGACCGACCGAACCATT	TTCTCCTCGTGAACACG
SWES995	aag	57	GCGACAGAAGCCAGTGAATC	CGAAGACGAGGAAGTAGAAGGT	SWES1074	ccaac	54	CGTGTTTCATCGTGGAGGC	AGACGGACGGTGGGGTT
SWES997	cgg	57	GGCTTCCCGTTCGTGTTT	GCGTCTCCTCTGGGTCAAT	SWES1075	ccct	53	CAGCACCCAATGTATCCC	CACCCAAACAAACCACCA
SWES999	gct	57	ACGCTGCGGATGAAGATG	GCTGGCAATGCTGAAAGG	SWES1076	gtcc	53	ACAAGTGACCGCAACCG	ACGAGCCAGCCAACAGCA
SWES1001	agc	57	GCTGGCAATGCTGAAAGG	GACGCTGCGGATGAAGAT	SWES1077	tgc	53	CGGGAGGAAGGAACCAAT	AGCCCATCTTATGAAACTGCT

续表4 Continued table 4

引物 Primer	基元 Motif	Tm (°C)	左端引物 (5'-3') Forward primer	右端引物 (5'-3') Reverse primer	引物 Primer	基元 Motif	Tm (°C)	左端引物 (5'-3') Forward primer	右端引物 (5'-3') Reverse primer
SWES1078	tgc	53	TCCGTCGTTTCTTCCCTC	TACCCATTGGTTGGTTCAGC	SWES1018	cgc	56	TTGATTAGACGACCAGACCG	CAACAACAGCCACCGACA
SWES1079	tgc	53	AAAAGGTCAAGTTAGGGAGT	GGATCAGAAAGAGGAGGC	SWES1019	gcc	56	ATCCTCCCGAGTCAACCA	CACTGCCTCAGCCAACAC
SWES1080	cgccg	53	AAAACCACCAATCAAAGGG	GATGTCGCCACCATAGCC	SWES1020	tgc	56	TGAAGCCAGAGCCGAACC	AACGCAGATGGACGAGGG
SWES1081	ttcag	53	AGTCGGAGGAGTGAGTAAACC	AATCAGACAAGCAGCGAAACA	SWES1021	gagg	56	CATCCCAGGTCCCTTTC	TGGTGGCACTGAGTCATACA
SWES1082	cag	53	TATGTTGAATCTTGACCCTC	CGCTGTTTGGGACTCTGT	SWES1022	tc	56	GAAGTCAATCACCACCAGC	CCACAAGGCAACAACAAAAG
SWES1084	tecct	53	CGATACGGCGGTCTCATT	TGTTGTCGTGCGTGTCC	SWES1023	tc	56	AAGCAAGGGCGTCTAGT	GGACCAGCAGAGCAAGAG
SWES1085	tgc	53	ACTGGAGCACCGTAAACA	TACAGCATCAAGGGCAAG	SWES1024	att	56	CGGAACCGCTACAAGGGAA	TCGGACCAAGCACCAACA
SWES1086	tecaa	53	AACCTCGCTCGCTTCTCC	TATCTCGCTGGCTGGCTC	SWES1026	cgcc	56	CGACCCAATCTCACATCAC	CGGGCAGAGGAGTTCATA
SWES1087	ac	53	GACAGGACGCACATCACTC	ATCAAAGCCCATAACCC	SWES1027	cag	56	ACCTACCCAACACCCATCA	TGTGCTGTCTGCTTTCG
SWES1088	ac	53	GGTCGGAGAATGAAGAAGC	GACGGGCGGATAATAAAGG	SWES1028	egg	56	CCCCTCCTCCACTACTCCT	GCCTACCAAGCGAATCAG
SWES1089	ctg	53	TCCAGGTTCTTTGTGACTT	AGCACTAACGGGTGAGGT	SWES1029	agaag	56	CACCATCCGCACGACATT	CTTGAGCACCGCCACCTT
