

ORIGINAL ARTICLE

Plant Genetic Resources

Genetic diversity and population structure among Ugandan shea tree (*Vitellaria paradoxa* subsp. *nilotica*) accessions based on DarTSeq markers

Juventine Boaz Odoi^{1,2}  | Emmanuel Amponsah Adjei^{2,3,4}  | Prasad Hendre⁵ |
Judith Ssali Nantongo¹ | Alfred Adibo Ozimati^{3,6}  | Arfang Badji^{2,3} |
Grace Nakabonge⁷ | Richard Edema^{2,3} | Samson Gwali¹ | Thomas Lapaka Odong²

¹Tree Improvement and Germplasm Research Programme, National Forestry Resources Research Institute, National Agricultural Research Organization, Kampala, Uganda

²Department of Agricultural Production, College of Agricultural and Environmental Sciences, Makerere University, Kampala, Uganda

³Makerere University Regional Centre for Crop Improvement, Makerere University, Kampala, Uganda

⁴Council for Scientific and Industrial Research, Savanna Agricultural Research Institute, Tamale, Ghana

⁵Center for International Forestry Research-World Agroforestry (CIFOR-ICRAF), United Nations Avenue, Nairobi, Kenya

⁶College of Natural Science, School of Bioscience, Department of Plant Science, Microbiology and Biotechnology, Makerere University, Kampala, Uganda

⁷School of Forestry, Environmental and Geographical Sciences, Makerere University, Kampala, Uganda

Correspondence

Juventine Boaz Odoi, Tree Improvement and Germplasm Research Programme, National Forestry Resources Research Institute, National Agricultural Research Organization, P.O. Box 7084, Kampala, Uganda.

Email: juventineboaz@gmail.com

Assigned to Associate Editor Tracie Matsumoto.

Funding information

The World Agroforestry provided funding under the Genebank Platform, for plant material sample collection from the field, Grant/Award Number: GCDT-1213; Makerere Regional Centre for Crop Improvement (MaRCCI) provided the student research grant for genetic

Abstract

Molecular markers such as single nucleotide polymorphisms (SNPs) and SilicoDarT are important in dissecting the genetic diversity of a population at DNA level. The two marker types were analyzed using the same genotyping platform. Although the two marker types were analyzed using the same genotyping platform, they were filtered using a different marker stringency. We determined the genetic diversity of 623 half-sibs of shea tree (*Vitellaria paradoxa* a C. F. Gaertn. *nilotica*) assembled from five geographical locations (Katakwi, Otuke, Amuru, Moyo, and Arua) in Uganda's shea parklands. A total of 27,063 Diversity Arrays Technology (DArT) SNP and 9018 SilicoDArT markers were used to assess genetic diversity and population structure. The populations showed a low genetic diversity ($G_{ST} = 0.21$), very low population differentiation ($F_{ST} = 0.02$), low-to-moderate linkage disequilibrium ($r = 0.2$), and Hardy–Weinberg Equilibrium ($HWE = 0.1982$). The study further revealed a higher genetic diversity within population (0.26) than among the population (0.21). A high level of heterozygosity was observed within individuals (0.26) and markers (0.32) revealing a high non-random association of alleles at different loci that offer

Abbreviations: AMOVA, analysis of molecular variance; BeCA-ILR, Biosciences eastern and central Africa- international Livestock Research Institute; CTAB, cetyltrimethylammonium bromide; DArT, Diversity arrays technology; DArT-seq, Diversity arrays technology sequencing; FST, genetic differentiation; He, expected heterozygosity; Ho, observed heterozygosity; MAF, minor allele frequency; PCoA, Principal coordinate analysis; PIC, polymorphic information content; RAPDs, Random amplified polymorphic DNA; SNPs, Single nucleotide polymorphisms; SSRs, Single sequence repeats.

fingerprinting and analysis, Grant/Award
Number: ACE 2 Project

opportunities to design association studies and allele transfer in marker-assisted selection in the population. The markers varied in their polymorphic information content values (SilicoDart = 0.11) and (SNPs = 0.15) on genetic diversity. This study reveals the importance of genetic diversity and population structure analysis using high-density Dart-Seq SilicoDart and SNP makers in the conservation and breeding of shea tree in Uganda.

1 | INTRODUCTION

Shea tree (*Vitellaria paradoxa* a C. F. Gaertn. *nilotica*) is a species with commercial and social importance in Sudano-Sahelian Africa (Hale et al., 2021). The species was also registered for its socioeconomic importance, where a number of products are used by the communities for rituals, treatments, and household income (Christophe et al., 2014). The tree is considered second to oil palm in supplementing plant oil necessities of the communities in Africa where oil palm is not grown (Hall, 2012). According to Odoi et al. (2020), the annual production throughout Africa stands at 750,000 tons of which between 55% and 75% is consumed locally, while the rest is exported (Lovett & Haq, 2000). From the shea export market, >90% of the products developed are for confectionary purposes, while about 10% of the shea oil goes for cosmetics industry, the exports translate to US \$200 million per annum across Africa (Maranz & Wiesman, 2003). With the increasing shea tree populations due to farmer-managed natural regeneration and available governing laws and regulations within Uganda's shea parklands (Odoi et al., 2020), trade in shea products is likely to increase. There is a need to further improve farmer-protected trees to guarantee increased productivity and quality product development for competitive markets.

To design effective shea tree breeding and conservation strategies, Gwali et al. (2015) pointed out the need to generate information on available shea tree genetic diversity among the ethnovarieties in Uganda. Information on available shea populations sheds light on the mating system, pollen dispersion syndrome, seed propagation, and effective population size, which are closely related to their geographic distribution and the nature of the surrounding species (Hamrick et al., 1979).

A number of molecular markers and quantitative methods such as amplified fragment length polymorphisms, random amplified polymorphic DNA (RAPDs), restriction fragment length polymorphisms, single sequence repeats (SSRs), inter simple sequence repeats, and single nucleotide polymorphisms (SNPs) have been widely used for the study of genetic diversity and population structure in tree species (Dadjo et al., 2020; Mohammad et al., 2018; Nantongo et al., 2010). These growing advances provide enormous opportunities to decipher the genetic diversity of genetically structured natural populations (Jombart & Ahmed, 2011). SNP markers, for

example, owing to their robustness and abundance in plant genomes, are being increasingly used for diversity analyses. The marker has a high level of sensitivity and accuracy and can be used for genotyping large populations compared to other markers.

Diversity arrays technology (Dart)-SNPs have been widely used for tree characterization due to limited resources such as SNP-chips that can be used for economic scaling up (Dadjo et al., 2020). In recent years, high throughput genotyping such as Dart is gaining the upper hand while choosing for genomic studies and genetic diversity analysis given its speed and low cost (Sodedji et al., 2021). In fact, diversity arrays technology sequencing (Dart-seq) is one of the economical and fast ways to generate these markers. Dart sequencing method allows the discovery of sufficient amounts of SNPs in many plant species and is now widely being deployed to analyze the genetic diversity and genetic structure of many crops around the world (Grzebelus et al., 2014; Nantongo et al., 2022). However, Gudeta (2018) recommends deploying a combination of marker types for the accurate assessment of intra- and interpopulation genetic diversity of natural populations of forest tree species. In the same vein, Fontaine et al. (2004) quantified the genetic variation of *V. paradoxa* using RAPDs and chloroplast microsatellites markers and were able to identify two major populations across the shea belt in Africa owing to the shea tree selection and dispersal during traditional agroforestry.

This study used SilicoDart and Dart-seq SNP markers to determine the genetic diversity and population structure among 623 selected half-sibs shea tree (*V. paradoxa* C.F. Gaertn. subsp *nilotica*) accessions in Uganda to plan for their conservation and breeding. Therefore, this study aimed at determining the level of genetic diversity and population structure in the selected elite shea tree accessions for the development of breeding seedling orchard as a source of a future true to type germplasm in Uganda.

