Genotype variant of soybean cultivars around Lake Victoria region, East Africa

Article · January 2010

CITATIONS
0

READS
35

3 authors:

Zedekiah Agina Okwany
National Museums of Kenya
3 PUBLICATIONS **1** CITATION
SEE PROFILE

Fredrick M Njoka
Embu University College
3 PUBLICATIONS **0** CITATIONS
SEE PROFILE

Elijah Kem Githui
Institute of Primate Research
11 PUBLICATIONS **50** CITATIONS
SEE PROFILE

Aspects of Applied Biology ***96****, 2010*
Agriculture: Africa's "engine for growth" - Plant science and biotechnology hold the key

Genotype variant of soybean cultivars around Lake Victoria region, East Africa

ZEDEKIAH A OKWANY[1,2], FREDRICK M NJOKA[1] and ELIJAH K GITHUI[2]

[1]*Department of Plant and Microbial Sciences, Kenyatta University, P.O. Box 43844, Nairobi, Kenya*
[2]*Molecular Genetics Laboratory, National Museums of Kenya, P. O. Box 40658, Nairobi, Kenya*

Summary

Glycine gracilis, which originated from North-East China, is morphologically intermediate between the cultivated *G. max* and the wild form, *G. soja*. These three forms are cross-compatible and their crosses provide a resource for biologically and agronomically important genes. Breeding genetics of soybean has developed slowly due to inherent difficulties in crossing. The soybean stamens and the pistils are present in the same flower resulting in a high percentage of self-fertilization. The cultivars grown around the Lake Victoria differ in agronomic performance but their genotypes have not been described at the molecular level. Amplification of 5S ribosomal gene from the cultivars yielded uniform PCR product of about 700 bps. Limited sample sequencing showed a restriction marker site within the 5S ribosomal gene. A phylogenetic tree based on 5S rDNA gene sequences of East African cultivars and different *Glycine* species suggests that the cultivar Mikumi is similar to, and SB 19 of Kenya and Maksoy of Uganda are closely related to, *G. max*.

Key words: *Glycine max*, East Africa, 5S rDNA, restricion enzyme markers

Introduction

Soybean (*Glycine max* (L.) Merrill.) is one of the popular pulses around the world and has been cultivated in China and Manchuria since 2500 BC (Morse, 1950). It is not only used as high protein food, but also as an oil source and forage. *Glycine gracilis*, which originated from north-east China is morphologically intermediate between the cultivated *G. max* and the (Hymowitz, 2004; Ngeze, 1993) wild form, *G. soja*. These three forms are cross-compatible and crosses between these cultivars provide good genetic resources for biologically and agronomically important genes.
Breeding genetics of soybean has developed slowly due to inherent difficulties in performing crosses, lack of genetic variation in the germplasm and lack of cytogenetic markers (Keim *et al*., 1990; Scott & Samuel, 1970). The method of producing hybrid soybean seed on the commercial scale is difficult to achieve since in soybean both the staminate and the pistillate are present in the same flower and pollination occurs before the flower opens (blooms). As a result soybeans exhibit a high percentage of self-fertilization and cross pollination is usually less than one percent (Caviness, 1966). DNA-based markers are important for selection and improvement of varieties and hybrids in plant breeding programs. (Gupta *et al*., 2001; Kota *et al*., 2003) and is particularly a more convenient method in producing hybrid soybean (Keim *et al*., 1990; Choi *et al*., 2007). If a larger number of simple sequence repeat (SSR) and restriction fragment length polymorphism (RFLP) and Single nucleotide polymorphic (SNP) makers can be identified in a single segregating

population, this may identify a genetic linkage map among makers and quantitative trait loci (QTLs).