SWES1090	ctg	53	ACCCACCATTTCTGTCT	TGATATTGCTGACCCTGA	SWES1030	tc	56	GAAGTCAATCACCACCAGC	CACAAGGCAACAACAAAAGT
SWES1091	gct	53	CAAGGCTTCAGTGGAGGT	ATGGATTTACAGGGACGA	SWES1031	tgc	56	AGTGCTATGAGGCTGCTACG	CGGACTTGATGGACGATT
SWES1092	tca	53	AAGGTGCTGCGTGTTAGG	CGGGACGGGATACAAGAT	SWES1032	cctcc	56	CTCGTGTGTTACCCTC	CCACCAGCAGCACTCTTA
SWES1093	ac	52	GCTGCCTTTCATACCAACC	ATCGCTGTACTCTGACCTG	SWES1033	cttet	55	CTTGAGCACCGCCACCTT	CCATCCGCACGACATTTT
SWES1094	tc	52	AAACGCAGTCCACAAAGG	CGACGAGGAAGATCAAGC	SWES1034	agc	55	AGCCAAAGAAGAACAAGAGC	GAAATAAAGACAGGGAGAACG
SWES1095	cac	52	ACAGCGGGAGTGATAGTGA	CAGGAGATGGAGGCAAAAT	SWES1035	ggac	55	AACAGGCAACGACGACAG	GCCAAGACCTTCCACGAC
SWES1096	tct	52	CCTTCCTTGTCGCCCTCT	GCAGCCTCGTAAGCCTTTT	SWES1036	gtcc	55	GCCAAGACCTTCCACGAC	AACAGGCAACGACGACAG
SWES1097	cgca	52	TAGTTATTACTCGCACCACCA	CCCGTCCATCTCTTGT	SWES1037	ctg	55	CGTTCTTCTGCGGGTAGT	GACGAGGGTTTCTTTGGA
SWES1098	tggc	52	CCCCTTCCGAGTTTAGTT	CGAGTCATTGTGCCTATTT	SWES1038	ctca	55	AGGAGGCAACACCAGGAC	TGGACAGCACCAACAACC
SWES1099	tgagg	52	TCGGAATAAATCGGAGAA	CTTGGAGGAAGGAGGAAT	SWES1040	atctc	55	CGATGCTCGCTTTGTGGT	TGATCGGGCAGATGGACT
SWES1100	ca	52	CTCCGAACGCAACCTAATA	ACCCGCAAGACTACCTCAT	SWES1041	gct	55	CGCTGTTGCCGTATTGCC	TGTGTCGCTGTCTTTC
SWES1101	gtt	52	CTAACAACCGAAGCCTCAT	CCAGTGGTGGCAACAATA	SWES1042	tgc	55	GAAATAAAGACAGGGAGAACG	GAAGAAGAGGCGGGTGAT
SWES1102	cgaca	52	CACATTCACCACCGCAACT	TCAAGATCGTCTTCTACAACCC	SWES1043	ct	55	AATCCCAGCCAGGCAAAC	TCATCCACGCCGTCATCT
SWES1103	caa	52	CAGCAGCCTCAGTAACAA	GTACCGAAGATGCCAAAT	SWES1046	agagg	55	CTACTCGCTTGTGCTTTGG	TCCGCATCTCTTCTTGG
SWES1104	ca	52	TGCCACGACGATCTTAAA	CGCAAGATGAAGCGAGAA	SWES1047	tgc	55	ACAGGAAATAAAGACAGGGAG	CGTCAACAGCCAGAAGGT
SWES1105	ct	52	CCGAGTCTCCGTCCAAA	TGCCCATCTTATCAACAAGTA	SWES1048	attc	55	TGAAAGACAAATGGACGGGAGC	CATAAGCCACGCCGCAAA
SWES1106	ctt	52	TCCCAGATGTCCGCCTCA	AAGCAGCCAGCCACCAAT	SWES1049	ctttt	55	GCACCTGTGCGTCAAACG	CTCCTGAAGCCAGCCATT
SWES1107	gca	52	GGACGAGGGTTCTTTGG	GCTGTGGCTGTGTGCTT	SWES1050	ccg	55	TCTCCACTTCTCCACTCC	CCCTGTAATCGGTCTTCC

