International Scholars Journals

African Journal of Plant Breeding ISSN 2375-074X Vol. 6 (12), pp. 001-007, December, 2019. Available online at www.internationalscholarsjournals.org © International Scholars Journals

Full Length Research Paper

Molecular genotyping of selected soybean (*Glycine max* L.) genotypes grown in different regions of East Africa

Fredrick M. Njoka[1]*, Zedekiah A. Okwany[1] and Elijah K. Githui[2]

[1]Kenyatta University, P. O. Box 4388, Nairobi Kenya.

[2]National Museums of Kenya, P. O. Box 78420-00500, Ngara Road, Nairobi, Kenya.

Accepted 10 August, 2019

Soybean (*Glycine max* (L.) Merril) is one of the most popular pulses because of its protein content and nodulating ability. However, most farmers and researchers do not have sufficient genetic information of the genotypes that they grow. This research aimed at determining the molecular characteristics of selected soybean genotypes grown in different regions of East Africa and make recommendations to the farmers and researchers. Genotypes were collected from Uyole, Ukiriguru, Kawanda, Namulonge, Njoro and Embu Agricultural Institutes as well as from farmers. Molecular analysis was done by DNA being extracted according to phenol chloroform method. This was followed with PCR process using custom ordered pair of primers that corresponded to the flanking ends of the targeted gene fragment (5S ribosomal genes). Restriction fragment length polymorphism (RFLP) and gene clean were then performed on the isolated PCR fragments. The elute was confirmed on agarose gel then sent for automated sequencing ABI prism (Applied Biosystems) at ILRI. The resulting gene sequences were compared with gene sequences of known *Glycine* species using various enzymes in computer based simulations. The gene sequences were then subjected to gene blast using MEGA 4 and resulted with a phylogenetic tree for the selected East African soybean. The studied cultivars were found to be closely related to *Glycine max* species. Enzyme *Sac11* was identified as a marker for the East African soybean genotypes. These results may also assist plant breeders to produce hybrids with the best performing cultivars based on their genetic diversity.

Key words: Soybean, DNA analysis, sequencing, polymorphism, markers.

INTRODUCTION

Genes greatly determine morphological characteristics of living organisms. Environmental conditions may also determine these characteristics, some of which are important in agronomy. The cultivars grown in East Africa show differences in agronomic performance but there has not been a genetic explanation at molecular level or based on DNA genotyping. According to Gupta et al. (2001) and Kota et al. (2003), development of DNA-based markers is important for selection and improvement of varieties. 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 hundred flowering plants, mainly owing to interest in using these data for phylogenetic analyses. Sequences include those from small-subunit (18S) rDNA (Nickrent and Soltis, 1995; Soltis et al., 1997), large subunit (26S) rDNA (Bult et al., 1995; Kuzoff et al., 1998), and a large number of internal transcribed spacer (ITS) and 5.8S rDNA sequences (Baldwin et al., 1995). ITS are the most variable portion of the rDNA cistron and are sometimes referred to as the NTS (non transcribed spacer), a misnomer, since large portions of it are transcribed. In contrast to the conserved 18S and 26S rDNA, which have phylogenetic utility at deeper divergence levels, ITS sequences often contain sufficient variation to allow exami-

*Corresponding author. E-mail: njokafm@yahoo.co.uk. Tel: (254)721426308.

nation of genetic relationships between cultivated varieties (cultivars), populations, and individuals (Rogers and Bendich, 1987; Hemleben et al., 1988; Jorgansen and Cluster, 1988; Schaal and Learn, 1988). The ITS rDNA varies widely in length among different plant groups, ranging in size from 0.1 to over 1.2 kb (Rogers and Bendich, 1987). Such length heterogeneity is a result of the presence of at least one and often many, tandem (direct) or dispersed sub-repeat domains. The advent of direct sequencing of PCR product has precluded the need for cloning, thus greatly facilitating broad-scale sequencing studies of many genes; however, the length of plant ITS sequences and the lack of universally conserved sites among different species within ITS's may explain the relative paucity of these sequences. Single nucleotide polymorphisms (SNPs) including insertion/ deletions (indels) can provide a rich source of useful molecular markers in genetic analysis. Because 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. In contrast, the lower levels of LD in out-crossing species such as maize (*Zea mays* sp. *mays* L.) will require the use of the candidate gene approach to discover the specific gene(s) underlying phenotypic changes (Rafalski, 2002; Tenaillon et al., 2001; Zhu et al., 2003). Mutations in coding DNA sequences (cSNPs) may change amino acid sequences and affect gene function and could therefore be valuable as markers of agronomic traits. 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 develop-ment of gene-derived SNP markers (Collins et al., 1998; Brookes 1999; Marth et al., 1999; Picoult-Newberg et al., 1999). This experiment was done in collaboration between researchers from Kenyatta University and the

