

大豆 EST-SNP 的挖掘鉴定及其 CAPS 标记的开发

束永俊 李勇 吴娜拉胡 柏锡 才华 纪巍 朱延明*

东北农业大学生命科学学院, 黑龙江哈尔滨 150030

摘要: 采用生物信息学方法将大豆 EST 序列联配到大豆基因组序列上, 挖掘到大豆 EST-SNP 位点 537 个。对其靶向基因进行功能注释分析, 发现它们主要参与亚细胞定位、蛋白质结合与催化以及代谢等与大豆重要农艺性状形成相关的生物过程。同时开发了简便易行的 SNP 检测方法, 利用 EMBOSS 软件筛选导致酶切位点改变的 EST-SNP, 分别以大豆绥农 14、合丰 25、Acher、Evans、Peking、PI209332、固新野生大豆、科丰 1 号、南农 1138-2 的 DNA 及其混合的 DNA 为模板, 设计引物进行 PCR 扩增, 发现 44 个 PCR 产物中有 36 个测序峰图在预期的 EST-SNP 位点表现出多态性。酶切分析发现 26 个 PCR 产物具有酶切多态性, 可以作为 CAPS 标记。结果表明该 EST-SNP 挖掘体系及其 CAPS 标记转化系统具有高效率、低成本等优点, 有利于促进大豆的遗传育种研究。

关键词: 大豆; 表达序列标签; 基因组序列; 单核苷酸多态性; 酶切扩增多态性序列

Mining and Identification of SNP from EST Sequences and Conversion of CAPS Markers in Soybean

SHU Yong-Jun, LI Yong, WU Na-La-Hu, BAI Xi, CAI Hua, JI Wei, and ZHU Yan-Ming*

College of Life Science, Northeast Agricultural University, Harbin 150030, China

Abstract: SNPs widely distribute throughout genomes from non-coding regions to coding regions, constituting the most abundant molecular markers used in animal and plant genetic breeding. With the rapidly growing genome sequencing projects, a large amount of genomic and EST sequences has become available to the public. Many SNPs are identified by comparing genome sequences or ESTs obtained from genetically diverse lines or individuals in plants. However the SNP assay always relies on expensive equipments or reagent, which has limited the application of SNPs in genetics and breeding especially in plants. The CAPS marker, also known as PCR-RFLP marker, is the technique combining PCR and restriction enzymes digestion to detect the restriction fragment length polymorphisms. With the development of high-throughput sequencing technology, more and more SNPs are identified, among them many mutations have altered the restriction enzymes recognition sites. This provides an opportunity for high-through development of CAPS markers. With soybean genome sequences becoming available, high-throughput SNP marker development will significantly improve genetic mapping, map based cloning in soybean. To discover new SNPs in soybean, we aligned the ESTs with whole genome sequences in different soybean varieties and identified 537 EST-SNPs. The function of genes targeted by these EST-SNPs was analysed, the results showed that these genes participated in subcellular localization, protein binding or catalyzing, metabolic process and cell rescue, defense and disease resistance, etc. Most of these functions are involving in various physiological and biochemical processes influencing important agronomic traits. To develop easy assay method for these EST-SNPs, we identified the EST-SNPs which alter the restrict enzyme recognition sites by software EMBOSS, and 48 pair primers were designed to detect these EST-SNPs. forty-four pair primers amplified single bands (400–800 bp) from genomic DNA of Suinong 14 widely planted in the Northeast China. To verify the SNP polymorphisms, we used these primer pairs for PCR amplification from genomic DNA of Suinong 14, Hefeng 25, Acher, Evans, Peking, PI209332, Guxin wild soybean, Kefeng 1, Nannong 1138-2 and pool DNA of the nine soybean varieties. The PCR amplicons were sequenced, the traces of the 36 discordant ones were detected as candidate SNPs, which were then validated by re-sequencing the individuals. SNPs were identified using restriction enzymes, and the products of 26 pair primers with unequivocal restriction pattern were identified as CAPS markers. The SNPs discovery and CAPS markers conversion system developed in this study is fast, low cost and effecient, and holds great promise for molecular assisted breeding of soybean.

Keywords: Soybean; EST; Genome sequences; SNP; CAPS

本研究由国家高技术研究发展计划(863 计划)项目(2006AA100104 和 2008AA10Z153)资助。

* 通讯作者(Corresponding author): 朱延明, E-mail: ymzhu2001@yahoo.com.cn; Tel: 0451-55190734

第一作者联系方式: E-mail: syjun2003@126.com

Received (收稿日期): 2009-11-16; Accepted (接受日期): 2010-01-08.

