

# EST 分子标记开发研究进展

陈全求, 詹先进, 蓝家祥, 黄云

(湖北省农业科学院经济作物研究所, 湖北武汉 430064)

**摘要:** EST(expressed sequence tags)计划作为植物基因组计划的一个重要组成部分, 已经在多种植物物种中开展起来, EST 资源库的不断扩充为利用这些数据来开发 EST 分子标记奠定了基础。概括介绍了几种主要 EST 标记(EST-SSR, EST-PCR, EST-SNP, EST-AFLP, EST-RFLP)的开发策略, 同时也提出了 EST 标记开发存在的一些问题及解决方法。

**关键词:** EST; 分子标记; 开发

**中图分类号:** Q341 **文献标识码:** A

## Study Progress in Development of EST (expressed sequence tags)

Chen Quanqiu, Zhan Xianjin, Lan Jiayang, Huang Yun

(Institute of Industrial Crops, Hubei Academy of Agricultural Sciences, Wuhan 430064)

**Abstract:** The plant EST (expressed sequence tags) program is one of the most important parts which compose the plant genome program. The expansion of EST data base is foundation of development of EST markers. In this paper, the development of five EST markers (EST-SSR, EST-PCR, EST-SNP, EST-AFLP, and EST-RFLP). Some problem exit in the development of EST markers is also discussed.

**Key words:** EST, molecular marker, development

EST (expressed sequence tags)是长约 150~500bp 的基因表达序列片段。EST 技术是将 mRNA 反转录成 cDNA 并克隆到质粒或噬菌体载体构建成 cDNA 文库后, 大规模随机挑选 cDNA 克隆, 对其 5' 或 3' 端进行一步法测序, 所获序列与基因数据库中已知序列进行比较, 从而获得对生物体生长、发育、代谢、繁殖、衰老死亡等一系列生理生化过程认识的技术。

植物 EST 计划作为植物基因组计划的一个重要组成部分, 已经在拟南芥、水稻、玉米、小麦、油菜等许多植物开展起来, 随着 EST 计划在不同物种间的不断扩展和研究的深入, DNA 测序技术的不断进步, EST 的数量也在呈指数级的惊人速度增长。1991 年各个公共数据库中的 EST 数目不足 2000 条<sup>[1]</sup>。1993 年, 美国国家生物技术信息中心(National center of Biotechnology information, NCBI) 建立了一个专门的 EST 数据库 dbEST (data base of EST)来收集和保存所有的 EST 数据。EST 资源库的不断扩充极大地方便和加快了人们

在生命科学领域的研究, 也为利用这些数据来开发 EST 分子标记奠定了基础。

## 1 EST 标记的开发策略

### 1.1 EST 数据的取得与前期处理

从数据库中直接获取的 EST 中包含一些低质量片段(<100bp), 同时存在着带有少量载体序列及末端存在 polyA/T“尾巴”的序列, 影响相关信息的分析, 所以在开发标记之前应去除这些“噪音”。这些处理可通过一些软件来进行, 如利用 EST-trimmer(<http://pgrc.ipk-gatersleben.de/>)和 cross-match([www.phrap.org](http://www.phrap.org))可分别用于去除“尾巴”和屏蔽载体序列。

### 1.2 EST 聚类

EST 是随机选取测序的, 因此导致同一基因重复测序的冗余现象也就不可避免。为此必须通过一些软件(如 Phrap 等)进行拼接和聚类来去除这些冗余的 EST, 以避免针对同一基因位点标记的重复建立而造成人力物力的浪费。而就 EST-SNP 的开发, 聚类的目

**第一作者简介:** 陈全求, 男, 1981 年出生, 湖北武汉人, 助理研究员, 硕士, 主要从事棉花育种工作。通信地址: 430064 湖北武汉市 湖北省农业科学院经济作物研究所。Tel: 027-87389507。E-mail: cq006@163.com。

