USE OF MORPHOLOGICAL AND MOLECULAR PROFILING FOR THE ESTABLISHMENT OF RELATEDNESS BETWEEN COMMERCIALLY AVAILABLE AND ELITE SOYBEAN GENOTYPES

Theophilus Achina [1] Peter Twumasi[2] Marian D. Quain[3]
[1,2]Department of Biochemistry and Biotechnology, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. [3]CSIR-Crops Research Institute, Fumesua-Ghana
[1]achinatbiotech@gmail.com, [2]twumasipeter@gmail.com, [3]marianquain@hotmail.com

ABSTRACT

Molecular profiling is a reliable tool for establishing distinctive differences between plant cultivars or lines and authenticating genomic data vital for tracking germplasm and breeding purposes. The objective of fingerprinting the soybean accessions was to characterize 11 soybean genotypes from the CSIR-Crop Research Institute Soybean breeding program for morphological traits characterization and to reveal genetic diversity using Simple sequence repeat (SSR) markers. Genomic DNA was isolated from young tender leaves and PCR-based techniques were used for SSR molecular analysis. A set of 25 SSR markers were used to determine the genetic diversity and relatedness among 11 soybean (*Glycine max* (L) Merr.) genotypes (5 released varieties and 6 elite lines) in Ghana. The cultivars under study were grown on the field till maturity to assemble quantitative and qualitative morphological data. Morphological description was performed with 10 qualitative and 11 quantitative traits. Cluster analysis based on qualitative morphological characters revealed clear separation of the genotypes on the basis of their seed coat colour. The polymorphic information content (PIC) among genotypes varied from 0.541 (Satt 171) to 0.023 (Satt 160, Satt 148 and Satt 285) with an average of 0.142. In addition, only one polymorphic SSR marker (Satt 171) was able to uniquely identify all 11 varieties and would be useful for DNA fingerprinting. The results suggest that SSR markers were found to be superior in measuring relatedness and identifying varieties of soybean.

Keywords: DNA fingerprinting, genetic diversity, Simple Sequence Repeats, polymorphic information content, Genomic DNA.

INTRODUCTION

It is on record that soybean, G*lycine max* (L.) Merrill, is a significant economic leguminous crop in terms of overall cultivation and global trade (Golbitz, 1995). It is known to have originated from China (Hymowitz and Newell, 1981). According to Hymowitz (1970), the first domestication of soybean in the eleventh century B.C was traced to northeastern China. It is said that the exchange of soybean seeds through human activities began from the centre of origin to Japan, Korea and different parts of South-Eastern Asia (Abe *et al.*, 2003). In 2014, international generation of soybean was just about 315.4 million metric tons (www.soystats.com/worldwide world-soybean-pdtn/), with the United States leading with 108.0 million metric tons. Soybean has a lot of therapeutic constituents such as vitamin C, K, D and folic acids which are antioxidants, vitamin B complex, nicotinic acid, thiamine, riboflavin, biotin and isoflavones such as genistein and daidzein, aside its quality protein and oil (Mathur, 2004). As a result of its numerous benefits, soybean crop is suitably termed as Golden Bean or Miracle crop of the 20th century (Mathur, 2004). Soybean crop has the ability to thrive in large environments apart from fixing atmospheric nitrogen to reduce

erosion of the soil, subdue wild plants and to fit inter and consecutive cropping arrangement. In terms of fodder, soybean likewise seems to be a major plant as it offers abundant protein, energy and fibre (Tran and Nyuyen, 2009). Among the grain legumes in Ghana, soybean ranks third after groundnut and cowpea in terms of production and utilization (Asafo-Adjei *et al.*, 2005). Soybean research and production in Ghana are quite recent and confronted with a lot of difficulties that cause incessant decline in cultivation. These constraints consist of inadequate funding, seed viability, drought stress, pests and diseases and pod shattering among others (Wuni, 2011). Narrow genetic base is also one of the major problems with consequent effect on the above mentioned factors which eventually negatively affect the crop's breeding and development in Ghana. The prospects of breeding soybean are wide on the grounds that a small part of the current varieties in germplasm accumulations currently add to the genetic source of the existing accessions (Chung and Singh, 2008). The broadening of the soybean genetic source may give clue to the influx of novel alleles to polygenic characters (Guzman *et al.*, 2007). The choice of cultivars to be integrated in a breeding system must embrace cultivars that are extremely unrelated to selected genotypes to offer new alleles to the traits of importance. Introducing new soybean lines can increase genetic diversity and facilitate the development of new varieties that can address some of the above mentioned constraints in soybean production. Thus, there is the need to assemble novel germplasm from different geographical locations to access their genetic diversity. This will lead to the selection of parental lines with desirable traits for hybridization to release new varieties that meet the breeder, industrial and consumer needs. Information on genetic diversity is vital to ease the entry of genes of interest into cultivars of commercial value, selection of parental lines for new crosses to be initiated and also for conservation of germplasm by plant breeders (Tatineni *et al*., 1996). Within the Sub Saharan African region, soybean genetic diversity has been traditionally evaluated by the dissimilarities in morphological traits and pedigree information (Ojo *et al*., 2012). However, these methods are highly influenced by certain factors from the environment, growing stages of the plant and also the type of plant material. For plants with a slim genetic source including soybean, DNA-based markers can provide a more accurate indication of genetic diversity within and among species (Acquaah, 2007). Molecular markers have the capacity to yield exceptional DNA profiles in several plants and marker-assisted selection, cultivar identification, and genomic studies are some of the noteworthy applications of SSR markers which have been used extensively in the genetic diversity studies of soybean genetic resource collections worldwide (Tantasawat *et al*., 2011). Aside the conventional breeding method, less work has been done to characterize the existing soybean germplasm to improve upon its usefulness in Ghana (Wuni, 2011). Studies evaluating the genetic diversity of soybean genetic resources using DNA markers especially SSR markers indicate that less cultivars have added to a greater number of genes in the existing accessions (Rana *et al*., 2012). The limited heritable variation in soybean therefore calls for the need to widen the hereditary base of the existing germplasm to accelerate soybean cultivation and development in Ghana. This work aims to: (i) evaluate the morphological make-up of eleven soybean accessions (ii) generate and analyse DNA fingerprints for eleven soybean accessions using SSR primers and phylogenetic analysis and (iii) determine the genetic profiles of the accessions for varietal identification and establish a dendrogram.

MATERIALS AND METHODS

Land preparation and plant material collection

The land was cleared, ploughed and harrowed and the planting materials were collected from the soybean breeding program of CSIR-Crops Research Institute (CRI) in Kumasi, Ghana. The accessions were planted on 12th October 2015 on an experimental field of 75m^2 area.

Each accession was planted on one row plot 3 m long with a spacing distance of 0.5 m within rows and 0.6 m between rows. Randomised Complete Block Design (RCBD) with three replications was used for the field work. Three seeds were sown per stand and thinned to one plant three weeks after emergence.

