





Article

Genotyping-by-Sequencing to Unlock Genetic Diversity and Population Structure in White Yam (*Dioscorea rotundata* Poir.)

Ranjana Bhattacharjee ^{1,*}, Paterne Agre ¹, Guillaume Bauchet ², David De Koeyer ³, Antonio Lopez-Montes ^{1,4}, P. Lava Kumar ¹, Michael Abberton ¹, Patrick Adebola ⁵, Asrat Asfaw ⁵ and Robert Asiedu ¹

¹ International Institute of Tropical Agriculture, PMB 5320, Ibadan 200001, Nigeria; P.Agre@cgiar.org (P.A.); mijuamarel@gmail.com (A.L.-M.); l.kumar@cgiar.org (P.L.K.); M.Abberton@cgiar.org (M.A.); r.asiedu@cgiar.org (R.A.)

² Boyce Thompson Institute, Cornell University, Ithaca, NY 14853, USA; gjb99@cornell.edu

³ Agriculture and Agri-Food Canada, Fredericton, NB E3B 4Z7, Canada; david.dekoeyer@canada.ca

⁴ International Trade Center, 29 Independence Avenue, Accra 00233, Ghana

⁵ International Institute of Tropical Agriculture, PMB 82, Kubuwa, Abuja 901101, Nigeria; p.adebola@cgiar.org (P.A.); a.amele@cgiar.org (A.A.)

* Correspondence: r.bhattacharjee@cgiar.org

Received: 7 August 2020; Accepted: 14 September 2020; Published: 22 September 2020



Abstract: White yam (*Dioscorea rotundata* Poir.) is one of the most important tuber crops in West Africa, where it is indigenous and represents the largest repository of biodiversity through several years of domestication, production, consumption, and trade. In this study, the genotyping-by-sequencing (GBS) approach was used to sequence 814 genotypes consisting of genebank landraces, breeding lines, and market varieties to understand the level of genetic diversity and pattern of the population structure among them. The genetic diversity among different genotypes was assessed using three complementary clustering methods, the model-based admixture, discriminant analysis of principal components (DAPC), and phylogenetic tree. ADMIXTURE analysis revealed an optimum number of four groups that matched with the number of clusters obtained through phylogenetic tree. Clustering results obtained from ADMIXTURE analysis were further validated using DAPC-based clustering. Analysis of molecular variance (AMOVA) revealed high genetic diversity (96%) within each genetic group. A network analysis was further carried out to depict the genetic relationships among the three genetic groups (breeding lines, genebank landraces, and market varieties) used in the study. This study showed that the use of advanced sequencing techniques such as GBS coupled with statistical analysis is a robust method for assessing genetic diversity and population structure in a complex crop such as white yam.

Keywords: white yam; West Africa; SNP markers; genetic diversity; population structure

1. Introduction

White yam or white guinea yam (*Dioscorea roundata* Poir.) is one of the most important staple tuber crops of West Africa [1], which is strongly associated with the food security, income, and social culture of >300 million people of this region, having a net value internationally of ≈\$15 billion [2]. It belongs to the section Enantiophyllum and Dioscoreaceae family consisting of approximately 600 species distributed in the tropical and subtropical regions of the world [3]. *D. rotundata* is an allogamous, polyploid, dioecious species with a basic chromosome number of $n = x = 20$. The ‘yam belt’ consisting of six countries in West and Central Africa including Nigeria, Ghana, Bénin, Côte d’Ivoire, Togo,

and Cameroon accounts for 97% of the total yam production of the world, while Nigeria alone accounts for 68% of global production [2]. In this region, there is the occurrence of several *Dioscorea* spp., including white yam. In West and Central Africa, where white Guinea yams were domesticated about 7000 years ago [4], farmers selected genotypes that suited their needs such as food, farming, etc. and thus generated a large number of traditional cultivars. Hence, white yam domestication took advantage of a huge reservoir of diversity in this region as a result of years of large production, consumption, and trade [5–8].

Several studies have been carried out to determine the phylogenetic relationships of Guinea yam (*D. rotundata*–*D. cayenensis*) using morphological features and molecular markers. The morphological characterization of guinea yams from Benin and Cameroon classified the accessions as *D. rotundata*, *D. cayenensis*, and *D. rotundata*–*D. cayenensis* groups [9]. However, the use of morphological features has remained a challenge to biosystematics for many years because of the limitations related with the continuous variation and plasticity of most of the features, making them less informative for species identification [5]. Furthermore, first-generation molecular markers such as isozymes [10–12], AFLPs [13], RAPDs [7,14,15], RFLPs [16] and SSRs [1,17,18] have also been used to assess taxonomic and phylogenetic diversity in cultivated and wild guinea yams. However, such assessment suggested the cultivars/accessions of *D. cayenensis* as a separate taxon from that of the *D. rotundata* [18]. Meanwhile, a recent study on guinea yam collections from Ethiopia using microsatellites suggested no clear distinction between cultivated and wild *Dioscorea* species [19]. In fact, there are very few studies that used morphological or molecular markers to assess genetic diversity in white yam per se [1,20,21].

In recent years, several next-generation sequencing (NGS)-based protocols have been developed for the discovery and generation of large sets of genome-wide SNP (single nucleotide polymorphism) markers for genetic diversity studies, linkage mapping, genomic selection, and QTL (quantitative trait loci) analysis [22]. NGS-based SNPs, due to their genome-wide abundance, are currently the most widely used molecular markers for germplasm characterization as well as the quantification of ancestry of cultivars [23]. Genotyping-by-sequencing (GBS) has emerged as one of the most inexpensive NGS-based genotyping platforms that allows for a high level of multiplexing and high marker density to reveal the extent of genetic relatedness and genetic variation within and between cultivated and wild species [24–27]. The GBS approach reduces the genome complexity using restriction enzymes for high-density SNP markers discovery [28]. In addition, the SNP calling and bioinformatics pipelines are well established and publicly available [29,30]. Next-generation sequencing (GBS-based/DArTseq-based) has been successfully applied in non-model species such as guinea yam [31], water yam [32–34], and trifoliolate yam [35], which demonstrated the suitability of this method for the high-throughput genotyping in yams. In the present study, GBS was used to characterize white/white Guinea yam (*D. rotundata*.) landraces, breeding lines, and market varieties to understand the genetic relationships and population structure for further improvement of this very important species in West and Central Africa.

2. Materials and Methods

2.1. Plant Materials

A total of 814 genotypes were used in this study that included 473 genebank landraces selected from the revised yam core collection [36], 314 breeding lines, and 27 popular market varieties or landraces collected from different markets across Nigeria and Ghana (Table S1). The genebank landraces and breeding lines were collected from the Genetic Resources Center (GRC) and from the Yam Breeding Unit (YBU) of the International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria, respectively. Of the 462 genebank landraces collected and conserved at the Genetic Resources Center at IITA-Ibadan, Nigeria, 198 were from Nigeria and 166 were from Togo, while other countries were Benin (32), Ghana (29), Cote d'Ivoire (27), Guinea (7), and 1 accession each were from Burkina Faso, Equatorial Guinea, and Sierra Leone (Table S1).

2.2. DNA Extraction and GBS

Genomic DNA was extracted from 100 mg of fresh young leaves using the Qiagen DNeasy Plant Mini kit (Qiagen, Germantown, MD, USA) following the manufacturer's protocol. A Nanodrop 8000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) was used to measure the quality of the DNA by comparing the 260 and 280 nm absorptions. DNA samples were further quantified using the Quant-iT™ PicoGreen® dsDNA assay kit (Invitrogen, Carlsbad, CA, USA) and diluted to 50 ng/μL with 1 × TE buffer. About 30 μL of DNA for each genotype was sent in 96-well plates to the Institute for Genomic Diversity (IGD) at Cornell University, Ithaca, New York, where GBS was done using a 96-plex *Pst* I GBS protocol [29]. In brief, for each library, purified genomic DNA was first digested with the restriction enzyme *Pst*I (New England Biolabs, Whitby, ON, Canada), and the ligation of customized adapters (barcodes) with T4 ligase was subsequently carried out. This was followed by PCR with flow-cell attachment site tagged primers. Single-end sequencing was performed using an Illumina HiSeq2000 (Illumina Inc. San Diego, CA, USA).