**收稿日期:** 2008-05-01, **修回日期:** 2008-05-29。

的并非剔除冗余 EST, 而是为得到多序列聚类簇(multi-member cluster), 从而发掘多态性的单位点。

## 2 各类 EST 标记的开发

根据开发的方法不同, EST 标记可分为 4 类: (1) EST-SSR 和 EST-PCR(微卫星)标记, 这一类以 PCR 技术为核心, 操作简便、经济, 是目前研究和应用最多的一类; (2) EST-SNP(单核苷酸多态性), 它是以特定 EST 区段内单个核苷酸差异为基础的标记, 可依托杂交、PCR 等较多种手段进行检测; (3) EST-AFLP, 它是以限制性内切酶技术和 PCR 相结合为基础的标记; (4) EST-RFLP, 它是以限制性内切酶和分子杂交为依托, 以 EST 本身作为探针, 与经过不同限制性内切酶消化后的基因组 DNA 杂交而产生的。

### 2.1 EST-SSR 标记的开发

SSR (Simple Sequence Repeats, 简单序列重复) 也称作微卫星 (microsatellite) 是指 1-6 个碱基长度的核苷酸单位以多次重复串联排列在基因组上的一段序列。SSR 标记具有多态性高、呈共显性遗传、数量丰富、

易于用 PCR 检测和在基因组上分布均匀等特点<sup>[2]</sup>。在用传统方法开发 SSR 标记的过程中, 需要经过文库的构建、含有 SSR 克隆的识别和筛选、序列测序并分析、引物设计、PCR 引物检测应用 SSR 标记六步, 步骤繁琐而且还需要投入大量的人力物力。其中基因组文库的构建和筛选是十分费时费力的, 而且要求的技术平台比较高。因此, 在改进 SSR 位点的分离方法时, 人们首先想到了如何能够避免文库的构建和筛选。

根据 EST 建立 SSR 标记是目前采用较多的一种策略。在经过前处理后, 对序列进行组装, 去除冗余序列和延长转录片段的长度, 利用软件根据一定的鉴定标准搜索 SSR, 根据返回结果分析 EST 中 SSR 的频率、特点和分布, 然后选取目标 SSR 设计引物, 即可通过 PCR 建立 EST-SSR 标记。表 1 中归纳了 SSR 位点搜寻专用软件, 这些软件都有着各自的特点, 有的软件可以成批处理数据, 而有的软件只能逐条处理序列, 而且这些软件的搜寻算法也不尽相同, 因此, 在研究中将不同的软件结合起来使用, 相互验证, 可以获得更可靠的结果。

表 1 SSR 位点鉴定软件

软件名称	来源
SSRFinder	Gao et al., 2003
BuildSSR	Rungis et al., 2004
SSRHunter	李强, 万建民, 2005
MIcroSATellite (MISA)	<a href="http://pgrc.ipk-gatersleben.de/misa/">http://pgrc.ipk-gatersleben.de/misa/</a>
SSR Identification Tool (SSRIT)	<a href="http://www.gramene.org/gramene/searehes/ssrtool">http://www.gramene.org/gramene/searehes/ssrtool</a>
CUGIssr	<a href="http://www.genome.clemson.edu/projects/ssr/">http://www.genome.clemson.edu/projects/ssr/</a>
Sputnik	<a href="http://www.abajian.net/sputnik">http://www.abajian.net/sputnik</a>
Modified Sputnik II	<a href="http://wheat.pw.usda.gov/ITMI/EST-SSR/LaRota/">http://wheat.pw.usda.gov/ITMI/EST-SSR/LaRota/</a>

表 2 EST-SSR 在不同作物中的发生频率

作物	EST-SSR 发生频率	参考文献	作物	EST-SSR 发生频率	参考文献
小麦	1/17.4kb	Gao et al. 2003	小麦	1/5.5kb	Peng et al. 2005
水稻	1/11.8kb		大麦	1/6.3kb	Thiel et al. 2003
玉米	1/28.3kb		大麦	1/4.8kb	Varshney et al. 2006
大豆	1/23.8kb	金基强等, 2006	柑橘	1/2.8kb	Jiang et al. 2006
茶树	1/2.61kb		咖啡	1/7.7kb	Poncet et al. 2006
白菜	1/3.9kb	葛佳等, 2005	棉花	1/14.8kb	王长彪等 2006

