Original Research Article

Genotyping by sequencing reveals genetic relatedness and duplicates amongst local cassava (*Manihot esculenta* Crantz) landraces and improved genotypes in Kenya

ABSTRACT

Future demand for cassava is expected to increase in order to mitigate climatic changes, sustain food security and provide raw materials for industry. To meet these demands, adoption of modern omics methods ensures reliability, precision and timely delivery of more productive and resilient varieties. A total of 112 mix of duplicate clones, diverse local cassava landraces (LARs) and improved genotypes (IMGs) were genotyped using single nucleotide polymorphisms (SNPs) generated through genotyping by sequencing (GBS) approach. About 17% (5808) of the 33672 SNPs were used for hierarchical clustering and *ADMIXTURE* analysis for ancestries. Approximately 48 and 52% of the germplasms respectively formed 17 independent clusters (identical clones or duplicates) and admixtures (unique or non-duplicated clones). Of the duplicates, 10 clusters were formed from LARs, four from IMGs and three from a mix of both LARs and IMGs, revealing their genetic relatedness. About 71 and 29% of clusters contained accessions from the same and different geographical regions respectively, with the geographical restriction of clusters adduced to the minimal movement of planting materials across the country, perhaps linked to either inefficient seed distribution system or disease-driven quarantine measures. Duplication of LARs was attributed to historical sharing or exchange of planting materials by farmers while duplicates of IMGs could be attributed to convergent evolution, selection, or sharing of common parentage. The high number of admixtures or unique clones implied minimal loss of genetic diversity. These findings can aid designing efficient and effective cassava improvement programs through development of a core set of diagnostic markers

Keywords: GBS, SNPs, landraces, improved genotypes, variety identification

1. INTRODUCTION

Cassava *(Manihot esculenta* Crantz) which originated around the Amazon basin [1-4] was introduced in sub-Saharan Africa (SSA) by the Portuguese traders in the 16th century [5] and in the East Africa coast in the 18th century [6]. The crop is a perennial woody shrub extensively grown in the tropical and subtropical regions of the world for its edible starchy tuberous roots, which are a major food source for developing countries [7, 8, 9]. The continuous rise in cassava popularity in Africa is attributed to the crop's low input requirement, tolerance to drought stress or low water requirement, survivability in marginal soils or soils with low nutrients, and flexible harvesting window that allows the crop to be left in the soil as a food reserve [10, 11, 12]. These make cassava a resilient crop important for food and nutritional

security in Africa, where half a billion people eat the crop daily [13, 14, 15, 16]. Despite its significance, cassava production in SSA still lags behind other parts of the world. This has largely been attributed to pests and diseases, low investments in breeding programs and inherent genetic challenges associated with the crop [58, 59]. For example genetic barriers such as high heterozygosity, inbreeding depression, allopolyploid, poor seed set, irregular flowering, and the polygenic and recessive nature of many desirable traits, constrain development of new or improved varieties especially via conventional breeding [17, 18, 19]. These are further compounded by a mixture of diverse local landraces and improved varieties that are often cultivated by most small-scale farmers on the same piece of land. Indeed, farmers often exchange stem cuttings or planting materials with their neighbors and neighboring communities, resulting in fields with a mixture of local cassava varieties [20, 21]. Commonly, this results in the same ethnic or local name being assigned to different cassava germplasms or the same germplasms assigned different local names.

Variety naming systems in the absence of formal seed systems can be quite temporally and spatially variable, leading to inconsistencies in the names of a particular variety [22]. All these hamper the selection of breeding lines. To overcome these limitations, molecular approaches can assist in reliable identification, characterization, and verification of genotypes or varieties and hasten selection of appropriate parental plants [23, 24, 25], thus improve the designing and delivery of tailored breeding objectives such as high yields [26]. Accurate identification of crop cultivars is crucial in assessing the impact of crop improvement research outputs and the two commonly used identification approaches, elicitation of variety names from farmer interviews and morphological plant descriptors, have inherent uncertainty levels [22]. The major aim of variety or cultivar identification is to catalog the crop's genetic diversity [26]. There are many reports on many landraces of cassava in SSA but with limited studies on the genetic relatedness between these landraces and elite or improved accessions [27]. Molecular marker technologies such as RFLPs, AFLPs, SSRs, DArTs, and SNPs among others have been used to detect polymorphisms and characterize genetic variation in cassava cultivars [26]. Rabbi et al. [22] successfully used SNPs derived from GBS to track and identify released cassava varieties and local landraces in Ghana, West Africa. The present study, therefore, applied the GBS approach to generate SNPs that revealed genetic relatedness amongst local landraces and improved cassava genotypes sampled from various cassava growing regions in Kenya. This is a preliminary step toward the acceleration of the cassava breeding process in the country.

2. MATERIALS AND METHODS

2.1 Sample Collection

A field survey was carried out in April 2018 in selected areas within major cassava growing regions of Nyanza, western, eastern, and coastal Kenya (Fig. 1). Systematic sampling was applied to identify cassava farmers or farms for cassava leaf collection [28]. This involved stopping at regular pre-determined intervals (~2-5 km) allowing wide coverage of the surveyed areas between farmer fields along the major motorable roads traversing each sampling location [29]. The local name of the

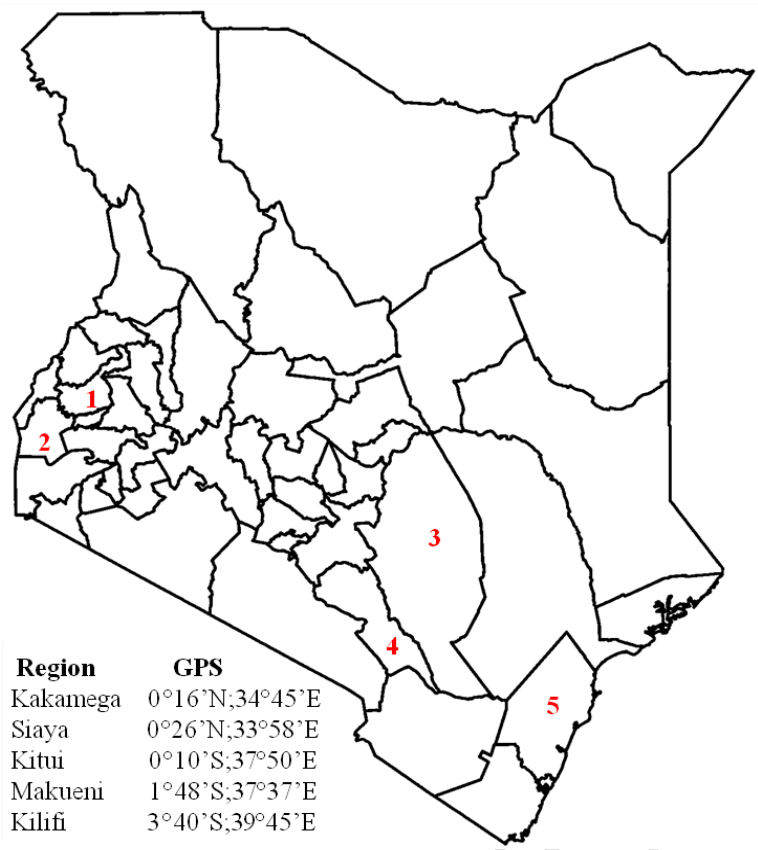
landraces and/or names of villages and GPS coordinates where the samples were collected were recorded (Table 1). Cassava leaves were harvested and pooled from five plants per landrace or genotype. The leaves were immediately transferred to falcon tubes half-filled with silica gels to preserve their integrity prior to DNA extraction.