The coding portion of the nuclear ribosomal rDNA cistron (18S, 5.8S, and 26S rDNA) has been extensively characterized at the sequence level for several flowering plants, mainly due to interest in using these data for phylogenetic analyses. Sequences include (18S) rDNA (Nickrent & Soltis, 1995; Soltis *et al*., 1997), (26S) rDNA (Bult *et al*., 1995; Kuzoff *et al*., 1998), internal transcribed spacer (ITS) and 5.8S rDNA sequences (Baldwin *et al*., 1995). In contrast to the conserved 5.8S, 18S and 26S rDNA, which have phylogenetic utility at deeper divergence levels, ITS sequences often contain sufficient variation to allow examination of genetic relationships between cultivated varieties, populations, and individuals (Rogers & Bendich, 1987; Hemleben *et al*., 1988; Jorgansen & Cluster 1988; Schaal & Learn, 1988).

The cultivars grown around the Lake Victoria region of East Africa show differences in agronomic performance, however, there has been no study describing their genotypes at molecular level. In this study, we sampled seedlings from different cultivars from the region and characterized the genotypes using 5S rDNA.

Materials and Methods

Sampling

Seeds were collected from three regions; Kenya Uganda and Tanzania. A nursery of seedlings from these different regions was established at National Museums of Kenya's green house. Shoots were plucked from the seedlings and used as materials for DNA extraction.

Genomic DNA extraction

DNA was extracted according to Phenol Chloroform method of Sambrook *et al*. (1989). Approximately 1.0 g young shoot of soybean was obtained from the set experimental blocks, taken to the laboratory, washed with distilled water and placed in a chilled sterile mortar and quickly ground into pulp using a pestle then immediately subjected to DNA extraction procedures. STE extraction buffer (200 µL) was added to approx 200 µL the ground leaf material. 500 µL of Phenol:Chloroform was used to separate DNA into the aqeous phase, then absolute ethanol added to precipitate it. The pellet was then re-suspended in 100µL TE (10 mM Tris-HCl, 1 mM ETDA). Quality of extracted DNA was determined by electroprhoresis on 1% agarose gel in TAE buffer (40mM Tris-Acetate, 1 mM EDTA) and visualized under UV illumination. The isolated DNA stored at -20°C.

PCR- polymerase chain reaction

PCR was done using a custom ordered pair of primers (*Syrib* F 5'-GCGGAAGGATCATTGTCGATG-3' and *Syrib* R 5'-TGACCTGAGGTCTCGTTG-3') that correspond to the flanking ends of 5S ribosomal gene including the intergenic sequences of the soy gene. The PCR amplification tube contained µL of the template DNA, 0.4 µM of each primer, 0.4 mM dNTPs, 2.0 mM $MgCl_2$, 1× buffer and 0.5U of *Taq* polymerase, made up to 25 µL with ddH_2O with cycling conditions: *Taq* activation 94˚C, 5 minutes, then 35 cycles of 94˚C, 1 min, annealing 53˚C, 1 min, extension 68˚C, 1 min, and final extension step 70°C, 3 mins. Perkin-Elmer GeneAmp PCR System 2400 (MJ research, INC USA) was used in all the reactions. PCR products were analyzed on agarose gel.

Gene clean

Target DNA bands were excised from agarose gel using a scalpel and placed in labeled microcentrifuge tubes for gene clean procedure (Vogelstein & Gillespie, 1979). Three volumes NaI (sodium iodide) solution was added to the gel slice and incubated at 56°C in water bath for

15–30 minutes, mixing after 5 minutes until all agarose had melted. 20 µL of Glassmilk (Silica) solution was added and mixed thoroughly by vortexing. The tubes containing the mixtures were incubated for 15 minutes on ice, with intermittent mixing. The tubes were then microcentrifuged for 5 s to pellet the glassmilk and DNA eluted in 25 µL TE analysed then forwarded to ILRI (International Laboratory for Research Institute, Kenya) for custom direct sequencing procedures. Sequencing was done using the BigDye® Terminator sequencing kit (Applied Biosystems Inc. Foster City, CA USA).

BLAST identification of fragment sequences and phylogenetic analysis

The identities of the generated nucleotide sequences were assigned in the NCBI's non-redundant nucleotide BLAST search (Altschul *et al*., 1997) and alignment done using BioEdit suite Clustal-W program.

