Plant Breeding

WILEY

Plant Breeding

ORIGINAL ARTICLE OPEN ACCESS

Genetic Diversity and Population Structure of Sweetpotato Accessions (*Ipomoea batatas* [L.] Lam.) Revealed by Single Nucleotide Polymorphism Markers

Nomusa Chizhande[1,2] | Julia Sibiya[1,3] | Edmore Gasura[4] | Bruce Mutari[5] | Davison Chaingeni[2] | Dumisani Kutywayo[2] | Lennin Musundire[6] | Isack Mathew[7] | Emeline Dossa[2]

[1]School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Pietermaritzburg, South Africa | [2]Department of Research and Specialist Services, Crops Research Division, Harare, Zimbabwe | [3]Vision for Adapted Crops and Soils (VACS) Capacity, SAS Program, International Maize and Wheat Improvement Center, (CIMMYT), Nairobi, Kenya | [4]Department of Plant Production Sciences and Technologies, University of Zimbabwe, Harare, Zimbabwe | [5]Alliance for a Green Revolution in Africa, Nairobi, Kenya | [6]Accelerated Breeding Initiative (ABI)-Transform, International Maize and Wheat Improvement Center, (CIMMYT), Nairobi, Kenya | [7]Department of Plant & Soil Sciences, FESA, University of Venda, Thohoyandou, South Africa

Correspondence: Nomusa Chizhande (nomsachizhande@gmail.com)

Received: 22 May 2025 | **Revised:** 19 July 2025 | **Accepted:** 12 August 2025

Funding: This work was supported by the Accelerated Breeding Initiative (ABI)-Transform ABI.

Keywords: genetic diversity | genotyping | polymorphic information content | SNP markers | sweetpotato

ABSTRACT

Use of molecular markers has improved the analysis of genetic variation by eliminating environmental influences on genotype performance. The objective of this study was to assess the genetic diversity (GD) and population structure of 327 sweetpotato genotypes sourced from the major sweetpotato-growing regions of Zimbabwe and from the International Potato Centre (CIP) in Mozambique using low-density Diversity Array Technology (DArTseq) SNP chip covering the 90 chromosomes of sweetpotato. The genotypes' GD varied from 0.12 to 0.50, with a mean of 0.36. The mean PIC value of the SNP markers was 0.29. As SNP markers are biallelic with a maximum PIC of 0.50, this value indicates a moderate level of polymorphism. There was a good representation of minor alleles within the population, with an average minor allele frequency (MAF) of 0.26. The average observed heterozygosity of 0.12 was consistent with the cross-pollinating system in sweetpotato but could perpetuate a narrow genetic base. There was limited interbreeding between the populations of sweetpotato, as indicated by a mean fixation index (*F*) of 0.68. The high *F* values indicated that most alleles per genotype were contributed by one parent, which is unusual in allogamous species such as sweetpotato. The sweetpotato genotypes in this study could be clustered into two sub-populations with significant differences within the sub-populations. Genetic variation among genotypes is essential for the improvement of sweetpotato. Still, significant genetic gain could be achieved by cross-pollinating divergent genotypes with high MAF to create segregants with rare alleles. It is, thus, important to capture the rare alleles as they help adapt to current and future environmental shifts.

1 | Introduction

Sweetpotato (*Ipomoea batatas* [L.] Lam.) is an important root crop used for food, feed and biofuel. The importance of sweetpotato has increased due to the rising prices of staple food crops such as maize (*Zea mays*), wheat (*Triticum aestivum L*), rice (*Oryza sativa*) and Irish potato (*Solanum tuberosum*) (Sapakhova et al. 2023). Although its high biomass productivity and adaptability to diverse production environments make it ideal for production in marginal environments (Chiona 2009;

Cole et al. 2014), several biotic (viruses and weevils) and abiotic (drought and low soil fertility) factors still challenge its production.

In addition, a lack of improved varieties for production has led to low average yields of 5 t/ha attained by farmers against a potential of 23–25 t/ha (van Vugt and Franke 2018). Many smallholder farmers still cultivate landraces and varieties that have low yield potential (Eriksson et al. 2018). However, landraces are important sources of genes for adaptation to biotic and abiotic stresses compared to improved varieties. Incorporating landraces in breeding programmes can broaden the gene pool for crop improvement but must be followed by effective and efficient selection for high yield and desirable agronomic performance (Lazaridi et al. 2024). Therefore, understanding the available genetic variation and population structure is imperative for devising suitable breeding strategies.

Developing new improved varieties depends on wide genetic variation for selection (Ulukan 2009). To evaluate genetic variation for significant traits and assess the population structure for conservation purposes, morphological and molecular markers have been employed (Emanuelli et al. 2013). Molecular markers are particularly useful for evaluating genetic diversity (GD) and relatedness among various genetic resources. To assess genetic variation in sweetpotato, several marker types have been used, such as simple sequence repeats (SSRs) or microsatellites (Buteler et al. 2002; Gichuru et al. 2006; Koussao et al. 2014; Imamura et al. 2014; Ngailo et al. 2016), inter-simple sequence repeats (ISSRs) (Huang and Sun 2000), random amplified polymorphic DNA (RAPD) (Gasura and Mukasa 2011; Maquia et al. 2013), amplified fragment length polymorphism (AFLP) (Zhang et al. 2000; Elameen et al. 2008), selective amplification of microsatellite polymorphic loci (SAMPL) (Yu-Tien Tseng et al. 2002) DNA amplification fingerprinting (He et al. 1995) and single nucleotide polymorphism (SNPs) (Arif et al. 2010). SNPs markers, in particular, facilitate genetic studies because they are amenable to automation and are highly reproducible (Jehan and Lakhanpaul 2006) compared to the other markers.

The Crop Breeding Institute (CBI) in Zimbabwe is involved in breeding for biofortified sweetpotato cultivars to improve nutritional and processing quality, and market preference. Accordingly, the CBI often introduces sweetpotato germplasm from diverse genetic backgrounds into Zimbabwe. The introductions are not well characterized for immediate incorporation into breeding programmes. It is imperative to conduct genetic studies on the introduced and local germplasm as a pre-breeding activity to devise appropriate breeding strategies. Therefore, this study aimed to assess the population structure and GD of local and introduced sweetpotato genotypes in Zimbabwe using SNP markers. The study will contribute to baseline information required to inform breeding efforts for sweetpotato improvement in Zimbabwe and its germplasm exchange partners.