续表 4 Continued table 4

引物 Primer	基元 Motif	Tm (°C)	左端引物 (5'-3') Forward primer	右端引物 (5'-3') Reverse primer	引物 Primer	基元 Motif	Tm (°C)	左端引物 (5'-3') Forward primer	右端引物 (5'-3') Reverse primer
SWES1051	cct	55	TCAACACCAACGCACGAA	GGGCATCAGCCAGAACAA	SWES1124	gtgc	51	GGAATCATCATCGGTTGT	GTCATAGGCTACGCACT
SWES1053	cag	54	TAACGCAGATGGACGAGG	CATGTTGAAGCCAGAGCC	SWES1125	taga	51	GCCAGGTTACAAAGGCAGAG	ACCACGACAGCGACGAAG
SWES1054	gact	54	TGATGAACCTAAAGACACCCTA	CAGTGCCAAGAACGAACA	SWES1126	catct	51	CACCCGAGAACCTAAACT	CTGAGGAAGATAGCACCC
SWES1055	gca	54	TGGAGGTGAAGAAGAGGC	CAGGAAATAAAGACAGGGAG	SWES1127	ct	51	GTTCACTACTTGCTTCCCATTA	CAGCCTTCTTCATTCTCTCA
SWES1056	gag	54	CCACCATCGTCTCCTTGA	GGCAGTTTGTCTCCTCGTT	SWES1128	tct	51	GCCTACCATCCCATCTTC	CCTCCCTTACCCTATCAA
SWES1058	gct	54	AACGGTAGAAATCACGGAGTA	ACAAGGACAAGGCTATGGA	SWES1129	agc	51	AAGGGCTCACTCTGTCTTATC	CTTGTGCCGTGCGTATTT
SWES1059	acgg	54	TGACGGTGCTGACGATGA	GGAGGCGATGCGATTCTA	SWES1131	ac	51	AGCCTACAGAAGGAACACG	AACACCGAGAAAGCGAGA
SWES1060	gggtt	54	ACAGACGGACGGTGAGGGT	GGCAGCAAAGCGAGTGGA	SWES1132	ac	51	AGCCATTGGTTCCTTCCT	TCGGGTGTCTGGGTGTAA
SWES1061	cag	54	AGAAGCCAAGAATACCAATAGC	TCCAGATGCCAGACCACC	SWES1133	agc	51	CAACGGGACATTGAAACA	TTGGCAAAGTAGCGTAGTG
SWES1062	cct	54	CCTTGCCATCACATAACAC	TGCCCTTGAGACATAACC	SWES1136	tag	51	AGTTTGAATCACCAGGGAA	AGCGAGACACGGAGGAAT
SWES1063	ttc	54	GACGAGGAAGAAGGAGCG	CAGCAGGAGCAAACGAAA	SWES1137	actg	51	TACATCCTTCGGGCATAG	GCAACCTTCTTCACCTC
SWES1064	gaatg	54	TGCTGCTATGCTCTGCTC	ACTCACCTCCACCAACCC	SWES1139	ccat	50	GGTCTTCAGACGACGATAA	GATGGGTGGATTGTTTT
SWES1108	aac	52	CAGTGGTGGAACAATAGA	AACAACCGAAGCCTCATC	SWES1140	agatg	50	TACAAGGCTGAGGAAGAT	ACCGCAGAACCTAAACTA
SWES1109	gaa	52	AGCAGCCAGCCACCAATG	TCCCAGATGTCCGCCTCA	SWES1141	caa	50	AGGGATGTCATCGTCTTG	GTTACTGAGGCTGCTGGT
SWES1110	getcc	52	TGGCTGAAACCACGATAA	CCAAACAACAACGACCCT	SWES1142	cta	50	AGCGAGACACGGAGGAAT	TGAAATCACCAGGGAAGG
SWES1111	tactg	52	AGTGATTGGGAACCTGGAGA	CACGCTGACACCGACATAAC	SWES1143	gat	50	TGTCCAGGGATACATTTCG	CCAGATTGGCACATTCA
SWES1112	ag	52	TACGGTAGAAGAAATGTAAGAG	CCTAAAGAGGAAGGAGCA	SWES1144	tcct	50	TTACAGCATCACGCAAAT	AGAAGGTAGCGGAGGAGT
SWES1113	tca	52	AAGGTGCTGCGTGTTAGG	CGGGATGGGATACAAGAT	SWES1145	agtc	50	GCAACCTTCTTCACCTC	ATACATCCTTCGGGCATA
SWES1114	cgca	52	AAATCACGCTCATAGGCT	GGAATCATCATCGGTTGT	SWES1146	cag	50	TTTCTCTTACTCTTCTTGTA	AGATTGCGGTGTTGTTGG
SWES1115	ac	52	CTGGTCGGAGAAGAAGCA	ACGGGCGGATAATAAGGT	SWES1147	atag	50	CATCTGTTGGTGCCTTTG	TCGGACGACTGAGCCTAT
SWES1116	tgc	52	TGGCTGTTGTGCTTGCTG	ATGGACGAGGTTTCTTTG	SWES1148	tga	50	AGGACCACCGTCATCTA	TCAAAGGAGCCCTCTGTA
SWES1117	tgc	52	GATGAGCCCAGTGAGAAC	GAATGGCAGTGACAAACA	SWES1151	tatcg	50	CGCCGAGAATAAGACGAG	ACGGTACAAGCCACAAGG
SWES1118	ctgaa	52	TCAGACAAGCAGCGAACAA	GGAGGAGTGAGTGAAACCC	SWES1152	tg	49	GGGCTACGACTACCTCAG	TGGTTATTACAATGGGATG
SWES1119	aag	52	GAAACCCACCTCCCTTAC	AGAGCCTACCATCCCATC	SWES1154	ga	49	AAGTACAAAGGAAGGCTAA	GGATTGCGGTAAGAAGTGAT
SWES1121	ctt	51	GTCATCCGCCAAACCAGA	TCATTGCCCTCCTCCATT	SWES1156	aca	49	GTTCTACCGCAACAATCAC	GCTGTGAGGGAATACTTT
SWES1122	gct	51	CTGCCACCATTGCTACTG	GTTCAAGGAAGAGCCACTA	SWES1158	ac	49	GCAGCAACTCAACCTTCT	ACCTCAGCGGCACTATTT
SWES1123	catct	51	CAAAGCGAGCGAGTGAAT	CAGCCAGTAGAAGTGAGACCA	SWES1159	ag	49	AGCCTCCTTCATTCTC	CACAACCTGCTTCCATT