2.2 Sequencing Cassava using DArTSeq

Cassava leaf samples were sent to Integrated Genotyping Service and Support (IGSS) platform located at the Biosciences eastern and central Africa (BecA-ILRI) Hub in Nairobi, Kenya for genotyping. DNA extraction was done using TANBead Plant extraction kit. The quality and quantity of genomic DNA were determined using NanoDrop ND-1000 (Thermo Fisher Scientific) and agarose gel electrophoresis. Libraries were constructed according to Kilian et al. [30] DArTSeq complexity reduction method through digestion of genomic DNA using a combination of *PstI* and *MseI* restriction enzymes and ligation of barcoded adapters followed by PCR amplification of adapter-ligated fragments. Libraries were sequenced using single read sequencing runs for 77 bases. Next generation sequencing was carried out using the Illumina Hiseq2500. DArTseq markers scoring was achieved using DArTsoft14 which is an in-house marker scoring pipeline based on algorithms. Two types of DArTseq markers were scored, SilicoDArT markers and SNP markers which were both scored as binary for presence /absence (1 and 0, respectively) of the restriction fragment with the marker sequence in genomic representation of the sample. Both SilicoDArT markers and SNP markers were aligned to the reference genomes of Cassava_v61 to identify chromosome positions.

2.3 Data Analysis

The quality of the SNP data was filtered using TASSEL and SNPs anchored on scaffold or missing chromosome information were discarded. TASSEL was also used to select SNPs with >0.05 minor allele frequencies (MAF) and SNPs with no more than 20% missing genotype data. For LD pruning and IBS matrix estimating, Plink 1.9 was used to select for SNP with less than 0.5 R2 LD value within each 50-SNP window size i.e. considering 50 SNPs at a time, the LD between them should be less than 0.5 LD R2. Two methods used for grouping the genotypes included hierarchical clustering using identity by state (IBS) matrix and a model-based maximum likelihood estimation of individual ancestries from multi-locus SNP genotype datasets using ADMIXTURE [22]. IBS examines if two lines are identical based on the nucleotide (SNP alleles) that they share. Using the pruned SNPs from Plink, IBS matrix was calculated with the distance function of Plink [31]. The matrix was used for hierarchical clustering using the Ward2 method for distance estimation. The critical distance threshold used to declare two genotypes are identical was 0.05 based on the empirically determined evidence suggested by Rabbi et al. [22] from the distribution of distances between duplicated DNA of 64 cassava samples. A ward's minimum variance hierarchical cluster dendrogram (Fig. 3) was then generated from IBS matrix using Analyses of Phylogenetics and Evolution (APE) package [32] implemented within *R* software (R Core Team, 2020).



#	Region	GPS
1	Kakamega	0°16'N;34°45'E
2	Siaya	0°26'N;33°58'E
3	Kitui	0°10'S;37°50'E
4	Makueni	1°48'S;37°37'E
5	Kilifi	3°40'S;39°45'E

Fig. 1. Five (5) major cassava growing regions of Kenya where leaf samples of local landraces and improved genotypes were collected. These regions represent 100% areas within Kenya where cassava is cultivated. GPS indicates the global positioning system for the coordinates of the regions.

After filtering, LD pruning and IBS matrix were used to determine the LD threshold and select SNPs accordingly. The same set of LD-pruned SNPs used for the hierarchical clustering was also used for ADMIXTURE to identify ancestries of the collected cassava germplasms [22]. The model-based clustering approach implemented in ADMIXTURE assumes linkage equilibrium among loci and Hardy-Weinberg equilibrium within ancestral populations [33, 22]. Considering a sub-population of 2 - 20, a 5-fold cross-validation procedure was used to select the optimum number of sub-populations present in the population as 14. The population structure was then modeled with the optimum number of underlying sub-population groups (Fig. 5).

Table 1: Cassava landraces and genotypes sampled during field surveys from different cassava growing regions of Kenya.

Local ID	Code	Region / GPS	Origin / Attributes	Local ID	Code	Region / GPS	Origin / Attributes	Local ID	Code	Region / GPS	Origin / Attributes
		0°16'N;34°45'E	Landrace / no information				Landrace / no information			1°48'S;37°37'	Landrace / no information
Shavirotsi	KK1			Nya-Yenga	SYA8	0°26'N;33°58'E		Kitwa_II	MK4	E	
Bwichina	KK2	"	Landrace / no information	Nya-Gang	SYA9	"	Landrace / no information	Kitwa_III	MK5	"	Landrace / no information
Lunyalala	KK3	"	Landrace / no information	Nyal-Kada	SYA10	"	Landrace / no information	Masokani_I	MK6	"	Landrace / no information
		"	Landrace / no information			"	Landrace / CMD			"	Landrace / no information
Shanina	KK4			Nya-Udai	SYA11		susceptible	Masokani_II	MK7		
		"	Landrace / no information	AdhiamboLer		"	Landrace / CMD			"	Landrace / no information
Mukulusu	KK5			a	SYA12		susceptible	Kaliluni	MK8		
		"	Landrace / no information	Nya-		"	Landrace / no information			"	Landrace / no information
Itenyi	KK6			Bungoma	SYA13			Muvila	MK9		
		"	Landrace / no information			"	Improved genotype			"	Cuba / CBSD & CMD
Shisembe	KK7			Lady Gay	SYA14			TC14	MK10		resistant
		"	Landrace / no information							"	Cuba / CBSD & CMD
Inzakula	KK8			Kiboko297	SEK1	0°10'S;37°50'E	KALRO / CBSD resistant	TC4-Katune	MK11		resistant
Shitaho	KK9	"	Landrace / no information	Thika272	SEK2	"	KALRO / CBSD resistant	99/0056	MK12	"	IITA / Improved genotype
Lugala	KK10	"	Landrace / no information	Thika273	SEK3	"	KALRO / CBSD resistant	Kalimbini_I	MK13	"	Landrace / no information
Lugusisti	KK11	"	Landrace / no information	Kiboko275	SEK4	"	KALRO / CBSD resistant	Kalimbini_II	MK14	"	Landrace / no information
Banasa	KK12	"	Landrace / no information	Kiboko274	SEK5	"	KALRO / CBSD resistant	Kalimbini_III	MK15	"	Landrace / no information
Isambe	KK13	"	Landrace / no information	Thika280	SEK6	"	KALRO / CBSD resistant	Kalimbini_IV	MK16	"	Landrace / no information
Isulu	KK14	"	Landrace / no information	Kiboko300	SEK7	"	KALRO / CBSD resistant	Katsuhanzala	MK17	"	KALRO / Improved genotype
Ikholi	KK15	"	Landrace / no information	Kiboko271	SEK8	"	KALRO / CBSD resistant	Kasukari (990127)	MK18	"	KALRO / Improved genotype
Ingotse	KK16	"	Landrace / no information	Thika279	SEK9	"	KALRO / CBSD resistant	Kitivo	MK19	"	Landrace / no information
Shikoti	KK17	"	Landrace / no information	Thika289	SEK10	"	KALRO / CBSD resistant	Kimutwa	MK20	"	Landrace / no information
Shipalo	KK18	"	Landrace / no information	Kiboko295	SEK11	"	KALRO / CBSD resistant	Mumbuni	MK21	"	Landrace / no information
		"	Landrace / no information			"	KALRO / CBSD resistant			3°40'S;39°45'	Landrace / no information
Shamiloli	KK19			Kiboko277	SEK12			Halu	KF1	E	
Madioli	KK20	"	Landrace / no information	Kiboko276	SEK13	"	KALRO / CBSD resistant	Kibandameno	KF2	"	Landrace / CMD susceptible
Shiswa	KK21	"	Landrace / no information	Thika278	SEK14	"	KALRO / CBSD resistant	Agriculture	KF3	"	IITA / improved genotype
MM96/1871	KK22	"	IITA / CMD resistant	Kiboko281	SEK15	"	KALRO / CBSD resistant	Tajirika/KME-0802	KF4	"	Landrace / CMD resistant
MM97/0293	KK23	"	KALRO / CMD resistant	Thika5	SEK16	"	Landrace / CMD resistant	Kaleso	KF5	"	Landrace / CMD resistant
		"	Landrace / CBSD	Serere	SEK17	"	CIAT / CBSD susceptible			"	Landrace / no information
Magana	KK24		resistant					Soyosoyo	KF6		
CK9	KK25	"	Landrace / no information	Kiboko9	SEK18	"	KALRO / CBSD resistant	Sokoke_I	KF7	"	Landrace / no information
		"	Landrace / CMD	Kiboko10	SEK19	"	KALRO / CBSD resistant			"	Landrace / no information
Matuja	KK26		susceptible					Sokoke_II	KF8		
Fumbachai	KK27	"	Landrace / no information	Kiboko11	SEK20	"	KALRO / CBSD resistant	Kakanjuni_I	KF9	"	Landrace / no information
MM98/1313-		"	KALRO / Improved	Kiboko159	SEK21	"	KALRO / CBSD resistant			"	Landrace / no information
HS	KK28							Kakanjuni_II	KF10		
MH95/0183	KK29	"	IITA / CMD resistant	Kiboko257	SEK22	"	KALRO / CBSD resistant	Kakanjuni_III	KF11	"	Landrace / no information
MM08/2206	KK30	"	IITA / Improved genotype	Kiboko258	SEK23	"	KALRO / CBSD resistant	Mkongo_I	KF12	"	Landrace / no information
MM96/0686	KK31	"	KALRO / CMD resistant	Kiboko259	SEK24	"	KALRO / CBSD resistant	Mkongo_II	KF13	"	Landrace / no information