Planting material

A total of eleven (11) soybean accessions obtained from the CSIR-CRI were used in this study. Out of these, five (5) accessions namely Anidaso, Nangbaar, Jenguma, Salentuya I and Quarshie were released varieties from the National research programs (Table 1). The origin of the remaining elite varieties which are TGX 1935-10E, TGX 1987-20F, TGX 1987-62F, TGX 1990-15F, TGX 1990-3F and TGX 1903-1F is the International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria. Three seeds from each accession were sampled and sown on the CRI Research field to get young leaves from the seedlings for DNA extraction. The detailed information regarding eleven soybean accessions, their pedigree, maturity period, year of introduction and source of collection are given in Table 1.

Table 1: List of 11 Soybean Accessions Used for Fingerprinting and Drought Tolerance Studies

S/No.	Accession Name	Pedigree/Line	Maturity Period	Year of Release	Source of Collection
1	Anidaso	-	105-115	1992	CRI, Fomesua
2	Nangbaar	TGX 1830-20E	110-115	2005	CRI, Fomesua
3	Jenguma	TGX 824-18D x TGX 814-27D	110-115	2003	SARI, Tamale
4	Salentuya I	Pure line	115	1985	SARI, Tamale
5	Quarshie	TGX 824-18D x TGX 814-27D	110-115	2003	SARI, Tamale
6	TGX 1935-10E	TGX 1935-10E		2013	IITA, Nigeria
7	TGX 1987-20F	TGX 1987-20F	-	2012	IITA, Nigeria
8	TGX 1987-62F	TGX 1987-62F	-	2013	IITA, Nigeria
9	TGX 1990-15F	TGX 1990-15F	-	2012	IITA, Nigeria
10	TGX 1990-3F	TGX 1990-3F	-	2012	IITA, Nigeria
11	TGX 1903-1F	TGX 1903-1F	-	2013	IITA, Nigeria

Morphological characterisation

Morphological data were detailed on three plants per accession from the total of three different plots. All observations were made in triplicates. Data were recorded on ten (10) qualitative and eleven (11) quantitative characters of soybean in each of the accessions at various stages of development.

Qualitative data collection

The qualitative traits which were described according to the International Union for the Protection of New Varieties of Plants (UPOV) descriptor (UPOV, 1998) were flower colour, pod colour, seed coat colour, seed shape, seed size, hilum colour, leaflet size, leaflet shape, plant growth habit and plant growth type.

Quantitative data collection

The quantitative traits include: days of germination, 50% day to flowering, days to maturity, number of pods per plant, pod length per plant, plant height (cm), pod yield per plant, number of seeds per plant, weight of seeds per plant, seeds per pod and 100-seed weight. The number of days to germination of seeds of each accession is the number of days taken for 50% -100% of seeds to sprout and emerge as seedlings. Flowering date was determined by counting the

number of days to the date that 50% of plants flowered after germination. Days to maturity is the date 90% of the pods turned brown colour and shattered easily. Mean number of pods per plant was counted using three randomly selected plants in each stand of eleven accessions per plot. Mean pod length (cm) per genotype was determined by taking length of three randomly selected pods of each plant in triplicates in each plot from the petiole to the pod tip at maturity. Similarly, mean plant height (cm) was measured at the field after the plant maturation (R7 and R8) stages. The height of all plants were measured in centimetres from just above the level of the soil to the tip of the plant. A ruler (40cm) was used to measure the height of the soybean plants. Mean number of pods per plant were counted from three randomly selected plants at harvesting stage. Mean number of seeds per plant were also counted as well as weight of seeds per plant using the analytical balance (Model: aeADAM AEP-6000G). A hundred (100)-seed weight well dried were selected at random from each plant per plot and measured with the analytical balance and the mean weight computed.

Molecular characterization Chemicals

The reagents and consumables used in this study were obtained from New England Biolabs Company, Zymo Research company or Seegene company. The solutions used in experiments were all prepared with Filtered Autoclaved Deionised Water (FADW).

Genomic DNA extraction

Total genomic DNA (gDNA) was isolated from samples using the DNeasy plant Mini kit (Qiagen, Hilden Germany) according to the manufacturer's instructions. One hundred milligram (100 mg) of young leaves were collected with sterile forceps and crushed in mortar with pestle using preheated AP1 buffer. The lysates were transferred to a 2.0 ml Eppendorf tubes kept on ice until returned to the laboratory for DNA extraction.

Genomic DNA Extraction using Dneasy plant mini kit

A total of 400 µl AP1 buffer and 4 µl of RNase were added to the lysates in the Eppendorf tubes with vigorous mixing. The samples were incubated for 10 minutes at 65^{o}C with frequent inversions. This was followed by the breakdown of the cell membranes to release the DNA into the extraction buffer, done by adding 130 µl AP2 buffer and incubating on ice for 5 minutes. After incubation, the lysate was spun at 3300 rpm (Genfuge 24D, Progen-UK) for 5 minutes at room temperature.

The supernatant was carefully pipetted and transferred into new 1.5 ml Eppendorf tubes and centrifuged at 3300 rpm for 5 minutes. The supernatant was then carefully transferred again into new 1.5 ml Eppendorf tubes without disturbing the cell debris pellet. Buffer AP3/E (1.5 volumes) was added to the lysate and gently mixed by inversion. After mixing, the samples were incubated in a water bath at -20^{o}C for 10 minutes. The samples were centrifuged at 8000 rpm for 5 minutes and the supernatant discarded carefully to avoid disturbing the DNA pellet. About 500 µl of buffer AW followed by 80% ethanol was used to spun the samples at 8000 rpm for 5 minutes. The last two steps above were repeated and the pellets were dried at room temperature for 30 minutes. 50 µl of buffer AE was added to dissolve the DNA pellet.

Quantification of genomic DNA

Genomic DNA isolated was visualized on 0.8% (w/v) agarose gel. The agarose gel (0.8%) was prepared with Ethidium bromide (7.5 µl in a volume of 200ml). Electrophoresis was carried out at a constant 100 volts for about 45 minutes. The gel was then captured on a computer screen using a transilluminator imaging system (DNR Alpha Bio-Imaging systems, Innotech cooperation, Israel) and the image was saved. Further quantification of DNA concentration and purity was done using Nanodrop^{TM} 2000 spectrophotometer (2009 Thermo

Fisher Scientific Inc.) 2 µl of genomic DNA was diluted to 1ml deionised water. The absorbance for all eleven soybean genotypes was measured at 260/280 nm.

Screening and testing of SSR primers

A total of thirty-two (32) pairs of SSR primers/markers distributed across the integrated linkage map of soybean (Cregan *et al*., 1999) were used in this study. They were supplied by Inqaba Biotec Industries Ltd, Pretoria, South Africa. The primers were tested using genomic DNA samples from eleven genotypes of soybean. The primers which gave very good amplicons were selected for the study. The primer information are available in SoyBase, the USDA-ARS Soybean Genome Database (http//:soybase.org) and published by Song *et al*. (2010).

PCR amplification of extracted DNA

PCR amplification was done with 10 µl reaction mixture for each of the twenty primer pairs, containing 1 µl of 50 ng/µl of soybean genomic DNA, 0.25 µl of each 5 µM primer (F/R), 2.5 µl nuclease free water and 6 µl of one Taq quick load 2x mMw/standard buffer (Biolabs reagents). PCR reactions were carried out in SeeAmp^{TM} thermocycler (Seegene PCR system 9600), Hangzhou Bioer Technology Co. Ltd, China. Thermocycler conditions for the SSR primers were programmed for an initial denaturation of 5mins at 94^{o}C, followed by 35 cycles of 1min at 94^{o}C, 1 min at 50^{o}C and 1 min at 72^{o}C. Final extension was 10 mins at 72^{o}C, and the product was stored at 20^{o}C for electrophoresis. The same cycling conditions were used for all the primer pairs.