2 | Materials and Methods

2.1 | Plant Material

A total of 327 sweetpotato genotypes were used in this study (Table S1). The genotypes consisted of landraces collected from the major sweetpotato-growing regions of Zimbabwe (Figure 1) and 181 genotypes from the International Potato Centre (CIP) in Mozambique. From Zimbabwe, 89 landraces

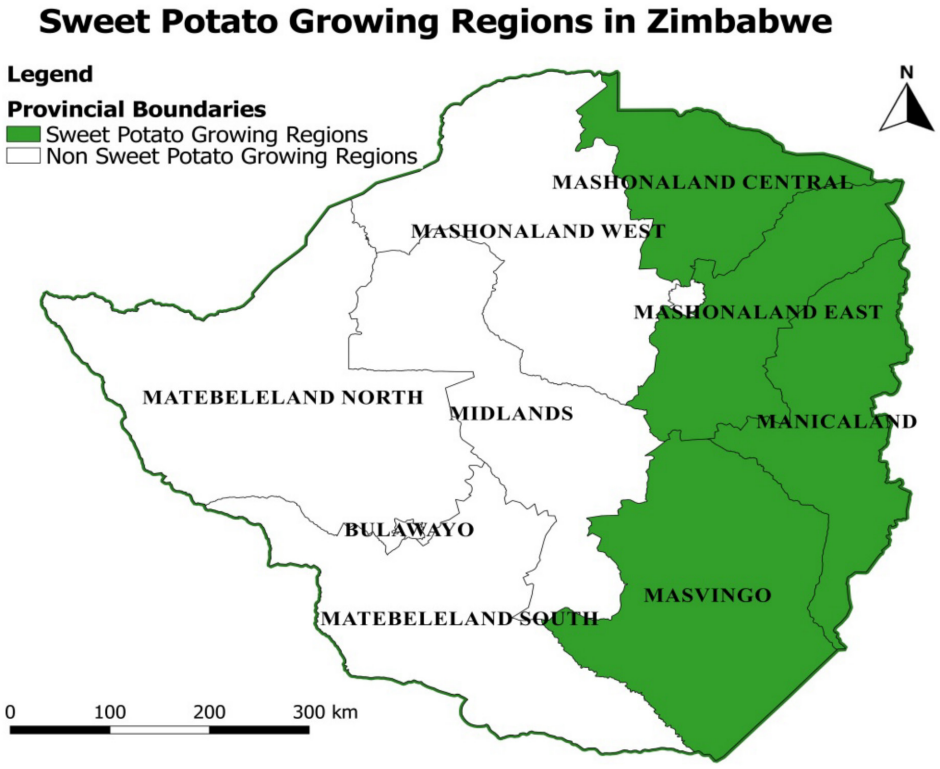

FIGURE 1 | Sweetpotato growing regions in Zimbabwe. [Color figure can be viewed at wileyonlinelibrary.com]

were collected from Mashonaland East, 27 from Manicaland, 16 from Masvingo and 10 from Mashonaland Central provinces. The genotypes were grouped based on the geographic region of their source for further analysis, with the groups comprising genotypes obtained from CIP-Mozambique as one group and the other group genotypes sourced in Zimbabwe. The group of genotypes from Zimbabwe was further subdivided into four sub-populations based on the province from which they were collected.

2.2 | Leaf Sampling and Shipment

Fresh leaf samples were collected from 3-week-old seedlings of each genotype for deoxyribonucleic acid (DNA) extraction. Supporting Information S3 provides the leaf sampling, drying, packaging and shipment protocol used in this study as guided by Intertek laboratory in Sweden https://www.intertek.com/agriculture/testing/. Four leaf disks were punched out for each accession using a leaf-cutting punch. To avoid cross-contamination between accessions, 70% alcohol was used to sterilize the puncher. Sampled leaf discs were oven-dried at 35°C for 24 h, and the plates were immediately secured with lids after drying. The sampling plates were packaged and shipped to the Intertek laboratory in Sweden https://www.intertek.com/contact/ema/sweden/. The samples were genotyped on a low-density Diversity Array Technology (DArTseq) SNP chip covering the 90 chromosomes of sweetpotato. Diversity Arrays Technology SNP (DArT SNP) markers were aligned to the reference genomes of wild sweetpotato (*Ipomoea Trifida v.3*).

2.3 | Data Filtering and SNP Calling

The snpReady software was used to filter the first 37 SNPs from the genotyping-by-sequencing (GBS) pipeline by eliminating SNPs with 20% missing data and less than 5% minor allele frequency (MAF) (Italo 2018). Individual genotypes with missing data, more than 20%, were eliminated. SNPs with a MAF of 0.05 and a call rate of 80% or above were kept. As a result, analysis was conducted using 327 genotypes and 30 SNPs, and of the 30, 27 were polymorphic (Table 1).

2.4 | Data Analysis

The Nei and Li (1979) technique was used to calculate the number of alleles per locus, the number of effective alleles per locus, Shannon's Information Index and gene diversity. An analysis of molecular variance (AMOVA) was conducted to partition total genetic variation within and among the different groups of the genotypes using GenAlex version 6.5 and R Studio version 4.3.0 (Peakall and Smouse 2012). Population structure analysis was performed using the Structure Software Version 2.3.4 (Pritchard et al. 2009).

The burn-in period and the Markov Chain Monte Carlo (MCMC) replications were set at 10,000. The structure analysis was done for K ranging from 1 to 10 with five iterations at each K to determine the optimum number of clusters. The ideal K value was identified using the Evanno method, which relies on ΔK (Evanno et al. 2005).

The GD parameters including expected heterozygosity (He), overall GD, polymorphic information content (PIC), the main allele frequency and the breeding coefficient (FIS) were calculated; the pairwise genetic distances and kinship matrix among the 327 sweetpotato genotypes were calculated using GenAlex version 6.5 and R studio Version 4.3.0. A dendrogram was then generated on the dissimilarity matrix using Gower's Ward.D2 hierarchical clustering method.