Aruaro	SYA1	0°26'N;33°58'E	Landrace / no information	Kiboko267	SEK25	"	KALRO / CBSD resistant	Cha-Vyango_I	KF14	"	Landrace / no information
Othigo-Diep	SYA2	"	Landrace / no information	Kiboko268	SEK26	"	KALRO / CBSD resistant	Cha-Vyango_II	KF15	"	Landrace / no information
Nyakatanegi_I	SYA3	"	Landrace / no information	Kiboko269	SEK27	"	KALRO / CBSD resistant	Chumani	KF16	"	Landrace / no information
Nyakatanegi_II	SYA4	"	Landrace / no information	Kiboko270	SEK28	"	KALRO / CBSD resistant	Matano-Manne	KF17	"	Landrace / no information
Nya-Uyoma	SYA5	"	Landrace / no information	Kasioni	MK1	1°48'S;37°37'E	Landrace / no information	KALRO	KF18	"	KALRO / Improved genotype
Kamis	SYA6	"	Landrace / CMD susceptible	Kisimba	MK2	"	Landrace / no information				
Nya-Uganda	SYA7	"	Landrace / CMD susceptible	Kitwa_I	MK3	"	Landrace / no information				

CMD = cassava mosaic disease; CBSD = cassava brown streak disease; KALRO = Kenya Agricultural & Livestock Research Organization; IITA = International Institute of Tropical Agriculture; CIAT = International center for tropical agriculture; KG = Kakamega (0°16'N;34°45'E) SYA = Siaya (0°26'N;33°58'E);SEK = SEKU / Kitui (0°10'S;37°50'E); MK = Makueni (1°48'S;37°37'E); KF = Kilifi (3°40'S;39°45'E). Information on germplasm attributes were sourced from several literature reviews

3. RESULTS

3.1 Cassava Germplasms

Out of 112 cassava germplasms collected from five cassava growing regions (Fig. 1), 71 (~63%) were local landraces and 41 (~37%) were improved genotypes (Fig. 2). Distribution showed more landraces were cultivated in all regions except in Kitui where more improved genotypes were collected (Fig. 2). Traits or characteristics of most landraces had not been documented compared to improved genotypes that were developed for resistance or tolerance against two (CMD & CBSD) major virus diseases (Table 1). However, farmers casually interviewed during sampling attributed their preferences to local landraces for sweet or bitter tubers, early maturity, and high yield (data not shown). Improved genotypes were introduced into these regions by research institutions such as International Center for Tropical Agriculture (CIAT), International Institute of Tropical Agriculture (IITA) and Kenya Agricultural and Livestock Research Organization (KALRO) (Table 1).

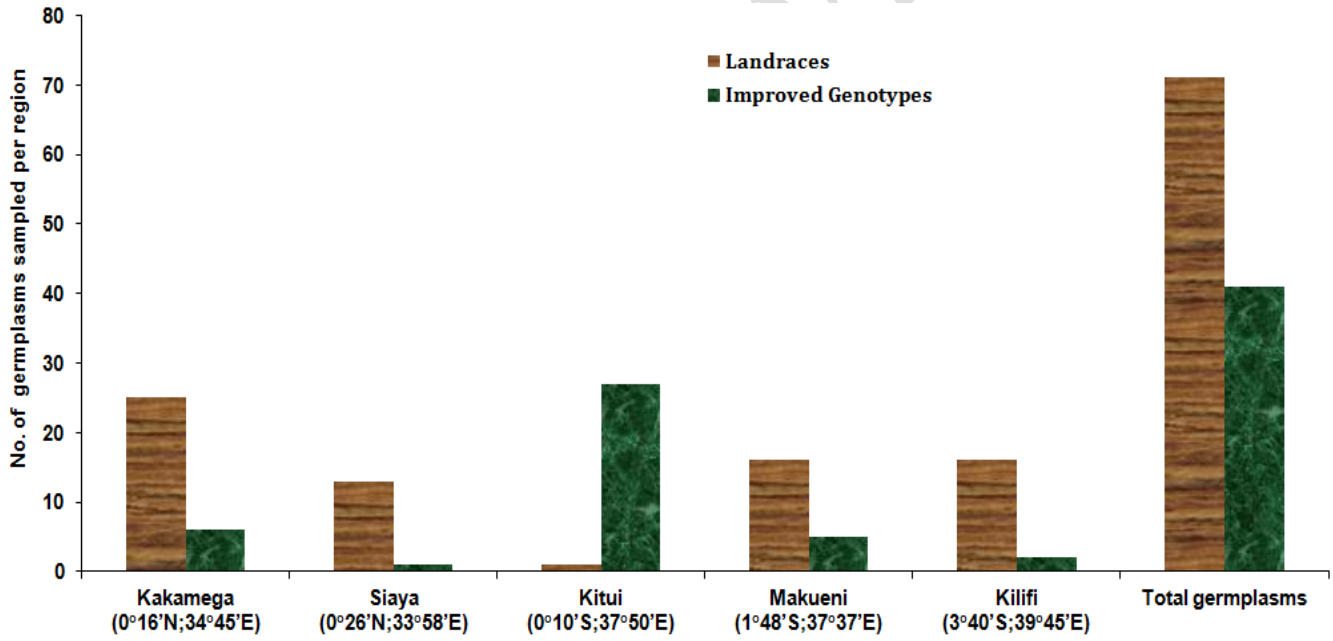


Fig. 2. Distribution of local cassava landraces and improved genotypes sampled across different cassava growing regions of Kenya. The two major germplasm (landraces and improved genotypes) were not uniformly cultivated in terms of numbers. For examples regions had more improved genotypes compared to landraces and vice versa.

1.2 Filtering and Selection of SNPs and Optimum Population Identification

A total of 33672 SNPs was identified. Out of this, 29614 SNPs (~88%) were anchored to chromosomes, 942 (~3%) were present in scaffolds, while the remaining 3116 SNPs (~9%) could not be mapped to any chromosome or scaffold. After quality filtering, 20846 SNPs were selected. LD pruning and IBS matrix estimation revealed that 5808 SNPs met the selected LD threshold criteria (Table 2). The 5-fold cross-validation procedure revealed the number of optimum populations to be 14 (Fig. 4).