2 | MATERIALS AND METHODS

2.1 | Genetic materials

Early studies identified *V. paradoxa* (then *Butyrospermum parkii*) as highly outcrossing with ploidy level of $2n = 24$

(Hale et al., 2021). A total of 1204 half-sibs were raised from the 180 pre-selected set of parent trees that were identified by their agronomic characteristics. The seeds were randomly collected from the parklands from the five districts of Katakwi, Otuke, Amuru, Moyo, and Arua (Figure 1) across the shea parklands in Uganda. The seedlings were then raised and managed in the tree nursery at Ngetta Zonal Agricultural Research and Development Institute between June 2018 and October 2019. The seedlings were watered every morning and evening with at least 10 L/m² of nursery bed for 18 months.

After the removal of duplicates based on the level of informativeness in the duplicated accessions, 708 lines remained, and on further filtration, the final number of lines used in this study was further reduced to 623 taxa. The shea seedlings were then planted in two 8-acre multilocal trials as breeding seed orchard in Lira and Serere districts (see Table 1).

2.2 | DNA extraction and DArT-sequencing

Young healthy leaf samples were randomly picked from 752 shea seedlings for DNA extraction and genotyping. Only 752 accessions that showed best growth characteristics were selected from the 1204 plantlets due to resource constraints. The genomic DNA from silica dried leaf samples was extracted at Biosciences Eastern and Central Africa-International Livestock Research Institute (BeCA-ILRI) following the cetyltrimethylammonium bromide (CTAB)/chloroform/isoamyl alcohol method (Doyle & Doyle, 1987). The DNA was quality checked using standard processes: 0.8% agarose gel electrophoresis, optical measurements for 260 and 280 nm using a NanoDrop 2000 spectrophotometer (ND-2000 V3.5; NanoDrop Technologies, Inc.), and Qubit 3.0 Fluorometer (Thermo Fisher Scientific) for quantification. The DNA from the 752 individuals was partial-genome sequenced using proprietary restriction enzyme tagged genotyping-by-sequencing approach using DArTseq methodology (DArT P/L) on a HiSeq2500 next-generation sequencer (Illumina Inc.) with 72 bases read length. In brief, the libraries were prepared using the PstI-SphI complexity reduction method (Sansaloni et al., 2011).

Sequences generated from each lane were processed using proprietary DArT analytical pipelines. In the primary pipeline, the FASTQ files were first processed to filter poor quality sequences, applying more stringent selection criteria to the barcode region compared to the rest of the sequence. Thus, the assignments of the sequences to specific samples carried out in the “barcode split” step were more consistent. Approximately 2,500,000 ($\pm 7\%$) sequences per barcode/sample were used in marker calling. Finally, identical sequences were collapsed into “fastqcall files.” These files were used in the secondary pipeline for DArT P/L’s proprietary SNP and SilicoDArT (Presence/Absence Markers in genomic representations) (present = 1 vs. absent = 0)

Core Ideas

- Shea trees in Uganda possess a moderate level of genetic diversity, essential for the species’ long-term survival.
- The low fixation index is a precursor of the good performance of the populations across Uganda’s agro-ecologies.
- Shea trees in Uganda are grouped into three populations.
- Deployment of single nucleotide polymorphisms revealed molecular markers and genetic diversity for putative selective footprints in shea tree.

calling algorithms (DArTsoft14). The analytical pipeline processed the sequence data. The reads were then aligned to the shea_V1 reference genome (<https://www.frontiersin.org/articles/10.3389/fpls.2021.720670/full>). Essentially, DArT-Seq technology relies on a complexity reduction method using restriction enzymes that are sensitive to DNA methylated sites and repetitive DNA (Raman et al., 2014). DNA samples were processed in digestion/ligation reactions as described by Kilian et al. (2016).

2.3 | Marker data quality

The marker data were then assessed for average SNP count, read depth, call rate, and reproducibility. Final filtering of the SNP dataset using minor allele frequency (MAF >0.05), polymorphism information content (>1%), the proportion of samples for which the genotype score is “1” (one ratio), the proportion of samples for which the genotype call is either “1” or “0,” rather than “–” (call rate >80%), and the proportion of technical replicate assay pairs for which the marker score is consistent (reproducibility > 95%) was done for SNPS and was performed with DArT R package (Gruber et al., 2018). Missing data of 10% were sought and imputed using the snpReady software (Granato & Fritsche-Neto, 2018).

Analysis was done separately over SNPs and SilicoDArT using similar filtration parameters as indicated in Table 2.

2.4 | SNP filtering and imputation of missing values

Data filtering and quality checks were performed in TASSEL (v5.2.52) (Bradbury et al., 2007). This enabled removal of SNP markers with >20% of missing data, < 0.05 MAF, and unknown positions on the genome. Imputation of SNP data was then performed using the k-nearest neighbor method, following Bradbury et al. (2007). A total of 7530 informative

TABLE 1 Shea tree ethnovarieties with their local and conventional descriptors.

No.	Characteristics	Geographical location				
		Katakwi	Otuke	Amuru	Moyo	Arua
1	Sweetness	Lo'jijit (sweet pulp) (3)	Abono (sweet pulp) (3)	Malim (sweet pulp) (3)	Limi (sweet pulp) (3)	Awa odu mbili mbili (sweet pulp) (3)
2	Unpleasant taste	Epiana (tasteless pulp) (3). Egeget (astringent taste) (3)	Yao Permit (tasteless pulp) (3). Yao cot (astringent taste) (3)	Yao mabot (tasteless pulp) (3). Macot (astringent taste) (3)	Asa (tasteless pulp) (3). Etri (astringent taste) (3)	Itritri (astringent taste) (3). Kpala kpala (tasteless pulped) (3)
3	Skin texture	Lo'ecekit (hairy) (3)	Jayer (hairy) (3)	Ayer (hairy) (3)	Enyi (hairy) (3) Anyua (hard fruit skin) (3)	
4	Pulp texture	Lo'karic (hard pulp) (3) Lo'nonok (soft pulp) (3)	Acogo (hard pulp) (3) Apocopoco (soft pulp) (3)	Acogo/Okwe (hard pulp) (3) Labokoboko (soft pulp) (3)	Urujua (hard pulp) (3) Muda muda (soft pulp) (3)	Kongolo (hard pulp) (3) Mele mele (soft pulp) (3)
5	Size	Bombom (big fruits) (3)	Yao Adongo (big shea fruits) (3). Yao atino (tiny seed) (3). Olindir (small fruits) (3)	Madongo (big fruits) (3). Lagili (small fruits) (3) Lalet (thin pulped fruits) (3)	Muda muda (soft pulp) (3)	Awa Iki nyiri (small fruits/seeds) (3). Omemea (tiny seed) (3). Awa odu perepere (thin pulped shea) (3)
6	Shape	Lo'somel (oval fruits) (3) Lo'lungur (round fruits) (3)	Acula (oval) (3) Aboco (elliptical) (3). Alulung (round fruits) (3)	Acula (oval) (3) Maboco (elliptical) (3)	Colobaa (elliptical) (3)	Gburu (round fruits) (3). Julu (oval fruits) (3)
7	Coloration	Narengun (red seeds) (3)	Arema (red seeds) (3). Molilo (black/dark seeded) (3)	Arema (red seeds) (3). Macol (black/dark seeded) (3)	Onia (black/dark seeded) (3)	Awa iki ika (red seeds) (3)
8	Tree trait		Yawo wigweng (shea tree growing on rocky places) (3)			

Note: The figures in parentheses show the number of families selected per ethnovariety.