SNP (single nucleotide polymorphism), 单核苷酸多态性, 是指基因组范围内单个碱基的插入、缺失、转换、颠换或小片段的突变等。它广泛分布于基因组范围内, 具有变异来源丰富、潜在数量巨大等优点, 是理想的第 3 代分子标记^[1]。传统的SNP的发现主要依靠对一些重要基因的测序, 比较同一基因在不同个体或品种间的差异, 这种方法鉴定SNP的速度较慢、成本较高, 因此发展缓慢。随着分子生物学的发展和基因组测序计划的进行, 研究人员发布了大量的基因组测序信息, 因此可以通过比较不同个体间的基因组序列, 发现新的SNP位点。如比较分析不同品系小鼠间^[2]、水稻的籼稻和粳稻两个亚种间^[3]、拟南芥的哥伦比亚和兰琳伯格生态型间的序列^[4], 发现了大量的SNP突变。而对于一些尚未完成基因组测序的物种, 或测序品种较少时, 可以通过比较分析不同个体间的EST序列信息, 发现新的SNP位点^[5], 目前在牛^[6]、玉米^[7]、小麦^[8-9]和大麦^[10-11]等物种中均有应用。

由于SNP标记检测方法操作复杂、成本昂贵, 或者需要高端的仪器设备, 如飞行时间质谱仪(MALDI-TOF), 或者需要特殊的检测试剂, 如TaqMan探针, 因此严重限制了SNP标记技术在动植物遗传育种中的应用。酶切扩增多态性序列(cleaved amplified polymorphic sequence, CAPS), 又称PCR-RFLP, 它是一种将特异引物PCR与限制性内切酶消化结合而产生的一种分子标记检测技术, 具有共显性、位点特异性、操作简单和成本低等特点, 广泛运用于植物基因的分型、定位、图位克隆和分子鉴定等方面的研究^[12]。但是, 一直以来, 由于突变的酶切位点发现得较少, CAPS标记难以大规模地开发应用。随着大豆基因组测序计划的进行, 研究人员鉴定了大量的SNP位点, 其中相当一部分SNP位点都是酶切位点突变, 这为高通量开发利用CAPS标记奠定了基础。

大豆是一种古四倍体作物, 是人类所需的植物油和蛋白的重要来源。大豆基因组复杂, 且高度自交, 导致其基因组遗传变异较低, 严重阻碍了其遗传育种的发展^[13-14]。随着大豆EST测序和基因组测序的进行, 公共数据库中积累了大量的序列信息^[15-16], 这为高通量开发SNP等分子标记提供了便利。

本研究利用大豆基因组测序和 EST 测序数据, 挖掘开发大豆 SNP 标记, 并对得到的 SNP 标记进行测序验证。根据获得的 SNP 标记位点, 进行酶切信息分析, 将其转化为方便实用的 CAPS 标记。

1 材料与amp;方法

1.1 植物材料

大豆品种合丰 25、绥农 14、Acher、Evans、Peking、PI209332 和固新野生大豆由中国农业科学院作物科学研究所惠赠。大豆品种科丰 1 号和南农 1138-2 由南京农业大学国家大豆改良中心惠赠。

1.2 基因组 DNA 的提取

采集参试大豆品种的幼叶 100 mg, 于研钵中用液氮速冻, 迅速研磨。采用CTAB法提取叶片总DNA, 保存于 50 μ L去离子水中, 并用分光光度计进行浓度测定, 稀释至 50 ng μ L⁻¹。

1.3 SNP 位点识别

大豆基因组测序数据来自Phytozome (<http://www.phytozome.net/soybean.php>, Glyma1.0), EST测序数据来自NCBI (<http://www.ncbi.nlm.nih.gov/>)。利用Blast将大豆EST序列定位到大豆基因组上, 确定EST序列所在的表达基因序列。利用EMBOSS程序包^[17]的程序Diffseq分析EST和基因组序列之间的差异, 筛选其中出现 3 次以上的差异位点, 作为候选SNP位点。利用EMBOSS中的程序Restrict分析经测序确认的SNP位点酶切信息, 选择 4 种内切酶(*Hind* III、*Pst* I、*Sca* I和*Vsp* I)的参数, 筛选突变引起酶切位点改变的候选SNP位点作为候选CAPS标记位点。利用软件Primer3^[18]在候选CAPS标记位点两侧设计PCR引物, 引物设计的标准是 T_m 值为 58 , 引物长度为 20~26 bp。

