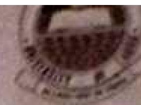

MOLECULAR CHARACTERIZATION OF SOYBEAN *Glycine Max* (L.) MERRIL GENOTYPES USING SSR MARKERS

Joseph[1]*, E. C., Tolorunse[1], K. D., Gana[1], A. S., and Azuh[2], V. O.

[1]Department of Crop Production, Federal University of Technology Minna, Nigeria
[2]Department of Botany, University of Ibadan, Nigeria.

*Corresponding Author: ethelchinaza.joseph@gmail.com; +2348039251146

ABSTRACT

Six SSR primers were used to amplify genomic DNA from 10 soybean genotypes, after which a high level of polymorphism was observed. Seventeen alleles with a mean value of 2.8 alleles per locus was detected. The polymorphic information content (PIC) among genotypes ranged from 0.60 to 0.16 with an average of 0.45 and a gene diversity of 0.51. A phylogenetic tree was produced from the allelic diversity data, which classified the genotypes into two groups. The major cluster was divided into two sub-groups, each consisting of 3 genotypes, while the minor cluster was divided into three sub-groups. The result suggested that the traits (emergence percentage and 100-seed weight) in which high broad sense heritability were observed can be considered as favourable attributes for improvement. The combination of morphological and molecular methods of characterization will be efficient in breeding programmes and that SSR is a useful tool for assessing genetic diversity.

KEYWORDS: Soybean, Characterization, SSR, Polymorphism, Gene diversity

INTRODUCTION

Soybean *Glycine max* (L.) Merril is a very important oil seed which is grown around the world. It belongs to the bean family Fabaceae the genus *Glycine* (Singh, 2017). It is grown basically for its seeds which is rich in oil and protein and is often referred to as the 'Miracle crop' or 'Golden bean' due its numerous uses and economic importance. It is used in the production of a nutritious drink called 'soya milk' and 'awara' (soybeans cake) in Nigeria, as a component of meals for infants and also as a very vital ingredient in the formulation of feeds for poultry and fishery. The oil is used as cooking oil and as the base for shortening, margarine, salad dressings and mayonnaise (Pratap et al., 2012). Lecithin extracted from the oil is also used in the manufacture of health and nutrition products, cosmetics and industrial coatings (List, 2015).

For effective breeding programs, information concerning the extent and nature of genetic diversity or variation within a crop species is essential as this knowledge may help improve desirable traits and can be used by plant breeders to determine suitable parents (Jain et al., 2017). Several methods have been employed in the determination of genetic diversity among species such morphological and biochemical methods but have been discovered to be limited in number and influenced by the environmental conditions (Chauhan et al., 2015). The use of PCR-based molecular markers is seen as an ideal way to conserve plant genetic resources, distinguish between genotypes and determine the breeding performance of species. Single sequence repeats (SSRs) also known as microsatellite markers consist of tandem repeats of nucleotides, with short sequences. They have advantageous attributes, which inform their use as tools for genetic diversity determination in species such as its co-dominant and multi-allelic nature, ability to detect higher levels of polymorphism, easy scorability, simplicity and high ability of being reproduced (Chauhan et al., 2015). Zhang et al. (2013) noted that SSR markers have been widely employed in studies of genetic diversity in soybean as it has been proven that molecular markers are of high advantage in the identification of variability among genotypes. The objective of this study was to group soybean genotypes based on their characteristics as portrayed by the SSR markers used for the study.

MATERIALS AND METHODS

DNA Extraction and Primer Selection

Leaf samples at 3 to 4 weeks after sowing were collected for this purpose. Deoxyribonucleic acid extraction from each soybean genotype was carried out using Dellaporta DNA extraction protocol (Dellaporta et al., 1983). A total of six Simple Sequence Repeats (SSRs) primer pairs (table 1), distributed across the integrated linkage map of soybean (Cregan et al., 1999) were used and synthesized by Integrated DNA Technologies. The SSR markers that were selected had an (ATT)n motif due to their abundance and polymorphic nature in soybeans and their easily interpretable allele patterns (Narvel et al., 2000). Reaction cocktail used for all PCR per primer set included (Reagent Volume µl) - 5X PCR SYBR green buffer (2.5), $MgCl_2$ (0.75), 10pM DNTP (0.25), 10 pM of each forward and backwards SSR primer (0.25), 8000 U of taq DNA polymerase (0.06) and made up to 10.5 with sterile distilled water to which 2 µl template was added. The primer sequences are listed thus;

Polymerase Chain Reaction (PCR)

PCR was carried out in a GeneAmp 9700 PCR System Thermalcycler (Applied Biosystem Inc., USA). A touch down PCR profile was used consisting of an initial denaturing for 5 minutes at 94 °C followed by 9 cycles consisting of a 15 seconds denaturing at 94 °C, 20 seconds annealing at 65 °C and 30 seconds extension at 72 °C followed by 35 cycles consisting of a 15 seconds denaturing at 94 °C, 20 seconds annealing at 55 °C and 30 seconds extension at 72 °C. The reaction was terminated at 72 °C for 7 minutes and kept on hold at 10 °C.

