Theor Appl Genet (2010) 121:1–8
DOI 10.1007/s00122-010-1285-x

ORIGINAL PAPER

Genome-wide identification of intron fragment insertion mutations and their potential use as SCAR molecular markers in the soybean

Yongjun Shu · Yong Li · Yanming Zhu · Zhenlei Zhu · Dekang Lv · Xi Bai · Hua Cai · Wei Ji · Dianjing Guo

Received: 30 July 2009 / Accepted: 24 January 2010 / Published online: 17 February 2010

Abstract Introns often have a high probability of mutation as a result of DNA insertions and deletions (indels). In this study, 503 introns with exon-derived insertions were identified using a comprehensive search of the soybean genome. Of the 375 pairs of PCR primer sets designed for the loci in question, 161 primer sets amplified length polymorphism among nine soybean varieties and were identified as soybean gene-intron-driven functional sequence characterized amplified region (SCAR) markers. These SCAR markers are distributed among all 20 of the soybean chromosomes, and they developed from numerous genes involved in various physiological and biochemical processes that influence important agronomic traits of the soybean. The development of these novel gene-driven functional SCAR markers was fast and cost effective, and their use will facilitate molecular-assisted breeding of the soybean.

Communicated by A. Schulman.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-010-1285-x) contains supplementary material, which is available to authorized users.

Y. Shu · Y. Li · Y. Zhu (✉) · Z. Zhu · D. Lv · X. Bai · H. Cai · W. Ji
College of Life Science, Northeast Agricultural University,
Harbin 150030, Heilongjiang, People's Republic of China
e-mail: ymzhu2001@neau.edu.cn

D. Guo
State Key Laboratory of Agrobiotechnology,
Department of Biology, The Chinese University of Hong Kong,
Shatin, N.T., Hong Kong

Introduction

Various types of molecular markers have been developed to construct genetic maps in the soybean, including RFLPs (Keim et al. 1997) and PCR-based markers such as RAPDs (Ferreira et al. 2000), SSRs (Cregan et al. 1999; Hisano et al. 2007; Shultz et al. 2007; Song et al. 2004), AFLP (Matthews et al. 2001; Meksem et al. 2001), CAPS, SCARs, and SNPs (Choi et al. 2007; Hyten et al. 2006; Saini et al., 2008; Zhu et al. 2003). Generally, RFLPs require large amount of pure DNA, and the procedure is time consuming and technique dependent. PCR-based markers are commonly used for QTL detection, mapping, and gene cloning because they are either simple or inexpensive. However, each marker may have its own disadvantage; for example, the amplicons of RAPD markers are poorly reproducible, SNP assays are technically complicated, and the SSR frequency in certain regions of the soybean genome is low. Compared with CAPS markers, genotyping with SCAR markers is relatively inexpensive and simple because additional treatment after PCR is unnecessary. SCAR markers can be used to detect variants between mutation sites by designing specific primer to amplify site-specific sequences using PCR and then using agarose gel electrophoresis for detection. SCARs are dominant or co-dominant markers with advantages such as being site-specific, reproducible, low cost, simple, easy, and fast. For these reasons, the use of SCARS has become common for genetic mapping, map-based cloning, and molecular-assisted breeding (Shimada et al. 2009). Traditional procedures for the development of SCAR markers are based on RAPD or AFLP markers (Noguera et al. 2005), and the SCAR is costly and inefficient to use.

Introns are non-coding region of genes (Berget et al. 1977). Although they can be transcribed into mRNA, they

are excised during pre-mRNA processing. Because intron sequences do not encode functional proteins, they are less conservative and contain significantly more mutations than exon regions. Thus, variants (such as an insertion) in intron sequences can be exploited as genetic markers (Presgraves 2006). The presence of such inserted fragments in exons of other genes indicates that exons can move into introns by some unknown mechanisms (unpublished data). Introns containing insertions of exon-driven fragments are easily identifiable by homologue search and can be developed easily as SCAR markers.

In this study, introns with exon-derived insertions were comprehensively identified in the soybean genome. The primers were designed to amplify the insertion loci, and the SCAR markers were screened in nine soybean varieties. The SCAR markers from the introns with exon-derived insertions were developed and mapped in silico. The function of marker-targeted genes also was analyzed.