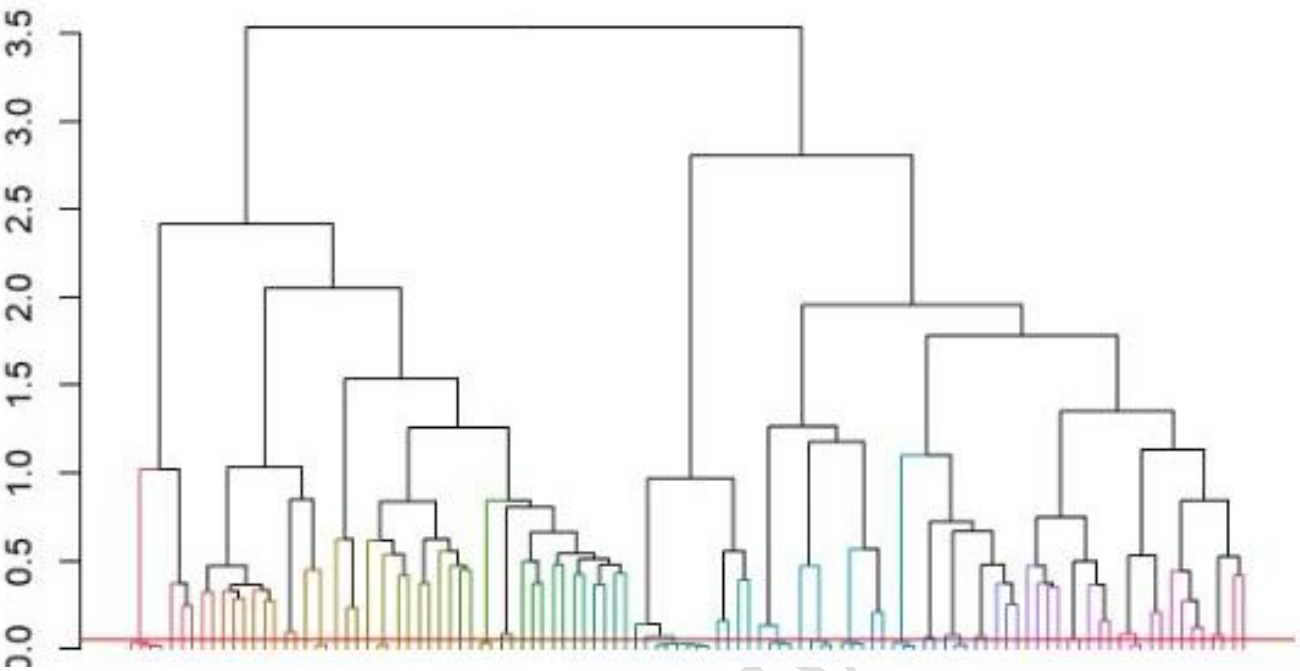


Fig. 3. Hierarchical clustering dendrogram from identity by state (IBS) matrix estimation. The Red line represents the empirically determined distance threshold

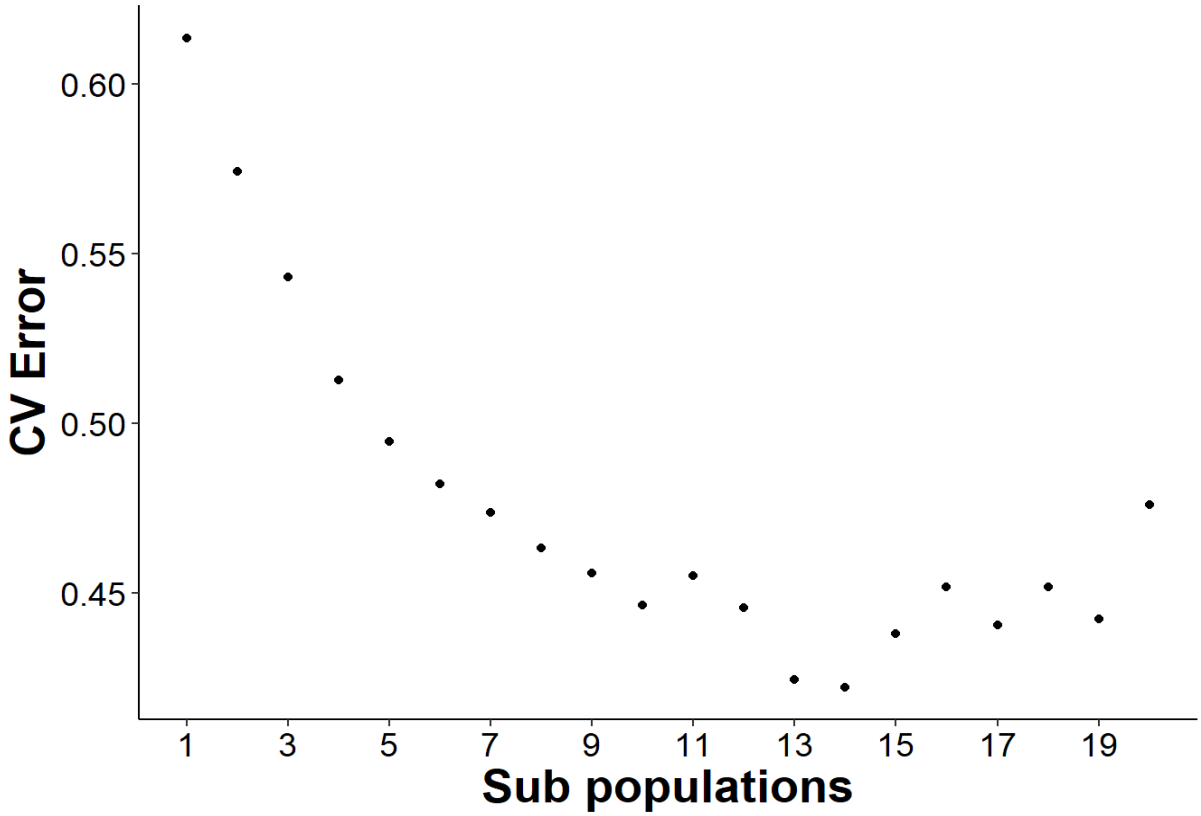


Fig. 4. Determination of optimal number of sub-population present in the population based on ADMIXTURE**.** Considering a sub-population of 2 - 20, a 5-fold cross-validation procedure was used to select the optimum number of sub-population present in the population as 14 as shown in the graph below (Fig. 5)

3.3 Admixture Analysis

Genetic relationships among genotyped cassava germplasms are shown on hierarchical clustering dendrogram (Fig. 3) while population structure depicting ancestries from admixture presented as a barplot (Fig. 5). The admixture clustering together with dendrogram topology enabled identification of clusters of genetically identical germplasms containing only landraces, only improved genotypes as well as clusters containing both landraces and improved genotypes (Table 3). A total of 54 germplasms (~48%) were grouped into 17 independent clusters (I - XVII) as identical clones or single pure lines (Table 3). They represented duplicated clones bearing different local names. Out of 17 clusters, 10 contained only landraces; four had only improved genotypes and the remaining three clusters had accessions from landraces and improved genotypes (Fig. 6). Of the 10 landrace clusters, cluster IX was the largest with eight accessions, followed by cluster XIV with five 5 accessions, cluster I and X each with four accessions, four clusters (XVII, XVI, XII, and XI) each with three accessions and two clusters (XV & VII) with two accessions each (Fig. 6). All the four clusters that contained only improved genotypes (VI, IV, III & II) had two accessions each while three clusters containing both landraces and improved genotypes (V, VIII & XIII) had three accessions each (Fig. 6).

Table 2: The distribution of the SNPs across the cassava genome

##	Chromosome	No_of_SNPs
## 1	01	495
## 2	02	431
## 3	03	396
## 4	04	367
## 5	05	335
## 6	06	416
## 7	07	254
## 8	08	307
## 9	09	258
## 10	10	363
## 11	11	392
## 12	12	262
## 13	13	215
## 14	14	315
## 15	15	350
## 16	16	249
## 17	17	199
## 18	18	204
Total		**5,808**

Geographically, majority of the clusters (12 of the 17 or ~71%) contained accessions sampled from the same region (Table 3). These included clusters II, III, IV, V, VI, VII, IX, XI, XIII, XIV, XV, and XVI. The remaining five of the 17 (~29%) clusters (I, VIII, X, XII & XVII) had accessions sampled from different regions (Table 3). For instance cluster I, VIII and XII were from regions in closer proximity (Siaya = 0°26'N, 33°58'E, and Kakamega = 0°16'N, 34°45'E) while cluster X (Kitui = 0°10'S, 37°50'E, and Kakamega = 0°16'N, 34°45'E) and XVII (Makueni = 1°48'S, 37°37'E, and Kakamega = 0°16'N, 34°45'E) represented clustering of accessions from far regions (Table 3). Landraces from Kilifi (3°40'S, 39°45'E) located in coastal Kenya did not cluster with landraces or improved genotypes from other regions (see cluster XIII and XIV) (Table

3). Compared to other regions, Kitui (0°10'S, 37°50'E) had a majority (5) of different clusters (II, III, IV, V & VI).

