

Article

Not peer-reviewed version

---

# Genetic Diversity of Subtropical Soybean [*Glycine max* (L.) Merrill] Genotypes Based on SSR Markers

---

[Hapson Mushoriwa](#)\*, [Isack Mathew](#), Pangirayi Tongoona, [Eastonce T. Gwata](#), [Godfree Chigeza](#), [Ibnou Dieng](#)\*, [Abush A. Tesfaye](#), John Derera

Posted Date: 1 July 2024

doi: 10.20944/preprints202406.1954.v1

Keywords: Diversity; Genotype; Soybean germplasm; SSR markers



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Article

# Genetic Diversity of Subtropical Soybean [*Glycine max* (L.) Merrill] Genotypes Based on SSR Markers

Hapson Mushoriwa <sup>1,\*</sup>, Isaack Mathew <sup>2</sup>, Pangirayi Tongoona <sup>3</sup>, Eastonce T. Gwata <sup>2</sup>, Godfrey Chigeza <sup>4</sup>, Ibnou Dieng <sup>1,\*</sup>, Abush A. Tesfaye <sup>1</sup>, John Derera <sup>5</sup>

<sup>1</sup> IITA, PMB 5320, Oyo Road, Ibadan 20000, Oyo State, Nigeria.

<sup>2</sup> University of Venda, Department of Plant & Soil Sciences, P. Bag X5050, Thohoyandou.

<sup>3</sup> West Africa Centre for Crop Improvement, College of Basic and Applied Sciences, University of Ghana PMB LG 30, Legon, Ghana.

<sup>4</sup> IITA, Southern Africa Research and Administration Hub Campus, P.O. Box 310142, Chelston, Lusaka, Zambia

<sup>5</sup> One CGIAR c/o IITA, PMB 5320, Oyo Road, Ibadan 20000, Oyo State, Nigeria

\* Correspondence: h.mushoriwa@cgiar.org; i.dieng@cgiar.org

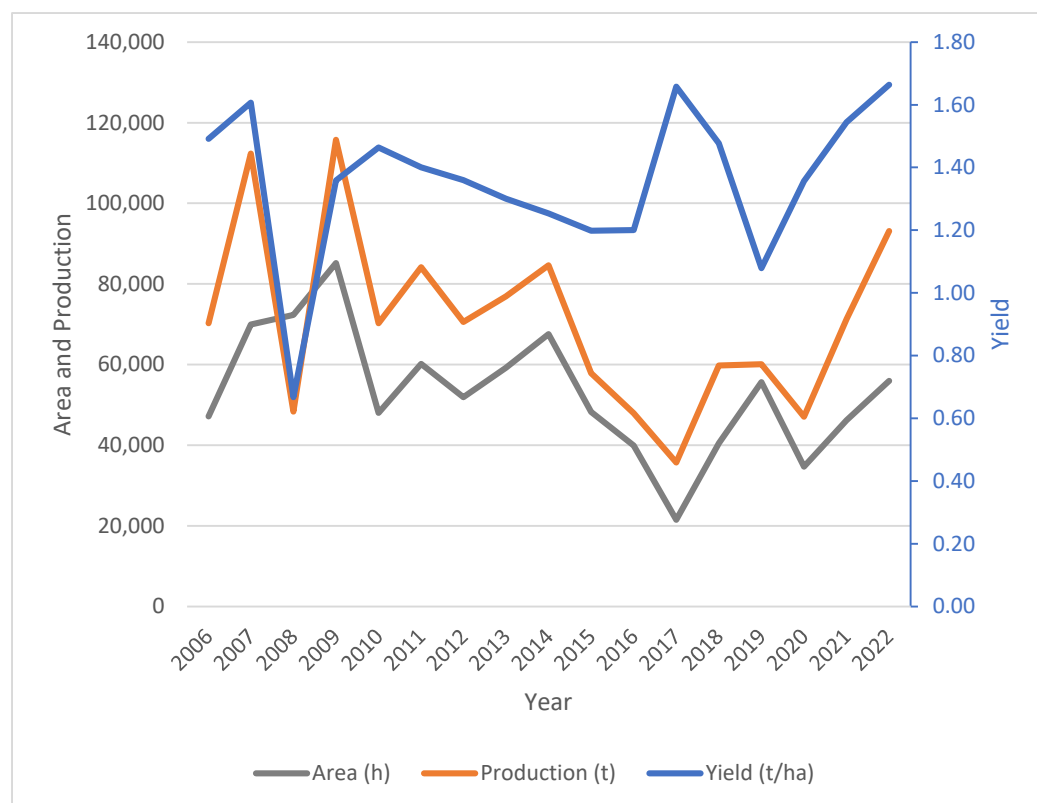
**Abstract:** Knowledge of the genetic diversity of the available germplasm is an important foundation for crop conservation, management, and improvement. Successful breeding strategies rely on a sound understanding of the genetic diversity of the crop germplasm. This study was conducted to evaluate and compare the genetic diversity of 49 soybean [*Glycine max* (L.) Merrill] genotypes using SSR markers as well as to evaluate parental potential of these genotypes in cultivar improvement. The bulk of the genotypes were developed in the national soybean breeding program in Zimbabwe between 1940 and 2013. The genetic diversity and relatedness were estimated using 30 SSR markers and produced 5 clusters that generally matched the pedigree information of the germplasm. The SSR analysis detected a total of 135 alleles. The polymorphic information content ranged from 0.10 to 0.77 with a mean of 0.45. The SSR primer, Satt012 was the most informative. The SSR marker analysis showed a wide genetic diversity in the germplasm. To date, there are no reports regarding the genetic diversity analysis of soybean germplasm in Zimbabwe. The genotypes, EL41 and EL7; EL41 and EL1; as well as EL41 and EL42 were the most divergent thus, indicating their potential use as parental lines in the genetic improvement of soybean in the region.