研究表明, 在不同物种的 EST 序列上, SSR 发生的频率不同。虽然不同物种 EST 序列中 SSR 的发生频率差异很大(表 2), 但 2% 以上的 EST 序列中都含有 SSR(Cordeiro et al., 2001)。在小麦 (Eujayl et al., 2001)、大麦 (Kota et al., 2001; Thiel et al., 2003)、玉米、高粱、水稻 (Kantety et al., 2002)、甘蔗 (Cordeiro et al., 2001)、

番茄 (Areshehenkova and Ganai, 2002)、葡萄 (Scott et al., 2000)<sup>[3-10]</sup>等多种植物中都建立了 EST-SSR 标记。

基于 EST 序列开发的 SSR 标记同基于基因组文库开发的 SSR 标记相比有很多内在的优点: (1) 由于利用的是公共序列, 省去了 SSR 引物开发过程中的克隆和测序步骤, EST-SSR 标记的开发过程简单成本低;

表 3 EST-SSR 在不同物种之间的可转移性

EST-SSR 源作物	可转移的作物	参考文献
大麦	小麦、黑麦、水稻	Thiel et al. 2003; Varshney et al. 2005
高羊茅	黑麦草属、水稻、小麦	Saha et al. 2004
	大麦、玉米、水稻	Yu et al. 2004a; 2004b
小麦	大麦、玉米、水稻、黑麦、燕麦	Gupta et al. 2003
	水稻、玉米、大豆	Gao et al. 2004
	山羊草属和小麦属复合体的 18 个种	Bandopadhyay et al. 2004
白菜	油菜、玉米、高粱、水稻、茶树	忻雅等, 2006

(2)EST-SSR 标记来自于比较保守的转录区,因此其在相关物种之间具有很高的可转移性和通用性(表 3),使之在比较基因组学研究、合并不同遗传图谱、定位候选基因等研究中比基因组 SSR 更有价值;(3)EST-SSR 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,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>[11]</sup>;(4)EST-SSR 标记通常都代表着某种功能,这种功能可以通过序列同源性比对获得;(5)EST-SSR 高质量标记的比率要比基因组 SSR 高。

EST-SSR 标记的多态性比来自于基因组的 SSR 的多态性低。Eujayl 等 (2002) 研究表明,在小麦中,EST-SSR 的多态性为 25%,比基因组 SSR 的多态性 (53%)低,在大麦中,来自于 3'-UTR 的 EST-SSR 的多态性比自于 5'-UTR 的 EST-SSR 的多态性高,而在甘蔗中则有相反的结果<sup>[7]</sup>。Scott 等(2000) 研究表明,多态性的高低取决于筛选引物的材料,若材料来源于同一物种,则自于 3'-UTR 的 EST-SSR 的多态性高,若来源于不同的物种,则自于 5'-UTR 的 EST-SSR 的多态性高。虽然,EST-SSR 的多态性低,但是仍然有少量引物的多态性较高<sup>[10]</sup>。

在应用上,EST-SSR 和基因组 SSR 都有相同的用途,但由于 EST-SSR 来自于转录区,与位置不确定的基因组 SSR 相比具有更高的使用价值:当 EST-SSR 用于种质资源评价时,它表现的是转录区的差异,因而能够反映出“真实的遗传多样性”;在应用于分子标记辅助选择 (marker assisted selection, MAS) 时,当 EST-SSR 位于控制目标性状的基因内部时,可以进行直接的等位基因选择;此外,由于 EST-SSR 具有很高的可转移性,因此非常适用于比较作图研究和合并不同遗传图谱的锚定位点。

与目前广泛使用的基因组 SSR 相比,EST-SSR 具有更多优良的性质,因此,随着 EST-SSR 标记被大量

的开发,EST-SSR 标记将有着更广泛的用途。

## 2.2 EST-PCR 标记的开发

根据 EST 建立常规 PCR 标记也是一条可行的途径。与 EST-SSR 标记一样,这两类标记都是以 PCR 为基础的,其技术难度较小,使用成本较低且准确度较高,所以易为人们所接受。

