

Article

Assessment of the Genetic Structure and Diversity of Soybean (*Glycine max* L.) Germplasm Using Diversity Array Technology and Single Nucleotide Polymorphism Markers

Abdulwahab S. Shaibu ^{1,*}, Hassan Ibrahim ¹, Zainab L. Miko ¹, Ibrahim B. Mohammed ¹, Sanusi G. Mohammed ², Hauwa L. Yusuf ³, Alpha Y. Kamara ⁴, Lucky O. Omoigui ⁴ and Benjamin Karikari ⁵

¹ Department of Agronomy, Bayero University Kano, Kano 700001, Nigeria; hassibonline1@gmail.com (H.I.); zlmiko.agr@buk.edu.ng (Z.L.M.); ibmohammed.agr@buk.edu.ng (I.B.M.)

² Centre for Dryland Agriculture, Bayero University Kano, Kano 700001, Nigeria; sanusi.gaya@buk.edu.ng

³ Department of Food Science and Technology, Bayero University Kano, Kano 700001, Nigeria; ladsufy83@gmail.com

⁴ International Institute of Tropical Agriculture, Ibadan 200211, Nigeria; a.kamara@cgiar.org (A.Y.K.); l.omoigui@cgiar.org (L.O.O.)

⁵ Department of Crop Science, Faculty of Agriculture, Food and Consumer Sciences, University for Development Studies, P.O. Box TL 1882, Tamale 00233, Ghana; benkarikari1@gmail.com

* Correspondence: asshuaibu.agr@buk.edu.ng; Tel.: +234-80-6601-9498

Abstract: Knowledge of the genetic structure and diversity of germplasm collections is crucial for sustainable genetic improvement through hybridization programs and rapid adaptation to changing breeding objectives. The objective of this study was to determine the genetic diversity and population structure of 281 International Institute of Tropical Agriculture (IITA) soybean accessions using diversity array technology (DArT) and single nucleotide polymorphism (SNP) markers for the efficient utilization of these accessions. From the results, the SNP and DArT markers were well distributed across the 20 soybean chromosomes. The cluster and principal component analyses revealed the genetic diversity among the 281 accessions by grouping them into two stratifications, a grouping that was also evident from the population structure analysis, which divided the 281 accessions into two distinct groups. The analysis of molecular variance revealed that 97% and 98% of the genetic variances using SNP and DArT markers, respectively, were within the population. Genetic diversity indices such as Shannon's diversity index, diversity and unbiased diversity revealed the diversity among the different populations of the soybean accessions. The SNP and DArT markers used provided similar information on the structure, diversity and polymorphism of the accessions, which indicates the applicability of the DArT marker in genetic diversity studies. Our study provides information about the genetic structure and diversity of the IITA soybean accessions that will allow for the efficient utilization of these accessions in soybean improvement programs, especially in Africa.

Keywords: DArT-seq; genetic diversity; population structure; soybean; AMOVA; Africa



Citation: Shaibu, A.S.; Ibrahim, H.; Miko, Z.L.; Mohammed, I.B.; Mohammed, S.G.; Yusuf, H.L.; Kamara, A.Y.; Omoigui, L.O.; Karikari, B. Assessment of the Genetic Structure and Diversity of Soybean (*Glycine max* L.) Germplasm Using Diversity Array Technology and Single Nucleotide Polymorphism Markers. *Plants* **2022**, *11*, 68. <https://doi.org/10.3390/plants11010068>

Academic Editor: Yan Li

Received: 12 October 2021

Accepted: 14 December 2021

Published: 26 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Soybean (*Glycine max* L. Merrill) is an important legume crop, providing approximately 68% of the world's supply of protein-rich meals and serving as an excellent source of vegetable oil and renewable fuel. The average worldwide yields of soybean increased from 2183 kg ha⁻¹ in 1994 to 2769 kg ha⁻¹ in 2019, with an annual increase of 26 kg ha⁻¹ [1]. This shows that there is still considerable potential for increasing soybean yield. The use of new sources of genetic variation to develop improved varieties is enhanced when the selection of parental genotypes is based on both phenotypic and genetic diversity [2]. Therefore, determining the variation in the soybean germplasm is important for the effective selection of parental lines for hybridization and the development of improved varieties.

Furthermore, the development of improved crop varieties that are resilient to the changing climate is one of the strategies for overcoming the challenges of climate change and population growth [3]. To achieve this, the genetic diversity of crops needs to be exploited and sustained [4]. Similarly, determining the genetic diversity available in the germplasm is crucial for developing new improved cultivars with desirable attributes. However, in the developing world, the low genetic diversity in the cultivated gene pool has significantly hindered soybean improvement [5–8]. Previous studies that were conducted on the genetic diversity of soybean utilized accessions from China and the USA [9], the Korean peninsula [10] and a few accessions from Africa [11]. Therefore, it is important to assess the genetic diversity of soybean germplasms comprising accessions from Africa for achieving a sustainable improvement in soybean yield and quality, as this crop is predicted to dominate crop production across Africa soon, due to its potential [12].

The genetic diversity of crops can be assessed at phenotypic and genotypic levels or by using statistical methods to partition phenotypic or genetic descriptors into genetic or environmental components [13]. The diversity among genotypes can be estimated using morphological markers [14]; however, these markers are less effective than DNA markers because of their limited number, subjectiveness and susceptibility to environmental influences [14].

The use of DNA markers to evaluate diversity aids in the effective utilization of germplasm for crop productivity improvement and conservation. The availability of a reliable, robust and economical marker platform is important for breeding programs aimed at utilizing genetic diversity. Genetic markers such as single nucleotide polymorphism (SNP) markers are more reliable in estimating crop diversities [15]. Recently, the application of diversity array technology (DArT), which is an advanced genotyping-by-sequencing (GBS) platform [16], has resulted in a more efficient method of detecting high-density SNP information with a significantly reduced time and cost [17]. Thus, studies of the genetic diversity of several crop species have a relatively better coverage and fewer missing data using DArT compared to other GBS platforms [18,19].

Recently, soybean germplasms from Africa have been reported to have a narrow genetic base, based on phenotypic evidence [20]. However, less information is available at the molecular level [21]. A few studies have reported the genetic diversity of soybean accessions from Africa using simple sequence repeats (SSR) [11], but there is no available information about these accessions based on GBS-SNP markers. Therefore, the objective of this study was to examine the genetic diversity and population structure of 281 soybean accessions of the International Institute of Tropical Agriculture (IITA) using DArT-SNP markers in order to gain a better understanding of their potential utilization in soybean improvement programs in Africa.