**Keywords:** Diversity; Genotype; Soybean germplasm; SSR markers

## 1. Introduction

Soybean [*Glycine max* (L.) Merrill] is an important grain legume with an immense economic value in many parts of the world including southern Africa. There has been a dramatic increase in demand due to the growing need for protein rich food and feed [1]. Soybean is fast becoming a major source cash income. Soybean has huge potential to raise and improve rural income [2]. Numerous studies have documented the capacity of soybean to fix nitrogen and holds considerable potential for arresting the declining soil fertility in the smallholder farming systems [2]. Despite the rise in demand, soybean production in Zimbabwe has been characterised by frequent and pronounced fluctuations (Figure 1). Over 2006 to 2022 period, the average annual production was 67 005 metric tonnes (MT) resulting in a production gap of 172 995 MT translating to a shortage of 72.1% soybean commodity. Precisely, Zimbabwe is a net importer of soybean [3]. Figure 1 also showed that the grain yield ranged from 0.67 to 1.66 t/ha with an average yield of 1.28 t/ha, lower than the global average (2.7 t/ha) [4,5]. The major challenge is that there is a wide gap between the on-farm yields and potential yield realised from research fields. The low yields are attributable to poor performing varieties in the smallholder sector, lack of access to certified seed in affordable packs, limited application of fertilisers and rhizobial inoculants in soils with no history of soybean and low adoption of best practices of management such as late planting, low plant populations, recycling seeds etc.

[4,6]. Inefficient and traditional production technologies also resulted in low productivity [2]. Furthermore, there are limited options of high yielding varieties available to farmers impacting yield [7]. Previous studies showed that soybean varieties with appreciable yield performance are needed to sustain production, but their improvement has been slowed down by the low genetic diversity [6,8]. Essentially, soybean is an autogamous crop i.e. self-pollinating and probably correlated with genetic diversity. Soybean production and nutritional values can be addressed by broadening germplasm diversity [8]. Determining the genetic diversity is critical for the development of new varieties with desirable traits or matching farmer and market preferred traits. In the same vein, the genetic diversity can be exploited to mitigate the effects of climate change through the development of climate smart varieties. This will aid the breeders to achieve sustainable improvement in yield and nutritional quality as soybean is poised to dominate crop production due to its multiple benefits [9].



**Figure 1.** Trends in soybean production, area and yields in Zimbabwe, 2006 to 2022. Source: Ministry of Lands, Agriculture, Fisheries, Water and Rural Development, 2023.

In Zimbabwe, the soybean breeding program commenced in 1940 and produced many breeding lines and commercial cultivars over the years [10,11]. Some of the germplasm from the program was introduced successfully into neighboring countries such as Malawi and Zambia and further afield [12]. However, the diversity of the soybean germplasm could be under threat due to the frequent use of the same genetic resources resulting in narrowing of the genetic base [13]. Previous studies involving pedigree analysis revealed that 35 ancestors contributed more than 95% of the alleles and only five lines of the germplasm that was released between 1947 and 1988 in North America accounted for more than 55% of the genetic background of public cultivars [14]. Similarly, in the Chinese germplasm released from 1923, <15.0% of the ancestral cultivars accounted for >50.0% genetic material [15]. Based on SSR markers, a low diversity among soybean genotypes was reported in other countries including Brazil and India [16,17]. These studies indicated that intensive breeding and selection contributed to the reduction in diversity hence the need to introgress new parental stock in the breeding pools.

In general, genetic diversity studies are important for understanding the evolutionary relationships among accessions or identifying diverse parental combinations for creating progenies with maximum genetic variability for further selection [18]. To overcome the effects of environmental factors on quantitative traits in assessing genetic diversity in soybean, molecular markers were applied widely in previous studies [19,20]. In Zimbabwe, there is no adequate information regarding the diversity in the soybean germplasm. Most importantly, there is lack of information at the molecular level for genetic relationships of the Zimbabwe germplasm. The use of DNA markers to evaluate diversity aids in the effective utilization of germplasm for crop productivity improvement and conservation [21]. During the 70-year period of its existence, the breeding program introduced exotic genetic material from several countries including the United States of America and South Africa as well as Nigeria and Taiwan. There are more than 10 000 improved and landrace cultivars in the Gene-banks. The breeding program primarily emphasized the mating of elite-by-elite parental lines that might have resulted in recombining the genes from a limited number of ancestral materials thus narrowing the genetic base of the gene pool [22]. Therefore, this study was designed to examine (i) the genetic diversity among 49 soybean genotypes as determined by SSR markers and (ii) the parental potential of these genotypes in future cultivar breeding and development.

## 2. Materials and Methods

### 2.1. Genetic Material

A total of 49 soybean genotypes were used in the study (Table 1). This sample of genetic material represented a collection of all the cultivars that were introduced, developed, and released in Zimbabwe from 1940 to 2013 [23]. Some of these cultivars were also registered in Malawi and Zambia (Table 1). The germplasm included all the cultivars from Crop Breeding Institute (Zimbabwe) and two commercial seed companies (Seed Co and Pannar Seed) and all based in Zimbabwe. The elite lines from Seed Co came from the cold store that located at Rattray Arnold Research Station, Mukwene Farm Enterprise, Arcturus, Harare; the other elite lines were from Crop Breeding Institute, 5th Street Extension, Harare and rest were from Pannar Seed domiciled at 164 Chihombe Road, Ruwa, Harare. The study was conducted during the summer season of 2011/2012, which stretched from November 2011 through to April of 2012.

**Table 1.** List of soybean genotypes used in the study.

Genotype Code	Designated Name	Growth Habit	Year of Release / Registration	Country
EL1	Hernon 147	Indeterminate	1940	Zimbabwe
EL2	Rhosa	Indeterminate	1966	Zimbabwe
EL3	Bragg	Determinate	1972	Zimbabwe
EL4	Oribi	Determinate	1973	Zimbabwe
EL5	Buffalo	Determinate	1974	Zimbabwe
EL6	Impala	Indeterminate	1977	Zimbabwe
EL7	Kudu	Determinate	1977	Zimbabwe
EL8	Sable	Indeterminate	1980	Zimbabwe
EL9	Duiker	Indeterminate	1982	Zimbabwe
EL10	Roan	Determinate	1985	Zimbabwe
EL11	Gazelle	Indeterminate	1988	Zimbabwe
EL12	Nyala	Indeterminate	1989	Zimbabwe
EL13	Nondo	Determinate	1992	Zimbabwe
EL14	Soma	Indeterminate	1992	Zimbabwe
EL15	SCS1	Determinate	1995	Zambia, Zimbabwe
EL16	Sonnet	Determinate	1994	Zimbabwe
EL17	Sonata	Determinate	1997	Zimbabwe
EL18	Viking	Indeterminate	1999	Zimbabwe