根据 EST 序列设计引物对特定区域进行扩增,这样就可能揭示出不同材料在编码区、非编码区及调控区序列的差异。目前,EST-PCR 标记已在云杉(Schubert et al., 2001)、火炬松(Temesgen et al., 2001)、洋葱 (McCallum et al., 2001)和白菜(忻雅等, 2006)<sup>[12-15]</sup>等植物中得到了应用。

扩增片段能否成为遗传标记的关键是其是否具有多态性。由于 EST 的高保守性,故以 PCR 为基础的 EST 标记在种内的多态性较低。通过以下几种策略可提高多态性的频率:(1)在选择 EST-PCR 和 EST-SSR 标记时,尽量靠近 3' 或 5' 端非编译区段,因为这些区域变异性较高<sup>[10]</sup>。(2)对无多态性的 EST-PCR 扩增产物先用限制性内切酶消化,然后对酶切产物进行电泳分离。这种策略的实质是在确定了扩增产物分子量大小差异的基础上,进一步检测片段内部序列的差异,其多态性的产生依赖于引物与限制性内切酶位点的组合,即通常所说的切割扩增多态性 (cleavage amplified polymorphism, CAP),故可提高多态性 EST-PCR 标记的频率,这在对本柳杉和洋葱的 EST-CAP 标记研究中已得到了证明<sup>[14,16]</sup>。(3)改进扩增产物的分析手段。采用分辨率较高的聚丙烯酰胺凝胶或变性梯度胶来进行分离<sup>[13]</sup>。

## 2.3 EST-SNP 标记的开发

单核苷酸多态性 (Single nucleotide polymorphisms, SNP) 是指同一位点的不同等位基因之间只有一个核苷酸的差异或只有一个碱基的插入、缺失。SNP 以十分高的频率在基因组上发生,非常适合于高通量



分析,更重要的是它能展示出不能被其它标记所检测到的隐藏的多态性。SNP为我们了解基因的复杂性提供了丰富的资源。有人将SNP标记称为第三代DNA分子标记。SNP不仅分布在非编码区,而且还分布在编码区(Collins et al., 1997; Harding et al., 1997; Nickerson et al., 1998),存在于编码区的SNP称为cSNP(Garg et al., 1999),被认为是应用前景最好的遗传标记物<sup>[17-19]</sup>。

EST是随机挑选cDNA克隆单侧测序得到的,来自于不同cDNA文库和不同个体的大量冗余的EST是发掘SNP很好的资源。但EST还没有被广泛地用于SNP标记的开发,其中一个主要原因是EST序列的低质量特性。通常情况下人们无法区分一个位点上碱基的变化到底是由测序误差引起的,还是由于自然存在的序列突变导致的。

虽然SNP在基因组中广泛存在,但其在不同的基因组中的分布频率并不一致,并且在同一个基因组中,编码区和非编码区的分布频率差异也很大。在大麦无内含子的Isa基因上SNP的发生频率是1/27bp(Bundock et al., 2004)<sup>[21]</sup>,在大麦P450基因家族成员的外显子区SNP的发生频率是1/131bp(Bundock et al., 2003)<sup>[22]</sup>,而在甘蔗的EST中,SNP的发生频率是1/50bp<sup>[23]</sup>。

EST-SNP可以通过比对多序列簇的EST以获取序列的多态性信息,包括候选的SNP位点。此外,NCBI的UniGene数据库提供的聚类基因簇(cluster),也是我们开发SNP的一个重要数据源。一些专用软件已被开发出来用于在EST序列中搜寻SNP位点,如SNiPpER(Kota et al., 2003)、SNPServer(Savage et al., 2005)等,这些软件通过一定的算法可以在一定程度上区分测序误差和自然变异,从而大大方便了研究人员利用公共EST序列信息开发EST-SNP标记<sup>[24,25]</sup>。虽然从EST中发掘SNP所用的软件不同,其基本思想都是一致的:先聚类同源序列,然后比较这些同源序列,phred, phrap, polyphred, consed是比较常用的分析这类问题的软件包。从EST中发掘SNP关键的一点是根据研究目的选择高质量的EST序列,如来源于同一物种的不同品种同一组织的EST序列,来研究同一物种不同品种同一组织的差异。Picoult-Newberg等(1999)用phred和phrap等软件也开发了一个从EST中发掘SNP的系统,并从300000条序列中辨识出了850个SNP候选位点<sup>[26]</sup>。