2. Results

2.1. Marker Quality and Genome Characterization

A total of 19,505 SNPs and 16,116 DArT markers were generated for the 281 soybean accessions. After the first stage of filtering, which excluded unmapped markers, 18,270 SNPs and 14,051 DArT markers were obtained. The second stage filtering returned 7244 SNPs and 4443 DArT markers. The SNPs were distributed across the 20 soybean chromosomes (Figure S1a), with Gm18 having the highest number of SNPs (533), while Gm12 had the lowest number (261). For the distribution of the DArT markers on the chromosomes, Gm16 had the highest number of markers (388) followed by Gm18 (387), while Gm12 had the lowest number (126) (Figure S1b). The polymorphism information content (PIC) ranged from 0.10–0.50 for both SNP and DArT markers (Figure S2a,b, respectively). Other quality parameters are presented in Table S1. For both the SNP and DArT markers, their minor allele frequency (MAF) ranged from 0.05–0.50 and their reproducibility ranged from 0.95–1.00 (Table S1).

2.2. Population Structure of the Soybean Accessions

A cluster analysis of the SNP and DArT markers revealed two groups, with each group having sub-groups (Figure 1A,B, respectively). For the SNP data, Group I had four sub-groups, with the USA accessions scattered throughout all the sub-groups, while Group II had 10 sub-groups (Figure 1A), comprising accessions from Taiwan and China. The DArT cluster analysis also revealed that Group I had nine sub-groups, and Group II had five sub-groups (Figure 1B). The accessions from the different countries were scattered throughout all the sub-groups. Principal component analysis (PCA) was conducted for the SNP and DArT markers (Figure 2A,B). The PCA for the SNP markers revealed two groups, with Group 2 having two sub-groups (Figure 2A). The accessions from the USA were scattered over both groups with the majority in Group 1, while the accessions from Africa were mainly in Group 2. Group 2 was further divided into two sub-groups, which were similar to Group 1 accessions. A similar trend was observed for the PCA of the DArT markers (Figure 2B).

The model-based (Bayesian) analysis of the SNP markers estimated two groups (Q1 and Q2) (Figure 3A), corresponding to the best-fit K-value of 2 (Figure 3B). There were 89 accessions in Q1 and 192 accessions in Q2, with cluster values 0.17 and 0.83, respectively. Most of the accessions from Africa were in Q1, while other accessions were in Q2, indicating a clear differentiation between the origin of the African accessions and accessions from other countries. The mean fixation index (F_{st}) was 0.23 and 0.40, while the expected heterozygosity was 0.45 and 0.33 for groups Q1 and Q2, respectively, (Table 1) indicating a high level of genetic variation. The allele frequency divergence (net nucleotide distance) between Q1 and Q2 was 0.10. The accessions in Q2 are sub-divided into two sub-groups, similarly to the cluster and PC analyses.

The model-based analysis of the DArT markers also estimated two groups (Q1 and Q2) (Figure 3C), which was supported by the Evanno criterion best-fit K-value of 2 as the highest population structure (Figure 3D). There were 227 accessions in group Q1 and 54 accessions in Q2 at cluster values of 0.32 and 0.63, respectively. Most of the accessions from Africa were in Q2, while accessions from countries outside Africa were in Q1. The mean fixation index was 0.32 and 0.19, while the expected heterozygosity was 0.27 and 0.30 for groups Q1 and Q2, respectively (Table 1). The allele frequency divergence (net nucleotide distance) between Q1 and Q2 was 0.09. The accessions in Q1 were sub-divided into two sub-groups corresponding to the clustering and PC analyses.

2.3. Analysis of Molecular Variance

The analysis of molecular variance (AMOVA) based on SNP markers revealed that 97.00% of the genetic variance was within the population and 3.00% was among populations (Figure S3a). The PhiPT value of the estimated population structures was 0.03 ($p < 0.01$), suggesting a high level of difference, and the number of migrants (N_m) was 18.02 (Table 2). The DArT markers revealed that 98.00% of the genetic variance was within the population and 2.00% was attributable to variance among populations (Figure S3b). The PhiPT value was 0.025 ($p < 0.01$), while the N_m was 19.54 (Table 3).

2.4. Allelic Patterns across the Populations

Based on SNP markers, Shannon's diversity index was used to quantify the genetic diversity of the 281 soybean accessions. The Shannon's diversity index ranged from 0.13 in population 4 (Chile) to 0.52 in populations 5 and 13 (China and USA, respectively) (Table S2). The lowest number of effective alleles was 1.19 in population 4 (Chile) and the highest was 1.59 in population 5 (China), followed by 1.58 in population 13 (USA). The highest number of alleles (N_a) was 2.00 in population 13 followed by 1.98 in population 5. The diversity (h) was highest in population 5 (0.34) followed by population 13 (0.34), while the unbiased diversity was highest in population 5 (0.35).