Polyacrylamide gel electrophoresis (PAGE)

The mini protean gel cassette unit (Bio-Rad Laboratories) was used for SSR profiling by PAGE. Acrylamide gel (6%) was prepared by mixing 25.865 ml of filtered autoclaved deionised water (FADW), 3.5 ml of 10x Tris-borate-EDTA (TBE), 5.25 ml of 40% acrylamide, 350 µl of 10% freshly prepared ammonium per sulphate (APS) and 35 µl of Tetramethylethylenediamine (TEMED) together and a final volume of 35 µl. Six microlitres (6 µl) of the PCR product were loaded into the sample wells. 100 bp DNA ladder was included to facilitate estimation of band size. Following electrophoresis, the gel was washed with water and subjected to ethidium bromide (0.0075%) staining in 800ml of FADW for 15 minutes. The bands were visualized, captured and scored for analyses using the AlphaImager HP (2011) software version 3.4.0, (Protein Simple, Santa Clara, CA). The scores were recorded in the form of a matrix with '1' and '0' which signifies the presence and absence of bands in each sample respectively. A dendrogram was constructed from the binary data scored using PowerMarker version 3.25.

RESULTS

Qualitative and quantitative traits variability

Qualitative morphological traits evaluation for the 11 soybean cultivars was conducted using UPOV (1998) chart. A summary of the analysis is presented in table 2. The genotypes exhibited genetic variation for all the traits studied. Ten accessions (91%) exhibited purple flower colouration while one (9%) was yellowish. The genotypes also exhibited differences in seed coat colour. Four accessions (36.4%) had cream colour, two accessions (18.2%) had brown colour, four (36.4%) accessions had light brown colour and one accession (9.1%) had yellow colour. Lateral leaflets were in three different sizes: small, medium and large. Two accessions (18.2%) possessed small lateral leaflets, four (36.4%) had medium size leaflet while five (45.4%) possessed leaflet with large size. The shape of lateral leaves were put into two groups namely broad and intermediate. Five accessions (45.4%) were broad while six

(54.5%) were intermediate. Plant growth type were either determinate (seven genotypes (63.63%) or indeterminate (four genotypes (36.4%). There were no semi-determinate growth type among the accessions. Different variations were observed among the accessions regarding their seed size: three accessions (27.3%) had small seed size, four accessions (36.4%) had medium size while the rest (36.4%) had large seed size.

Hilum of the accessions had varying colours which were light brown, brown and imperfect black (black with buff outer ring). Four genotypes (36.4%) possessed light brown colour, six genotypes (54.5%) were of brown colour and one genotype (9.1%) of imperfect black colour. The colour of the pod after maturity were found to be brown (five accessions) and light brown (six accessions). The plant growth habit varied between spreading, erect and semi erect. Three accessions (27.3%) were spreading; seven (63.6%) were erect and one (9.1%) was semi-erect. Shape of the seeds varied among two forms; oval and round. Three accessions (27.3%) had oval seeds while the eight other accessions (72.7%) had round seed. The genetic similarity of the eleven soybean accessions was analysed using Excel Software to determine the Shannon diversity index which gives a measure of phenotypic variation of each character trait (Table 1). Seed coat colour had the highest variability (1.264) among all qualitative traits followed by seed size (1.090) and hilum colour (0.916). Flower colour had the lowest variation (0.218); where 90.9% accessions were of the purple colour.

Table 2: Frequency, %age and Shannon Index of Nine qualitative Traits Studied in Eleven Genotypes of Soybean (*Glycine max*)

S/N	Trait	Category	Frequency	%age	Shannon Index
1	Flower Colour	Purple	10	90.90	0.218
		Yellowish	1	9.09	
2	Seed Coat Colour	Cream	4	36.36	1.264
		Brown	2	18.18	
		Light Brown	4	36.36	
		yellow	1	9.09	
3	Seed Size	Small	3	27.27	1.090
		Medium	4	36.36	
		Large	4	36.36	
4	Hilum Colour	Light brown	4	36.36	0.916
		Brown	6	54.54	
		Imperfect Black	1	9.09	
5	Leaflet Size	Small	2	18.18	1.040
		Medium	4	36.36	
		Large	5	45.45	
6	Pod Colour	Brown	5	45.45	0.689
		Light Brown	6	54.54	
7	Seed Shape	Oval	3	27.27	0.586
		Round	8	72.72	
8	Leaflet Shape	Broad	5	45.45	0.689
		Intermediate	6	54.54	
9	Plant Growth Type	Determinate	7	63.63	0.689
		Indeterminate	4	36.36	
10.	Plant Growth Habit	Spreading	3	27.27	0.860
		Erect	7	63.63	
		Semi-erect	1	9.09	

Table 3: Mean Performance of Released Soybean Accessions Sown Over the Period Oct – Jan. 2016

Trait	Accessions											
	Anidaso	Nangbaar	Jenguma	Salentuya I	Quarshie	Tgx 10E	Tgx 20F	Tgx 62F	Tgx 15F	Tgx 3F	Tgx 1F	Mean
50% days to flowering	44	42	43	42	42	44	44	43	44	42	41	42.82
Days to maturity	91	87	88	88	101	84	90	91	87	88	94	89.91
Number of pods/plant	61	77	35	49	73	33	53	33	60	37	65	52.36
Pod length/plant	3.3	3.3	11.6	11.9	11.7	3.8	3.2	3.6	3.6	3.7	3.7	5.74
Plant height(cm)	32.5	31.4	29.3	31.6	35.7	35.8	33.0	32.5	34.3	37.3	50.6	34.91
Pod yield per plant	64	80	37	53	75	36	55	35	62	39	68	54.91
No. of seeds/plant	57	67	66	103	90	84	56	85	98	96	100	82
Weight of seeds(g)/plant	5.6	6.4	6.4	10.8	9.8	9.7	7.2	12.8	14.8	11.8	10	9.57
Seeds/pod	2	2	2	2	2	2	2	2	2	2	2	2.09
100-seed weight (g)	8.6	8.4	8.7	11.0	10.6	11.4	12.4	15.2	14.6	12.3	10	11.20

Table 3 represents the mean performance of 11 soybean genotypes sown during the period of the experiment. The lowest and highest value for 50% day to flowering was 41 (TGX 1903-1F) and 44 (Anidaso, TGX 1987-20F, TGX 1990-15F and TGX 1935-10E), respectively with a mean value of 42.82 (Table 3). Considerable extent of variation was noticed four days to 50% flowering. With regards to number of pods per plant, a wide range of variation was noticed among accessions. The range was from 36 pods per plant (TGX 1935-10E) to 75 pods per plant (Quarshie) with a mean of 52.36. Days to maturity is a significant characteristic that determine whether an accession can be cultivated with success. This trait showed a range of 84 days (TGX 1935-10E) to 101 days (Quarshie) with a mean value of 89.91. The highest 100 seed-weight of seeds per plant (103 g) was observed in the genotype Salentuya I followed by genotype TGX 1903-1F (100 g). Minimum weight of seeds per plant (57 g) was recorded in the genotype Anidaso with the mean value of 82 g. The lowest plant height (29.3cm) was noticed in the accession Jenguma followed by Nangbaar (31.4cm) and the highest plant height (50.6cm) was noted in the accession TGX 1987-62F. A mean value of 34.91 was observed from the genotypes. For 100-seed weight data, the genotype TGX 1987-62F recorded the maximum value of 15.2g and the genotype Nangbaar had a minimum value of 8.4g, with a mean of 11.2. There was no significant difference ($P>0.05$) among the eleven soybean quantitative traits, with the exception of plant height (p-value = 0.001) and 100-seed weight (p-value = 0.002) which showed some level of significance. Figure 1 below illustrates a graph showing the number of days it took for the seeds of the accessions to emerge from the soil after sowing.