值得注意的是,EST-SNP多态性有些可能只是测序时碱基响应(base call)的问题,或cDNA反转录和大

肠杆菌聚合酶复制过程产生的错配。所以筛选这些SNP时应注意以下几点:(1)一些聚类的重叠区域质量比较低,往往发生连续的不配对,这些一般确认率比较低,所以应尽量选择两侧有一定数量严格配对(perfect match)的碱基的SNP。(2)由于使用聚类程序时,为了得到更高的比对分值往往需要在一条序列中加入空格,这些人为空格有可能造成低质量数据信号。为此, Picoult-Newberg等(1999)建议,甚至可不考虑这些可能是插入或缺失的SNP位点而只关注替换类型<sup>[26]</sup>。(3)由于每个碱基响应质量与其所在的位置相关,如测序前区段的序列质量往往较低,因此可选择序列质量比较高的前100个碱基以后的SNP。(4)由于同一个碱基位置上随机发生两次错误的概率很低,为了得到高质量的SNP,可以规定重叠区域必须要有两条或两条以上的序列来佐证这个SNP位点,一些学者已经编写了相应的程序可自动筛选SNP,如AutoSNP和SNiP<sup>[27,28]</sup>。

目前,在玉米(Useche et al., 2001; Barker et al., 2003)、拟南芥(Iwata et al., 2001)、大麦(Kota et al., 2001)等植物中已有EST-SNP的开发报道<sup>[5,16,27,29]</sup>。

#### 2.4 EST-AFLP 标记的开发

EST-AFLP技术结合了RT-PCR和AFLP技术,以cDNA作为操作对象,用限制性内切酶酶切双链,酶切片段加人工接头后,利用与接头序列互补的引物进行预扩增和选择性扩增,并进行电泳显示。在选择内切酶时,可以充分利用已知的EST序列进行分析,或根据EST的3'端和5'端非翻译区富含A/T的特性,选择识别序列的碱基组成集中在有义的区段的内切酶,实现有目的的选择扩增。其最大的特点就是特异性强,对低丰富度的表达产物也比较敏感,并且在转录产物高度表达时,扩增条带的强度还能准确反映基因间表达量的差别。Qin等(2001)开发了GenEST软件,可以利用EST数据对<sup>[30]</sup>DNA进行AFLP操作,有效地实现EST数据与cDNA-AFLP的信息、的双向连接和转化。他们将感染了马铃薯胞囊线虫的材料根据感染程度分5个阶段取材,用cDNA-AFLP方法结合256对引物进行分析,共获得了8200个TDF(transcripts-derived fragment)。但是目前由于相关基因的EST信息量的制约,而且完成此类分析需要昂贵的实验设备和试剂,技术要求高,在一般实验室仍不易推广。

#### 2.5 EST-RFLP 标记的开发

EST-RFLP标记与一般的RFLP标记相似,只是所用的探针是cDNA,即EST本身,其多态性的产生依

赖于探针与不同限制性内切酶之间的组合。在建立 EST-RFLP 标记时, 应尽量选择那些单拷贝或低拷贝的片段作探针, 以降低后续标记分析的复杂性。这也是早期建立 EST 标记并将其绘制到遗传图谱的主要方法。在大麦上, Schwarz 等(2002)已用 cDNA 作为筛选克隆的有效探针来构建染色体的物理图谱<sup>[31]</sup>。

EST-RFLP 标记是共显性标记, 可靠性高, 在揭示植物的遗传信息和加速比较基因组研究等方面都起到了重要作用。但开发这类标记涉及到分子杂交和探针标记等繁琐的过程, 技术要求较高, 而且费用昂贵, 现在已很少用。