The remaining 58 germplasms (~52%) were classified as admixtures and thus unique or non-duplicated clones as they did not cluster (Table 3, Fig. 6). Under this category, 31 accessions (~53%) were landraces and 27 (~47%) were improved genotypes (Fig. 6). In terms of known traits (from literature reviews), clusters containing either improved genotypes alone or a mix of improved genotypes with local landraces were described as resistant or tolerant to two major virus diseases i.e. cassava mosaic disease (CMD) and cassava brown streak disease (CBSD) compared to the majority of landrace-based clusters with no information available on their known traits (Table 1). Only clusters I and X (all landraces) had CMD and CBSD susceptible accessions. In summary, the majority of landraces clustered as identical clones or accessions compared to improved genotypes while regionally, most clusters contained accessions sampled within the same region. The unique or non clustered accessions (58) plus clustered or duplicates (17) reduced the total accessions surveyed to 73 from 112 that were originally genotyped.

Table 3: Classification of cassava accessions into clusters based on underlying sub-population groups derived from Figure 5

Local ID	Type	Class	Cluster #	Region / GPS	Local ID	Type	Class	Cluster #	Region / GPS
				0°16'N;34°45'E			Uniqu	NCL	0°16'N;34°45'E
Matuja	LAR	All	I		Shavirotsi	LAR	e		
Othigo-Diep	"	Identical		0°26'N;33°58'E	Bwichina	LAR	"	"	"
Aruaro	"			"	Lunyalala	LAR	"	"	"
Nya-Udai	"			"	Mukulusu	LAR	"	"	"
Kiboko276	IMG	All	II	0°10'S;37°50'E	Shisembe	LAR	"	"	"
Kiboko297	"	Identical		"	Shitaho	LAR	"	"	"
Kiboko274	IMG	All	III	0°10'S;37°50'E	Lugusisti	LAR	"	"	"
Thika278	"	Identical		"	Banasa	LAR	"	"	"
Kiboko271	IMG	All	IV	0°10'S;37°50'E	Ingotse	LAR	"	"	"
Thika289	"	Identical		"	Shiswa	LAR	"	"	"
Kiboko300	IMG	All	V	0°10'S;37°50'E	MM96/1871	IMG	"	"	"
Thika273	"	Identical		"	MM97/0293	IMG	"	"	"
Thika5	LAR			"	Magana	LAR	"	"	"
Kiboko281	IMG	All	VI	0°10'S;37°50'E	CK9	LAR	"	"	"
Thika280	"	Identical		"	MM98/1313-HS	IMG	"	"	"
Itenyi	LAR	All	VII	0°16'N;34°45'E	MM08/2206	IMG	"	"	"
Inzakula	"	Identical		"	MM96/0686	IMG	"	"	"
Lady Gay	LAR	All	VIII	0°26'N;33°58'E	Nyakatanegi-II	LAR	"	"	0°26'N;33°58'E
Shanina	"	Identical		0°16'N;34°45'E	Nya-Uyoma	LAR	"	"	"
MH95/0183	IMG			"	Kamis	LAR	"	"	"
Kalimbini-I	LAR			1°48'S;37°37'E	Nya-Uganda	LAR	"	"	"
Kasioni	"			"	AdhiamboLera	LAR	"	"	"
Kitwa-II	"			"	Nya-Bungoma	LAR	"	"	"
Kitwa-III	"	All		"	Thika272	IMG	"	"	0°10'S;37°50'E
Kitivo	"	Identical	IX	"	Kiboko275	IMG	"	"	"
Kitwa-I	"			"	Thika279	IMG	"	"	"
Kimutwa	"			"	Kiboko295	IMG	"	"	"
Mumbuni	"			"	Kiboko277	IMG	"	"	"
Serere	LAR			0°10'S;37°50'E	Kiboko9	IMG	"	"	"
Madioli	"	All		0°16'N;34°45'E	Kiboko10	IMG	"	"	"
Shikoti	"	Identical	X	"	Kiboko11	IMG	"	"	"
Ikholi	"			"	Kiboko159	IMG	"	"	"
Lugala	LAR	All	XI	0°16'N;34°45'E	Kiboko257	IMG	"	"	"
Shamiloli	"	Identical		"	Kiboko258	IMG	"	"	"
Shipalo	"			"	Kiboko259	IMG	"	"	"
Fumbachai	LAR	All	XII	0°16'N;34°45'E	Kiboko267	IMG	"	"	"
Isambe	"	Identical		"	Kiboko268	IMG	"	"	"
Nyal-Kada	"			0°26'N;33°58'E	Kiboko269	IMG	"	"	"
KALRO				3°40'S;39°45'E	Kiboko270	IMG	"	"	"
Matano-	IMG	All	XIII	"	Masokani-I	LAR	"	"	1°48'S;37°37'E
Manne	LAR	Identical		"		LAR	"	"	"
Kakanjuni-II	"				Masokani-II				

Tajirika				3°40'S;39°45'E	Muvila	IMG	"	"	"
Kaleso	LAR			"	TC14	IMG	"	"	"
Cha-Vyango-II	"	All Identical	XIV	"	TC4-Katune	IMG	"	"	"
Sokoke-I	"			"	99/0056	IMG	"	"	"
Chumani	"					LAR	"	"	"
Kalimbini-III		All	XV	1°48'S;37°37'E	Kalimbini-II Katsuhanzala	IMG	"	"	"
Kalimbini-IV	LAR	Identical		"	Kasukari (99/0127)	IMG	"	"	"
Nya-Gang	LAR			0°26'N;33°58'E	Halu	LAR	"	"	3°40'S;39°45'E
Nya-Yenga	"	All	XVI	"	Kibandameno	LAR	"	"	"
Nyakatanegi-I	"	Identical		"	Agriculture	LAR	"	"	"
Kaliluni	LAR			1°48'S;37°37'E	Soyosoyo	LAR	"	"	"
Kisimba	"	All	XVII	"	Sokoke-II	LAR	"	"	"
Isulu	"	Identical		0°16'N;34°45'E	Kakanjuni-I	LAR	"	"	"
					Kakanjuni-III	LAR	"	"	"
					Mkongo-I	LAR	"	"	"
					Mkongo-II	LAR	"	"	"
					Cha-Vyango-I	LAR	"	"	"

LAR = Landrace; IMG = Improved Genotype; Unique = non-duplicated clone; NCL = Non-clustered landraces / improved genotypes

2. DISCUSSION

Most of the sampled materials (approximately 63%) were local landraces compared to improved cassava genotypes that constituted 37%. This implied cultivation of more local cassava varieties or landraces which have been attributed to farmer preferred characteristics such as culinary attributes and cooking quality, sweet or bitter tastes, early maturity, pests and disease resistance, high yield, root storability in the ground, drought tolerance among other traits [21, 34, 35]. Farmers often hold several generations of knowledge concerning the attributes of landraces and sometimes have specific reasons why they retain particular cultivars [36]. On the reverse, the results implied minimal adoption and cultivation of the improved varieties in Kenya, a potential drawback for the management of cassava diseases as most of the improved genotypes had been bred and introduced for resistance or tolerance to CMD and CBSD. This was corroborated by earlier studies on low dissemination, adoption, and production of improved cassava varieties in Africa, a situation that was linked to lack of involvement of farmers and end-users in designing, planning, and execution of breeding strategies and objectives [21, 37, 38, 39]. Farmer preferences and varietal attributes influence the adoption of new cassava varieties [40, 41, 42, 43]. It is however noted that farmer preferences or attributes of the genotyped landraces and improved varieties were not assessed in the present study.

The SNPs marker data generated using GBS was successfully used to determine genetic relatedness among sampled cassava germplasms. From a total of 33672 SNPs identified, 5808 SNPs (~17%) obtained after LD pruning and IBS matrix estimation were used for hierarchical clustering and *ADMIXTURE* analysis to identify ancestries. This enabled the identification of germplasms that clustered together as well as unique or non-duplicated germplasms. Thus, a large number of SNPs may not be needed to achieve accurate identification of cassava varieties, whether in farmers' fields or formal germplasms collections [22, 26; 44]. A further study could be initiated to identify these SNPs and design KASP markers for varietal identification.