1.4 SNP 位点的功能分析

从 TAIR (<http://www.arabidopsis.org/>, TAIR7)下载拟南芥的基因组数据和注释信息, 提取大豆 SNP 所在的基因序列, 用 Blastp 比对拟南芥蛋白质序列, 提取相似性最高的序列注释信息, 作为 SNP 靶向基因的功能注释, 对 SNP 位点的注释信息进行分类。

1.5 PCR扩增和测序验证

以9个大豆品种的基因组和混合基因组DNA为模板, 25 μL 体系含 2.5 mmol L^{-1} MgCl_2 、1 U *Taq*酶、800 $\mu\text{mol L}^{-1}$ dNTP、0.2 $\mu\text{mol L}^{-1}$ 引物、模板DNA 50 ng。采用降落PCR (touchdown PCR, TD-PCR)进行扩增, 程序为 94 预变性 7 min, 94 变性 20 s, 60 退火 20 s, 每个循环降 0.5 , 72 延伸 30 s, 30 个循环, 94 变性 20 s, 45 退火 20 s, 72 延伸 30 s, 10 个循环, 72 延伸 7 min。对PCR产物采用 2%琼脂糖电泳, 将条带单一的PCR产物送交南京金思特公司测序, 对挖掘的SNP位点进行验证。

1.6 酶切位点识别和酶切分析

酶切分析(CAPS分析)参照TaKaRa的内切酶操作指南, 用限制性核酸内切酶酶切PCR产物, 酶切体系包括 10 U μL^{-1} 限制酶 0.3 μL 、1 μL buffer、5.7 μL ddH₂O、3 μL PCR产物, 酶切过夜, 采用 2%琼脂糖电泳酶切产物。

2 结果与分析

2.1 EST-SNP 突变位点的挖掘

将来源于大豆品种(科丰1号和 Nourin 2)的 62939 条 EST 序列联配到大豆基因组上, 发现 15 971 条 EST 序列与品种 Williams 82 的测序序列存在差异, 编写 Perl 脚本将这些突变进行归类分析, 发现 537 个 EST-SNP (图 1), 突变发生的类型主要是转换突变, 达到 73.0% (392/537), 颠换突变只占一小部分, 约 27.0% (145/537)。

1. Glyma01g45110.1	T	A	T	A	G	C	T	G	C	T	G	T	C	A	T	G	A	T	C	T	C	A	T	C	A	G	C	C	T	T	T	G	T	T	G	G	T
2. gi 209727901 gb BW663415.1	T	A	T	A	G	C	T	G	C	T	G	T	C	A	T	G	A	G	C	T	C	A	T	C	A	G	C	C	T	T	T	G	T	T	G	G	T
3. gi 209724465 gb BW666674.1	T	A	T	A	G	C	T	G	C	T	G	T	C	A	T	G	A	G	C	T	C	A	T	C	A	G	C	C	T	T	T	G	T	T	G	G	T
4. gi 209722038 gb BW671781.1	T	A	T	A	G	C	T	G	C	T	G	T	C	A	T	G	A	G	C	T	C	A	T	C	A	G	C	C	T	T	T	G	T	T	G	G	T
5. gi 209718954 gb BW674807.1	T	A	T	A	G	C	T	G	C	T	G	T	C	A	T	G	A	G	C	T	C	A	T	C	A	G	C	C	T	T	T	G	T	T	G	G	T
6. gi 209709101 gb BW658817.1	T	A	T	A	G	C	T	G	C	T	G	T	C	A	T	G	A	G	C	T	C	A	T	C	A	G	C	C	T	T	T	G	T	T	G	G	T
7. gi 209706260 gb BW668086.1	T	A	T	A	G	C	T	G	C	T	G	T	C	A	T	G	A	G	C	T	C	A	T	C	A	G	C	C	T	T	T	G	T	T	G	G	T
8. gi 209704894 gb BW673492.1	T	A	T	A	G	C	T	G	C	T	G	T	C	A	T	G	A	G	C	T	C	A	T	C	A	G	C	C	T	T	T	G	T	T	G	G	T
9. gi 209702399 gb BW656375.1	T	A	T	A	G	C	T	G	C	T	G	T	C	A	T	G	A	G	C	T	C	A	T	C	A	G	C	C	T	T	T	G	T	T	G	G	T
10. gi 209701444 gb BW655825.1	T	A	T	A	G	C	T	G	C	T	G	T	C	A	T	G	A	G	C	T	C	A	T	C	A	G	C	C	T	T	T	G	T	T	G	G	T

