Genetic diversity of local and introduced cassava germplasm in Burundi using DArTseq molecular analyses

Niyonzima Pierre [1, 2, *], Lydia Nanjala Wamalwa[1], William Maina Muiru[1], Bigirimana Simon[2], Edward Kanju[3], Morag Elizabeth Ferguson [4], Malu Muia Ndavi[5] and Silver Tumwegamire[6]

[1]Department of Plant Science and Crop Protection, University of Nairobi, P.O. Box 29053-00625, Nairobi, Kenya

[2]Institut des Sciences Agronomiques du Burundi (ISABU), P.O. Box 795, Bujumbura, Burundi

[3]International Institute of Tropical Agriculture (IITA), P.O. Box 7878, Kampala, Uganda

[4]International Institute of Tropical Agriculture (IITA), P.O. Box 30709, Nairobi, Kenya

[5]International Fund for Agriculture Development (IFAD), Rome, Italy

[6]International Institute of Tropical Agriculture (IITA), P.O. Box 1269, Kigali, Rwanda

*Corresponding author: pierreniyonzima@ymail.com; pierreniyonzima2014@gmail.com

Abstract

In Burundi, most of small-scale farmers still grow traditional cassava landraces that are adapted to local conditions and have been selected for consumer preferred attributes. They tend to be susceptible, in varying degrees, to devastating cassava viral diseases such as Cassava Brown Steak Disease (CBSD) and Cassava Mosaic Disease (CMD) with production annual losses of US$1 billion annually. For long term resistance to the disease, several breeding strategies have been proposed. A sound basis for a breeding program is to understand the genetic diversity of both landraces and elite introduced breeding cultivars. This will also assist in efforts to conserve landraces ahead of the broad distribution of improved varieties which have the possibility of replacing landraces. Our study aimed at

determining the genetic diversity and relationships of local landraces and introduced elite germplasm using morphological and simple nucleotide polymorphism (SNP) markers as well as identifying a core set of germplasm from the local varieties to be used in the cassava breeding program. A total of 118 cultivars were characterized for morphological trait variation based on leaf, stem and root traits, and genetic variation using SNP markers. Results of morphological characterization based on Ward's Method revealed three main clusters and five accessions sharing similar characteristics. Molecular characterization identified over 18,000 SNPs and six main clusters and three pairs of duplicates which should be pooled together as one cultivar to avoid redundancy. Results of population genetic analysis showed low genetic distance between populations and between local landraces and elite germplasm. Accessions that shared similar morphological traits were divergent at the molecular level indicating that clustering using morphological traits was inconsistent. Despite the variabilities found within the collection, it was observed that cassava germplasm in Burundi have a narrow genetic base.

Key words: Cassava, Diversity Array Technology (DArT), Genetic diversity, SNP markers, Morphological traits, Morphological descriptors.

Introduction

Cassava was the most important staple crop in Burundi in 2019, with production of 2.41 million tons followed by fruits, bananas, sweet potatoes and vegetables [1]. It is grown mainly by small scale farmers throughout low, medium and high-altitude areas for human consumption. The root crop is eaten in the form of "imikembe", "ubuswage" and processed into flour for 'ugali' while leaves are used as vegetables or sauce [2]. Production of cassava doubled between 2010 to 2013 [1]. but since then, there has been a steady reduction in production, mainly due to cassava brown streak disease (CBSD) and cassava mosaic disease

(CMD). This is exacerbated by a lack of improved resistant cassava cultivars and the continued use of local susceptible landraces. The need to determine the genetic composition of local landraces and enhance the frequency of resistance genes within the local gene pool is a priority.

Breeding approaches for clonally propagated crops include variety introduction, germplasm assembly and maintenance, clonal selection and hybridization [3]. Breeding methods of cassava are defined by its genetic variability, the mode of reproduction and the breeding objectives. Cassava is a highly heterozygous species and presents substantial segregation in the first generation progenies, that are then evaluated through phenotypic mass selection [4]. The methods developed for self-pollinating crops are applicable to cassava with some modifications because of its specific characteristics. There is no classic genetic improvement methods initiated for vegetative propagated crops [5]. The main genetic improvement methods used in cassava are the assembly of the germplasm and selection followed by hybridization among selected elite clones [3, 6]. The introduction of varieties and selection are the most important breeding methods used in most of African countries [7]. However, crossing followed by selection of superior genotypes in the segregating population is the most universal method employed in cassava genetic improvement.