TABLE 2 Marker quality parameters of shea tree populations in Uganda.

Marker quality parameters	SilicoDaRT	SNPs
PIC	0.15	0.26
One ratio	0.15	0.29
Call rate	0.98	0.78
Reproducibility	0.99	0.98

Abbreviations: PIC, polymorphic information content; SNPs, single nucleotide polymorphisms.

SNP markers were then obtained for further analysis in this study.

The observed heterozygosity (Ho) and expected heterozygosity (He) were performed using the “adegenet” package in R (R Core Team, 2020), and the call rate and reproducibility were determined using “dartR” in R environment (Jombart & Ahmed, 2011).

2.5 | Determination of genetic variation among populations

The markers summary statistics that included MAF, He, Ho, polymorphic information content (PIC), and gene diversity were generated in PowerMarker v3.25 (Lui & Muse, 2005). The populations used for the initial fixation index analysis comprised the selected genotypes from the five study districts described above. The explained genetic variation among populations was determined from the SNP markers using analysis of molecular variance (AMOVA), and genetic differentiation (FST) was performed in GenAlEx 6.5 (Peakall & Smouse, 2012). The AMOVA was generated after numerically ($A = 1$, $C = 2$, $G = 3$, $T = 4$, and missing data = 0) coding the markers as in Blyton and Flanagan (2006).

2.6 | Population structure

A hierarchical clustering analysis using the SNP markers was performed and a dendrogram generated as in

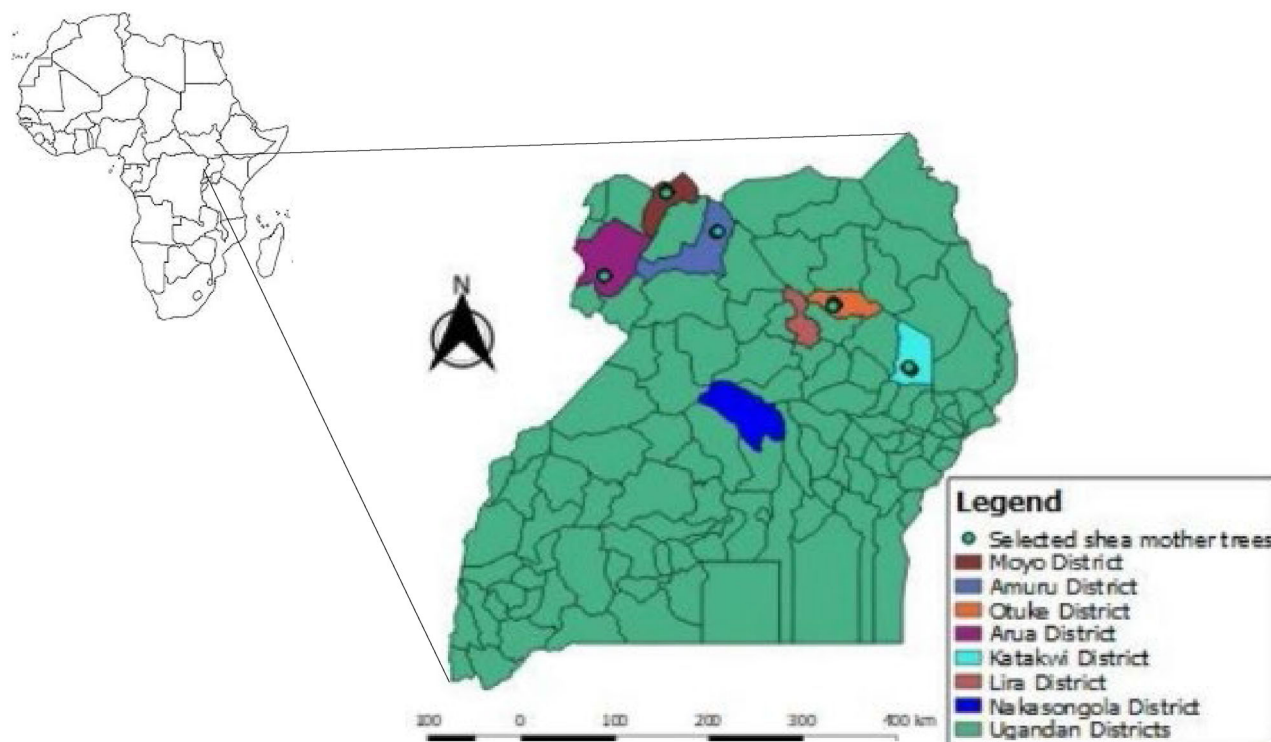


FIGURE 1 Map of Uganda indicating the study sites.

Dadjo et al. (2020) and Odong et al. (2011). Using filtered markers, the population structure was performed with Bayesian clustering approach in STRUCTURE V2.3.4 (Bennett et al., 2008). Burn-in period of 10,000 Markov Chain Monte Carlo iterations, 100,000-run length were performed, and the best K value with the highest likelihood was employed to estimate the most suitable population size following Evanno et al. (2005). Ten runs per K value (number of clusters) from 1 to 10 were considered in Structure Harvester (Earl & vonHoldt, 2012). The shea accessions with inferred ancestry ≥ 0.80 were assigned to their respective groups. The hierarchical clusters and dendrogram were performed on the 623 shea lines using Analyses of Phylogenetics and Evolution (ape) package implemented in R (R Core Team, 2020). The principal coordinate analysis (PCoA) was generated from GenAlEx 6.5 (Peakall & Smouse, 2012) and plotted in R (R Core Team, 2020).

Phylogenetic and molecular evolutionary analyses were conducted using MEGA version X (Kumar et al., 2018). Neighbor Joining was performed to identify the subpopulations with their genetic diversity using PowerMarker V3.25 and GenAlEx 6.5 (Peakall & Smouse, 2012).

3 | RESULTS

3.1 | Marker identification, alignment, and quality control

The total number of SNP markers with MAF of <0.1 was 3895 (51.72%) (Figure 1). The informative SNPs with PIC values ≥ 0.4 represent 10.08% ($n = 759$) of the markers used (Figure 1). A total of 7530 putative co-dominant SNP markers were used in this study, and 3991 (53%) were found in the coding region, of these 3518 (24.5%) were non-synonymous SNPs. The co-dominant markers revealed a high reproducibility (0.98), mean call-rate value of 0.78, and MAF of 0.13, with the majority (54.23%) having PIC value above 0.1 and mean value of 0.26. The minor and major allele frequencies of the SNP markers ranged from 0.001 to 0.001 and 0.9986 to 0.779, respectively.

A total of 30,733 SNP markers data were initially generated, and 11.9% (4670 SNPs) markers were removed during filtering due to the lack of informativeness and duplicates; thus, 27,063 SNP (88.1%) markers remained. These were further filtered to 7530 SNPs which matched the quality criteria and were used in this study.

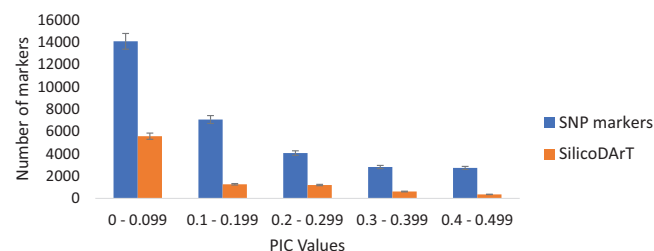


FIGURE 2 Distribution of polymorphism information content (PIC) in 27,063 single nucleotide polymorphisms (SNPs) and 9019 SilicoDArT SNP markers derived from 708 shea tree accessions. X-axis indicates PIC value classes, whereas Y-axis indicates the number of the diversity arrays technology sequencing (DArTseq) markers in each PIC value class.