图1 EST-SNP位点的发掘

Fig. 1 Identification of EST-SNP sites

2.2 SNP 靶向基因的功能分析

提取 SNP 靶向基因的蛋白质序列, 比对拟南芥基因组蛋白质序列, 利用同源的拟南芥蛋白质功能对大豆 EST-SNP 靶向基因进行注释(图 2)。537 个 EST-SNP 定位在 333 个基因上, 这些基因主要参与亚细胞定位过程(261/1109, 23.53%)、蛋白质的结合和催化活性(152/1109, 13.71%)和代谢过程(124/1109, 11.18%)。另外, 这些 EST-SNP 靶向基因还参与一些特异的细胞过程, 如蛋白质合成、蛋白质折叠修饰、对外界刺激的反应、信号转导、细胞修复和对病毒的防御等过程, 表明这些 EST-SNP 基因广泛地与大豆的各种代谢反应过程, 影响着大豆各种农艺性状的形成。

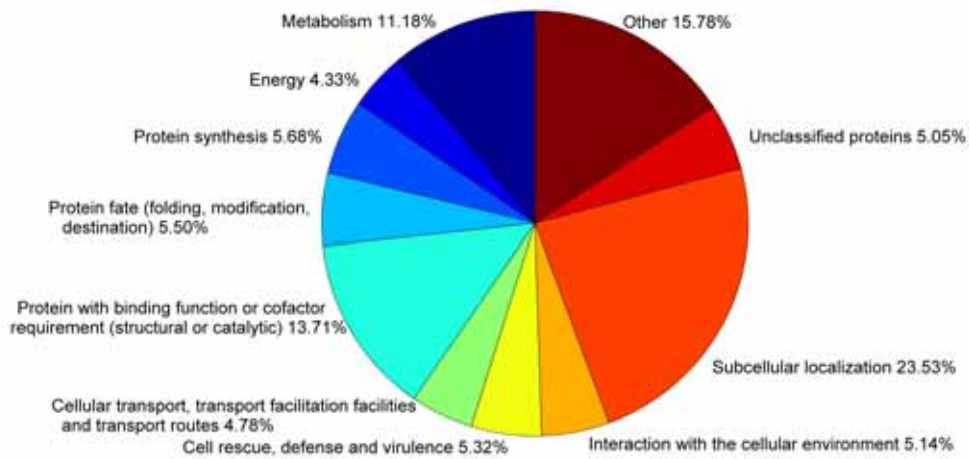


图2 EST-SNP靶向基因的功能分析

Fig. 2 Functional analysis of the genes containing EST-SNP sites

2.3 EST-SNP 的验证以及作为 CAPS 标记的开发

由于受到特殊设备或检测试剂的限制，SNP 检测较为困难。利用 EMBOSS 的 Restrict 程序分析挖掘得到的 EST-SNP 突变位点的内切酶识别信息，筛选得到酶切位点改变的候选 EST-SNP 位点 48 个，即为候选 CAPS 标记位点。设计引物扩增这些 EST-SNP 位点，其中 44 对引物具有单一条带的扩增产物。以混合 DNA 和 9 个大豆品种的 DNA 为模板进行扩增，发现 36 对引物的混合 DNA 扩增产物在 EST-SNP 位点出现紊乱，且在不同品种间出现碱基差异(图 3)。对测序验证 EST-SNP 位点的 PCR 产物进行相应的酶切，酶切产物电泳具有多态性即为 CAPS 标记(图 4)，获得的 26 个 CAPS 标记的引物以及相应的酶切信息见表 1。