National Museums of Kenya.

METHODOLOGY

Molecular work was carried out at the Molecular Biology Laboratories of the National Museums of Kenya and the International Livestock Research Institute (ILRI). Seedlings of thirteen genotypes from different East African regions that had been grown the previous season were used in characterizing their 5S rDNA. Genotypes were collected between January and April, 2006 from Uyole Agricultural Research Institute, Mbeya and Ukiriguru Agricultural Research Institute all in Mwanza in Tanzania; Kawanda Agricultural Research Institute, Namulonge Research Institute and Makerere University Agricultural Research Institute, Kabanyoro in Uganda and National Plant Breeding Research Institute, Njoro, Kenya National Gene Bank, Muguga and Kenya Agricultural Research Institute all in Embu in Kenya. In addition some genotypes were collected from farmers in different parts of the East African countries. They were planted the same season.

Genomic DNA extraction

Approximately 1.0 g young shoot of soybean was obtained from the field, washed with distilled water and placed in sterile mortar and quickly ground into pulp using a pestle then immediately subjected to DNA extraction procedures. STE extraction buffer (200 µℓ) (1M Tris (pH 7.5), 5MNaCl, 0.5M EDTA, 10% SDS and 7µℓ of mecaptoethanol) was added to the ground leaf materials then transferred to eppendorf tubes. An aliquot of 500 µℓ of the Phenol : Chloroform (1:1) was added and vortex for 10 s, then left at room temperature for about 45 min. It was then centrifuged at 10000 rpm for 10 min. Supernatant (500 µℓ) was transferred into a fresh eppendorf tube and an equal volume of cold chloroform: Isoamyl alcohol (24:1) added and mixed well by inversion to emulsify then centrifuged at 10000 rpm for 10 min. Chloroform: Isoamyl wash was repeated twice. Supernatant (500 µℓ) was transferred into a fresh eppendorf tube then an equal volume of chilled isopropanol (propan-2-ol) was added, mixed and incubated at -20°C for 15 min. It was then centrifuged at 14000 rpm for 5 min at room temperature to pellet the DNA and the pellets washed with 450 µℓ of 70% ethanol then centrifuging at 14000 rpm for 5 min. Alcohol was poured off and eppendorf tubes were inverted to air dry the DNA pellet. The pellet was then suspended in 100 µℓ T.E (10 mM Tris-HCl, 1mM ETDA [pH 7.5]). The isolated genomic DNA was stored at -20°C. Quality of extracted DNA was determined by electro-prhoresis on 1% agarose gel run for about 30 min at 90 mA (150 V), and then visualized under UV illumination. The image was captured with Polaroid film type 667.

PCR- Polymerase chain reaction

PCR on the extracted DNA was done using 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. Amplification reaction was performed in a 25 µℓ volume, containing 14 µℓ double distilled water, 2.5 µℓ PCR buffer (pH 8.4), 2.0 µℓ $MgCl_2$, 2.0 µℓ of dNTPs, 1.0 µℓ of each primer, 2.0 µℓ of genomic DNA and 0.5 µℓ of Taq DNA polymerase. The temperature profile was as follows: initial denaturation at 94°C for 7 min; 35 cycles of denaturation at 94°C for 1 min, annealing at 53°C for 1 min, and extension at 68°C for 1 min, and final extension at 72°C for 7 min. PCR products were loaded into separate lanes of a 1% agarose gel. The gel ran for about 30 min at 90 mA (150 V), and then was visualized under UV illumination and photographed with Polaroid film type 667.