3.2 | Characterization of the SNP markers

Six SNP types were determined in this study (Table 1). Out of the 7530 SNPs retained after removal of duplicates and further filtration, 4305 (57.17%) were transitions ($C \leftrightarrow T$ or $A \leftrightarrow G$), and 3225 (42.83%) were transversions ($C \leftrightarrow A$ or $T \leftrightarrow G$). The resultant transition/transversion ratio was 1.34. Accordingly, out of the 4305 (57.2%) genic SNPs located in the coding regions, 2282 (53%) were non-synonymous substitutions and 2023 (47%) were synonymous, and the ratio of non-synonymous to synonymous substitutions (K_a/K_s) was 1.13.

3.3 | Comparison between SilicoDArT and SNP markers

A total of 7530 quality SNPs markers and 9018 SilicoDArT were genotyped from 708 shea tree accessions. The proportion of technical replicate assay pairs for which the marker score is consistent (reproducibility) differed between SilicoDArT (99%) and SNP markers (98%). The PIC range for SilicoDArT and SNP markers types was between 0.006 and 0.5 and 0.003 and 0.5, respectively (Figure 2).

The average marker call rates for SilicoDArT and SNPs after filtration were 98% and 78%, respectively, while their reproducibility was 0.99% and 0.98%, respectively, indicating a consistent marker score (Table 3).

3.4 | Genetic diversity

The results revealed a mean H_o of 0.26, the mean H_e of 0.21, and moderate HWE of 0.198. The results indicate a higher H_o (0.26), H_e (0.21), and H_t (0.32) in SNP marker type than the SilicoDArT (0.15, 0.1, and 0.1, respectively) and a negative inbreeding coefficient ($SNP F_{IS} = -0.46$ and SilicoDArT

TABLE 3 Proportion of single nucleotide polymorphism (SNP) types.

SNP	Number (n)	Proportion (%)	Type
A/C	783	10.4	Transversion
A/G	2160	28.68	Transition
A/T	847	11.11	Transversion
C/G	836	11.10	Transversion
C/T	2145	28.49	Transition
G/T	762	10.22	Transversion
Total	7530	100	

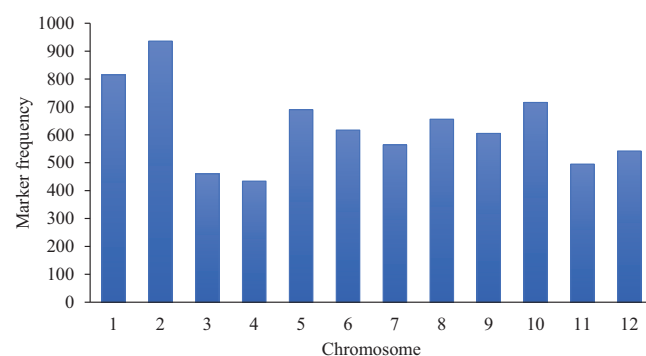


FIGURE 3 Chromosome single nucleotide polymorphisms (SNP) marker diversity.

$F_{IS} = -0.49$) (Table 4). In general, the results revealed low genetic differentiation among locations but a significant level of genetic differentiation within populations (Table 5).

3.5 | SNP marker gene diversity

The SNP marker frequencies of the MAF ranged from 0.03 to 0.5 with the average frequency of 0.13 (Table 6 and Figure 3).

The PIC and gene diversity also ranged from a minimum of 0.05 and a maximum of 0.63 and 0.68, respectively (Table 6).

Chromosome 2 had the highest number of markers (990), followed by chromosome 1 (808), chromosome 10 (790), chromosome 5 (650), and chromosome 8 (645).

Meanwhile, chromosomes 4 and 3 had the lowest (425 and 432, respectively) marker frequencies (Figure 4). This indicates a non-random distribution of SNPs with varying SNP frequencies in the 12 chromosomes of shea tree in Uganda.

3.6 | Population structure

The shea tree populations in Uganda are genetically grouped into two clusters of Eastern group (B) and West Nile/Northern Uganda group (A) (Figures 5 and 6).

TABLE 4 Gene diversity and population structure of shea tree populations in Uganda.

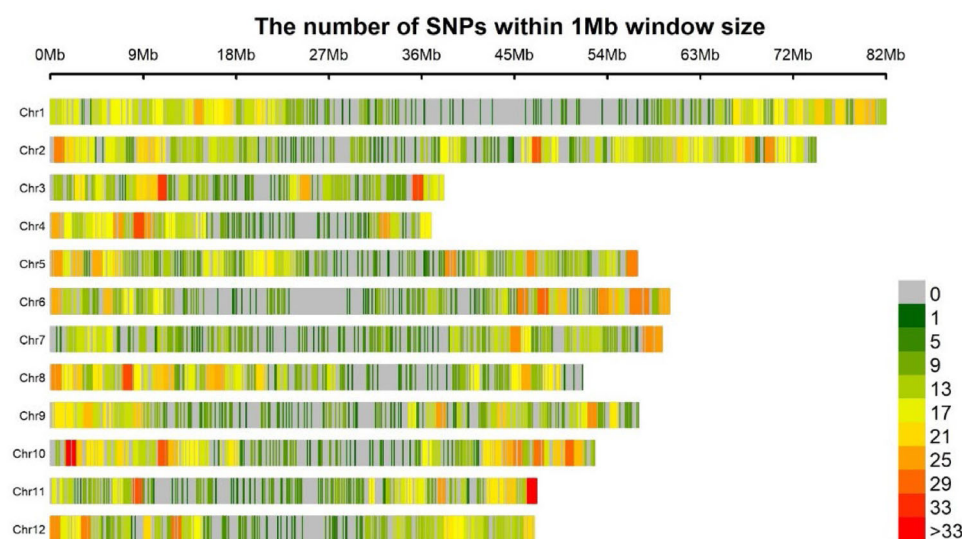
Marker diversity parameters		SilicoDaRT (<i>n</i> = 2018)	DArT-SNP (<i>n</i> = 7530)
Ho	Observed heterozygosity	0.15	0.26
He	Expected heterozygosity	0.1	0.21
Ht	Overall gene diversity within population	0.1	0.32
Dst	Gene diversity between population	0	0
Htp	Corrected gene diversity within population	0.1	0.2
Dstp	Corrected gene diversity between population	0	0
Fst	Population differentiation	0.02	0.01
Fstp	Corrected population differentiation	0.02	0.02
Fis	Inbreeding coefficient	−0.49	−0.46

Abbreviations: DArT-SNP, Diversity Arrays Technology-single nucleotide polymorphisms; He, expected heterozygosity; Ho, observed heterozygosity.

TABLE 5 Analysis of molecular variance (AMOVA) among the 623 shea lines assessed with 7530 single nucleotide polymorphism (SNP) markers.

Source	df	SS	MS	Est. Var.	Percentage	PhiPT	<i>p</i> Value
Among pops	4	6931.08	1732.77	4.35	0		
Within pops	618	740,136.16	1197.63	1197.63	100	0.004	0.000
Total	622	747,067.24		1201.98	100		

Abbreviations: MS, mean square; SS, sums of squares.

**FIGURE 4** The number and size of single nucleotide polymorphisms (SNPs) within 1 Mb window size of *V. paradoxa* subsp. *nilotica* genome.

Generally, lines arranged themselves according to their areas of origin (population) apart from few that were grouped within a different population forming admixtures. The population at $K = 2$, representing two clusters (Figure 6), was chosen in consonant with the PCoA grouping in Figure 5 (see also Table 7).

The different populations were assessed for their proportion contribution of individuals to the overall number of studied

accessions. The study reveals that cluster 2 contributed to the highest (57%) proportion of individuals with cluster 1 (43%) having the lists (Table 8).