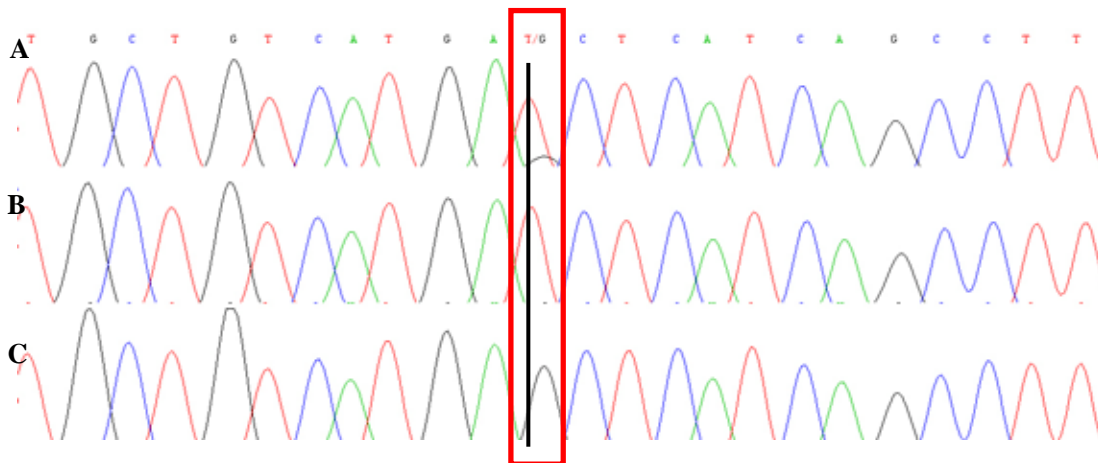


图3 EST-SNP 位点的测序验证

Fig. 3 Validation of the EST-SNP site by Sanger sequencing

A : 杂合子基因型 T/G 的测序结果 ; B : 纯合子基因型 T 的测序结果 ; C : 纯合子基因型 G 的测序结果。

A: sequencing result of heterozygote T/G; B: sequencing result of homozygote T; C: sequencing result of homozygote G.

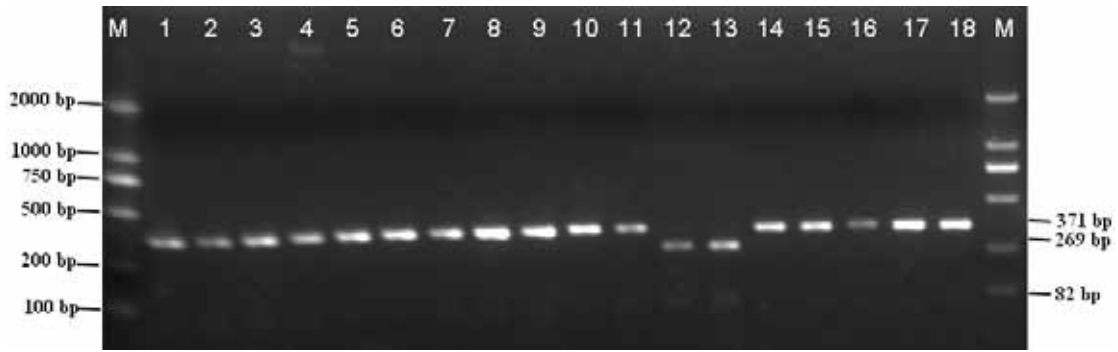


图 4 标记 ESNP02 在 9 个大豆品种中的多态性分析

Fig. 4 Polymorphism analysis of markers ESNP02 in nine soybean varieties

1~9: 基因组 DNA 的 PCR 产物; 10~18: PCR 产物的酶切产物, 引物以及内切酶信息见表 1; M: DNA marker DL2000.

1-9: PCR amplicons; 10-18: digested products of PCR amplicons; the primers and enzymes information are as given in Table 1; M: DNA molecular marker DL2000.