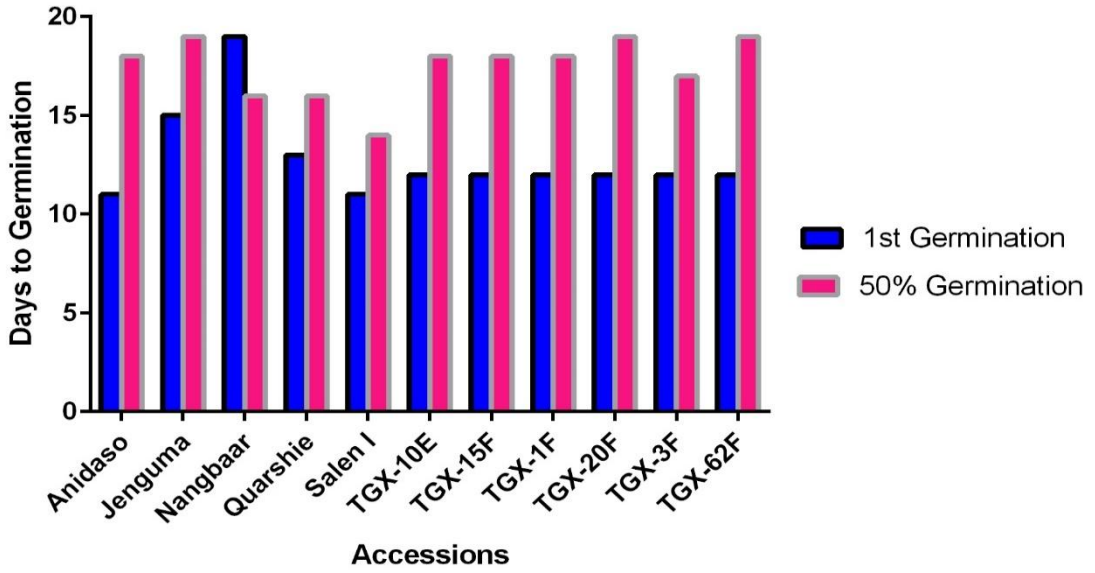


Figure 1: Days to Germination among 11 Soybean Genotypes

Salentuya I seeds emerged after 11 days of sowing and can be said to be an early germinating accession. On the other hand, Nangbaar seeds emerged 16 days after sowing and can be seen as a late germinating accession. Fifty percent (50%) of seeds sown for the genotype Salentuya I germinated after 14 days but Jenguma, TGX 1987-20F and TGX 1987-62F took 19 days for half of the seeds sown to emerge from the soil.

Cluster analysis based on dendrogram

Figure 2 is a dendrogram of the based morphological data UPGMA for the 11 accessions. The dendrogram disclosed that on the basis of morphological characteristics, the genotypes are related on a scale of 0 – 0.4. There were two major clusters; (I and II). The Cluster I consisted of three of the five released varieties namely Salentuya, Anidaso, Quarshie and TGX 1903-1F. However, cluster II had four elite lines and one released line (Nangbaar), with the genotypes Jenguma and TGX1990-15F in separate sub-clusters.

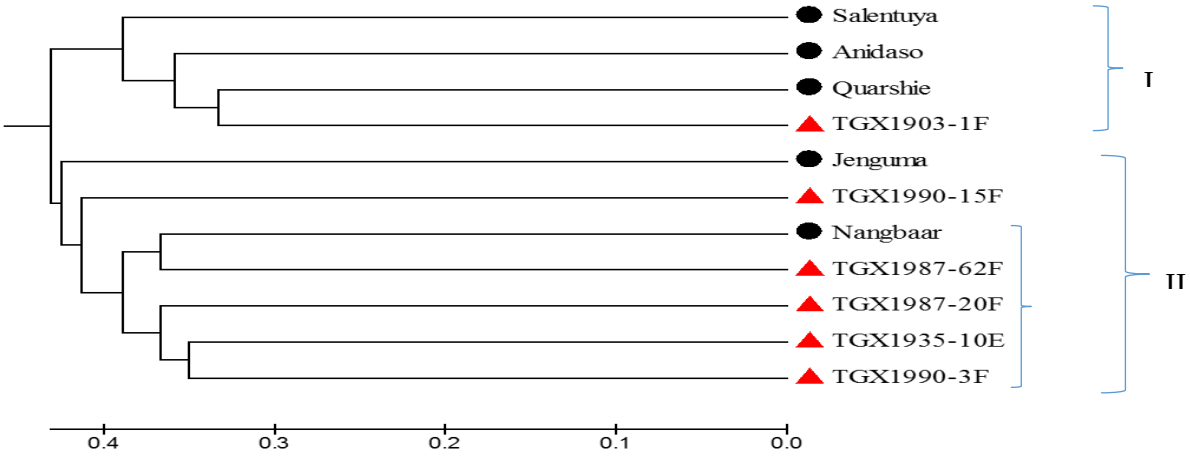


Figure 2: Dendrogram based on morphological traits of 11 Soybean Genotypes

Principal component analysis

Principal component analysis was performed using all the ten quantitative traits. The values of the Eigen vectors and their contribution to variation are given in table 4.

Table 4: Principal Components (PCs) for 10 Quantitative Traits in 11 Genotypes of Soybean (*Glycine max*)

S/N	Trait	PC1	PC2	PC3
	Eigenvalue	993.58	779.07	239.50
	Proportion	0.475	0.372	0.114
	Cumulative	0.475	0.847	0.962
1	Days to 50% Flowering	-0.033	0.044	-0.273
2	Days to Maturity	0.166	-0.117	0.928
3	Number of Pods per Plant	0.585	-0.366	-0.145
4	Pod Length	0.003	0.005	0.008
5	Plant Height	0.053	0.040	-0.150
6	Pod Yield per Plant	0.586	-0.368	-0.144
7	Number of Seeds per Plant	0.529	0.836	0.026
8	Weight of Seeds per Plant	0.054	0.113	-0.003
9	Seeds per Pod	-0.001	0.002	0.003
10	100-Seed Weight (g)	-0.001	0.042	-0.017

The first principal component (PC1) accounted for 47.5% of the total variance and had high contributing loadings from number of pods per plant, pod yield per plant and number of seeds per plant. The second principal component (PC2) was strongly positively associated with number of seeds per plant, pod yield per plant and number of pods per plant and contributed to 37.2% of the total variation. The principal component analysis data revealed that two major principal components contributed 84.7% of total variation among the 11 accessions, accounting for 47.5% and 37.2% respectively (Table 4).