Materials and methods

Plant materials

DNA from the 'Acher', 'Evans', 'Peking', 'PI209332', 'Hefeng No. 25', 'Suinong No. 14', 'Guxin wild soybean' (*G. soja*), 'Kefeng No. 1', and 'Nannong 1138-2' soybean varieties was used for the PCR amplification. The Acher, Evans, Peking, and PI209332 varieties were described by Zhu et al. (2003). The Hefeng No. 25 and Suinong No. 14 varieties are two commercial varieties commonly grown in Northeast China, and the Suinong No. 14 variety was derived from the progenies of the cross between Hefeng No. 25 and Suinong No. 8. Guxin is a semi-wild genotype soybean. Kefeng No. 1 and Nannong 1138-2 were described in Zhang et al. (2004), and the seeds were provided by Nanjing Agricultural University. The other varieties were obtained by request from the Chinese Crop Germplasm Information System. DNA from the nine genotypes was extracted from fresh leaf tissue using the cetyl trimethyl ammonium bromide procedure (Murray and Thompson 1980) and was dissolved with 50 µL distilled water. Quality of the DNA was examined using 0.8% agarose gels, and the DNA concentration was determined with a spectrophotometer (SHIMADZU, UV-2550).

Identification of soybean gene introns with exon-derived insertions

Genome sequences and gene information were downloaded from the DOE JGI of USA (Website: http://www.phytozome.net/soybean, version: Glyma0.1c). Perl script was written to retrieve soybean genomic, intron, and exon sequences. Then soybean intron sequences were blasted against exon sequences using BLASTN (Altschul et al. 1997) with an *E*-value of 1E-20. The introns that aligned with more than one homologous exon sequences were kept. Gene sequences containing the sequences of such introns and theirs best aligning exon were aligned to the genes per se by BL2SEQ, and the size of the overlapping sequences was identified. Only introns with an exon-derived insertion size of 50–500 bp were used for primer design.

Primers designed for SCAR markers

The insertion and flanking sequences were blasted against the whole genome sequences. Next, the insertion and flanking sequences were excluded if they aligned with the repeated and homologue regions of the genome sequences. Primers were designed to these unique sequences, remain sequences after excluded using Primer3 software (Rozen and Skaletsky 2000) and the primer names were prefixed with SCAR. The parameters used for primer design were: (1) primer length 20–26 bp with an optimum of 22 bp; (2) annealing temperature of 50–60°C, with an optimum of 58°C; (3) percentage GC in the range of 40–55%; and (4) product size in the range of 400–800 bp. All of the oligonucleotides (Supplementary Table 1) were synthesized by GeneScript Inc. (Nanjing, China).

Amplification and detection of insertion mutations

All primer sets were used to amplify the genomic DNA of the Suinong No. 14 soybean variety. Amplification reactions were performed in a 25 µl reaction mixture containing 30 ng template DNA, 0.2 µmol L^{-1} of each primer, 800 µmol L^{-1} dNTP, 1.5 mM $MgCl_2$, 1 U *rTaq* polymerase, and 2.5 µl of 10× PCR buffer. A touchdown PCR (TD-PCR) program was used and the cycling conditions were as follows: 10 min at 94°C; 30 cycles of 30 s at 94°C, 30 s at 60 − 0.5°C per cycle, 40 s at 72°C; and 10 min at 72°C for final extension. PCR products were resolved on a 2% agarose gel stained with ethidium bromide. Primer pairs that successfully amplified discrete product on a 2% agarose gel were selected and to amplify DNA from the eight other varieties in order to identify polymorphism among the nine soybean varieties. All soybean varieties were scored in binary digits based on the fragment size variation of the PCR, and the UPGMA-based dendrogram was constructed using NTSYS-pc (v2.02a).

Functional analysis of SCAR targeted genes and mapping of the SCAR markers in silico

The soybean whole genome sequences (version: Glyma1), the *Arabidopsis* genome sequences, and protein annotation

information were downloaded from Phytozome of the DOE JGI, TAIR (Website: http://www.arabidopsis.org/, version: TAIR8) (Swarbreck et al., 2008) and from MIPS (Website: http://www.helmholtz-muenchen.de/en/mips/, version 2.0) (Mewes et al. 2006). Flanking sequences of the SCAR markers were BLASTN-searched against the soybean genome sequences to locate the physical position of SCAR target genes on the chromosomes. Protein sequences of the SCAR targets then were BLASTP-searched against *Arabidopsis* proteins for SCAR marker annotation. The SCAR markers were classified into different groups according to the function map of the MIPS. The positions of soybean gene-driven SCAR markers were mapped to the soybean chromosomes of cultivar 'Williams 82' in silico, and the map was drawn with MapChart 2.0 (Voorrips 2002). For the SCAR markers positioned on the physical map of cultivar 'Forrest', the sequences of the SCAR marker-targeted genes were blasted against the BAC-end sequences (BES) in SoyGD (Website: http://www.soybeangenome.siu.edu/) (Shultz et al. 2006). Minor errors likely were made due to the method of chromosome measurement in an interspecific cross and to the arbitrary nature of assignment.