本文新作图的位点加下划线 (SWES966—1159, LOD ≥ 3.0)。图距在染色体的左边, 标记名称在染色体的右边; 标记的最可能的区间在远右边, 其中, LOD ≥ 3.0 的标记用粗体表示, LOD 为 3.0-1.0 用正常字体表示。星号 (\*) 表示该位点用中国春缺体——四体定位在某染色体。  
 The new mapped loci in this study (SWES966-1159, LOD ≥ 3.0) were underlined. Map distances are indicated on the left side of each chromosome, and the names of markers on the right side. The most likely intervals are on the far right side, of which the loci with LOD ≥ 3.0 were showed in bold letters and with LOD from 3.0 to 1.0 in normal letters. Asterisk (\*) indicates that the loci were located on chromosomes using Chinese Spring nulli-tetrasomic lines

图 EST-SSR 标记在遗传图谱上的位置 (SWES966-1159, 下划线的位点)  
 Fig. Location of EST-SSR makers in genetic map (SWES966-1159, loci with underline)

### 3 讨论

随着 EST 和 cDNA 大规模测序的开展, EST-SSR 引物的开发展现了广阔的前景。截至 2008 年 11 月 14 日, 在 dbEST 数据库 (<http://www.ncbi.nlm.nih.gov/dbEST>) 中普通小麦的 EST 序列已达到 1 250 932 条, 与 2007 年 2 月 4 日的 1 015 662 条序列相比, 增加了 235 270 条。在利用 EST 进行小麦的 EST-SSR 引物开发方面, Eujayl 等<sup>[18]</sup>根据 DuPont 公司的 EST 数据设计了 137 个 EST-SSR 引物对, 有 22 个引物对在 64 个硬粒小麦中扩增出多态性; Nicot 等<sup>[22]</sup>利用 170 746 条 EST 检测到 3 530 个 SSR, 设计了 688 个 EST-SSR 引物对, 483 个引物对有扩增产物, 其中 68% 的扩增质量较好; Gao 等<sup>[23]</sup>利用 71 495 条 EST 检测到 1 228 个 SSR, 设计了 597 个 EST-SSR 引物对, 有 478 个引物对可以成功地扩增出产物; 陈军方等<sup>[25]</sup>利用国际小麦族 EST 协作网 (ITEC) 上公布的 10 380 条 EST 序列共设计 135 个 EST-SSR 引物对, 其中 82 对在小麦上有扩增产物, 占所设计引物总数的 60.8%。陈海梅等<sup>[26]</sup>利用 GenBank/dbEST 普通小麦 151 695 条 EST 序列, 设计了 249 个 EST-SSR 引物对, 其中 166 个有扩增产物。上述 EST-SSR 标记可以作为小麦和相关物种的分子标记。本文根据 265 362 条普通小麦 EST 序列, 共设计了 596 个 EST-SSR 引物对, 选择其中 194 个合成, 165 个引物对 (85%) 可以扩增出 PCR 产物, 为小麦提供了新的 EST-SSR 标记。

关于 EST 中不同 SSR 出现频率, 多数结果表明三核苷酸重复出现频率最高, 本文的结果中二核苷酸和三核苷酸重复出现频率最高, 其中二核苷酸稍高于三核苷酸。这有可能与笔者查找 SSR 标准为重复序列长度  $\geq 20$  bp 有关, 因为在前人的研究结果中, 18 bp 长度的重复序列中三核苷酸重复出现频率高于二核苷酸重复<sup>[28]</sup>。在水稻、玉米、大豆、高粱中二核苷酸重复基元出现频率最多的都是 GA/CT<sup>[27]</sup>; 在大麦中二核苷酸重复基元 AG/TC 和 AC/TG 出现频率较高<sup>[28]</sup>。在水稻、玉米、大麦中 CCG/GGC 和 AGG/TCC 出现频率高<sup>[27-28, 33]</sup>。在本研究中, 小麦中二核苷酸重复基元出现频率最多的是 GA/CT 和 AG/TC, 三核苷酸重复基元出现频率最多的是 CAA/GTT、CGG/GCC、CGC/GCG 和 GGC/CCG, 此结果与陈海梅等<sup>[26]</sup>一致。二核苷酸重复基元出现的重复次数类型多、跨度大, 如 GA/CT 和 AG/TC 分别有 63 种和 60 种重复次数类型, 跨度从重复 10 次到 76 次、69 次。而三核苷酸重

复基元出现重复次数类型最多为 12 种, 跨度从重复 7 次到 20 次。Cho 等<sup>[33]</sup>在水稻中发现 SSR 的重复次数与 SSR 的变异呈正相关, 本研究结果为二核苷酸 SSR 具有更高多态性<sup>[22]</sup>提供支持。

在 EST-SSR 遗传图谱绘制方面, Gao 等<sup>[23]</sup>利用 RIL 群体将 101 个 EST-SSR 位点整合到遗传图谱上。在笔者以前的研究中, Li 等<sup>[4]</sup>将 94 个 EST-SSR 位点绘制到 RIL 群体的遗传图谱上, 本研究将 19 个新的 EST-SSR 位点整合到 RIL 群体的遗传连锁图谱上, 增加了图谱的长度, 丰富了功能标记的数目。

### 4 结论

本研究在 265 362 条普通小麦 EST 序列中发现 6 314 个微卫星, 占整个 EST 数据库的 2.38%。设计了 596 个 EST-SSR 引物对, 选择其中 95 分以上的 194 个合成, 165 个引物对 (85%) 可以扩增出稳定清晰的带型, 即开发了 165 个小麦 EST-SSR 新标记。21 个引物的 26 个位点在 RIL 群体有多态性, 将其中的 23 个位点整合到已有小麦遗传图谱上。这表明 EST 序列是小麦 SSR 标记的重要来源。