However, agricultural genetic diversity is imperative to provide a robust food security systems able to adapt to pest, diseases and environmental stresses [8] as well as to make genetic gains in plant breeding. It allows breeders to develop superior cultivars adapted to changing climatic conditions to meet the user demands. Understanding genetic diversity of species is the basis for a breeding program and to develop strategies for germplasm collection, management, conservation and improvement for food security and sustainable agricultural development [9, 10, 11]. Genetic diversity studies have widely been done for

78 cassava using both morphological and molecular methods in other countries such as Brazil
79 [12], Chad [13], Benin Republic [14], Nigeria [15] etc. but not Burundi.

80 Morphological markers are widely used to characterize cassava germplasm but it takes
81 long to get results compared to use of molecular markers. The morphological markers have
82 been used by [14] to study the genetic diversity and relationships among elite cassava
83 cultivars in Benin and highlighted a significant diversity and the most discriminating
84 morphological parameters within the germplasm.

85 Molecular markers have also been widely used for genetic diversity studies of cassava.
86 Molecular marker associated with agronomic traits have contributed significantly in marker
87 assisted cassava breeding programs [16]. De Souza [16] also reported the identification of
88 Simple Sequence Repeat (SSR) and Amplified Fragment Length Polymorphisms (AFLP)
89 markers linked to the CMD-resistance gene in cassava landraces and Random Amplified
90 Polymorphic DNAs (RAPD) markers linked to resistance to anthracnose. SSR markers were
91 developed and utilized to construct the genetic linkage map of cassava [17] and evaluate the
92 genetic diversity of cassava [16, 18, 19]. The first genetic linkage map of cassava was
93 constructed from F1 intra-specific cross using SSR, Restriction Fragment Length
94 Polymorphisms (RFLP) and RAPD [17]. The frequency and number of alleles per SSR
95 marker in the Puerto Rican cassava collection were determined [20]. RFLP, AFLP and RAPD
96 markers were used to analyze the genetic diversity of cassava [21, 22, 23]. Furthermore,
97 study on the genetic diversity and relationships within cassava germplasm using SNPs
98 markers, was done by [24]. The utilization of SNPs has gained popularity in recent years
99 due to their abundance, ubiquitous nature, polymorphism and amenability to automation
100 [25]. In cassava, SNPs have been used for genetic linkage mapping [26, 27], genome-wide
101 association studies [15] and genetic diversity assessments [28]. Diversity Arrays
102 Technology (DArT) markers for cassava were developed and reported as a tool for

genotyping large germplasm collections [29] but this has not been used on Burundian cassava genotypes. DArT performs well in polyploid species and does not require any existing DNA-sequence information and can be used with little resources required for SNP platforms [30]. Thus, it is a sequence-independent genotyping method designed to detect genetic variation at several hundred genomic loci in parallel without relying on sequence information.

Our study sought to characterize and determine relationships of cassava local landraces and introduced elite genotypes, as well as identifying any putative duplicates, using morphological and molecular markers.

Materials and methods

Germplasm collection and establishment

One hundred local landraces of cassava were collected from four agro-ecological zones in Burundi, namely Imbo plain, Mumirwa slopes, east and north depressions, and Central plateau (Table 1), selected on the basis of their importance in growing cassava. The identification of cassava landraces was done in farmers' fields jointly by farmers and investigators based on a short discussion. The landraces were recorded under the name as given by the farmers. Once collected, the landraces were planted in field gene banks at two sites (Moso and Murongwe research stations) representing two major cassava growing regions in Burundi for morphological and molecular characterization. Eighteen elite cassava genotypes earlier introduced to Burundi (Table 2) were also planted at the same sites. Single row plots with five plants spaced one meter between and within rows were used for each genotype in the trial. No fertilizer or irrigation was provided and weeds were managed throughout the growing period.