Results

Development of the soybean SCAR markers

A total of 454,545 soybean sequences containing 202,624 introns and 251,921 exons were retrieved from the soybean genome sequences. Among these, 503 introns with an exon-derived insertion of 50–500 bp as were identified via BLASTN and BL2SEQ and were selected for primer design. More than 400 pairs of primers were designed, and 375 remained after elimination of the introns that contained alternative splicing sequences. Of the 331 pairs of primers (88.2%) that successfully amplified stable products from soybean variety Suinong 14, a total of 161 pair primers amplified polymorphic products among all nine soybean varieties (Fig. 1). Of these, 128 (79.5%) primer pairs ampli-

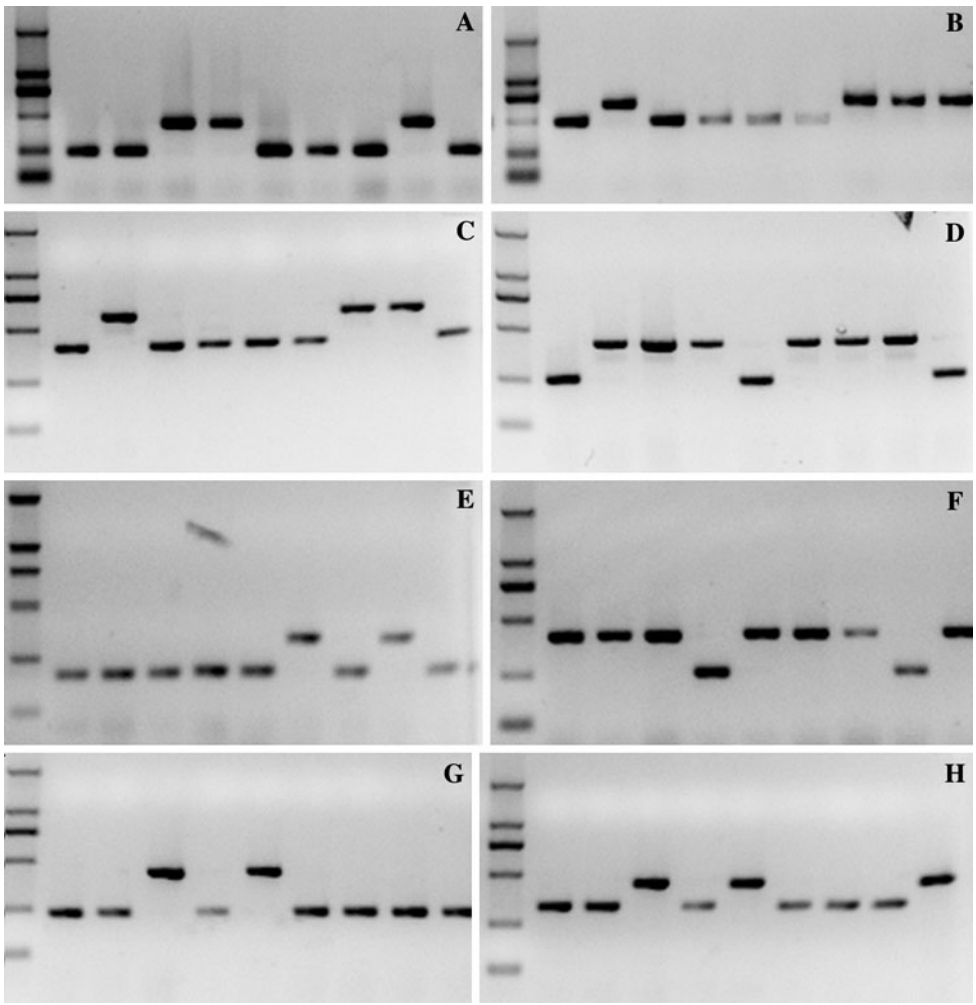

Fig. 1 Agarose gel electrophoresis of sequence characterized amplified region (*SCAR*) markers. **a** SCAR44, **b** SCAR56, **c** SCAR195, **d** SCAR183, **e** SCAR72, **f** SCAR88, **g** SCAR188, and **h** SCAR181. Lane order is (*left* to *right*) DNA marker DL2000, Suinong 14, Kefeng 1, Nannong 1138-2, Guxin wild soybean, Acher, Evans, Peking, PI209332, and Hefeng 25

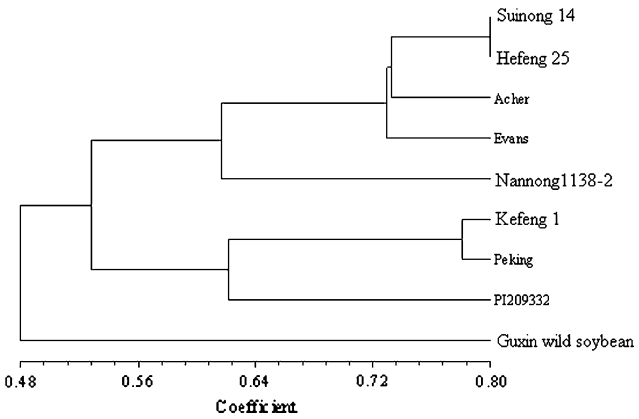

Fig. 2 UPGMA dendrogram of the nine soybean varieties based on SCAR markers

fied a single unique fragment; this is in contrast to the RFLP and RAPD markers, which frequently amplified multiple bands.