表 1 CAPS 标记引物及内切酶信息

Table 1 Summary information of CAPS primers and relative restrict enzymes

引物编号 Primer ID	上游引物 Forward primer	下游引物 Reverse primer	内切酶 Restrict enzyme
ESNP01	GTACCTGGAGATTGGAATGGA	TCCAGGTAAGCCTCAGGTA AAA	<i>Hind</i> III
ESNP02	CCAACAATCTCTGTTAGACAGGC	CTTCTTTGGGACTGTTGATCCT	<i>Sac</i> I
ESNP03	GACATTACAAAGACCGAAGCC	TGTCTGGAGTATATGAATGCCG	<i>Sac</i> I
ESNP04	CGTGCTTTCTTCTAGCAACT	TGTGGTGTAGGGA ACTTGATTG	<i>Sac</i> I
ESNP05	CACTTTTCCAACAAGCCTAACC	GACACGAGCTGGATAATGGTIT	<i>Vsp</i> I
ESNP06	GATCATCCCCAGCACTAAAAAC	AGTCAGCAATTTCTTCAGGAGG	<i>Vsp</i> I
ESNP07	CAAGAATTCCAGCAAGGGA	GCAATAGCTTCTGGTGTGACAAT	<i>Vsp</i> I
ESNP08	CACATCACTACCAGGAAGTCCC	CCTGCAGATGCACTACAAGTCT	<i>Vsp</i> I
ESNP09	TGTAGTGCATCTGCAGGGTAAG	AGTCAGCAAGCCAAGTAACACA	<i>Hind</i> III
ESNP10	TGGATAGCCTCAGTTGCTGATA	TAGGACCAGGTCCTGGGATAC	<i>Pst</i> I
ESNP11	CGTGCATAGACTGTTGGAGAAA	ACTTATGTGCTGTTTGCCCT	<i>Vsp</i> I
ESNP12	ACTTGGGCCTGTTAGGAAAAAC	CAAAGAACCAAGAGAATCGTCC	<i>Vsp</i> I
ESNP13	CTACAACATTCCAAGTGGGTCA	GCACACA ACTCCACGTAAACTC	<i>Vsp</i> I
ESNP14	GTTTTGGTGGTCTAAGGAGGAA	AGCTCCTCACAGAACCAACTTA	<i>Vsp</i> I
ESNP15	CGTAAACACATCCTTGGTTCC	TACAGAGTGAATCCCATTTGTC	<i>Vsp</i> I
ESNP16	GGCTTCCGTTCTTCTTTACCTT	TCTACGTCAGGGGTTCAATTCT	<i>Vsp</i> I
ESNP17	AAATGTGAGCAGTGATCCAGAG	GCCACTCAATTTGTTGCTTG	<i>Hind</i> III
ESNP18	GAACAAAGGTTGTCCTTGCACT	CAGCAAAATGGCCTCTAACTCT	<i>Vsp</i> I
ESNP19	GATGGGGTTGAGGTAAGTTTTG	GTCTGAGAAGCAAATTCGG	<i>Vsp</i> I
ESNP20	TGCAGTGGAGAGACTGAGGATA	AATTCTCAAGAGCTGCCAAGAG	<i>Vsp</i> I
ESNP21	GTGAGGAGGGATTGACTTTCAG	CTAGTTGCACTTCACAGCAGC	<i>Vsp</i> I
ESNP22	AGAGCCACCCTTTATGTCATGT	ATGAGGGTGTCTTATCGTTGG	<i>Vsp</i> I
ESNP23	ATGGGGGAGAAATGAGTAATCC	AACTGCCACTACAGGTAGCACA	<i>Hind</i> III
ESNP24	TAAAAGGAAGCTGAAGGTGGAG	GAAACACCGTTAAGAAGGTCCA	<i>Vsp</i> I
ESNP25	AATGCAGTTGATAGTGTGTGC	TGTCAGCTCAGTATTCAGTGGG	<i>Pst</i> I
ESNP26	CACACTTCAACCTTGTCAGGA	AGAGACACACATCCAGCAGAAA	<i>Pst</i> I

3 讨论

3.1 EST-SNP 标记的挖掘

EST测序是研究功能基因的重要方法,公共数据库中已经积累了大量的EST序列。在海量的EST序列中,

很多来源于不同的个体或品种，它们在组装成基因时往往会出现不一致的碱基，即为EST-SNP突变位点，这些位点可以通过生物信息学方法检测到^[5]。自从Picoult-Newberg等提出从EST数据中挖掘SNP位点的方法以来，从模式物种人类^[19]、小鼠和大鼠^[20]等，到家畜的牛^[6]和猪^[21]等都有所应用。近些年来，在植物方面从模式生物拟南芥^[22]，到主要粮食作物水稻^[3]、玉米^[7]、小麦^[8-9]和大麦^[10-11]，及一些小物种植物，如番茄^[23]、松树^[24]和甘蔗^[25]等，该方法也得到了普遍应用。

虽然 EST-SNP 开发方法具有高效率、低成本的优点，但是，也有如下缺点：(1) EST 测序的质量普遍不高，在进行 EST-SNP 挖掘时会经常得到一些假阳性的 SNP，使挖掘 SNP 的正确率不高。本研究在利用大豆 EST 数据挖掘 SNP 时，添加了大豆基因组测序数据，由于基因组序列是高覆盖率的测序结果，序列的质量较高，将 EST 序列和基因组序列联配挖掘 SNP，大大提高了挖掘 SNP 的正确率，达到了 81.82%。(2) EST 测序中存在大量的直系同源基因序列，会干扰 EST 序列的正确联配，导致 EST-SNP 挖掘会出现较高的假阳性。本研究先进行严格的 Blast 比对，筛选出正确的匹配模式，降低直系同源基因序列的干扰，然后再进行 EST 序列的联配，最后对联配结果进行计数，去除一些频率低的突变位点，大大减少了由于 EST 序列错配带来的假阳性 SNP，提高了挖掘 EST-SNP 的正确率和开发效率。