The two clusters had similar (0.032) genetic variability (net nucleotide distance). Higher normalized variance was observed in allele frequencies between populations, F_{ST} (0.14) in cluster 1 than cluster 2 (0.06) (Table 8).

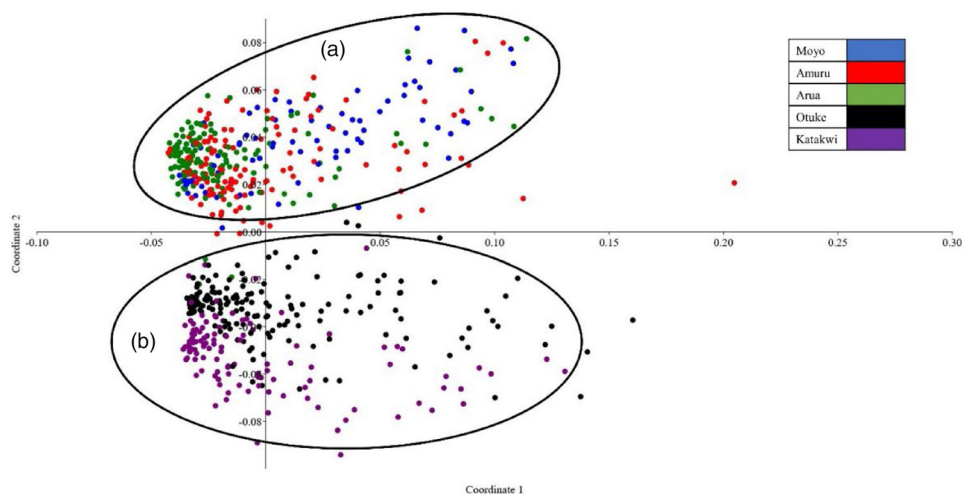


FIGURE 5 Scatter plot of the accessions along the first two principal coordinates PCoA1 and PCoA2. The colors of the dots along the principal coordinate analysis (PCoA) coordinates 1 and 2 represent the accessions from different locations. A and B indicate the subpopulations identified.

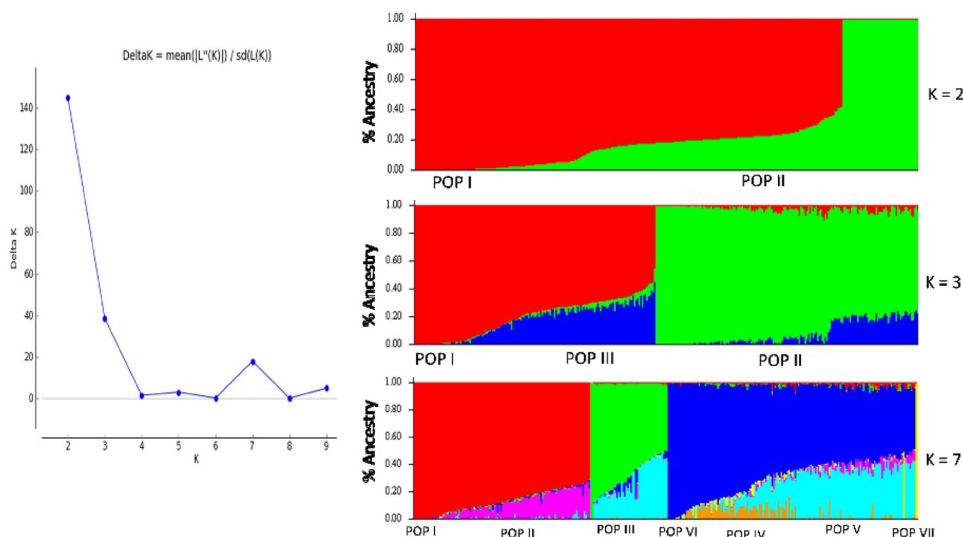


FIGURE 6 Likelihood of ΔK calculated by Evanno et al.'s (2005) method, revealing maximum value at $K = 2$ and further subpopulations at $K = 3$, conducted using the Bayesian clustering approach in structure V2.3.4. The bar plots of line assigned to $K = 2$ populations (pop I and pop II) are represented by red and green colors as identified in the structure and structure harvester.

TABLE 6 Single nucleotide polymorphism (SNP) marker allele frequency.

Variables	Mean	Minimum	Maximum
MnAF	0.13	0.03	0.5
MAF	0.78	0.57	0.97
H_e	0.21	0.01	0.57
H_o	0.26	0.01	0.96
PIC	0.26	0.05	0.63
Gene diversity	0.32	0.05	0.68

Abbreviations: MAF, minor allele frequency; PIC, polymorphic information content.

TABLE 7 Number of accessions assigned to the different clusters and admixture for the inferred K . The highest most appropriate inferred K was observed at $K = 3$.

Population	Metapopulation			
	Cluster 1	Cluster 2	Cluster 3	Admixture
Arua	0	0	143	3
Amuru	109	0	13	13
Moyo	76	0	1	1
Katakwi	2	116	0	2
Otuke	0	160	0	0
Total	187	276	157	19

TABLE 8 Genetic variability among the inferred K clusters (net nucleotide distance) computed using point estimates of p and within (expected heterozygosity, H_e) populations, proportion of membership, and mean value of the fixation index (F_{ST}) observed from the population structure of 623 shea tree lines.

Population	Net nucleotide distance		H_e	Mean F_{ST}	Proportion of membership (%)
	Cluster 1	Cluster 2			
Cluster 1	–	0.032	0.29	0.14	0.43
Cluster 2	0.032	–	0.31	0.06	0.57

Abbreviations: H_e , expected heterozygosity; F_{ST} , fixation index.

There was a high level of heterozygosity within individuals (0.31) and markers (0.32).

4 | DISCUSSION

4.1 | Marker diversity and DNA substitution mutations

Determination of the genetic diversity and population structure in *V. paradoxa* presents an important opportunity in shea research and development (breeding, conservation, management, processing, and business) in Uganda. This research further presents the first-ever molecular study of the genetic diversity and population structure of *V. paradoxa* subspecies *nilotica* using a high throughput genome-wide SNP dataset.

A total of 7530 informative SNPs markers were used in this study to determine the genetic diversity and population structure of shea tree accessions from five populations in Uganda. Two types of DNA substitution mutations (transitions and transversions) were observed in this study. The results indicated higher transition mutation than the transversion type with mutation rates in the likely order of 10–8 to 10–9 (Shepherd et al., 2005). The frequent occurrence of the transition type could be due to the higher rates of 5-methylcytosine deamination from the CpG dinucleotides (Yang et al., 2016). Such a higher transition mutation within the coding regions is due to the degeneration and selective nature of genetic code that results in gene conservation.

This low fixation index in a population has been associated with high gene flow and genetic drift so as to minimize allele loss (Barchi et al., 2011). SNPs may change the encoded amino acids (nonsynonymous SNP) and therefore change the amino acid sequence or be silent (synonymous SNP), thereby maintaining the amino acid sequence or simply occurring in the noncoding regions. They may also influence promoter activity (gene expression), messenger RNA (mRNA) conformation (stability), and subcellular localization of mRNAs and/or proteins (Yang et al., 2016). The non-synonymous SNPs are of interest to this study because they generate new variants that are important for varietal development during breeding programs. Such mutations, when selected by the environment, are important for conservation agenda of the

species since the variants get more adapted to the environmental changes (Barchi et al., 2011). In fact, C/T transition is the most common type of SNP found in higher plants, which was also the case in this study (57.1%). The shea tree lines used in this study revealed a high SNP marker variation with A↔G and C↔T than the other SNP types (C↔A and T↔G). This finding aligns with other studies on *Garcinia kola* (Dadjo et al., 2020) and cowpea (Huynh et al., 2013; Xiong et al., 2016). However, such markers are important in evaluating the amount, extent, and distribution of genetic variation within and among populations (Gudeta, 2018). Such variations may be due to high frequencies of shifts from A to G and C to T mutations following methylation (Ketema et al., 2020). These sites of cytosine methylation offer hot spots for C to T mutations important in studying the process of transcript elongation resulting in spontaneous mutations (Beletskii & Bhagwat, 1996).