Phylogenetic analysis based on soybean SCAR markers

The allelic data obtained from the 161 soybean SCAR markers were computed to construct a dendrogram of the nine soybean varieties using UPGMA in the NTSYS-pc. As shown in Fig. 2, the nine species were divided into three groups. The first group consists of five soybean varieties, including Hefeng No. 25 and Suinong No. 14; these two had the smallest distance among the nine varieties in the phylogenetic tree. This was consistent with the pedigree showing that Suinong No. 14 was the offspring of the Hefeng No. 25. Three land races, which were the ancestors of cultivated soybean varieties (Zhu et al. 2003) constituted the second group. The genetic distance of the Guxin variety to the other *G. max* was the greatest observed in this study. In conclusion, the UPGMA dendrogram of the nine soybean varieties based on SCAR allelic data was consistent with the genetic relationship of these soybean types, and thus it can be used for the study of soybean genetic evolution.

The Kefeng No. 1, Peking, PI209332, and Guxin wild soybean varieties all have a black seed coat and are resistant to abiotic or biotic stress, whereas the other five varieties do not have these phenotypes. Because SCAR markers can distinguish between these different phenotypes, we believe that SCAR markers may have potential application in soybean phenotypic identification and classification.

Functional analysis of the soybean SCAR targeted genes

The SCAR targeted genes were identified by BLASTN against the soybean genome sequences, and functional annotations were assigned by BLASTP against the *Arabidopsis* genome proteins (Fig. 3). Although SCAR targeted genes mostly are involved in the process of subcellular localization (80/360, 22.2%), protein binding and catalytic activities (59/360, 16.4%), and metabolism (41/360, 11.4%), some of the genes also are involved in other cellular processes, such as protein synthesis, protein folding modification, response to outside stimuli, signal transduction, cell repair, and defense against viruses. For example, the marker SCAR10 targets on the third intron (from 2,372 to 2,559 bp) of the gene Glyma05g28350, which encodes a protein with 99% identity to the Rhg4 protein (Lightfoot et al. 2001) (Fig. 3). Rhg4 contributes to improved resistance to sudden death syndrome and to the cyst nematode of soybeans. This finding indicates that this SCAR marker (SCAR10) could be used to distinguish cyst nematode-resistant varieties of Peking and PI209332 from the nine soybean varieties. This maker may have potential application in molecular breeding of soybean variety resistance to sudden death syndrome and the soybean cyst nematode. Marker SCAR73 targeted gene Glyma18g52560 encodes phospholipase D, which plays a

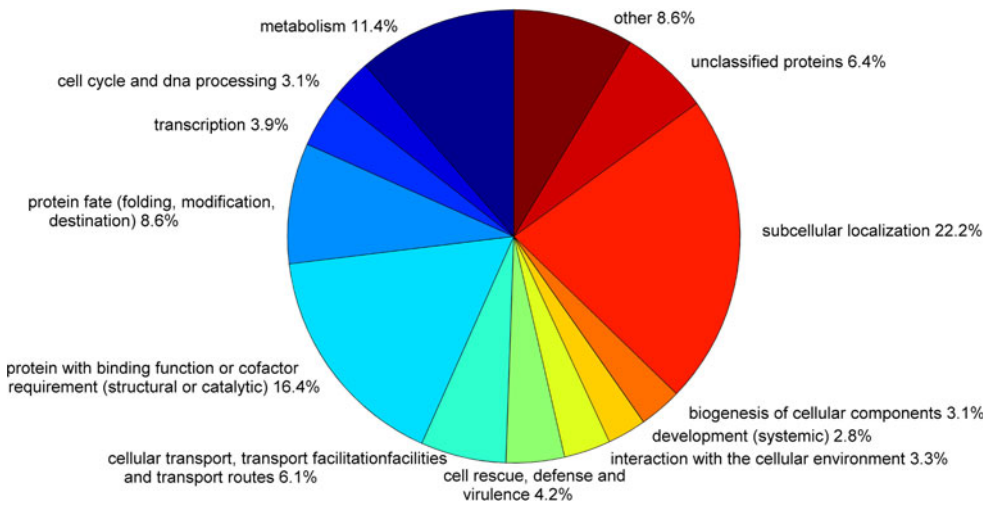

Fig. 3 Molecular functional analysis of genes containing SCAR markers

key role in the regulation and signal transduction pathway of plant response to abiotic and biotic stress (Wang 2002). Moreover, association analysis of the frogeye leaf spot resistance locus with the SCAR73 marker in the soybean showed that the rate of the classification consistency based on the variants of SCAR73 and the resistance or susceptibility of the genotypes was as high as 80% (unpublished data). This results suggests that this marker is related to pathogen resistance and could be used in marker-assisted selection for breeding lines resistant to pathogens.