SSR and redox marker analysis for polymorphism

Among 20 SSR primers evaluated in this study, 13 (52%) showed distinct DNA profiles and were used for evaluation of genetic relatedness among the genotypes. All the redox primers used showed clear DNA profiles. Eleven SSR and 1 redox marker loci were polymorphic (table 5), and the number of alleles per locus varied from one to five with an average of 2.1 alleles per locus (table 5).

Table 5: Band Size, Number of Polymorphic Alleles and Allele Size of SSR and Redox markers Used in DNA Profiling

Primer	AT(^{0}C)	Band Size (bp)	Number of Alleles	Polymorphic Alleles	Observed Allele Size (bp)
Satt 001	50	103-145	2	2	80, 107
Satt 171	58	247-277	5	3	190, 201, 240, 161,173
Satt 173	58	200-288	2	2	270, 310
Satt 177	50	98-200	2	2	112, 126
Satt 160	58	251	2	2	262, 291
Satt 148	58	150-171	1	-	174
Satt 534	50	188	3	3	180, 195, 220

Satt 285	58	204-236	1	-	215
Satt 005	58	80-210	4	4	128, 148, 155, 162
Satt 185	58	247	4	4	207, 245, 258, 282
Satt 038	50	155-190	2	2	187, 199
Satt 307	58	150-250	2	2	173, 196
Satt 183	58	293	2	2	244, 283
40S	60	250	1	-	201
ELF Soybean	60	-	1	-	225
DHAR	60	-	2	2	800, 900
Glyma 07g	60	957	1	-	647
Glyma 09g	60	968	1	-	590
Glyma 10g	60	716	1	-	358
Glyma 14g	60	191	1	-	194

Eleven (11) SSR primers (84.6%) were found to be polymorphic and one redox maker (14.2%) was polymorphic (table 5). The genetic diversity which indicate the efficacy of the SSR loci information was quite high, ranging from 0.0000 to 0.9727 with a mean value of 0.6245 (Table 6). 13 SSR and seven redox loci provided 40 genetic profiles of the 11 accessions studied on the average, there were 2 alleles per loci. The total number loci were 20, of which 12 (60%) had alleles which showed complete genetic heterogeneity and 8 (40%) revealed genetic homogeneity (Table 5). The highest range (800-900bp) of allele size was found in DHAR, however the lowest (80-107bp) was recorded in Satt 001 (Table 5)

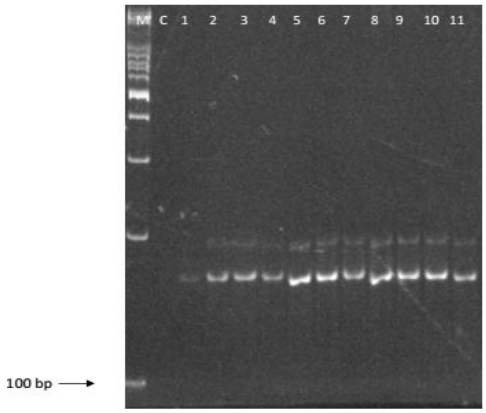
Table 6: Functional SSR and Redox Primers Used to Access the Genetic Diversity among 11 Soybean Accessions

Marker	Allele Frequency	Gene Diversity	PIC
Satt 001	0.8182	0.7858	0.127
Satt 171	0.6546	0.0776	0.541
Satt 173	0.7728	0.7541	0.165
Satt 177	0.8182	0.8153	0.124
Satt 160	0.9546	0.9727	0.023
Satt 148	0.9091	0.9727	0.023
Satt 534	0.8485	0.7541	0.165
Satt 285	0.9091	0.9727	0.023
Satt 005	0.7955	0.3301	0.263
Satt 185	0.7500	0.4438	0.352
Satt 038	0.9091	0.9115	0.064
Satt 307	0.9091	0.9115	0.000
Satt 183	0.7728	0.7541	0.000
40S	1.0000	0.0000	0.000
ELF Soybean	0.9091	0.9727	0.023
DHAR	0.8182	0.8230	0.128
Glyma 07g	0.8637	0.8426	0.101
Glyma 09g	1.0000	0.0000	0.000
Glyma 10g	0.7273	0.3967	0.000

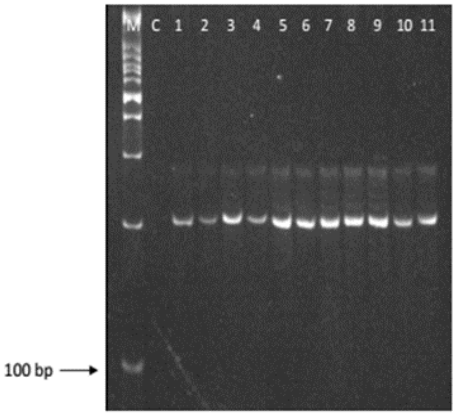
Glyma 14g	1.0000	0.0000	0.000
Mean	0.8570	0.6246	0.1061

The presence of unique bands enables easy identification of specific genotypes in the germplasm. The number of alleles per locus as shown in table 5 varied from 5 (locus Satt 171) to 1 (locus Satt 148, Satt 285, 40S, ELF Soybean, Glyma 10g and Glyma 14g). The gene diversities of these seven loci were 0.0776, 0.9727, 0.9727, 0.0000, 0.9727, 0.3967 and 0.0000 respectively. The mean values obtained for the alleles and the polymorphic information content (PIC) were respectively 0.8570 and 0.1061 with an average gene diversity of 0.6246 per locus. Polymorphic information content (PIC) values, a measure of the allelic diversity of the markers ranged from 0.541 in Satt 171 to 0.023 in Satt 160, Satt 148, Satt 285 and ELF Soybean (table 6). The primers Satt 307, Satt 183, 40S, Glyma 07g, Glyma 10g and Glyma 14g with a PIC of 0.000 were strictly not informative. The locus Satt 171 (PIC = 0.541) is the most informative for distinguishing among the soybean genotypes used in this study. The observed product size ranged from 80 bp in Satt 001 to 900 bp in DHAR. Figure 3 shows four examples of DNA profiles at the Satt 307, Satt 185, DHAR and Satt 038 loci with three different alleles among diverse soybean genotypes. The noticeable mono bands generated by the DHAR marker were identified at about 800 and 900 bp in the genotypes Quarshie, Nangbaar, Anidaso, TGX-1987-20F, TGX 1903-1F, TGX 1987-62F, TGX 1990-3F, TGX 1990-15F and TGX 1935-10E.

