Research Article

Author for correspondence:
Paterne A. Agre,
E-mail: p.agre@cgiar.org

Diversity of white Guinea yam (*Dioscorea rotundata* Poir.) cultivars from Benin as revealed by agro-morphological traits and SNP markers

Paterne A. Agre[1] (iD), Anicet G. Dassou[2], Laura E. Y. Loko[3], Roger Idossou[2], Eric Dadonougbo[2], Anicet Gbaguidi[2], Jean M. Mondo[1,4], Yusuf Muyideen[5], Patrick O. Adebola[6], Robert Asiedu[1], Alexandre A. Dansi[2] and Asrat Asfaw[1] (iD)

[1]International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria; [2]Laboratory of Biotechnology, Genetic Resources and Plant and Animal Breeding (BIORAVE), National University of Sciences, Technologies, Engineering and Mathematics, P.O. Box 14, Dassa-Zoumé, Benin; [3]Laboratory of Applied Entomology, National High School of Biosciences and Applied Biotechnologies (ENSBBA), National University of Sciences, Technologies, Engineering and Mathematics, P.O. Box 14, Dassa-Zoumé, Benin; [4]Department of Crop Production, Université Evangélique en Afrique, P.O. Box 3323, Bukavu, Democratic Republic of Congo; [5]Department of Horticultural Sciences, University of Minnesota, Saint Paul, MN 55108, USA and [6]International Institute of Tropical Agriculture (IITA), Abuja, Nigeria

Abstract

White Guinea yam (*Dioscorea rotundata* Poir.) is indigenous to West Africa, a region that harbours the crop's tremendous landrace diversity. The knowledge and understanding of local cultivars' genetic diversity are essential for properly managing genetic resources, conservation, sustainable use and their improvement through breeding. This study aimed to dissect phenotypic and molecular diversity of white yam cultivars from Benin using agro-morphological and single nucleotide polymorphism (SNP) markers. Eighty-eight Beninese white Guinea yam cultivars collected through a countrywide ethnobotanical survey were phenotyped with 53 traits and genotyped with 9725 DArT-SNP. Multivariate analysis using phenotypic traits revealed 30 traits as most discriminative and explained up to 80.78% of cultivars' phenotypic variation. Assessment of diversity indices such as Shannon–Wiener (H'), inverse Shannon (H.B.), Simpson's (λ) index and Pilou evenness (J) based molecular and phenotypic data depicted a moderate genetic diversity in Beninese white Guinea yam cultivars. Genetic differentiation of cultivars among country production zones was low due to the high exchange of planting materials among farmers of different regions. However, there was high genetic diversity within regions. Hierarchical clusters (HCs) on phenotypic data revealed the presence of two groups while HCs based on the SNP markers and the combined analysis identified three genetic groups. Our result provided valuable insights into the Beninese white Guinea yam diversity for its proper conservation and improvement through breeding.

Introduction

Yam is a popular staple in West Africa (Asiedu and Sartie, 2010; Darkwa *et al.*, 2020*a*). Its value chain sustains ∼5 million people's livelihood, including the rural farming households, traders, transporters and processors (Mignouna *et al.*, 2020). Yam is an essential source of carbohydrates, vitamins, essential minerals, fibres, with a low glycemic index (Akinola *et al.*, 2019). Benin is part of the African yam belt, a region producing more than 90% of the world yam production. It produces ∼2.9 million tons of yam annually and ranks fourth after Nigeria, Ghana and Côte d'Ivoire (FAO, 2020). Yam is a primary staple food and symbolic in many key socio-cultural and religious events (Zannou *et al.*, 2004; Loko *et al.*, 2019). The country produces several yam species, including *Dioscorea cayenensis*, *D. rotundata*, *D. alata* and *D. dumetorum* (Scarcelli *et al.*, 2006). Among these species, *D. rotundata* is the most popular and preferred for its economic profitability and suitability for a range of local food recipes.

As part of the crop improvement and genetic conservation efforts, several genetic diversity studies have been conducted on yam across the globe in general and in the sub-region in particular. These include an inventory and characterization of local cultivars using phenotypic traits (Dansi *et al.*, 1997; Loko *et al.*, 2013, 2015; Etchiha *et al.*, 2019), random amplified polymorphism DNAs (RAPD) and simple sequence repeat (SSR) molecular markers (Dansi *et al.*, 2000; Tostain *et al.*, 2007; Missihoun *et al.*, 2009; Loko *et al.*, 2017). Although a high genetic diversity was reported in Benin, only a few cultivars are widely grown for their superior agronomic performance and market value (Dansi *et al.*, 2000; Etchiha *et al.*, 2019). Besides, yam yield in this country is still meagre (13 t/ha) compared to the crop's yield potential in the

region (40–50 t/ha) (Frossard *et al.*, 2017). Developing improved varieties combining high tuber yield and superior food quality traits is an effective strategy to enhance yam productivity and increase farmers' interest in yam farming (Darkwa *et al.*, 2020*a*). An effective yam improvement programme through breeding, to develop such varieties, requires proper knowledge of the genetic resources available to guide the selection of cross-compatible and suitable parents (with desirable traits) to be involved in hybridization (Mondo *et al.*, 2020). As above-mentioned, both molecular and phenotypic analyses have been routinely used in yam for cultivar characterization prior to hybridization to achieve breeding objectives. However, there are increasing evidences on the limitations of each of these characterization approaches.

Molecular markers previously used for yam genetic diversity studies include RAPD (Asemota *et al.*, 1995), amplified fragment length polymorphism (AFLP) (Mignouna *et al.*, 1998, 2002*a*, 2002*b*; Terauchi and Kahl, 1999), SSR (Arnau *et al.*, 2009; Girma *et al.*, 2017; Loko *et al.*, 2017; Mulualem *et al.*, 2018; Babil *et al.*, 2021), inter SSRs (Ousmael *et al.*, 2019) and single nucleotide polymorphism (SNP) markers (Agre *et al.*, 2019; Bhattacharjee *et al.*, 2020; Darkwa *et al.*, 2020*b*). However, the reproducibility and reliability of markers like AFLP are limited, time-consuming and have low distribution across the genome. From the last decade, SNPs generated through the Next Generation Sequencing approaches (Diversity Arrays Technology (DArT) and genotyping-by-sequencing) are the most widely used in yam diversity studies because of their stability and abundance in the genome (Girma *et al.*, 2014; Akakpo *et al.*, 2017; Siadjeu *et al.*, 2018; Scarcelli *et al.*, 2019; Bhattacharjee *et al.*, 2020; Darkwa *et al.*, 2020*b*).

According to Andrade *et al.* (2017); Alves *et al.* (2017); Agre *et al.* (2019); Nkhoma *et al.* (2020); Stanley *et al.* (2020), genotypic and phenotypic data display very low or negligible correlations and produce non-duplicate information. A large portion of variation detected by molecular markers is non-adaptive compared with phenotypic characters which are influenced by the environment. In yam, the effectiveness of combining phenotypic and molecular marker information for dissecting genetic diversity has been reported (Sartie *et al.*, 2012; Agre *et al.*, 2019; Darkwa *et al.*, 2020*b*). Studies in yam and other crops suggest that combined analysis for phenotypic and genotypic information is useful for assessing the functional genetic diversity in crop plants and for minimizing limitations from either analysis approach. Genetic diversity assessment studies on Beninese white Guinea yam cultivars were conducted independently using either phenotypic or genotypic data. However, a joint analysis has not yet been deployed to explore the diversity of this crop in Benin. Hence, this study's objective was to assess the genetic structure and diversity in a panel of farmers' cultivars across the country using both phenotypic and SNP markers. Our study therefore complements previous studies in the country that used either morphological or molecular data alone. Proper genetic resource characterization provided by this joint analysis will be instrumental for efficient parental selection for yam breeding programmes in West Africa, including Benin. Besides, it will strengthen the crop genetic resource conservation efforts which were previously relying on either phenotypic or genotypic information alone.