2.3. Processing of Illumina Raw Sequence Read Data, SNP Calling, and Filtering

The raw sequencing reads (of read length 1 × 100 bp) containing the barcode were sorted, de-multiplexed, and trimmed to the first 64 bases starting from the enzyme cut site. All the reads containing 'N' within the first 64 bases and tags with less than 64 bases were removed. We used the first draft of the *Dioscorea rotundata* reference genome [37] and Bowtie2 [38] to align the sequencing reads. We implemented the best practices of the GATK pipeline to call SNPs [39] GATK 2.4 version and GATK-UnifiedGenotyper were used in this study for the SNP calling. We used multi-sample variant calling by GATK-UnifiedGenotyper considering the large number of samples involved in the study. The SAM files generated from the alignment were converted to BAM format and sorted by name using SAMtools. The final variant calling was generated through GATK (2.4) (using HaplotypeCaller in the gVCF mode) and joint genotyping (using GenotypeGVCFs). The VCF file developed was filtered for using criteria of MAF (minor allele frequency) >0.05 and missing data >80% both at the genotypes and SNP markers level. Only bi-allelic SNP markers with genotype quality >20 and read depth >5 were retained after using Vcftools v.0.1.12b [40] and PLINK v1.07 [41] for filtering. The resulting SNPs were subjected to linkage disequilibrium (LD) and SNP markers in LD were removed, and a total of 3432 SNP markers were retained for all subsequent analysis.

2.4. Population Structure, Genetic Diversity, and Relationships

The filtered SNP data were used to assess the population structure and genetic diversity in white yam. A set of parametric and non-parametric methods including the model-based maximum likelihood estimation of ancestral subpopulations using ADMIXTURE [42], assumption-free discriminant analysis of principal components (DAPC) [43,44] and fixation index (*F*_{st})-based population differentiation were used for analysis. Genetic diversity analysis was carried out using parameters such as minor allele frequency (MAF), polymorphic information content (PIC), expected heterozygosity (*H*_E), and observed heterozygosity (*H*_O) using R [45]. Additionally, genetic diversity was assessed by calculating Shannon–Weaver (*H'*), Simpson, inverse Simpson, and Pielou's evenness indices [46] for the three genetic categories (genebank landraces, breeding lines, and market varieties) using Vegan package in R [47].

A pairwise identity by state (IBS) genetic distance matrix was calculated from 3432 SNP markers using PLINK, and a critical distance threshold [48] was used to declare whether two genotypes are identical based on the pairwise distances between them. Then, the dissimilarity matrix was used to construct the network relationships among the three genetic groups using QGRAPH [49] implemented in R. The aim was to detect important key nodes representing genetic relationships between two genetic groups (breeding lines–market varieties, genebank landraces–breeding lines, and genebank landraces–market varieties) in the network. The results from ADMIXTURE was

complemented with DAPC analysis using the R package ‘adegenet’ [50] in a two-step process. First, the optimal number of clusters was inferred using k-means analysis [51,52] of PCA (principal component analysis)-transformed genome-wide SNP data by varying the possible number of clusters from one to 100. Bayesian Information Criterion (BIC) was used to assess the best supported model as well as the number and nature of clusters. DAPC scatter plots were later developed on the clusters identified through k-means using the first 50 principal components. The information based on ADMIXTURE analysis was used to determine the most appropriate K, and accessions with membership proportions (Q-value) $\geq 70\%$ were assigned to groups, while those with membership probabilities less than 70% were declared as admixtures. Then, the results from DAPC analysis and ADMIXTURE were compared. The coefficient of genetic differentiation among the population was calculated based on pairwise F_{st} (fixation index) to estimate the genetic distance and the relationships among the three genetic groups of *D. rotundata* used in the study. Analysis of molecular variance (AMOVA) was also conducted to assess the population differentiation among the three genetic groups used in the study. Then, a phylogenetic tree was built by following the procedure of IBS with 1000 bootstrap replicates in PowerMarker v3.25 [53]. The resulting tree was visualized in Molecular Evolutionary Genetics Analysis (MEGA) software version X [54].

3. Results

3.1. SNP Summary

The FastQ files of the generated sequences were aligned to the *D. rotundata* reference genome [37], of which 43.4% tags aligned uniquely to the reference, 9.8% aligned to multiple positions, and 46.8% did not successfully align. Uniquely aligned tags were used for calculating the distribution of tag density at each position in *D. rotundata* genome and for SNP distribution. A total of 137,800 unfiltered SNPs was detected as raw SNP markers. A total of 3432 filtered SNPs were obtained and distributed across the twenty-one pseudo chromosomes of *D. rotundata* (Table 1; Figure 1). The genome-wide SNP density plot (Figure 1) revealed that the highest number of SNPs was in chromosome 5 (11.1%, 380 SNPs), while the lowest number of SNPs was mapped in chromosome 11 (2.6%, 90 SNPs). Then, the transition and transversion SNPs were calculated. Transition SNPs (66.9%, 2296 SNPs) were more frequent than transversions (33.1%, 1136 SNPs). The C/T transitions (34.3%) accounted for the highest frequency, while C/G transversions (5.1%) occurred at the lowest frequency among all the 3432 SNPs (Figure 2). The average PIC value across all the markers was 0.135, while the observed heterozygosity ranged from 0.138 to 0.190 with an average of 0.165 (Table 1). The expected heterozygosity ranged between 0.133 and 0.190, and the mean was 0.161. Similarly, the minor allele frequency (MAF) ranged between 0.090 and 0.133 with an average of 0.111.

Table 1. SNP marker summary statistics across twenty-one chromosomes.

Chromosome	No. of SNPs	PIC	MAF	Ho	He
1	81	0.143	0.119	0.179	0.172
2	144	0.131	0.105	0.154	0.156
3	147	0.147	0.121	0.173	0.176
4	271	0.146	0.123	0.184	0.175
5	380	0.135	0.112	0.168	0.161
6	179	0.122	0.098	0.138	0.144
7	117	0.133	0.104	0.156	0.157
8	273	0.151	0.126	0.184	0.180
9	110	0.132	0.113	0.165	0.159
10	123	0.152	0.126	0.190	0.182
11	90	0.130	0.090	0.139	0.133

Table 1. Cont.

Chromosome	No. of SNPs	PIC	MAF	Ho	He
12	118	0.123	0.106	0.162	0.154
13	106	0.121	0.093	0.134	0.142
14	185	0.127	0.102	0.154	0.150
15	165	0.146	0.122	0.181	0.175
16	193	0.120	0.098	0.139	0.142
17	186	0.151	0.131	0.182	0.182
18	153	0.127	0.103	0.161	0.150
19	188	0.159	0.133	0.190	0.191
20	120	0.133	0.106	0.172	0.158
21	103	0.124	0.097	0.142	0.145
Total/Average	3432	0.135	0.111	0.165	0.161

PIC: polymorphic information content; MAF: minor allele frequency; Ho: observed heterozygosity; He: expected heterozygosity.