Integrity and Gel Electrophoresis

The integrity of the amplified fragment was checked on a 2 % Agarose gel. The buffer (1XTAE buffer) was prepared and subsequently used to prepare 2 % agarose gel. The suspension was boiled in a microwave for 5 minutes. The molten agarose was allowed to cool to 60 °C and stained with 3 µl of 0.5 g/ml ethidium bromide (which absorbs invisible UV light and transmits the energy as visible orange light). A comb was inserted into the slots of the casting tray and the molten agarose was poured into the tray. The gel was allowed to solidify for 20 minutes to form the wells. The 1XTAE buffer was poured into the gel tank to barely submerge the gel, after which 6 µl of each PCR product was loaded into the wells after the 100 bp DNA ladder was loaded into well 1. The gel was electrophoresed at 120 V for 45 minutes visualized by ultraviolet trans-illumination and photographed. The sizes of the PCR products were estimated by comparison with the mobility of a 100 bp molecular weight ladder that was ran alongside experimental samples in the gel and the presence or absence of a binding pattern was estimated for each primer and used for analysis.

Data scoring and analysis

The PCR products from SSR were analyzed by scoring qualitatively for presence or absence of amplification with 1 representing presence and 0 representing absence. This was used to generate a score matrix on which bases the genetic similarity between the genotypes was estimated. This matrix was imported into the Power marker V3.25 and used to estimate the major allelic frequency, gene diversity and Polymorphism information content. Phylogenetic analysis was estimated using the minimum likelihood estimation as provided by the Molecular Evolutionary Genetics Analysis Version 6.0 (MEGA6) software (Tamura et al., 2013).

RESULTS AND DISCUSSION

In this study, all the 6 SSR markers used for the analysis of genetic diversity amplified scorable loci with none being monomorphic as shown in Table 2. This supports the findings of Chauhan et al. (2015) who characterized 48 soybean genotypes and found all the 21 SSR primers used for the study to be polymorphic. The high rate of polymorphism (100 %) detected in this study can also be compared with the

findings of Tantasawat et al (2011); Giriraj et al. (2015), where similar results were obtained. This high rate of polymorphism in SSR locus may be ascribed to the selected set of primers used for the study.

The genetic diversity which shows the probability that two alleles selected at random are different from others in the population, ranged from 0.66 (satt135, satt294 and satt581) to 0.18 (satt309) with a mean value of 0.51. This is considered to be higher than the results obtained by Anchal et al. (2015) where the genetic diversity ranged from 0.61 to 0.05 with a mean value of 0.23 and the low level of genetic diversity was reported to be attributed to direct introduction, selection from introduced germplasm and single cross hybrids in the soybean breeding programs and therefore proposed the inclusion of more diverse soybean germplasms in breeding programs to boost genetic diversity.

A total of 17 alleles were detected among all the 10 genotypes in the study. The highest allele number was recorded for satt581 (4), the lowest number of alleles were recorded in satt309 and satt601 (2), while satt123, satt135 and satt294 fell in-between, having 3 alleles each. However, low mean values for allele number and polymorphic information content (Table 2) were observed. The allele number for each SSR locus varied from 2 to 4 alleles with a mean value of 2.8. The genetic diversity for each primer was evaluated by the value of polymorphic information content (PIC). The highest PIC value (0.5958) was recorded for satt581 while the lowest was recorded for satt309 (0.1638) with a mean PIC value of 0.45. The major allele frequency was lowest in satt135, satt294 and satt581 (0.4), while the highest (0.9) was recorded for satt309 with a mean value 0.58. This indicates low diversity in the set of soybean genotypes considered (Giriraj et al., 2015) but agrees with the study of Hisano (2008), who observed the same mean value for number of alleles (2.8) and a comparable mean PIC value of 0.40.

Data Scoring

The PCR products were scored qualitatively for presence or absence of amplification with 1 representing presence and 0 representing absence as shown in table 3. Phylogenetic analysis was estimated after which a dendrogram was generated (Figure 1). The allelic diversity data was used to produce a dendrogram (phylogenic tree) to explain the genetic relationships among the tested genotypes (Figure 1). The maximum similarity co-efficient (60 %) was found between TGX 1951-3F versus TGX 1987-62F, JG, TGX 1835-10E and TGX 1904-6F versus TGX 1448-2E. Additionally, the minimum similarity coefficient (40 %) was recorded between genotype TGX 1955, NCRI SOY 56 and NCRI SOY 16 versus TGX 1987-10F. The cluster analysis divided the total 10 genotypes into two groups: 6 genotypes in major group and 4 in minor group respectively. The major group was further subdivided into two subgroups. Each of these subgroups consisted of three genotypes with TGX 1951-3F, TGX 1987-62F and JG making up the subgroup A and subgroup B comprising of TGX 1835-10E, TGX 1904-6F and TGX 1448-2E. The minor group was divided into three subgroups, with TGX 1955 and NCRI SOY 56 found to be distinctly different from NCRI SOY 16 and TGX 1987-10F.