Table 1 Cassava landraces and their region of origin within Burundi

Name of accession	Agro-ecological zone	Name of accession	Agro-ecological zone	Name of accession	Agro-ecological zone
Nakarasi ya congo	1	Gatarina	3	Mpamba	4
Nakarasi y'ikirundi	1	Serereka	3	Mabare	4
Gitamisi_1	1	Bugiga annonciate_1	3	Imiduga_1	4
Muzinda	1	Yongwe_2	3	Tabika	4
Kwezikumwe	1	Gitikatika	3	Yongwe ederi	4
Rumonge	1	Gifunzo caritsa_1	3	Umukurajoro	4
Mbubute	1	Gifunzo caritsa_2	3	Rukokora	4
Yagata	1	Fyiroko	3	Kinazi dorothee1	4
Niga	1	Munebwe	3	Gasu	4
Ibigororoka	1	Ndoha	3	Inagitembe	4
Maguruyinkware_1	1	Maguruyinkware_2	3	Umutuburano	4
Mwarabu	1	Rumarampunu	3	Gitamisi_2	4
Rushishwa	1	Imikabika	3	Rubona_2	4
Sosomasi	1	Hanyesi	3	Nakarasi_1	4
Myezisita	1	Rubona_1	3	Surupiya	4
Zegura	1	Bwome devote1	3	Sogota	4
Igipila	1	Umuyobera	4	Nabuseri	4
Igikoshi	1	Gasahira	4	Imirundi	4
Nakarasi_2	1	Mbwayasaze	4	Imizariya	4
Solange	2	Kidihe_1	4	Maguruyinkware_3	4
Yongwe_1	2	Bunwa	4	Umutakabumba	4
Kibembe_1	2	Inarubono	4	Mugerera Yvonne_1	4
Criolina	2	Ntunduguru	4	Mugerera Yvonne_2	4
Matara	2	Kigoma	4	Kidihe_2	4
Sisiriya	2	Imijumbura	4	Nyawera	4
Ruvuna	2	Nyabisindu anastasie_1	4	Nyamugari sophie_1	4
Butoke	2	Kabumbe	4	Mukecuru	4
Kiganda	2	Gasasa	4	Fundiko	4
Ntabahungu	2	Yongwe_3	4	Umuhendangurube	4
Kibembe_2	2	Mutsindekwiburi	4	Sagarara	4
Munengera	3	Murozi	4	Imiduga_2	4
Mwotsi_2	3	Umusimbaruzi	4	Mwotsi_1	4
Berita	3	Bukarasi	4	Kavyiro	4
Ntegagakoko	3	-	-	-	-

1 = Imbo plain, 2 = Mumirwa slopes, 3 = East and north depressions, 4 = Central plateau

Table 2 Introduced elite germplasm in Burundi and their country of origin

Variety name	Country of origin
KBH2002/066	Tanzania
Pwani	Tanzania
Mkumba	Tanzania
KBH2006/026	Tanzania
Kizimbani	Tanzania
Kiroba	Tanzania
Albert	Tanzania
Okhumelela	Mozambique
Orera	Mozambique
Eyope	Mozambique
Tajirika	Kenya

F10-30-R2	Kenya
Kibandameno	Kenya
TZ 130	Uganda
Nase14	Uganda
Nase1	Uganda
Nase3	Uganda
MM96/5280	Burundi

131

132 **Morphological Characterization of Cassava Local Landraces and**

133 **Introduced Elite Germplasm**

134 Seventeen qualitative agro-morphological traits were evaluated (Table 3) using cassava
135 descriptors described by Fukuda *et al.* [31]. Data was collected at 3, 6, 9 and 12 months
136 after planting (MAP) using 17 descriptors with score codes that varied between 0 and 10
137 (Table 3). Color and pubescence on apical leaves were recorded earlier rather than later to
138 avoid obscured traits due to damage by cassava green mites that normally infest cassava at
139 later stages of plant growth. At 6 MAP, data on the shape of central leaf lobe and color of the
140 leaf and petiole, and petiole orientation were recorded by taking a leaf from the mid-height
141 stem position. At 9 MAP, data on prominence of foliar scars, color of stem cortex and color
142 of stem exterior were recorded from the middle third of the plant. Color of stem cortex was
143 visualized by shallow cut and peel back of the epidermis as described by Fukuda *et al.* [31].
144 Distance between leaf scars was measured from the middle part of stem on the middle third
145 of the plant, where the scars are not flat. Measurement was made along the stem and the
146 distance was divided by the number of nodes in the measured section to obtain the mean
147 internode length. Data on the stem's growth habit was recorded either as straight or zig-zag,
148 and color of the end branches of the adult plant was observed on the top 20 cm of the plant.
149 At 12 MAP, observations on color of root cortex, color of root-pulp, external color of storage

150 root and root taste were taken. Root cortex color and color of root-pulp were visualized by
151 removing the skin of the root and by transversal cutting of the root.