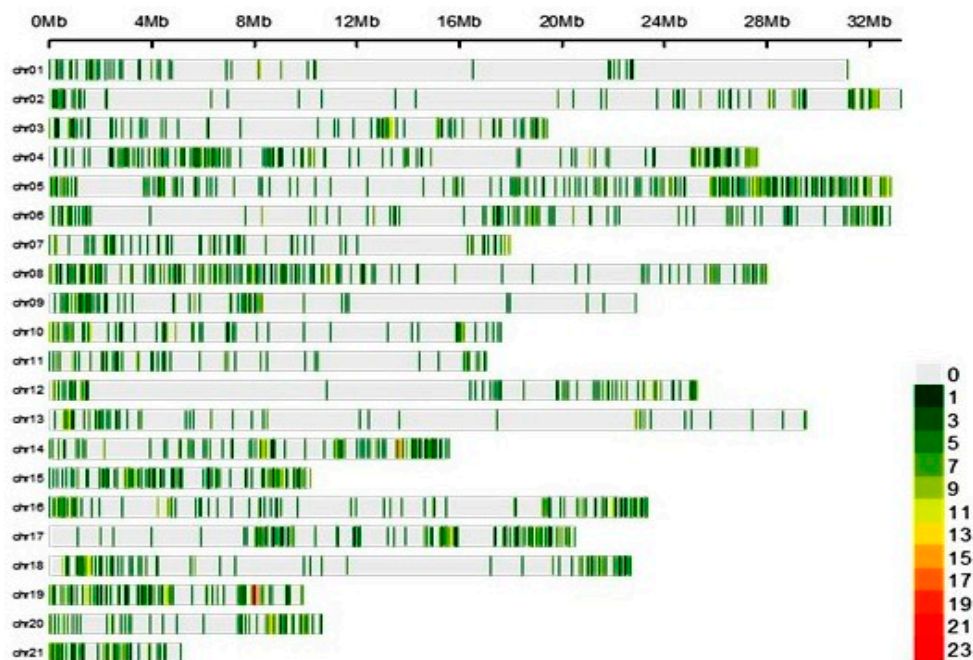


Figure 1. Distribution and density of filtered single nucleotide polymorphism (SNPs) across 21 pseudo chromosomes, as suggested by Tamiru et al. [38]. The horizontal axis displays the chromosome length. The number of SNPs in a given region is indicated at the bottom right.

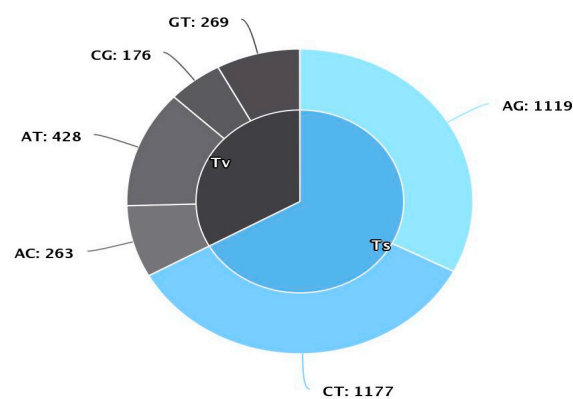


Figure 2. Transition and transversion based on bi-allelic SNP markers. Tv: Transversions; Ts: Transitions; A: Adenine; T: Thymine; G: Guanine; C: Cytosine. Chart developed using SNIPLAY software [55].

3.2. Population Structure and Genetic Diversity

Based on missing data (>20%), 11 genebank landraces of geographical origin of Nigeria were removed from further analysis. Among the 803 accessions analyzed, 314 were breeding lines that were generated through open pollination and bi-parental crossing over several years by the Yam Breeding Unit at IITA-Ibadan, Nigeria. The details are available in Table S1. The majority of the bi-parental crosses were made using a limited number of parental lines (see Table S1), and this could be attributed to the flowering behavior (dioecious, shy female flowering, limited seed set, among others) in white yam [56]. Being clonally propagated, the progenies of bi-parental crosses in white yam represented a segregating population (F2) and were genetically different from each other. This is evident from the grouping of progenies derived from same bi-parental crosses in different clusters (Table S1).

The genetic distance among 803 genotypes varied between 0 and 0.27 (Table S2). Based on the genetic estimation, two accessions were considered identical or representative of the same clone if their pairwise genetic distance was lower than 0.02. Based on this criterion, a total of 767 unique genotypes were recorded. To understand the pattern of population structure, a Bayesian Information Criterion (BIC) and complementary coordination analysis by DAPC were performed. The BIC results suggested the best clustering at $K = 2$ (with a probability of cluster membership assignment of 100) based on delta K values (Figure 3A, Table S3). Clusters 1 and 2 consisted of 739 (309 breeding lines, 403 genebank landraces, and 27 market varieties) and 64 genotypes (five breeding lines and 59 genebank landraces) (Figure 3B, Table S3), respectively. DAPC analysis was further carried out to assess the subclusters at $K = 3$ (Figure 3C), $K = 4$ (Figure 3D), $K = 5$ (Figure 3E), and $K = 12$ (Figure 3F). The summary of DAPC cluster grouping and probability of cluster membership assignment of genotypes at $K = 2, 3, 4, 5,$ and 12 is presented in Table S2. Based on the probability of cluster membership assignment, DAPC clusters both at $K = 2$ and $K = 3$ represented a good fit. At $K = 3$, Cluster 1 consisted of 67 genotypes including five breeding lines, 59 genebank landraces majorly representing two countries such as Nigeria (28 landraces) and Togo (21 landraces), and three market varieties (Tables S1 and S3). Cluster 3 was the largest, consisting of 654 genotypes (genebank landraces: 372; breeding lines: 282) while Cluster 2 consisted of 82 genotypes representing 31 genebank landraces, 27 breeding lines, and 24 market varieties (Table S3). A comparative analysis of genetic diversity among the three genetic groups (genebank landraces, breeding lines, and market varieties) revealed that the PIC (0.141), H_o (0.173), and H_E (0.168) values were relatively higher for genebank landraces while they were the lowest for the market varieties (Table 2). Among the different genetic groups of *D. rotundata*, the Shannon–Weaver index and Simpson’s index were the highest for genebank landraces, while the highest Pielou’s evenness value was for the market varieties (Table 2).

A significant level of population divergence based on pairwise F_{st} ($p < 0.0001$) was also observed between different genetic groups (breeding lines, genebank landraces, and market varieties, while strong genetic relationships were observed within each group (Table 3). The F_{st} -based population differentiation was highest among breeding lines and genebank landraces (0.038), and it was the minimum between genebank landraces and market varieties (0.024). The AMOVA analysis revealed that the variability was divided into 96% within genetic groups and 4% between the three genetic groups (Table 4).