3 | Results

3.1 | Genetic Parameters of the Markers and Genotypes

The PIC for the markers ranged from 0.11 to 0.37, with an average of 0.29 for the 30 SNP markers (Table 1). The most informative markers were snpIB00001, snpIB000010, snpIB00012, snpIB00032 and snpIB00051, while snpIB00057 was the least informative. The minor allele frequency ranged between 0.3 and 0.5, averaging 0.26. The average expected heterozygosity (He) and observed heterozygosity (Ho) were 0.12 and 0.36, respectively. The marker snpIB00032 had the maximum heterozygosity of 0.69. The maximum Nei GD index was 0.5 and the minimum was 0.12, with an average of 0.36 (Table 1).

The genotypes' GD varied from 0.12 to 0.50, averaging 0.36 (Table 1). The PIC showed a broad range from 0.11 to 0.37, with a mean of 0.29. The minor allele frequency (MAF) spanned from 0.06 to 0.50, averaging 0.26. Observed heterozygosity ranged from 0.00 to 0.23, with a mean of 0.12. The fixation index (F) had a mean of 0.68, ranging from 0.36 to 1.00 (Table 2).

3.2 | Population Structure as Revealed by SNP Markers

Evanno's method deduced that the highest ΔK value occurred at $K = 2$ (Figure 2). The scree plot shows a significant drop in ΔK after $K = 2$, and all the other values of K show no change in ΔK in the scree plot, confirming that K was optimal at 2. The two populations were divided into the largest group with 250 genotypes and the smaller group with 77 genotypes (Figure 3). The distribution of the genotypes into clusters was based on kinship. Very few individual genotypes exhibited kinship to both groups. Different colours represent separate clusters, while the length of the coloured segment shows the estimated proportion of the membership of a genotype in a particular population.

3.3 | Cluster Analysis of the Genotypic Data for the 327 Genotypes

The neighbour-joining dendrogram clustered the 327 genotypes into five clusters: C1, C2, C3, C4 and C5 based on the SNP markers (Figure 4). The biggest cluster was Cluster 2, which

TABLE 1 | Genetic diversity parameters were revealed by analysing 30 SNP markers among the 327 sweetpotato accessions.

SNP_Name	Chr	Position (ref. *Ipomoea trifida v.3*)*	Allele	MAF	PIC	Allele no.	Ho	He	Nei
snpIB00031	1	24,315,805	C/T	0.22	0.28	2	0.02	0.34	0.34
snpIB00032	1	31,682,527	C/T	0.49	0.37	2	0.69	0.5	0.5
snpIB00021	2	6,247,633	T/G	0.19	0.26	2	0.09	0.3	0.3
snpIB00027	2	13,184,152	G/T	0.24	0.3	2	0.02	0.37	0.37
snpIB00035	3	12,595,421	C/G	0.07	0.12	2	0.02	0.13	0.13
snpIB00026	3	24,217,089	A/T	0.29	0.33	1	0	0.41	0.41
snpIB00036	4	8,685,123	C/T	0.25	0.31	2	0.13	0.38	0.38
snpIB00037	4	29,217,872	A/G	0.14	0.21	2	0.01	0.24	0.24
snpIB00038	5	738,566	C/T	0.1	0.16	2	0.02	0.17	0.17
snpIB00040	6	13,312,429	A/C	0.37	0.36	2	0.05	0.47	0.47
snpIB00030	6	19,672,316	G/C	0.09	0.16	2	0.08	0.17	0.17
snpIB00042	7	969,986	A/T	0.39	0.36	1	0	0.48	0.48
snpIB00043	7	17,963,596	C/T	0.27	0.32	2	0.06	0.39	0.39
snpIB00029	7	23,485,155	T/G	0.2	0.27	2	0.1	0.32	0.32
snpIB00044	8	3,253,902	C/T	0.34	0.35	2	0.47	0.45	0.45
snpIB00045	8	3,449,546	C/T	0.19	0.26	2	0.05	0.31	0.31
snpIB00001	8	6,218,106	A/G	0.42	0.37	2	0.01	0.49	0.49
snpIB00046	9	437,404	C/T	0.23	0.29	2	0.03	0.36	0.36
snpIB00010	10	4,446,069	G/A	0.5	0.37	2	0.03	0.5	0.5
snpIB00049	10	22,810,642	G/T	0.23	0.29	2	0.02	0.36	0.36
snpIB00050	11	1,435,859	A/G	0.11	0.18	1	0	0.2	0.2
snpIB00051	11	5,905,491	G/T	0.47	0.37	2	0.03	0.5	0.5
snpIB00012	12	1,719,732	C/A	0.47	0.37	2	0.61	0.5	0.5
snpIB00052	12	8,660,781	A/G	0.28	0.32	2	0.06	0.4	0.4
snpIB00053	12	23,669,355	A/T	0.25	0.3	2	0.01	0.37	0.37
snplB00054	13	221,791	C/G	0.23	0.29	2	0.04	0.36	0.36
snpIB00056	14	16,102,914	C/T	0.3	0.33	2	0.53	0.42	0.42
snpIB00057	14	18,292,043	C/G	0.06	0.11	2	0.02	0.12	0.12
snpIB00058	15	619,627	A/G	0.2	0.27	2	0.02	0.32	0.32
snpIB00060	15	7,544,880	C/T	0.34	0.35	2	0.25	0.45	0.45
Minimum				0.06	0.11	1	0.23	0.12	0.12
Maximum				0.5	0.37	2	0	0.5	0.5
Mean				0.26	0.29		0.12	0.36	0.36

Abbreviations: He = expected heterozygosity; Ho = observed heterozygosity; MAF = minor allele frequency; Nei = Nei genetic distance; PIC = polymorphic information content.
*A note explaining that the chromosome position is based on the reference genome*Ipomoea trifida v.3* as indicated in the table heading.

comprised 170.04 shares. The next biggest cluster was Cluster 3, with 71.94, and Cluster 1, with 42.51. Clusters 4 and 5 were the smallest, with 32.7 and 9.81 memberships, respectively. The names of individuals belonging to each cluster are provided in Table S2.