Materials and methods

Collection of plant materials, field establishment and phenotyping

A total of 88 white Guinea yam (*D. rotundata*) cultivars were collected through an ethnobotanical survey in 20 villages and 16 markets of Benin's four largest yam production regions (Atacora, Donga, Borgou, Collines) (online Supplementary Fig. S1). The collection targeted the most popular cultivars with high market value to be used for yam crop improvement. The farmers' preference criteria for tuber yield potential, tolerance to weed competition, drought and diseases, long shelf life, good taste and aroma of processed food products and slow tuber flesh oxidation were considered in the sampling process.

These varieties were planted at the experimental farm of the Laboratory of Biotechnology, Genetic Resources and Plant and Animal Breeding (BIORAVE) located at Massi, Zogbodomey district, during the 2019 growing season (February–November). The field was established using an 11×8 lattice design with two replications. The plot consisted of an 11 m column with five plants per variety, at a spacing of $1 \text{ m} \times 1 \text{ m}$. No fertilizer was applied, while manual weeding was done regularly to maintain the plot free of weed. A total of 53 agro-morphological variables covering different plant parts (leaves, stem, flowers, fruits, roots, tubers) (online Supplementary Table S1) were assessed following the procedures described in the yam crop ontology (Asfaw, 2016).

Genotyping

About 10 g of young, healthy and fully expanded leaves from each genotype were collected using silica-gel granules with a colour indicator. Leaves were stored in the silica-gel for 72 h to remove the moisture. Subsequently, DNA extraction was carried out at the Bioscience centre, International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria, using the CTAB procedure with slight modification (Dellaporta *et al.*, 1983). The DNA quality was ascertained by running the gDNA in a 1% agarose gel while NanoDrop 2000 spectrophotometer was used to estimate its concentration and purity.

DArT genotyping was performed as described by Sansaloni *et al.* (2010). For the sequencing-based DArT genotyping, complexity reduction methods optimized for yam at the DArT Pty Ltd., Australia, were used. PstI_ad/TaqI/HpaII_ad with TaqI restriction enzyme was used to eliminate a subset of PstI–HpaII. The pstI-site-specific adapter was tagged for the 88 accessions with different barcodes enabling encoding a plate of DNA samples to run within a single lane on an Illumina GAIIx.

Data analysis

Phenotypic analysis

Prior the Least Squares means (LSmeans) estimation, the data were scaled and the LSmeans generated for the 88 yam varieties were used for principal component analysis (PCA) with FactoMineR (Le *et al.*, 2008) and missMDA (Josse and Husson, 2016). The optimum number of discriminant factors was determined through Peres-Neto *et al.* (2003) principle by considering factors with Eigenvalues >1. On each factor, the variables were declared discriminative and retained for subsequent analysis with a correlation above 0.5. Retained variables were then subjected to correlation plot using corrplot function implemented in R. Considering the discriminant variables, Gower dissimilarity matrix was generated using cluster package implemented in R. Count and nominal and or categorical variables (plant vigour, sprout colour, plant type, stem number per plant, number of internodes, spines on stem above base, flowering degree/intensity, plant sex, inflorescence type, tuber size, number of tubers

harvested per plant, number of marketable tubers, number of tubers damaged by rot and disease, tuber shape, tuber skin thickness, tuber flesh oxidation, flesh colour, absence/presence of corms on tubers) were listed in the model and transformed using log ratio function prior generating the dissimilarity matrix. The final hierarchical cluster (HC) was then performed based on ward.D2 method in cluster R package (Maechler *et al.*, 2021). Gower dissimilarity matrix was subjected to various diversity indexes such as Shannon–Wiener Index (H'), Inverse Simpson's (HB), Simpson's Index (λ) and Pilou evenness (J) using vegan library (Oksanen *et al.*, 2019). The reason for using different indices for diversity assessment was solely for increasing the reliability of conclusions.

SNP analysis

For quality control, DArTseq SNP-derived markers were filtered to remove the unwanted SNP markers using the software PLINK 1.9 and VCFtools. Markers and genotypes with >20% missing data were eliminated. Rare SNPs with <5% minor allele frequencies and low coverage read depth (<5) were also removed. In the end, only informative 9725 DArT-SNP markers and 88 cultivars were used for the subsequent analysis.

The summary statistics such as observed and expected heterozygosity, minor allele frequency (MAF) and polymorphic information content (PIC) were estimated using VCFtools (Danecek *et al.*, 2011) and PLINK 1.9 (Purcell *et al.*, 2007). Mutation transversion and transition were determined using the SniPlay web base (Purcell *et al.*, 2007). Dosage SNP format (0, 1, 2) was generated using recodeA function implemented in Plink where 0 is the homozygote reference, 1 the heterozygote and 2 the homozygote alternative. Dosage format was then subjected to Jaccard dissimilarity matrix using phylentropy R package (Drost, 2018). Jaccard dissimilarity matrix was then used to estimate the genetic diversity indexes such as Shannon–Wiener Index (H'), Inverse Simpson's (HB), Simpson's Index (λ) and Pilou evenness (J) to quantify the level of genetic diversity countrywide in Benin as well as within and among surveyed geographical zones using the vegan library (Oksanen *et al.*, 2019).

Besides, three multivariate analyses, including admixture for population structure, PCA, and cluster analysis with Ward method, were employed. Binary file generated from vcf file was then subjected to admixture analysis using 'adegenet' R package (Jombart *et al.*, 2010). The optimal number of clusters was inferred using k-means analysis after varying the number of clusters from 2 to 40. Through the admixture analysis, genotypes with membership proportions (Q-value) ≥60% were assigned to groups. In comparison, those with membership probabilities <60% were designated as admixtures (Salazar *et al.*, 2017). For the HC analysis, generated Jaccard dissimilarity matrix was used and the HC was plotted using Ward.D2 method. Further, PCA was conducted to reveal the genetic relationships among the yam varieties using GenAlEx v. 6.503 software (Peakall and Smouse, 2012) by considering the collection zone as a factor. Jaccard dissimilarity matrix was further subjected to the analysis of molecular variance (AMOVA) using GenAlEx v. 6.503 (Peakall and Smouse, 2012) to partition components of genetic variance among and within the populations (k).

Joint analysis of phenotypic and molecular data

The Gower and Jaccard dissimilarity matrices from both phenotypic and molecular data were combined to define potential genetic groups. The joint matrix was calculated as the algebraic sum of the Gower and Jaccard matrix generated through the phenotypic and the SNP marker data, respectively (Alves *et al.*, 2013; Andrade *et al.*, 2017). Mantel test was performed using the Monte-Carlo method with 9999 permutations to assess the correlation among the three matrices (phenotypic, genotypic and the combined). Phenotypic data of the different groups obtained through the joint analysis were subjected to k-mean analysis and the performance of each cluster was compared by Rstatix package (Kassambara, 2020).

Results

Phenotypic diversity

The first 14 principal components were identified as most discriminative and accounted for 80.78% of the cultivars' total phenotypic variation. Thirty (30) traits were correlated to the first 14 principal components (online Supplementary Table S2). The first principal component explained 13.98% of the total variation and was significantly associated with tuber size (TBRSZ), the number of marketable tubers (NMT), marketable tuber weight (MTW), marketable tuber length (MTLS), marketable tuber width (WMT) and tuber yield. The second principal component accounted for 10.16% of the total variation and was significantly correlated with six variables (online Supplementary Table S1). The third component explained 7.62% of the total variation and was positively associated with four phenotypic traits. Number of rotten tubers (NRDB) and weight of rotten tubers (RDTW) positively contributed to the fourth principal component, which explained 7.16% of the total variation.