Restriction fragment length polymorphisms (RFLPs) analysis

Restriction endonuclease digestion of genomic DNA with Hae III, Hpa II, Mse I, Taq I, Sac II and Sma I enzymes was conducted according to manufacturer's instructions. Digests were incubated for 3 ho using 2 to 5 units of enzymes per µg DNA. DNA fragments were separated according to size by electrophoresis in a 1% agarose gel using TAE running buffer and Polyacrylamide Gel Electrophoresis (PAGE) in TBE (90 mM Tris, 90 mM borate, 2 mM EDTA) for approximately 30 min at 90 mA (150 V). The restriction pattern was visualized under UV illumination and photographed with Polaroid film type 667. 123 base pair ladder was used as molecular size marker.

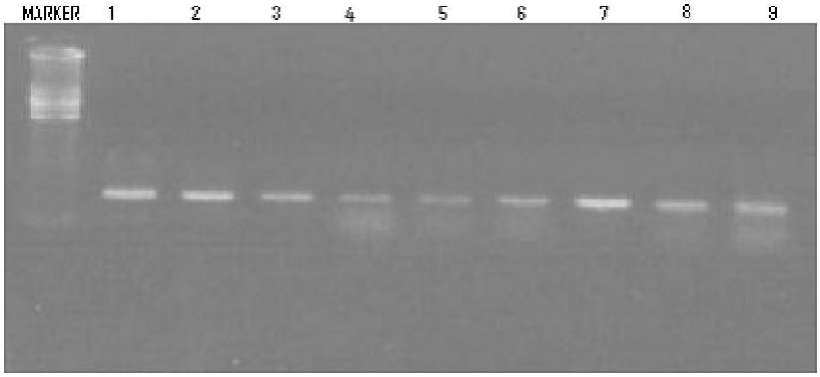

Figure 2. PCR product of the East African soybean *Glycine max*. 5S rDNA gene on an agarose gel.

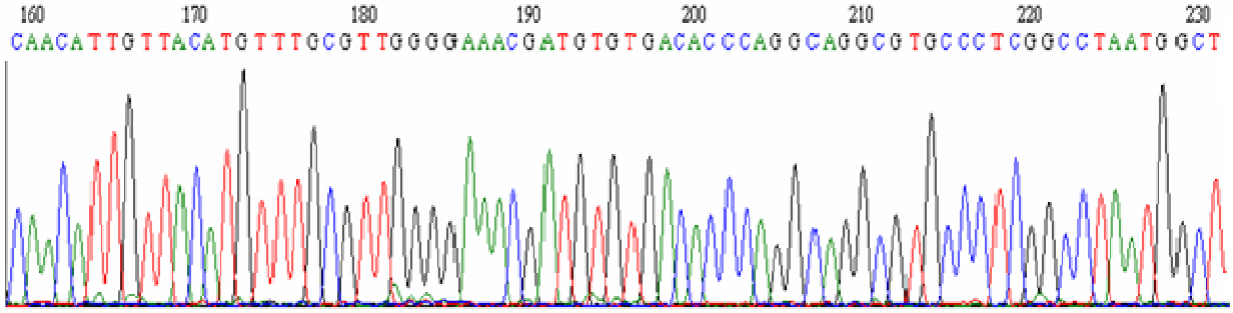

Figure 1. A chromatogram file showing a gene sequence.

Gene clean

Bands were excised from agarose gel using straight-edge razor blade and placed in different well labeled microcentrifuge tubes. Three volumes (1.2 ml for 0.4 g slice) NaI stock solution was added. It was then incubated at 56°C in water bath for 5 min, mixing after two minutes until all agarose had melted. Glassmilk stock solution was vortexed to mix well then 20 µl glassmilk was added to the microcentrifuge tubes. The tubes containing the mixtures were incubated for 15 min on ice, mixing every 3 to 5 min to keep glassmilk in suspension. The tubes were then microcentrifuged on high for 5 s to pellet the glassmilk. Supernatant was removed and placed into new tubes. 500 µl (10-50 volumes) ice-cold New Wash was added to silica gel pellet and the pellets resuspended. The tubes then microfuged for 10 s on high, and the supernatant discarded. The wash was repeated two more times (3 washes total). After discarding supernatant from last wash, the pellets were spinned for 10 s; and the residual new wash removed with pipette. The pellets were then resuspended in 15 µl dH2O; incubated for 5 min at 45 to 55°C in a water bath; microfuged on high for 30 s to pellet silica gel. DNA-containing supernatant was then transferred to new tubes. The tubes were then spinned for a few seconds on high to remove any residual glassmilk carried over with supernatant transfer. Supernatant was then transferred to new tubes.