In silico mapping of the soybean SCAR markers

The positions of the identified SCAR targeted genes on the chromosomes were drawn using MapChart (Fig. 4). All 161 SCAR markers are distributed among the 20 soybean chromosomes. The number of markers on chromosomes ranged from 2 (chromosome 20) to 15 (chromosome 9). The SCAR markers were mainly distributed on the two arms of each chromosome, but a few occurred in the centromere regions, which was consistent with the distribution of genes in the genome. Fourteen SCAR marker targeted genes were identified to have homologous sequences in the cultivar Forrest when blasted against the BES sequences in SoyGD (Supplementary Table 2) (Fig. 5).

Compared with traditional SSR markers, whose mutations are neutral and are distributed lowly in the gene-rich region of genome, soybean gene-driven SCAR markers have significant advantages. They can be used for linkage analysis and QTL mapping. Moreover, because they are designed based on the functional genes, which may have experienced selection pressure during evolution, SCAR markers may be feasible for use in candidate-gene association mapping (Gupta et al. 2005; Wilson et al. 2004), whose polymorphisms may result in the phenotypic variation of the traits.

Discussion

In the past few decades tremendous progress has been made in soybean genetic mapping, from the first RFLP-based map of the soybean genome (Keim et al. 1990) to the integrated maps generated with RAPD (Ferreira et al. 2000), AFLP (Keim et al. 1997), SSR (Cregan et al. 1999; Hisano et al. 2007; Shultz et al. 2007; Song et al. 2004), and SNP (Choi et al. 2007; Saini et al. 2008) markers. However, due to non-specificity problems, the application of RFLP and RAPD markers was limited. The PCR-based SSR marker was broadly used for its simplicity, but the development efficiency of SSR markers in gene-rich regions was low (Choi et al. 2007). SNPs were introduced for the development of gene-targeted markers, but due to the duplication of the soybean genome, SNP discovery and SNP detection were difficult.

Because many more mutations accumulate in introns during evolution, introns are considered less conservative than exons. Intron mutations include nucleotide type mutation and length polymorphism (indel mutations). Length polymorphism is the most easily recognizable, especially when the indel size is large. Exon-primed intron-crossing PCR (EPIC-PCR) (Bierne et al. 2000) has been used to develop a large number of EPIC-PCR markers in many species (Berrebi et al. 2005; Tay et al. 2008). In EPIC-PCR, primers are designed based on the flanking sequence in exons to amplify introns, and intron length mutations are detected by electrophoresis. Although many intron length mutations can be detected by this method, most of the mutations contain only a few nucleotides. A difference of less than 3 bp is difficult to visualize on an agarose gel, and a PAGE has to be used to separate and visualize DNA bands (Wang et al. 2005).

Intron insertion mutations, or exogenous sequences of introns, can result from insertion of an exon sequence. Exons are the coding regions, and their structural characteristics differ from those of the non-coding regions; therefore, they can be easily distinguished from introns. In this study, our strategy was to scan the soybean genome to identify exogenous fragments in introns derived from exons, design primers to amplify the insertion mutation sites, and then detect polymorphism among the nine soybean varieties. Using this approach, 161 soybean gene-driven functional SCAR markers were developed. All of the SCAR markers

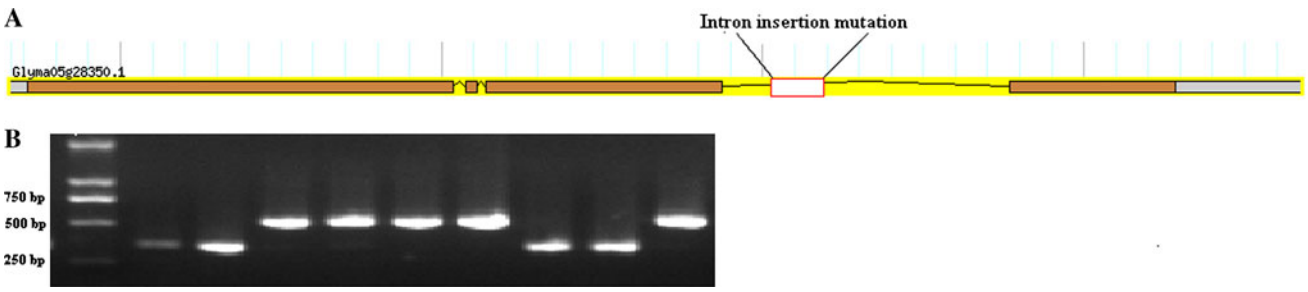

Fig. 4 Information about molecular marker SCAR10. **a** The gene structure of gene Glyma05g28350 and **b** agarose gel electrophoresis of SCAR10