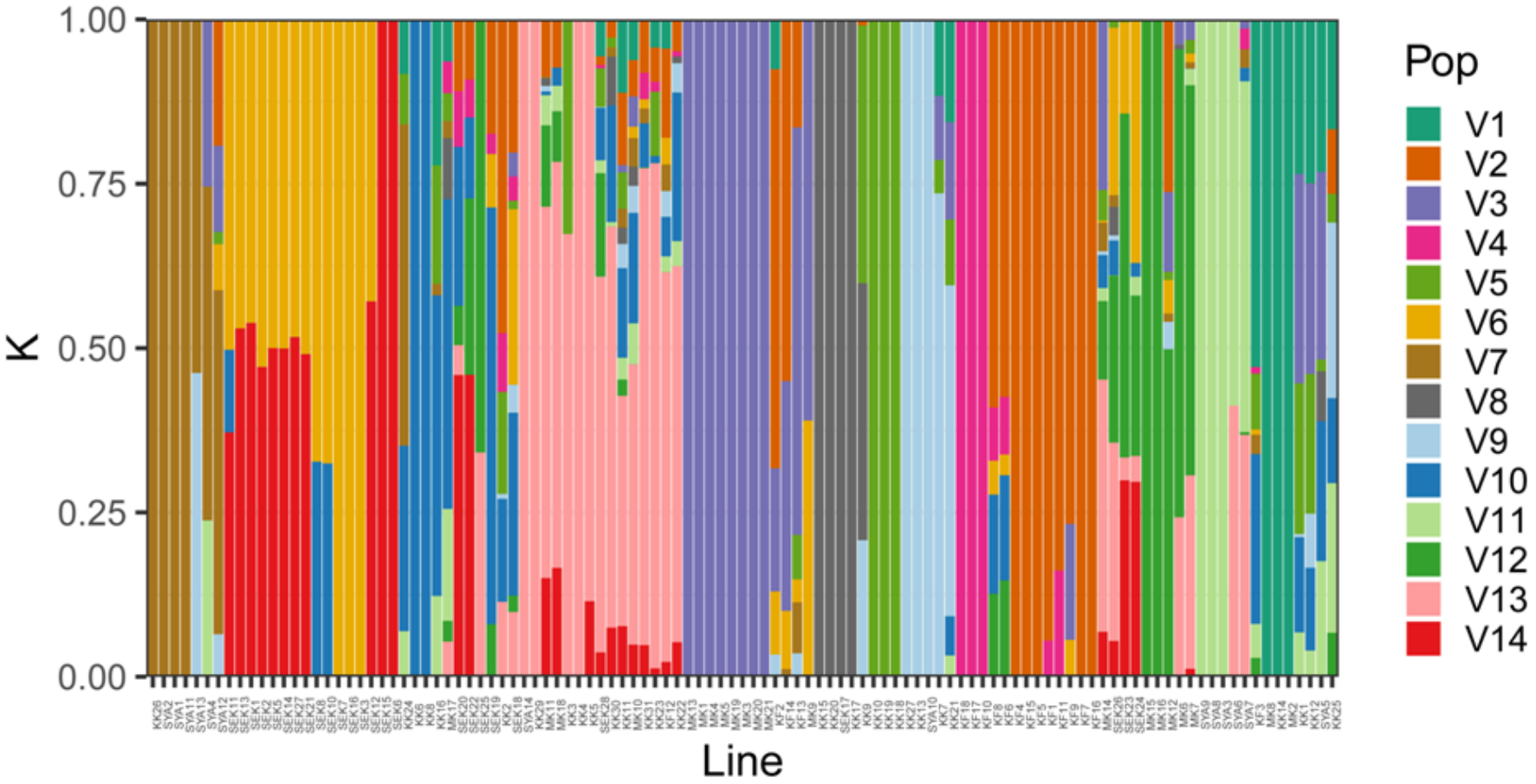


Fig. 5. Barplot showing population structure modeled with 14 underlying sub-population groups from *ADMIXTURE.* The sample order of the hierarchical clustering was maintained for the *ADMIXTURE* plot for easy comparison of the out from the

two grouping method. For the ADMIXTURE plot, the different colors represent the different sub-population while each bar represents each individual sample. Samples with just one color are pure lines from a single sub-population. Samples with more than one color are admixture from different sub-populations.

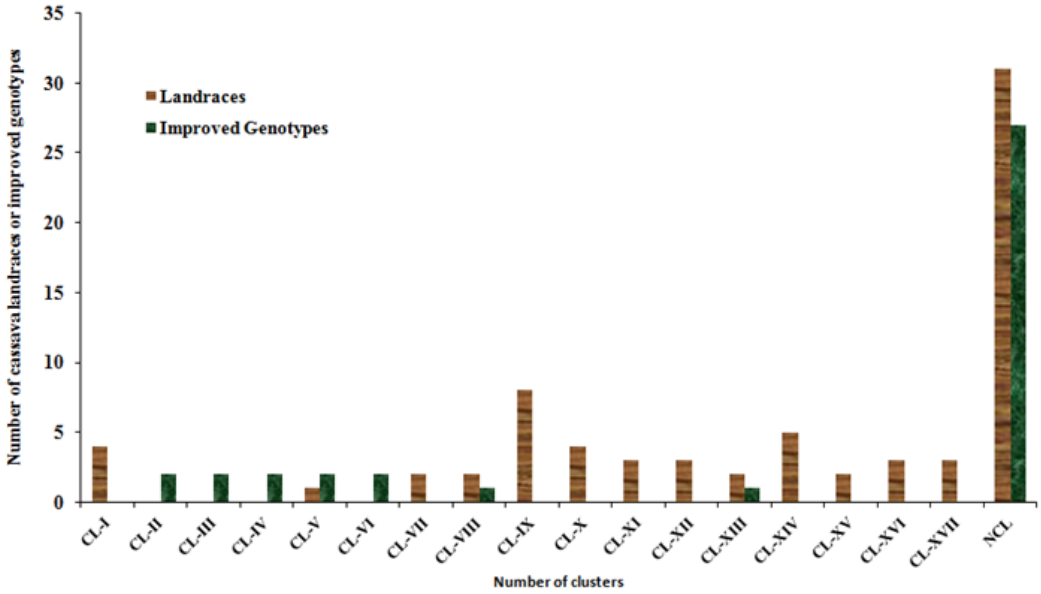


Fig. 6. Number and type of cassava accessions (local landrace & improved genotypes) grouped in each cluster. The data used to generate this figure were derived from Table 3. CL = cluster; NCL = non clustered / unique accessions.

Knowledge of the existence of duplicates in the field is important during the collection of variability and evaluation and selection of parents for cassava improvement or breeding purposes. Similarly, genomic or SNPs markers have been used to confirm that particular cassava accessions are not identical, and others are possible duplicates [44, 45]. They have also been used to track local landraces and assess the adoption of improved varieties [22, 27, 46]. Accurate identification of crop cultivars is crucial in assessing the impact of crop improvement research outputs [22]. Generally, the genomic approach contributes to further characterization of cassava genetic resources, an important step in improving cassava production in Kenya.

Further results from the present study showed that the majority of the duplicated clones were local landraces while geographically; most of the duplicated accessions were sampled either from the same region or from different regions of closer proximity. These redundancies were previously attributed to the historical sharing of cassava accessions or the same germplasms exchange between farmers with different genotype names [47]. Farmers often exchange planting materials with their neighbors or different neighboring communities, resulting in fields with a mixture of local cassava varieties [20, 21]. Thus the same ethnic or local name could be assigned to different cassava germplasms or the same germplasms assigned different local names. Variety naming systems in the absence of formal seed systems can be quite temporally and spatially variable leading to inconsistencies in the names of a particular variety [22]. The informal farmer-farmer seed distribution system is often inefficient, denying farmers in far flung areas access or a share of alternative planting materials.