152 **Table 3** Qualitative traits used to characterize 118 cassava genotypes

Trait observed	Trait acronym	Sore code	Data entry
Color of apical leaves	CAL	3 = light green; 5 = dark green; 7 = purplish green; 9 = purple	3 MAP
Pubescence on apical leaves	PAL	0 = absent, 1 = present	3 MAP
Shape of central leaflet	SCL	1 = ovoid; 2 = elliptical-lanceolate; 3 = obovate-lanceolate; 4 = oblong-lanceolate; 5 = lanceolate; 6 = linear; 7 = pandurate; 8 = linear-pyramidal; 9 = linear-pandurate; 10 = linear-hostatilobalate	6 MAP
Petiole color	PC	1 = yellowish-green, 2 = green, 3 = reddish-green, 5 = greenish-red, 7 = red, 9 = purple	6 MAP
Leaf color	LC	3 = light green; 5 = dark green; 7 = purple green; 9 = purple	6 MAP
Petiole orientation	PO	1 = inclined upwards, 3 = horizontal, 5 = inclined downwards, 7 = irregular	6 MAP
Prominence of foliar scars	PFS	3 = semi-prominent, 5 = prominent	9 MAP
Color of stem cortex	CSC	1 = orange, 2 = light green, 3 = dark green	9 MAP
Color of stem epidermis	CSEp	1 = cream, 2 = light brown, 3 = dark brown, 4 = orange	9 MAP
Color of stem exterior	CSEx	3 = orange, 4 = green-yellowish, 5 = golden, 6 = light brown, 7 = silver, 8 = gray, 9 = dark brown	9 MAP
Distance between leaf cars	DBLS	3 = short (≤ 8 cm), 5 = medium (8–15 cm), 7 = long (≥15 cm)	9 MAP
Growth habit of stem	GHS	1 = Straight, 2 = Zig-zag	9 MAP
Color of end branches of adult plant	CEBAP	3 = Green, 5 = Green-purple, 7 = Purple	9 MAP
Color of root cortex	CRC	1 = White or cream, 2 = Yellow, 3 = Pink, 4 = Purple	12 MAP
Color of root-pulp	CRP	1 = white; 2 = cream; 3 = yellow; 4 = orange; 5 = pink	12 MAP
External color of storage root	ECSR	1 = white or cream; 2 = yellow; 3 = light brown; 4 = dark brown	12 MAP
Root taste	RT	1 = Sweet, 2 = Intermediate, 3 = Bitter	12 MAP

153 MAP = Months after planting

154

155 Molecular characterization of cassava local landraces and
156 introduced elite germplasm

157 In terms of DNA extraction; four disks of approximately 5mm diameter were collected from
158 young fresh leaf samples. These were dried in an oven overnight at 45^{o}C and shipped to
159 Intertek in Australia for DNA extraction, before being forwarded to Diversity Array
160 Technologies for genotyping using DArTseq. DNA quality and quantity were checked on a
161 0.8% agarose gel. Libraries were constructed at Diversity Arrays Technology in Canberra,
162 Australia according to DArTseq^{TM} complexity reduction method through digestion of

genomic DNA and ligation of barcoded adapters [32]. DArT uses a genotyping by sequencing DArTseq™ technology, providing rapid, high quality and affordable genome profiling, even from the most complex polyploid genomes [32, 33]. SNP marker scoring was achieved using DArTsoft14 which is an in-house marker scoring pipeline based on algorithms [32]. Two types of DArTseq: SilicoDArT and SNP markers were both scored as binary markers for presence or absence (1 and 0 respectively).

Data analyses

Morphological data was analyzed using the Statistical Package for the Social Sciences (SPSS) software (IBM SPSS Statistics for Windows version 20.0, IBM Corp, Armonk, NY.). Dissimilarity matrix was used to determine the relationship among accessions. The structure of morphological variation was visualized using ascending hierarchical clustering (AHC) based on data and Ward's Method to plot a dendrogram [24, 13]. Morphological traits distribution was determined using Microsoft (MS) Excel (2016). Generated SNP data were cleaned in MS Excel by removing all genotypes with >5% missing data and monomorphic SNPs. Hamming's single distance between genotypes was calculated using KDCompute, Version 1.5.2 beta and hierarchical clustering done by Ward's method for dendrogram (https://www.rdocumentation.org/packages/dartR). Generated SNP data were imported into DartR and then filtered for repeatability, monomorphic loci, call rate per locus, single locus per sequence tag and call rate per individual [34]. To better identify putative duplicated genotypes and to determine cut-off, known duplicate cassava genotypes were included with the samples genotyped. To assess the population statistics, the observed heterozygosity (Ho) was calculated using mean 'hobs' function and expected heterozygosity (He) using Hs function in the R package 'Adegenet' [35, 36, 37]. The pair wise fixation index (Fst) among populations was calculated using StAMPP package in R

(https://www.rdocumentation.org/packages/dartR) and the output value indicated existence or not of differentiation between populations where <15% indicate low differentiation, 0.15<Fst<0.25 indicate moderate differentiation and >25% indicate high differentiation [38]. Genetic relationships of landrace and introduced cassava genotypes were assessed by estimation of hamming distance between genotypes using dartR in KDCompute as described by Hoque et al. [39]. Single distance matrix was exported as a csv file and imported into DARwin v6.0.21 [40] to construct a dendrogram to estimate the genetic relationships.