3.4 | Principal Coordinate Clustering

The principal coordinate analysis (PCoA) showed variability among sweetpotato accessions based on the 30 SNP markers. The 30 SNP markers were used as a cost-effective rapid QA/QC set for

sweetpotato (Gemenet et al. 2020). The two PCs on the PCoA biplot accounted for 90% of the genotype variation (Figure 5). The genotypes from CIP-Mozambique were mostly clustered in the bottom two quadrants, while the genotypes from Zimbabwe were scattered in the top quadrants. The genotypes from Zimbabwe did not show further clustering by their source province.

3.5 | Analysis of Molecular Variance

The analysis of molecular variance (AMOVA) among the 327 sweetpotato genotypes showed that all the variance observed was due to genotypic differences within the two populations (Table 3). The expected heterozygosity of genes within populations was 0.09 for Population I and 0.15 for Population II (Table 4). The corresponding fixation indices (F_{ST}) for both Populations I and II were 0.08. The estimated Wright's fixation index (PhiPT max) of 0.591 and 0.291 was obtained for Subpopulations I and II, respectively.

TABLE 2 | Genetic parameters of 327 sweetpotato genotypes measured with 30 SNP markers.

Statistics	GD	PIC	MAF	Ho	*F*
Mean	0.36	0.29	0.26	0.12	0.68
Lower	0.12	0.11	0.06	0.00	0.36
Upper	0.50	0.37	0.50	0.23	1.00

Abbreviations: FST = fixation Index; GD = gene diversity; Ho = observed heterozygosity; MAF = minor allele frequency; PIC = polymorphic information content.

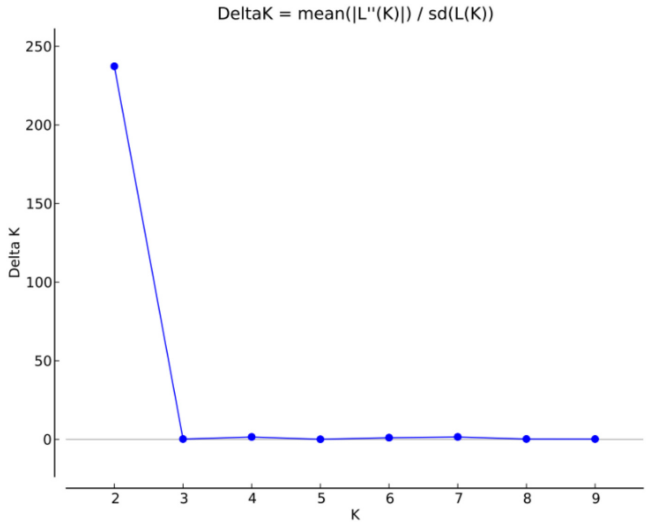

FIGURE 2 | A scree plot showing the Delta K values for the number of clusters determined among 327 sweetpotato genotypes measured using 30 SNP markers. [Color figure can be viewed at wileyonlinelibrary.com]

4 | Discussion

The average PIC value of 0.29 is regarded as moderate to high for SNP markers. Values above 0.25 are typically considered meaningful for GD research since biallelic SNPs have a maximum PIC of 0.50 (Botstein et al. 1980; Baloch et al. 2017). As a result, the SNP markers employed here showed adequate discriminatory power to evaluate genetic variation and successfully distinguish between the genotypes. Usually, bigger sets of markers may have higher resolution. Smaller sets of markers can offer some advantages by balancing accuracy and cost-effectiveness in breeding programmes. Given that the markers are SNP, small and negligible errors can occur compared to a larger data set but they are more affordable and provide useful information for characterization and identifying variation. The 30 markers used in this study were selected for genome coverage, threshold on missing values, chromosome position, genetic distances and validation in progeny. The markers used in this study were highly informative based on high PIC values and MAF. For the purposes of broad characterization, the markers were sufficient because the sweetpotato genotypes were diverse and being a hexaploid. Gemenet et al. (2020) used 30 SNP in a quality control study and found that the small number of markers was sufficient with only a 4% difference compared to using 300 markers. The PIC values and expected heterozygosity (He) are indicators of GD within breeding populations, evolutionary pressures and mutation rates at specific gene locations (Botstein et al. 1980; Shete et al. 2000). PIC values are particularly helpful in linkage analysis, indicating how well markers track inheritance between parents and offspring (Zhang et al. 2016). Though SNPs have lower PIC values due to their bi-allelic nature compared to SSR markers, they offer higher resolutions for genetic differentiation among genotypes because they detect differences at specific loci (Helyar et al. 2011; Suvi et al. 2020). The three markers that showed higher polymorphisms can be used in the future genotyping of sweetpotato clones to identify distant parents for developing and deploying varieties. The markers with low PIC may still be linked to important traits or regions of interest. Low PIC of a marker does not discount its potential value in genetic studies (Gupta et al. 2001).

This study's heterozygosity (Ho) mean of 0.12 was very low, suggesting that the sweetpotato genotypes were highly homozygous. This is unusual for sweetpotato since it is allogamous. Sweetpotato is a hexaploidy with generally high heterozygosity for most loci (Monden and Tahara 2017; Lindqvist-Kreuze

FIGURE 3 | Population structure of 327 sweetpotato genotypes based on 30 SNP markers. [Color figure can be viewed at wileyonlinelibrary.com]