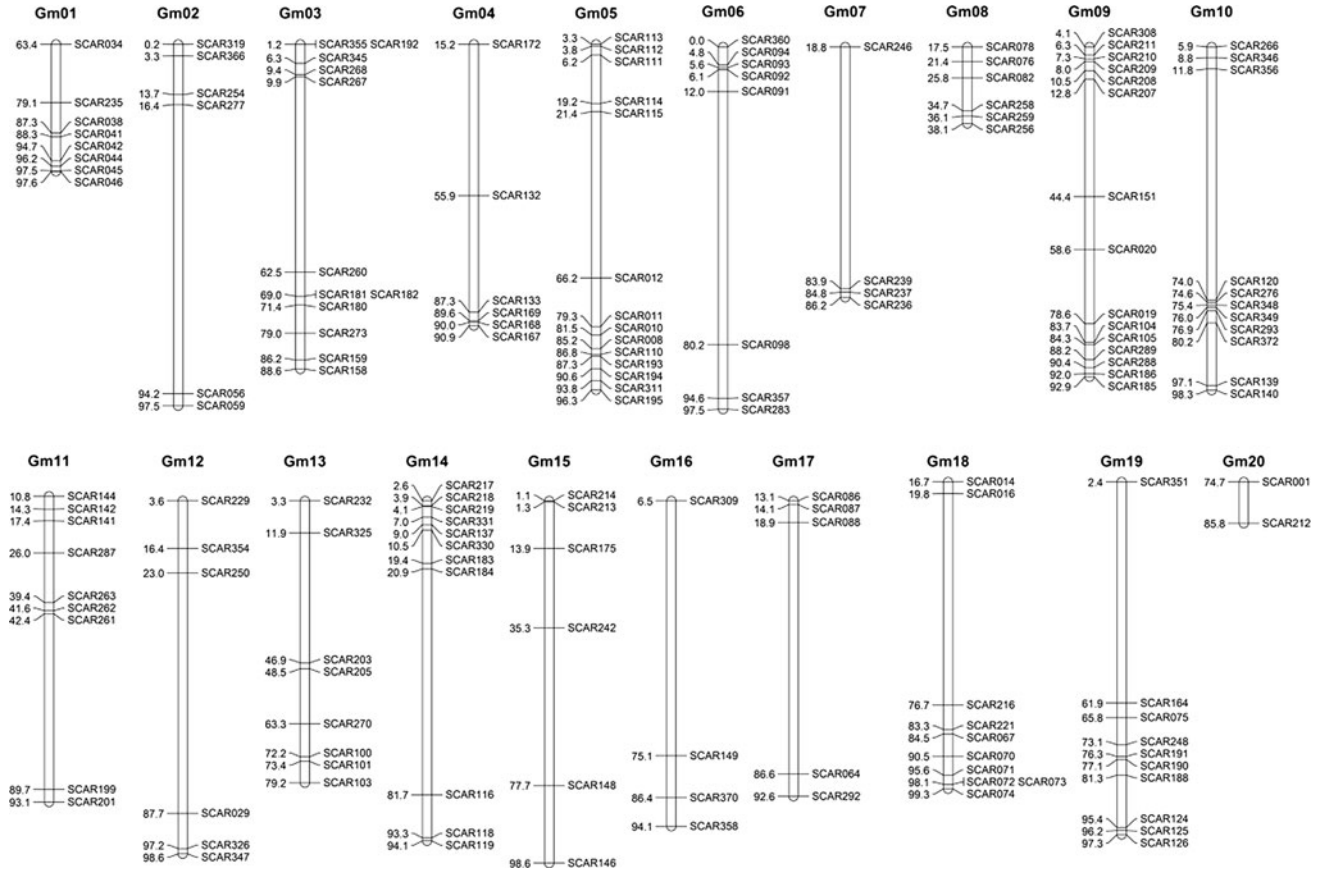

Fig. 5 Distribution of the 161 SCAR marker on soybean chromosomes

were DNA fragment length polymorphic and co-dominant. Moreover, because the mutations of the SCAR markers resulted from a long fragment insertion, they could be easily detected by PCR and agarose gel. Unlike traditional markers, mutations detected in this experiment were within genes. In addition to the genetic linkage analysis and QTL mapping, soybean gene-driven SCAR markers may be used for candidate-gene association mapping.