Ferguson et al. [36] reported that individual cassava landraces were not widely distributed across Tanzania with limited farmer-to-farmer diffusion with implications for seed systems. Indeed, smallholder farmers recycle stem cuttings of traditional landrace cultivars [48] and there is a flow of seed within and outside the villages, with little introduction of new cultivars [49]. The absence of an effective seed distribution system [50] has limited farmers' access to planting materials from improved genotypes**.** Additionally, elicitation of cassava variety names from farmer interviews during surveys and/or use of morphological plant descriptors have had inherent uncertainty levels [22]. Morphological descriptors are also greatly influenced by the environment and show continuous variation and high plasticity, with most of them only scorable at maturity [51]. Restrictions of clusters to the same geographical areas where accessions were sampled could also be attributed to quarantine measures that restricted the movement of planting materials in order to stop the spread of virus diseases such as CMD and CBSD.

Similarities in cassava accessions can also arise due to convergent evolution, selection, or sharing of common parentage [51]. This was probably the case in Kitui region where the majority of duplicates were improved genotypes that had shared the same parents during their breeding for resistance to cassava brown streak disease [52]. Crops gradually lose their genetic variability through domestication and breeding, resulting in more uniform cultivars and reducing their recombination rates [53]. This could perhaps be used to explain clusters that included both improved genotypes and local landraces. It is however noted that no recent evidence has shown loss of genetic variation from genetic drift during the introduction of cassava to Africa [54]. The relatively low levels of diversity reported in the previous study were only observed in IITA breeders' germplasms and may represent rather a genetic bottleneck [54]. For future breeding programs involving hybridization or selection, de Oliveira et al. [55] recommended the introduction of new genetic variability into commercial cultivars to avoid low genetic variation and to improve the quality of cassava roots. The unique or non-duplicated landraces and improved genotypes in the present study represented a more expanded cassava genetic pool from which variability can be derived for future breeding purposes. It might also be important to build the core collection of the 73 unique genotypes studied in this study for further efficient conservation and cassava breeding. High genetic diversity drives better crop adaptation to emerging environmental cues.

5. CONCLUSION

In conclusion, molecular markers have an important role to play as farmers frequently give different names to the same cultivar or landraces, making identification difficult, particularly as cassava varieties are not easy to distinguish morphologically [45]. This enables the correct assessment of adoption rates, which in turn, influences breeding priorities and agricultural policies [56]. Knowledge on the extent of genetic diversity among cassava landraces and improved genotypes in Kenya using GBS-derived SNP markers may promote their conservation and/or efficient selection and utilization as parental lines for breeding for biotic and abiotic stress tolerance. Although local landraces may be low-yielding, they may have high

genetic diversity that could promote gene flow through hybridization [27], enabling crop improvement and adaptability of species to changing climatic conditions, new pests, and diseases [57].

REFERENCES

1. Léotard G, Duputié A, Kjellberg F, Douzery E, Debain C. Phylogeography and the origin of cassava: New insights from the northern rim of the Amazonian basin. Molecular Phylogenetics Evolution, 2009; 53(1):329-334
2. Hirst KK. The History and Domestication of Cassava." ThoughtCo, 2020, thoughtco.com/cassava-manioc-domestication-170321
3. Olsen KM. SNPs, SSRs and inferences on cassava's origin. Plant Molecular Biology, 2004; 56:517–526
4. Olsen KM, Schaal B. Evidence on the origin of cassava: Phylogeography of cassava Manihot esculenta Crantz. PNA*S,* 1999; 96:5586–5591.
5. Spencer D, Ezedinma C. Cassava cultivation in sub-Saharan Africa, 2017; https://doi.org/10.19103/AS.2016.0014.06
6. Were HK, Winter S, Maiss E. Viruses infecting cassava in Kenya. Plant Disease, 2004; 88:17-22
7. Clifton P, Keogh J. Starch: Encyclopedia of Food and Health, 2016; 146-151
8. Liu Q, Liu J, Zhang P, He S. Root and tuber crops, in: Encyclopedia of Agric. & Food Syst.2014; https://doi.org/10.1016/B978-0-444-52512-3.00151-0
9. Kuete V. Physical, Hematological and Histopathological Signs of Toxicity Induced by African Medicinal Plants. Toxicological Survey of African Medicinal Plants 2014; 635 – 657
10. Orek C, Gruissem W, Ferguson M, Vanderschuren H. Morpho-physiological and molecular evaluation of drought tolerance in cassava (*Manihot esculenta* Crantz). Field Crops Research, 2020; 255. http://doi.org/10.1016/j.fcr.2020.107861
11. Rabbi IY, Kayondo SI, Bauchet G, Yusuf M, Aghogho CI. Genome-wide association analysis reveals new insights into the genetic architecture of defensive, agro-morphological and quality-related traits in cassava. Plant Molecular Biology 2022; 109:195-213 https://doi.org/10.1007/s11103-020-01038-3
12. Shigaki T. Cassava: The Nature and Uses. Encyclopedia of Food and Health, 2016; 687-693
13. Amelework AB, Bairu MW, Obakeng M, Venter SL, Laing M. Adoption and promotion of resilient crops for climate risk mitigation and import substitution: a case analysis of cassava for South African agriculture. Frontiers in Sustainable Food Systems, 2021; 5:105
14. Tize I, Fotso AK, Nukenine EN, Masso C, Ngome FA, Suh C, et al. New cassava germplasm for food and nutritional security in Central Africa. Scientific Reports 2021; 11:7394
15. Agre AP, Bhattacharjee R, Rabbi IY, Alaba OA, Unachukwu NN et al. Classification of elite cassava varieties (*Manihot esculenta* Crantz) cultivated in Benin Republic using farmers' knowledge, morphological traits and simple sequence repeat markers. Genetic Resources and Crop Evolution 2018; 65:513–525

16. CIAT. Cassava: 2019 ; https://ciat.cgiar.org/what-we-do/breeding-better-crops/rooting-for-cassava/
17. Elegba W, McCallum EJ, Wilhelm G, Vanderschuren H. Genetic transformation and regeneration of a farmer-preferred cassava cultivar from Ghana. Frontiers in Plant Science 2021; 12:909
18. Ceballos H, Rojanaridpiched C, Phumichai C, Becerra LA, Kittipadakul P, et al. Excellence in Cassava Breeding: Perspectives for the Future. Crop Breeding Genetics Genome, 2020; 2:e200008. https://doi.org/10.20900/cbgg20200008
19. Makwarela M, Rey EMC, 2006. Cassava Biotechnology, a Southern African Perspective. *Biotechnology and Molecular Biology Review* 1(1):2-11
20. Andersson MS, de Vicente MC. Cassava, manioc, yuca", Chapter 6, in: Andersson, M.S. and M.C. de Vicente (eds.), Gene Flow Between Crops and Their Wild Relatives, Johns Hopkins University Press, Baltimore, Maryland, 2010; 125-146.
21. Nakabonge G, Samukoya C, Baguma Y. Local varieties of cassava: conservation, cultivation and use in Uganda. Environmental Development Sustainability 2018; 20: 2427–2445.
22. Rabbi IY, Kulakow PA, Manu-Aduening JA, Dankyi AA, Asibuo JY, et al. Tracking crop varieties using genotyping-by-sequencing markers: a case study using cassava (Manihot esculenta Crantz). BMC Genetics 2015; 6:115 https://doi.org/10.1186/s12863-015-0273-1
23. OECD. Cassava (*Manihot esculenta* Crantz), in Safety Assessment of Transgenic Organisms in the Environment, 6: OECD Consensus Documents, OECD Publishing and Paris.2016; https://doi.org/10.1787/9789264253421-6-en
24. Otti G, Fakoya A, Andrew I, Gedil M. Development of genomic tools for verification of hybrids and selfed progenies in cassava (*Manihot esculenta* Crantz). African Journal of Biotechnology 2011; 10(76): 17400-17408
25. Lebot V. Tropical Root and Tuber Crops: Cassava, Sweet Potato, Yams, and Aroids, Crop Production Science in Horticulture Series, 2009; Vol. 17, CABI, Wallingford, United Kingdom.
26. Lopez-Lavalle LAB, Bohorquez-Chaux A, Zhang X. Identification of Cassava Varieties in Ex-Situ Collections and Global Farmer's Fields: An Update from 1990 to 2020. IntechOpen, 2021; 1-30
27. Turyagyenda LF, Kizito EB, Ferguson ME, Baguma Y, Harvey JW. Genetic diversity among farmer-preferred cassava landraces in Uganda. African Crop Science Journal 2012; 20:15–30.
28. Koima IN, Orek CO, Nguluu SN. Distribution of Cassava Mosaic and Cassava Brown Streak Diseases in agro-ecological zones of lower Eastern Kenya. IJISRT, 2018; 3(1):391–399.
29. Mware B, Narla R, Amata R, Olubayo F, Songa J, et al. Efficiency of cassava brown streak virus transmission by two whitefly species in coastal Kenya. Journal of General Molecular Virology 2009; 1:40-45.
30. Kilian A, Wenzl P, Huttner E, Carling J, Xia L, et al. Diversity arrays technology: a generic genome profiling technology on open platforms. Methods Molecular Biology 2012; 88: 67–89.
31. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics 2007; 81:559–575.