Phylogenetic relationships were inferred from the aligned nucleotide sequences by the neighbour-joining method implemented in the Phylip package (Felsentein, 1997) as implemented in MEGA software package (Tamura *et al*., 2007). Consensus trees were bootstrapped for 100 replicates and rooted to the Glycine soja to determine divergence from the other taxa.

Restriction fragment length polymorphisms (RFLPs) analysis

Choice of restriction enzymes was defined by predicted simulation restriction sites across the known *Glycine* spp. 5S rDNA sequences using BioEdit suite programs. Restriction endonuclease digestion (Hae III, Hpa II, Mse I, Taq I, Sac II and Sma I) of the amplified PCR product was conducted according to manufacturer's instructions. DNA fragments were separated by electrophoresis in a 1% agarose gel in TAE running buffer or 8% Poly-acrylamide Gel Electrophoresis in TBE (45 mM Tris-borate, 1 mM EDTA) for approximately 30 mins at 90 mÅ.

Results

PCR amplification of the 5S ribosomal gene in the cultivars yielded invariably PCR product of about 650 bps equivalent to expected size based on the targeted DNA region. (Fig. 1). This suggests that no gross differences in the in 5S gene among the cultivars.

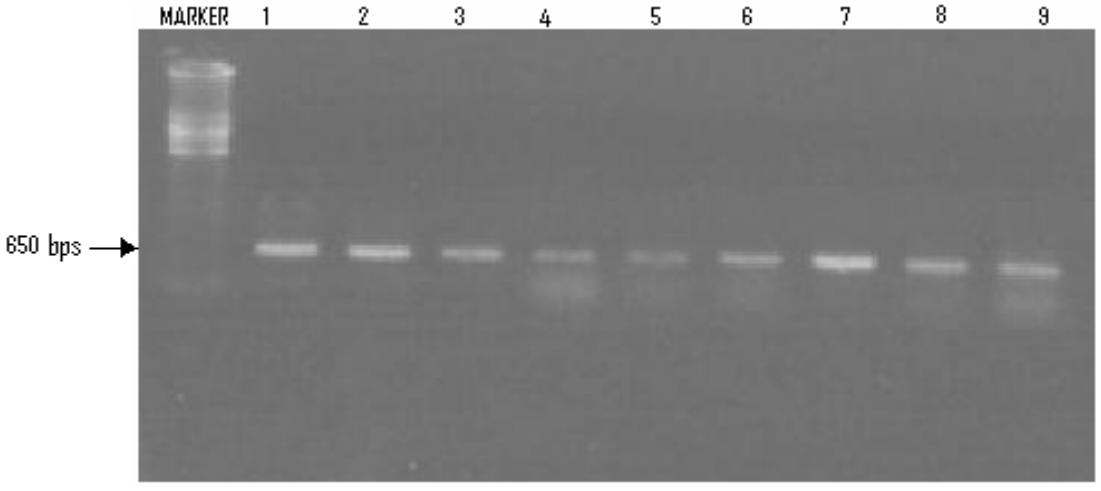

Fig. 1. 1% agarose gel of PCR product of the East African soybean *Glycine max* 5S rDNA gene. 1,Nyala, 2,Mikumi, 3, Duicker, 4,Mikese, 5,SB20, 6,Nam 4M, 7,SB9, 8, SB8, 9Maksoy IN.

Restriction fragment length polymorphism

Digestion with restriction enzymes *Hae*III, *Hpa*II and *Mse*I gave similar patterns across the cultivars (Fig. 2a).