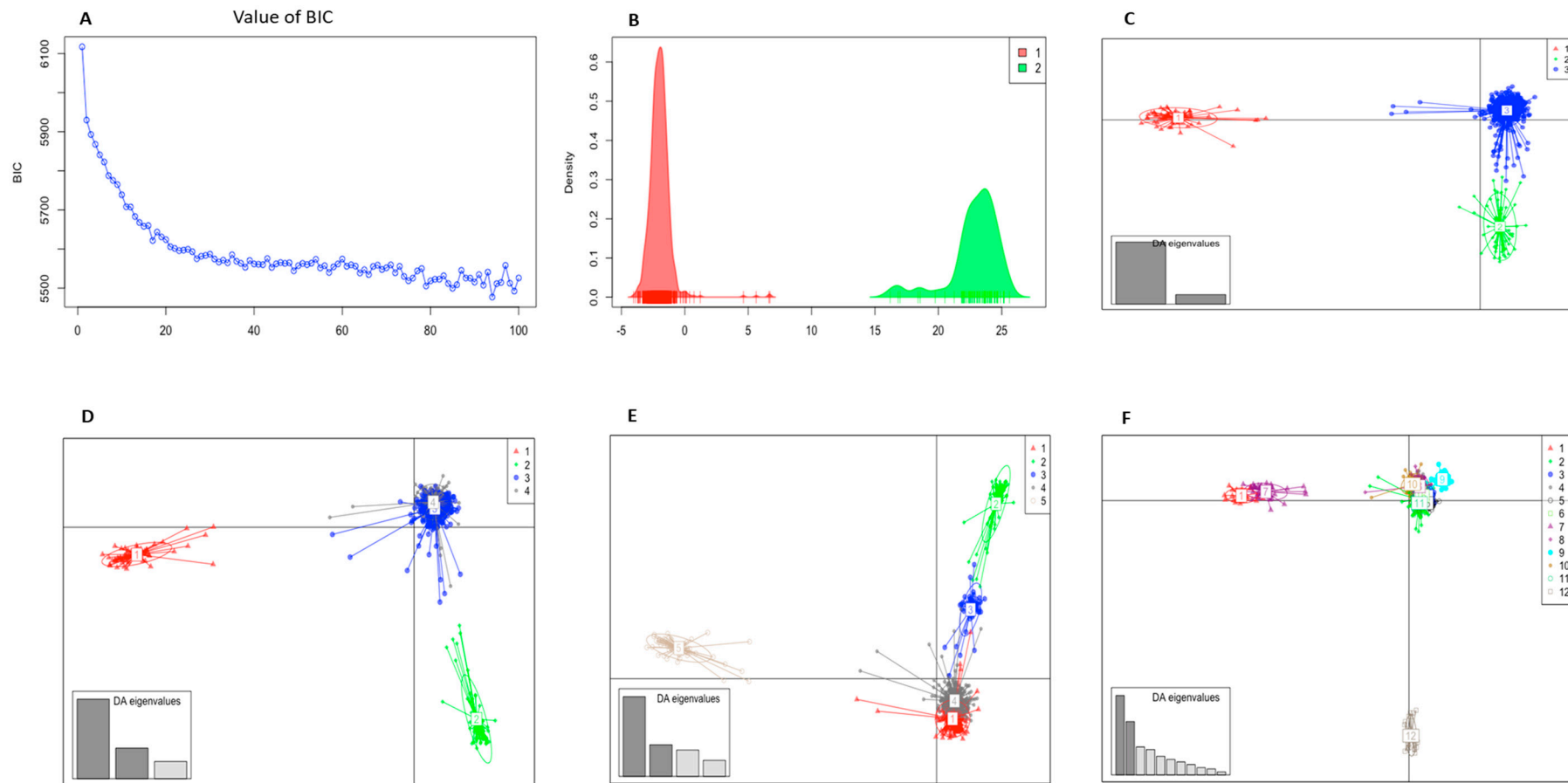


Figure 3. (A) Graph representing number of clusters vs. Bayesian Information Criterion (BIC). In the x-axis, a different number of clusters that could be considered in the population are presented. In the y-axis, the BIC value associated with each number of clusters is presented. (B) Discriminant analysis of principal components (DAPC) with $K = 2$. (C) Discriminant analysis of principal components (DAPC) with $K = 3$. (D) Discriminant analysis of principal components (DAPC) with $K = 4$. (E) Discriminant analysis of principal components (DAPC) with $K = 5$. (F) Discriminant analysis of principal components (DAPC) with $K = 12$. The axes represent the first two linear discriminants. Each color represents a cluster and each dot represents an individual.

Table 2. Comparison of genetic diversity parameters among the three genetic groups of *D. rotundata* used in the study.

Summary Statistics				
Genotypes	PIC	MAF	Ho	He
All	0.135	0.111	0.165	0.161
Breeding line	0.126	0.106	0.156	0.151
Genebank landraces	0.141	0.115	0.173	0.168
Market varieties	0.117	0.100	0.157	0.142
Genetic Diversity Parameters				
Genotypes	H'	Simpson	Inverse Simpson	Pielou's Evenness
All	6.666	0.998	767	0.149
Breeding lines	5.736	0.996	308	0.173
Landraces	5.968	0.997	434	0.163
Market varieties	3.242	0.960	25	0.294

PIC: polymorphic information content; MAF: minor allele frequency; Ho: observed heterozygosity; He: expected heterozygosity; H': Shannon–Weaver index.

Table 3. Pairwise fixation index (Fst) values among the genetic groups.

Fst-Based Genetic Groups			
	Breeding Lines	Market Varieties	Genebank Landraces
Breeding Lines	0.000		
Market Varieties	0.031	0.000	
Genebank Landraces	0.038	0.024	0.000

Table 4. Analysis of molecular variance (AMOVA) within/among genetic groups.

Source of Variation	df	SS	MS	Est. Var.	%
Among genetic groups	2	7051.37	3525.68	15.13	4
Within genetic groups	800	318,465.22	398.08	398.08	96
Total	802	325,516.59		413.21	100

df: degrees of freedom; SS: sum of squares; MS: mean square; Est. Var.: estimated variance; %: percentage of variation.

To elucidate the clustering of 803 genotypes using the ADMIXTURE program, a varying number of subpopulations from $K = 2$ to 50 was plotted (Figure S1). This resulted in the most appropriate number of subpopulations at $K = 2, 3, 4,$ and 10. Figure 4 represented the genetic relationships among 803 genotypes as the estimated ancestries (Q) from ADMIXTURE analysis represented as barplot at $K = 2, 3, 4,$ and 10. The summary of ADMIXTURE cluster composition at $K = 2, 3, 4,$ and 10 is presented in Table S4. At $K = 2$, two major clusters were obtained consisting of 735 genotypes in Cluster 1 (Red) and 59 genotypes in Cluster 2 (Green) (Figure 4). After assessing the number of subpopulations (K) from 2, 3, 4, and 10, the most appropriate number was found to be $K = 4$ and $K = 10$, which produced the lowest cross-validation error compared to other K values (Figure 4 and Figure S1). At $K = 4$, 119 genotypes were found to be admixed consisting of 29 breeding lines, 82 landraces, and 8 market varieties. Similarly, at $K = 10$, the majority of the market varieties (17) were found to be admixed (Table S4).