Re-sequencing of three mutation sites revealed that the mutations have obvious characteristics of retrotransposons: (1) there are two terminal inverted repeat (TIR) sequences with a length of 18–20 bp at both ends of the insertion sequences. The same sequence also appears in soybean species but with the absence of the insertion sequence. (2) Length of insertion sequences ranges from 161 to 191 bp, which is consistent with the characteristics of the retrotransposon SINE. Thus, we conclude that the soybean SCAR markers are the result of retrotransposon movement (unpublished data). Transposons and retrotransposons, especially the short mobile sequence types such as SINE (Mochizuki et al. 1992; Xu et al. 2005) and MITE (Casa et al. 2000, 2002; Monden et al. 2009) transposons, are widely used as molecular markers for plant and animal breeding. Moreover, transposon insertion mutagenesis is evidence of species evolution. Molecular markers targeting transposon insertion sites therefore can reveal effectively the genetic relationship of species. Compared with other molecular markers, the soybean SCAR markers are more suitable for study of the origin, evolution, and genetic diversity of the soybean. In this research, the SCAR markers could distinguish the genetic relationship of the Hefeng No. 25 and Suinong No. 14 varieties. The cultivated soybean and wild soybean also were distinguished by SCAR markers, but not with other molecular markers (e.g., CAPS markers; Shu et al. 2009). These results demonstrate the potential of SCAR markers in soybean germplasm classification and evolutionary research.

In summary, intron insertion mutations were identified and 161 soybean gene-driven functional SCAR markers were developed by genome sequence analysis. The SCAR markers are widely distributed in the soybean genome and target numerous genes involved in various physiological and biochemical processes that influence important agronomic traits of the soybean. The novel gene-driven functional SCAR marker system developed in this study is fast and cost effective and offers great potential for molecular-assisted breeding of the soybean.

Acknowledgments The soybean genome sequence data were produced by the US Department of Energy Joint Genome Institute (http://www.jgi.doe.gov/) in collaboration with the user community, and we would thank to them. The project was supported by the National "863" Program (2006AA100104, 2008AA10Z153), the National Transgenic Research Project (2008ZX08004), the Key Research Plan of Heilongjiang Province (GA06B103), and the Innovation Research Group of NEAU (CXT004).

References

Altschul S, Madden T, Schaffer A, Zhang J, Zhang Z, Miller W, Lipman D (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Berget SM, Moore C, Sharp PA (1977) Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci USA 74:3171–3175

Berrebi P, Boissin E, Fang F, Cattaneo-Berrebi G (2005) Intron polymorphism (EPIC-PCR) reveals phylogeographic structure of *Zacco platypus* in China: a possible target for aquaculture development. Heredity 94:589–598

Bierne N, Lehnert SA, Bedier E, Bonhomme F, Moore SS (2000) Screening for intron-length polymorphisms in penaeid shrimps using exon-primed intron-crossing (EPIC)-PCR. Mol Ecol 9:233–235

Casa AM, Brouwer C, Nagel A, Wang L, Zhang Q, Kresovich S, Wessler SR (2000) Inaugural article: the MITE family heartbreaker (Hbr): molecular markers in maize. Proc Natl Acad Sci USA 97:10083–10089

Casa AM, Mitchell SE, Smith OS, Register JC 3rd, Wessler SR, Kresovich S (2002) Evaluation of Hbr (MITE) markers for assessment of genetic relationships among maize (*Zea mays* L.) inbred lines. Theor Appl Genet 104:104–110

Choi I-Y, Hyten DL, Matukumalli LK, Song Q, Chaky JM, Quigley CV, Chase K, Lark KG, Reiter RS, Yoon M-S, Hwang E-Y, Yi S-I, Young ND, Shoemaker RC, van Tassell CP, Specht JE, Cregan PB (2007) A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis. Genetics 176:685–696

Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Kaya N, VanToai TT, Lohnes DG, Chung J, Specht JE (1999) An integrated genetic linkage map of the soybean genome. Crop Sci 39:1464–1490

Ferreira A, Foutz K, Keim P (2000) Soybean genetic map of RAPD markers assigned to an existing scaffold RFLP map. J Hered 91:392–396

Gupta PK, Rustgi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol Biol 57:461–485

Hisano H, Sato S, Isobe S, Sasamoto S, Wada T, Matsuno A, Fujishiro T, Yamada M, Nakayama S, Nakamura Y, Watanabe S, Harada K, Tabata S (2007) Characterization of the soybean genome using EST-derived microsatellite markers. DNA Res 14:271–281

Hyten DL, Song Q, Zhu Y, Choi I-Y, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB (2006) Impacts of genetic bottlenecks on soybean genome diversity. Proc Natl Acad Sci USA 103:16666–16671

Keim P, Diers BW, Olson TC, Shoemaker RC (1990) RFLP mapping in soybean: association between marker loci and variation in quantitative traits. Genetics 126:735–742

Keim P, Schupp JM, Travis SE, Clayton K, Zhu T, Shi L, Ferreira A, Webb DM (1997) A high-density soybean genetic map based on AFLP markers. Crop Sci 37:537–543

Lightfoot DA, Gibson PT, Merkem K (2001) Soybean sudden death syndrome resistant soybeans, soybean cyst nematode resistant soybeans and methods of breeding and identifying resistant plants. US Patent: 6300541