Results

Morphological traits of Cassava local landraces and Improved Elite Germplasm

Leaf traits

There was a diversity of color on apical leaves for the cassava genotypes. Most accessions (68%) had purplish green color as the dominant color for apical leaf (Fig 1) mostly dominated by landraces (64.4%) (Fig 2), and few of the elite germplasm (3.4%) (Fig 3). About 19% and 15% of the accessions had dark-green and purple apical leaf color, respectively. Less than 7% of the accessions had pubescence on apical leaves (Fig 1). The shape of the leaves also had variations where 50% of the accessions had elliptic-lanceolate as the dominant shape (Fig 1) and mostly among the landraces (44.1%) (Fig 2). Obovate lanceolate, pandurate, lanceolate-pandurate and linear-pyramidal leaf shapes were rare, and altogether observed in 6.8% of the accessions (Fig 2). The color of petioles varied among the accessions, where purple color was the most frequent (50%). Most landraces

had purple color (44.9%) compared to only 5.1% of the elite germplasm (Fig 2-3) respectively. Other petiole colors were observed, including yellowish-green, green, green purple, purple yellow, red-green, and red (Fig 1). Dark green color (72.9%) (Fig 1) was the dominant leaf color observed in most accessions of which 68.6% were landraces (Fig 2). The most frequent petiole observed was horizontally oriented (56%) (Fig 1) and more so for landraces (52.5%) (Fig 2). Color of the end branches of adult plants was mostly greenish purple among accessions (47.5%) (Fig 1), which was the most frequent color among the landraces (39%) (Fig 2). However, green and purple colors were also observed among the accessions (Fig 1).

Stem traits

Most accessions (41.5%) had light green stem cortex color (Fig 1), mostly dominated by landraces (35.6%) (Fig 2) and a few (5.9%) by the elite germplasm (Fig 3). Dark green stem color was found on 27.1% of the 118 accessions (Fig 1). Epidermis color was diverse, where more than 56% accessions were dark brown (Fig 1) mostly landraces (32.2%) (Fig 2). The rest of the accessions (43%) had light brown stem epidermis. Color of stem exterior was mostly green yellow with 33.9% of the accessions (Fig 1) mostly landraces (Fig 2), followed by grey (27.1%) (Fig 1), also dominated by landraces (Fig 2). Silver color of stem exterior then followed at 21.2% (Fig 1) mostly landraces (Fig 2). Other stem exterior colors recorded were dark brown, orange and green colors (Fig 1).

Foliar scars were prominent among 78.8% of the accessions while 21.2% had semi prominent foliar scars. Accessions with prominent foliar scars were mostly landraces (65.2%) (Fig 2) while the elite germplasm with prominent foliar scars comprised only 13.6%. The distance between leaf scars varied within the cassava accessions where 66% had medium distance (8–15 cm) (Fig 1) comprised mostly by landraces (Fig 2), 34% had

long (≥ 15 cm) and short (≤ 8 cm) distance. All the accessions had straight stem growth habit except one (Orera) that had the zigzag stem habit (Fig 1).

Root traits

Cream root cortex color was recorded among 85.7% of the accessions (Fig 1) and almost all the elite germplasm (14.4%) belonged to this group (Fig 3). Pink root cortex has been observed on some accessions (Fig 1). All the accessions had white pulp color except Solange that had yellowish root pulp (Fig 1-2-3). Dark brownish external storage root color was the most frequent (39.0%) among the accessions (Fig 1) mostly the landraces (31.4%) (Fig 2). Bitter taste was noted for 79.5% of the accessions (Fig 1), mainly among both the landraces (Fig 2) and the elite germplasm (Fig 3).