EL19	Solitaire	Indeterminate	1998	Zambia, Malawi, Zimbabwe
EL20	Soprano	Determinate	1998	Zambia, Malawi, Zimbabwe
EL21	Bimha	Determinate	1999	Zimbabwe
EL22	Mhofu	Determinate	1999	Zimbabwe
EL23	Nyati	Indeterminate	1999	Zimbabwe
EL24	Scorpio	Indeterminate	2000	Zimbabwe
EL25	Storm	Determinate	2000	Zambia, Malawi, Zimbabwe
EL26	Safari	Indeterminate	2001	Zambia, Zimbabwe
EL27	Score	Indeterminate	2003	Zambia, Zimbabwe
EL28	Santa	Determinate	2005	Zambia, Zimbabwe
EL29	Siesta	Determinate	2005	Zimbabwe
EL30	Serenade	Indeterminate	2005	Malawi, Zimbabwe
EL31	Sirocco	Indeterminate	2006	Zambia, Malawi, Zimbabwe
EL32	Scribe	Indeterminate	2007	Zambia, Zimbabwe
EL33	Satellite	Indeterminate	2007	Zambia, Zimbabwe
EL34	Squire	Determinate	2007	Malawi, Zimbabwe
EL35	Saga	Determinate	2008	Zimbabwe
EL36	Sequel	Indeterminate	2008	Zimbabwe
EL37	Sputnik	Indeterminate	2008	Zimbabwe
EL38	Sovereign	Indeterminate	2008	Zimbabwe
EL39	S810/6/10	Determinate	2010	Experimental line
EL40	Semeki	Indeterminate	2012	Zambia, Zimbabwe
EL41	Spike	Determinate	2012	Zambia, Zimbabwe
EL42	PN 1867	Determinate	2012	Zimbabwe
EL43	PN 1856	Determinate	2013	Malawi, Zambia, Zimbabwe
EL44	PN 891	Indeterminate	2005	Zimbabwe
EL45	Duocrop	Indeterminate	1999	Exotic line
EL46	Tulumayo	Determinate	2013	Zimbabwe
EL47	SB88	Indeterminate	2013	Exotic line
EL48	SB90	Determinate	2013	Exotic line
EL49	SB92	Determinate	2013	Exotic line

## 2.2. Plant Material and DNA Extraction

Viable seed of the genotypes was grown in pots in a greenhouse and each pot consisted of four plants from which DNA was extracted subsequently. DNA Landmarks, Canada, supplied the DNA kit. It is important to report that DNA sampling was done according to the protocol that was provided. Fresh leaf tissue was harvested from the young leaves (four weeks after planting) from each plant, hence, four leaf discs were sampled and bulked together to represent each genotype. The four sampled leaf discs were placed in a single hole of the 96-well block or plate. After sampling all the genotypes, the entire plate was sealed with air pore tape and placed into a plastic bag together with 50 g of silica gel in order to facilitate drying the leaf discs for 48 hours. The extracted DNA was transferred for profiling in the laboratory (by DNA Landmarks, Canada).

## 2.3. Selection of Primers and Polymerase Chain Reaction Amplification

Thirty SSR markers were used to genotype the cultivars (Table 2). The markers were selected because of their uniform distribution across the soybean genome and amplification quality [24]. Although previous studies have also shown that there is a high correlation between RFLPs, AFLPs and SSRs, the SSRs were found to generate hyper variable polymorphisms [25,26]. The assumption



in selecting the primers was that 10 to 30 highly polymorphic markers would suffice to provide quality or precise data [27]. DNA samples were used for polymerase chain reaction (PCR) amplification with 30 SSR markers. The PCR amplification reactions was accomplished in a total volume of 5µl containing 0.40µl of 2.5mM MgCl<sub>2</sub>, 1.50µl of 10X assay buffer, 1.00µl of 2mM dNTPs, 0.10µ of 5 U/µl Taq polymerase (Thermo Fisher Scientific), 0.06µl of 20µM forward, 0.06µl of 20µM reverse of primers, 1.00µl of each DNA (50 ng) and remaining 0.88µl of pure water. Amplification was performed in PTC Thermal Cycle 100 (MJ Research Inc., 1987) programmed for an initial denaturation of 94°C for three minutes, followed by 35 cycles of 1 minute denaturation at 94°C, one minute annealing at 49°C and extension of 1minute at 72°C. Final extension was done to a period of 10 minutes at 72°C and the product was then stored at 4°C. Data were scored on a scale of -1 to 1 where, -1 denoted missing alleles, 0 represented allele absence and one allele presence.

Table 2. The SSR primer sequences used for soybean genotyping.

Primer name	Forward 5' → 3	Reverse 3' → 5
Satt012	GCAATTAGTTTTAAAATGTTTC	AGAATAGAGCCTACATATAATCATA
Satt148	AATCCGGGAC- GCAAAATTATTATTAA	TGCAAATTCCTAATTAACACCCTTTATAC
Satt156	CGCACCCCTCATCCTATGTA	CCAACTAATCCCAGGGACTTACTT
Satt172	AGCCTCCGGTATCACAG	CCTCCTTTCTCCCATTTT
Satt180	TCGCGTTTGTGTCAGC	TTGATTGAAACCCAACTA
Satt182	GGTCCACATGAAATGAAGGT	TCTCAGCCTGCAAAGAAAA
Satt184	GCGCTATGTAGAT- TATCCAAATTACGC	GCCACTTACTGTTACTCAT
Satt215	GCGCCTTCTTCTGCTAAATCA	CCCATTCAATTGAGATCCAAAATTAC
Satt242	GCGTTGATCAGGTCGAT- TTTTATTGT	GCGAGTGCCAACTAACTACTTTTATGA
Satt294	GCGGGTCAAATGCAAATTATTTT	GCGCTCAGTGTGAAAGTTGTTTCTAT
Satt372	CAGAAAAGGAATAA- TAACAACATCAC	GCGAAAACATAATTCACACAAAAGACAG
Satt387	GCGTTACGTTTCAC- TATTTATTTAACAT	GCGGCAGGCTAGCTACATCAAGAG
Satt394	GCGTTTTTCAATTTAAAGA- GAATTGAC	GCGTAACTTGCATGTGGTATATCGAGATG
Satt397	TCTCGGGATCCTTGTTAGAT	GCGAAGAAGAAGAGAACATGTGAA
Satt414	GCGTATTCCTAGTCACATGC- TATTTC	GCGTCATAATAATGCCTAGAACATAAA
Satt429	GCGAC- CATCATCTAATCACAATCTACTA	TCCCCATCATTTATCGAAAATAATAATT
Satt434	GCGTTCGGATATACTATA- TAATCCTAAT	GCGGGGTTAGTCTTTTTATTAACTTAA
Satt441	AAACCCACCCTCAAAAATAAAAA	AAATGCACCCATCAATCACA
Satt459	TCGTGTTAGAT- TTTTACTGTCACATT	AACTGCATACCCTTTGTGTTGAA
Satt477	GTTGGGAAAAGGTTACTAC- CATATC	GGTCCGTATGCAATTCCTGACTAATA
Satt490	GCGGCACGAG- TCAACTTTCTGTTTCCT	GCGGAAGAAGATTTTCGTTTTTAT
Satt509	GCGCTACCGTGTGGTGGTGTGC- TACCT	GCGCAAGTGGCCAGCTCATCTATT
Satt511	GCGACTTTACTGAAAACCTGGAAAGCTTCAAACCAACAAACAACCTA	
Satt522	GCGAACTGCCTAGGTTAAAA	TTAGGCGAAATCAACAAT
Satt530	CATGCATATTGACTTCATTATT	CCAAGCGGGTGAAGAGGTTTTT
Satt577	CAAGCTTAAGTCTTGGTCTTCTCT	GGCCTGACCCAAAACCTAAGGGAAGTG