在大豆中，Chio等^[26]于 2007 年对 6 个大豆品种(Minsoy, Noir 1, Archer, Peking, Evans和Essex)测序获得了 5 551 个 SNP，但到目前为止大豆中尚未有从公共数据中挖掘发现 SNP 的报导。本研究通过联配 EST 序列和基因组测序序列建立了一种高效、可行的基因靶向的 SNP 标记挖掘方法。利用该方法，挖掘了 537 个 EST-SNP，其中大多数是转换突变，这与 Chio 等^[26]和 Zhu 等^[27]对大豆的研究结果一致。由于大量的 C 突变成 T，改变基因组序列甲基化，影响功能基因的表达。

3.2 CAPS 标记的转化

SNP 检测一直是 SNP 标记应用于植物遗传育种研究中的限制因素，如何开发简便、高效、快速的 SNP 标记检测方法是首要解决的问题。在所有的 SNP 位点检测方法中，CAPS 标记方法可以通过简单 PCR、限制性内切酶酶切和琼脂糖电泳即可完成，因此，通过将 SNP 为标记转换为 CAPS 标记可大大促进 SNP 标记在植物遗传育种，特别是大豆遗传育种中的应用。SNP 发生在内酶切识别位点时，SNP 位点就可以通过转化为 CAPS 标记进行检测。在本研究中，通过对 36 个引物的 PCR 产物酶切，发现 26 对引物具有酶切多态性，即 CAPS 标记的转化成功率约为 72.22%，这略高于 Kota 等^[11]在大麦中的转化成功率，成功地将大豆中的 SNP 转化成为 CAPS 标记。

4 结论

建立了从公共数据库的大豆 EST 等数据中开发 SNP 的体系，并成功挖掘出 537 个 EST-SNP 标记。获得 36 个真实的 SNP 位点，正确率达到 81.82%。获得 26 个方便、实用的 CAPS 标记。建立的高效率、低成本的 SNP 标记开发和检测体系将极大地促进大豆和其它植物的遗传育种研究。

References

- [1] Ganai M W, Altmann T, Roder M S. SNP identification in crop plants. *Curr Opin Plant Biol*, 2009, 12: 211–217
- [2] Wiltshire T, Pletcher M T, Batalov S, Barnes S W, Tarantino L M, Cooke M P, Wu H, Smylie K, Santrosyan A, Copeland N G, Jenkins N A, Kalush F, Mural R J, Glynn R J, Kay S A, Adams M D, Fletcher C F. Genome-wide single-nucleotide polymorphism analysis defines haplotype patterns in mouse. *Proc Natl Acad Sci USA*, 2003, 100: 3380–3385
- [3] Feltus F A, Wan J, Schulze S R, Estill J C, Jiang N, Paterson A H. An SNP resource for rice genetics and breeding based on subspecies indica and japonica genome alignments. *Genome Res*, 2004, 14: 1812–1819
- [4] Jander G, Norris S R, Rounsley S D, Bush D F, Levin I M, Last R L. Arabidopsis map-based cloning in the post-genome era. *Plant Physiol*, 2002, 129: 440–450
- [5] Picoult-Newberg L, Ideker T E, Pohl M G, Taylor S L, Donaldson M A, Nickerson D A, Boyce-Jacino M. Mining SNPs from EST databases. *Genome Res*, 1999, 9: 167–174