https://doi.org/10.1086/519795
32. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 2004; 20(2):289–90.
33. Frichot E, Mathieu F, Trouillon T, Bouchard G, Francois O. Fast and efficient estimation of individual ancestry coefficients. *Genetics,* 2014; 196(4):973–83
34. Bentley J, Olanrewaju A, Madu T, Olaosebikan O, Abdoulaye T, et al. Cassava farmers' preferences for varieties and seed dissemination system in Nigeria: Gender and regional perspectives. IITA Monograph, IITA, Ibadan, Nigeria. 2017; https://cgspace.cgiar.org/handle/10568/80554
35. Woyengo WV. Cassava breeding through complementary conventional and participatory approaches in Western Kenya. 2011; PhD thesis, University of KwaZulu-Natal, South Africa
36. Ferguson ME, Tumwegamire S, Chidzanga C, Shah T, Mtunda K, et al. Collection, genotyping and virus elimination of cassava landraces from Tanzania and documentation of farmer knowledge. *PLoS ONE* 2021; 16(8): e0255326. https://doi.org/10.1371/journal.pone.0255326
37. Bechoff A, Tomlins K, Fliedel G, Becerra Lopez-lavalle LA, Westby A, et al. Cassava traits and end-user preference: Relating traits to consumer liking, sensory perception and genetics. Critical Review in Food Science Nutrition 2018; 58:547–567.
38. Woyengo WV, Melis R, Shanahan P, Odongo OM. Participatory evaluation methods of cassava varieties preferred in mild-altitude tropical climate conditions of western Kenya. African Journal of Agricultural Research 2014; 9(17):1326-1333.
39. Kamau J, Melis R, Laing M, Derera J, Shanahan P, Ngugi ECK. Farmers' participatory selection for early bulking cassava genotypes in semi-arid Eastern Kenya. Journal of Plant Breeding and Crop Science 2011; 3(3):44-52
40. Okuku IO, Nyikal RA, Otieno DJ. An assessment of the effect of varietal attributes on the adoption of improved cassava in Homa Bay County, Kenya. 2018; MSc. Thesis, University of Nairobi.
41. Nduwumuremyi A, Melis R, Shanahan P, Asiimwe T. Participatory appraisal of preferred traits, production constraints & postharvest challenges for cassava farmers in Rwanda. Food Security 2016; 8:375–388 https://doi.org/10.1007/s12571-016-0556-z
42. Kamau J, Melis R, Laing M, Derera J, Shanahan P. Farmers' Perceptions of Production Constraints and Preferences in Cassava Grown in Semi-Arid Areas of Kenya. International Journal of Current Microbiology Applied Science 2016; 5(3): 844-859.
43. Khonje M, Mkandawire P, Manda J, Alene DA. Analysis of adoption and impacts of improved cassava varieties in Zambia. International Conference of Agricultural Economics, 2015; 1 – 28
44. Ferguson ME, Hearne SJ, Close TJ, Wanamaker S, Moskal WA, et al. Identification, validation and high-throughput genotyping of transcribed gene SNPs in cassava. Theoretical and Applied Genetics 2012; 124:685–695 https://doi.org/10.1007/s00122-011-1739-9
45. Mbanjo EGN, Rabbi IY, Ferguson ME, Kayondo SI, Hwa EN, et al. 2021. Technological Innovations for Improving Cassava Production in Sub-Saharan Africa. *Frontiers in Genetics* 11:1829 https://doi.org/10.3389/fgene.2020.623736

46. Assfaw WT, Girma TG, Abdoulaye T, Rabbi IY, Olanrewaju A, et al. The cassava monitoring survey in Nigeria final report. 2017; IITA, Ibadan, Nigeria. ISBN 978-978-8444-81-7.66
47. Albuquerque HYG, de Oliveira EJ, Brito AC, de Andrade LRB, do Carmo CD et al. Identification of duplicates in cassava germplasm banks based on single nucleotide polymorphisms (SNPs). Genetics and Plant Breeding Sci. Agric. (Piracicaba, Braz.) 2019; 76: (4)
48. Nweke FI, Spencer DSC, Lynam JK. The Cassava Transformation: Africa's Best-Kept Secret, Michigan State University Press, 2002**.**
49. Mtunguja MK, Laswai HS, Muzanila YC, Ndunguru J. Farmers knowledge on selection and conservation of cassava (*Manihot esculenta*) genetic resources in Tanzania. Journal Biology Agriculture 2014; 4:120–129.
50. Kyamanywa S, Kashaija IN, Getu E, Amata R, Senkesha N, Kullaya A. Enhancing Food Security through Improved Seed Systems of Appropriate Varieties of Cassava, Potato and Sweetpotato Resilient to Climate Change in Eastern Africa. 2011; Nairobi, Kenya, ILRI.
51. Ndung'u JN, Wachira FN, Kinyua MG, Lelgut DK, Njau P et al. Genetic diversity study of Kenyan cassava germplasm using simple sequence repeats. African Journal of Biotechnology 2014; 13(8):926-935
52. Koima IN, Orek CO. Response to Cassava Brown Streak Disease infections in local and improved cassava genotypes under field and greenhouse assays in lower eastern Kenya. International Journal of Pathology Research 2018; 1(3):1-14 https://doi.org/10.9734/ijpr/2018/v1i329616
53. Rufo R, Alvaro F, Royo C, Soriano JM. From landraces to improved cultivars: Assessment of genetic diversity and population structure of Mediterranean wheat using SNP markers. PLoS ONE 2019; 14(7): e0219867 https://doi.org/10.1371/journal.pone.0219867
54. Ferguson ME, Shah T, Kulakow P, Ceballos H. A global overview of cassava genetic diversity. PloS ONE, 2019; 14(11): e0224763. https://doi.org/10.1371/journal.pone.0224763
55. de Oliveira EJ, Alves F, Alves SL, de Oliveira V, da Silva S. Genetic variation of traits related to quality of cassava roots using affinity propagation algorithm. Genetics & Plant Breeding, Sci. Agric. 2015; 72 (1) https://doi.org/10.1590/0103-9016-2014-0043
56. Kretzschmar T, Mbanjo EGN, Magalit GA, Dwiyanti MS, Habib MA. DNA fingerprinting at farm level maps rice biodiversity across Bangladesh and reveals regional varietal preferences. Science Reports 2018; 8:14920.
57. Prempeh WNA, Manu-Aduening JA, Quain MD, Asante IK, Offei SK, Danquah EY. Assessment of genetic diversity among cassava landraces using single nucleotide polymorphic markers. African Journal of Biotechnology 2020; 19(6):383-391
58. Orek C. Farmer-cultivated landraces and improved cassava genotypes exhibit varied response to cassava brown streak disease under natural infections in Kenya. Journal of Plant Pathology Research, 2022; 4(1): 30-44 https://doi.org/10.36959/394/624
59. Orek C, Kyallo M, Yao N, 2023. Analysis of local cassava landraces and improved genotypes in response to infections by cassava mosaic begomoviruses under field conditions in Kenya. Tropical Plant Pathology, 2023; 48: 182-198.

https://doi.org/10.1007/s40858-023-00558-9