Satt590	GCGCG- CATTTTTTAAGTTAATGTTCT	GCGCGAGTTAGCGAATTATTTGTC
Satt598	CGATTGGAATATACTTACCGTC- TATA	CACAATACCTGTGGCTGTTATACTAT
Sct_034	AATTCTCACTCTCACAACTTC	CCATGGGAATAGTTGGGT
Sct_067	CTCCCCATCTCTCTCTAAC	GGATTTTGTTATTTATTTATTGA

2.4. SSR Data Analysis

The binary data matrix was used to calculate the genetic similarity matrix using Dice Similarity Coefficient [28] with the help of the Numerical Taxonomy Multivariate Analysis System for personal computer (NTSYS-pc) version 2.1 [29]. The resultant similarity distance matrix data was used to construct a dendrogram using the agglomerative hierarchical un-weighted pair-group method with an arithmetic average (UPGMA). Powermarker V3.25 was used to determine major allele frequency, gene diversity, observed heterozygosity and polymorphic information content (PIC) values for each SSR marker used in the study [30]. The expected heterozygosity ( $H_e$ ) and observed heterozygosity ( $H_o$ ) were used to evaluate the genetic diversity within the set of genotypes. Expected heterozygosity, i.e. the probability that two alleles from the same locus would be different when selected at random was estimated for each SSR locus according to the following equation [31];

$$H_e = 1 - \sum p_i^2,$$
 (1)

where  $\Sigma$  stands for summation over all alleles;  $p_i$  is the expected frequency of the  $i$ th allele at a locus for individual  $p$ .

Observed heterozygosity was estimated by dividing the number of heterozygous individuals by the number of individuals scored. Polymorphic information content for the SSR markers in the sample DNA was calculated as follows:

$$PIC = 1 - \sum p_i^2,$$
 (2)

where  $p_i$  is the expected frequency of the  $i$ th allele at a locus for individual  $p$ .

3. Results and Discussion

A total of 135 alleles were recorded and the number of alleles scored per locus ranged from two to nine, with a mean of 4.57 alleles per locus (Table 3). In addition, the gene diversity ranged from 0.10 to 0.79 with an average of 0.50 for the genotypes studied. The polymorphic information content (PIC) estimated for all loci ranged between 0.10 and 0.77 of which primer (Sct\_034) showed a minimum of 0.10, and Satt012 showed a maximum of 0.77. The PIC was average of 0.45. Results also showed that the means for the major allele frequency and observed heterozygosity ( $H_o$ ) were 0.63 and 0.11 respectively. Among the primers that were used, Satt012, Satt414 and Satt372 were highly informative with PIC values of 0.77%, 0.71% and 0.68% respectively.

The pair wise genetic similarity coefficients among the 49 genotypes varied from 10% to 98% (Table 4). The highest similarity coefficient was observed between genotypes EL20 and EL19, while the lowest similarity coefficient was between genotypes, EL41 and EL7 and EL41 and EL1 implying that genetic distance (dissimilarity coefficient) was 90%. Clustering analysis grouped the cultivars into 5 clusters at 46% dice similarity (Figure 2). The closest distance was between genotypes EL25 and EL26 while the greatest genetic distance (dissimilarity) was between genotypes EL41 and EL7; and EL41 and EL1. Cluster 12 consisted of the majority of genotypes (17) accounting for 40% of the total population.