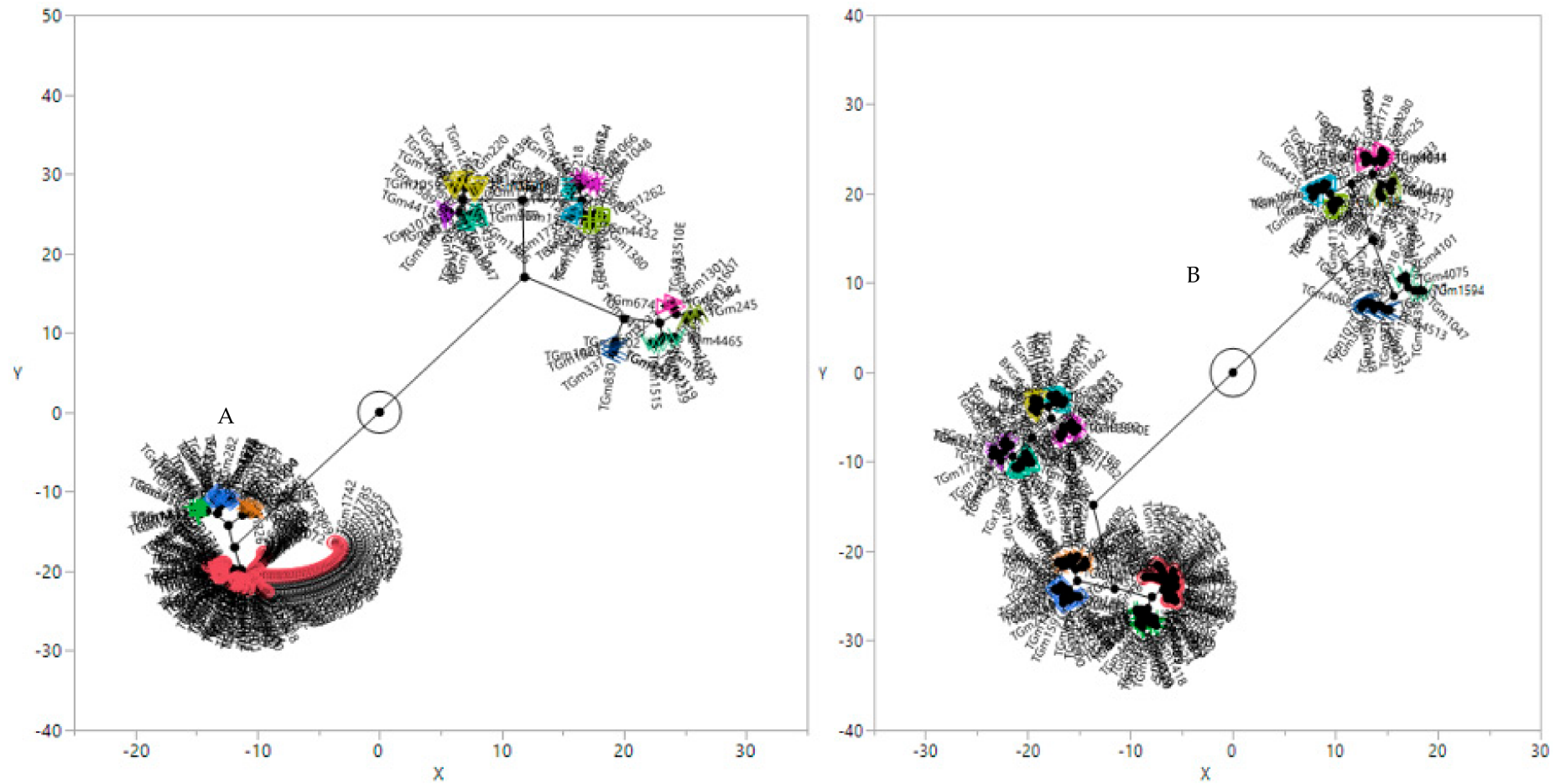


Figure 1. Cluster analysis of the 281 soybean genotypes based on unweighted pair-group method with arithmetic means: (A) SNP markers and (B) DArT markers. The different colors indicate the different clusters identified.

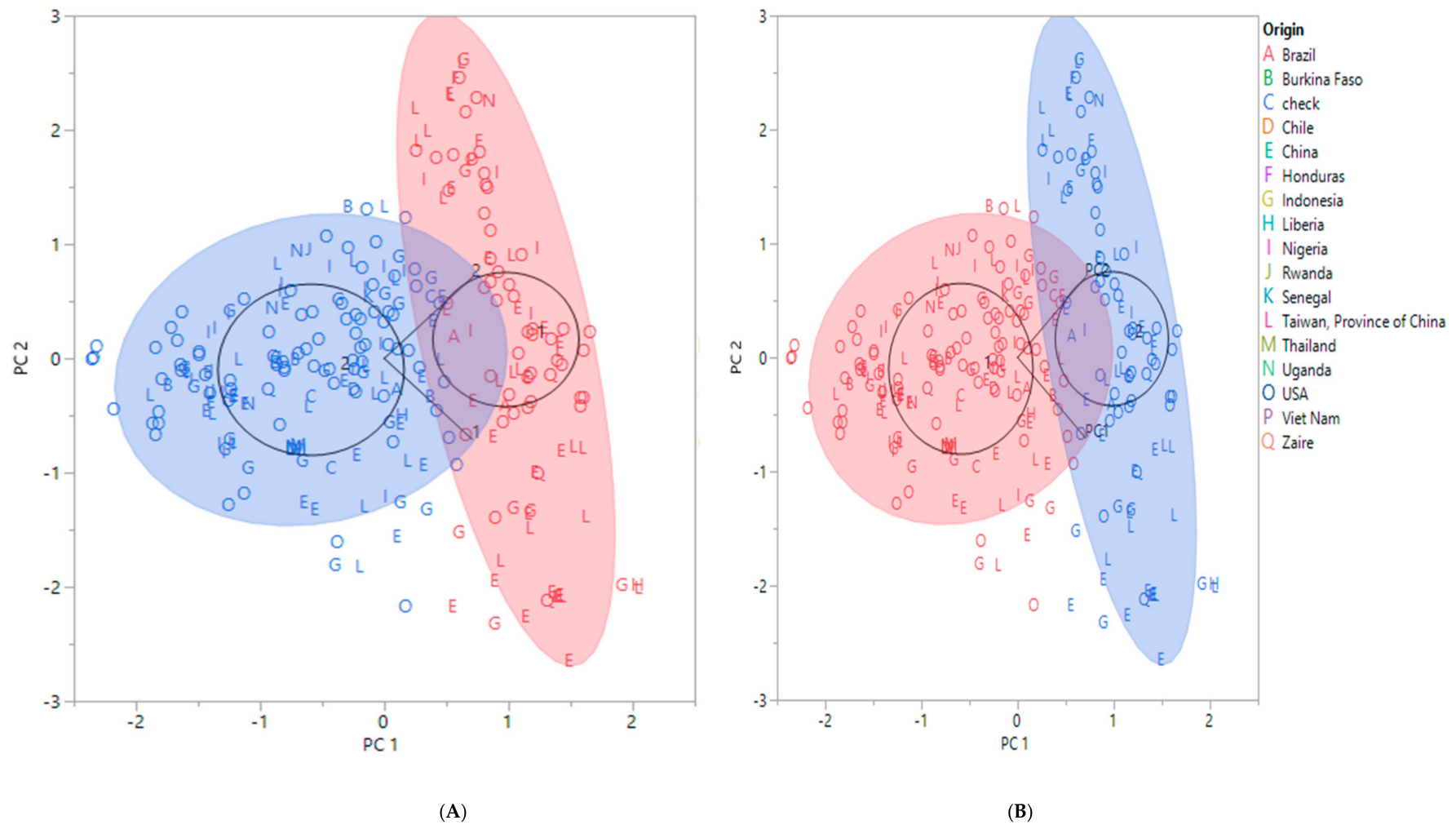


Figure 2. Principal component analysis biplot showing the first two components: (A) SNP markers and (B) DArT markers. The blue- and red-filled circles indicate the two groups identified. The black circles also represent the two groups and show the labels of the groups (1 and 2). The black line indicates the sub-group.