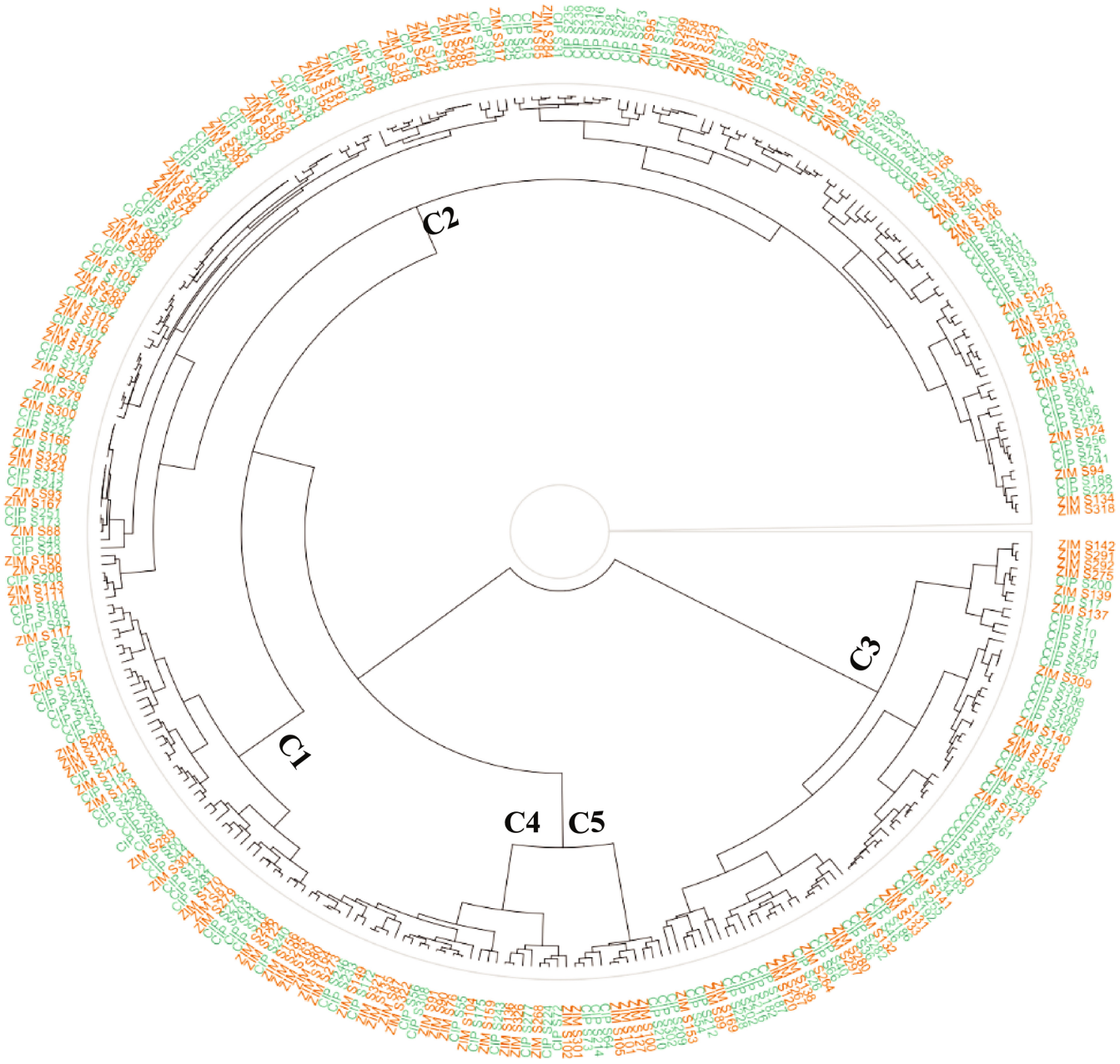

FIGURE 4 | Neighbour-joining dendrogram showing the clustering of 327 accessions genotyped with SNP markers. C1–C5 = Clusters 1 to 5. [Color figure can be viewed at wileyonlinelibrary.com]

et al. 2024). However, most sweetpotato genotypes in Zimbabwe produce sterile flowers due to unfavourable photoperiodicity and altitude. As such, sweetpotato is predominantly reproduced vegetatively and cloned, which could have caused the low level of heterozygosity (Gurmu et al. 2013). Low heterozygosity and high inbreeding coefficient (F) values are better understood as representing the distinct genetic makeup of the sampled germplasm, which was influenced by clonal propagation as opposed to intrinsic self-fertilization. Prior research has documented similar trends, attributing low GD in sweetpotato collections to vegetative propagation, rather than reproductive biology (Roullier et al. 2013; Mabaya et al. 2025). These results highlight the difficulty of interpreting diversity measurements in clonally generated polyploids, even though they would seem unusual for a self-incompatible, outcrossing species like sweetpotato.

In contrast, other reports have shown high heterozygosity in sweetpotato (Monden and Tahara 2017; Lindqvist-Kreuze et al. 2024) Ultimately, the lack of heterozygosity in the study genotypes could pose a serious bottleneck for breeding and can threaten sweetpotato production should new biotic and abiotic stresses emerge. Low levels of heterozygosity are often associated with limited genetic variation for resistance to emerging biotic or abiotic stresses (Baafi et al. 2016). The study showed that the inbreeding coefficient (FIS) varied widely and had a high mean value, indicating that the genotypes in the population underwent significant inbreeding. The F value estimates the likelihood that two alleles at every locus within an organism are descendants of the same parent (Wright and Adams 2012). High F values in the study indicated that one parent contributed the most alleles per genotype. This further emphasizes that most genotypes were

clones or developed from self-pollination, where the same parent contributes all the chromosomes to the offspring. This is common in sweetpotato, which suffers from a lot of cross-incompatibility. Ultimately, there are high levels of inbreeding (Gurmu et al. 2013). Due to a lack of diversity within individual genotypes, inbreeding could lead to suppressed yield potential or unfavourable nutritional content.

Generally, with some exceptions, genotypes were randomly allocated to different groups and subgroups. Some closely related genotypes were grouped into the same cluster or subcluster (cluster 2), confirming the existence of a relationship between the pedigree and SNP marker groupings in this study. Some of these genotypes appeared to be clustered according to the pedigree group (DOF/31, DOF/34, DOF/52, DOF/53, DOF/56, DOF/61, DOF/63 and DOF/64), but there were some inconsistencies. For instance, Chemugomo, Masvingo 8,

TABLE 3 | Analysis of molecular variance of sweetpotato populations based on 30 SNPs.

Source	df	SS	MS	Est. Var.	%
Among pops	1	9.391	9.391	0.029	0%
Within pops	324	1951.367	6.023	6.023	100%
Total	325	1960.758		6.052	100%

Abbreviations: % = percentage variation; df = degree of freedom; Est. Var = estimated variance; SS = sum of squares.