**Table 3.** Details of polymorphisms and genetic analysis of 49 soybean genotypes using 30 SSR markers.

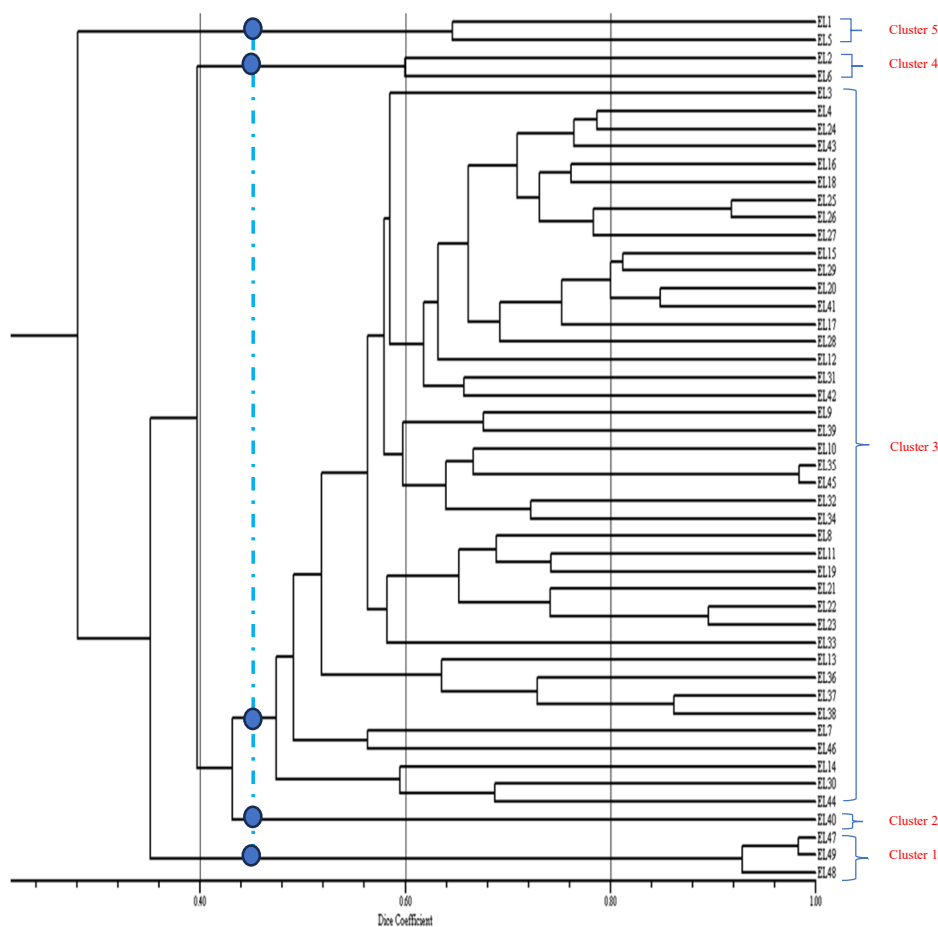
Marker	Major allele frequency	Number of alleles	Gene diversity	Observed heterozygosity (Ho)	Polymorphic information content
Satt394	0.85	5.00	0.27	0.00	0.26
Satt577	0.68	3.00	0.47	0.15	0.40
Satt180	0.74	5.00	0.43	0.08	0.40
Satt459	0.76	3.00	0.38	0.06	0.35
Satt522	0.44	4.00	0.65	0.19	0.59
Satt598	0.67	2.00	0.44	0.13	0.34
Satt242	0.56	5.00	0.57	0.17	0.49
Satt429	0.40	6.00	0.69	0.13	0.64
Satt182	0.91	3.00	0.17	0.08	0.17
Satt397	0.50	2.00	0.50	0.06	0.38
Satt148	0.62	4.00	0.55	0.19	0.50
Satt477	0.84	2.00	0.27	0.09	0.23
Satt530	0.75	7.00	0.43	0.09	0.41
Satt590	0.58	8.00	0.64	0.15	0.61
Satt012	0.34	8.00	0.79	0.11	0.77
Satt172	0.61	3.00	0.53	0.17	0.46
Satt414	0.41	7.00	0.74	0.09	0.71
Satt215	0.78	4.00	0.37	0.06	0.35
Satt387	0.50	3.00	0.55	0.13	0.45
Satt441	0.49	9.00	0.69	0.19	0.66
Satt156	0.65	5.00	0.52	0.11	0.46
Satt184	0.67	4.00	0.51	0.19	0.47
Satt294	0.43	6.00	0.68	0.13	0.63
Satt434	0.75	6.00	0.43	0.09	0.41
Satt511	0.49	4.00	0.67	0.17	0.63
Satt490	0.50	3.00	0.58	0.09	0.40
Satt509	0.65	5.00	0.54	0.06	0.51
Sct_034	0.95	3.00	0.09	0.02	0.10
Satt372	0.35	6.00	0.73	0.21	0.68
Sct_067	0.91	2.00	0.17	0.00	0.16
Total	-	135	-	-	-
Mean	0.63	4.57	0.50	0.11	0.45



**Table 4.** The similarity matrix of the 49 soybean genotypes generated by Numerical Taxonomy Multivariate Analysis System for personal computer (NTSYS-pc).

	EL1	EL2	EL3	EL4	EL5	EL6	EL7	EL8	EL9	EL10	EL11	EL12	EL13	EL14	EL15	EL16	EL17	EL18	EL19	EL20	EL21	EL22	EL23	EL24
EL1	1.00																							
EL2	0.19	1.00																						
EL3	0.29	0.30	1.00																					
EL4	0.22	0.27	0.69	1.00																				
EL5	0.65	0.36	0.26	0.19	1.00																			
EL6	0.30	0.60	0.46	0.54	0.27	1.00																		
EL7	0.28	0.33	0.56	0.49	0.29	0.42	1.00																	
EL8	0.28	0.40	0.60	0.74	0.26	0.64	0.57	1.00																
EL9	0.28	0.29	0.53	0.70	0.31	0.53	0.43	0.68	1.00															
EL10	0.31	0.38	0.66	0.67	0.34	0.53	0.66	0.64	0.64	1.00														
EL11	0.16	0.37	0.57	0.77	0.26	0.48	0.49	0.71	0.52	0.55	1.00													
EL12	0.28	0.30	0.58	0.61	0.31	0.52	0.51	0.62	0.59	0.56	0.52	1.00												
EL13	0.27	0.38	0.49	0.49	0.33	0.52	0.59	0.60	0.58	0.56	0.46	0.56	1.00											
EL14	0.22	0.37	0.39	0.43	0.35	0.37	0.36	0.53	0.58	0.46	0.40	0.61	0.57	1.00										
EL15	0.31	0.20	0.56	0.69	0.25	0.47	0.48	0.57	0.63	0.54	0.52	0.60	0.62	0.33	1.00									
EL16	0.24	0.44	0.58	0.74	0.33	0.58	0.73	0.72	0.68	0.68	0.71	0.66	0.59	0.52	0.61	1.00								
EL17	0.29	0.24	0.60	0.70	0.23	0.45	0.56	0.55	0.58	0.67	0.50	0.58	0.54	0.37	0.75	0.62	1.00							
EL18	0.32	0.39	0.55	0.69	0.33	0.59	0.54	0.58	0.64	0.60	0.59	0.73	0.57	0.48	0.75	0.76	0.62	1.00						
EL19	0.25	0.36	0.52	0.74	0.34	0.49	0.54	0.67	0.68	0.73	0.74	0.59	0.50	0.58	0.54	0.75	0.61	0.67	1.00					
EL20	0.36	0.25	0.59	0.66	0.24	0.56	0.55	0.60	0.55	0.52	0.53	0.58	0.62	0.38	0.83	0.61	0.72	0.71	0.52	1.00				
EL21	0.09	0.40	0.53	0.69	0.25	0.44	0.52	0.60	0.54	0.66	0.72	0.44	0.48	0.36	0.52	0.70	0.72	0.51	0.70	0.52	1.00			
EL22	0.23	0.48	0.57	0.69	0.32	0.54	0.53	0.67	0.55	0.61	0.72	0.52	0.52	0.39	0.56	0.72	0.63	0.52	0.67	0.54	0.79	1.00		
EL23	0.22	0.44	0.54	0.60	0.32	0.48	0.52	0.58	0.49	0.55	0.63	0.48	0.46	0.33	0.52	0.68	0.60	0.48	0.58	0.47	0.69	0.90	1.00	
EL24	0.22	0.23	0.65	0.79	0.25	0.53	0.48	0.65	0.69	0.63	0.62	0.73	0.54	0.49	0.81	0.73	0.72	0.78	0.63	0.77	0.58	0.59	0.52	1.00