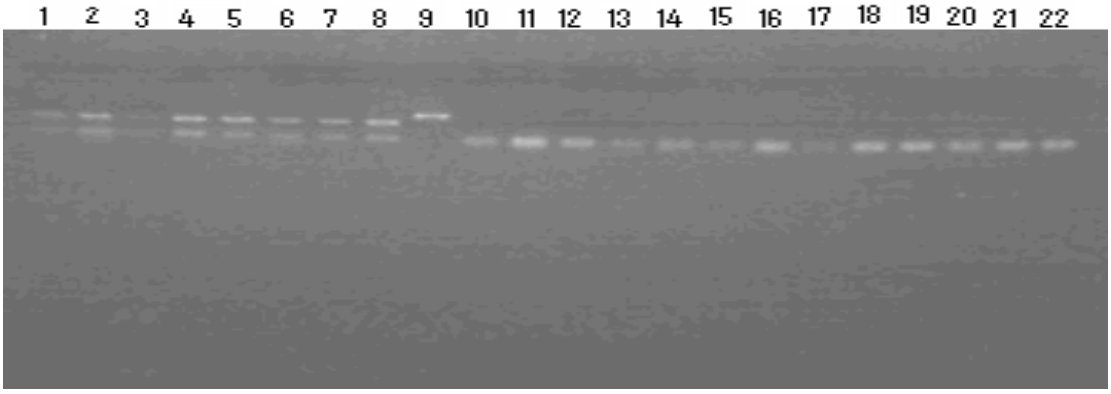

Fig. 2a. Restriction digest of 5S rDNA gene with enzyme M*se*I. (Lanes 1-8) and H*ae*III (Lanes 10-22) Lane 1,2,3,4,5,6,7 and 8 are Nyala, Mikumi, Maksoy 1N, SB20, Mikese, TGX-1876-2E and Nam II respectively. Lane 9 is undigeted PCR product. Lines 10 to 22 are Nyala, SB8, SB9, SB19, SB20, Mikumi, Mikese, TGX-1876-2E, Duicker, Nam I, Nam II Nam 4M and Maksoy 1N.

The restriction enzyme did not discriminate between the local cultivars indicating that they closely related. However, in the computer simulated restriction digests amongst distantly related cultivars, these enzymes could indicate markers sites.

Restriction digest with restriction enzymes *Hpa*II was analysed on agarose gel. All these cultivars produced similar bands (Fig. 2b).