Fig 1. Morphological traits distribution among both landraces and elite germplasm with error bars indicating whether differences are statistically significant

Fig 2. Morphological traits distribution among the cassava landraces with error bars indicating whether differences are statistically significant

Fig 3. Morphological traits distribution among the elite germplasm with error bars indicating whether differences are statistically significant

Hierarchical clustering of Cassava local landraces and Improved Elite Germplasm

Ascending hierarchical clustering analysis based on morphological traits and Ward's method showed three major clusters (I, II and III) (Fig 4) following the horizontal line at a dissimilarity level of 6. Cluster I containing 31 accessions (all local landraces) had two sub-clusters. Cluster II consisted of 26 accessions (3 sub-clusters) and was composed of local

landraces and elite genotypes. Sub-cluster III of cluster II consisted of five elite genotypes; Tajirika, Nase 1, Nase 3, KBH2002/066 and Orera, and two local landraces (Nakarasi and Igipila), sub-cluster II of 3 elite genotypes namely Kizimbani, Kiroba and Eyope while sub-cluster I consisted of 16 local landraces (Fig 4). Cluster III was the largest with two sub-clusters consisting of 51 local landraces and ten elite genotypes. Elite genotypes under this category were KBH2006/026, Okhumelela, MM96/5280, Nase14, F10-30-R2, TZ130, Albert, Mkumba, Kibandameno and Pwani (Fig 4).

Fig 4. Phenotypic classification of cassava accessions based on the Ward's method at a dissimilarity level of 6

Genetic relationship among cassava genotypes using DArT analyses

Results from DartR analysis showed 72 unique genotypes, 39 genotypes presented similar SNP profile (Fig 5) following the cut off (green line) calculated from the distance matrix based on an average value of known duplicates. Putative duplicates accessions were grouped in 16 classes, each of them with different clones (Fig 5). Genotypic classification of accessions based on Ward's method showed six major clusters (Fig 5) at dissimilarity level of 1.0 (red line). Cluster I had two elite genotypes, Pwani and Mkumba alongside five known duplicate checks: Pwani_2, Pwani_3_SB101, Mkumba_1, Pwani_1, Mkumba_2_SB102 (Fig 5). Cluster II had nine genotypes consisting of four local (Nakarasi ya congo, Rumonge, Munembwe and Gitamisi) and five elite genotypes, namely KBH2006/026, Tajirika, KBH2002/066, Kizimbani and Okhumelela (Fig 5). Cluster II had five duplicates namely, Tajirika-2, KBH 2002/026/1, KBH 2002/026/2, Tajirika-5CP-Kephis and KBH 2002-066-SB103 (Fig 5). Cluster III and V consisted of eight and seven accessions, respectively, all local landraces. Cluster IV was composed of 58 accessions sub clustering into two mains

groups that were sub clustered in different subgroups showing many similarities (Fig 5). Cluster IV consisted of 50 local landraces and eight elite genotypes including Orera, F10-30-R2, Kibandameno, Albert, Okhumelela, MM96/5280, Nase 14 and TZ130 (Fig 5). Cluster VI consisted of 33 local landraces. Paired similar accessions that fell into this category were Igikoshi and Munengera, Sosomasi and Igipila, Mwotsi, Mwarabu and Mwzisita, Bunwa and Kigoma, Maguruyinkware-2 and Rumaramuntu, Ndoha and Imikabika, and Bugiga annociate 1 and Gifunzo caritas 2 (Fig 5).

Fig 5. Genotypic classification of accessions based on the Ward's method at dissimilarity level of 1.0 (red line), the green line determining the threshold for putative and known duplicates

Assessment of the population statistics of the genotypes

Assessment was done within and between populations to determine existence of any relationships. The output values of calculated pair wise fixation index (Fst) among all populations were <15% indicating low differences between populations (Table 4, Fig 6). Results showed pair wise fixation index of 0.071, 0.095, 0.073 and 0.083 between elite genotypes and local landraces of Imbo plain, Mumirwa slopes, North east depressions and Central plateau, respectively, that showed little variation (Table 4). Between local landraces of Imbo plain and Mumirwa slopes, North East (NE) Depressions and Central plateau, the pair wise fixation index was 0.010, 0.023 and 0.020, respectively, indicating very low differentiation between populations. Pair wise fixation index between landraces of Mumirwa Slopes and NE Depressions, between landraces of Central plateau and NE Depressions were 0.027 and 0.028 respectively (Table 4).