It indicates the level of diversity of a gene or DNA segment in a population useful in revealing any evolutionary pressure that could have occurred on the alleles or mutation in a locus. Therefore, the identification of SNP marker diversity and variations in shea tree may lead to a better understanding of their function in mutation causing variation within shea tree populations and ethnovarieties across its rangeland (Dadjo et al., 2020). Similar findings were also reported in SNP-based genetic diversity studies in crops (Shimizu et al., 2016).

4.2 | Genetic diversity and population structure

Genetic diversity and population structure of a species is a good precursor of its management and conservation status. The results revealed low genetic differentiation, suggesting the existence of anthropogenic gene flow and admixture in the area. The study further found moderate (0.21) genetic diversity in *V. paradoxa*, similar to what was reported in West Africa (Hunter et al., 2012) and Uganda using SSR microsatellite markers (Gwali et al., 2015).

The reduced genetic variation could possibly be a result of selection based on farmer preferred traits (tasty pulp, large fruit size, and high oil content) across populations (Lovett & Haq, 2000; Odoi et al., 2020). This lack of variation among

shea populations offers opportunities for generating shea varieties that can be grown successfully in shea belts across geographical regions in Uganda. On the other hand, Table 1 indicates a high variation of phenotypic characteristics among shea tree in Uganda, although genetic results revealed a lack of variability indicating that the characteristics are not genetic in nature but shaped by the environmental and anthropogenic selection for domestication.

Findings from the assessment of genetic diversity in shea tree across its range in Africa (Fontaine et al., 2004) were closely similar (0.29) to the results of this study (0.21). The level of population differentiation (F_{st}) among shea tree populations in this study is also not different from other studies across shea belt in Africa (Hunter et al., 2012; Lovett & Haq, 2000). These findings, therefore, are useful in designing shea conservation strategies within the Sudano-Sahelian Africa to conserve and guarantee the continuity of the shea belt from the West African coast, through Central to Eastern highlands of Africa.

The genetic variability in such a population, represented by either observed or H_e , is essential for the genetic conservation and breeding of the species. The higher H_o (0.26) than the H_e (0.21) observed in the populations of this study points to high heterozygosity in *V. paradoxa* accessions in Uganda. This level of heterozygosity may have been caused by anthropogenic forces and the outcrossing nature of shea tree that encouraged gene flow among the populations (Hale et al., 2021). Similar results have been reported in self-pollinated tree species such as *Garcinia kola* (Dadjo et al., 2020; Olawuyi & Azeez, 2019). However, at the individual level, heterozygosity may lead to the selection of dominant tree accessions which may in turn increase heterozygosity at the marker level of the selected gene (Black et al., 2001). This lack of variation among shea trees has been associated with its allogamous nature, gene flow between populations, and the selective nature of trees with community-preferred traits such as tasty pulp (Odoi et al., 2020). However, with high anthropogenic effects through this selection and domestication of the species, bottlenecks shall occur that may disrupt heterozygosity through forced selfing and hybridization that reduce genetic diversity among domesticates compared to their wild progenitors (Ndjondjop et al., 2017).

The negative inbreeding coefficient indicates that the shea tree is a highly outcrossing species, which could be due to Null alleles that led to false discovery of homozygotes and cause heterozygote deficiency (Waples, 2015). This low fixation index in a population has been associated with high gene flow and genetic drift to minimize allele loss (Lee et al., 2017). Gene flow, inbreeding, natural selection, the Wahlund effect, and random genetic drift are some of the evolutionary processes identified as impacting the genetic diversity of a population (Wang et al., 2008).

The results of this study offer a good selection potential of putative shea tree genotypes as possible parent trees for future shea tree breeding programs in Uganda.

4.3 | Conservation implications

In consonant with the discussion, there is moderate genetic variability in *V. paradoxa*. However, anthropogenic factors affect its long-term conservation in Uganda. Therefore, there is an urgent need to embrace in situ and ex situ conservation strategies like the ones in the multilocal seed gene banks established by this study. The admixing gene pools should be promoted to increase their diversity, population size, and representation of the seven inferred gene pools. This therefore means that the selection of *V. paradoxa* parent plants from Arua, Amuru, and Ouke would suffice to capture maximum diversity for ex situ conservation. With the growing international and local demand for the shea products, the National Forestry Authority should include *V. paradoxa* in the list of tree species promoted under community tree planting program for the Teso, Lango, Acholi, and West Nile regions.

5 | CONCLUSION

This study provides insights into the molecular diversity and population structure of *Vitellaria paradoxa* in Uganda. We, therefore, conclude that the Eastern and West Nile ecotypes of *V. paradoxa* offer a genetic resource of importance to conservation and future breeding program in Uganda. Further studies are required to identify the hot spots of genetic diversity of the species in the two populations for the selection of parent materials for germplasms multiplication and conservation. This will inform shea conservation and management policy formulation to protect the hot spots from destruction and promote germplasm production for in situ and ex situ conservation.

Therefore, it is important to establish a collection of the different provenances and ethnovarieties in a multilocal breeding seed orchard germplasm to study their appropriate agronomic practices.

Determining the functional genes responsible for secondary metabolic processes conferring molecular mechanisms of preferred phenotypic traits in the shea tree is a key gap. These genomic markers have widened our understanding of shea genomes and other related species for future genomic studies, conservation, and breeding.

AUTHOR CONTRIBUTIONS

Juventine Odoi: Conceptualization, data curation, formal analysis, investigation, methodology, writing—original draft.

Emmanuel Amponsah Adjei: Formal analysis, methodology, software, writing–review and editing. **Judith Nantongo** and **Arfang Badji:** Formal analysis, software, writing–review and editing. **Alfred Ozimati:** Supervision, writing–review and editing. **Arfang Badji:** Formal analysis, software, writing review and editing. **Grace Nakabonge:** Methodology, supervision, writing–review and editing. **Thomas Odong:** Conceptualization, methodology, supervision, writing–review and editing. **Samson Gwali:** Conceptualization, methodology, supervision, writing–original draft, writing–review and editing. **Richard Edema:** Funding acquisition, Supervision, writing–review and editing. **Prasad Hendre:** Methodology, software, supervision, writing– original draft, writing–review and editing.