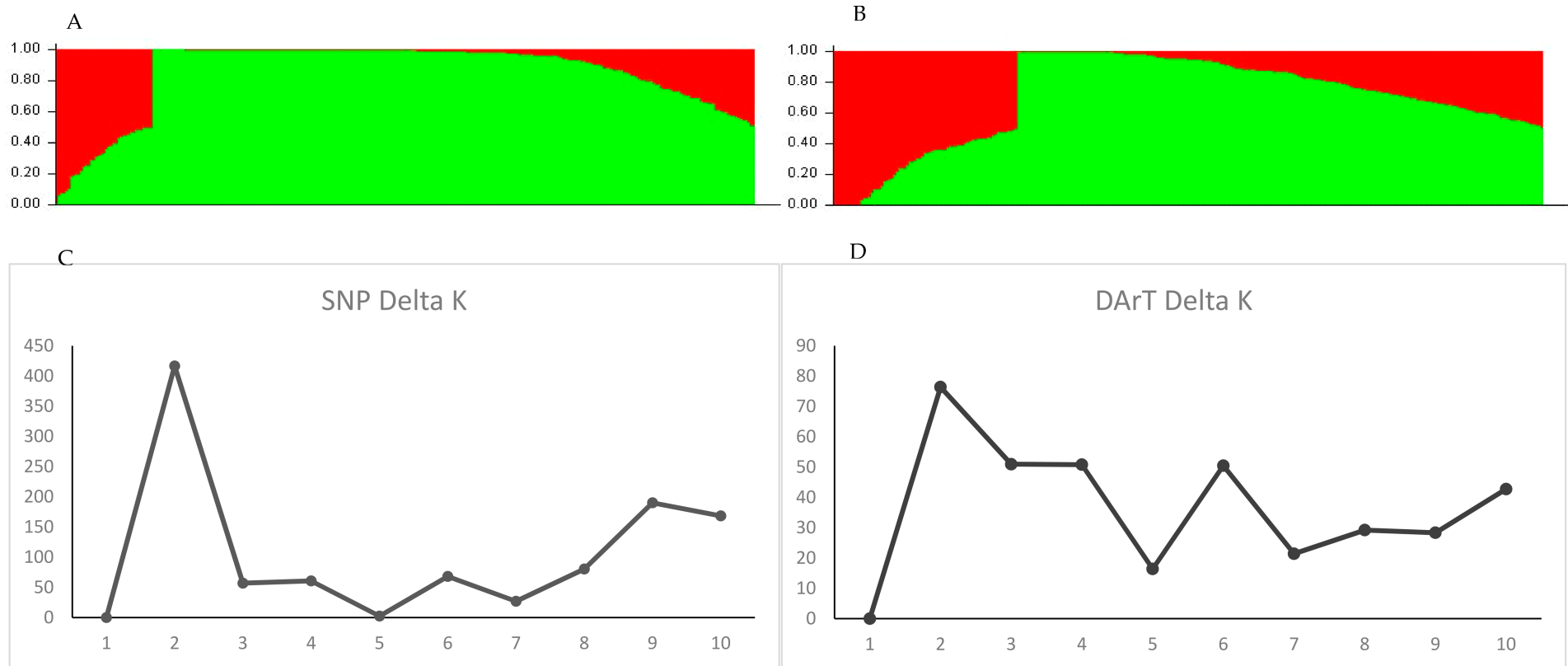


Figure 3. Population structure of the soybean accessions: (A) summary plot of estimates of Q for SNP markers; (B) rate of change of the likelihood (ΔK) for SNP markers; (C) summary plot of estimates of Q for DArT markers; (D) rate of change of the likelihood (ΔK) for DArT markers. The ΔK shows a clear peak at the true value of K.

Table 1. Model-based Bayesian analysis of the SNP and DArT markers.

| | Cluster Value | Mean Value of <i>Fst</i> | Expected Heterozygosity | Net Nucleotide Distance |
|------|---------------|--------------------------|-------------------------|-------------------------|
| SNP | | | | |
| Q1 | 0.17 | 0.23 | 0.45 | 0.10 |
| Q2 | 0.83 | 0.40 | 0.33 | 0.10 |
| DArT | | | | |
| Q1 | 0.32 | 0.32 | 0.27 | 0.09 |
| Q2 | 0.63 | 0.19 | 0.30 | 0.09 |

Fst = Fixation index.

Table 2. Analysis of molecular variance (AMOVA) for 281 soybean germplasms based on SNP markers.

| Source | df | SS | MS | Estimated Variance | % |
|-----------------|-------|---------------------|----------|--------------------|------|
| Among Pops | 13 | 27,640.82 | 2126.227 | 40.45 | 3% |
| Within Pops | 267 | 391,863.18 | 1467.65 | 1467.65 | 97% |
| Total | 280 | 419,504.00 | | 1508.10 | 100% |
| Stat | Value | P(rand \geq data) | | | |
| PhiPT | 0.027 | 0.001 | | | |
| Nm (Haploid) | 18.02 | | | | |

Nm = number of migrants.

Table 3. Analysis of molecular variance (AMOVA) for 281 soybean germplasms based on DArT markers.

| Source | df | SS | MS | Estimated Variance | % |
|-----------------|-------|---------------------|---------|--------------------|------|
| Among Pops | 13 | 15,980.95 | 1229.30 | 22.21 | 2% |
| Within Pops | 267 | 231,682.11 | 867.72 | 867.72 | 98% |
| Total | 280 | 247,663.06 | | 889.93 | 100% |
| Stat | Value | P(rand \geq data) | | | |
| PhiPT | 0.025 | 0.001 | | | |
| Nm (Haploid) | 19.54 | | | | |

Nm = number of migrants.

Based on the DArT markers, Shannon's diversity index ranged from 0.09 in population 11 (Thailand) to 0.64 in populations 10 and 13 (Taiwan, Province of China and USA, respectively), followed by 0.62 in population 5 (China) (Table S2). The lowest number of effective alleles was 1.29 in population 9 (Senegal), and the highest was 1.71 in population 10 (Taiwan, Province of China) followed by 1.69 in populations 5 and 13 (China and USA, respectively). The N_a was highest at 2.77 in populations 5 and 10. The value of h was highest in populations 5, 10 and 13 (0.37), while u_h was highest in population 7 (0.42).