311 **Table 4** Pairwise fixation index between landraces from different locations

	Elite genotypes	Landraces of Imbo Plain	Landraces of Mumirwa Slopes	Landraces of NE Depressions	Landraces of Central Plateau
Elite genotypes	-				
Landraces of Imbo Plain	0.071	-			
Landraces of Mumirwa Slopes	0.095	0.010	-		
Landraces of NE Depressions	0.073	0.023	0.027	-	
Landraces of Central Plateau	0.083	0.020	0.001	0.028	-

312

313 **Fig 6.** Genetic relationships between cassava populations based on Nei's genetic distance

314 Within population, output values for pair wise fixation index were greater than 25% for all
315 populations indicating high differentiation between genotypes (Table 5). Pair wise fixation
316 index of 0.59, 0.60, 0.57, 0.59 and 0.56 was noted within elite genotypes and landraces of
317 Imbo plain, Mumirwa slopes, NE depressions and Central plateau, respectively, indicating
318 high variation between genotypes (Table 5). The heterozygosity was calculated per marker
319 and population, where observed heterozygosity (Ho) was greater than expected
320 heterozygosity (He) in all populations except elite genotypes, indicating a suspected mixing
321 of previously isolated populations (Table 5).

322 **Table 5** Fixation index and heterozygosity within population

Population	Fixation Index F within population	Observed heterozygosity (Ho)	Expected heterozygosity (He)
Elite genotypes	0.59	0.25	0.27
Landraces of Imbo Plain	0.60	0.27	0.25
Landraces of Mumirwa slopes	0.57	0.27	0.25
Landraces of NE Depressions	0.59	0.26	0.25
Landraces of Central Plateau	0.56	0.26	0.25

324

Comparison of results from morphological and molecular dendrograms

Morphological classification clustered accessions into three main groups, whereas molecular analysis clustered accessions into six groups. All genotypes in clusters I and II for morphological classification method and clusters III, V and VI for genetic classification method were local landraces. Cluster I in the genetic classification method only consisted of elite genotypes (Pwani and Mkumba) while clusters II and IV for the same method and cluster III for the morphological clustering method contained both local landraces and elite genotypes.

Discussion

Morphological traits

Morphological characterization based on leaf traits (color of apical leaves, color of end branches of adult plant, pubescence on apical leaves, shape of central leaf lobe, petiole color, prominence of foliar scars, distance between leaf scars, leaf color, petiole orientation), stem traits (color of stem cortex, color of stem epidermis, color of stem exterior, growth habit of stem) and root traits (color of root cortex, color of root-pulp, external color of storage root and root taste) were diverse among the cassava landraces as well as the elite germplasm studied. These traits are very interesting and can be used in breeding and in identifying varieties.

Leaf traits

The leaf traits play an important role in cultivar identification and are more relevant for selection of cassava varieties suitable for the leafy vegetable markets. Leaf shape is one such an important trait as it affects leaf area and hence light interception which can directly

affect root yield [41]. [42, 14] reported respectively that leaf shape and color were the most important variables to distinguish cassava accessions and that farmers identify their cassava cultivars based on the traits related to leaf and stem color. The analysis revealed that apical leaves of 68% of cultivars were colored purplish green as dominant color and the mature leaves of 72.9% of cultivars are colored dark green as dominant color. [43] reported that leaf color plays an important role in predicting fresh root weight as nearly 90% of the dry matter (or biomass) of a plant is produced by leaves. Study of Khumaida et al. [44] revealed that dark green leaf color would increase the weight of tubers per plant, thus, can be useful in predicting root yield estimates of several cassava genotypes. Analysis also revealed few accessions colored light green on apical leaves, having hairs and with central leaflet shaped linear- pyramidal, which were comparable to those obtained by Nadjiam et al. [13]. According to Ehleringer et al. [45], presence of hair on apical leaves reduces leaf light absorbance, heat load, and consequently lower leaf temperatures and transpiration rates. On the other hand, presence of hairs on leaves lowers photosynthetic activity, and therefore lowers the yield. In addition, most of the end branches of adult plants of most of the genotypes were colored greenish purple. Comparable results were obtained by Eze et al. [46] who reported the predominant greenish purple color of end branches in adult plants in cassava varieties from Nigeria.

Stem traits

Most of the landrace accessions had stem epidermis and stem exterior colored light brown and grey, respectively. Similar findings were reported by Kosh-Komba et al. [47] who studied the diversity of cassava in Central Africa Republic and underlined clusters of accessions characterized by stems colored light brown and gray. Our results revealed that more than 78% of characterized cassava genotypes had prominent foliar scars while more than 21% had semi-prominent foliar scars. According to [48, 49], when cuttings are planted

in soil, roots developed from the foliar scars as well as lateral branches. Furthermore, the distance between leaf scars determines the number of scars per unit stem length and indeed the number of lateral branches.