	EL25	EL26	EL27	EL28	EL29	EL30	EL31	EL32	EL33	EL34	EL35	EL36	EL37	EL38	EL39	EL40	EL41	EL42	EL43	EL44	EL45	EL46	EL47	EL48	EL49
EL25	1.00																								
EL26	0.92	1.00																							
EL27	0.74	0.83	1.00																						
EL28	0.62	0.65	0.67	1.00																					
EL29	0.70	0.69	0.62	0.66	1.00																				
EL30	0.50	0.49	0.55	0.66	0.67	1.00																			
EL31	0.70	0.62	0.52	0.56	0.67	0.57	1.00																		
EL32	0.53	0.52	0.58	0.69	0.70	0.67	0.53	1.00																	
EL33	0.56	0.58	0.57	0.52	0.63	0.62	0.49	0.59	1.00																
EL34	0.61	0.55	0.51	0.55	0.64	0.53	0.56	0.72	0.49	1.00															
EL35	0.52	0.45	0.51	0.65	0.53	0.49	0.46	0.66	0.45	0.60	1.00														
EL36	0.53	0.52	0.52	0.62	0.54	0.43	0.37	0.63	0.43	0.53	0.46	1.00													
EL37	0.58	0.57	0.56	0.67	0.61	0.51	0.41	0.58	0.47	0.45	0.47	0.78	1.00												
EL38	0.61	0.60	0.52	0.57	0.58	0.41	0.44	0.44	0.37	0.45	0.40	0.68	0.86	1.00											
EL39	0.54	0.53	0.49	0.60	0.58	0.41	0.37	0.64	0.43	0.56	0.57	0.61	0.69	0.62	1.00										
EL40	0.47	0.46	0.39	0.43	0.44	0.30	0.40	0.50	0.36	0.44	0.36	0.47	0.47	0.58	0.47	1.00									
EL41	0.65	0.63	0.59	0.70	0.80	0.68	0.71	0.65	0.63	0.62	0.48	0.55	0.59	0.56	0.52	0.52	1.00								
EL42	0.63	0.56	0.56	0.56	0.63	0.57	0.66	0.49	0.56	0.59	0.51	0.43	0.49	0.52	0.43	0.37	0.69	1.00							
EL43	0.72	0.65	0.67	0.61	0.66	0.56	0.56	0.59	0.48	0.66	0.61	0.46	0.47	0.50	0.50	0.46	0.60	0.65	1.00						
EL44	0.56	0.49	0.48	0.43	0.51	0.69	0.47	0.44	0.49	0.53	0.40	0.31	0.38	0.41	0.35	0.41	0.52	0.54	0.55	1.00					
EL45	0.55	0.48	0.53	0.67	0.55	0.52	0.48	0.68	0.48	0.62	0.98	0.45	0.46	0.39	0.56	0.35	0.47	0.50	0.60	0.42	1.00				
EL46	0.46	0.45	0.47	0.59	0.52	0.40	0.46	0.46	0.34	0.51	0.42	0.40	0.35	0.38	0.38	0.34	0.53	0.45	0.54	0.38	0.44	1.00			
EL47	0.43	0.42	0.38	0.35	0.41	0.30	0.43	0.26	0.29	0.33	0.26	0.36	0.37	0.43	0.40	0.26	0.35	0.34	0.32	0.28	0.29	0.54	1.00		
EL48	0.41	0.41	0.37	0.38	0.39	0.29	0.41	0.22	0.28	0.29	0.22	0.35	0.35	0.45	0.35	0.32	0.37	0.33	0.31	0.27	0.25	0.55	0.94	1.00	
EL49	0.40	0.39	0.39	0.33	0.41	0.30	0.43	0.27	0.30	0.33	0.26	0.33	0.34	0.41	0.37	0.23	0.35	0.34	0.33	0.28	0.29	0.54	0.98	0.92	1.00



**Figure 2.** Genetic relationships among 49 soybean genotypes evaluated using 30 SSR markers.

The total number of alleles found among the soybean genotypes and the range of effective allele per marker shows that the markers were effective in assessing the available genetic diversity. It also suggests that there was allelic polymorphism in the soybean gene that can be exploited for soybean improvement or germplasm management. The average number of alleles per locus (4.57) found in this study demonstrated high levels of genetic diversity in the gene pool and was consistent with the findings reported in other studies [18,32]. The number of effective alleles reported is partly determined by the number and nature of molecular markers used for genotyping. Genotyping a panel of germplasm with a higher number of markers could increase the number of effective alleles found. On the other hand, markers such as single nucleotide polymorphism (SNP) markers can also identify a more effective allele because they assess polymorphism at one locus while differences identified by SSR markers could be less abundant. While comparisons between studies are important as a guide, there will always be differences due to the choice and number of markers used and inherent genetic differences in the germplasm. The expected heterozygosity of ( $H_e = 0.11$ ) found in this study is within range of self-pollinating species. Soybean is a highly autogamous species, which increases homozygosity and a corresponding decrease in heterozygosity. The strict autogamous nature can be a barrier to introduction of new gene combinations in the gene pool. Low heterozygosity is a challenge in creating new recombinants as it shows that most of the tested loci are homozygous. Homozygosity is also a challenge especially when recessive deleterious alleles occur in the chromosome. Thus, genetic variation may gradually be eroded due to inbreeding if natural or induced mutations do not occur. The PIC in this study was 0.45 with only 10 markers with PIC values over 0.50 showing that most of the markers had moderate ability to discriminate the genotypes. The moderate PIC values are also a reflection of the genetic diversity in the germplasm under investigation. The PIC values are good indicators of possibilities in finding parental donor alleles in the progeny and can be used to trace parental pedigrees and marker-trait associations.

The SSR marker data separated the genotypes into 5 clusters which, concurred with the pedigree information. EL41 was assigned to the same cluster that contained its progeny EL37. Also, genotypes EL19 and EL20 which are sister lines (full sibs) were grouped together. The results demonstrated that the SSR markers could discriminate the genotypes. The ability of the SSR markers to detect polymorphisms at molecular level (DNA) justified their effectiveness and power in differentiating the genotypes (Powell et al., 1996). The genetic relationships that were displayed in this study could facilitate the selection of parental stock for hybridization purposes in addition to the other breeding lines available in the program. Selection of the most divergent parents would suggest that unique alleles for the desirable traits could be exploited. In other studies, SSR markers were used successfully to determine genotypic diversity [33–35]. In soybean, SSR markers were used for constructing linkage maps and identifying a quantitative trait locus [36]. Genetically distant and high performing genotypes should be selected as parental lines for genetic recombination to reduce inbreeding depression and improve genetic gain. Inbreeding depression results in reduced fitness among the offspring. Parents that have small genetic distances should be avoided because they have low chances of producing transgressive segregates and the recombinants will be less variable [37]. Ideally, genetically diverse parents will help to accelerate genetic gains as they are assumed to possess complementary genes. The use of genetic markers is a reliable method to identify relationships among genotypes to use in breeding programs.