### 3 小结

虽然 EST 标记除具有一般分子标记的特点外, 还有其特殊优势, 有着多方面的利用价值, 但是 EST 标记的开发也存在着一些问题: (1) 目前注册的 EST 为一次性测序, 其中存在着一定的错误信息; (2) mRNA 存在选择性剪接, 事实上利用软件进行序列拼接时错拼是很难避免的; (3) EST 研究中有相当一部分为未知基因, 利用这些 EST 开发的分子标记, 不易很快与功能建立联系; (4) 由于生物信息学的有关软件使用的算法不同以及设置的参数严谨度不同, 得出的结果不尽相同, 如 SNP 的颠换与转换、SSR 出现的频率等; (5) 基于 PCR 的 EST 分子标记是以长度多态性为基础的, 其分辨取决于高分辨率的凝胶, 然而由于高频率非长度变异的等位基因的存在, 这些信息检测存在一定难度<sup>[32]</sup>; (6) EST 的保守性在一定程度上也限制了 EST 标记的多态<sup>[33]</sup>。

尽管存在以上问题, 但随着 EST 计划开展、深入以及相关研究技术和分析手段的不断改进并走向成熟, EST 数据资源将不断丰富, 而其本身又具备独特的优势和多方面的利用价值, 开发比较简便, 我们相信 EST 标记将会大规模开发并广泛应用于生物之间关系的研究。

### 参考文献

- [1] 邱咏梅, 夏庆友. EST 规模测序在家蚕功能基因组中的应用. 蚕学通报, 2002, 13(4): 413-418.
- [2] Powell W, Gordon C Machray, Provan J. Polymorphism revealed by simple sequence repeats. Trends in Plant Science, 1996, 1(7): 215-222.
- [3] Cordeiro, GM, Casu R, McIntyre CL. et al. Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to helianthus and sorghum. Plant Science, 2001, 160(6): 1115-1123.
- [4] Eujayl I, Sorrells M, Baum M, Wolters P, Powell W. Assessment of genotypic variation among cultivated durum wheat based on EST-SSRs and genomic SSRs. Euphytica, 2001, 119(12): 39-43.
- [5] Kota R, Varshney R K, Thiel T, et al. Generation comparison of EST-derived SSRs and SNPs in barley (*Hordeum vulgare* L.). Hereditas, 2001, 135: 145-151.
- [6] Thiel T, Miechalek W, Varshney R K, et al. Exploiting EST database for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theoretical and Applied Genetics, 2003, 106: 411-422.
- [7] Eujayl I, Sorrells M E, Baum M, et al. Isolation of EST derived microsatellite markers for genotyping the A and B genomes of wheat. Theoretical and Applied Genetics, 2002, 104(23): 399-407.
- [8] Kantety R V, LaRota M, Matthews D E, et al. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Molecular Biology, 2002, 48(5): 501-510.
- [9] Areshchenkova T, Ganal M W. Comparative analysis of Polymorphism and chromosomal location of tomato microsatellite markers isolated from different sources. Theoretical and Applied Genetics, 2002, 104: 229-235.
- [10] Scott K D, Eggler P, Seaton G, et al. Analysis of SSRs derived from grape ESTs. Theoretical and Applied Genetics, 2000, 100(5): 723-726.
- [11] Chen X, Salamini F, Gebhardt C. A potato molecular function map for carbohydrate metabolism and transport. Theoretical and Applied Genetics, 2001, 102(2): 284-295.
- [12] Schubert R, Starek G M, Riegel R. Development of EST-PCR markers and monitoring their intra populational genetic variation in *Picea abies* (L.) Karst. Theoretical and Applied Genetics, 2001, 103: 1223-1231.
- [13] Temesgen B, Brown G R, Harry D E, et al. Genetic mapping of expressed sequence tag polymorphism (ESTP) markers in loblolly pine (*Pinus taeda* L.). Theoretical and Applied Genetics, 2001, 102: 664-675.
- [14] McCallum J, Leite D, Pither-Joyee M, et al. Expressed sequence markers for genetic analysis of bulb onion (*Allium cepa* L.). Theoretical and Applied Genetics, 2001, 103: 979-991.
- [15] 忻雅, 崔海瑞, 张明龙, 等. 白菜 EST-SSR 标记的通用性. 细胞生物学杂志, 2006, 28(2): 248-252.
- [16] Iwata H, Ujino-Ihara T, Yoshimura K, et al. Cleaved amplified polymorphic sequence markers in sugi and their locations on a linkage map. Theoretical and Applied Genetics, 2001, 103: 881-895.
- [17] Collins E S, Guyer M S, Chakravati A. Variations on a theme: Cataloging human DNA sequence variations, Science, 1997, 278: 1580-1581.
- [18] Harding R M, Rullerton S M., Griffiths R C. Archaic African and Asian lineages in the genetic ancestry of modern humans. Am. J. Hum. Genet, 1997, 60: 772-789.
- [19] Nickerson D A, Tayloe S L, Weiss K M, et al., DNA sequence diversity in a 9.7kb region of the human lipoprotein lipase gene. Nat. Genet, 1998, 19: 233-240.
- [20] Garg K, Green P, Deborah A. Nickerson, identification of Candidate Coding Region Single Nucleotide Polymorphisms in 165 Human Genes Using Assembled Expressed Sequence Tags, Genome Re-