3. Discussion

Studies of genetic diversity of the IITA soybean accessions are important for conservation and utilization. These studies will assist soybean breeders in the development of suitable procedures in attaining wide diversity in soybean breeding programs in Africa [21]. Analyses of the genetic diversity and structure within a population of individuals are performed to determine the magnitude of genetic variations and their usage in breeding programs, e.g., marker-trait associations and linkage analyses. Recently, molecular markers such as the GBS-SNP markers have been used in estimating genetic diversity and have been found to be particularly efficient [15]. Notable among the GBS platforms is DArT, which

generates large numbers of SNPs that are used in genotyping and genetic analyses [22]. Thus, in this study, the genetic diversity and population structure of soybean accessions from different countries were investigated using the DArT GBS technology. The DArT has the advantages of rapidity, low cost and high throughput and has been utilized to study the genetic diversity of some crops [23]. This technology has been applied in a recent study on the genetic diversity of rice, revealing a high genetic variation [24]. The current study is the first report on a successful application of DArT-seq for investigating the genetic diversity of soybean accessions, including those from Africa, using 7244 SNP and 4443 DArT-seq markers. The SNP and DArT markers were found to be effective in describing the genetic diversity of the 281 soybean accessions. Previous studies conducted on the genetic diversity and population structure of soybean used accessions from China and the USA [9], and the Korean peninsula [10].

In the present study, 19,505 SNPs and 16,116 DArT markers were analyzed and filtered to remove redundant markers that may not be functional. After filtering, 7244 polymorphic SNPs and 4443 DArT SNPs were retained and distributed across the 20 chromosomes of the soybean genome. Redundant markers are mostly not useful for specific genetic analyses, as reported previously [25,26]. Furthermore, according to Anderson et al. [27], important reasons for the filtering out of markers are lower variability, large amounts of missing data and genotyping errors. Thus, all the 7244 SNP and 4443 DArT markers retained passed the filtering criteria and were deemed appropriate for assessing the genetic diversity and structure of the studied soybean accessions.

From the SNP marker distribution, chromosome 12 had the least number of polymorphic SNP and DArT markers. Fewer polymorphic markers on chromosomes have been reported to be an indication of lower genetic diversity [24,28]; thus, chromosome 12 was identified as a less diverse chromosome in the soybean accessions. The PIC was used to estimate how informative the SNP and DArT markers were. Approximately 58% and 68% of the SNP and DArT markers, respectively, had PIC values within the range of 0.30–0.50, and they were therefore considered informative, revealing a higher genetic diversity compared to that in a previous report [9]. These results showed the applicability of SNP and DArT markers for better characterization of collections and for other genomic studies in soybean.

The cluster and PC analyses identified two groups that were similar to the model-based population structure analysis for the 281 accessions. The highest level of structure observed ($k = 2$) was similar to the structure reported by [9]. This similarity may be as a result of the predominance of the USA and Chinese accessions in the current study and the previous report. However, a population structure of five sub-groups has been reported [10] in soybean germplasms that comprise mostly South Korean accessions. Thus, it can be inferred that there is a low diversity between the Chinese and USA accessions. Karikari et al. [11] also reported a population structure of soybean comprising USA and Asian accessions with a few accessions from Africa, with four sub-groups. The mean *Fst* values of 0.23 (SNP Q1), 0.40 (SNP Q2) and 0.32 (DArT Q1) showed the existence of genetic diversity within these sub-populations, and parental materials can be selected to develop variable populations for breeding programs. Wright [29], suggested that an *Fst* value of 0.25 is an indication of good differentiation between population groups, from 0.15 to 0.25 is an indication of moderate differentiation and the differentiation is negligible if the *Fst* value is 0.05 or less. The similar grouping of some accessions from different countries indicated their close relatedness through common ancestry but may also be attributed to improper documentation during germplasm exchange; thus, the same accession may be recorded with different origins [30]. Hybridization between accessions in sub-populations 1 and 2 could offer the needed variation for enhancing genetic gain through active selection.

The AMOVA revealed that 97% and 98% of all the genetic variances using SNP and DArT markers were within the population, and only 3% were attributed to population structure (variance among populations). Previous studies on soybeans [9,10] and other crops, such as *Camelina sativa* [31], wheat [28], Bangladesh rice [32] and rice [24] attributed

a larger proportion of genetic variation to individual differences rather than the variance among populations.

The PhiPT value from the AMOVA showed that there were significant differences between the groups (PhiPT = 0.03 and 0.03, $p = 0.001$, for SNP and DArT markers, respectively). These significant differences may be as a result of the low genetic exchange and gene flow which was evident from the low value of the number of migrants. A low value for the number of migrants has been described as evidence of significant differences among populations [28,31]. Furthermore, the differentiation (structures) may be as a result of selection, genetic drift [33,34] and gene exchange [28,31], which will have impacts on the use of these populations in plant breeding programs [28]. In genome-wide association studies, the identification of quantitative trait loci and genomic selections, for assessing the patterns of genetic structure, is vital. The information obtained from such an assessment increases the confidence in the results of the associations detected.

The levels of diversity among population groups have been estimated using indices such as Shannon's information index, number of private alleles, number of different alleles, number of effective alleles, diversity and unbiased diversity [24]. A higher value for these indices indicates a higher level of genetic diversity [28,31]. Based on these indices, the Chinese and USA accessions had higher diversity, while the accessions from Chile had the lowest diversity. In our study, these indices also revealed the existence of genetic diversity among the populations. The level of diversity identified may be attributed to the low gene-flow index [35] and represents valuable resources for future soybean improvement programs. The information revealed in this study could be useful for soybean breeding efforts in Africa.

4. Materials and Methods

4.1. Plant Materials

The plant materials used in this study consisted of 281 soybean accessions from different countries, including improved cultivars (Table S3). The accessions were collected from IITA. Seeds of each of the accessions were packed and sent to SEQART AFRICA, Nairobi, Kenya, for GBS using DArT [36,37]. SEQART AFRICA is a non-profit organization formed from the integration of DArT Pty Ltd., Canberra, Australia and the Biosciences eastern and central Africa (BeCA) ILRI under the Integrated Genotyping Service and Support (IGSS) platform.