Sequencing

The eluate was confirmed on agarose gel (Figure 2) then sent for automated sequencing ABI prism (Applied Biosystems) at ILRI.

Sequencing procedures utilizes fluorescent tagged NTP's in the sanger's dideoxy chain termination method. The PCR reaction consists of DNA template, primer, DNA polymerase, unlabeled dNTP's, fluorescently labeled ddNTP's, and appropriate buffer in ultrapure sterile water. All customer samples are run using the same standard PCR cycle on a GeneAmp 9600. During the PCR reaction, the double stranded DNA template is denatured one time, before the cycles start, for 2 min. at 96^{O}C. After this "hot start", the temperature cycle begins with denaturation at 96^{O}C for 10 s, a ramp down to 50^{O}C for 15 s. to allow the primers to anneal to the template's priming site, then a ramp up to 60^{O}C for 4 min. to allow extension of the primer by the polymerase. This temperature cycle is repeated for 25 cycles. The DNA polymerase used is "Amplitaq FS", a thermostable modified form of Thermus aquaticus DNA polymerase. Amplitaq has no 3'-5' exonuclease activity, enabling rapid and efficient incorporation of dNTP's and the fluorescent ddNTP's.

When the PCR is complete, the reaction mix contains a population of PCR fragments of different lengths, each terminating in a fluorescent-dye-containing dideoxynucleotide. Each dideoxy-nucleotide base contains a different fluorescent dye which emits a characteristic wavelength, thus the identity of the dye corresponds to the final base on that fragment. The entire reaction is purified, then run in a single lane on a polyacrylamide gel in an ABI 377 Sequencer, so that the fragments separate according to size. As the fragments are electrophoresed, they run past a laser detector at the bottom of the gel, and the emission wavelength of each fragment is detected. The data is compiled into a gel image, analyzed by ABI software and the resulting sequence is written into a text file and a chromatogram file as shown in Figure 1.

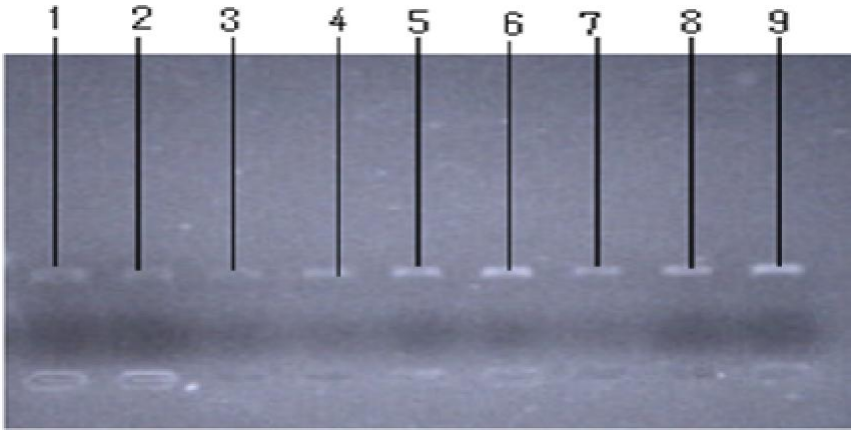

Figure 3. Gene clean of the East African soybean *Glycine max.* 5S rDNA gene on an agarose gel.