- search,1999,9:1087-1092.
- [21] Bundock P C, Henry R J. Single nucleotide polymorphism, haplotype diversity and recombination in the *Isa* gene of barley. *Theoretical and Applied Genetics*,2004,109(3):543-551.
- [22] Bundock P C, Christopher J T, Eggler P, et al. Single nucleotide polymorphisms in cytochrome P450 genes from barley. *Theoretical and Applied Genetics*,2003,106(4):676-682.
- [23] Cordeiro GM., Eliot F, McIntyre CL, et al. Characterization of single nucleotide polymorphisms in sugarcane ESTs. *Theoretical and Applied Genetics*,2006,113(2):331-343.
- [24] Kota R, Rudd S, Facius A, et al. Snipping polymorphisms from large EST collections in barley (*Hordeum vulgare* L.). *Molecular Genetics and Genomics*,2003,270(1):24-33.
- [25] Savage D, Batley J, Erwin T, et al. SNPServer: a real-time SNP discovery tool. *Nucleic Acids Research*,2005,33:493-495.
- [26] Picoult-Newberg L, Ideker TE, Pohl MG, et al. Mining SNPs from EST databases. *Genome Research*.1999,9:167-174.
- [27] Barker G, Batley J, Sullivan H O, et al. Redundancy based detection of sequence Polymorphisms in expressed sequence tag data using auto SNP. *Bioinformatics*,2003,19:421-422.
- [28] Rudd S, Schoof H, Mayer K. PlantMarkers: a database of predicted molecular markers from Plants. *Nucleic Acids Research*,2005,33:628-632.
- [29] Useche F J, Gao G, Hanafey M, et al. High throughput identification database storage and analysis of SNPs in EST sequence. *Genome Informatics*,2001,12:194-203.
- [30] Qin L, Prins P, Jones J T, et al. a powerful bidirectional link between cDNA sequence data and gene expression profiles generated by cDNA-AFLP. *Nucleic Acids Res.*2001,29(7):1616-1622.
- [31] Schwarz G, Miehalek W, Jahoor A, et al. Direct selection locus of expressed Sequence on a YAC clone revealed praline-rich-like genes and BAKE-1 sequences physically linked to the complex *Mla* powdery mildew resistance locus of barley (*Hordeum vulgare* L.). *Plant Science*,2002,163:307-311.
- [32] Decroocq V, Fave M G, Hagen L, et al. Development and transferability of apricot and grape EST microsatellite markers across taxa. *Theoretical and Applied Genetics*,2003,106:912-922.
- [33] Zhang P, Dreisigacker S, Melchinger A E, et al. Quantifying novel sequence variation and selective advantage in synthetic hexaploid wheat and their backcross-derived lines using SSR markers. *Molecular Breeding*,2005,15(1):1-10.