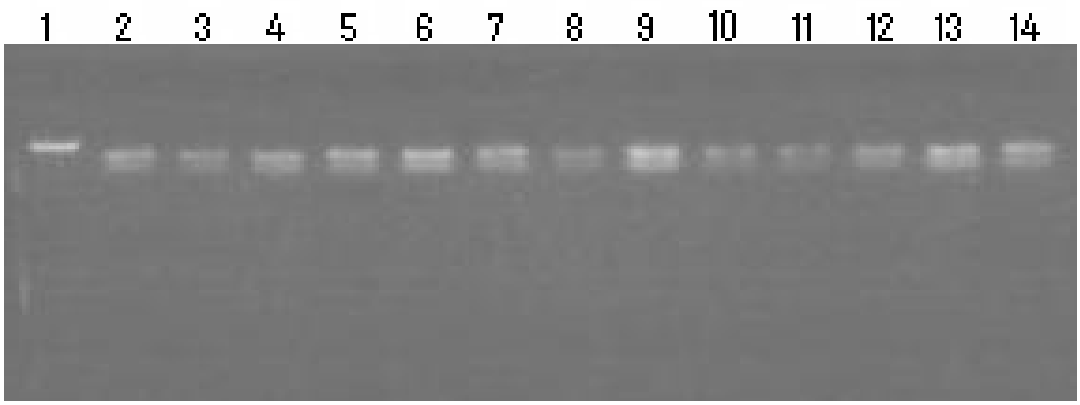

Fig. 2b. Restriction digests of 5S rDNA gene with enzyme *Hpa*II on Agarose gel. Line 1, PCR product. Lines 2 to 14 were Nyala, SB8, SB9, SB19, SB20, Mikumi, Mikese, TGX-1876-2E, Duicker, Nam I, Nam II Nam 4M and Maksoy 1N in that order.

```
SB 19 F seq       TCCCGGGGGCCCGGAGACGGTGTCCCGTGGGAGTCGTCACGACACAACATTTACA
SB 19 R seq       TCCCGGGGGCCCGGAGACGGTGTCCCGTGGGAGTCGTCACGACACAACATTTACA
MIKUMI F seq      TCCCGGGGGCCCGGAGACGGTGTCCCGCGGGAGTCGTCACGACACAACATTTACA
MIKUMI R seq      TCCCGGGGGCCCGGAGACGGTGTCCCGCGGGAGTCGTCACGACACAACATTTACA
MAKSOY IN F seq   TCCCGGGGGCCCGGAGACGGTGTCCCGTGGGAGTCGTCACGACACAACATTTACA
MAKSOY IN R seq   TCCCGGGGGCCCGGAGACGGTGTCCCGTGGGAGTCGTCACGACACAACATTTACA
```

Fig. 3. Section of gene sequences of one cultivar each from East African countries.

Sequence analysis of 5S rDNA genes

Limited sample sequencing of the 5S gene fragment was carried out to determine possible sequence variations. Sequence analysis identified a restriction marker site within the 5S ribosomal gene in the local cultivars (Fig. 3). The 5S rDNA sequences for the East African soybean were identical except for one base that appeared different in Mikumi.

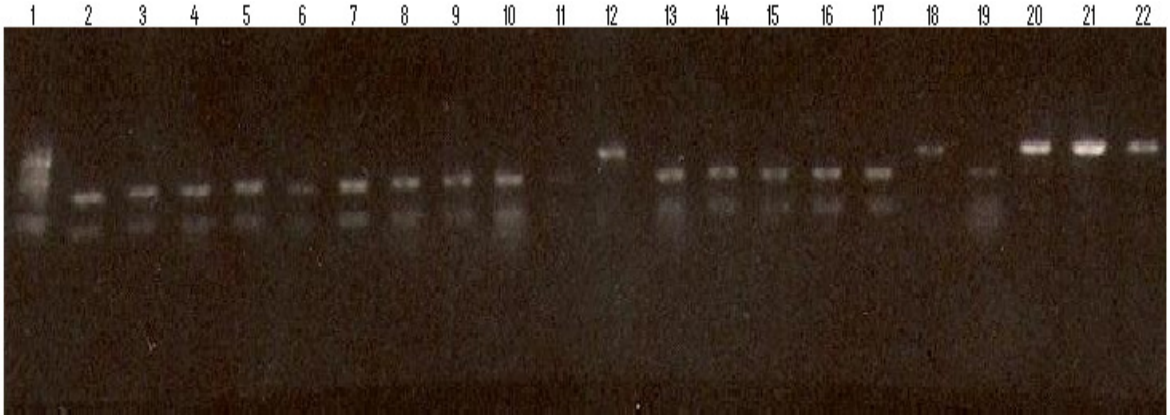

Fig. 4. Restriction digests of 5S rDNA gene of the soybean cultivars using an enzyme *Sac* II on agarose gel. Lane 1 is Marker, lanes 2-4 Duicker, 5 and 6 Mikese, 7 Mikumi, 8-10 TGX-1876-2E, 11 and 12 Maksoy, 13 Nam 4M, 14-16 Nam II, 17-19 Nam I, 20-22 Nyala.

SacII marker site

The sequence analysis identified *Sac*II digestion with restriction enzyme gave distinct restriction patterns (Fig. 4). However, these mutations did not correspond to differences as shown by agronomic performance.

Gene BLAST analysis

Blast analysis of 5S ribosomal gene of the East African cultivars against different species of genus *Glycine* show that all the cultivars studied have high (≥99%) homology with E = 0.00 within *Glycine* spp.