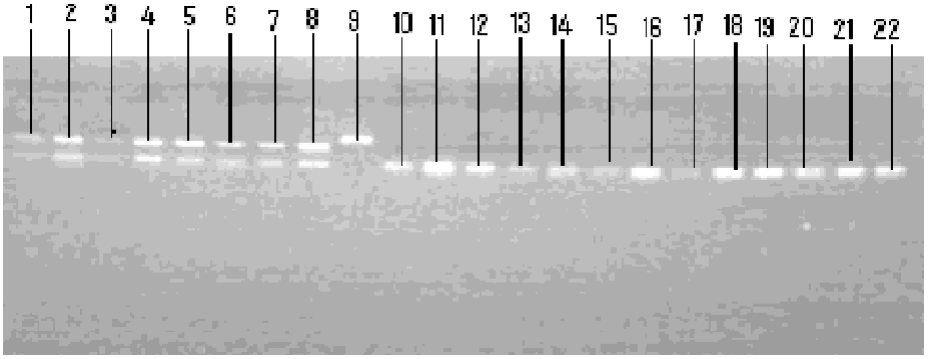

Figure 4. Restriction digest of 5S rDNA gene with enzyme *Mse*I. (Lanes 1-8) and *Hae*III (Lanes 10-22).

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

Phylogenetic analysis

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

RESULTS AND DISCUSSION

PCR amplification of the 5S ribosomal gene from the cultivars yielded uniform PCR product of about 1 kb. The patterns were uniform (Figure 2).

The order of the cultivars is as follows; Nyala, Mikumi, Duicker, Mikese, SB20, Nam 4M, SB9, SB8 and Maksoy IN.

The order of the cultivars from 1-9 is as follows; Nyala, SB8, SB19, Mikumi, TGX-1876-2E, Duicker, Maksoy IN, Nam 4M and Nam I respectively (Figure 3). Digestion with restriction enzymes *Hae*III, *Hpa*II and *Mse*I gave similar patterns across the cultivars (Figures 4 - 6).

The choice of these enzymes had been determined by simulation restriction digests across the known *Glycine* sp.). Digestion with restriction enzymes *Mse*I was done and electrophoresis done on agarose gel. The bands produced similar patterns (Figure 4). The bands from digestion with enzyme *Hae*III also produced similar pattern in agarose gel. Lane 1,2,3,4,5,6,7 and 8 were Nyala, Mikumi, Maksoy 1N, SB20, Mikese, TGX-1876-2E and Nam II respectively. Lane 9 was PCR product. This was loaded to act as a marker. Lanes 10 to 22 were Nyala, SB8, SB9, SB19, SB20, Mikumi, Mikese, TGX-1876-2E, Duicker, Nam I, Nam II Nam 4M and Maksoy 1N in that order. These could not be analysed further thus the digests were electrophoresised on polyacry-lamige gel for better analysis. Restriction digest with enzymes *Hpa*II was performed and electrophoresis done

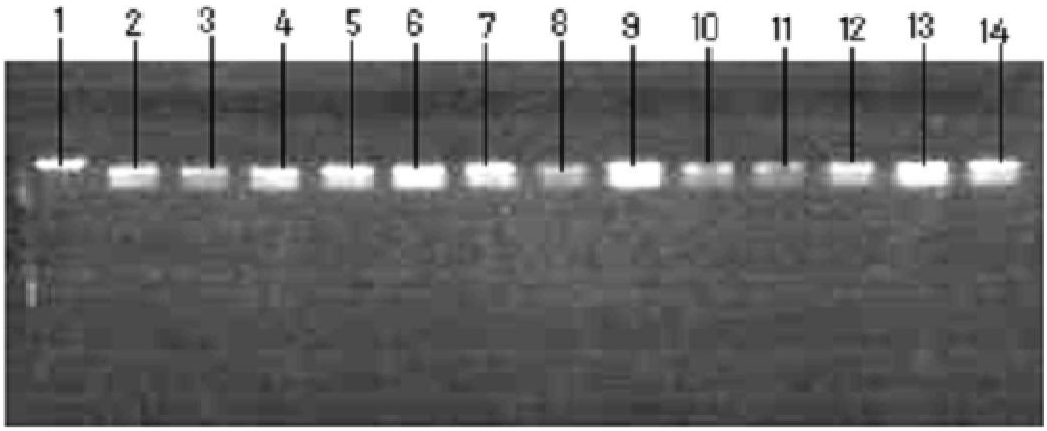

Figure 5. Restriction digests of 5S rDNA gene with enzyme *Hpa*II on agarose gel.