### References

- [1] Röder M S, Korzun V, Wandehake K, Planschke J, Tixier M H, Leroy P, Ganal M W. A microsatellite map of wheat. *Genetics*, 1998, 149: 2007-2023.
- [2] Somers D J, Isaac P, Edwards K. A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics*, 2004, 109: 1105-1114.
- [3] Xue S L, Zhang Z Z, Lin F, Kong Z X, Cao Y, Li C J, Yi H Y, Mei M F, Zhu H L, Wu J Z, Xu H B, Zhao D M, Tian D G, Zhang C Q, Ma Z Q. A high-density intervarietal map of the wheat genome enriched with markers derived from expressed sequence tags. *Theoretical and Applied Genetics*, 2008, 117: 181-189.
- [4] Li S S, Jia J Z, Wei X Y, Zhang X C, Chen H M, Sun H Y, Fan Y D, Li L Z, Zhao X H, Lei T D, Xu Y F, Jiang F S, Wang H G, Li L H. A intervarietal genetic map and QTL analysis for yield traits in wheat. *Molecular Breeding*, 2007, 20: 167-178.
- [5] Hao Y F, Liu A F, Wang Y H, Feng D S, Gao J R, Li X F, Liu S B, Wang H G. Pm23: a new allele of Pm4 located on chromosome 2AL in wheat. *Theoretical and Applied Genetics*, 2008, 117: 1205-1212.
- [6] Sun Q, Ni Z, Liu Z, Gao J, Huang T. Genetic relationships and diversity among Tibetan wheat, common wheat and European spelt