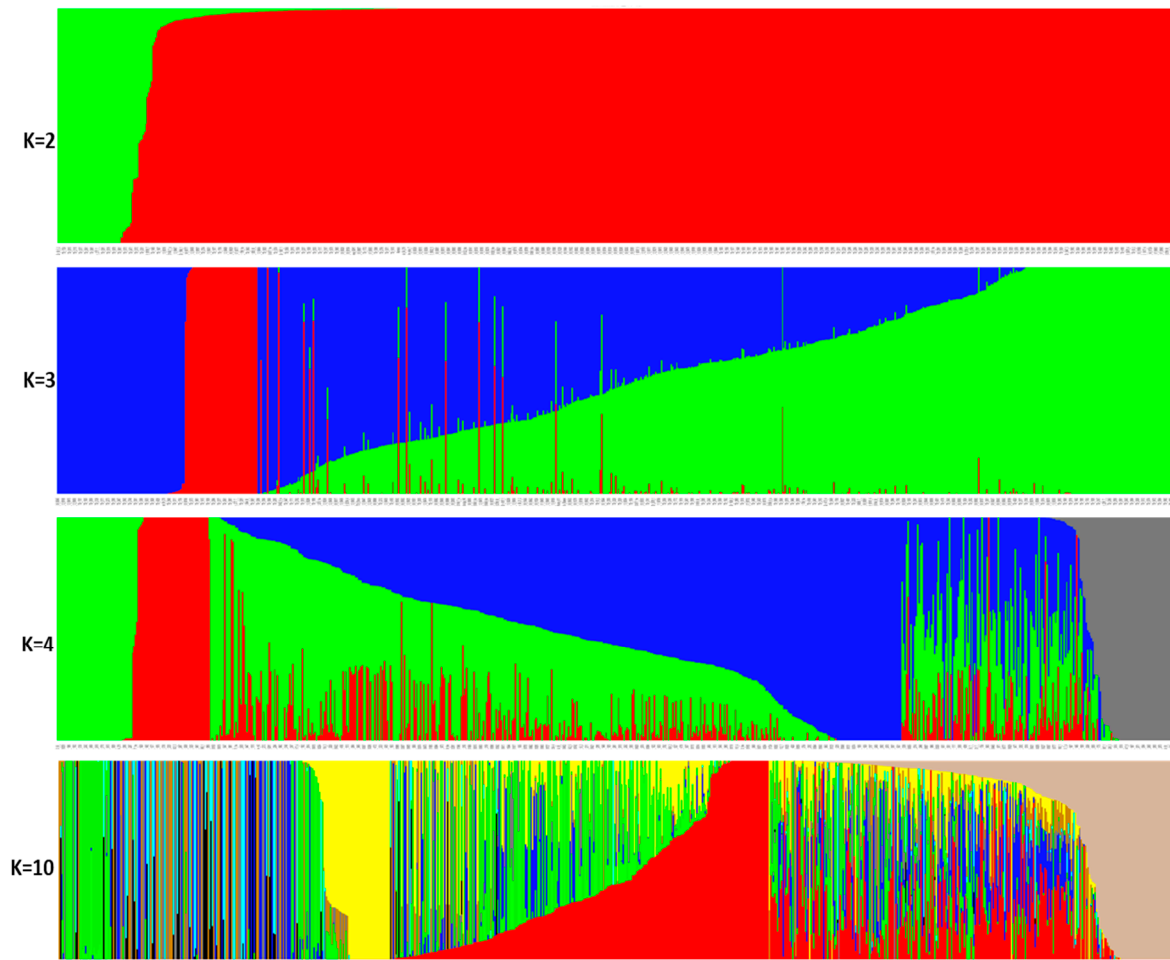


Figure 4. Population structure with $K = 2, 3, 4,$ and 10 for 803 *Dioscorea* genotypes using 3432 SNPs. The genotypes represented by vertical bars along the horizontal axis were classified into K color segments based on their estimated membership fraction in each K cluster. Accessions on the x-axis were sorted in the same order for each K .

The results of ADMIXTURE-based clustering at $K = 4$ was strongly supported by the topology of the distance-based phylogenetic tree (Figure S2). A major difference between the results of DAPC and ADMIXTURE clustering was the tendency of DAPC analysis to assign all genotypes to a single cluster compared to ADMIXTURE, which assigned admixed genotypes to multiple clusters based on K values (Tables S2 and S4).

3.3. Hierarchical Clustering and Network Analysis

A phylogenetic tree was further generated that grouped 803 genotypes into four major clusters (Figure S1) with several subclusters within Cluster 1 (Green) and Cluster 2 (Red). Cluster 1 was the largest cluster consisting of a mixture of genebank landraces (400), market varieties (26), and breeding lines (311) corresponding to Cluster 3 of DAPC clustering at $K = 3$ and Clusters 3 and 4 of ADMIXTURE clustering at $K = 4$. Clusters 3 (Black) and 4 (Blue) were comparatively smaller groups consisting mainly very few breeding lines and market varieties, respectively. In addition, a network analysis was further carried out to assess the genetic relationship between the three genetic groups (breeding lines, genebank landraces, and market varieties). Although the phylogenetic tree grouped 803 genotypes into four distinct clusters, the grouping was unable to highlight any particular pattern for the three genetic groups under study. On the contrary, the network analysis among these genetic groups (Figure 5) revealed a centralized structure and genetic contribution of each genotype within a genetic group (genebank landrace, breeding lines, and market varieties). The network analysis between breeding lines

and market varieties showed no direct genetic relationship among them, although some of the breeding lines may have been used directly as market varieties (Figure 5A), since there is a tendency with the farmers/traders to re-name the genotypes/breeding lines using their own naming system, which is associated with a popular market variety name. The naming system is also complex and depends upon the region within the country. For example, one of the popular market varieties in Nigeria is Hembakwase, which corresponds to several breeding lines (TDr 09/00023) and is associated with other market varieties such as Makakuasa and Omi_efun. Similarly, market variety Alumaco_1 was found to be similar to genebank landrace TDr 3584 and other market varieties such as Alumaco_2, TDr_Adaka, and TDr_Idu_Ekpeye. In addition, market variety TDr_Ehuru was identical to the breeding line TDr 08/00628. The network analysis between breeding lines and genebank landraces (Figure 5B) showed a strong genetic relationships indicating the use of genebank landraces in the yam breeding program to generate the selected breeding lines. This can be further elucidated from Table S1, wherein the pedigree information of the breeding lines has been provided. The central core of the QGRAPH (Figure 5B) represented a set of genebank landraces that were genetically similar and probably not used in the yam breeding program to generate breeding lines. Similarly, the network analysis between genebank landraces and market varieties indicated that these are genetically closer (Figure 5C). These findings highlighted the genetic relationships among different genetic groups of *D. rotundata* and the limited use of genebank landraces and market varieties in the breeding program.

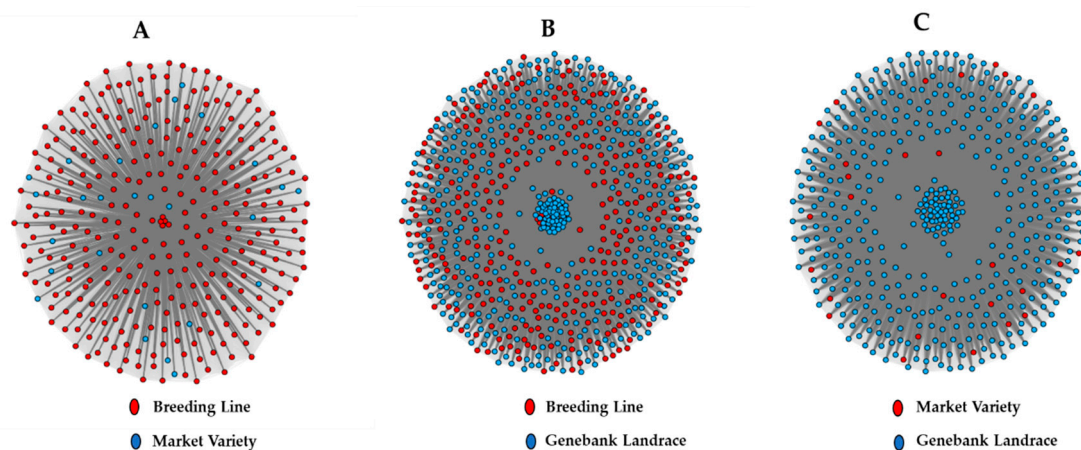


Figure 5. The genetic networks obtained using QGRAPH. The figure represents networks for all genetic groups with the node size depicting genetic relationships among different genotypes based on the observed heterozygosity and allelic richness. (A) Network analysis between breeding lines and market varieties; (B) Network analysis between breeding lines and genebank landraces; (C) Network analysis between genebank landraces and market varieties.

4. Discussion

The present study dissected the genetic relationships between different genetic groups (breeding lines, genebank landraces, and market varieties) of white yam/white guinea yam (*D. rotundata*) so that diverse genetic materials are utilized in the yam breeding program to introgress gene(s) of interest for its improvement. Despite several studies to assess the genetic diversity of white yam, little is known about the population structure or genetic diversity in contrast to other crops. A previous study [31] used a total of 94 landrace/gene bank accessions across seven guinea yam species to understand genetic diversity and their evolution. The allelic diversity, admixed patterns, and differential genome-wide population structure assayed by GBS-SNPs in three diverse genetic groups of *D. rotundata* further implied their efficacy in genomics-assisted breeding applications, which is one of the food security crops in the yam belt of West and Central Africa. The 3432 SNP markers identified in the current study were distributed across 21 pseudo chromosomes as per Tamiru et al. [37]. Sansaloni et al. [57] concluded in their study that the clustering of SNPs within certain regions across the genome is an

issue due to the reduced representation method used in developing the probes in GBS sequencing while resulting in low genome coverage.