4.2. DArT-Based Genotyping by Sequencing

Genotyping of the soybean accessions was performed using the DArT-based GBS platform of SEQART AFRICA. Briefly, genomic DNA was extracted from young, fresh soybean leaves using the NucleoMag Plant genomic DNA extraction kit (Takara Bio, Shiga, Japan). The quality and quantity of DNA in each sample were checked via 0.8% *w/v* agarose gel and a NanoDrop spectrophotometer, respectively. Genomic DNA was digested using the DArT-seq complexity-reduction method and ligated to the barcoded adapters, after which PCR amplification of adapter-ligated fragments was performed. The genotypic analysis and SNP calling were carried out using the DArT-based GBS platform. The procedures for DNA extraction, DNA quality assessment, cleaning, complexity reduction, cloning and library construction have been previously described [37]. Samples were sequenced using a HiSeq 2500 system [36] and the final SNPs were called using the DArTsoft analytical pipeline (<http://www.diversityarrays.com/darttechnologypackage-dartSoft>, accessed on 11 October 2021).

4.3. Statistical Analysis

4.3.1. SNP Filtering and Genome Characterization

The SNP characteristics (missing rate, PIC and frequency of heterozygosity) were generated by the DArT GBS platform. The filtering of the SNPs was performed in two stages; the first stage filtering removed all SNPs that were unmapped to any of the 20 soybean

chromosomes, while the second stage filtering removed the redundant markers. The redundant markers were markers with missing rates of more than 10%, MAF below 5% and PIC < 10%. The genome summary plugin in the TASSEL v.5.2.37 software [38] was used to generate the allele frequency and MAF of the markers.

4.3.2. Cluster and Population Structure Analyses

Cluster and principal component analyses of the 281 tropical soybean accessions were performed in TASSEL v.5.2.37 software [38] using the neighbour-joining algorithm. The grouping of accessions into clusters was achieved using the unweighted pair-group method with arithmetic means (UPGMA) and displayed via the Archaeopteryx Tree plugin.

The structure of the populations was assessed using a model-based (Bayesian) analysis in STRUCTURE v2.3.4 [39]. The k-values generated were from four independent analyses of the fixed number of population sub-groups (1–10). The analysis was based on 100,000 Markov chain Monte Carlo (MCMC) iterations and an initial burn-in period of 10,000, with no previous information on the source of accession. The best-fit k-value for the population was calculated using Structure Harvester [40]. In addition, Structure Harvester was used to generate the CLUMPP files for each k-value based on the log probability of the data [$\ln P(D)$] and derived statistics (ΔK). The number of the sub-populations was determined when the ΔK value reached a maximum [41]. The individual Q-values for the accessions were obtained from the generated CLUMPP file, and accessions with Q-values > 0.5 are in the same group.

4.3.3. Analysis of Molecular Variance

Analysis of molecular variance (AMOVA) was conducted using the results of population structures to determine the variation within and among the populations using GenAlEx 6.5 [42]. Genetic differentiation between populations was determined using the phi-statistics (PhiPT) value. The number of migrants (N_m value) was also estimated. The probability value used to test the significance of the variance was estimated using 1000 standard permutations.

4.3.4. Analysis of Allelic Pattern

The allelic pattern analysis across the population was performed using GenAlEx 6.5 [42]. The Shannon's information index (I), diversity (h), unbiased diversity (uh), number of private alleles (Pa), number of different alleles (Na) and number of effective alleles (Ne) were used to assess the genetic diversity within and among the populations [40].

5. Conclusions

The use of different marker types in evaluating the genetic diversity of crops provides opportunities for a comparison of the marker types and gives a more reliable estimate of the diversity present. In the present study, the estimated genetic diversities were comparable between SNP and DArT markers. The few inconsistencies observed may be a result of the characteristics of the markers. The SNP and DArT markers provided similar evidence regarding the genetic diversity and population structure of the accessions. The DArT GBS platform produced high-density markers that were distributed across the 20 chromosomes of the soybean genome. Both markers revealed two population structures for the 281 soybean accessions. Greater genetic diversity existed within the population than among populations; thus, these populations could be useful in breeding programs aimed at recombining favourable alleles within adapted gene pools for the improvement of grain yield, stress resistance and seed quality traits. Additionally, the introgression of novel genes from diverse sources using advanced genomic techniques such as CRISPR/Cas9 technology, RNA interference (RNAi) and overexpression could be explored to develop more diverse populations. Our study provides an understanding of the genetic structure and diversity of the IITA soybean accessions, which will allow for the efficient utilization of these accessions in soybean improvement programs, especially in Africa.

Our next objectives are to identify promising genotypes and to conduct genome-wide association studies for seed quality traits, yield and biotic and abiotic stresses.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/plants11010068/s1>, Table S1: Quality parameters of the DArT-SNP markers; Figure S1: Distribution of SNPs across the 20 chromosomes of soybean. (a) SNP markers and (b) DArT markers; Figure S2: The polymorphic information content of markers. (a) SNP markers and (b) DArT markers; Figure S3: Percentages of molecular variance within and among populations. (a) SNP markers (b) DArT markers; Table S2: Allelic patterns across the populations for SNP and DArT markers; Table S3: List of soybeans and their origin.