The SSR primers used in the study exhibited high levels of polymorphism, which is an important attribute for ascertaining genetic diversity, hence it serves as an effective and reliable tool for analysis of genetic relationship among cultivars, identifying soybean cultivars and selection of better soybean lines for further research work. Based on the primers used, the ten soybean genotypes were also classified into two major groups. Six genotypes (TGX1951-3F, TGX1987-62F, JG, TGX1835-10E, TGX1904-6F and TGX1448-2E) where found to be genetically similar and classified in group I, while the other four genotypes (TGX1955, NCRISOY 56, NCRISOY 16 and TGX1987-10F) made up group II. The use of tightly linked molecular markers with distinguishable expression of morphological traits such as single nucleotide polymorphisms (SNPs) can also be employed in further research for a more precise classification and identification of diversity in genotypes.

REFERENCES

Anchal B, Dhirendra K, Priya N and Niraj T (2015) SSR analysis of 38 genotypes of soybean (Glycine Max (L.) Merril) genetic diversity in India. Physiology and Molecular Biology of Plants 21(1): 109–115.

Chauhan DK, Bhat J, Thakur A, Kumari S, Hussain Z and Satyawathi CT (2015) Molecular characterization and genetic diversity assessment in soybean (*Glycine max* (L.) Merril) varieties using SSR markers. Indian Journal of Biotechnology 14: 504-510.

Cregan PB, Jarvik T, Bush AL. Shoemaker RC, Larl KG, Kahler AL, Kaya N, Van-Toai TT, Lohnes DG, Chung J and Specht JE (1999) An Integrated Genetic Linkage Map of the Soybean. Crop Science39: 1464-1490.

Dellaporta SL, Wood J and Hicks JB (1983) A plant DNA mini-preparation: Version II. Molecular Biology Reports 1: 19-21.

Giriraj K, Gourav S, Gireesh C, Shivakumar M, Mamta A, Dinesh K, and Syed M (2015) Molecular Characterization and Genetic Diversity Analysis of Soybean (Glycine max (L.) Merr.) germplasm accessions in India. Physiology and Molecular Biology of Plants 21(1): 101–107.

Hisano H, Sato S, Isobe S, Sasamoto S, Wada T, Matsuno A, Fujishiro T, Yamada M, Nakayama S, Nakamura Y, Watanabe S, Harada K and Tabata S (2008) Characterization of the soybean genome using EST-derived microsatellite markers. DNA research. International Journal for Rapid Publication of Reports on Genes and Genomes 14(2): 71-81.

Jain R, Joshi A and Rajpurohit D (2017) ISSR based molecular characterization of soybean (*Glycine max* (L.) Merrill) genotypes. Bulletin of Environment, Pharmacology and Life Sciences 7(1): 46-54.

List GR (2015) Polar Lipids. In Moghis UA and Xuebing X (eds) Soybean lecithin: food, Industrial Uses and Other Applications p. 1-3.

Narvel JM, Fehr WR, Chu W, and Grant D (2000) Simple sequence repeat diversity among soybean plant introductions and elite genotypes. Crop Science 40(5).

Pratap A, Gupta S, Jitendra K, and Ramesh S (2012). Soybean. In Gupta SK (eds.) Technological Innovations in Major World Oil Crops p. 293-231.

Singh RJ (2017) Botany and cytogenetics of soybean. USDA-Agricultural Research Service. Soybean/Maize Germplasm, Pathology, and Genetics Research Unit, Department of Crop Sciences, University of Illinois, p 11-17.

Tamura K, Stecher G, Peterson D, Filipski A, and Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution 30(12): 2725-2729.

Tantasawat P, Trongchuen J, Prajongjai T, Jenweerawat S, and Chaowiset W (2011) SSR analysis of soybean (Glycine max (L.) Merril) genetic relationship and variety identification in Thailand. Australian Journal of Crop Science 5: 283–290.

Zhang G, Xu S, Mao W, Hu Q, and Gong Y (2013) Determination of the genetic diversity of vegetable soybean (*Glycine max* (L.) Merril) using EST-SSR markers. Journal of Zheijang University of Science B 14: 279-288.