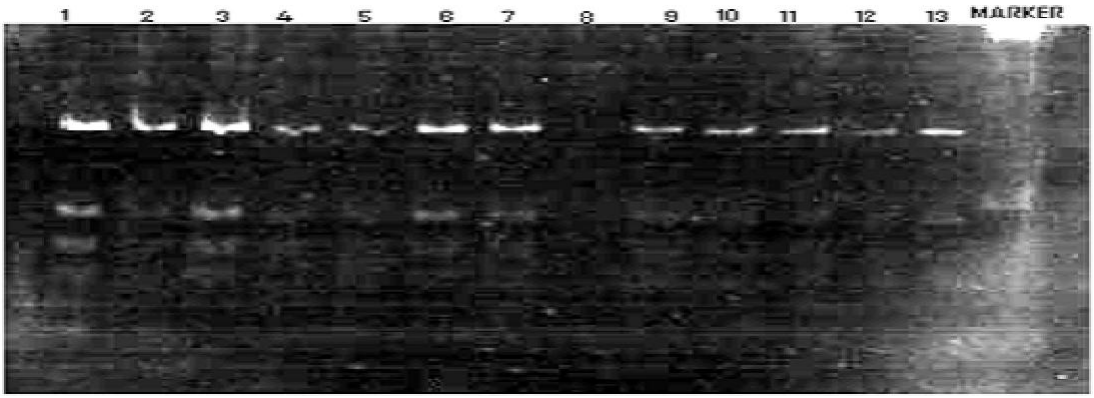

Figure 6. Restriction digests of 5S rDNA gene with An enzyme *Hpa*II. Lanes 1-13 show Nyala, SB8, SB9, SB19, SB20, Duicker, and Mikese. TGX1876-2E, Maksoy IN, Nam I, Nam II and Nam 4M respectively.

on agarose gel. The bands could not be easily analysised since they were too close to one another. Lane 1 was PCR DNA product (Figure 5). Lanes 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. All these produced bands with similar patterns.

When PAGE was done five bands were observed as predicted from simulation restriction digests across the known *Glycine* sp.

However, digestion with restriction enzyme *Sac*II gave distinct restriction patterns (Figure 7).

Sequence analysis of 5S rDNA genes

Sequencing results were analyzed, a restriction enzyme marker site within the 5S ribosomal gene was determined. Similar sequences for the East African soy beans except for one region that appeared different in Mikumi. This process enables the determination of a marker for the East African soybean cultivars.

Blast analysis of the East African cultivars and different species of genus *Glycine* (Figure 8) confirmed that all the cultivars studied belong to the genus *Glycine.* 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 *G. max* while SB 19 of Kenya and Maksoy of Uganda are the same species and also closely related to *G. max*

This study therefore managed to trace the East African soybeans origin in comparison with *Glycine* species globally Figure 9.

CONCLUSION AND RECOMMENDATIONS

A phylogenetic tree based on 5S rDNA gene sequences suggested that Mikumi is similar to *G. max*. SB19 and Maksoy are the same species and closely related to *G. max.* The study therefore managed to trace East African soybeans origin in comparison with *Glycine* species globally. Genotypic diversity among the studied genotypes was not influenced by the region of origin.