- wheat revealed by RAPD markers. *Euphytica*, 1998, 99: 205-211.
- [7] Reif J C, Zhang P, Dreisigacker S, Warburton M L, van Ginkel M, Hoisington D, Bohn M, Melchinger A E. Wheat genetic diversity trends during domestication and breeding. *Theoretical and Applied Genetics*, 2005, 110: 859-864.
- [8] Wang H Y, Wang X E, Chen P D, Liu D J. Assessment of genetic diversity of Yunnan, Tibetan, and Xinjiang wheat using SSR markers. *Journal of Genetics and Genomics*, 2007, 34: 623-633.
- [9] Gupta P K, Rustgi S, Sharma S, Singh R, Kumar N, Balyan H S. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Molecular Genetics and Genomics*, 2003, 270: 315-323.
- [10] Cooke R J. Gel electrophoresis for the identification of plant variety. *Journal of Chromatography A*, 1995, 698: 281-299.
- [11] 贾继增. 分子标记种质资源鉴定和分子标记育种. 中国农业科学, 1996, 29(4): 1-10.  
Jia J Z. Molecular germplasm diagnostics and molecular marker assisted breeding. *Scientia Agricultura Sinica*, 1996, 29(4): 1-10 (in Chinese).
- [12] John D, William A B, Gene D L. A comparison of marker-assisted and phenotypic selection for high grain protein content in spring wheat. *Euphytica*, 2006, 152: 117-134.
- [13] Huang X Q, Cöster H, Ganai M W, Röder M S. Advanced backcross QTL analysis for the identification of quantitative trait loci alleles from wild relatives of wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics*, 2003, 106: 1379-1389.
- [14] Powell W, Machray G C, Provan J. Polymorphism revealed by simple sequence repeats. *Trends in Plant Science*, 1996, 1: 215-222.
- [15] Gupta P K, Balyan H S, Sharma P C, Ramesh B. Microsatellites in plants: A new class of molecular markers. *Current Science*, 1996, 70: 45-54.
- [16] Han Z G, Wang C B, Song X L, Guo W Z, Gou J Y, Li C H, Chen X Y, Zhang T Z. Characteristics, development and mapping of *Gossypium hirsutum* derived EST-SSRs in allotetraploid cotton. *Theoretical and Applied Genetics*, 2006, 112: 430-439.
- [17] Wang C B, Guo W Z, Cai C P, Zhang T Z. Characterization, development and exploitation of EST-derived microsatellites in *Gossypium raimondii* Ulbrich. *Chinese Science Bulletin*, 2006, 51: 557-561.
- [18] Eujayl I, Sorrells M E, Baum M, Wolters P, Powell W. Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. *Theoretical and Applied Genetics*, 2002, 104: 399-407.
- [19] Yu J K, Dake T M, Singh S, Benscher D, Li W, Gill B, Sorrells M E. Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat. *Genome*, 2004, 47: 805-818.
- [20] Yu J K, La R M, Kantety R V, Sorrells M E. EST-derived SSR markers for comparative mapping in wheat and rice. *Molecular Genetics and Genomics*, 2004, 271: 742-751.