Matthews BF, Devine TE, Weisemann JM, Beard HS, Lewers KS, MacDonald MH, Park Y-B, Maiti R, Lin J-J, Kuo J, Pedroni MJ, Cregan PB, Saunders JA (2001) Incorporation of sequenced cDNA and genomic markers into the soybean genetic map. Crop Sci 41:516–521

Meksem K, Ruben E, Hyten D, Triwitayakorn H, Lightfoot DA (2001) Conversion of AFLP bands into high-throughput DNA markers. Mol Genet Genomics 265:207–214

Mewes HW, Frishman D, Mayer KF, Munsterkotter M, Noubibou O, Pagel P, Rattei T, Oesterheld M, Ruepp A, Stumpflen V (2006) MIPS: analysis and annotation of proteins from whole genomes in 2005. Nucleic Acids Res 34:D169–D172

Mochizuki K, Umeda M, Ohtsubo H, Ohtsubo E (1992) Characterization of a plant SINE, p-SINE1, in rice genomes. Jpn J Genet 67:155–166

Monden Y, Naito K, Okumoto Y, Saito H, Oki N, Tsukiyama T, Ideta O, Nakazaki T, Wessler SR, Tanisaka T (2009) High potential of a transposon mPing as a marker system in japonica × japonica cross in rice. DNA Res 16:131–140

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4325

Noguera FJ, Capel J, Alvarez JI, Lozano R (2005) Development and mapping of a codominant SCAR marker linked to the andromonoecious gene of melon. Theor Appl Genet 110:714–720

Presgraves DC (2006) Intron length evolution in *Drosophila*. Mol Biol Evol 23:2203–2213

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365–386

Saini N, Shultz J, Lightfoot DA (2008) Re-annotation of the physical map of *Glycine max* for polyploid-like regions by BAC end sequence driven whole genome shotgun read assembly. BMC Genomics 9:323

Shimada N, Nakatsuka T, Nakano Y, Kakizaki Y, Abe Y, Hikage T, Nishihara M (2009) Identification of gentian cultivars using SCAR markers based on intron-length polymorphisms of flavonoid biosynthetic genes. Sci Hortic 119:292–296

Shu Y, Li Y, Bai X, Cai H, Ji W, Zhu Y (2009) Development of soybean gene-driven functional CAPS markers from the genes' re-sequencing information. Acta Agron Sin 35:2015–2021

Shultz JL, Kurunam D, Shopinski K, Iqbal MJ, Kazi S, Zobrist K, Bashir R, Yaegashi S, Lavu N, Afzal AJ, Yesudas CR, Kassem MA, Wu C, Zhang HB, Town CD, Meksem K, Lightfoot DA (2006) The soybean genome database (SoyGD): a browser for display of duplicated, polyploid, regions and sequence tagged sites on the integrated physical and genetic maps of Glycine max. Nucleic Acids Res 34:D758–D765

Shultz J, Kazi S, Bashir R, Afzal J, Lightfoot D (2007) The development of BAC-end sequence-based microsatellite markers and placement in the physical and genetic maps of soybean. Theor Appl Genet 114:1081–1090

Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB (2004) A new integrated genetic linkage map of the soybean. Theor Appl Genet 109:122–128

Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, Radenbaugh A, Singh S, Swing V, Tissier C, Zhang P, Huala E (2008) The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucleic Acids Res 36:D1009–D1014

Tay WT, Behere GT, Heckel DG, Lee SF, Batterham P (2008) Exon-primed intron-crossing (EPIC) PCR markers of *Helicoverpa armigera* (Lepidoptera: Noctuidae). Bull Entomol Res 98:509–518

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93:77–78

Wang X (2002) Phospholipase D in hormonal and stress signaling. Curr Opio Plant Biol 5:408–414

Wang X, Zhao X, Zhu J, Wu W (2005) Genome-wide investigation of intron length polymorphisms and their potential as molecular markers in rice (*Oryza sativa* L.). DNA Res 12:417–427

Wilson LM, Whitt SR, Ibanez AM, Rocheford TR, Goodman MM, Buckler EST (2004) Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell 16:2719–2733

Xu JH, Kurata N, Akimoto M, Ohtsubo H, Ohtsubo E (2005) Identification and characterization of Australian wild rice strains of *Oryza meridionalis* and *Oryza rufipogon* by SINE insertion polymorphism. Genes Genet Syst 80:129–134

Zhang WK, Wang YJ, Luo GZ, Zhang JS, He CY, Wu XL, Gai JY, Chen SY (2004) QTL mapping of ten agronomic traits on the soybean (*Glycine max* L. Merr.) genetic map and their association with EST markers. Theor Appl Genet 108:1131–1139

Zhu YL, Song QJ, Hyten DL, Van Tassell CP, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB (2003) Single-nucleotide polymorphisms in soybean. Genetics 163:1123–1134