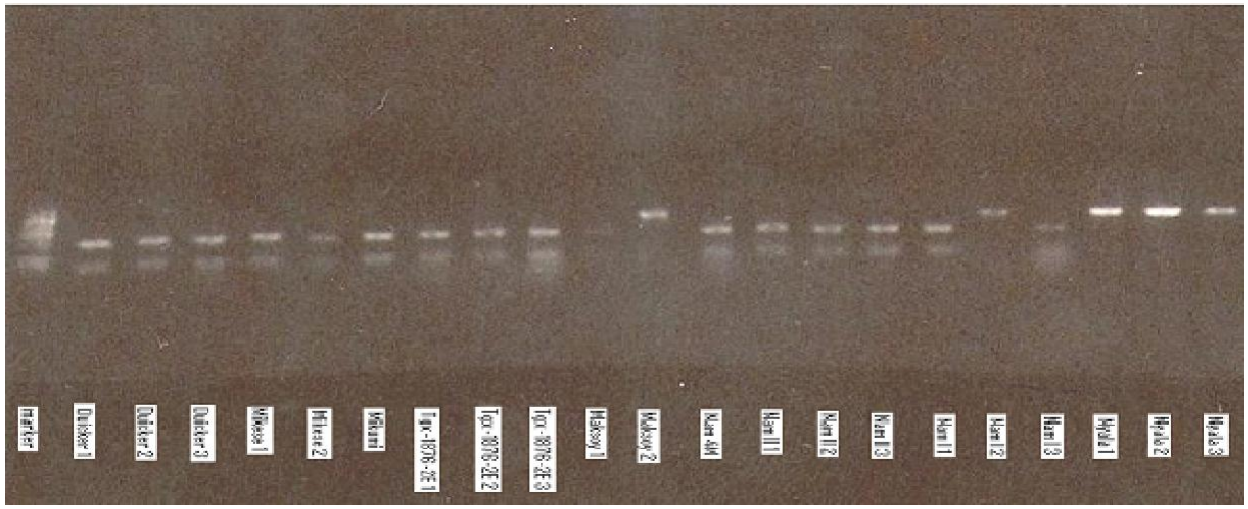

Figure 7. Restriction digests of 5S rDNA gene of the soybean cultivars using an enzyme *Sac* II.

```
SB19 F seq      TCCCGGGGGCCCGGAGACGGTGTCCCGTGGGAGTCGTCACGACACAACATTTACA
SB19 R seq      TCCCGGGGGCCCGGAGACGGTGTCCCGTGGGAGTCGTCACGACACAACATTTACA
MIKUMI F seq    TCCCGGGGGCCCGGAGACGGTGTCCCGCGGGAGTCGTCACGACACAACATTTACA
MIKUMI R seq    TCCCGGGGGCCCGGAGACGGTGTCCCGCGGGAGTCGTCACGACACAACATTTACA
MAKSOY IN F seq TCCCGGGGGCCCGGAGACGGTGTCCCGTGGGAGTCGTCACGACACAACATTTACA
MAKSOY IN R seq TCCCGGGGGCCCGGAGACGGTGTCCCGTGGGAGTCGTCACGACACAACATTTACA
```

Figure 8. Section of gene sequences of one cultivar each from East African countries.

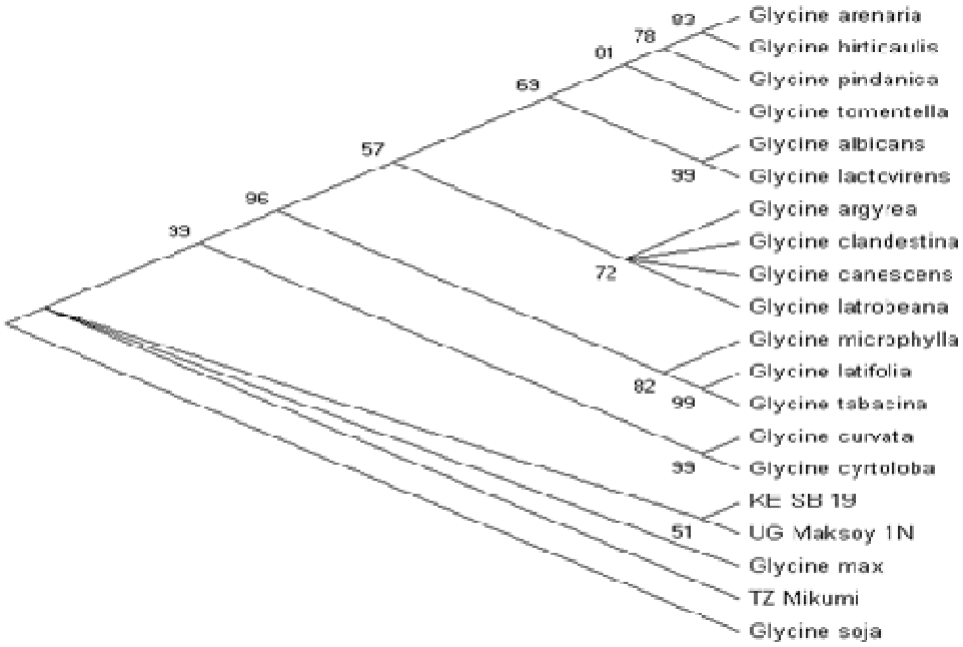

Figure 9. Phylogenetic relationship between the East African cultivars and global *Glycine* species.