Author Contributions: Conceptualization, A.S.S. and I.B.M.; methodology, A.S.S., H.I., Z.L.M. and H.L.Y.; software, A.S.S.; validation, A.S.S. and L.O.O.; formal analysis, A.S.S.; investigation, H.I., Z.L.M., A.Y.K., S.G.M. and I.B.M.; resources, I.B.M.; data curation, A.S.S.; writing—original draft preparation, A.S.S.; writing—review and editing, I.B.M., A.Y.K., L.O.O. and B.K.; visualization, A.S.S. and H.I.; supervision, A.S.S. and I.B.M.; project administration, I.B.M.; funding acquisition, A.S.S., S.G.M. and I.B.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Nigerian Tertiary Education Trust Fund (TETFUND), grant number TETFund/DR&D/CE/NRF/STI/09/VOL1.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. FAOSTAT Databases: Import and Export Values of Soybean-Based Products; Food Balance Sheets for Soybean 2021. Available online: <https://www.fao.org/faostat/en/#data/QCL> (accessed on 11 December 2021).
2. Bisen, A.; Khare, D.; Nair, P.; Tripathi, N. SSR Analysis of 38 Genotypes of Soybean (*Glycine max* (L.) Merr.) Genetic Diversity in India. *Physiol. Mol. Biol. Plants* **2015**, *21*, 109–115. [[CrossRef](#)]
3. Wilson, R.F. *Designing Soybeans for 21st Century Markets*; Elsevier: Amsterdam, The Netherlands, 2015; ISBN 1630670111.
4. Salgotra, R.K.; Gupta, B.B.; Bhat, J.A.; Sharma, S. Genetic Diversity and Population Structure of Basmati Rice (*Oryza sativa* L.) Germplasm Collected from North Western Himalayas Using Trait Linked SSR Markers. *PLoS ONE* **2015**, *10*, e0131858.
5. Hyten, D.L.; Song, Q.; Zhu, Y.; Choi, I.-Y.; Nelson, R.L.; Costa, J.M.; Specht, J.E.; Shoemaker, R.C.; Cregan, P.B. Impacts of Genetic Bottlenecks on Soybean Genome Diversity. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 16666–16671. [[CrossRef](#)]
6. Kofsky, J.; Zhang, H.; Song, B.-H. The Untapped Genetic Reservoir: The Past, Current, and Future Applications of the Wild Soybean (*Glycine soja*). *Front. Plant Sci.* **2018**, *9*, 949. [[CrossRef](#)] [[PubMed](#)]
7. Lam, H.-M.; Xu, X.; Liu, X.; Chen, W.; Yang, G.; Wong, F.-L.; Li, M.-W.; He, W.; Qin, N.; Wang, B. Resequencing of 31 Wild and Cultivated Soybean Genomes Identifies Patterns of Genetic Diversity and Selection. *Nat. Genet.* **2010**, *42*, 1053–1059. [[CrossRef](#)] [[PubMed](#)]
8. Li, Y.; Zhao, S.; Ma, J.; Li, D.; Yan, L.; Li, J.; Qi, X.; Guo, X.; Zhang, L.; He, W. Molecular Footprints of Domestication and Improvement in Soybean Revealed by Whole Genome Re-Sequencing. *BMC Genom.* **2013**, *14*, 579. [[CrossRef](#)] [[PubMed](#)]
9. Liu, Z.; Li, H.; Wen, Z.; Fan, X.; Li, Y.; Guan, R.; Guo, Y.; Wang, S.; Wang, D.; Qiu, L. Comparison of Genetic Diversity between Chinese and American Soybean (*Glycine max* (L.)) Accessions Revealed by High-Density SNPs.v. *Front. Plant Sci.* **2017**, *8*, 2014. [[CrossRef](#)] [[PubMed](#)]
10. Jeong, S.-C.; Moon, J.-K.; Park, S.-K.; Kim, M.-S.; Lee, K.; Lee, S.R.; Jeong, N.; Choi, M.S.; Kim, N.; Kang, S.-T.; et al. Genetic Diversity Patterns and Domestication Origin of Soybean. *Theor. Appl. Genet.* **2019**, *132*, 1179–1193. [[CrossRef](#)]
11. Karikari, B.; Bhat, J.A.; Denwar, N.N.; Zhao, T. Exploring the Genetic Base of the Soybean Germplasm from Africa, America and Asia as Well as Mining of Beneficial Allele for Flowering and Seed Weight. *3 Biotech* **2020**, *10*, 195. [[CrossRef](#)] [[PubMed](#)]
12. Foyer, C.H.; Siddique, K.H.M.; Tai, A.P.K.; Anders, S.; Fodor, N.; Wong, F.-L.; Ludidi, N.; Chapman, M.A.; Ferguson, B.J.; Considine, M.J.; et al. Modelling Predicts That Soybean Is Poised to Dominate Crop Production across Africa. *Plant. Cell Environ.* **2019**, *42*, 373–385. [[CrossRef](#)] [[PubMed](#)]
13. Cobb, J.N.; DeClerck, G.; Greenberg, A.; Clark, R.; McCouch, S. Next-Generation Phenotyping: Requirements and Strategies for Enhancing Our Understanding of Genotype–Phenotype Relationships and Its Relevance to Crop Improvement. *Theor. Appl. Genet.* **2013**, *126*, 867–887. [[CrossRef](#)]
14. Holbrook, C.C.; Stalker, H.T. Peanut Breeding and Genetic Resources. *Plant Breed. Rev.* **2003**, *22*, 297–356.