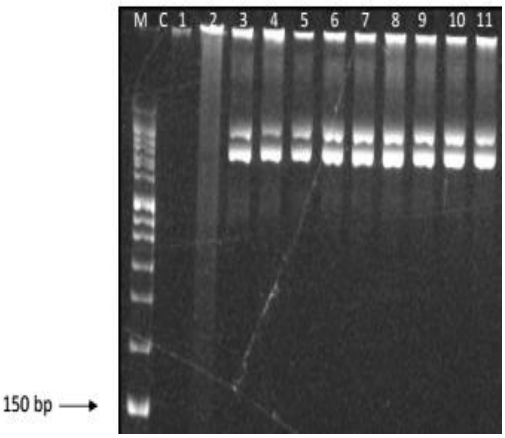
(A)



(B)



(C)



(D)

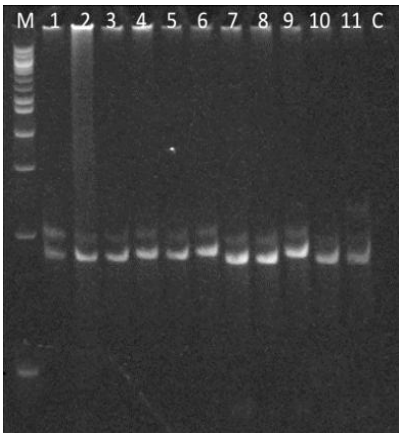


Figure 3: DNA Profiles at Different Loci: (A): Primer Satt 307, (B): Primer Satt 185 and (C): Primer DHAR (D): Satt 038 showing 11 accessions. M denotes the DNA marker, C –control sample and 1-11 refers to the germplasm accessions namely: 1-Salentuya I, 2 Jenguma, 3-Quarshie, 4-Anidaso, 5-Nangbaar, 6-TGX 1987-20F, 7-TGX 1903-1F, 8-TGX 1987-62F, 9-TGX 1990-3F, 10-TGX 1990-15F and 11-TGX 1935-10E.

Genetic relationships among soybean accessions using cluster analysis

Evaluating the genetic relatedness among the genotypes based on the molecular markers, Jaccard's similarity coefficients were calculated for all 40 alleles scored among the 11 accessions. The similarity coefficient varied on a scale of 0.05 to 0.25 **(**Figure 4**)** and the similarity coefficients matrix was used for UPGMA cluster analysis to obtain a dendrogram. All 40 alleles scored were used for the genetic fingerprinting analysis. Similarity matrices of all the 11 genotypes were generated using PowerMaker version 3.25. The genetic similarity coefficients (Table 7) found in the cultivars comparison matrix were relatively low, as all the values were between 0.125-0.450 with an average of 0.267 indicating limited diversity among genotypes.

Table 7: Genetic Similarity Matrix for the Soybean Accessions Studied

Accession	Anid	Jeng	Nang	Quar	Salen	TGX 10E	TGX 15F	TGX 1F	TGX 20F	TGX 3F
Anid										
Jeng	0.275									
Nang	0.200	0.275								
Quar	0.125	0.200	0.175							
Salen	0.325	0.300	0.325	0.300						
TGX 10E	0.250	0.325	0.200	0.225	0.325					
TGX 15F	0.325	0.300	0.275	0.250	0.450	0.225				
TGX 1F	0.225	0.250	0.175	0.200	0.350	0.125	0.300			
TGX 20F	0.175	0.300	0.075	0.150	0.350	0.175	0.300	0.150		
TGX 3F	0.425	0.400	0.275	0.350	0.450	0.275	0.200	0.300	0.300	
TGX 62F	0.325	0.350	0.275	0.300	0.400	0.225	0.200	0.150	0.250	0.250

The highest genetic similarity coefficient (0.450) was found between the genotypes TGX 1990-3F and Salentuya I while the lowest genetic similarity coefficient (0.125) was observed between the genotypes Quarshie and Anidaso as well as TGX 1903-1F and TGX 1935-10E. There were no duplicates among the genotypes. The dendrogram (Figure 4) based on genetic similarities between genotypes showed that the 11 genotypes formed five major clusters.

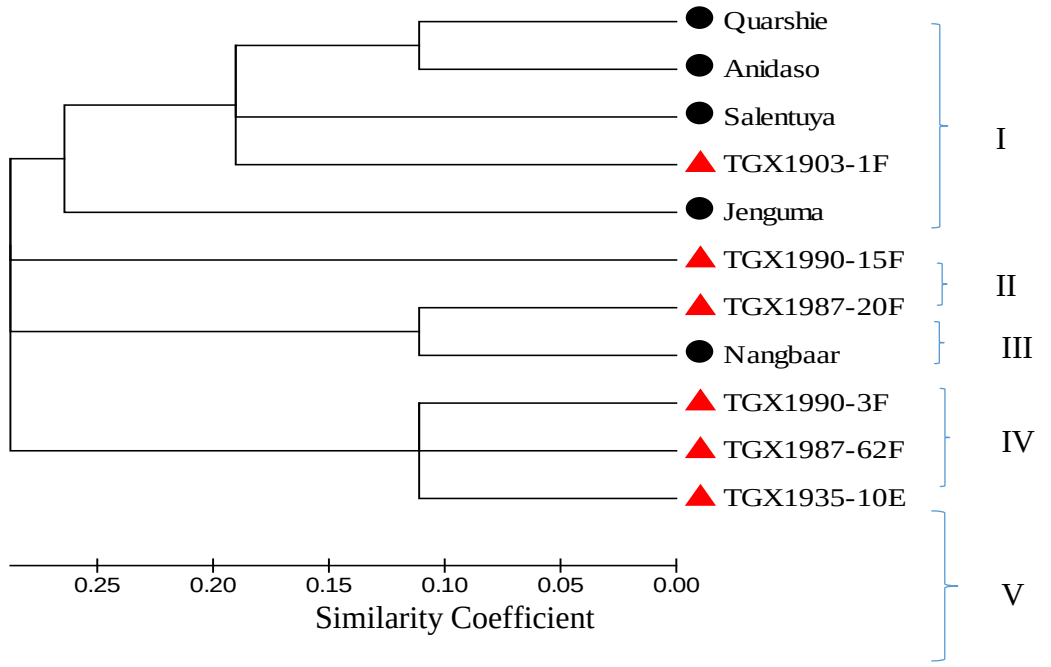


Figure 4: Dendrogram showing Similarity Coefficients and Genetic Relationships among 11 Genotypes of Soybean analysed by SSR and Redox Markers

Cluster I (Figure 4) was the largest and consists of four accessions (three released lines and one elite line). The cluster consisted of the accessions Quarshie, Anidaso, Salentuya I and TGX 1903-1F. Cluster II and III are the smallest having only one released and one elite genotypes which are TGX 1990-15F and Jenguma respectively. Cluster IV consisted of one released and one elite genotypes which included Nangbaar and TGX 1987-20F. Cluster V was highly homogenous and consisted of three elite lines namely TGX 1990-3F, TGX 1987-62F, and TGX 1935-10E.

DISCUSSION

Morphological and molecular data obtained from field and laboratory work were used for the analysis. All the 21 morphological traits showed significant ($P \leq 0.05$) variation signifying the existence of genetic variation among the genotypes for all the traits studied. Considerable variations have also been similarly reported in soybean genotypes by other researchers such as Tavaud-Pirra *et al*., (2009) and Ojo *et al*., (2012) who recorded average Shannon index (0.8041) indicative of medium diversity among the genotypes (Priolli *et al*., 2002). This was in consonance with the Shannon index (0.8041) obtained in this study and that reported by Malik *et al*., (2011).