This study elucidated that the majority of genetic variance exists within countries instead of between countries in the *D. rotundata* core collection. This was evident in DAPC clustering wherein the landraces from different countries were grouped together (Table S1). Several other studies including cowpea did not observe significant correlation between molecular clustering and geographic origins of genebank landraces [58]. The DAPC method revealed more clusters than ADMIXTURE (Tables S3 and S4), while the latter method provided information on genotypes with ancestries. The DAPC approach relies on discriminant functions and maximizes the diversity between clusters and minimizes within-cluster diversity [44]. However, DAPC-based clustering was found to be less efficient in clonally propagated crops such as white yam because of their continuous and complex population structure. This has been also reported in other clonally propagated crops such as cassava [59]. In general, DAPC cluster membership assignment was in agreement with ADMIXTURE clusters. Admixed ancestry was observed among breeding lines, which probably reflected their complex breeding history involving open pollination and bi-parental crossing coupled with strong adaptive selection pressure [60]. The admixed ancestry observed in the genebank landraces could be due to complex domestication patterns of *D. rotundata* landraces during evolutionary divergence. It has been explicitly demonstrated that West Africa at the forest/savannah ecotone is the cradle of yam domestication [8,11]. Furthermore, the inclusion of diverse landraces as common parents in the yam improvement program to develop/breed for valuable agronomic traits and higher yield might have influenced their population group assignment, resulting in numerous admixtures among these breeding lines (Table S1). In the present study, we have successfully unraveled the underlying genetic relationships and population structures among different genetic groups through network analysis. In the absence of complete pedigree records (across several years of breeding) or where breeding lines were selected from open-pollinated seeds, the dissection of genetic relationships among different genetic groups through network analysis was a reliable process. This has been successfully elucidated in cassava [59], wherein GBS-SNPs and complementary cluster analysis were used to assess the population structure and variety identification.

In clonal crops such as yams, improved varieties or breeding lines are often generated through intergenerational crosses (mainly bi-parental), which depends again on the flowering behavior of the parents used in those crosses across years. This has necessitated the use of the same parents with known flowering behavior in crossing programs at the Yam Breeding Unit, thus narrowing down the genetic base used. The genetic diversity observed among genebank landraces was high based on the Shannon–Weaver index, observed and expected heterozygosity, and this could be attributed to the extent of diversity captured within the core collection [37] and the percent of unique accessions (94%) (Table 2) identified in the present study. The genebank landraces were distributed across all the clusters generated through DAPC, ADMIXTURE, and phylogenetic tree, representing a significant amount of genetic diversity in the *D. rotundata* core collection. Hence, the diverse genebank landraces can be used as parents after the preliminary evaluation of flowering behavior and trait profiling, in white yam breeding programs to broaden the genetic base. Furthermore, this study was able to identify the complex naming system followed by farmers/traders to market varieties that were independently collected from different markets in Nigeria and Ghana, resulting in discrepancies from synonymy and homonymy in the tracking of released breeding lines when relying on use of names alone. Therefore, there is a need for a broader study including market varieties from a wide range of markets from different regions within Nigeria and Ghana to establish the inconsistencies with varietal names and its effect on the formal seed system in yams.

5. Conclusions

In conclusion, our study on white yam/white guinea yam represented a larger set of genotypes representing different genetic groups such as genebank landraces/breeding lines/market varieties. The genetic relationships dissected among different genetic groups in this study could be further explored in the white yam improvement by identifying diverse parents to generate mapping populations for target traits. Further studies are clearly needed to introgress gene(s) of target traits for white yam improvement. This study confirmed the reliability and accuracy of high-density SNP markers generated from next-generation sequencing-based genotyping coupled with complementary statistical analyses for genetic diversity and population structure.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4395/10/9/1437/s1>, Table S1. List of genotypes analyzed with information on their name, respective genetic group, geographic origin/pedigree, and results of DAPC clustering. Table S2. Identity by state (IBS)-based dissimilarity matrix. Table S3. Summary of DAPC cluster groups at different K values. Figure S1. Determination of optimal number of ADMIXTURE clusters using cross-validation error rates for K = 2 to K = 50. Table S4. Summary of ADMIXTURE cluster groups at different K values. Figure S2. Phylogenetic tree consisting of 803 genotypes using 3432 SNPs.

Author Contributions: Conceptualization, R.B., A.L.-M., P.L.K., M.A. and R.A.; Data curation, P.A. (Paterne Agre), G.B. and D.D.K.; Formal analysis, P.A. (Paterne Agre) and G.B.; Funding acquisition, R.B.; Methodology, R.B., P.A. (Paterne Agre), A.L.-M. and P.L.K.; Supervision, R.B.; Validation, R.B.; Writing—original draft, R.B.; Writing—review and editing, R.B., P.A. (Paterne Agre), G.B., D.D.K., A.L.-M., P.L.K., M.A., P.A. (Patrick Adebola), A.A. and R.A. All authors have read and agreed to the published version of the manuscript.

Funding: The research study was partly funded through complementary funding from the CGIAR Coordinated Research Project on Roots, Tubers and Bananas (CRP-RTBs) and partly through a research grant (OPP1052998) to IITA.

Acknowledgments: The authors would like to thank Ibukun Ogunleye and other Bioscience Center staff at IITA for the collection of samples and DNA extraction. We acknowledge Andreas Gisel and Yusuf Muideen for their critical advice on statistical analyses. The authors would also like to thank the Boyce Thompson Institute (BTI) Ithaca, Cornell University for providing server space for raw data deposit.

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

Data Availability: The datasets generated during and/or analysed during the current study are available on yambase (<https://yambase.org>).

Abbreviations

BWA	Burrow–Wheeler aligner
DAPC	Discriminant analysis of principal components
DNA	Deoxyribonucleic acid
GATK	Genome analysis tool kit
GBS	Genotyping-by-sequencing
Ho	Observed heterozygosity
He	Expected heterozygosity
H'	Shannon–Weaver index
IBS	Identity-by-state
IGD	Institute for Genomic Diversity
IITA	International Institute of Tropical Agriculture
LD	Linear discriminants
MAF	Minor allele frequency
NGS	Next-generation sequencing
PCA	Principal coordinate analysis
PCR	Polymerase chain reaction
PIC	Polymorphic information content
SNP	Single nucleotide polymorphism
VCF	Variant call format