- [21] Leigh P, Lea V, Wolters P, Powell W, Donini P. Assessment of EST- and genomic microsatellite markers for variety discrimination and genetic diversity studies in wheat. *Euphytica*, 2003, 133: 359-366.
- [22] Nicot N, Chiquet V, Gandon B, Amilhat L, Legeai F, Leroy P, Bernard M, Sourdille P. Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs). *Theoretical and Applied Genetics*, 2004, 109: 800-805.
- [23] Gao L F, Jing R J, Huo N X, Li Y, Li X P, Zhou R H, Chang X P, Tang J F, Ma Z Y, Jia J Z. One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat. *Theoretical and Applied Genetics*, 2004, 108: 1392-1400.
- [24] Zhang LY, Bernard M, Leroy P, Feuillet C, Sourdille P. High transferability of bread wheat EST-derived SSRs to other cereals. *Theoretical and Applied Genetics*, 2005, 111: 677-687.
- [25] 陈军方, 任正隆, 高丽锋, 贾继增. 从小麦 EST 序列中开发新的 SSR 引物. 作物学报, 2005, 31(2): 154-158.  
Chen J F, Ren Z L, Gao L F, Jia J Z. Developing new SSR markers from EST of wheat. *Acta Agronomica Sinica*, 2005, 31(2): 154-158. (in Chinese).
- [26] 陈海梅, 李林志, 卫宪云, 李斯深, 雷天东, 胡海洲, 王洪刚, 张宪省. 小麦 EST-SSR 标记的开发、染色体定位和遗传作图. 科学通报, 2005, 50(20): 2208-2216.  
Chen H M, Li L Z, Wei X Y, Li S S, Lei T D, Hu H Z, Wang H G, Zhang X S. Development, chromosome location and genetic mapping of EST-SSR markers in wheat. *Chinese Science Bulletin*, 2005, 50(20): 2328-2336.
- [27] Li L Z, Wang J J, Guo Y, Jiang F S, Xu Y F, Wang Y Y, Pan H T, Han G Z, Li R J, Li S S. Development of SSR markers from ESTs of gramineous species and their chromosome location on wheat. *Progress in Natural Science*, 2008, 18: 1485-1490.
- [28] Thiel T, Michalek W, Varshney R K, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics*, 2003, 106: 411-422.
- [29] Chabane K, Ablett G A, Cordeiro G M, Valkoun J, Henry R J. EST versus genomic derived microsatellite markers for genotyping wild and cultivated barley. *Genetic Resources and Crop Evolution*, 2005,

- 52: 903-909.
- [30] Varshney R K, Grosse I, Hähnel U, Siefken R, Prasad M, Stein N, Langridge P, Altschmied L, Graner A. Genetic mapping and BAC assignment of EST-derived SSR markers shows non-uniform distribution of genes in the barley genome. *Theoretical and Applied Genetics*, 2006, 113: 239-250.
- [31] Hackauf B, Wehling P. Identification of microsatellite polymorphisms in an expressed portion of the rye genome. *Plant Breeding*, 2002, 121: 17-25.
- [32] Khlestkina E K, Than M H M, Pestsova E G, Röder M S, Malyshev S V, Korzun V, Börner A. Mapping of 99 new microsatellite-derived loci in rye (*Secale cereale* L.) including 39 expressed sequence tags. *Theoretical and Applied Genetics*, 2004, 109: 725-732.
- [33] Cho Y G, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch S R, Park W D, Ayres N, Cartinhour S. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics*, 2000, 100: 713-722.

(责任编辑 郭银巧)