Table 1: SSR primers and sequences

Primer code	Forward sequence	Reverse sequence
satt123	TTC AAC CAT TCA AAC ATG	AAT TGA AAA GAT GTG AAA CAT
satt135	CAC GGA TTT TAA ATC ATT ATT ACA T	TTC CAA TAC CTC CCA ACT AAC
satt294	GCG GGT CAA ATG CAA ATT ATT TTT	GCG CTC AGT GTG AAA GTT GTT TCT AT
satt309	GCG CCT TCA AAT TGG CGT CTT	GCG CCT TAA ATA AAA CCC GAA ACT
satt581	CCA AAG CTG AGC AGC TGA TAA CT	CCC TCA CTC CTA GAT TAT TTG TTG T
satt609	CCT ACG CAA GTA ACA TTG GTT GTC AT	GCG GCT AAA ACG TAT TAA ATT AAG A

Table 2: Frequency of major allele, allele number, availability, gene diversity, and PIC of SSR primers in soybean

Marker	Major. Allele Freq.	Allele No	Gene Diversity	PIC
Satt123	0.7	3	0.46	0.4102
Satt135	0.4	3	0.66	0.5862
Satt294	0.4	3	0.66	0.5862
Satt309	0.9	2	0.18	0.1638
Satt581	0.4	4	0.66	0.5958
Satt601	0.7	2	0.42	0.3318
Mean	0.58	2.8	0.51	0.45

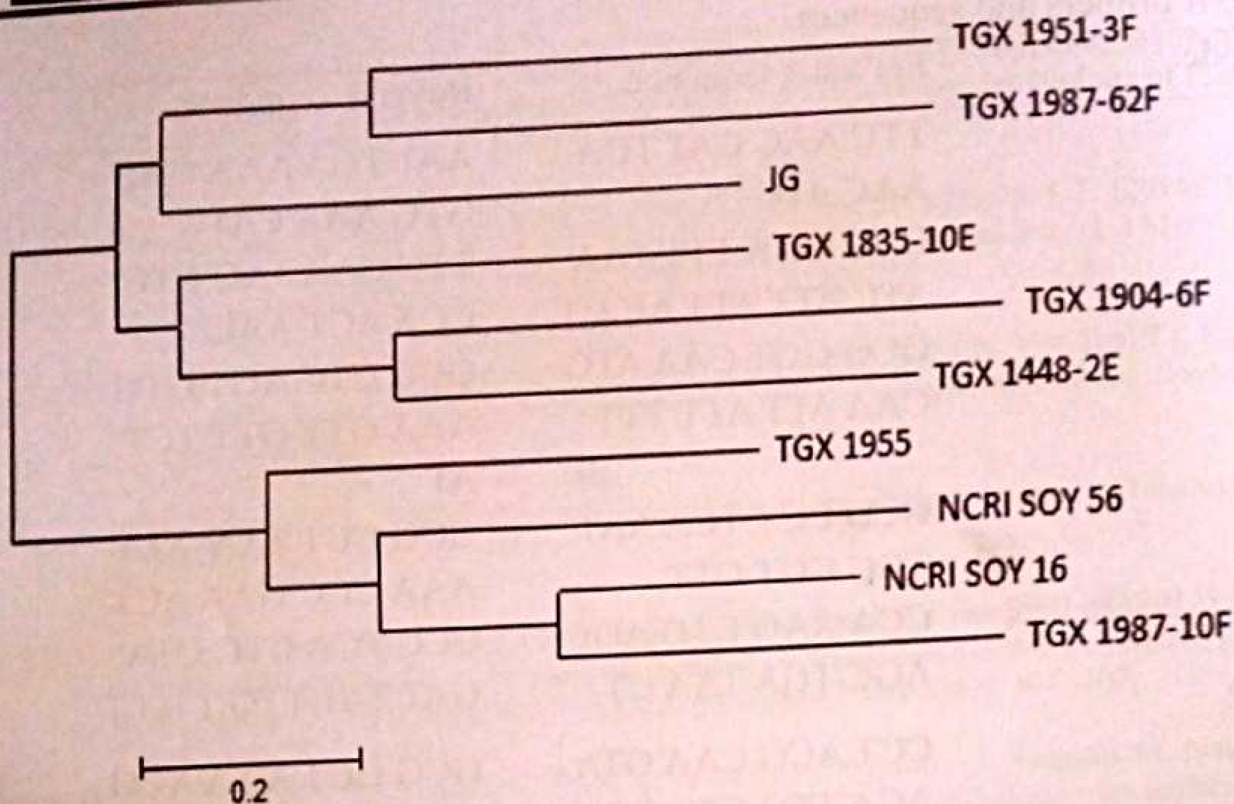

Figure 1: UPGMA dendrogram of soybean genotypes based on SSR markers