References

- Mignouna, H.D.; Abang, M.M.; Fagbemi, S.A. A comparative assessment of molecular marker assays (AFLP, RAPD and SSR) for white yam (*Dioscorea rotundata*) germplasm characterization. *Ann. Appl. Biol.* **2003**, *142*, 269–276. [CrossRef]
- FAO. Food and Agriculture Organization of the United Nations Statistics Database, FAOSTAT. 2018. Available online: <http://www.fao.org/faostat/en/#data/QC> (accessed on 20 July 2020).
- Wilkin, P.; Schols, P.; Chase, M.W.; Chayamarit, K.; Furness, C.A.; Huysmans, S.; Rakotonasolo, F.; Smets, E.; Thapayai, C. A plastid gene phylogeny of the yam genus, *Dioscorea*: Roots, fruits and Madagascar. *Syst. Bot.* **2005**, *30*, 736–749. [CrossRef]
- Burkill, I.H. The organography and the evolution of Dioscoreaceae, the family of the yams. *Bot. J. Linn. Soc.* **1960**, *56*, 319–412. [CrossRef]
- Mignouna, H.D.; Dansi, A. Yam (*Dioscorea* spp.) domestication by the Nago and Fon ethnic groups in Benin. *Genet. Resour. Crop Evol.* **2003**, *50*, 519–528. [CrossRef]
- Dumont, R.; Dansi, A.; Vernier, P.; Zoundjihékpon, J. *Biodiversité et Domestication des Ignames en Afrique de l'Ouest. Pratiques Traditionnelles Conduisant à Dioscorea Rotundata*; CIRAD, Ed.; Collection Repère: Montpellier, VT, USA, 2005.
- Zannou, A.; Agbicodo, E.; Zoundjihékpon, J.; Struik, P.C.; Ahanchédé, A.; Kossou, D.K.; Sanni, A. Genetic variability in yam cultivars from the Guinea-Sudan zone of Benin assessed by random amplified polymorphic DNA. *Afr. J. Biotechnol.* **2009**, *8*, 26–36. [CrossRef]
- Scarcelli, N.; Cubry, P.; Akakpo, R.; Thuillet, A.-C.; Obidiegwu, J.; Baco, M.N.; Otoo, E.; Sonke, B.; Dansi, A.; Djedatin, G.; et al. Yam genomics supports west Africa as a major cradle of crop domestication. *Sci. Adv.* **2019**, *5*, eaaw1947. [CrossRef]
- Dansi, A.; Mignouna, H.D.; Zoundjihékpon, J.; Sangaré, A.; Asiedu, R.; Quin, F.M. Morphological diversity, cultivar groups and possible descent in the cultivated yams (*Dioscorea cayenensis*–*Dioscorea rotundata* complex) of Benin Republic. *Genet. Resour. Crop Evol.* **1999**, *46*, 371–388. [CrossRef]
- Mignouna, H.D.; Dansi, A.; Zok, S. Morphological and isozymic diversity of the cultivated yams (*Dioscorea cayenensis*/*Dioscorea rotundata* complex) of Cameroon. *Genet. Resour. Crop Evol.* **2002**, *49*, 21–29. [CrossRef]
- Bressan, E.A.; Briner Neto, T.; Zucchi, M.I.; Rabello, R.J.; Veasey, E.A. Genetic structure and diversity in the *Dioscorea cayenensis*/*D. rotundata* complex revealed by morphological and isozyme markers. *Genet. Mol. Res.* **2014**, *13*, 425–437. [CrossRef]
- Dansi, A.; Mignouna, H.D.; Zoundjihékpon, J.; Sangaré, A.; Asiedu, R.; Ahoussou, N. Using isozyme polymorphism to assess genetic variation within cultivated yams (*Dioscorea cayenensis*/*Dioscorea rotundata* complex) of the Republic of Benin. *Genet. Resour. Crop Evol.* **2000**, *47*, 371–383. [CrossRef]
- Scarcelli, N.; Tostain, S.; Mariac, C.; Agbangla, C.; Da, O.; Berthaud, J.; Pham, J.-L. Genetic nature of yams (*Dioscorea* spp.) domesticated by farmers in Benin (West Africa). *Genet. Resour. Crop Evol.* **2006**, *53*, 121–130. [CrossRef]
- Mignouna, H.D.; Abang, M.M.; Wanyera, N.W.; Chikaleke, V.A.; Asiedu, R.; Thottapally, G. PCR marker-based analysis of wild and cultivated yams (*Dioscorea* spp.) in Nigeria: Genetic relationships and implications for ex situ conservation. *Genet. Resour. Crop Evol.* **2005**, *52*, 755–763. [CrossRef]
- Dansi, A.; Mignouna, H.D.; Zoundjihékpon, J.; Sangaré, A.; Ahoussou, N.; Asiedu, R. Identification of some Benin Republic's Guinea yam (*Dioscorea cayenensis*/*Dioscorea rotundata* complex) cultivars using randomly amplified polymorphic DNA. *Genet. Resour. Crop Evol.* **2000**, *47*, 619–625. [CrossRef]
- Terauchi, R.; Chikalele, V.A.; Thottapally, G.; Hahn, S.K. Origin and phylogeny of guinea yams as revealed by RFLP analysis of chloroplast DNA and nuclear ribosomal DNA. *Theor. Appl. Genet.* **1992**, *83*, 743–751. [CrossRef] [PubMed]
- Obidiegwu, J.E.; Kolesnikova-Allen, M.; Ene-Obong, E.; Muoneke, C.; Asiedu, R. SSR markers reveal diversity in Guinea yam (*Dioscorea cayenensis*/*D. rotundata*) core set. *Afr. J. Biotechnol.* **2009**, *8*, 2730–2739.
- Loko, L.Y.; Bhattacharjee, R.; Agre, A.P.; Dossou-Aminon, I.; Orobisi, A.; Djedatin, G.; Dansi, A. Genetic diversity and relationship of Guinea yam (*Dioscorea cayenensis* Lam.–*D. rotundata* Poir. complex) germplasm in Benin (West Africa) using microsatellite markers. *Genetic Resour. Crop Evol.* **2017**, 1205–1219. [CrossRef]

19. Wendawek, A.M.; Demissew, S.; Fay, M.F.; Smith, R.J.; Nordal, I.; Wilkin, P. Genetic diversity and population structure of guinea yams and their wild relatives in south and south west Ethiopia as revealed by microsatellite markers. *Genet. Resour. Crop Evol.* **2013**, *60*, 529–541. [[CrossRef](#)]
20. Tostain, S.; Agbangla, C.; Scarcelli, N.; Mariac, C.; Berthaud, J.; Pham, J.-L. Genetic diversity analysis of yam cultivars (*Dioscorea rotundata* Poir.) in Benin using simple sequence repeat (SSR) markers. *Plant Genet. Resour.* **2007**, *5*, 71–81. [[CrossRef](#)]
21. Harikumar, P.; Sheela, M.N. Genetic diversity in white yam (*Dioscorea rotundata* Poir.) using random amplified polymorphic DNA (RAPD) markers. *Indian J. Pure Appl. Biosci.* **2019**, *7*, 30–35. [[CrossRef](#)]
22. Spindel, J.; Wright, M.; Chen, C.; Cobb, J.; Gage, J.; Harrington, S.; Lorieux, M.; Ahmadi, N.; McCouch, S. Bridging the genotyping gap: Using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. *Theor. Appl. Genet.* **2013**, *126*, 2699–2716. [[CrossRef](#)]
23. Rowe, H.C.; Renaut, S.; Guggisberg, A. RAD in the realm of next-generation sequencing technologies. *Mol. Ecol.* **2011**, *20*, 3499–3502. [[CrossRef](#)] [[PubMed](#)]
24. Deschamps, S.; Llaca, V.; May, G.D. Genotyping-by-sequencing in plants. *Biology* **2012**, *1*, 460–483. [[CrossRef](#)] [[PubMed](#)]
25. Poland, J.A.; Rife, T.W. Genotyping-by-Sequencing for Plant Breeding and Genetics. *Plant Genome* **2012**, *5*, 92–102. Available online: <https://dl.sciencesocieties.org/publications/tpg/abstracts/5/3/92> (accessed on 2 September 2020). [[CrossRef](#)]
26. He, J.; Zhao, X.; Laroche, A.; Lu, Z.-X.; Liu, H.; Li, Z. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Front. Plant Sci.* **2014**, *5*, 484. [[CrossRef](#)] [[PubMed](#)]
27. Hamblin, M.T.; Rabbi, I.Y. The effects of restriction-enzyme choice on properties of genotyping-by-sequencing libraries: A study in cassava (*Manihot esculenta*). *Crop Sci.* **2014**, *54*, 2603–2608. [[CrossRef](#)]
28. Elshire, R.J.; Glaubitz, J.C.; Sun, Q.; Poland, J.A.; Kawamoto, K.; Buckler, E.S.; Mitchell, S.E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **2011**, *6*, e19379. [[CrossRef](#)]
29. Lu, F.; Lipka, A.E.; Glaubitz, J.; Elshire, R.J.; Cherney, J.H.; Casler, M.D.; Buckler, E.S.; Costich, D.E. Switchgrass genomic diversity, ploidy, and evolution: Novel insights from a network-based SNP discovery protocol. *PLoS Genet.* **2013**, *9*, e1003215. [[CrossRef](#)]
30. Eaton, D.A.; Ree, R.H. Inferring phylogeny and introgression using RADseq data: An example from flowering plants (Pedicularis: Orobanchaceae). *Syst. Biol.* **2013**, *62*, 689–706. [[CrossRef](#)]
31. Girma, G.; Hyma, K.E.; Asiedu, R.; Mitchell, S.E.; Gedil, M.; Spillane, C. Next-generation sequencing based genotyping, cytometry and phenotyping for understanding diversity and evolution of guinea yams. *Theor. Appl. Genet.* **2014**, *127*, 1783–1794. [[CrossRef](#)]
32. Sasaki, C.A.; Bhattacharjee, R.; Scheffler, B.E.; Asiedu, R. Genomic resources for water yam (*Dioscorea alata* L.): Analyses of EST-sequences, de novo sequencing and GBS libraries. *PLoS ONE* **2015**, *10*, e0134031. [[CrossRef](#)]
33. Cormier, F.; Mournet, P.; Causse, S.; Arnau, G.; Maledon, E.; Gomez, R.-M.; Pavis, C.; Chair, H. Development of a cost-effective single nucleotide polymorphism genotyping array for management of greater yam germplasm collections. *Ecol. Evol.* **2019**, *9*, 5617–5636. [[CrossRef](#)] [[PubMed](#)]
34. Agre, P.; Asibe, F.; Darkwa, K.; Edemondou, A.; Bauchet, G.; Asiedu, R.; Adebola, P.; Asfaw, A. Phenotypic and molecular assessment of genetic structure and diversity in a panel of winged yam (*Dioscorea alata*) clones and cultivars. *Sci. Rep.* **2019**, *9*, 18221. [[CrossRef](#)] [[PubMed](#)]
35. Siadjeu, C.; Mayland-Quellhorst, E.; Albach, D.C. Genetic diversity and population structure of trifoliate yam (*Dioscorea dumetorum* Kunth) in Cameroon revealed by genotyping-by-sequencing (GBS). *BMC Plant Biol.* **2018**, *18*, 359. [[CrossRef](#)]
36. Girma, G.; Bhattacharjee, R.; Lopez-Montes, A.; Gueye, B.; Ofodile, S.; Franco, J.; Abberton, M. Redefining the yam (*Dioscorea* spp.) core collection using morphological traits. *Plant Genet. Resour.* **2018**, *16*, 193–200. [[CrossRef](#)]