Enzyme Sac11 has been identified as a good marker for the East African soybean genotypes. Plant breeders can now base their selection on the diversity shown in this research. Further study should be conducted on the ribosomal gene using SNP marker.

ACKNOWLEDGEMENTS

We are grateful to VicRes for the funding. We also thank the following institutions for all the support: National Museums of Kenya, Kenyatta University, International Centre for Insect Physiology and Ecology (ICIPE- Mbita Point).

REFERENCES

Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ (1995). The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. Ann. Mo. Bot. Gard. 82: 247-277.

Brookes AJ (1999). The essence of SNPs. Gene. 234: 177-186.

Bult CJ, Sweere JA, Zimmer EA (1995). Cryptic sequence simplicity, nucleotide composition bias, and molecular coevolution in the large subunit of ribosomal DNA in plants: implications for phylogenetic analyses. Ann. Mo. Bot. Gard. 82: 235-246.

Collins FS, Brooks LD, Charkravarti A (1998). A DNA polymorphism discovery resource for research on human genetic variation. Genome Res. 8: 1229-1231.

Gupta PK, Roy JK, Prasad M (2001). Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. Curr. Sci. 80: 524-535.

Hemleben V, Ganal M, Gerstner J, Schiebel K, Torres RA (1988). Organization and length heterogeneity of plant ribosomal RNA genes. In The architecture of eukaryotic genes. Edited by G. Kahl. VHC, Weinheim. pp. 371-383.

Jorgansen RA, Cluster PD (1988). Modes and tempos in the evolution of nuclear ribosomal DNA: new characters for evolutionary studies and new markers for genetic and population studies. Ann. Mo. Bot. Gard. 75: 1238-1247.

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.). Mol. Gen. Genomics 270: 24-33.

Kuzoff RK, Seveere JA, Soltis DE, Soltis PS, Zimmer EA (1998). The phylogenetic potential of entire 26S rDNA sequences in plants. Mol. Biol. Evol. 15: 251-263.

Marth GT, Korf I, Yandell MD, Yeh RT, Gu Z, Zakeri H, Stitziel NO, Hillier L, Kwok NK, Stocker H, Schindelholz B, Hafen E (1999). High-resolution SNP mapping by denaturing HPLC. Proc. Natl. Acad. Sci. USA 99: 10575-10580

Nickrent DL, Soltis DE (1995). A comparison of angiosperm phylogenies based upon complete 18S rDNA and *rbc*L sequences. Ann. Mo. Bot. Gard. 82: 208-234.

Picoult-Newberg L, Ideker TE, Pohl MG, Taylor SL, Donaldson MA, Nickerson DA, Boyce-Jacino M (1999). Mining SNPs from EST databases. Genome Res. 9: 167-174.

Rafalski A (2002). Applications of single nucleotide polymorphisms in crop genetics. Curr. Opin. Plant Biol. 5: 94-100.

Rogers SO, Bendich AJ (1987). Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer. Plant Mol. Biol. 9: 509-520.

Schaal BA, Learn GHJ (1988). Ribosomal DNA variation with and among plant populations. Ann. Mo. Bot. Gard. 75: 1207-1216.

Soltis DE, Soltis PS, Nickrent DL, Johnson LA, Hahn WJ, Hoot SB, Sweere JA, Kuzof RK, Kron KA, Chase MW, Swensen SM, Zimmer EA, Chaw SM, Gillespie LJ, Kress JW, Sytsma KJ (1997). Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. Ann. Mo. Bot. Gard. 84: 1-49.

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF (2001). Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proc. Natl. Acad. Sci. USA 98: 9161-9166.

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. Genetic 163: 1123-1134.