A phylogenetic tree was constructed based on 5S rDNA gene sequences of East African cultivars and different Glycine species. This phylogenetic analysis suggests that cultivar Mikumi is similar to *Glycine max* while SB 19 of Kenya and Maksoy of Uganda are the same species and also closely related to *Glycine soja.*

This type of study can, therefore, trace the East African soybeans origin in comparison with *Glycine* species globally (Fig. 5). Further analysis is needed to reflect genotype and agronomic data of specific region of the world.

Discussion

Improvement of our local soybean cultivars will require identifying molecular makers that define their relationship to other improved varieties and also makers linked to quantitative trait loci (Keim *et al.,* 1990). This study is a pioneering effort in this region, to study molecular makers that correlates to the differences in agronomic performances observed in East African soybean cultivars (Nassiuma & Wasike, 2002).

There is negligible size variation in soybean cultivars rDNA (Nickrent & Patrick, 1998). Similarly, PCR product from cultivars in East African region was uniform (Fig. 1). Restriction enzyme data show that makers that can identify disparate varieties of genus Glycine but did not show differences among the local cultivars (Figs 2 and 3). This is not unexpected since soybean flower do not cross pollinate domesticated cultivars are developed from a few hybrid crosses.

Sequence alignment of the local cultivars identified a single *Sal II* restriction site (Fig. 3) within 5S rDNA ITS region. This enabled genotyping of large population size of soybean cultivars from Lake Victoria region (Fig. 4). The data show that most cultivars in Kenya and Uganda has *Sal II* site in their 5S rDNA while those from Tanzania does not. This may indicate differences in lineages of cultivars that were introduced in the region.

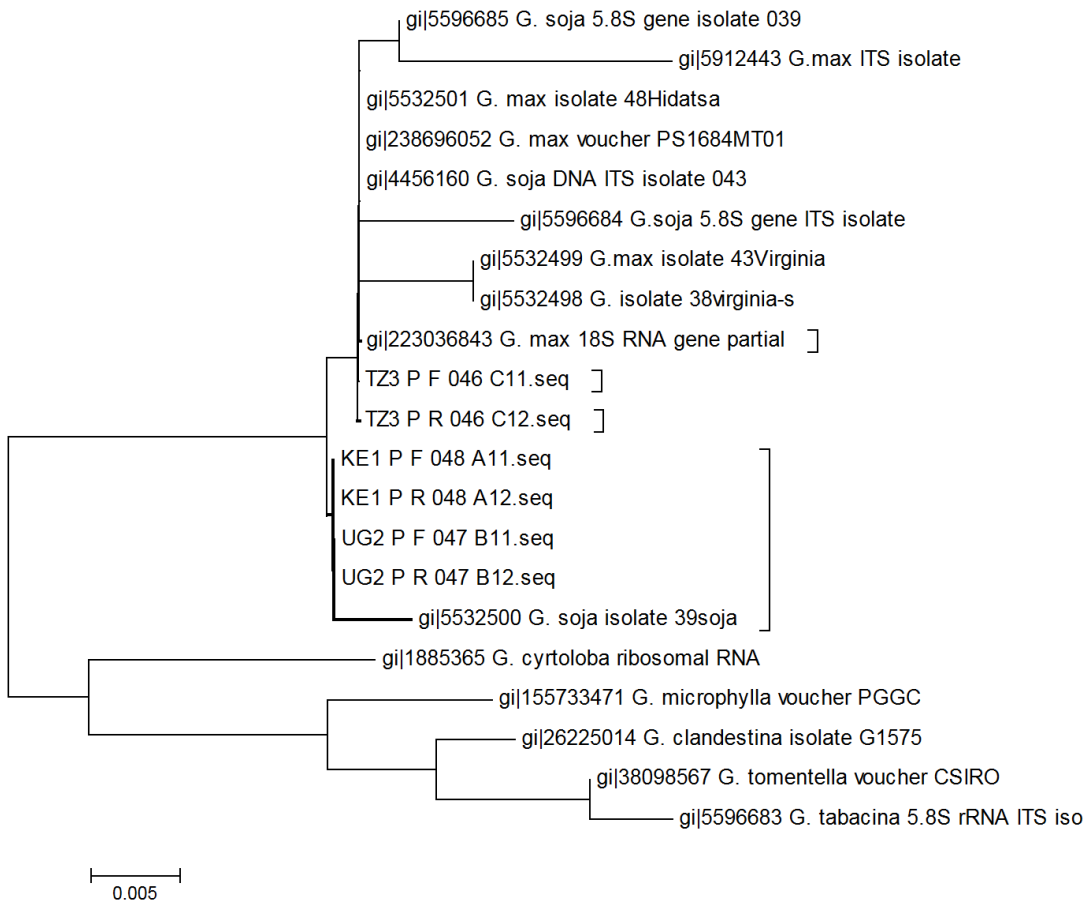

Fig. 5. Phylogenetic relationship between the East African cultivars and global *Glycine* species.

The *Sal II* maker did not correlate with agronomic performance observed in the cultivars (data not shown). The maker reflects an underlying genotype difference that may not have phenotypic expression. Moreover, the maker site is within a non coding (ITS) region. Further SNP and SSR genotyping of the two varieties of cultivars in the East African region can elucidate other genotype differences than can relate the local cultivars to the original hybrid lines. A phylogenetic analysis of rDNA sequences of the local cultivars against similar sequences available in public databases (NCBI: http://www.ncbi.nlm.nih.gov/) indicate that the genotypes cultivars grown in the Lake Victoria region of East Africa are related to *Glycine max* and *Glycine soja* isolates and distantly related to the indigenous cultivars. (Fig. 2). Further studies are necessary to identify QTL markers that relate to agronomic performances. Single nucleotide polymorphisms (SNPs) provide a rich source of useful molecular markers in genetic analysis.

SNPs can be analyzed using high-throughput and cost effective systems, they are useful for construction of high-density genetic maps as well as for genetic association studies (Cho *et al*., 1999, Picoult-Newberg *et al*., 1999; Nairz *et al*., 2002; Rafalski 2002; Kota *et al*., 2003). The relatively high level of linkage disequilibrium (LD) that would be anticipated in self-fertilizing plant species such as soybean may permit whole genome scans using SNPs for QTL discovery (Rafalski, 2002; Tenaillon *et al*., 2001; Zhu *et al*., 2003). Expressed sequence tag (EST) data serve as a useful source of DNA sequences in which SNPs can be discovered. The Soybean EST Project database contained more than 342,000 publicly available ESTs from 84 libraries as of December 2004. This resource provides an excellent source for the development of gene-derived SNP markers (Collins *et al*., 1998; Brookes, 1999; Marth *et al*., 1999; Picoult-Newberg *et al*., 1999).

References

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

Baldwin B G, Sanderson M J, Porter M, Wojciechowski M F, Campbell C S, Donoghue M J 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of The Missouri Botanical Garden* **82**:247–277.

Brookes A J. 1999. The essence of SNPs. *Gene* **234**:177–186.

Bult C J, Sweere J A, Zimmer E A . 1995. Cryptic Sequence Simplicity, Nucleotide Composition Bias, and Molecular Coevolution in the large subunit of Ribosomal DNA in Plants: implications for phylogenetic analyses. *Annals of The Missouri Botanical Garden* **82**:235–246.

Caviness C E. 1966. Estimates of natural cross-pollination in Jackson soybeans in Arkansas. *Crop Science* **6**:211.

Cho R J, Mindrinos M, Richards D R, Sapolsky R J, Anderson M, Drenkard E, Dewdney Stammers M J, Reuber T L, Federspiel N *et al.* 1999. Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *National Genetics* **23**:203–207.

Collins F S, Brooks L D, Charkravarti A. 1998. A DNA polymorphism discovery resource for research on human genetic variation. *Genome Research* **8**:1229–1231.

Felsenstein J. 1997. An alternating least squares approach to inferring phylogenies from pairwise distances. *Systems Biology* **46**:101–111

Gupta P K, Roy J K, Prasad M. 2001. Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. *Current Science* **80**:524–535.

Hemleben V, Ganal M, Gerstner J, Schiebel K, Torres R.A. 1988. Organization and length heterogeneity of plant ribosomal RNA genes. In *The architecture of eukaryotic genes*, pp. 371–383. Ed. G Kahl. Weinheim: VHC.

Hymowitz T, 2004. Speciation and cytogenetics. In *Soybeans: Improvement, Production, and Uses*, pp. 97–136. Ed. H R Boerma and J . Specht. Madison, WI: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America.

Jorgansen R A. Cluster P D. 1988. Modes and tempos in the evolution of nuclear ribosomal DNA: New characters for evolutionary studies and new markers for genetic and population studies. *Annals of The Missouri Botanical Garden* **75**:1238–1247.

Keim P Diers B W, Olson T C, Shoemaker R. C. 1990. RFLP Mapping in Soybean Association: Association Between Maker Loci and Variation in Quantitative traits. *Genetics* **126**:755–742.

Kota R, Rudd S, Facius A, Kolesov G, Thiel T, Zhang H, Stein N, Mayer K, Graner A. 2003. Snipping polymorphisms from large EST collections in barley (*Hordeum vulgare* L.). *Molecular & Genetic Genomics* **270**:24–33.

Kuzaff R K, Seveere J A, Soltis D E, Soltis P S, Zimmer E A. 1998. The phylogenetic potential of entire 26S rDNA sequences in plants. *Molecular & Biological Evolution* **15**:251–263.

Marth G T, Korf I, Yandell M D, Yeh R T, Gu Z, Zakeri H, Stitziel N O, Hillier L, Kwok P Y, Gish W R. 1999. A general approach to single-nucleotide polymorphism discovery. *National Genetics* **23**:452–456.

Morse W J. 1950. History of Soybean Production In *Soybean and soybean products*, Vol. 1, pp. 3–59. Ed. K S Markley. Easton, PA, USA: Interscience Publishers.

Nairz K, Stocker H, Schindelholz B, Hafen E. 2002. High-resolution SNP mapping by denaturing HPLC. *Proceedings of the National Academy of Science, USA* **99**:10575–10580.

Ngeze P B. 1993. *Learn how to grow soybean.* Acacia Stantex. 1 pp.

Nickrent D L, Soltis D E. 1995. A comparison of angiosperm phylogenies based upon complete 18S rDNA and rbcL sequences. *Annals of The Missouri Botanical Garden* **82**:208–234.

Picoult-Newberg L, Ideker T E, Pohl M G, Taylor S L, Donaldson M A, Nickerson D A, Boyce-Jacino M. 1999. Mining SNPs from EST databases. *Genome Research* **9**:167–174,
Rafalski A. 2002. Applications of single nucleotide polymorphisms in crop genetics. *Current Opinions in Plant Biology* **5**:94–100.
Rogers S O, Bendich A J. 1987. Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer. *Plant Molecular Biology* **9**:509–520.
Schaal B A, Learn G H J. 1988. Ribosomal DNA variation with and among plant populations. *Annals of The Missouri Botanical Garden* **75**:1207–1216.
Soltis D E, Soltis P S, Nickrent D L, Johnson L A, Hahn W J, Hoot S B, Sweere J A, Kuzof R K, Kron K A, Chase M W, Swensen S M, Zimmer E A, Chaw S M, Gillespie L J, Kress J W, Sytsma K J. 1997. Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. *Annals of The Missouri Botanical Garden* **84**:1–49.
Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA 4: Molecular Evolutionary Genetics Analysis (MEGA) software Version 4.0. *Molecular Biology Evolution* **24**:1596–1599.
Tenaillon M I, Sawkins M C, Long A D, Gaut R L, Doebley J F. 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proceedings of the National Academy of Science, USA* **98**:9161–9166.
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. 2003. Single-nucleotide polymorphisms in soybean. *Genetics* **163**:1123–1134.