ACKNOWLEDGMENTS

The authors would like to extend their sincere thanks to the Biosciences Eastern and Central Africa-International Livestock Research Institute (BeCA-ILRI) and Integrated Genotyping Service & Support (IGSS) Unit for the waivers for genomic sequencing at BeCA-ILRI. The authors would also like to thank Dr. Hillary Agaba (the Director, National Forestry Resources Research Institute (NaFORRI) for all the support and guidance. The World Agroforestry provided funding under the Genebank Platform for plant material sample collection from the field. Makerere Regional Centre for Crop Improvement (MaRCCI) provided the student research grant for genetic fingerprinting and analysis at the Biosciences Eastern and Central Africa-International Livestock Research Institute (BeCA-ILRI).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

ORCID

Juventine Boaz Odoi  <https://orcid.org/0000-0002-1504-9304>

Emmanuel Amponsah Adjei  <https://orcid.org/0000-0002-3138-6060>

Alfred Adibo Ozimati  <https://orcid.org/0000-0003-0503-8107>

REFERENCES

- Barchi, L., Lanteri, S., Portis, E., Acquadro, A., Valè, G., Toppino, L., & Rotino, G. L. (2011). Identification of SNP and SSR markers in eggplant using RAD tag sequencing. *BMC Genomics*, 12(1), 304. <https://doi.org/10.1186/1471-2164-12-304>
- Beletskii, A., & Bhagwat, A. S. (1996). Transcription-induced mutations: Increase in C to T mutations in the nontranscribed strand during transcription in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 93(24), 13919–13924. <https://doi.org/10.1073/pnas.93.24.13919>
- Bennett, R. A., Thiagarajah, M. R., King, J. R., & Rahman, M. H. (2008). Interspecific cross of *Brassica oleracea* var. *alboglabra* and *B. napus*: Effects of growth condition and silique age on the efficiency of hybrid production, and inheritance of erucic acid in the self-pollinated backcross generation. *Euphytica*, 164(2), 593–601. <https://doi.org/10.1007/s10681-008-9788-0>
- Black, W. C., Baer, C. F., Antolin, M. F., & DuTeau, N. M. (2001). Population genomics: Genome-wide sampling of insect populations. *Annual Review of Entomology*, 46(1), 441–469. <https://doi.org/10.1146/annurev.ento.46.1.441>
- Blyton, M. D. J., & Flanagan, N. S. (2006). *A comprehensive guide to GenAlEx 6.5*. Australian National University.
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19), 2633–2635. <https://doi.org/10.1093/bioinformatics/btm308>
- Burkardt, K., Pettenkofer, T., Ammer, C., Gailing, O., Leinemann, L., Seidel, D., & Vor, T. (2021). Influence of heterozygosity and competition on morphological tree characteristics of *Quercus rubra* L.: A new single-tree based approach. *New Forests*, 52(4), 679–695. <https://doi.org/10.1007/s11056-020-09814-1>
- Christophe, D., Diaga, D., Seyni, S., & Kandiora, N. (2014). Morphological characterization of shea tree (*Vitellaria paradoxa* subsp. *paradoxa*) populations in the region of Mandoul in Chad. *International Journal of Biodiversity and Conservation*, 6(2), 184–193. <https://doi.org/10.5897/IJBC2013.0662>
- Dadjo, C., Nyende, A. B., Yao, N., Kiplangat, N., & Assogbadjo, A. E. (2020). Genome-wide genetic diversity and population structure of *Garcinia kola* (Heckel) in Benin using DArT-Seq technology. *PLoS One*, 15(9), e0238984. <https://doi.org/10.1371/journal.pone.0238984>
- Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, 19, 1–15.
- Earl, D. A., & vonHoldt, B. M. (2012). Structure harvester: A website and program for visualizing structure output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359–361. <https://doi.org/10.1007/s12686-011-9548-7>
- Egea, L. A., Mérida-García, R., Kilian, A., Hernandez, P., & Dorado, G. (2017). Assessment of genetic diversity and structure of large garlic (*Allium sativum*) germplasm bank, by diversity arrays technology “genotyping-by-sequencing” platform (DArTseq). *Frontiers in Genetics*, 8, 98. <https://doi.org/10.3389/fgene.2017.00098>
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology*, 14, 2611–2620.
- Fontaine, C., Lovett, P. N., Sanou, H., Maley, J., & Bouvet, J.-M. (2004). Genetic diversity of the shea tree (*Vitellaria paradoxa* C.F. Gaertn), detected by RAPD and chloroplast microsatellite markers. *Heredity*, 93(6), 639–648. <https://doi.org/10.1038/sj.hdy.6800591>
- Granato, I., Galli, G., Couto, E., Sousa, M., Mendoca, L., & Fritsch-Neto, R. (2018). snpReady: A tool to assist breeders in genomic analysis. *Molecular Breeding*, 38(8).
- Gruber, B., Unmack, P., Berry, O., & Georges, A. (2018). dartR: an R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Molecular Ecology Resources*, 18(3). <https://doi.org/10.1111/1755-0998.12745>
- Grzebelus, D., Iorizzo, M., Senalik, D., Ellison, S., Cavagnaro, P., Macko-Podgorni, A., Heller-Uszynska, K., Kilian, A., Nothnagel, T., Allender, C., Simon, P. W., & Baranski, R. (2014). Diversity, genetic mapping, and signatures of domestication in the carrot (*Daucus carota* L.) genome, as revealed by diversity arrays technology (DArT) markers. *Molecular Breeding*, 33(3), 625–637. <https://doi.org/10.1007/s11032-013-9979-9>