In this study, all cultivars were found to differ from each other in one or more individual traits without any duplication. The 10 qualitative traits (Table 2) used for characterizations were sufficient to differentiate between the 11 cultivars, suggesting that they may be very useful in cultivar discrimination and identification, despite the weaknesses and limitations of morphological markers reported by some studies (Li *et al*; 2008; Wang *et al*., 2010). Development of new plant type with improved characters is the main focus of plant breeding. In soybean breeding, however, the significant attributes include the number of pods per plant, seeds per pod, and seed weight which determine the seed yield. In this study it was observed

that among the 11 genotypes, 6 performed maximally with respect to number of seeds per plant, weight of seeds per plant and 100-seed weight along with other quantitative traits such as pod yield per plant and number of seeds per pod. These results are in harmony with that of Kundi *et al*., 1997, who reported improvement in yield attributes of soybean genotypes. Among the traits studied, pod yield per plant, weight of seeds per plant, 100-seed weight and flower colour could be used as traits to distinguish the genotypes. Anna Durai (2005) reported that the total number of pods per plant, seed yield per plant, hundred seed weight and number of clusters per plant could be used to differentiate cultivars. The UPGMA procedure defined cluster 2 of the morphological dendrogram (Figure 2) into 3 sub-clusters. The analysis identified interrelated clusters of accessions and many of the elite soybean genotypes were found to be genetically different from the released accessions. For example, cluster II consists mainly of elite genotypes while cluster I consisted of most of the released cultivars, however, one elite genotype was found in cluster I. The morphological dendrogram (Figure 2) formed two major groups with the 11 genotypes whereas the molecular dendrogram (Figure 4) showed a clear separation of the genotypes into five groups. Salentuya I, Anidaso, Quarshie and TGX 1903-1F tended to group together as a single group (cluster I) in both dendrograms. These genotypes could be similar in their origin although they have considerable diversity. The molecular dendrogram is preferentially finer, clearer and sharper in contrast with the morphological dendrogram. However, the clusters were very similar indicating the effectiveness of both morphological and molecular traits in separating cultivars. The phenotypic uniqueness between the released and elite accessions suggests the introduction of new germplasm to the breeding program soybean gene pool, and this has the potential of broadening the local soybean germplasm diversity. Genetic relatedness observed in this study could support soybean breeders to make healthier selections when choosing among large numbers of accessions.

Principal component analysis is a procedure used to highlight variation and bring out strong patterns in a given set of data. It is a suitable method to evaluate germplasm (Falcinelli *et al*., 1988). Ten (10) quantitative traits created three independent principal components, namely PC1, PC2 and PC3 (Table 4). Traits having high positive loadings in PC1 and PC2 were number of pods per plant, pod yield per plant and number of seeds per plant. These results suggests that the genotypes with high PC1 and PC2 values were not only high yielding but also have high 100-seed weight and oil content. This was in agreement with the findings of Toker and Cagirgan (2004) who observed that, number of pods per plant, number of seeds per plant and 100-seed weight were significant contributing traits to the total variation. Three principal traits under PC1 and PC2 with Eigen values < 1 contributed 47.5% and 37.2% of the variation respectively. It can be observed that the first two principal components with Eigen values < 1 were able to explain 84.7% of the total variation (Table 4). This is in consonance with Mardia *et al*. (1979), who posit that the total variance accrued by principal component close to 80% clarifies reasonably the variability established between individuals. The obtained gene diversity which ranged from 0.000 to 0.9727 (table 6) with an average of 0.6246 was in agreement with the data of Diwan and Cregan (1997) who found mean gene diversity values close to 0.69 in a group of 36 commercial soybean lines; and in agreement with the results of Rongwen *et al*. (1995) who found a mean value of 0.74 in a group of 96 soybean genotypes. Four markers generated relatively extensive polymorphism within the soybean cultivars studied. Satt 005 and Satt 185 (Figure 3), and were more powerful in cultivar identification. Both markers generated four distinct banding patterns for all soybean cultivars under study and can be used to discriminate between them. These results are not in line with that of AiMin *et al*. (1998) who reported that "no single primer out of seven could distinguish 16 *Brassica juncea* cultivars through single analysis".

The highest similarity values (0.450 and 0.425) were recorded between the two cultivars TGX 1990-3F vrs Anidaso and TGX 1990-3F vrs Salentuya I respectively. This indicates that these cultivars were the most distantly related genotypes in the collection used in this study as an indice value of "1" indicates the two varieties are duplicates. On the other hand, the lowest value (0.125) was recorded between the cultivars Quarshie vrs Anidaso and TGX 1903-1F vrs TGX 1935-10E, indicating that they were the most closely related cultivars. Polymorphic information content (PIC) was highest for the SSR primer Satt 171 (0.541) and was lowest (0.000) for the redox markers Glyma 09g, Glyma 10g and Glyma 14g (table 6). The higher the PIC value, the more informative is the marker. Hence primer Satt 171 is highly informative in the study and might be an effective and useful tool to determine the genetic differences among the soybean accessions and to study phylogenetic relationships. Twelve primers capable of producing polymorphic bands (60%) were identified with 11 SSR primers and 1 redox marker in this study. Similar results were shown by Baranek *et al*., (2002), who detected 122 highly reproducible bands with 22 RAPD markers out of which 55 of them were polymorphic. The polymorphism detected in this study was moderate because the materials analysed were all from breeding programs that shared the same germplasm, thus having a relatively narrow genetic base. The high rate of SSR polymorphism could be linked to the chosen set of SSR markers which were already examined for polymorphism among the set of genotypes (Cregan *et al*., 1999). Nevertheless, the lower allele number and PIC values signifies low allelic diversity in the soybean genotypes studied.

CONCLUSION

The 11 soybean genotypes evaluated demonstrated morphological and genetic variations among themselves. It was observed that the number of pods per plant, pod yield per plant and 100-seed weight contributed a greater proportion of the variations that were present among the cluster groups. The principal component analysis showed that all the 11 soybean genotypes were ordered into three separate clusters. Salentuya I and TGX 1903-1F were indicated to be the high yielding genotypes whereas the accessions Nangbaar and TGX 1935-10E were the early maturing genotypes. Associating variation at the phenotypic level with molecular level enhanced easy differentiation of the soybean accessions studied. Both morphological and SSR marker data were able to reveal the existence of wide genetic variability among the soybean accessions studied. SSR analysis was more effective in discriminating between the 11 soybean genotypes and revealed thorough identification than perceiving changes at the morphological level. These markers can successfully be employed to produce variety specific fingerprints.

ACKNOWLEDGEMENTS

This work was supported by the West Africa Agricultural Productivity Programme (WAAPP II) and the Biotechnology Division of the Crops Research Institute (CRI)-Ghana. Special thanks goes to Agnes Bosompem and Belinda Akomeah of the CRI Molecular Biology Lab for their support and assistance in the course of my work. Acheampong Samuel of North Carolina State University, USA deserves my appreciation for discussions and critical scrutiny of the manuscript.