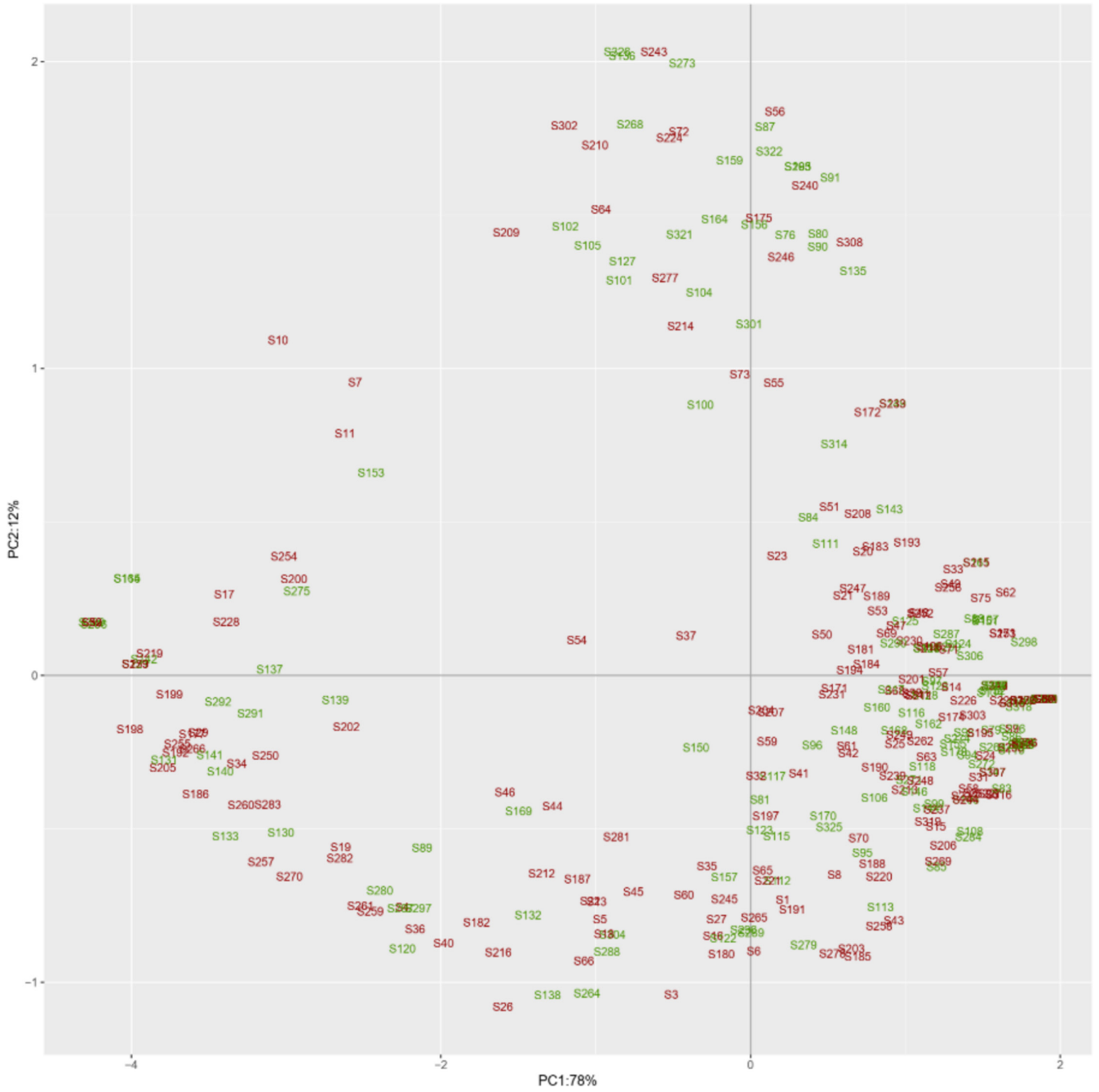

FIGURE 5 | Principal coordinate analysis showing variability among sweetpotato accessions from different regions in Zimbabwe based on 30 SNP markers. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 4 | Expected heterozygosity, fixation indices and membership of the two populations found among the 327 genotypes assessed with SNP markers.

	FST		He	Mean Wright FST	No of genotypes	Membership
I	—	0.08	0.09	0.59	250	0.75
II	0.08	—	0.15	0.29	77	0.25

Abbreviations: FST = allele frequency divergence among subpopulations; He = expected heterozygosity.

Hwedza 2, DOF/37 and Murehwa clustered together in cluster 5 despite being unrelated by pedigree. Similar findings of genotypes clustering according to their pedigree grouping were earlier reported (Dhliwayo et al. 2009; Semagn et al. 2012; Yang et al. 2017).

Genotypes were not clustered based on geographical location, which could be due to the exchange of genetic material in geographical regions. Therefore, crossing more distant sweetpotato genotypes might create wider genetic variation and exploit heterosis during breeding for new varieties (Molin et al. 2013). The molecular variability indicated that genetic variation existed only between genotypes within populations, and no variation was observed between different populations.

This might be attributed to the limited number of markers that were used in this study. On the other hand, the exchange of genetic material among the sources means that there is potentially no differentiation based on the collection areas. Consequently, the observed relationships among genotypes within each subpopulation could be exploited to design an effective sweetpotato breeding programme in Zimbabwe.

5 | Conclusion and Recommendations

The underlying GD and population structure in the sweetpotato introductions from CIP and local germplasm can potentially be exploited for improving sweetpotato cultivars in Zimbabwe and other countries with similar germplasm. The SNP markers used in this study were polymorphic and exhibited discriminating capability to diagnose allelic differences among the 327 genotypes. The sweetpotato genotypes were genetically divergent, providing adequate allelic diversity for recombination. Genotypes such as G40 and G9, G142 and G73, G114 and G73, I40 and G210 and G18 and G114 were distantly related and could be useful parents for recombination and developing breeding populations.

Author Contributions

Nomusa Chizhande: conceptualization, investigation, methodology, writing – original draft. **Julia Sibiya:** supervison, writing – review and editing. **Edmore Gasura:** supervision, methodolgy, writing – review and editing. **Bruce Mutari:** writing – reveiw and editing. **Davison Chaingeni:** writing – reveiw and editing. **Dumisani Kutywayo:** writing – reveiw and editing. **Lennin Musundire:** writing – reveiw and editing. **Isack Mathew:** writing – reveiw and editing. **Emeline Dossa:** data curation, formal analysis, writing – review and editing. All authors have read and approved the final manuscript.