- Gudeta, T. B. (2018). Molecular marker based genetic diversity in forest tree populations. *Forestry Research and Engineering: International Journal*, 2(4), 176–182. <https://doi.org/10.15406/freij.2018.02.00044>
- Gwali, S., Vaillant, A., Nakabonge, G., Okullo, J. B. L., Eilu, G., Muchugi, A., & Bouvet, J.-M. (2015). Genetic diversity in shea tree (*Vitellaria paradoxa* Subspecies nilotica) ethnovarieties in Uganda assessed with Microsatellite markers. *Forests Trees and Livelihoods*, 24(3), 163–175.
- Hale, I., Ma, X., Melo, A. T. O., Padi, F. K., Hendre, P. S., Kingan, S. B., Sullivan, S. T., Chen, S., Boffa, J.-M., Muchugi, A., Danquah, A., Barnor, M. T., Jamnadass, R., van de Peer, Y., & van Deynze, A. (2021). Genomic resources to guide improvement of the shea tree. *Frontiers in Plant Science*, 12, 720670. <https://doi.org/10.3389/fpls.2021.720670>
- Hall, A. E. (2012). Phenotyping cowpeas for adaptation to drought. *Frontiers in Physiology*, 3, 155. <https://doi.org/10.3389/fphys.2012.00155>
- Hamrick, J. L., Mitton, J. B., & Linhart, Y. B. (1979). Levels of genetic variation in trees: Influence of life history characteristics. *Proceedings of the Symposium on Isozymes of North American Forest Trees and Forest Insects*, Berkeley, CA. https://www.fs.usda.gov/psw/publications/documents/psw_gtr048/psw_gtr048_hamrick.pdf
- Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T. K., Bateman, A., Bernard, T., Binns, D., Bork, P., Burge, S., de Castro, E., Coghill, P., Corbett, M., Das, U., Daugherty, L., Duquenne, L., Finn, R. D., Fraser, M., Gough, J., ... Yong, S.-Y. (2012). InterPro in 2011: New developments in the family and domain prediction database. *Nucleic Acids Research*, 40(D1), D306–D312. <https://doi.org/10.1093/nar/gkr948>
- Huynh, B., Close, T. J., Roberts, P. A., Hu, Z., Wanamaker, S., Lucas, M. R., Chiulele, R., Cissé, N., David, A., Hearne, S., Fatokun, C., Diop, N. N., & Ehlers, J. D. (2013). Gene pools and the genetic architecture of domesticated cowpea. *The Plant Genome*, 6(3), plantgenome2013.03.0005. <https://doi.org/10.3835/plantgenome2013.03.0005>
- Jombart, T., & Ahmed, I. (2011). ADEGENET 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21), 3070–3071. <https://doi.org/10.1093/bioinformatics/btr521>
- Ketema, S., Tesfaye, B., Keneni, G., Amsalu Fenta, B., Assefa, E., Greliche, N., Machuka, E., & Yao, N. (2020). DArTseq SNP-based markers revealed high genetic diversity and structured population in Ethiopian cowpea [*Vigna unguiculata* (L.) Walp] germplasms. *PLoS One*, 15(10), e0239122. <https://doi.org/10.1371/journal.pone.0239122>
- Kilian, A., Sanewski, G., & Ko, L. (2016). The application of DArTseq technology to pineapple. *Acta Horticulturae*, 1111, 181–188.
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35, 1547–1549.
- Lee, S. J., Ban, S. H., Kim, G. H., Kwon, S., Kim, J. H., & Choi, C. (2017). Identification of potential gene-associated major traits using GBS-GWAS for Korean apple germplasm collections. *Plant Breeding*, 136(6), 977–986. <https://doi.org/10.1111/pbr.12544>
- Lovett, P. N., & Haq, N. (2000). Diversity of the sheanut tree (*Vitellaria paradoxa* C.F. Gaertn.) in Ghana. *Genetic Resources and Crop Evolution*, 47(3), 293–304. <https://doi.org/10.1023/A:1008710331325>
- Lui, K. J., & Muse, S. V. (2005). Power marker: An integrated analysis environment for genetic marker analysis. *Bioinformatics*, 21, 2128–2129.
- Maranz, S., & Wiesman, Z. (2003). Evidence for indigenous selection and distribution of the shea tree, *Vitellaria paradoxa*, and its potential significance to prevailing parkland savanna tree patterns in sub-Saharan Africa north of the equator. *Journal of Biogeography*, 30(10), 1505–1516. <https://doi.org/10.1046/j.1365-2699.2003.00892.x>
- Meirmans, P. G. (2012). AMOVA-based clustering of population genetic data. *Journal of Heredity*, 103(5), 744–750. <https://doi.org/10.1093/jhered/ess047>
- Mohammad, N., Dahayat, A., Yadav, M., Shirin, F., & Ansari, S. A. (2018). Genetic diversity and population structure of *Litsea glutinosa* (Lour.) in central India. *Physiology and Molecular Biology of Plants*, 24(4), 655–663. <https://doi.org/10.1007/s12298-018-0556-x>
- Nantongo, J., Okullo, J. B. L., Eilu, G., Ramonta, I. R., Odee, D., & Cavers, S. (2010). Structuring of genetic diversity in *Albizia gummifera* C.A.Sm. among some East African and Madagascan populations. *African Journal of Ecology*, 48, 841–843.
- Nantongo, J., Odoi, J. B., Agaba, H., & Gwali, S. (2022). SilicoDART and SNP markers for genetic diversity and population structure analysis of *Trema orientalis*: A fodder species. *PLoS One*, 17, e0267464.
- Ndjiondjop, M.-N., Semagn, K., Gouda, A. C., Kpeki, S. B., Dro Tia, D., Sow, M., Goungoulou, A., Sie, M., Perrier, X., Ghesquiere, A., & Warburton, M. L. (2017). Genetic variation and population structure of *Oryza glaberrima* and development of a mini-core collection using DArTseq. *Frontiers in Plant Science*, 8, 1748. <https://doi.org/10.3389/fpls.2017.01748>
- Odoi, J. B., Muchugi, A., Okia, C. A., Gwali, S., & Odong, T. L. (2020). Local knowledge, identification and selection of shea tree (*Vitellaria paradoxa*) ethnovarieties for pre-breeding in Uganda. *The Journal of Agriculture and Natural Resources Sciences*, 7(1), 22–33.
- Odong, T. L., van Heerwaarden, J., Jansen, J., van Hintum, T. J. L., & van Eeuwijk, F. A. (2011). Determination of genetic structure of germplasm collections: Are traditional hierarchical clustering methods appropriate for molecular marker data? *Theoretical and Applied Genetics*, 123(2), 195–205. <https://doi.org/10.1007/s00122-011-1576-x>
- Olawuyi, O. J., & Azeez, A. A. (2019). Molecular evaluation of *Garcinia kola* Heckel accessions using RAPD markers. *American Journal of Molecular Biology*, 9(2), 41–51. <https://doi.org/10.4236/ajmb.2019.92004>
- Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: Genetic analysis in excel. Population genetic software for teaching and research—An update. *Bioinformatics*, 28(19), 2537–2539. <https://doi.org/10.1093/bioinformatics/bts460>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Raman, H., Raman, R., Kilian, A., Detering, F., Carling, J., Coombes, N., Diffey, S., Kadkol, G., Edwards, D., McCully, M., Ruperao, P., Parkin, I. A. P., Batley, J., Luckett, D. J., & Wratten, N. (2014). Genome wide delineation of natural variation for pod shatter resistance in *Brassica napus*. *PLoS ONE*, 9, e101673. <https://doi.org/10.1371/journal.pone.0101673>
- Sansaloni, C., Petroli, C., Jaccoud, D., Carling, J., Detering, F., Grattapaglia, D., & Sansaloni, A. K. (2011). Diversity Arrays Technology (DArT) and next generation sequencing combined: Genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. *BMC Proceedings*, 5, 54. <https://www.biomedcentral.com/1753-6561/5/S7/P54>

- Shepherd, D., Martin, D., McGivern, D., Boulton, M., Thomson, J., & Rybicki, E. (2005). A three-nucleotide mutation altering the maize streak virus rep pRBR-interaction motif reduces symptom severity in maize and partially reverts at high frequency without restoring pRBR-Rep binding. *Journal of General Virology*, 86, 803–813.
- Shimizu, T., Kitajima, A., Nonaka, K., Yoshioka, T., Ohta, S., Goto, S., Toyoda, A., Fujiyama, A., Mochizuki, T., Nagasaki, H., Kaminuma, E., & Nakamura, Y. (2016). Hybrid origins of citrus varieties inferred from DNA marker analysis of nuclear and organelle genomes. *PLoS One*, 11(11), e0166969. <https://doi.org/10.1371/journal.pone.0166969>
- Sodedji, F. A. K., Agbahoungba, S., Agoyi, E. E., Kafoutchoni, M. K., Choi, J., Nguetta, S. A., Assogbadjo, A. E., & Kim, H. (2021). Diversity, population structure, and linkage disequilibrium among cowpea accessions. *The Plant Genome*, 14(3), e20113. <https://doi.org/10.1002/tpg2.20113>
- Waples, R. S. (2015). Testing for hardy–weinberg proportions: Have we lost the plot? *Journal of Heredity*, 106(1), 1–19. <https://doi.org/10.1093/jhered/esu062>
- Wang, M. L., Barkley, N. A., Gillaspie, G. A., & Pederson, G. A. (2008). Phylogenetic relationships and genetic diversity of the USDA *Vigna* germplasm collection revealed by gene-derived markers and sequencing. *Genetics Research*, 90(6), 467–480. <https://doi.org/10.1017/S0016672308009889>
- Xiong, H., Shi, A., Mou, B., Qin, J., Motes, D., Lu, W., Ma, J., Weng, Y., Yang, W., & Wu, D. (2016). Genetic diversity and population structure of cowpea (*Vigna unguiculata* L. walp). *PLoS One*, 11(8), e0160941. <https://doi.org/10.1371/journal.pone.0160941>
- Yang, H., Wei, C.-L., Liu, H.-W., Wu, J.-L., Li, Z.-G., Zhang, L., Jian, J.-B., Li, Y.-Y., Tai, Y.-L., Zhang, J., Zhang, Z.-Z., Jiang, C.-J., Xia, T., & Wan, X.-C. (2016). Genetic divergence between camellia sinensis and its wild relatives revealed via genome-wide SNPs from RAD Sequencing. *PLoS One*, 11(3), e0151424. <https://doi.org/10.1371/journal.pone.0151424>

How to cite this article: Odoi, J. B., Adjei, E. A., Hendre, P., Nantongo, J. S., Ozimati, A. A., Badji, A., Nakabonge, G., Edema, R., Gwali, S., & Odong, T. L. (2023). Genetic diversity and population structure among Ugandan shea tree (*Vitellaria paradoxa* subsp. *nilotica*) accessions based on DarTSeq markers. *Crop Science*, 1–13. <https://doi.org/10.1002/csc2.21013>