37. Tamiru, M.; Natsume, S.; Takagi, H.; White, B.; Yaegashi, H.; Shimizu, M.; Yoshida, K.; Uemura, A.; Oikawa, K.; Abe, A.; et al. Genome Sequencing of the Staple Food Crop White Guinea Yam Enables the Development of a Molecular Marker for Sex Determination. *BMC Biol.* **2017**, *15*, 86. Available online: https://www.ncbi.nlm.nih.gov/assembly/GCA_002240015.2 (accessed on 2 September 2020). [[CrossRef](#)]
38. Li, H.; Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **2010**, *26*, 589–595. [[CrossRef](#)] [[PubMed](#)]
39. Available online: https://software.broadinstitute.org/gatk/best-practices/bp_3step.php?case=GermShortWGS (accessed on 2 September 2020).
40. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [[CrossRef](#)]
41. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.W.; Daly, M.J.; et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [[CrossRef](#)]
42. Alexander, D.H.; Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinform.* **2011**, *12*, 246. [[CrossRef](#)]
43. Alexander, D.H.; Novembre, J.; Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **2009**, *19*, 1655–1664. [[CrossRef](#)] [[PubMed](#)]
44. Jombart, T.; Devillard, S.; Balloux, F. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genet.* **2010**, *11*, 94. [[CrossRef](#)] [[PubMed](#)]
45. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018. Available online: <https://www.R-project.org/> (accessed on 2 September 2020).
46. Hill, M.O. Diversity and evenness: A unifying notation and its consequences. *Ecology* **1973**, *54*, 427–432. [[CrossRef](#)]
47. Oksanen, J.; Kindt, R.; Legendre, P.; O’Hara, B.; Simpson, G.L. The vegan package. *Community Ecol. Package* **2007**, *10*, 631–637.
48. Noli, E.; Teriaca, M.S.; Conti, S. Criteria for the definition of similarity thresholds for identifying essentially derived varieties. *Plant Breed.* **2013**, *132*, 525–531. [[CrossRef](#)]
49. Epskamp, S.; Cramer, O.J.C.; Waldorp, L.J.; Schmittmann, V.D.; Borsboom, D. qgraph: Network Visualizations of Relationships in Psychometric Data. *J. Stat. Software* **2012**, *48*, 1–18. Available online: <http://www.jstatsoft.org/v48/i04/> (accessed on 2 September 2020). [[CrossRef](#)]
50. Jombart, T. adegent: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **2008**, *24*, 1403–1405. [[CrossRef](#)]
51. Parida, S.K.; Mukerji, M.; Singh, A.K.; Singh, N.K.; Mohapatra, T. SNPs in stress-responsive rice genes: Validation, genotyping, functional relevance and population structure. *BMC Genom.* **2012**, *13*, 426. [[CrossRef](#)]
52. Nimmakayala, P.; Levi, A.; Abburi, L.; Abburi, V.L.; Tomason, Y.R.; Saminathan, T.; Vajja, V.G.; Ma, S.A.; Reddy, R.; Wehner, T.C.; et al. diversity, linkage disequilibrium, and selective sweeps in cultivated watermelon. *BMC Genom.* **2014**, *15*, 767. [[CrossRef](#)]
53. Liu, K.; Muse, S.V. Power Marker: Integrated analysis environment for genetic marker data. *Bioinformatics* **2005**, *21*, 2128–2129. [[CrossRef](#)]
54. Kumar, S.; Stecher, G.; Li, M.; Nnyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [[CrossRef](#)]
55. Available online: https://sniply.southgreen.fr/cgi-bin/snp_statistics.cgi?session=3036277183304&result=result (accessed on 2 September 2020).
56. Sartie, A.; Asiedu, R.; Franco, J. Genetic and phenotypic diversity in a germplasm working collection of cultivated tropical yams (*Dioscorea* spp.). *Genet. Resour. Crop Evol.* **2012**, *59*, 1753–1765. [[CrossRef](#)]
57. Sansaloni, C.P.; Petrolini, C.D.; Carling, J.; Hudson, C.J.; Steane, D.A.; Myburg, A.A.; Grattapaglia, D.; Vaillancourt, R.E.; Kilian, A. A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in Eucalyptus. *Plant Methods* **2010**, *6*, 16. [[CrossRef](#)]
58. Xiong, H.; Shi, A.; Mou, B.; Qin, J.; Motes, D.; Lu, W.; Ma, J.; Weng, Y.; Yang, W.; Wu, D. Genetic diversity and population structure of cowpea (*Vigna unguiculata* L. Walp). *PLoS ONE* **2016**, *11*, e0160941. [[CrossRef](#)] [[PubMed](#)]

59. Rabbi, I.Y.; Kulakow, P.A.; Manu-Aduening, J.A.; Dankyi, A.A.; Asibou, J.Y.; Parkes, E.Y.; Abdoulaye, T.; Girma, G.; Gedil, M.A.; Ramu, P.; et al. Tracking crop varieties using genotyping-by-sequencing markers: A case study using cassava (*Manihot esculenta* Crantz). *BMC Genet.* **2015**, *16*, 115. [[CrossRef](#)]
60. Darkwa, K.; Olasanmi, B.; Asiedu, R.; Asfaw, A. Review of empirical and emerging breeding methods and tools for yam (*Dioscorea* spp.) improvement: Status and prospects. *Plant Breed.* **2020**, *139*, 474–497. [[CrossRef](#)]



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).