Acknowledgements

The authors thank the staff at Marondera Horticultural Research Institute and the Roots and Tuber Crops technical team at Harare Research Station for their technical assistance. They also thank the International Potato Center (CIP) for providing some of the material.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The datasets generated and/or analysed during the current study are not publicly available but are available from the corresponding author on reasonable request.

References

Arif, I. A., M. A. Bakir, H. A. Khan, et al. 2010. "A Brief Review of Molecular Techniques to Assess Plant Diversity." *International Journal of Molecular Sciences* 11, no. 5: 2079–2096.

Baafi, E., K. Ofori, E. E. Carey, V. E. Gracen, E. T. Blay, and J. Manu-Aduening. 2016. "Genetic Control of Beta-Carotene, Iron and Zinc Content in Sweetpotato." *Journal of Plant Studies* 6, no. 1: 1–11.

Baloch, F. S., A. Alsaleh, M. Q. Shahid, et al. 2017. "A Whole Genome DArTseq and SNP Analysis for Genetic Diversity Assessment in Durum Wheat From Turkey." *Genes* 8, no. 5: 126.

Botstein, D., R. L. White, M. Skolnick, and R. W. Davis. 1980. "Botstein." *American Journal of Human Genetics* 32: 314–331.

Buteler, M. I., D. R. LaBonte, R. L. Jarret, and R. E. Macchiavelli. 2002. "Microsatellite-Based Paternity Analysis in Polyploid Sweetpotato." *Journal of the American Society for Horticultural Science* 127, no. 3: 392–396.

Chiona, Martin 2009. "Towards Enhancement of β-Carotene Content of High Dry Mass Sweetpotato Genotypes in Zambia By." (December).

Cole, J. R., Q. Wang, J. A. Fish, et al. 2014. "Ribosomal Database Project: Data and Tools for High Throughput RRNA Analysis." *Nucleic Acids Research* 42, no. D1: 633–642.

Dhliwayo, T., K. Pixley, A. Menkir, and M. Warburton. 2009. "Combining Ability, Genetic Distances, and Heterosis Among Elite CIMMYT and IITA Tropical Maize Inbred Lines." *Crop Science* 49: 1201–1210.

Elameen, A., S. Fjellheim, A. Larsen, et al. 2008. "Analysis of Genetic Diversity in a Sweet Potato (*Ipomoea batatas* L.) Germplasm Collection From Tanzania as Revealed by AFLP." *Genetic Resources and Crop Evolution* 55, no. 3: 397–408.

Emanuelli, F., et al. 2013. "Genetic Diversity and Population Structure Assessed by SSR and SNP Markers in a Large Germplasm Collection of Grape." *BMC Plant Biology* 13, no. 1: 1–17.

Eriksson, D., M. Akoroda, G. Azmach, M. Labuschagne, N. Mahungu, and R. Ortiz. 2018. "Measuring the Impact of Plant Breeding on Sub-Saharan African Staple Crops." *Outlook on Agriculture* 47, no. 3: 163–180.

Evanno, G., S. Regnaut, and J. Goudet. 2005. "Detecting the Number of Clusters of Individuals Using the Software STRUCTURE: A Simulation Study." *Molecular Ecology* 14, no. 8: 2611–2620.

Gasura, E., and S. B. Mukasa. 2011. "Prevalence and Implications of Sweetpotato Recovery From Sweet Potato Virus Disease in Uganda." *African Crop Science Journal* 18, no. 4: 195–205.

Gemenet, D. C., M. N. Kitavi, M. David, et al. 2020. "Development of Diagnostic SNP Markers for Quality Assurance and Control in Sweetpotato [*Ipomoea batatas* (L.) Lam.] Breeding Programs." *PLoS ONE* 15, no. 4: e0232173.

Gichuru, V., V. Aritua, G. W. Lubega, et al. 2006 "A Preliminary Analysis of Diversity Among East African Sweetpotato Landraces Using Morphological and Simple Sequence Repeats (SSR) Markers." 159–164.

Gupta, P. K., J. K. Roy, and M. Prasad. 2001. "Single Nucleotide Polymorphisms: A New Paradigm for Molecular Marker Technology and DNA Polymorphism Detection With Emphasis on Their Use in Plants." *Current Science* 80, no. 4: 524–535.

Gurmu, F., S. Hussein, and M. Laing. 2013. "Self- and Cross-Incompatibilities in Sweetpotato and Their Implications on Breeding." *Australian Journal of Crop Science* 7, no. 13: 2074–2078.

He, J., S. Choe, R. Walker, P. di Marzio, D. O. Morgan, and N. R. Landau. 1995. "Human Immunodeficiency Virus Type 1 Viral Protein R (Vpr) Arrests Cells in the G2 Phase of the Cell Cycle by Inhibiting P34cdc2 Activity." *Journal of Virology* 69, no. 11: 6705–6711.

Helyar, S. J., J. Hemmer-Hansen, D. Bekkevold, et al. 2011. "Application of SNPs for Population Genetics of Nonmodel Organisms: New Opportunities and Challenges." *Molecular Ecology Resources* 11, no. Suppl. 1: 123–136.

Huang, J. C., and M. Sun. 2000. "Genetic Diversity and Relationships of Sweetpotato and Its Wild Relatives in Ipomoea Series Batatas (Convolvulaceae) as Revealed by Inter-Simple Sequence Repeat (ISSR) and Restriction Analysis of Chloroplast DNA." *Theoretical and Applied Genetics* 100, no. 7: 1050–1060.

Imamura, K., N. Imamachi, G. Akizuki, et al. 2014. "Long Noncoding RNA NEAT1-Dependent SFPQ Relocation From Promoter Region to Paraspeckle Mediates IL8 Expression Upon Immune Stimuli." *Molecular Cell* 53, no. 3: 393–406. https://doi.org/10.1016/j.molcel.2014.01.009.

Italo. 2018. SnpReday and BGGE: R Packages to Prepare Genomic Datasets and Perform Genome-Enabled Predictions.

Jehan, T., and S. Lakhanpaul. 2006. "Single Nucleotide Polymorphism (SNP) - Methods and Applications in Plant Genetics: A Review." *Indian Journal of Biotechnology* 5, no. 4: 435–459.