15. Fischer, M.C.; Rellstab, C.; Leuzinger, M.; Roumet, M.; Gugerli, F.; Shimizu, K.K.; Holderegger, R.; Widmer, A. Estimating Genomic Diversity and Population Differentiation—An Empirical Comparison of Microsatellite and SNP Variation in Arabidopsis Halleri. *BMC Genom.* **2017**, *18*, 69. [[CrossRef](#)] [[PubMed](#)]
16. Sansaloni, C.; Petroloni, C.; Jaccoud, D.; Carling, J.; Detering, F.; Grattapaglia, D.; Kilian, A. Diversity Arrays Technology (DART) and next-Generation Sequencing Combined: Genome-Wide, High Throughput, Highly Informative Genotyping for Molecular Breeding of Eucalyptus. In Proceedings of the BMC Proceedings; BioMed Central: London, UK, 2011; Volume 5, pp. 1–2.
17. Poland, J.A.; Rife, T.W. Genotyping-by-sequencing for Plant Breeding and Genetics. *Plant Genome* **2012**, *5*, 92–102. [[CrossRef](#)]
18. Chen, J.; Zavala, C.; Ortega, N.; Petroloni, C.; Franco, J.; Burgueño, J.; Costich, D.E.; Hearne, S.J. The Development of Quality Control Genotyping Approaches: A Case Study Using Elite Maize Lines. *PLoS ONE* **2016**, *11*, e0157236.
19. Yang, X.; Ren, R.; Ray, R.; Xu, J.; Li, P.; Zhang, M.; Liu, G.; Yao, X.; Kilian, A. Genetic Diversity and Population Structure of Core Watermelon (*Citrullus lanatus*) Genotypes Using DARTseq-Based SNPs. *Plant Genet. Resour.* **2016**, *14*, 226–233. [[CrossRef](#)]
20. Denwar, N.N.; Awuku, F.J.; Diers, B.; Addae-Frimpomaah, F.; Chigeza, G.; Oteng-Frimpong, R.; Puozaa, D.K.; Barnor, M.T. Genetic Diversity, Population Structure and Key Phenotypic Traits Driving Variation within Soyabean (*Glycine max*) Collection in Ghana. *Plant Breed.* **2019**, *138*, 577–587. [[CrossRef](#)]
21. Chigeza, G.; Boahen, S.; Gedil, M.; Agoyi, E.; Mushoriwa, H.; Denwar, N.; Gondwe, T.; Tesfaye, A.; Kamara, A.; Alamu, O.E.; et al. Public Sector Soybean (*Glycine max*) Breeding: Advances in Cultivar Development in the African Tropics. *Plant Breed.* **2019**, *138*, 455–464. [[CrossRef](#)]
22. Beissinger, T.M.; Hirsch, C.N.; Sekhon, R.S.; Foerster, J.M.; Johnson, J.M.; Muttoni, G.; Vaillancourt, B.; Buell, C.R.; Kaeppler, S.M.; de Leon, N. Marker Density and Read Depth for Genotyping Populations Using Genotyping-by-Sequencing. *Genetics* **2013**, *193*, 1073–1081. [[CrossRef](#)] [[PubMed](#)]
23. Deschamps, S.; Llaca, V.; May, G.D. Genotyping-by-Sequencing in Plants. *Biology* **2012**, *1*, 460–483. [[CrossRef](#)]
24. Adeboye, K.A.; Oyedeji, O.E.; Alqudah, A.M.; Börner, A.; Oduwaye, O.; Adebambo, O.; Daniel, I.O. Genetic Structure and Diversity of Upland Rice Germplasm Using Diversity Array Technology (DART)-Based Single Nucleotide Polymorphism (SNP) Markers. *Plant Genet. Resour.* **2020**, *18*, 343–350. [[CrossRef](#)]
25. Pootakham, W.; Shearman, J.R.; Ruang-Areerate, P.; Sonthirod, C.; Sangsrakru, D.; Jomchai, N.; Yoocha, T.; Triwitayakorn, K.; Tragoonrung, S.; Tangphatsornruang, S. Large-Scale SNP Discovery through RNA Sequencing and SNP Genotyping by Targeted Enrichment Sequencing in Cassava (*Manihot esculenta* Crantz). *PLoS ONE* **2014**, *9*, e116028. [[CrossRef](#)]
26. Yang, G.; Chen, S.; Chen, L.; Sun, K.; Huang, C.; Zhou, D.; Huang, Y.; Wang, J.; Liu, Y.; Wang, H. Development of a Core SNP Arrays Based on the KASP Method for Molecular Breeding of Rice. *Rice* **2019**, *12*, 21. [[CrossRef](#)] [[PubMed](#)]
27. Anderson, C.A.; Pettersson, F.H.; Clarke, G.M.; Cardon, L.R.; Morris, A.P.; Zondervan, K.T. Data Quality Control in Genetic Case-Control Association Studies. *Nat. Protoc.* **2010**, *5*, 1564–1573. [[CrossRef](#)] [[PubMed](#)]
28. Eltahir, S.; Sallam, A.; Belamkar, V.; Emara, H.A.; Nower, A.A.; Salem, K.F.M.; Poland, J.; Baenziger, P.S. Genetic Diversity and Population Structure of F3: 6 Nebraska Winter Wheat Genotypes Using Genotyping-by-Sequencing. *Front. Genet.* **2018**, *9*, 76. [[CrossRef](#)] [[PubMed](#)]
29. Wright, S. *Evolution and the Genetics of Populations, Volume 4: Variability within and among Natural Populations*; University of Chicago Press: Chicago, IL, USA, 1984; Volume 4, ISBN 0226910415.
30. Zhang, S.; Li, B.; Chen, Y.; Shaibu, A.S.; Zheng, H.; Sun, J. Molecular-Assisted Distinctness and Uniformity Testing Using SLAF-Sequencing Approach in Soybean. *Genes* **2020**, *11*, 175. [[CrossRef](#)]
31. Luo, Z.; Brock, J.; Dyer, J.M.; Kutchan, T.; Schachtman, D.; Augustin, M.; Ge, Y.; Fahlgren, N.; Abdel-Haleem, H. Genetic Diversity and Population Structure of a *Camelina sativa* Spring Panel. *Front. Plant Sci.* **2019**, *10*, 184. [[CrossRef](#)]
32. Islam, M.Z.; Khalequzzaman, M.; Prince, M.; Siddique, M.A.; Rashid, E.; Ahmed, M.S.U.; Pittendrigh, B.R.; Ali, M.P. Diversity and Population Structure of Red Rice Germplasm in Bangladesh. *PLoS ONE* **2018**, *13*, e0196096. [[CrossRef](#)]
33. Buckler IV, E.S.; Thornsberry, J.M. Plant Molecular Diversity and Applications to Genomics. *Curr. Opin. Plant Biol.* **2002**, *5*, 107–111. [[CrossRef](#)]
34. Breseghello, F.; Sorrells, M.E. Association Mapping of Kernel Size and Milling Quality in Wheat (*Triticum aestivum* L.) Cultivars. *Genetics* **2006**, *172*, 1165–1177. [[CrossRef](#)]
35. Arora, A.; Kundu, S.; Dilbaghi, N.; Sharma, I.; Tiwari, R. Population Structure and Genetic Diversity among Indian Wheat Varieties Using Microsatellite (SSR) Markers. *Aust. J. Crop Sci.* **2014**, *8*, 1281–1289.
36. Kilian, A.; Wenzl, P.; Huttner, E.; Carling, J.; Xia, L.; Blois, H.; Caig, V.; Heller-Uszynska, K.; Jaccoud, D.; Hopper, C.; et al. Diversity Arrays Technology: A Generic Genome Profiling Technology on Open Platforms. *Methods Mol. Biol.* **2012**, *888*, 67–89. [[CrossRef](#)] [[PubMed](#)]
37. Shaibu, A.; Sneller, C.; Motagi, B.; Chepkoech, J.; Chepngetich, M.; Miko, Z.; Isa, A.M.; Ajeigbe, H.A.; Mohammed, S.G. Genome-Wide Detection of SNP Markers Associated with Four Physiological Traits in Groundnut (*Arachis hypogaea* L.) Mini Core Collection. *Agronomy* **2020**, *10*, 192. [[CrossRef](#)]
38. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for Association Mapping of Complex Traits in Diverse Samples. *Bioinformatics* **2007**, *23*, 2633–2635. [[CrossRef](#)] [[PubMed](#)]
39. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **2000**, *155*, 945–959. [[CrossRef](#)]

40. Earl, D.A. STRUCTURE HARVESTER: A Website and Program for Visualizing STRUCTURE Output and Implementing the Evanno Method. *Conserv. Genet. Resour.* **2012**, *4*, 359–361. [[CrossRef](#)]
41. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the Number of Clusters of Individuals Using the Software STRUCTURE: A Simulation Study. *Mol. Ecol.* **2005**, *14*, 2611–2620. [[CrossRef](#)]
42. Peakall, R.O.D.; Smouse, P.E. GENALEX 6: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research. *Mol. Ecol. Notes* **2006**, *6*, 288–295. [[CrossRef](#)]