Phenotypic correlation among the discriminating variables revealed high and positive correlations among total tuber yield, number of tubers, tuber weight, plant vigour, tuber length and tuber width while a negative correlation was recorded between the total tuber yield and tuber flesh oxidation (online Supplementary Fig. S2). Diversity indexes based on Gower dissimilarity matrix revealed the presence of moderate genetic diversity across the country (Table 1).

Cluster analysis based on phenotypic traits differentiated the 88 varieties into two major clusters (Fig. 1). The first cluster (red) was composed of 62 yam varieties with intermediate to late maturity cycle, moderate flowering intensity, moderate tuber yield (4.10 kg per plant) and high tuber flesh oxidation. The second cluster (green) comprised 26 early maturing yam varieties with low flowering intensity and good agronomic performance. Varieties in this cluster had high tuber yield (7.96 kg per plant) and high marketable tuber weight.

Molecular diversity and population structure

A total of 20,000 SNP markers was initially generated, from which 9725 were retained after the removal of low-quality SNP markers. The SNP markers were unequally distributed across the 20 yam chromosomes with a minimum of 212 SNPs on chromosome 13 to a maximum of 1469 SNPs on chromosome 5 (online Supplementary Fig. S3a). The SNP mutation showed that transition (T.S.) was greater (59%) than the transversion (T.V.) in the yam genome. Among the transition mutations, A/G had the highest occurrence rate in the genome (online Supplementary Fig. S3b). In contrast, for the transversion mutations, A/C appeared at the highest rate. Values of 0.22 and 0.24 were recorded as averages for the

Table 1. Genetic diversity indices based on SNP data across collection zones of Benin

	Based phenotypic data	Based molecular SNP data				
		Across (88)	Atacora (21)	Borgou (24)	Donga (21)	Collines (22)
(HB)	81.54	79.34	18.03	19.04	17.59	19.72
(H′)	4.43	4.39	2.92	2.99	2.88	2.99
(J)	0.22	0.22	0.31	0.30	0.32	0.31
(λ)	0.98	0.99	0.94	0.95	0.94	0.95

Inverse Simpson's (HB); Shannon–Wiener Index (H′); Pilou evenness (J); Simpson's Index (λ).
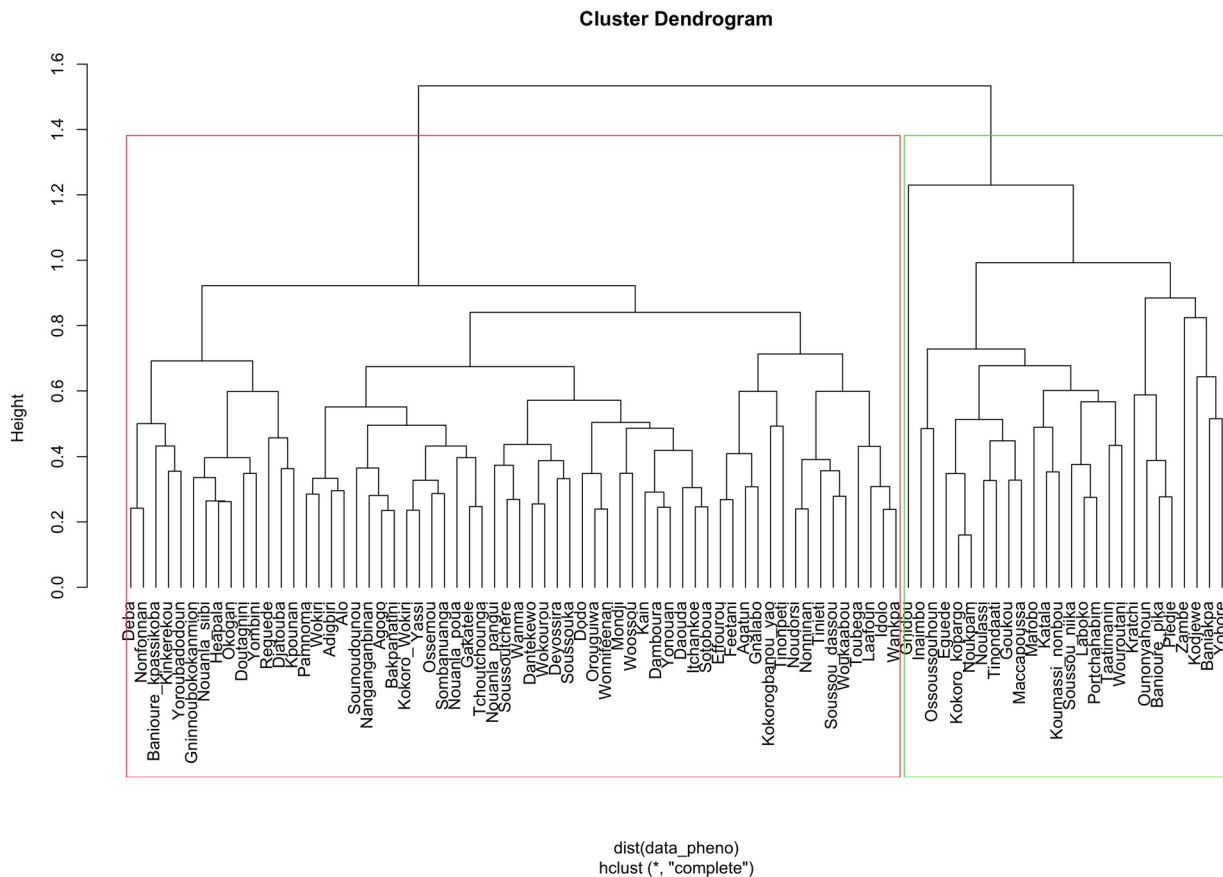
Fig. 1. Phenotypic characterization of Beninese white Guinea yam varieties based on Gower dissimilarity matrix using the 30 most discriminating traits. The colours represent the two clusters: Cluster 1 (red) and cluster 2 (green).

observed and expected heterozygosity, respectively. The PIC varied from 0 to 0.37, with an average of 0.20. The MAF was 0.17. Using the 9725 SNP markers, all cultivars evaluated were estimated to be diploid.
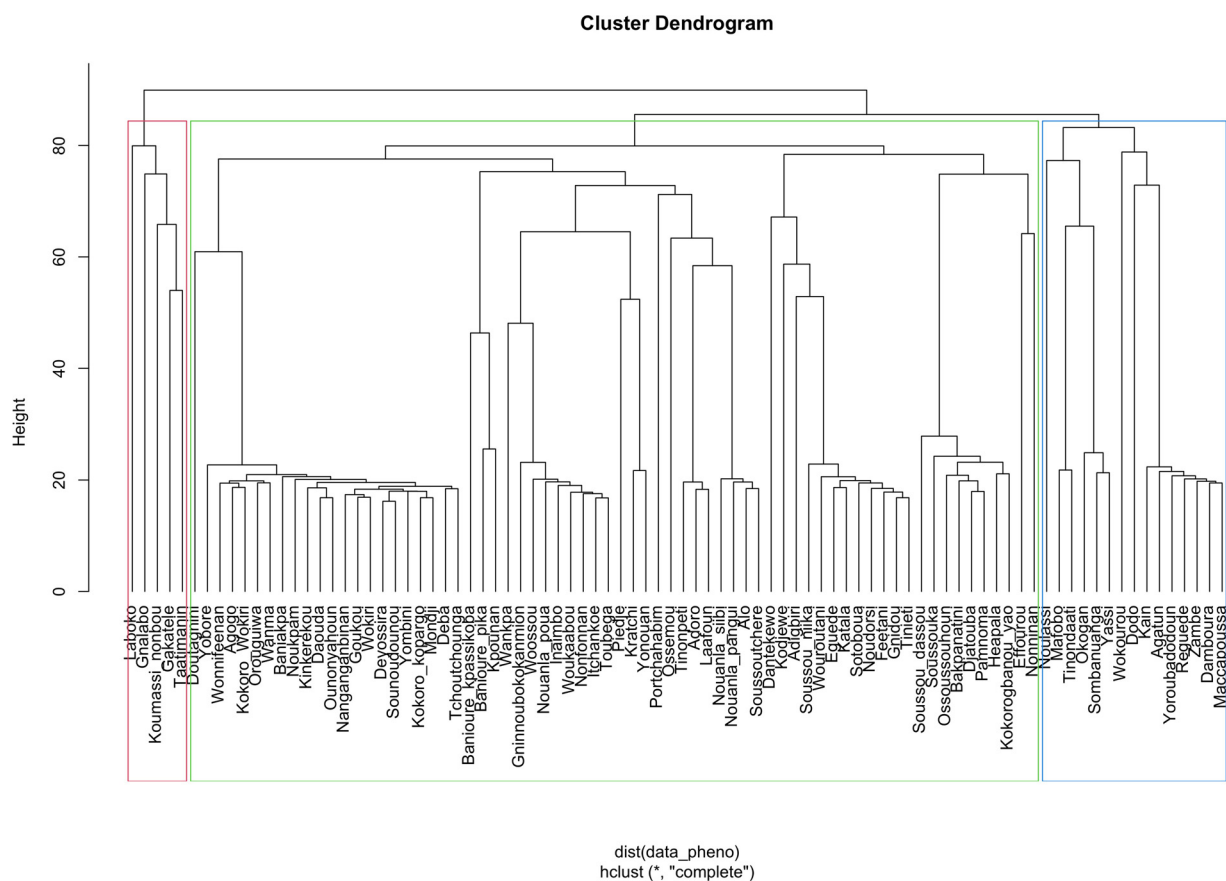
The Jaccard dissimilarity matrix displayed a high range of genetic distance from 0 to 0.89. Hierarchical clustering-based DArT-SNP marker grouped the 88 yam varieties into three major clusters (Fig. 2). Five cultivars were grouped in cluster 1 (red) (Fig. 2) with the lowest genetic distance (0.10) obtained between Taatimanin and Gakatele. In this cluster, the highest genetic distance (0.21) was observed between *Laboko* and Taatimanin. Cluster 2 (green) had the highest membership (68), with its members widely distributed across the different yam agro-ecologies. In this cluster, member cultivars were majorly

identified as males except *Gnidou*, *Katala*, *Eguede* and *Sotoboua* which had female flowers. Cluster 3 (blue) was made of 15 cultivars (Fig. 2) and displayed 0.22 as an average genetic distance within the cluster. Members of this cluster were identified as females with low to high flowering intensity.

Through the Bayesian Information Criteria (BIC) analysis, population structure revealed the presence of deflection at *k* equal to 3 (online Supplementary Fig. S4, Fig. 3). Using a 60% membership probability threshold, 74 cultivars (84.09%) were successfully assigned to the three clusters. In contrast, 14 cultivars with an association probability of <60% were designated as admixtures (Fig. 3).

Shannon–Wiener, Inverse Shannon, Simpson and Pilou diversity indices revealed the presence of moderate yam genetic

Fig. 2. Assessment of the genetic diversity of the 88 Beninese white Guinea yam varieties based on Jaccard dissimilarity matrix using the DArT-SNP markers. The colours represent the three clusters: Cluster 1 (red), cluster 2 (green) and cluster 3 (blue).

diversity in Benin (Table 1). The level of genetic diversity was relatively high in Collines, followed by Borgou and Atacora regions. At the same time, the lowest index was recorded in Donga (Table 2).

The dispersion of the genotypes on the first two principal components, which explained 44.2% of the total molecular variation, showed that the 88 yam varieties were clustered irrespective of their geographical origins (online Supplementary Fig. S5). The pairwise fixation index (F_{st}) value further confirmed the low genetic differentiation between geographical origins (Table 2). The AMOVA also revealed high genetic variability within populations (95%), while only 5% of the total variability was among populations (online Supplementary Table S3).

Assessment of genetic diversity based on joint analysis

Clustering based on the joint analysis for phenotypic and molecular marker information formed three groups (Fig. 4). Cluster 1 (red) comprised 21 cultivars. Thirty and six cultivars formed the second cluster (green), with 36% of them from the Department of Collines (Central Benin), and the remaining from Atacora and Donga Departments (North-West). Cluster 3 (blue) comprised of 31 cultivars majority collected from the Donga department and the rest from the other regions.

Discriminant analysis based on phenotypic variables revealed high phenotypic variation among the clusters (online Supplementary Table S4). Cultivars in cluster 1 were characterized by high plant

vigour and few spines at the stem's base. The average number of tubers varied from 2 to 3 per plant with a tuber length between 15 and 25 cm and an average yield of 2.18 kg/plant. Cultivars from cluster 2 had vigorous plants, generally no spines on the tuber except *Gnidou*, and produced long and large tubers (1–2 per plant). Many of the cluster 2 members had high tuber flesh oxidation intensity such as *Gnidou*, popularly known for this trait. All the cultivars locally known as a 'Kokoro' type (intermediate to late maturity cycle) were grouped into cluster 3. Cultivars in cluster 3 were characterized by multiple and small tubers (<15 cm long) per plant that hardly showed the parenchyma's oxidization at peeling.

The distance matrices from the phenotypic and molecular data were weakly correlated (online Supplementary Fig. S6). However, such a relationship was relatively improved with the joint matrix and the respective phenotypic and molecular matrices. The phenotypic distance matrix had a low correlation (0.125) with the joint dissimilarity matrix. In contrast, the correlation between the SNP marker-based Jaccard dissimilarity matrix and the joint matrix was high and positive ($r = 0.98$) (online Supplementary Fig. S6).

Discussion

Of the 53 phenotypic traits assessed in our study, 30 appeared very informative in discriminating the yam cultivars (online Supplementary Table S2). To save time and resources during
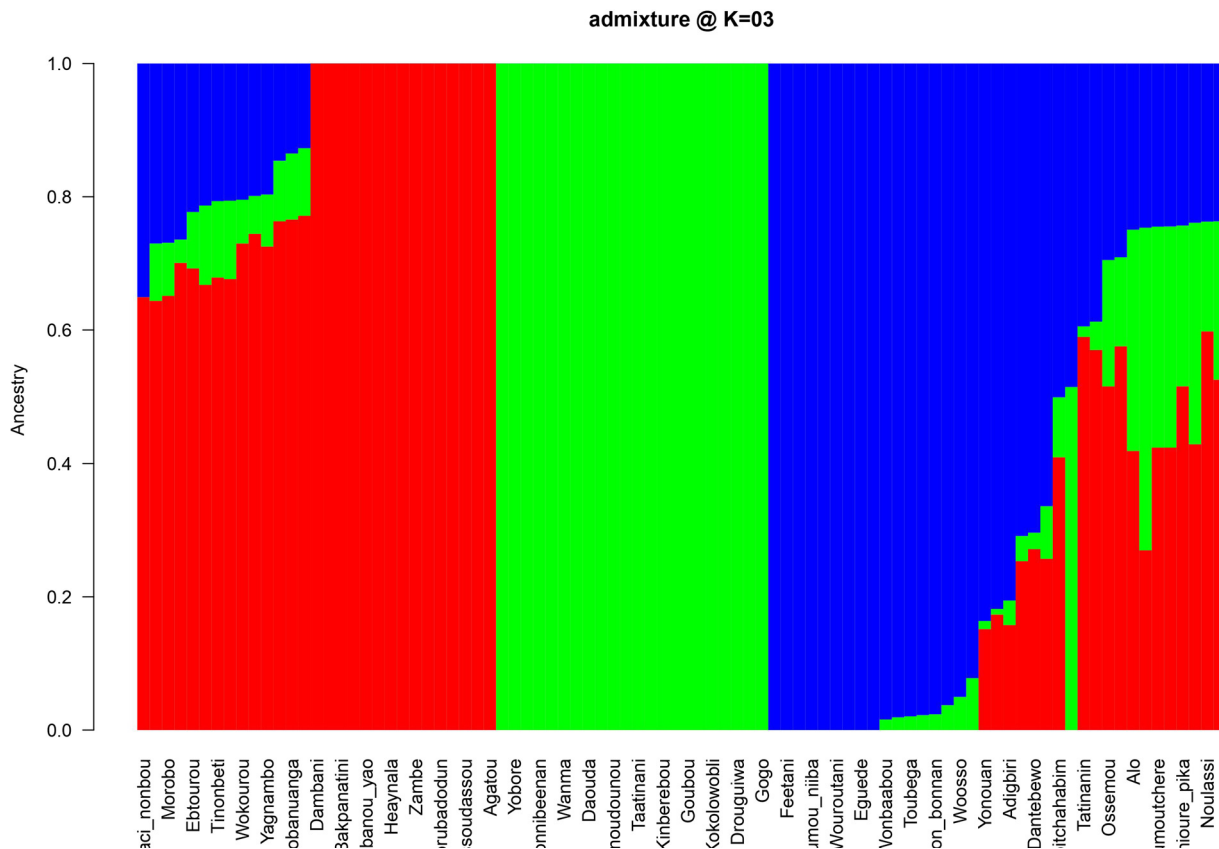
admixture @ K=03

Fig. 3. Graphical representation of the 88 Beninese yam varieties' population structure based on admixture analysis. Subpopulations were set at *k* = 3. The colours represent the three clusters: Cluster 1 (red), cluster 2 (green) and cluster 3 (blue) based on a membership coefficient of ≥60%.

Table 2. Pairwise fixation index (F_{st}) values among the genetic groups

| Population	Atacora	Borgou	Collines	Donga
Atacora	–			
Borgou	0.090	–		
Collines	0.044	0.080	–	
Donga	0.054	0.013	0.012	–

yam germplasm characterization, we would recommend focusing on these informative traits in future studies.

The level of diversity in yam cultivars from Benin was found moderate, especially with SNP markers. Our results identified three subgroups of yam cultivars with a low level of admixture. The members of the three sub-groups were generally clustered based on the genetic distance, although those in clusters 1 and 2 were also grouped based on their sex information (Fig. 2) The observed genetic divergence with a low admixture level is due to original differences in domestication and subsequent vegetative propagation by farmers. In Benin, farmers often collect tubers from wild yams, plant them in their fields and select suitable ones through clonal propagation using tubers (Scarcelli *et al.*, 2006; Akakpo *et al.*, 2017). This practice is referred to as 'ennoblement' and takes 3–6 years until the suitable tuber morphology is achieved. The tradition of raising seedlings from random out-crossing seeds in a farmer's field and further selecting the adapted

ones to production niches and uses has not been reported. As a result, the development of new cultivars from cross-breed seeds is a less likely phenomenon for the yam germplasm, even if the varietal mixture of different phenotypes is a common farming practice in many yam growing regions of Africa. Plants of a yam clone in a farmer's field are genetically homogenous with negligible recombination rates, as farmers select their planting material from tubers and not from botanical seed. However, the molecular evidence on ennobled cultivars showed that the tubers collected by farmers from wild environments are often a mixture of wild (*D. abyssinica* and *D. praehensilis*) and interspecific hybrid (*D. rotundata* × wild species) yams (Scarcelli *et al.*, 2006, 2017; Sugihara *et al.*, 2020). This could partly explain the origin of the genetic admixture identified among the Beninese white Guinea yam accessions. Besides, Loko *et al.* (2013) showed that farmers often establish yam fields within savannah and forests, a practice that favours gene flows between cultivated yams and their wild relatives.

Our results further demonstrated a high level of gene flow between regions of Benin with no apparent pattern of geographical differentiation in farmer cultivars. This result supports Loko *et al.*'s (2017) findings that used microsatellite markers on Beninese white yam cultivars. Although Tostain *et al.* (2007) agreed on gene flows among regions, they established a geographical pattern between *D. rotundata* genetic diversity, cropping region and farmers' crop management practices. However, that geographic pattern hypothesis has not been supported by other diversity studies on Beninese white Guinea yam accessions

Cluster Dendrogram
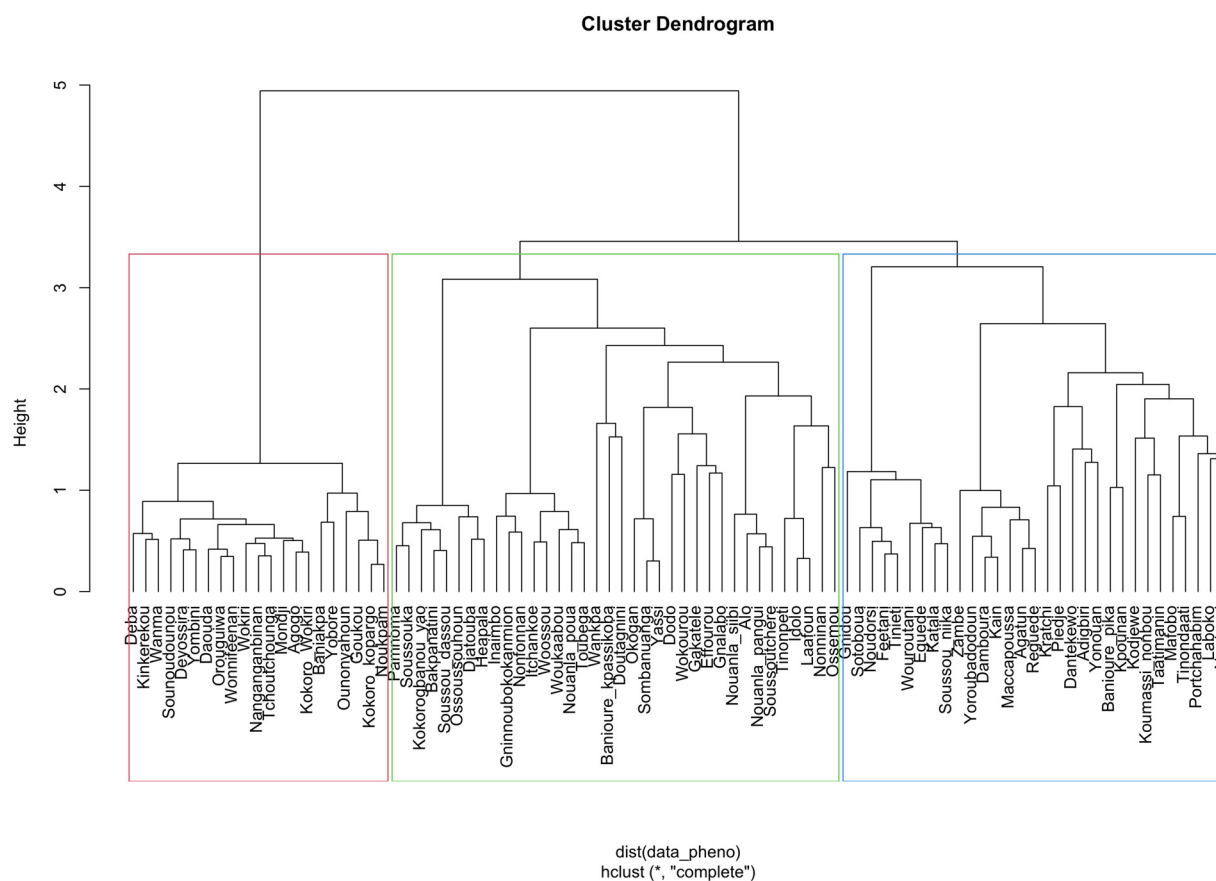
Fig. 4. Assessment of the genetic diversity among Beninese white Guinea yam accessions through the joint dissimilarity matrix. The colours represent the three clusters: Cluster 1 (red), cluster 2 (green) and cluster 3 (blue).

(Mignouna and Dansi, 2003; Loko *et al.*, 2017). This gene flow reflects high inter-region seed yam exchange or social and commercial networks for yam in Benin. For instance, farmers often exchange seeds of landraces with other farmers within areas or involve outlining localities, and to some extent, with neighbouring countries such as Nigeria and Togo. They gain access to new landraces, which were adapted to similar environments. Several other studies on cowpea (Fatokun *et al.*, 2018; Nkhoma *et al.*, 2020), maize (Nelimor *et al.*, 2020) and even yam (Loko *et al.*, 2017; Adewumi *et al.*, 2020, 2021; Bhattacharjee *et al.*, 2020) showed the existence of gene flow among regions and the absence of significant correlations between molecular clustering and geographic origins.

Based on the diversity indexes, the accessions from the department of Collines were more diverse. Traditionally, farmers of this area cultivate multiple yam varieties with different maturity groups such as extra early (5–6 months), early (6–7 months), intermediate (8 months) and late-maturing (9–10 months) to satisfy the steady demand of the international yam market based in Glazoue throughout the year.

Mantel test revealed a low correlation between the dissimilarity matrices originated from the phenotypic and the molecular data. This finding was supported by the membership inconsistency between genomic-based and phenotypic-based clustering. Similar results were reported by several authors using the same approaches on sugar cane, yellow fruit passion, white yam,

water yam and cowpea (Sartie *et al.*, 2012; Alves *et al.*, 2013; Orgogozo *et al.*, 2015; Agre *et al.*, 2019; Darkwa *et al.*, 2020*b*; Nkhoma *et al.*, 2020). The low correlation between the phenotypic and the genotypic data could have resulted from the natural and artificial selections on phenotypic variables, as these are under selection and influence of environmental factors. In contrast, the variation detected by molecular markers is commonly non-adaptive, and hence, not subject to natural and or artificial selections (Arnau *et al.*, 2017). This highlights the importance of combining phenotypic and molecular information while selecting parental lines in breeding programmes to account for their agro-morphological performance and genetic background. Phenotypic traits have the advantage of revealing the agronomic performance of a variety in a particular environment but have limited polymorphism, and they are subjected to changes in environmental conditions (Darkwa *et al.*, 2020*b*; Nkhoma *et al.*, 2020). Harnessing the advantages of phenotypic and molecular markers improves the grouping of entries in a germplasm collection (da Silva *et al.*, 2017), which provides a piece of valuable base information for parental selection to realize and sustain genetic gain. The different genetic groups developed through the joint analysis used in this study will be of good use for yam breeding community in West Africa through germplasm exchange. For instance, cultivars with desirable traits such as high yield, tuber quality, disease resistance, etc., have been identified and clustered, and the genetic distance and sex information was provided. This

information is critical in selecting parental clones and designing efficient crossing plans. It is noteworthy that more than 90% of the *D. rotundata* cultivars are diploids (Gatarira, 2020; Babil *et al.*, 2021) and the ploidy-level influence on gene flow within this species has not been significant (Mondo *et al.*, 2020, 2021*a*, 2021*b*). Therefore, cultivars' agronomic performance, genetic distance and sex information as provided in this study will be a useful asset for yam breeding programmes.

Genetic diversity parameters such as PIC, MAF, observed and expected heterozygosity levels displayed low variability from one region to another, which could be due to the presence of natural gene flow among the study zones and possibly from different neighbouring countries.

Conclusion

This study revealed a low to moderate genetic diversity among different yam-growing areas of Benin as a result of seed exchange among farmers of different regions. However, high genetic diversity was observed within regions due to a range of farmers and other end-users' preferences within a region. High genetic diversity among plant materials collected within areas as well as information related to their genetic distance and sex provide an opportunity and a good source of selection for plant breeding programmes. We ascertained the relevance of combining molecular markers with phenotypic data for a refined genetic diversity assessment, as it revealed an independent nature of morphological and molecular variations.

References

Adewumi AS, Taah KJ, Asare PA, Adu MO and Agre PA (2020) Assessment of genetic diversity of *Dioscorea praehensilis* (Berth.) collected from Central Region, Ghana using simple sequence repeat (SSR) markers. E.S. *Journal of Agriculture and Current Research* **1**, 1005–1012.

Adewumi AS, Asare PA, Adu MO, Taah KJ, Akaba S, Mondo JM and Agre PA (2021) Farmers' perceptions on varietal diversity, trait preferences and diversity management of bush yam (*Dioscorea praehensilis* Benth.) in Ghana. *Scientific African* **12**, e00808.

Agre P, Asibe F, Darkwa K, Edemodu A, Bauchet G, Asiedu R and Asfaw A (2019) Phenotypic and molecular assessment of genetic structure and diversity in a panel of winged yam (*Dioscorea alata*) clones and accessions. *Scientific Reports* **9**, 18221–182232.

Akakpo R, Scarcelli N, Chaïr H, Dansi A, Djedatin G, Thuillet AC, Rhoné B, François O, Alix K and Vigouroux Y (2017) Molecular basis of African yam domestication: analyses of selection point to root development, starch biosynthesis, and photosynthesis related genes. *BMC Genomics* **18**, 782–791.

Akinola AA, Ezeorah SN and Nwoko EP (2019) Modelling the rehydration characteristics of White Yam. *The West Indian Journal of Engineering* **2**, 70–76.

Alves AA, Bhering LL, Rosado TB, Laviola BG, Formighieri EF and Cruz CD (2013) Joint analysis of phenotypic and molecular diversity provides new insights on the genetic variability of the Brazilian physic nut germplasm bank. *Genetics and Molecular Biology* **36**, 371–381.

Alves RM, Silva CR, De S, Albuquerque PSB and Santos VS (2017) Phenotypic and genotypic characterization and compatibility among genotypes to select elite clones of cupuassu. *Acta Amazonica* **3**, 175–184.

Andrade LA, Barbosa NA and Pereira J (2017) Extraction and properties of starches from the non-traditional vegetables Yam and Taro. *Polímeros* **2**, 151–157.

Arnau G, Némorin A, Maledon E and Abraham K (2009) Revision of ploidy status of *Dioscorea alata* L.(Dioscoreaceae) by cytogenetic and microsatellite segregation analysis. *Theoretical and Applied Genetics* **7**, 1239–1249.

Arnau G, Bhattacharjee R, Sheela MN, Malapa R, Lebot V, Abraham K and Pavis C (2017) Understanding the genetic diversity and population structure of yam (*Dioscorea alata* L.) using microsatellite markers. *PLoS ONE* **3**, 1–17.

Asemota HN, Ramser J, Lopez-Peralta C, Weising K and Kahl G (1995) Genetic variation and cultivar identification of Jamaican yam germplasm by random amplified polymorphic DNA analysis. *Euphytica* **3**, 341–351.

Asfaw A (2016) Standard operating protocol for yam variety performance evaluation trial. *IITA, Ibadan, Nigeria* **27**, 1–33.

Asiedu R and Sartie A (2010) Yams for income and food security. *Food Security* **2**, 305–315.

Babil P, Yamanaka S, Girma G, Matsumoto R, Tamiru-Oli M, Bhattacharjee R, Abberton M, Muranaka S, Asiedu R, Terauchi R and Takagi H (2021) Simple sequence repeat-based mini-core collection for white Guinea yam (*Dioscorea rotundata*) germplasm. *Crop Science* **61**, 1269–1279. https://doi.org/10.1002/csc2.20431.

Bhattacharjee R, Agre P, Bauchet G, De Koeyer D, Lopez-Montes A, Kumar P, Abberton M, Adebola P, Asfaw A and Asiedu R (2020) Genotyping-by-sequencing to unlock genetic diversity and population structure in white yam (*Dioscorea rotundata* Poir.). *Agronomy* **9**, 1437.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E and DePristo MA (2011) The variant call format and VCFtools. *Bioinformatics (Oxford, England)* **15**, 2156–2158.

Dansi A, Zoundjihekpon J, Mignouna HD and Quin FM (1997) Collecte d'ignames cultivées du Complexe *Dioscorea cayenensis – rotundata* au Benin. *Plant Genetic Resources Newslett* **112**, 81–85.

Dansi A, Mignouna HD, Zoundjihékpon J, Sangare A, Ahoussou N and Asiedu R (2000) Identification of some Benin Republic's Guinea yam (*Dioscorea cayenensis/Dioscorea rotundata* complex) accessions using randomly amplified polymorphic DNA. *Genetic Resources and Crop Evolution* **6**, 619–625.

Darkwa K, Olasanmi B, Asiedu R and Asfaw A (2020*a*) Review of empirical and emerging breeding methods and tools for yam (*Dioscorea* spp.) improvement: status and prospects. *Plant Breeding* **3**, 474–497.

Darkwa K, Agre P, Olasanmi B, Kohtaro I, Matsumoto R, Powell A, Bauchet G, De Koeyer D, Muranaka S, Adebola P, Asiedu R, Terauchi R and Asfaw A (2020*b*) Comparative assessment of genetic diversity matrices and clustering methods in white Guinea yam (*Dioscorea rotundata*) based on morphological and molecular markers. *Scientific Reports* **10**, 1319–1323.

da Silva GM, Soares ACF, Menezes RSC, Sampaio EVDS, da Silva Souza EM and do Nascimento Macedo BL (2017) Quality and yield of yam tubers cultivated with green manure in the Northeastern Brazil. *Comunicata Scientiae* **2**, 275–280.

Dellaporta SL, Wood J and Hicks JB (1983) A plant DNA minipreparation: version II. *Plant Molecular Biology Report* **1**, 19–21.

Drost (2018) Philentropy: information theory and distance quantification with R. *Journal of Open Source Software* **26**: 765–769.

Etchiha AP, Affokpon A, Loko YLE, Agbangla C and Dansi A (2019) Yam (*Dioscorea rotundata* Poir and *D. cayenensis* Lam complex) in the traditional agriculture of benin: present-day cultivar diversity and farmers' perception on their tolerance to tuber dry rot caused by the nematode *Scutellonema bradys*. *International Journal of Current Microbiology and Applied Science* **8**, 1119–1138.

FAO FAOSTAT (2020) Available online: http://www.fao.org/faostat/en/#data/QL (accessed on 1 December 2020).

Fatokun C, Girma G, Abberton M, Gedil M, Unachukwu N, Oyatomi O and Boukar O (2018) Genetic diversity and population structure of a mini-core subset from the world cowpea (*Vigna unguiculata* (L.) Walp.) germplasm collection. *Scientific Reports* **8**, 16036–16045.

Frossard E, Aighewi BA, Aké S, Barjolle D, Baumann P, Bernet T, Dao D, Diby LN, Floquet A, Hgaza VK and Ilboudo LJ (2017) The challenge of improving soil fertility in yam cropping systems of West Africa. *Frontiers in Plant Science* **8**, 1953–1162.

Gatarira PC (2020) *Ploidy determination, flowering behaviour and genome wide association studies in Dioscorea species* (PhD thesis). Pan African University-Institute of Life and Earth Sciences, University of Ibadan, Ibadan, Nigeria, 115.

Girma G, Hyma KE, Asiedu R, Mitchell SE, Gedil M and Spillane C (2014) Next-generation sequencing based genotyping, cytometry and phenotyping for understanding diversity and evolution of Guinea yams. *Theoretical and Applied Genetics* **127**, 1783–1794.

Girma G, Gedil M and Spillane C (2017) Morphological, SSR and ploidy analysis of water yam (*Dioscorea alata* L.) accessions for utilization of aerial tubers as planting materials. *Genetic Resources and Crop Evolution* **64**, 291–305.

Jombart T, Devillard S and Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* **11**, 1–15.

Josse J and Husson F (2016) missMDA: a package for handling missing values in multivariate data analysis. *Journal of Statistical Software* **1**, 1–31.

Kassambara A (2020) rstatix: Pipe-friendly framework for basic statistical tests. R package version 0.6.0. Available at https://CRAN.R-project.org/package=rstatix.

Le S, Josse J and Husson F (2008) FactoMineR: an R package for multivariate analysis. *Journal of Statistical Software* **1**, 1–18.

Loko YL, Dansi A, Linsoussi C, Tamo M, Vodouhè R, Akoegninou A and Sanni A (2013) Current status and spatial analysis of Guinea yam (*Dioscorea cayenensis* Lam.–*D. rotundata* Poir. complex) diversity in Benin. *International Research Journal of Agricultural Science and Soil Science* **7**, 219–238.

Loko YL, Adjatin A, Dansi A, Vodouhe R and Sanni A (2015) Participatory evaluation of Guinea yam (*Dioscorea cayenensis* Lam.–*D. rotundata* Poir. complex) landraces from Benin and agromorphological characterization of accessions tolerant to drought, high soil moisture and chips storage insects. *Genetic Resources and Crop Evolution* **62**, 1181–1192.

Loko LY, Ranjana B, Agre AP, Dossou-Aminon I, Orobiyi A, Djedatin GL and Dansi A (2017) Genetic diversity and relationship of Guinea yam (*Dioscorea cayenensis* Lam.– *D. rotundata* Poir. complex) germplasm in Benin (West Africa) using microsatellite markers. *Genetic Resources and Crop Evolution* **64**, 1205–1219.

Loko LEY, Fagla SM, Orobiyi A, Glinma B, Toffa J, Koukoui O, Djogbenou L and Gbaguidi F (2019) Traditional knowledge of invertebrates used for medicine and magical–religious purposes by traditional healers and indigenous populations in the Plateau Department, Republic of Benin. *Journal of Ethnobiology and Ethnomedicine* **1**, 66–78.

Maechler M, Rousseeuw P, Struyf A, Hubert M and Hornik K (2021) Cluster: cluster analysis basics and extensions. R package version 2.1.2—For new features, see the 'Changelog' file (in thepackage source), https://CRAN.R-roject.org/package=cluster.

Mignouna H and Dansi A (2003) Yam (*Dioscorea* ssp.) domestication by the Nago and Fon ethnic groups in Benin. *Genetic Resources and Crop Evolution* **50**, 519–528.

Mignouna HD, Ellis NTH, Knox MR, Asiedu R and Ng QN (1998) Analysis of genetic diversity in Guinea yams (*Dioscorea* spp.) using AFLP fingerprinting. *Tropical Agriculture (Trinidad)* **75**, 224–229.

Mignouna H, Mank R, Ellis T, Van Den Bosch N, Asiedu R, Ng S and Peleman J (2002a) A genetic linkage map of Guinea yam (*Dioscorea rotundata* Poir.) based on AFLP markers. *Theoretical and Applied Genetics* **5**, 716–725.

Mignouna H, Mank R, Ellis T, Van Den Bosch N, Asiedu R, Abang M and Peleman J (2002b) A genetic linkage map of water yam (*Dioscorea alata* L.) based on AFLP markers and QTL analysis for anthracnose resistance. *Theoretical and Applied Genetics* **5**, 726–735.

Mignouna DB, Akinola AA, Abdoulaye T, Alene AD, Manyong V, Maroya NG, Aighewi BA, Kumar LP, Balogun M, Lopez-Montes A, Rees D and Asiedu R (2020) Potential returns to yam research investment in sub-Saharan Africa and beyond. *Outlook on Agriculture* **4**, 1–10.

Missihoun AA, Ahanhanzo C, Agbangla C, Dansi A and Akpagana KK (2009) Evaluation génétique du processus de la domestication des ignames sauvages *Dioscorea abyssinica*, dans l'état de Nasarawa (Nigéria) par les marqueurs microsatellites (SSR). Comparaison des ignames prédomestiquées avec les variétés locales des ignames cult. *Journal de la Recherche Scientifique de l'Université de Lomé* **1**, 39–48.

Mondo JM, Agre PA, Edemodu A, Adebola P, Asiedu R, Akoroda MO and Asfaw A (2020) Floral biology and pollination efficiency in yam (*Dioscorea* spp.). *Agriculture* **10**, 560.

Mondo JM, Agre PA, Asiedu R, Akoroda MO and Asfaw A (2021a) Genome-wide association studies for sex determination and cross-compatibility in water yam (*Dioscorea alata* L.). *Plants* **10**, 1412.

Mondo JM, Agre PA, Edemodu A, Asiedu R, Akoroda MO and Asfaw A (2021b). Cross-compatibility analysis in intra- and interspecific breeding relationships in yam (*Dioscorea* spp.). Under review for publication in Scientific Reports.

Mulualem T, Mekbib F, Shimelis H, Gebre E and Amelework B (2018) Genetic diversity of yam (*Dioscorea* spp.) landrace collections from Ethiopia using simple sequence repeat markers. *Australian Journal of Crop Science* **8**, 1835–2707.

Nelimor C, Badu-Apraku B, Garcia-Oliveira AL, Tetteh A, Agre PA, N'guetta ASP and Gedil M (2020) Genomic analysis of selected maize landraces from Sahel and Coastal West Africa reveals their variability and potential for genetic enhancement. *Genes* **9**, 1054–1065.

Nkhoma N, Shimelis H, Laing MD, Shayanowako A and Mathew I (2020) Assessing the genetic diversity of cowpea [*Vigna unguiculata* (L.) Walp.] germplasm collections using phenotypic traits and SNP markers. *BMC Genetics* **21**, 110–126.

Oksanen J, Guillaume BF, Friendly M, Roeland KP, Dan ML, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Henry H, Szoecs E and Wagner H (2019) vegan: Community ecology package. R package version 2.5–5. Available at https://CRAN.R-project.org/package=vegan.

Orgogozo V, Morizot B and Martin A (2015) The differential view of genotype and phenotype relationships. *Frontiers in Genetics* **6**, 179–193.

Ousmael KM, Tesfaye K and Hailesilassie T (2019) Genetic diversity assessment of yams (*Dioscorea* spp.) from Ethiopia using inter simple sequence repeat (ISSR) markers. *African Journal of Biotechnology* **30**, 970–977.

Peakall R and Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics (Oxford, England)* **28**, 2537–2539.

Peres-Neto PR, Jackson DA and Somers KM (2003) Giving meaningful interpretation to ordination axes: assessing loading significance in principal component analysis. *Ecology* **84**, 2347–2363.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D and Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* **3**, 559–575.

Salazar E, González M, Araya C, Mejía N and Carrasco B (2017) Genetic diversity and intra-racial structure of Chilean Choclero corn (*Zea mays* L.) germplasm revealed by simple sequence repeat markers (SSRs). *Scientia Horticulturae* **225**, 620–629.

Sansaloni CP, Petroli CD, Carling J, Hudson CJ, Steane DA, Myburg AA and Kilian A (2010) A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in *Eucalyptus*. *Plant Methods* **1**, 16–27.

Sartie A, Asiedu R and Franco J (2012) Genetic and phenotypic diversity in a germplasm working collection of cultivated tropical yams (*Dioscorea* spp.). *Genetic Resources and Crop Evolution* **8**, 1753–1765.

Scarcelli N, Tostain S, Vigouroux Y, Agbangla C, Daïnou O and Pham JL (2006) Farmers' use of wild relative and sexual reproduction in a vegetatively propagated crop. The case of yam in Benin. *Molecular Ecology* **9**, 2421–2431.

Scarcelli N, Chaïr H, Causse S, Vesta R, Couvreur TL and Vigouroux Y (2017) Crop wild relative conservation: wild yams are not that wild. *Biological Conservation* **210**, 325–333.

Scarcelli N, Cubry P, Akakpo R, Thuillet AC, Obidiegwu J, Baco MN and Vigouroux Y (2019) Yam genomics supports West Africa as a major cradle of crop domestication. *Science Advances* **5**, 56–68.

Siadjeu C, Mayland-Quellhorst E and Albach DC (2018) Genetic diversity and population structure of trifoliate yam (*Dioscorea dumetorum* Kunth) in Cameroon revealed by genotyping-by-sequencing (GBS). *BMC Plant Biology* **18**, 359–373.

Stanley A, Menkir A, Agre P, Ifie B, Tongoona P, Unachukwu N, Meseka S, Mengesha W and Gedil M (2020) Genetic diversity and population structure of maize inbred lines with varying levels of resistance to striga hermonthica using agronomic trait-based and SNP markers. *Plants* **9**, 1223–1239.

Sugihara Y, Darkwa K, Yaegashi H, Natsume S, Shimizu M, Abe A, Hirabuchi A, Ito K, Oikawa K, Muluneh T, Ohta A, Matsumoto R, Agre P, De Koeyer D, Pachakkil B, Yamanaka S, Muranaka S, Takagi T, White B, Asiedu R, Innan H, Asfaw A, Adebola P and Terauchi R (2020) Genome analyses reveal the hybrid origin of the staple crop white Guinea yam (*Dioscorea rotundata*). *Proceedings of the National Academy of Sciences* **50**, 31987–31992.

Terauchi R and Kahl G (1999) Mapping of the *Dioscorea tokoro* genome: AFLP markers linked to sex. *Genome* **4**, 752–762.

Tostain S, Agbangla C, Scarcelli N, Mariac C, Daïnou O, Berthaud J and Pham JL (2007) Genetic diversity analysis of yam cultivars (*Dioscorea rotundata* Poir.) in Benin using simple sequence repeat (SSR) markers. *Plant Genetic Resources* **2**, 71–81.

Zannou A, Ahanchédé A, Struik PC, Richards P, Zoundjihékpon J, Tossou R and Vodouhè S (2004) Yam and cowpea diversity management by farmers in the Guinea-Sudan transition zone of Benin. *NJAS-Wageningen Journal of Life Sciences* **4**, 393–420.