REFERENCES

1. Abe, J., Xu, D. H., Susuki Y, Kanazawa, A. and Shimomoto, Y., (2003). Soybean germplasm pools in Asia revealed by SSRs. *Theoretical and Applied Genetics* 106:445-453.
2. Acquaah, G., (2007): Principle of Plant Genetics and Breeding. Blackwell Publishing Ltd., U.S.A., pp. 520–521.
3. AiMin, Q., L. PeiYing, L. Jianjun, H. XueLin, F. JiaRui, Q. AM, P.Y. Liu, J.J. Lei, X.L. Huang and J.R. Fu, (1998): Identification of mustard (*Barassica juncea* Coss.) cultivars with RAPD markers. *Acta Scientiarum Naturalium Universitatis Sunyatseni*, 37: 73–6.
4. Anna Durai, A. (2005). Genetic diversity, diallele analysis and induced mutagenesis in soybean [*Glycine max* (l.) Merrill**]** *M.Sc (Agri) Thesis*. Tamilnadu Agricultural University. Coimbatore.
5. Asafo-Adjei, B., Ansah, I. O. O., Asuboah, R.A., Dapaah, H., Haruna, M. and Oti-Boateng, C. (2005). *Soybean Production Guide*. CSIR-CRI, Kumasi, Ghana.
6. Baranek, M., M. Kadlec, J. Raddova, M. Vachun, M. Pidra (2002): Evaluation of genetic diversity in 19 *Glycine max* (L.) Merr. accessions included in the Czech National Collection of Soybean Genotypes. *Czech J.Genet. Plant Breed*., *38*:69–74.
7. Botstein D., White, R.L., Skolnick, M., and Davis, R.W. (1980): Construction of a genetic linkage map in man using restriction fragment length polymorphisms Am. J. Hum. Genet. 32:314-331.
8. Chung, G.; Singh, R.J. (2008): Broadening the genetic base of soybean: a multidisciplinary approach. *Critical Reviews in Plant Science*, v.27, p.295-341.
9. *Cregan P.B., Jarvik T., Bush A.L., Shoemaker R.C., Lark K.G., Kahler A.L., et al. (1999): An integrated genetic linkage map of the soybean. Crop Sci. 39:1464–90.*
10. Diwan, N. and Cregan, P. B. (1997): Automated sizing of fluorescent labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. *Theoretical and Applied Genetics* 95**:**723-733.
11. Falcinelli, M., F. Veronesi and S. Lorenzetti, (1988): Evaluation of an Italian germplasm collection of *Lolium perenne* L. through a multivariate approach. *Proc. Eucarpia Fodder Crops Section Meeting*, Lusignan, France. p 22-24.
12. Golbitz P (ed) (1995) 95–96 Soya Bluebook Plus. Soyatech, Inc., Bar Harbour
13. Guzman, P.S.; Diers, B.W.; Neece, D.J.; Martin, S.K.; Leroy, A.R.; Grau, C.R.; Hughes, T.J.; Nelson, R.L. (2007): QTL associated with yield in three backcross-derived populations of soybean. *Crop Science*, v.47, p.111-122. http//:soybase.org
14. Hymowitz, T. (1970): On the domestication of the Soybean *Econ Bot*., 23:408-421
15. Hymowitz, T., & Newell, C. A. (1981). Taxonomy of the Genus *Glycine*, Domestication and Uses of Soybeans. *Economic Botany*, *35*(3), 272–288.33.
16. Jaccard, P. (1908) : Nouvelles rescerches sur la distribution florale. Bull. Soc. Vaud. *Sci. Nat.*, 44: 233-270.
17. Kundi, M. S. Gill, T. P. Singh, and P. S. Phul: (1997): "Radiation induced variability for quantitative traits in soybean (*Glycine max* (L.) Merrill)," *Crop Improvement*, vol. 24, pp. 231-234.
18. Malik, F.A., M. Ashraf, A.S. Qureshi, M.R. Khan (2011): Investigation and comparison of some morphological traits of the soybean populations using cluster analysis. Pak. J. *Bot.* 43(2):1249-1255.
19. Mardia, K. V., J. T. Kent and J. M. Bibby, (1979): Multivariate analysis. *Academic press*, London.
20. Mathur, S. (2004): Soybean the Wonder Legume. *Beverage Food World*, 31, 61-62.49.

21. Nei M. (1987): Molecular Evolutionary Genetics. *Columbia University Press*, New York, Volume 75, Issue 3, pp.428–429.
22. Ojo, Ajayi A. O., and Oduwaye O. A., (2012): "Genetic relationships among soybean accessions based on morphological and RAPDs techniques," Pertanika *Journal of Tropical Agricultural Science*, vol. 35, no. 2, pp. 237-248.
23. Priolli, R.H.G.; Mendes-Junior, C.T.; Arantes, N.E.; Contel, E.P.B. (2002): Characterization of Brazilian soybean cultivars using microsatellite markers. *Genetics and Molecular Biology*, v.25, p.185-193.
24. Rana, M.K., Archak, S. and Singh, R. (2012): An E-Manual on "Molecular Marker Techniques for Genotype Identification". *NBPGR, Pusa Campus*, New Delhi-110 012
25. Rongwen J, Akkaya MS, Lavi U and Cregan PB: (1995): The use of
26. microsatellite DNA markers for soybean genotype identification. *Theor Appl Genet* 19:43-48.
27. Song Q., Jia G., Zhu Y., Grant D., Nelson R.T., Hwang E., Hyten D.L., Cregan P.B., (2010): Abundance of SSR motifs and development of candidate polymorphic SSR markers (BARCSOYSSR_1.0) in Soybean crop Sci. 50:1950-1960.
28. Tang, Q.U., (1997): Data Processing System. Chinese Agricultural Press, Beijing.
29. Tantasawat, P., Trongchuen, J., Prajongjai, T., Jenweerawat, S., and Chaowiset, W., (2011): SSR analysis of soybean (*Glycine max* (L.) Merr.) Genetic relationships and variety identification in Thailand. *Australian Journal of Crops Science* 5 (3): 283-290
30. Tatineni, V., Cantrell, R.G., Davis, D.D., (1996): Genetic diversity in elite cotton germplasm determined by morphological characteristics and RAPDs. *Crop Sci.* 36, 186–192.
31. Tavaud-Pirra, P. Sartre, R. Nelson, S. Santoni, N. Texier, and P. Roumet, (2009): "Genetic diversity in a soybean collection," *Crop Science*, vol. 49, no. 3, pp. 895-902.
32. Toker, C. and M.I. Cagirgan (2004): The use of phenotypic correlations and factor analysis in determining characters for grain yield selection in chickpea (*Cicerarietinum* L.). *Hereditas* 140:226-228.
33. Tran LS, Nguyen HT. (2009): Future biotechnology of legumes. In: Emerich WD and Krishnan H, eds. Nitrogen Fixation in Crop Production. Madison, WI: ASA-CSA-SSSA. pp. 265-307.
34. UPOV, TG/80/6 (1998): Guidelines for the conduct of tests for distinctness, uniformity and stability: Soya Bean (*Glycine max* (L) Merr.), *International union for the protection of new varieties of plants.*
35. Wang M, Li R, Yang W, Du W (2010): Assessing the genetic diversity of cultivars and wild soybeans using SSR markers. *African J of Biotechnol* 9:4857-4866.
36. Wuni, M., (2011): Assessment of the constraints in soybean production: a case of Northern Region, Ghana. *J. Dev. Sustainable Agric*. 6:198:214.