In conclusion, it was evident from this study that there was genetic diversity at the molecular level among the 49 genotypes that were evaluated. The genetic patterns obtained from the study could help the breeders to make better and reliable choices of distant parents when planning a crossing program to obtain higher genetic variation among segregates. However, it would be prudent to introduce exotic germplasm in order to further enrich the available diversity. It is also important to complement the molecular marker analysis with phenotyping to identify genotype environment interactions. This would also allow an opportunity to identify high performing and adapted genotypes and establish marker-trait associations for marker assisted selection.

**Author Contributions:** Conceptualization, H.M., I.M., P.T., E.T.G.; Investigation, H.M., I.M., P.T., E.T.G., G.C., I.D., A.A.T., J.D.; Methodology, H.M., I.M., P.T., E.T.G., G.C., I.D., A.A.T., J.D.; Writing – original draft, H.M., I.M., P.T., E.T.G., G.C., I.D., A.A.T., J.D.; Project administration, P.T.

**Funding:** This research received no external funding.

**Data Availability Statement:** Despite our best efforts to retrieve and maintain the original datasets used to support the findings reported in our manuscript, unforeseen circumstances beyond our control have resulted in data loss. We are committed to transparency and scientific integrity and are prepared to welcome any guidance or suggestions regarding the next steps.

**Acknowledgments:** The authors would like to acknowledge the support provided by several organizations: the University of Venda, South Africa; the West Africa Centre for Crop Improvement of the University of Ghana; One CGIAR; and the International Institute of Tropical Agriculture.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Siamabele, B. Soya Beans Production in Zambia: Opportunities and Challenges. *Am J Agric Biol Sci* **2019**, *14*, 55–60, doi:10.3844/AJABSSP.2019.55.60.
2. Tufa, A.H.; Alene, A.D.; Manda, J.; Akinwale, M.G.; Chikoye, D.; Feleke, S.; Wossen, T.; Manyong, V. The Productivity and Income Effects of Adoption of Improved Soybean Varieties and Agronomic Practices in Malawi. *World Dev* **2019**, *124*, 104631, doi:10.1016/J.WORLDDEV.2019.104631.
3. Technoserve Southern Africa Regional Soybean Roadmap Final Report; 2011;
4. Khojely, D.M.; Ibrahim, S.E.; Sapey, E.; Han, T. History, Current Status, and Prospects of Soybean Production and Research in Sub-Saharan Africa. *Crop J* **2018**, *6*, 226–235, doi:10.1016/J.CJ.2018.03.006.
5. FAOSTAT Available online: <https://www.fao.org/faostat/en/#home>.
6. Omondi, J.O.; Mkuhlani, S.; Mugo, J.; Chibeba, A.M.; Chiduwa, M.S.; Chigeza, G.; Kyei-Boahen, S.; Masikati, P.; Nyagumbo, I. Closing the Yield Gap of Soybean (*Glycine Max* (L.) Merrill) in Southern Africa: A Case of Malawi, Zambia, and Mozambique. *Frontiers in Agronomy* **2023**, *5*, 1219490, doi:10.3389/FAGRO.2023.1219490/BIBTEX.
7. Santos, M.; Da, M.; Santos, F. SOYBEAN VARIETIES IN SUB-SAHARAN AFRICA. *Afr. J. Food Agric. Nutr. Dev* **2019**, *19*, 15136–15139, doi:10.18697/ajfand.88.SILFarmDoc06.