- [6] Snelling W M, Casas E, Stone R T, Keele J W, Harhay G P, Bennett G L, Smith T P. Linkage mapping bovine EST-based SNP. *BMC Genomics*, 2005, 6: 74
- [7] Batley J, Barker G, O'Sullivan H, Edwards K J, Edwards D. Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. *Plant Physiol*, 2003, 132: 84–91
- [8] Somers D J, Kirkpatrick R, Moniwa M, Walsh A. Mining single-nucleotide polymorphisms from hexaploid wheat ESTs. *Genome*, 2003, 46: 431–437
- [9] Mao X-G(毛新国), Tang J-F(汤继凤), Zhou R-H(周荣华), Jing R-L(景蕊莲), Jia J-Z(贾继增). Wheat cSNP mining based on full-length cDNA sequences. *Acta Agron Sin* (作物学报), 2006, 32(12): 1836–1840 (in Chinese with English abstract)
- [10] Kota R, Rudd S, Facius A, Kolesov G, Thiel T, Zhang H, Stein N, Mayer K, Graner A. Snipping polymorphisms from large EST collections in barley (*Hordeum vulgare* L.). *Mol Genet Genomics*, 2003, 270: 24–33
- [11] Kota R, Varshney R, Prasad M, Zhang H, Stein N, Graner A. EST-derived single nucleotide polymorphism markers for assembling genetic and physical maps of the barley genome. *Funct Integr Genomics*, 2008, 8: 223–233
- [12] Komori T, Nitta N. Utilization of the CAPS/dCAPS method to convert rice SNPs into PCR-based markers. *Breed Sci*, 2005, 55: 93–98
- [13] Cregan P B, Jarvik T, Bush A L, Shoemaker R C, Lark K G, Kahler A L, Kaya N, VanToai T T, Lohnes D G, Chung J, Specht J E. An integrated genetic linkage map of the soybean genome. *Crop Sci*, 1999, 39: 1464–1490
- [14] Song Q J, Marek L F, Shoemaker R C, Lark K G, Concibido V C, Delannay X, Specht J E, Cregan P B. A new integrated genetic linkage map of the soybean. *Theor Appl Genet*, 2004, 109: 122–128
- [15] Tian A G, Wang J, Cui P, Han Y J, Xu H, Cong L J, Huang X G, Wang X L, Jiao Y Z, Wang B J, Wang Y J, Zhang J S, Chen S Y. Characterization of soybean genomic features by analysis of its expressed sequence tags. *Theor Appl Genet*, 2004, 108: 903–913
- [16] Yang Z-Y(杨振宇), Ma X-P(马晓萍), Yamanaka N(山中直树). Polymorphism of phosphoenolpyruvate carboxylase gene in soybean cultivars from northeastern China and Japan. *Acta Agron Sin* (作物学报), 2005, 31(9): 1233–1235 (in Chinese with English abstract)
- [17] Rice P, Longden I, Bleasby A. EMBOSS: The european molecular biology open software suite. *Trends Genet*, 2000, 16: 276–277
- [18] Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol*, 2000, 132: 365–386
- [19] Garg K, Green P, Nickerson D A. Identification of candidate coding region single nucleotide polymorphisms in 165 human genes using assembled expressed sequence tags. *Genome Res*, 1999, 9: 1087–1092
- [20] Guryev V, Berezikov E, Malik R, Plasterk R H, Cuppen E. Single nucleotide polymorphisms associated with rat expressed sequences. *Genome Res*, 2004, 14: 1438–1443
- [21] Kerstens H H, Kollers S, Kommadath A, Del Rosario M, Dibbits B, Kinders S M, Crooijmans R P, Groenen M A. Mining for single nucleotide polymorphisms in pig genome sequence data. *BMC Genomics*, 2009, 10: 4
- [22] Torjek O, Berger D, Meyer R C, Mussig C, Schmid K J, Rosleff Sorensen T, Weisshaar B, Mitchell-Olds T, Altmann T. Establishment of a high-efficiency SNP-based framework marker set for Arabidopsis. *Plant J*, 2003, 36: 122–140
- [23] Yamamoto N, Tsugane T, Watanabe M, Yano K, Maeda F, Kuwata C, Torki M, Ban Y, Nishimura S, Shibata D. Expressed sequence tags from the laboratory-grown miniature tomato (*Lycopersicon esculentum*) cultivar Micro-Tom and mining for single nucleotide polymorphisms and insertions/deletions in tomato cultivars. *Gene*, 2005, 356: 127–134
- [24] Dantec L L, Chagné D, Pot D, Cantin O, Garnier-Géré P, Bedon F, Frigerio J M, Chaumeil P, Léger P, Garcia V, Laigret F, de Daruvar A, Plomion C. Automated SNP detection in expressed sequence tags: Statistical considerations and application to maritime pine sequences. *Plant Mol Biol*, 2004, 54: 461–470
- [25] Cordeiro G M, Elliott F, McIntyre C L, Casu R E, Henry R J. Characterisation of single nucleotide polymorphisms in sugarcane ESTs. *Theor Appl Genet*, 2006, 113: 331–343
- [26] Choi I Y, Hyten D L, Matukumalli L K, Song Q, Chaky J M, Quigley C V, Chase K, Lark K G, Reiter R S, Yoon M S. A soybean transcript map: Gene distribution, haplotype and single-nucleotide polymorphism analysis. *Genetics*, 2007, 176: 685–696
- [27] Zhu Y L, Song Q J, Hyten D L, Van Tassell C P, Matukumalli L K, Grimm D R, Hyatt S M, Fickus E W, Young N D, Cregan P B. Single-nucleotide polymorphisms in soybean. *Genetics*, 2003, 163: 1123–1134