Root traits

Presence of accessions showing cream color on root cortex and cream color on root pulp confirmed the presence of genotypes with low levels of beta-carotene, the precursor of Vitamin A [50]. Furthermore, yellow cassava roots are associated with high levels of proteins in leaves and, therefore, improving cassava for high beta-carotene content could also improve the overall nutritional value of the crop [50]. Almost 80% of the accessions were found to have a bitter taste, suggesting that roots from these cultivars may have high levels of toxic cyanogenic glucosides and therefore, must be processed prior to consumption as suggested by Chiwona-Karltun et al. [51]. Furthermore, the presence of different colors of the external storage roots can be used to differentiate cassava accessions.

Molecular characterization

The analysis based on molecular characterization clustered accessions into six main clusters indicating varying genetic distances between genotypes. Clusters I contained two elite genotypes, Pwani and Mkumba that shared all genetic characteristics, suggestive of putative duplicates. The putative duplicates clones detected within cluster I such as Pwani and Mkumba, which were distributed by the "New Cassava varieties and Clean Seed to Combat CMD and CBSD" (5CP) project have previously been shown to be the same genotype by Ferguson et al. (in press) (Supplementary file S12).

Clusters III, V and VI had eight, seven and 33 of all landraces, respectively, suggesting that landraces in each group shared similar genetic traits. Under cluster II, five elite genotypes clustered together with four local landraces while in cluster IV, seven elite genotypes clustered together with 51 local landraces, suggesting that some local landraces

399 and elite genotypes have similar genetic characteristics. Thus, the local landraces that
400 belonged to the two clusters (II and IV) could be possible sources of resistance or they could
401 be related due to genetic elements that control other traits other than resistance.
402 Furthermore, in cluster II, a pair of accessions namely: Nakarasi ya congo and Rumonge
403 shared all genetic characteristics, indicating that these accessions can be assumed to be
404 putative duplicates clones.

405 Morphological and molecular analysis

406 Accessions that shared similar morphological characteristics were distinct at the molecular
407 level, indicating that the resolution provided by morphological traits is lower than with
408 molecular markers. These results are in agreement with findings of [52, 53] who reported
409 that plants showing similar morphological characteristics could be very distinct at
410 molecular level. In addition, [52, 54] reported that clustering using morphological traits is
411 less reliable due to the influence of the environment and plant growth stage on their
412 expression and the limited number of markers to distinguish entities. This phenomenon
413 could explain the changing and clustering observed in comparing the hierarchical cluster
414 dendrograms for the morphological and molecular traits [54]. Accessions Kizimbani,
415 Rumonge, Munembwe and Tajirika were grouped in clusters II for both morphological and
416 molecular characterization. Pwani and Mkumba are grouped in cluster I for molecular
417 characterization while cluster III in morphological characterization together with
418 Kibandameno, Albert, TZ130, F10-30-R2, Nase 14 and MM96/5280 in one of the sub
419 clusters, demonstrating the varied discriminative powers of the two methods of
420 characterization. The putative duplicates clones detected within clusters shows that some
421 genotypes such as Pwani and Mkumba, belonging to elite germplasm and Imiduga,
422 Mutsindekwiburi and Rubona belonging to local landraces are several copies, thus could be
423 pooled together as one cultivar. Difference of number of clusters between the two methods

of characterization was due to the number of specific traits and genetic variations for phenotypic and genotypic classification respectively. However, the phenotypic classification dealt with a small number of traits while the genotypic classification dealt with more than 18 000 SNPs, hence the genotypic classification showed a lot of differences between accessions. Comparable results have been reported by [55, 56] while identifying duplicate accessions based on multi-locus analysis, and concluded that accessions presenting similar SNP profiles were assumed to be putative duplicates as each multi locus genotype corresponded to a single genotype.

Conclusion

The aim of this study was to determine the morphological and genotypic polymorphism in the local landraces and characterize elite cassava genotypes as well as identifying duplicates. Morphological and molecular characterization showed distinct classes of cultivars and within each class, sub classes with similar SNP profiles were identified. Accessions having very close similar characteristics namely Pwani and Mkumba, and Imiduga, Mutsindekwiburi and Rubona should be considered as putative duplicates, hence, need to be pooled together as one cultivar. Despite the variabilities found within the collection, it was concluded that cassava landraces in Burundi as well as the introduced clones present a narrow genetic base.

Acknowledgments

This work was sponsored by the International Institute of Tropical Agriculture (IITA) under the 'CBSD (Cassava Brown Streak Disease) Control Project'. The project is funded by the International Fund for Agricultural Development (IFAD) and the "Institut des Sciences

Agronomiques du Burundi" (ISABU) is the main partner in Burundi. The cassava research team from ISABU are acknowledged for their support in trial management and data collection. