Vol. 15(2), pp. 74-85, April-June 2023
DOI: 10.5897/JPBCS2023.1013
Article Number: 0FE227370737
ISSN 2006-9758

http://www.academicjournals.org/JPBCS

ACADEMIC JOURNALS
expand your knowledge

Journal of Plant Breeding and Crop Science

Full Length Research Paper

Diversity of *Dioscorea bulbifera* Linn in Uganda assessed by morphological markers and genotyping-by-sequencing technology (GBS)

Hilda Ikiriza[1]*, Hedmon Okella[1], Naasson Tuyiringiye[5], Ali Milton[4], Nuwagira Catherine[1], Raphael Wangalwa[1], Grace Kagoro-Rugunda[1], Abubaker Muwonge[2], Mai Abdalla Ali Abdalla[3], and Casim Umba Tolo[1]

[1]Department of Biology, Faculty of Science Mbarara University of Science and Technology, Mbarara, Uganda.
[2]Department of Biochemistry and Molecular Biology, Faculty of Science University of Gezira, Wad Madani, Sudan.
[3]Department of Biological Sciences, Faculty of Science, Kyambogo University, Kampala, Uganda.
[4]National Agricultural Research Laboratories (NARL), Kawanda, Uganda.
[5]School of Nursing and Midwifery, College of Medicine and Health Sciences, University of Rwanda, Rwanda.

Received 3 February, 2023; Accepted 3 April, 2023

***Dioscorea bulbifera* L. is a widely distributed plant in the humid and sub-humid tropics. The taxonomic treatment of this species is ambiguous due to its extreme polymorphic morphological characters. Therefore, it is imperative to assess the species distribution, and the inter and intraspecific variability relationships in *D. bulbifera*. In this study, the genetic diversity and species distribution of 41 accessions of *D. bulbifera* in Uganda were investigated. Diversity among the 41 accessions was assessed using distribution of accessions, morphological descriptors and SNP markers. To investigate the relationship between individuals and populations, a heat map, phylogenetic tree, and principal components analysis were used. A total of 221455432 paired-end raw reads were generated out of which 162662 SNP loci were called with 150 bp reads (2×150) to a targeted depth of 112 million reads. Analysis of SNP data has proved a heterogenous population of *D. bulbifera* with an exception of Abur accessions that stand out as homogenous populations. This study has revealed that the utilization of advanced sequencing techniques such as GBS in combination with statistical analysis is a robust method for evaluating genetic diversity in a complex crop such as aerial yam.**

Key words: Aerial yam (*Dioscorea bulbifera* Linn), genetic diversity, SNP markers, Uganda.

INTRODUCTION

Yams are classified under genus *Dioscorea* which contains approximately 600 species (Hahn et al., 1995),

*Corresponding author. E-mail: hilidaikiriza2@gmail.com.

of which *Dioscorea rotundata*, *Dioscorea alata*, *Dioscorea bulbifera*, *Dioscorea esculenta*, and *Dioscorea dumenterum* are the most economically important staple foods in the tropical and sub-tropical regions (Ranjana et al., 2011). They provide food to over 100 million people in many developing countries (Lebot, 2009). Eaten for its carbohydrates (Tetchi et al., 2007), protein, vitamins, potassium, sodium, magnesium, copper, and zinc (Shajeela et al., 2011), yams also have medicinal properties derived from the steroidal sapogenins that are used in the production of cortisone and synthetic steroid hormones (Narula et al., 2007). *D. bulbifera* in particular has numerous therapeutic applications in various patho-physiological conditions namely ulcers, sores, wound, spasms, dysentery, diabetes, and cancer (Sharma and Code, 2017).

The antioxidant, anti-inflammatory, antibacterial, plasmid curing, anti-diabetic, anticancer (Ikiriza et al., 2019; Worou, 2018; Anandpara and Tirgar, 2017) and anti-fertility properties (Priya et al., 2002) of *D. bulbifera* are ascribed to a bioactive component called diosgenin. Given its food and medicinal importance, *D. bulbifera* has been over exploited in Uganda, necessitating conservation of its genetic resources for sustainable use. Trimanto and Hapsari (2015) suggested that promoting forest conservation and conducting germplasm collection backup can improve *D. bulbifera* conservation status. To effectively enhance the use of genetic resources, foremost it is important to increase our knowledge of the conserved germplasm through characterization. Unfortunately, the germplasm of *D. bulbifera* in Uganda has not been characterized, yet germplasm characterization is an initial step to conservation (Bioversity International, 2003). Elsewhere, several studies have been done to characterize morphological traits of *Dioscorea* species including *D. alata* L (Anokye et al., 2014), Amerindian yam (Beyerlein et al., 2018), and *D. bulbifera* (Atieno et al., 2020). Morphological traits are however less informative and exhibit a lot of variations attributed to agro-climatic fluctuations. Consequently, molecular markers are extensively used to evaluate the genetic relatedness in germplasm. The first-generation molecular markers namely Random Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphism (AFLP), Inverse Sequence Tagged Repeat (ISTR) and inter simple sequence repeat (ISSR) have been used to analyze genetic variation within *Dioscorea* spp. (Rao et al., 2020, Velasco-Ramírez et al., 2014). These markers are however not cost effective, and with poor discrimination within *Dioscorea* species (Cormier et al., 2019). The advent of next generation sequencing (NGS) protocols has revolutionized phylogenetic diversity studies. One such protocol is Genotype by Sequencing (GBS) that has been used for genome-wide genotyping of organisms for molecular markers discovery. It is a reliable, simple, and cost-effective method of SNP calling and mapping. The method is useful in fishing out SNPs within a population to better understand the relationships between and within species for genomic selection, prediction, or trait analysis. This method has been utilized successfully on Cassava (*Manihot esculenta* Crantz) (Rabbi et al., 2015), guinea yam (Girma et al., 2014), water yam (Saski et al., 2015), and *Daucus carota* (Arbizu et al., 2016), justifying its power to as a suitable technology for high throughput genotyping in yam.

Due to many ecological restrictions and study neglect, the genetics of yams is the least understood and has remained mostly ignored among the major staple food crops (Mignouna et al., 2003). In this regard, *D. bulbifera* from various regions in Uganda have not been characterized, not to mention using single nucleotide polymorphism genotyping to analyze species distribution and genetic diversity of *D. bulbifera*. Therefore, this study investigated the genetic relationships and species distribution of Aerial yam (*D. bulbifera*) accessions in Uganda using GBS for further improvement of this very important species in Uganda.

MATERIALS AND METHODS

Sample collections sites

Uganda is located in Eastern Africa, between latitudes of 4.5°N and 1.5°S, and longitudes of 29.5°E and 35°E (Figure 1), with an average altitude of 1,100 m above sea level. The sample collection of *D. bulbifera* was conducted in six geolocations, that is, Mabira Central Forest Reserve, Bugoma Central Forest Reserve, Bwindi Impenetrable National Park, Kasyoha Kitomi forest reserve, Mt Elgon forest reserve, and Abur forest (Figure 1). The six geolocations were selected based on the presence of *D. bulbifera* and agroecological zone (AEZ) classification (MAAIF, 2014). The six geolocations are encircled by varying agroecological zones. These zones have been determined by soil types, climate, landforms, and socio economic and cultural factors (Kabi et al., 2014).

Sample collection and processing

The plant materials used for morphological characterization comprised of forty-one accessions of *D. bulbifera* conserved in the selected forest reserves of Uganda namely Mabira, Bugoma, Bwindi, Kasyoha Kitomi, Mt Elgon, and Abur (Figure 1). The collected leaf samples and bulbils were transported to the National Herbarium in Makerere University Department of Plant science, Microbiology and Biotechnology, Kampala Uganda where the samples were identified and given the Accession number of 50974 and voucher number 20. Data on Morphological differences was collected using available descriptors i.e., tuber shape, tuber cortex color, tuber flesh color, skin texture, and presence or absence of warts on the tuber cortex, young leaf color (Ipgri and Iita 1997). Each tuber was cut longitudinally to assess the colour of flesh and the texture was assessed by hand feeling and observation. For genetic characterization and molecular analyses, young leaves from each accession were silica dried and sent to the University of Wisconsin-Madison Biotechnology Centre, U.S. Genomic DNA was extracted using QIAGEN DNeasy mericon 96 QIAcube HT Kit. DNA was quantified using the Quant-iT™ PicoGreen® dsDNA kit (Life Technologies, Grand Island, NY).

Preparation of libraries and sequencing

D. bulbifera GBS Libraries were prepared as in Elshire et al. (2011)

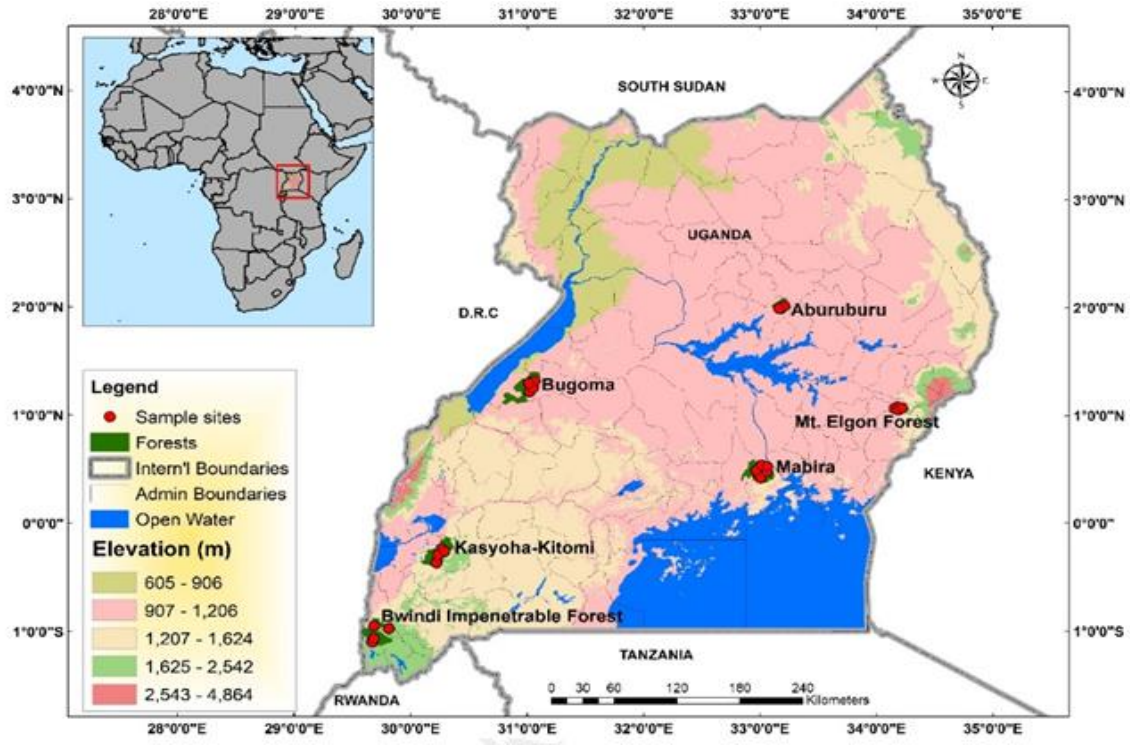

Figure 1. Location map of the six forest reserves of Uganda where the 41 accessions were sampled.
Source: Drawn using ArcGIS version 10.5

with minimal modification; GBS libraries were prepared and analysed at the University of Wisconsin-Madison Biotechnology Center's DNA Sequencing Facility. Apek 1 restriction enzyme was used for digestion and for creating a library containing 96 unique barcodes. The barcodes were PCR amplified to provide library quantities amenable for sequencing, and adapter dimers were removed by solid phase reverse immobilization (SPRI) bead purification. Quality and quantity of the finished libraries were assessed using the Agilent Tapestation (Agilent Technologies, Inc., Santa Clara, CA) and Qubit® dsDNA HS Assay Kit (Life Technologies, Grand Island, NY), respectively. Prepared Libraries were sequenced on Illumina NovaSeq 6000 2×150 S2. Resulting images were analysed using the standard Illumina Pipeline, version 1.8.2 to generate 150 bp raw reads for SNP calling.

Processing of illumina raw sequence read data and SNP calling

A modified version of the non-reference GBS SNP calling pipeline Universal Network Enabled Analysis Kit (UNEAK) as implemented in Tassel Version 3.0.160 (Lu et al., 2013), was used for SNP calling using *VCF* tools version v0.1.10 (Danecek et al., 2011). The generated SNPs were by filtered to remove genotypes with quality scores less than 98 (-GQ 98), and SNP loci with more than 90% missing data (-geno 0.1). Multi-dimensional scaling analysis (MDS) was conducted using *PLINK* version v1.07 (Purcell et al., 2007). The VCF file developed was filtered using criteria of minor allele frequency (MAF) >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 (Danecek et al., 2011) and *PLINK* v1.07 (Purcell et al., 2007) for filtering.

Data analysis

Data on morphological characterization was analyzed using *SPSS* version 21 (IBM Corp., Armonk, N.Y., USA) and a Chi square test was used to determine the relationship between place of plant collection and morphological traits. Pairwise correlations between the SNP profiles of all collected samples were done to investigate the relationship between individuals and populations. To investigate the evolution of the six populations, phylogenetic inference was performed by maximum likelihood using the IQ-TREE software (Minh et al., 2020), with the filtered VCF file acquired in the previous step as input. IQ-TREE requires input files in the PHYLIP format, and so VCF was converted to PHYLIP using the software *Vcf2phylip* (Ortiz, 2019). IQ-TREE was run on default settings according to the documentation, and the appropriate model used for the phylogenetic inference was selected using *Model Finder* (Ortiz, 2019). The obtained phylogenetic tree was then plotted using ggtree in R software (Yu, 2020). To summarize the variance observed between samples, the principal component analysis (PCA) was performed.

RESULTS

Assessment of morphological traits

Phenotypic descriptors have been extensively used for plant genetic resources management and conservation (Zamir, 2013). The use of morphological traits in the classification of crops is very essential due to their relative stability over quantitative traits (Jyothy et al.,

Table 1. Morphological traits of accessions from different locations.

Morphological traits	Category	Place of plant collection						Total%	*p*-value
		Db	M	Bu	E	KK	A		
Young leaf color	Dark Green	5	6	0	4	2	7	58.5	< 0.0001
	Yellowish green	0	1	13	3	0	0	41.5	
	Total	5	7	13	7	2	7		
Leaf surface	Smooth	0	7	0	0	0	7	34.1	< 0.0001
	Rough	5	0	13	7	2	0	65.9	
	Total	5	7	13	7	2	7		
Tuber shape	Irregular	5	5	13	6	2	7	92.7	>0.05
	Bent V	0	1	0	0	0	0	2.4	
	Flattened top v	0	1	0	0	0	0	2.4	
	Heart shape	0	0	0	1	0	0	2.4	
	Total	5	7	13	7	2	7		
Tuber cortex color	Dark Brown	5	7	12	3	0	0	65.9	< 0.0001
	Greenish brown	0	0	1	1	1	0	7.3	
	Greenish purple	0	0	0	1	0	0	2.4	
	Light brown	0	0	0	2	1	7	24.4	
	Total	5	7	13	7	2	7		
Tuber flesh color	Yellow	5	6	12	2	2	0	65.9	< 0.0001
	Purple	0	1	1	3	0	0	12.1	
	Brownish yellow	0	0	0	1	0	7	19.5	
	Yellowish green	0	0	0	1	0	0	2.4	
	Total	5	7	13	7	2	7		
Tuber texture	Rough	5	7	11	6	2	7	92.7	>0.05
	smooth	0	0	2	1	0	0	7.3	
	Total	5	7	13	7	2	7		
Warts on the tuber cortex surface	Absent	5	5	12	6	2	7	90.2	>0.05
	Present	0	2	1	1	0	0	9.8	
	Total	5	7	13	7	2	7		

N=41, E=Elgon, BU=Bugoma, M=Mabira, A=Abur, KK=Kasyoha Kitomi, Db=Bwindi.
Source: Drawn using SPSS version 21.

2017). Wide variation was observed in the seven qualitative characters used in the differentiation of the accessions. There were two different young leaf colours observed namely, yellowish green, and dark green. In all, 41.5% showed yellowish green colour, while 58.5% had dark green colour. Two different leaf surfaces were observed among *D. bulbifera* accessions. It was observed that 34.1% was smooth while 65.9% of the leaf surfaces were rough. Variation was observed in the tuber shapes among the aerial yam accessions including Irregular shape, Bent V shape, Flattened top v shape and Heart shape with the aerial yam accessions shapes distributed as irregular (92.7%), bent v (2.4%), flattened top v (2.4%), and heart (2.4%) (Table 1). The tuber cortex colour showed high variability among the accessions studied and four different cortex colours were observed. Among the accessions, majority (65.9%) had dark brown, followed by greenish brown (7.3%), greenish purple (2.4%), and light brown (24.4%) cortex colour.

Wide variation was observed in tuber flesh colour among 41 accessions of *D. bulbifera* accessions studied. Majority of the accessions (65.9%) had yellow colour, 12.1% showed purple colour, 19.5% had brownish yellow colour and 2.4% exhibited yellowish green coloured flesh. Significant phenotypic variability was identified among the individual accessions with respect to texture of tuber.

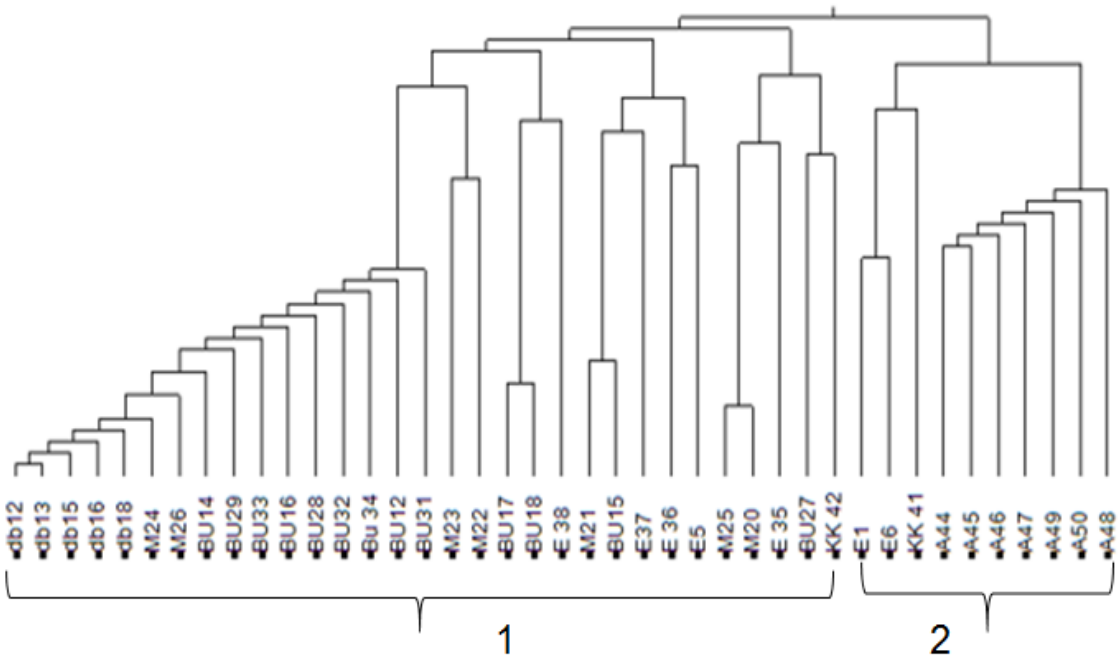

Figure 2. Phylogenetic clustering of *D. bulbifera* accessions based on morphological traits.
Source: Drawn using SAS JMP ver. 10.0, Hierarchical Ward Method

Two types of tuber texture were recorded; rough and smooth. Among the accessions 92.7% had rough texture, while 7.3% had exhibited smooth texture. Significant phenotypic variability was identified among the individual accessions with respect to presence or absence of warts on the tuber cortex surface. Among the accessions 90.2% had no warts on the tuber cortex surface, while 9.8% had warty tuber cortex surface (Table 1).

Morphological phylogeny of *D. bulbifera*

Based on morphological traits, the accessions were grouped into two major clusters (Figure 2). The first cluster contains mixed populations from Bwindi (Db), Mabira (M), Bugoma (Bu), Kasyoha Kitomi (K.K), and Elgon (E). The second Cluster grouped together all accessions all from Abur (A), 2 accessions from E and 1 accession from KK. These major clusters were also further grouped into sub clusters and the first major cluster comprised 5 sub clusters with sub cluster 1 consisting of all the *D. bulbifera* accessions from Db, 2 from M and 9 accessions from Bu between Bu 14 and Bu 31. Sub cluster 2 consisted of only two *D. bulbifera* accessions from Mabira central forest reserve, followed by sub cluster 3 that contained Bu 17, Bu 18 and E 38 that standing separately. Sub cluster 4 was, also mixed up with *D. bulbifera* accessions of M21, Bu 15, E37, E36, and E5 standing separately. Sub cluster 5 groups *D. bulbifera* accessions of M25, M20, E35 standing separately and Bu27 and KK42 grouping together. The second major cluster contained only two subclusters; the first one containing accessions of E1 and E2 grouping together while K.K stands separately and the second subcluster contained all accessions from Abur forest. Generally, accessions of *D. bulbifera* from Abur seem to be different from other accessions from other forests, while those from the rest of the forests including Bwindi, Mabira, Bugoma, Elgon and Kasyoha Kitomi mixed up and show relatedness.

Genetic diversity analysis using Single Nucleotide Polymorphism markers

The SNP markers were also used to study the variation of germplasm at a molecular level. A total of 164049 SNPs were identified. Poor quality scores with less than 98% and SNP loci with more than 90% missing data were removed remaining with a total of 162662 SNP loci. Of the remaining filtered SNPs, 68.98% SNPs were transition and 31.02% were transversions giving a ratio of 2.22. The C/T transitions (20.39%) accounted for the highest frequency while T/G transversions (2.99%) occurred for the lowest frequency among all the SNP scenarios. The frequencies of two transition types were similar (that is G/A 17.66% and A/G 17.26%) while the frequencies of the eight transversion types ranked as follows: C/G 5.06%, G/T 4.37%, G/C 4.03%, A/C 3.86%, C/A 3.83%, A/T 3.69%, T/A 3.17%, and T/G 2.99%

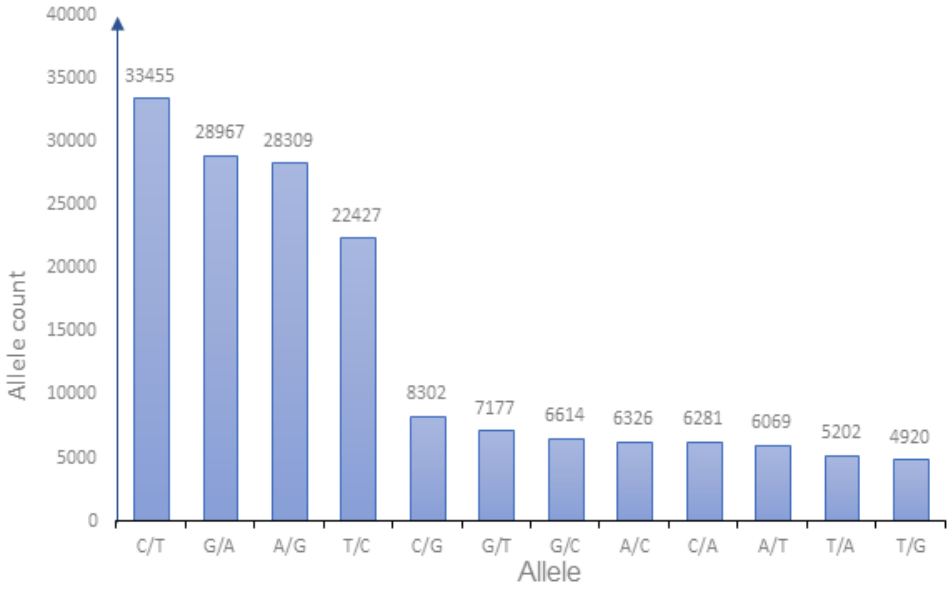

Figure 3. Statistics of the Single Nucleotide Polymorphism data of *D. bulbifera*. (C/T) indicate that C is a reference allele and T is the alternate allele.
Source Drawn using minitab version 17

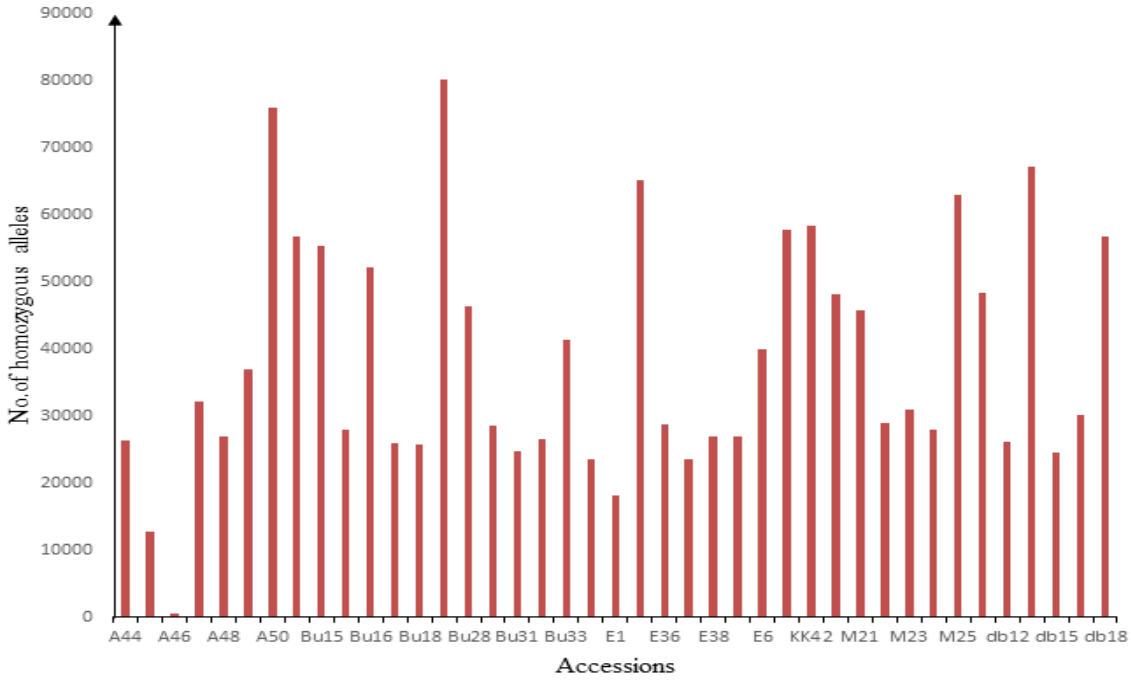

Figure 4. Number of homozygous single-nucleotide polymorphisms per accession.
Source: Drawn using minitab version 17

(Figure 3). To further investigate the diversity of *D. bulbifera*, the SNP types were investigated. A total of 533718 heterozygous and 1569708 homozygous SNPs were identified among the accessions. The detailed numbers of heterozygous and homozygous SNPs per accession are as shown in Figures 4 and 5. Heterozygous SNPs were used to compute the Shannon's Information index which is a measure of genetic diversity. The

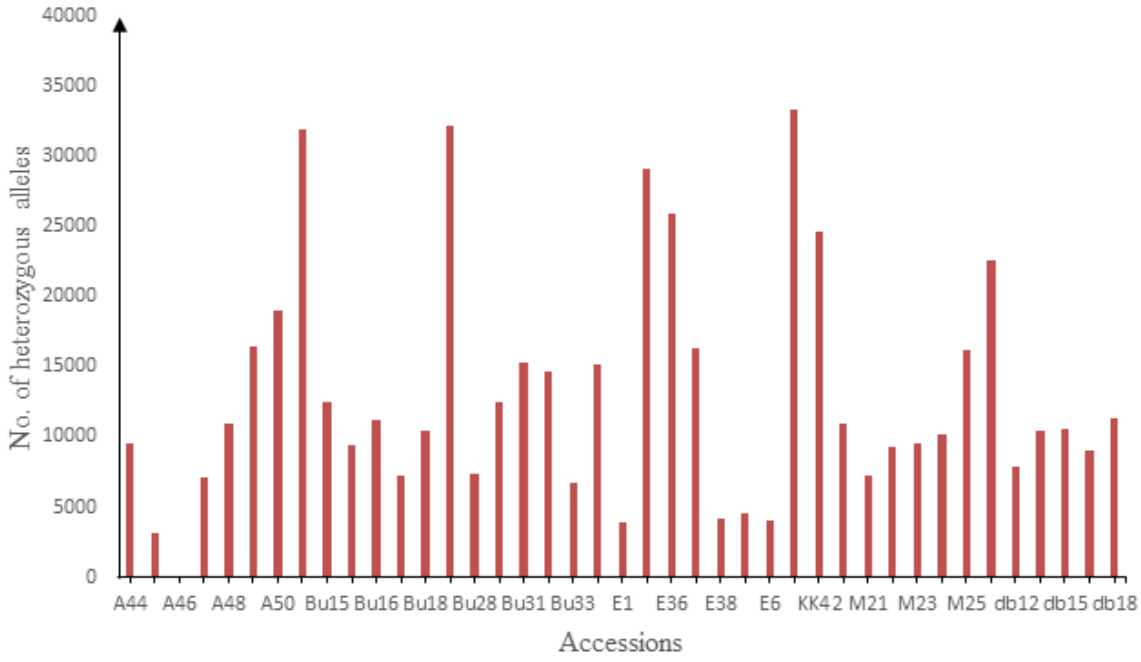

Figure 5. Number of heterozygous single-nucleotide polymorphisms per accession.
Source: Drawn using minitab version 17

Table 2. Genetic diversity indices for *Dioscorea bulbifera* populations from different forest.

Popn	N	Na	% Na	Ho	%Ho	He	%He	I
Abur	7	124396.9	75.8	30198	18.4	9454.1	5.8	1.9
Bugoma	13	110045.4	67.1	39649.2	24.2	14354.5	8.8	2.5
Elgon	07	118728.1	72.4	32762.6	19.9	12558.3	7.7	1.9
K.K	02	76994.5	46.9	58065	35.4	28989.5	1.7	0.7
Mabira	07	109962.4	67.0	41819.7	25.5	12266.9	7.5	1.9
Bwindi	05	113278.4	69.1	40935.4	24.9	9835.2	5.9	1.6

N, Number of accessions, Na, number of alleles from each *D. bulbifera* population, Ho, homozygosity, He, heterozygosity estimated with computer program GenAlex 6.503, I, Shannon Information Index estimated with computer program Past 4.11 (Peakall and Smouse, 2012), calculated mean I = 1.75.
Source: Drawn using Microsoft excel.

Shannon's Information index (mean = 1.75, Table 2) revealed high genetic diversity among *D. bulbifera* accessions. Among the *D. bulbifera* populations sampled, Abur registered the highest number of alleles (124396.9) followed by Elgon (118728.1) while Kasyoha Kitomi had the lowest (76994.5) (Table 2). The Shannon information index (I = 2.5) and the genetic diversity (He = 0.088), portrayed accessions from Bugoma to be the most genetically diverse while accessions from Kasyoha Kitomi (He = 0.017, I = 0.7) registered the lowest genetic diversity.

The variation of accessions within a population and among populations was computed using a heat map. Each row and column are same set of accession annotated by the same cluster dendrogram of similarity distance. Additionally, each row and column are color coded to its associated correlation module. The deep blue color indicated close similarity in accessions and the yellow color indicated dissimilarity of accessions. The correlation map revealed that accessions belonging to the same population tend to be more strongly correlated to each other than to other populations on the genotype level (Figure 6). Samples A44-50 form a strongly correlated cluster with the exceptions of A49 and A50.

This means that 5/7 samples from the Abur population were strongly correlated, showing only weak correlations with samples collected from other populations. However, the genotypes of A49-50 seem more strongly correlated

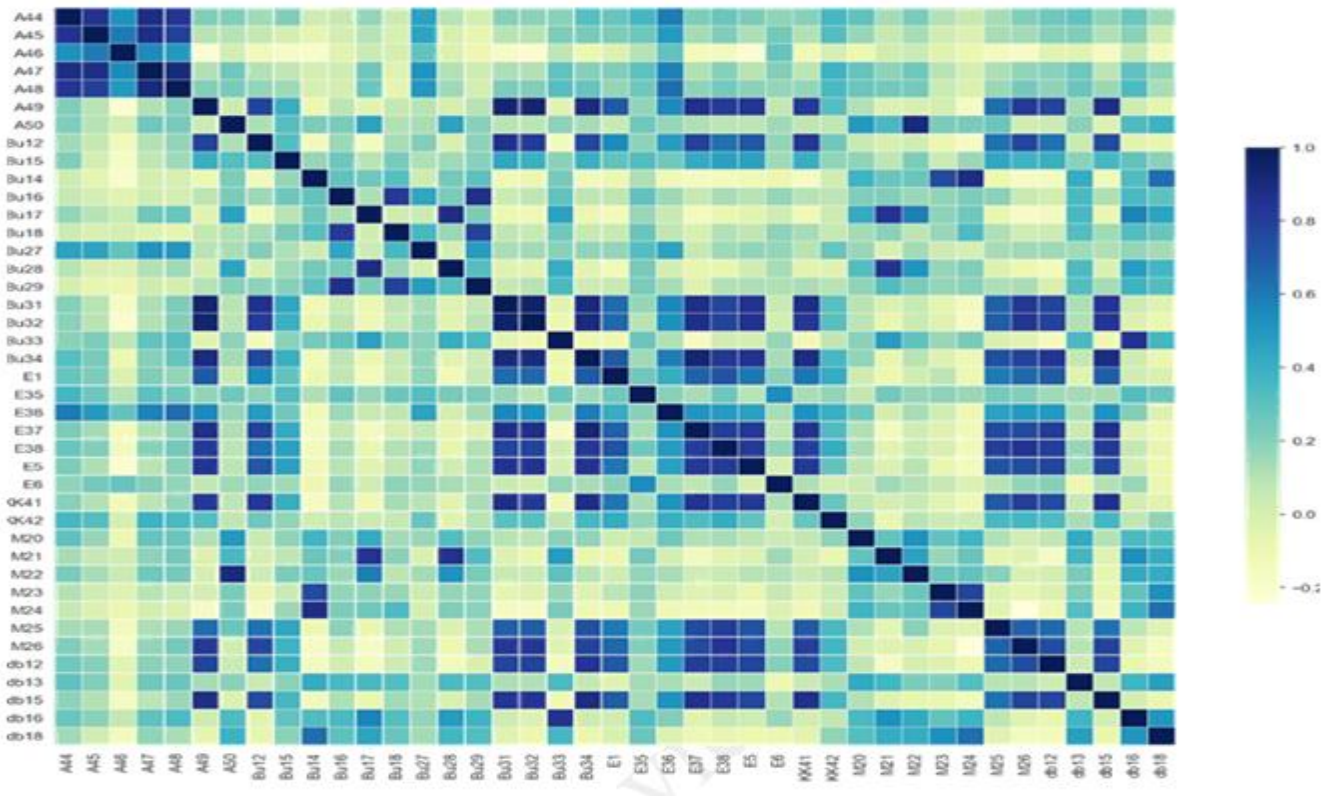

Figure 6. Heat map showing the variation within and between populations.
Source:Drawn using SPSS version 21.

with several members of the Bugoma population. Similar patterns were observed for other samples as well (Bu33, E35). The genotype correlations of samples from the Bugoma population (Bu12-Bu34) are harder to interpret as they seem to form a merged cluster with samples from the Elgon population (E1-E38). This suggests that the genotypes of individuals from the Bugoma and Elgon populations are quite similar, leading to high correlation. Notably, Bu12-15 are moderately correlated with A49-50 while samples Bu14-29 formed a weakly correlated cluster, separate from the merged Bugoma/Elgon clusters. This might suggest a large genotype diversity within the Bugoma population. Samples from the Kasyoha Kitomi, (KK41-42), Mabira (M20-26) and Bwindi (db12-18) populations show similar patterns of moderate to strong correlations with samples from other forest populations.

To illustrate the genetic diversity among the populations, the phylogenetic tree was drawn (Figure 7). Three major clusters stood out with the first major cluster grouping accessions from M21-Bu27, the second major cluster contained accessions between M25-A48 and the last one contained accessions between E6-A44. The tree had a number of clusters, often composed of accessions from different populations. However, accessions from the Abur (A44-A47) stood out as the most homogenous. Notably is that accessions from Bwindi impenetrable national park and Mount Elgon Forest reserve are distributed throughout all the major clusters. Generally, the dendrogram inferred from genetic data base was better resolved than the one from morphological samples since it clearly brings out the diversity of *D. bulbifera* from forests in different agro ecological zones in Uganda and can thus be used to always complement the traditional morphology-based method for phylogenetic studies (Amit Roy, 2014).

To further elucidate the genetic relatedness among accessions and populations, a principal component analysis (PCA) was done. The two first principal components captured ~38% of the variance observed between samples (Figure 8). However, it was difficult to distinguish the populations from the PCA results as the samples formed a large central cluster, indicating high similarity between the genotypes. The Abur population seemed to be the most distinct population as it separated most clearly from the other population in the PCA.

DISCUSSION

In this study, *D. bulbifera* from Uganda were found to vary both at morphological and genetic level. The study

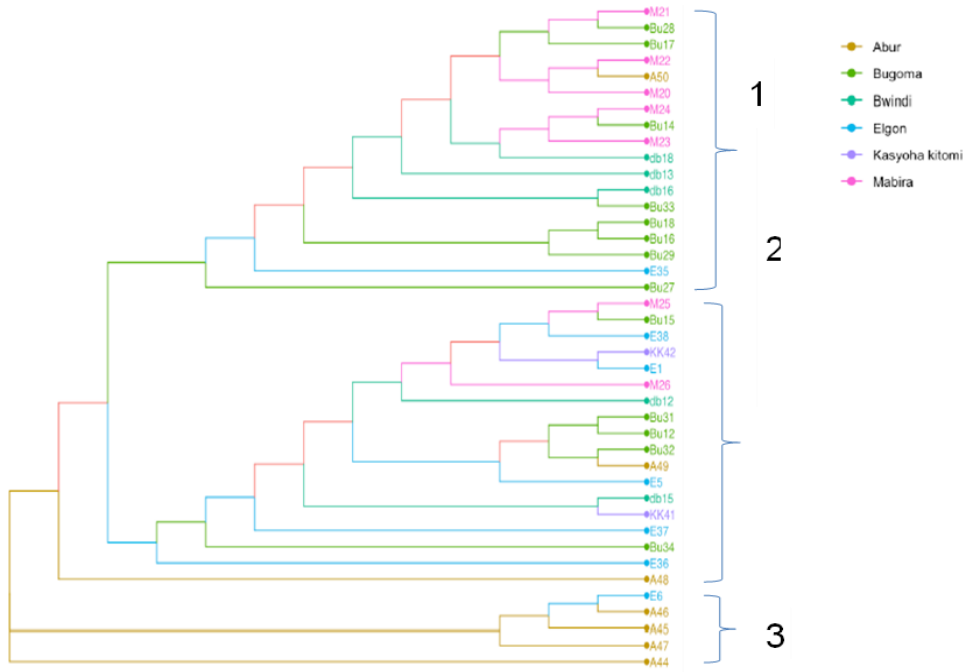

Figure 7. Phylogenetic tree of accessions based on Single nucleotide polymorphism markers.
Source: Drawn using R software.

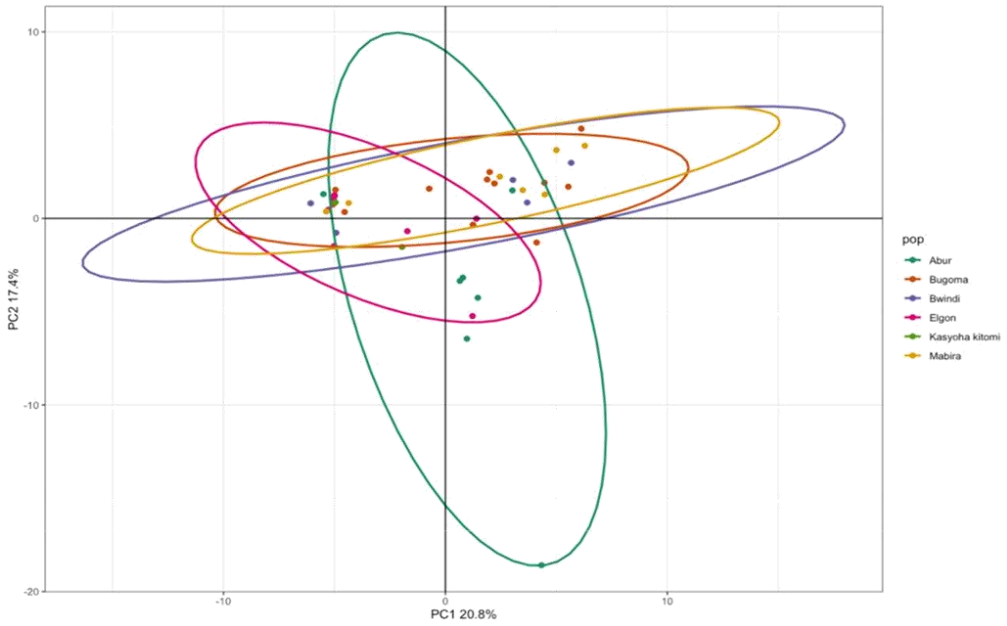

Figure 8. Principal components analysis of *D. bulbifera* accessions using SNP markers
Source: Drawn using SPSS version 21.

has its strength on the fact that GBS being one of the latest and reliable molecular techniques that utilizes organism's whole-genome to characterize and draw conclusion about its diversity. At a morphological level, tuber shape, tuber flesh color, tuber skin texture, tuber cortex color, leaf form, and young leaf color were found to vary with the collection sites ($p < 0.0001$) and similar results were reported in Jyothy et al. (2017) in *D. alata*.

This could mean that these morphological traits may be influenced by numerous environmental factors in the region from which the accessions were collected for example, water shortage and excess, temperature, nutrient availability, and light (Ji et al., 2017). Because these environmental factors influence morphological traits, there is more reason to utilize molecular methods to characterize this plant *D. bulbifera*.

The *D. bulbifera* accessions from Abur forest grouped separately and seemed to be a homogeneous population most probably because it has undergone divergent evolution leading to speciation. Also, *D. bulbifera* accessions from Elgon, Bugoma and Kasyoha Kitomi forests showed relatedness most probably because D. *bulbifera* is a fall back plant in famine, people can move long distances to preserve it and this has brought about plant migrations as confirmed by Siqueira et al. (2014). Conversely, genotyping by sequencing (GBS) has been increasingly used for genetic diversity analyses, gene identification, and plant breeding. This technique has been applied to genomic characterization of Ethiopian sorghum germplasm (Girma et al., 2020) analysis of switchgrass genomic diversity (Lu et al., 2013), development of genetic maps in barley and wheat (Alipour et al., 2019) and genome wide association studies in maize (Xiao et al., 2017). This study for the first time used GBS generated SNP markers for genetic characterization of *Dioscorea bulbifera.* Consistent with previous studies involving *Camellia sinensis* (Yang et al., 2016) and *Brassica napus* (Bus et al., 2012), transition SNPs were more frequent than transversions as observed *D. bulbifera*, indicating that transition mutations are better tolerated than transversion mutations during natural selection (Luo et al., 2017). This phenomenon is common in other plant species (Bus et al., 2012) and maybe due to synonymous mutations in protein-coding sequences (Guo et al., 2017).

However, it was difficult to distinguish the Kasyoha Kitomi and Bwindi populations on the genotype level basing on SNP data while, the Mabira population shows slightly higher internal correlation. The correlation map presented here indicated that the individual genotypes sampled from each population tend to be slightly more similar to each other than to those of other populations. However, a significant amount of correlation between the different populations was also observed, suggesting that phylogenetic inference might be difficult to perform. There was some similarity between and within the populations, hence the PCA was used. However, the PCA could not distinguish the accessions based on the populations. Similar results of such a PCA were reported by (Mbanjo et al., 2019). Most strikingly, the Abur population separated from the other populations in all the performed analyses. This indicates that the Abur population has undergone mutation and consequently divergent evolution compared to other *D. bulbifera* accessions other Agro Ecological Zones and this could be attributed to the difference in the environmental factors (Elena and Visser, 2003). This homogeneity in Abur populations was even observed using the morphological characteristics as indicated in Figure 2. The close correlation of Bugoma and Bwindi accessions could be attributed to the fact that both forests lie in the same region (south and mid-western), and people could have moved with them. *D. bulbifera* being a fallback plant in famine, people could move long distances to preserve it, leading to hybridization (Guo, 2014). The clustering together of Elgon and Abur accessions could be because there is no significant barrier separating these two forests. Thus, the mixing up of genotypes may not be ruled out. The diversity of accessions from Bwindi impenetrable forest (Db) and Mt Elgon forest reserve (E) could be as a result of protection status of these two collection forest reserves. This is because these two forest reserves are located in national Parks which are more strictly protected leading to ecological islands of the species.

Limitations

The lack of a reference genome for yams limited us from further assessment of molecular diversity in aerial yam. Specifically, the limitation of identifying biallelic SNPs that differ at only one base pair within a 64 base pair tag may lead to biases when estimating true rates of divergence within or among species due to mis identified or unobserved loci, especially when divergence rates are high. Also, it was not easy to locate *D. bulbifera* plants from most parts of the study forests and therefore the number of accessions from different forests could not be the same.

Conclusions

In this study, diversity of *D. bulbifera* accessions both in terms of phenotypic and genetic diversity parameters was demonstrated. Narrow genetic diversity among the *D. bulbifera* accessions in Uganda displays underutilization of the existing genetic potential. The studied accessions represent the total population of *D. bulbifera* in Uganda, and will inform crop improvement studies involving phenotyping across multiple environments, future genomics studies, and ultimately, selection of genotypes with desirable traits.

Recommendation

The GBS raw sequence data generated in this study should be re-analysed in the future using a reference sequence-based pipeline for calling SNPs once the genome sequence of *Dioscorea* becomes available. More so, this baseline study is expected to be an initial step

that may contribute into making a reference genome for *Dioscorea* spp.

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

ABBREVIATIONS

AEZ, Agroecological zone; **IPGRI,** International Plant Genetic Resources Institute; **IUCN,** international union for conservation of nature; **GBS,** genotyping-by-Sequencing; **MAF,** minor allele frequency; **MDS,** multi-dimensional scaling analysis; PCA, principal component analysis; **SNP,** single nucleotide polymorphism. **SPRI,** solid phase reverse immobilization; **UNEAK,** Universal Network Enabled Analysis Kit; **VCF,** variant call format.

ACKNOWLEDGEMENT

This work was supported by the Pharm-Biotechnology and Traditional Medicine Center (PHARMBIOTRAC).

REFERENCES

Alipour H, Bai G, Zhang G, Bihamta M R, Mohammadi V, Peyghambari SA (2019). Imputation accuracy of wheat genotyping-by-sequencing (GBS) data using barley and wheat genome references. PLoS One 14(1):0208614.

Anandpara RC, Tirgar PR (2017). Investigation of therapeutics benefits of *Dioscorea bulbifera* in breast cancer. Journal of Chemical and Pharmaceutical Research 9(6):313-326.

Anokye M, Tetteh JP, Otoo E (2014). Morphological characterization of some water yam (*Dioscorea alata* L.) germplasm in Ghana. Journal of Agriculture, Science and Technology ***4*:**518-532.

Arbizu CI, Ellison SL, Senalik D, Simon PW, Spooner DM (2016). Genotyping-by-sequencing provides the discriminating power to investigate the subspecies of Daucus carota (Apiaceae). BMC Evolutionary Biology 16:1-16.

Atieno V, Gatheri G W, Kamau JW, Muthini M (2020). Morphological and molecular characterization of cultivated yam (*Dioscorea* species) in selected counties in Kenya. African Journal of Plant Science 14(7):270-279.

Beyerlein P, Pereira HS (2018). Morphological diversity and identification key for landraces of the Amerindian yam in central Amazon. *Pesquisa Agropecuária Brasileira* 53*(*4):405-418.doi:10.1590/s0100-204x2018000400002.

Bioversity International/ILRI, Addis Ababa, Ethiopia (2003). Characterization and Evaluation Approaches. IPGRI Handbooks for Genebanks 6:73-74.

Bus A, Hecht J, Huettel B, Reinhardt R, Stich B (2012). High-throughput polymorphism detection and genotyping in Brassica napus using next-generation RAD sequencing. BMC genomics 13(1):1-11.

Cormier F, Mournet P, Causse S, Arnau G, Maledon E, Gomez RM, Chair H (2019). Development of a cost-effective single nucleotide polymorphism genotyping array for management of greater yam germplasm collections. Ecology and Evolution 9(10):5617-5636.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA (2011). The variant call format and VCF tools. Bioinformatics 27(15):2156-2158.

Elena SF, De Visser JAG (2003). Environmental stress and the effects of mutation. Journal of biology 2:1-4.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS One 6(5):19379.

Girma G, Hyma KE, Asiedu R, Mitchell SE, Gedil M, 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, Nida H, Tirfessa A, Lule D, Bejiga T, Seyoum A, Mengiste T (2020). A comprehensive phenotypic and genomic characterization of Ethiopian sorghum germplasm defines core collection and reveals rich genetic potential in adaptive traits. The Plant Genome 13(3):20055.

Guo C, McDowell IC, Nodzenski M, Scholtens D M, Allen AS, Lowe WL, Reddy TE (2017). Transversions have larger regulatory effects than transitions. Bmc Genomics 18:1-9.

Guo Q (2014). Plant hybridization: the role of human disturbance and biological invasion. Diversity and Distributions 20(11):1345-1354.

Hahn SK, Osiru D, Otoo J, Akoroda M (1995). Production of yams: present role and future prospect.

Ikiriza H, Ogwang PE., Peter EL, Hedmon O, Tolo CU, Abubaker M, Abdalla AAM (2019). *Dioscorea bulbifera*, a highly threatened African medicinal plant, a review. Cogent Biology 5(1):1631561.

IPGRI/IITA (1997) Descriptors for Yam (*Dioscorea* spp.). International Institute of Tropical Agriculture, Ibadan, Nigeria/International plant genetic resources institute, Rome, Italy.

Ji M, Deng J, Yao B, Chen R, Fan Z, Guan J, Niklas KJ (2017). Ecogeographical variation of 12 morphological traits within Pinus tabulaeformis: the effects of environmental factors and demographic histories. Journal of Plant Ecology 10(2):386-396.

Jyothy A, Sheela MN, Radhika NK, Anwar I, Kumar V, Abhilash PV (2017). Morphological characterization of greater yam (*Dioscorea alata* L.) landraces in Kerala. Journal of Root Crops 43(1):3-10.

Kabi F, Masembe C, Muwanika V, Kirunda H, Negrini R (2014). Geographic distribution of non-clinical Theileria parva infection among indigenous cattle populations in contrasting agro-ecological zones of Uganda: implications for control strategies. Parasites and Vectors *7*(1):1-9.

Lebot V (2009). Tropical Root and Tuber Crops*:* Cassava, Sweet Potato*,* Yams and Aroids. https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Lebot+V+%282009%29.+Tropical+root+and+tuber+crops&btnG=.

Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, Casler MD, Costich DE (2013). Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. PLoS Genetics 9(1):1003215.

Luo Z, Iaffaldano BJ, Zhuang X, Fresnedo-Ramírez J, Cornish K (2017). Analysis of the first Taraxacum kok-saghyz transcriptome reveals potential rubber yield related SNPs. Scientific Reports 7(1):1-13.

Mignouna HD, Abang MM, Asiedu R (2003). Harnessing modern biotechnology for tropical tuber crop improvement: Yam (*Dioscorea* spp.) molecular breeding. African Journal of Biotechnology 2(12):478-485.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution 37(5):1530-1534.

Ministry of Agriculture, Animal Industry and Fisheries, and Agriculture (MAAIF) (2014). Agriculture Sector Development Strategy and Investment Plan: (March 2010).

Mbanjo EGN, Jones H, Caguiat XGI, Carandang S, Ignacio JC, Ferrer MC, Kretzschmar T (2019). Exploring the genetic diversity within traditional Philippine pigmented Rice 12(1):1-18.

Narula A, Kumar S, Srivastava PS (2007). Genetic fidelity of in vitro regenerants, encapsulation of shoot tips and high diosgenin content in *Dioscorea bulbifera* L., a potential alternative source of diosgenin. Biotechnology Letters 29:623-629.

Ortiz EM (2019). GitHub Vcf2phylip v2.0: Convert a VCF Matrix into Several Matrix Formats for Phylogenetic Analysis. (Zenodo).

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

https://doi.org/10.1093/bioinformatics/bts460
Priya RS, Kumuthakalavalli R, Karuppusamy S (2002). Some Traditional Folk Formularies against Fertility in Kerala. Journal of Human Ecology 13(4):335-336.
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Sham PC (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics 81(3):559-575.
Rabbi IY, Kulakow PA, Manu-Aduening JA, Dankyi AA, Asibuo JY, Parkes EY, Maredia MK (2015). Tracking crop varieties using genotyping-by-sequencing markers: a case study using cassava (Manihot esculenta Crantz). BMC Genetics 16(1):1-11.
Rao GK, Kapadia C, Patel NB, Desai KD, Murthy PN (2020). Genetic diversity analysis of greater yam (*Dioscorea alata* L.) genotypes through RAPD and ISSR markers. Biocatalysis and Agricultural Biotechnology 23:101495.
Saski CA, Ranjana B, Brian ES, Robert A (2015). "Genomic Resources for Water Yam (*Dioscorea Alata* L.): Analyses of Estsequences, de Novo Sequencing and GBS Libraries 10(7):1-14.
Siqueira MV, Bonatelli ML, Günther T, Gawenda I, Schmid KJ, Pavinato VA, Veasey EA (2014). Water yam (*Dioscorea alata* L.) diversity pattern in Brazil: an analysis with SSR and morphological markers. Genetic Resources and Crop Evolution 61:611-624.
Shajeela PS, Mohan L, Louis J, Tresina S (2011). "Nutritional and Antinutritional Evaluation of Wild Yam (*Dioscorea* Spp.) [Evaluaciòn Del Valor Nutricional y Factores Antinutricionales de *Dioscorea* Spp. Silvestre]." Tropical and Subtropical Agroecosystems 14:723-30. http://www.scielo.org.mx/pdf/tsa/v14n2/v14n2a30.pdf.
Sharma AK, Code QR (2017) Phytochemical screening (qualitative) of certain medicinal plants of Saharanpur city. International Journal of Pharmaceutical Chemistry 7(4):58-62.
Tetchi FA, Rolland-Sabaté A, Amani GNG, Colonna P (2007). Molecular and physicochemical characterisation of starches from yam, cocoyam, cassava, sweet potato and ginger produced in the Ivory Coast. Journal of the Science of Food and Agriculture 87(10):1906-1916.
Trimanto T, Hapsari L (2015). Diversity and utilization of *Dioscorea* spp. tuber as alternative food source in Nganjuk Regency, East Java. *AGRIVITA,* Journal of Agricultural Science 37(2):97-107.
Velasco-Ramírez AP, Torres-Morán MI, Molina-Moret S, Sánchez-González JDJ, Santacruz-Ruvalcaba F (2014). Efficiency of RAPD, ISSR, AFLP and ISTR markers for the detection of polymorphisms and genetic relationships in camote de cerro (*Dioscorea* spp.). Electronic Journal of Biotechnology 17(2):65-71.
Xiao Y, Liu H, Wu L, Warburton M, Yan J (2017). Genome-wide association studies in maize: praise and stargaze. Molecular Plant 10(3):359-374.
Yang H, Wei CL, Liu HW, Wu JL, Li ZG, Zhang L, Wan XC (2016). Genetic divergence between Camellia sinensis and its wild relatives revealed via genome-wide SNPs from RAD sequencing. PLoS One 11(3):151424.
Yu G (2020). Using ggtree to visualize data on tree-like structures. Current Protocols in Bioinformatics *69*(1), e96.
Zamir D (2013). Where have all the crop phenotypes gone? PLoS Biology 11(6):1001595.