8. Mukuze, C.; Tukamuhabwa, P.; Maphosa, M.; Dari, S.; Dramadri, I.O.; Obua, T.; Kongai, H.; Rubaihayo, P. Genetic Diversity Analysis among Soybean Genotypes Using SSR Markers in Uganda. *African Journal of Biotechnology* **2020**, *19*, 439–448, doi:10.5897/AJB2020.17152.
9. Jo, H.; Lee, J.Y.; Cho, H.; Choi, H.J.; Son, C.K.; Bae, J.S.; Bilyeu, K.; Song, J.T.; Lee, J.D. Genetic Diversity of Soybeans (Glycine Max (L.) Merr.) with Black Seed Coats and Green Cotyledons in Korean Germplasm. *Agronomy* **2021**, Vol. 11, Page 581 **2021**, *11*, 581, doi:10.3390/AGRONOMY11030581.
10. Gwata, E.T.; Wofford, D.S.; Boote, K.J.; Blount, A.R.; Pfahler, P.L. Inheritance of Promiscuous Nodulation in Soybean. *Crop Sci* **2005**, *45*, 635–638, doi:10.2135/CROPSCI2005.0635.
11. Gwata, E.T.; Nziramasanga, N. Seed Protein and Oil Content in Zimbabwean Soyabean (Glycine Max L.) Varieties. *Plant Varieties and Seeds* **2001**, *14*, 125–128.
12. Nassiuma, D.; Wasike, W. Stability Assessment of Soybean Varieties in Kenya. *Afr Crop Sci J* **2002**, *10*, doi:10.4314/ACSJ.V10I2.27546.
13. Reyna, N.; Sneller, C.H. Evaluation of Marker-Assisted Introgression of Yield QTL Alleles into Adapted Soybean. *Crop Sci* **2001**, *41*, 1317–1321, doi:10.2135/CROPSCI2001.4141317X.
14. Gizlice, Z.; Carter Jnr, T.E.; Burton, J.W. Genetic Base for North American Public Soybean Cultivars Released between 1947 and 1988. *Crop Sci* **1994**, *34*, 1143–1151, doi:10.2135/CROPSCI1994.0011183X003400050001X.
15. Junyi, G.; Tuanjie, Z. The Core Ancestors of Soybean Cultivars in China. *Journal of Nanjing Agricultural University* **2001**, *24*, 20–23.
16. Bhat, S.; Basavaraja, G.; Salimath, P. Analysis of Variability in Segregating Generation of Soybean (Glycine Max (L.) Merrill). *Karnataka Journal of Agricultural Sciences* **2012**.
17. Mulato, B.M.; Möller, M.; Zucchi, M.I.; Quecini, V.; Pinheiro, J.B. Genetic Diversity in Soybean Germplasm Identified by SSR and EST-SSR Markers. *Pesqui Agropecu Bras* **2010**, *45*, 276–283.
18. Tantasawat, P.; Trongchuen, J.; Prajongjai, T.; Jenweerawat, S.; Chaowiset, W. SSR Analysis of Soybean (“Glycine Max” (L.) Merr.) Genetic Relationship and Variety Identification in Thailand. *Aust J Crop Sci* **2011**.
19. Baruah, J.; Pandey, S.K.; Sarmah, N.; Lal, M. Assessing Molecular Diversity among High Capsaicin Content Lines of Capsicum Chinense Jacq. Using Simple Sequence Repeat Marker. *Ind Crops Prod* **2019**, *141*, 111769, doi:10.1016/J.INDCROP.2019.111769.
20. Behera, T.K.; Sharma, P.; Singh, B.K.; Kumar, G.; Kumar, R.; Mohapatra, T.; Singh, N.K. Assessment of Genetic Diversity and Species Relationships in Eggplant (Solanum Melongena L.) Using STMS Markers. *Sci Hortic* **2006**, *107*, 352–357, doi:10.1016/J.SCI.2005.11.004.
21. Franco, J.; Crossa, J.; Ribaut, J.M.; Bertran, J.; Warburton, M.L.; Khairallah, M. A Method for Combining Molecular Markers and Phenotypic Attributes for Classifying Plant Genotypes. *Theoretical and Applied Genetics* **2001**, *103*, 944–952, doi:10.1007/s001220100641.
22. Manjarrez-Sandoval, P.; Carter, T.E.; Webb, D.M.; Burton, J.W. RFLP Genetic Similarity Estimates and Coefficient of Parentage as Genetic Variance Predictors for Soybean Yield. *Crop Sci* **1997**, *37*, 698–703, doi:10.2135/CROPSCI1997.0011183X003700030002X.
23. Mushoriwa, H.; Mathew, I.; Gwata, E.T.; Tongoona, P.; Derera, J. Grain Yield Potential and Stability of Soybean Genotypes of Different Ages across Diverse Environments in Southern Africa. *Agronomy* **2022**, Vol. 12, Page 1147 **2022**, *12*, 1147, doi:10.3390/AGRONOMY12051147.
24. Cregan, P.B.; Jarvik, T.; Bush, A.L.; Shoemaker, R.C.; Lark, K.G.; Kahler, A.L.; Kaya, N.; VanToai, T.T.; Lohnes, D.G.; Chung, J.; et al. An Integrated Genetic Linkage Map of the Soybean Genome. *Crop Sci* **1999**, *39*, 1464–1490, doi:10.2135/CROPSCI1999.3951464X.
25. Gogoi, A.; Munda, S.; Paw, M.; Begum, T.; Siddiqui, M.H.; Gaafar, A.R.Z.; Kesawat, M.S.; Lal, M. Molecular Genetic Divergence Analysis amongst High Curcumin Lines of Golden Crop (Curcuma Longa L.) Using SSR Marker and Use in Trait-Specific Breeding. *Scientific Reports* **2023**, *13*, 1–18, doi:10.1038/s41598-023-46779-5.
26. Rongwen, J.; Akkaya, M.S.; Bhagwat, A.A.; Lavi, U.; Cregan, P.B. The Use of Microsatellite DNA Markers for Soybean Genotype Identification. *Theoretical and Applied Genetics* **1995**, *90*, 43–48.
27. Guichoux, E.; Lagache, L.; Wagner, S.; Chaumeil, P.; Léger, P.; Lepais, O.; Lepoittevin, C.; Malausa, T.; Revardel, E.; Salin, F.; et al. Current Trends in Microsatellite Genotyping. *Mol Ecol Resour* **2011**, *11*, 591–611, doi:10.1111/J.1755-0998.2011.03014.X.
28. Dice, L.R. Measures of the Amount of Ecologic Association Between Species. *Ecology* **1945**, *26*, 297–302, doi:10.2307/1932409.
29. Rohlf, F.J.; Applied Biostatistics, Inc.; Exeter Software (Firm) NTSYS-Pc: Numerical Taxonomy and Multivariate Analysis System. **2009**.
30. Liu, K.; Muse, S. V PowerMarker: An Integrated Analysis Environment for Genetic Marker Analysis. *BIOINFORMATICS APPLICATIONS NOTE* **2005**, *21*, 2128–2129, doi:10.1093/bioinformatics/bti282.
31. Nei, M. Analysis of Gene Diversity in Subdivided Populations. *Proc Natl Acad Sci U S A* **1973**, *70*, 3321–3323, doi:10.1073/PNAS.70.12.3321.
32. Priolli, R.H.G.; Pinheiro, J.B.; Zucchi, M.I.; Bajay, M.M.; Vello, N.A. Genetic Diversity among Brazilian Soybean Cultivars Based on SSR Loci and Pedigree Data. *Brazilian Archives of Biology and Technology* **2010**, *53*, 519–531, doi:10.1590/S1516-89132010000300004.
33. Powell, W.; Morgante, M.; Andre, C.; Hanafey, M.; Vogel, J.; Tingey, S.; Rafalski, A. The Comparison of RFLP, RAPD, AFLP and SSR (Microsatellite) Markers for Germplasm Analysis. *Molecular breeding* **1996**, *2*, 225–238, doi:10.1007/BF00564200.

34. Thompson, J.A.; Nelson, R.L.; Vodkin, L.O. Identification of Diverse Soybean Germplasm Using RAPD Markers. *Crop Sci* **1998**, *38*, 1348–1355, doi:10.2135/CROPSCI1998.0011183X003800050033X.
35. Ude, G.N.; Kenworthy, W.J.; Costa, J.M.; Cregan, P.B.; Alvernaz, J. Genetic Diversity of Soybean Cultivars from China, Japan, North America, and North American Ancestral Lines Determined by Amplified Fragment Length Polymorphism. *Crop Sci* **2003**, *43*, 1858–1867, doi:10.2135/CROPSCI2003.1858.
36. Thovhogi, F.; Gwata, E.T.; McHau, G.R.A.; Safodien, S.S.; Koopman, T. Molecular Characterization of Spider Plant (Cleome gynandra) Accessions Using SSR Markers. *Agronomy* **2021**, *11*, 2206, doi:10.3390/AGRONOMY11112206/S1.
37. Biswas, M.K.; Mondal, M.A.A.; Hossain, M.; Islam, R. Utilization of Genetic Diversity and Its Association with Heterosis for Progeny Selection in Potato Breeding Programs. *American-Eurasian J. Agric. Env. Sci* **2008**, *3*, 882–887.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.