Koussao, S., V. Gracen, I. Asante, et al. 2014. "Diversity Analysis of Sweet Potato (*Ipomoea Batatas* [L.] Lam) Germplasm From Burkina Faso Using Morphological and Simple Sequence Repeats Markers." *African Journal of Biotechnology* 13, no. 6: 729–742.

Lazaridi, E., A. Kapazoglou, M. Gerakari, et al. 2024. "Crop Landraces and Indigenous Varieties: A Valuable Source of Genes for Plant Breeding." *Plants* 13, no. 6: 1–23.

Lindqvist-Kreuze, H., M. Bonierbale, W. J. Grüneberg, T. Mendes, B. De Boeck, and H. Campos. 2024. "Potato and Sweetpotato Breeding at the International Potato Center: Approaches, Outcomes and the Way Forward." *Theoretical and Applied Genetics* 137, no. 1: 1–25. https://doi.org/10.1007/s00122-023-04515-7.

Mabaya, N., D. A. Odeny, F. Kosmowski, W. de Boef, and A. Mude. 2025. "Genetic Diversity of Farmer-Preferred Sweetpotato in West Africa Using DArTseq SNP Markers." *PLoS ONE* 20, no. 4: e0312384. https://doi.org/10.1371/journal.pone.0312384.

Maquia, I., I. Muocha, A. Naico, et al. 2013. "Molecular, Morphological and Agronomic Characterization of the Sweet Potato (*Ipomoea batatas* L.) Germplasm Collection From Mozambique: Genotype Selection for Drought Prone Regions." *South African Journal of Botany* 88: 142–151. https://doi.org/10.1016/j.sajb.2013.07.008.

Molin, D., C. J. Coelho, D. S. Máximo, F. S. Ferreira, J. R. Gardingo, and R. R. Matiello. 2013. "Genetic Diversity in the Germplasm of Tropical Maize Landraces Determined Using Molecular Markers." *Genetics and Molecular Research* 12, no. 1: 99–114.

Monden, Y., and M. Tahara. 2017. "Genetic Linkage Analysis Using DNA Markers in Sweetpotato." *Breeding Science* 67: 41–51.

Nei, M., and W. H. Li. 1979. "Mathematical Model for Studying Genetic Variation in Terms of Restriction Endonucleases." *Proceedings of the National Academy of Sciences of the United States of America* 76, no. 10: 5269–5273.

Ngailo, S., H. Shimelis, J. Sibiya, B. Amelework, and K. Mtunda. 2016. "Genetic Diversity Assessment of Tanzanian Sweetpotato Genotypes Using Simple Sequence Repeat Markers." *South African Journal of Botany* 102: 40–45. https://doi.org/10.1016/j.sajb.2015.08.001.

Peakall, R., and P. E. Smouse. 2012. "GenALEx 6.5: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research—An Update." *Bioinformatics* 28, no. 19: 2537–2539.

Pritchard, Jonathan K, X Wen, and D Falush 2009. "Documentation: STRUCTURE Version 2.3." In Practice: 29.

Roullier, C., R. Kambouo, J. Paofa, D. McKey, and V. Lebot. 2013. "On the Origin of Sweetpotato (*Ipomoea batatas* [L.] Lam.) Genetic Diversity in New Guinea, A Secondary Centre of Diversity." *Heredity* 110, no. 6: 594–604. https://doi.org/10.1038/hdy.2013.1.

Sapakhova, Z., N. Raissova, D. Daurov, et al. 2023. "Sweet Potato as a Key Crop for Food Security Under the Conditions of Global Climate Change: A Review." *Plants* 12, no. 13: 1–24.

Semagn, K., C. Magorokosho, B. S. Vivek, et al. 2012. "Molecular Characterization of Diverse CIMMYT Maize Inbred Lines From Eastern and Southern Africa Using Single Nucleotide Polymorphic Markers." *BMC Genomics* 13, no. 1: 113.

Shete, S., H. Tiwari, and R. C. Elston. 2000. "On Estimating the Heterozygosity and Polymorphism Information Content Value." *Theoretical Population Biology* 57, no. 3: 265–271.

Suvi, W. T., H. Shimelis, M. Laing, I. Mathew, and A. I. T. Shayanowako. 2020. "Assessment of the Genetic Diversity and Population Structure of Rice Genotypes Using SSR Markers." *Acta Agriculturae Scandinavica Section B Soil and Plant Science* 70, no. 1: 76–86. https://doi.org/10.1080/09064710.2019.1670859.

Tseng, Y. T., H. F. Lo, and S. Y. Hwang. 2002. "Genotyping and Assessment of Genetic Relationships in Elite Polycross Breeding Cultivars of Sweet Potato in Taiwan Based on SAMPL Polymorphisms." *Botanical Bulletin of Academia Sinica* 43: 99–105.

Ulukan, H. 2009. "The Evolution of Cultivated Plant Species: Classical Plant Breeding Versus Genetic Engineering." *Plant Systematics and Evolution* 280, no. 3–4: 133–142.

van Vugt, D., and A. C. Franke. 2018. "Exploring the Yield Gap of Orange-Fleshed Sweet Potato Varieties on Smallholder Farmers' Fields in Malawi." *Field Crops Research* 221, no. 36: 245–256. https://doi.org/10.1016/j.fcr.2017.11.028.

Wright, I., and P. Adams 2012. "Changing History." *Drapers* (21-APRIL-2012): 18–20.

Yang, J., S. Mezmouk, A. Baumgarten, et al. 2017. "Incomplete Dominance of Deleterious Alleles Contributes Substantially to Trait Variation and Heterosis in Maize." *PLoS Genetics* 13, no. 9: e1007019.

Zhang, J., M. Terrones, C. R. Park, et al. 2016. "Carbon Science in 2016: Status, Challenges and Perspectives." *Carbon* 98, no. May 2018: 708–732. https://doi.org/10.1016/j.carbon.2015.11.060.

Zhang, Z., S. Schwartz, L. Wagner, and W. Miller. 2000. "A Greedy Algorithm for Aligning DNA Sequences." *Journal of Computational Biology* 7, no. 1–2: 203–214.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Table S1:** List of Genotypes Used in the Study. **Table S2:** List of Genotypes in Sub-cluisters.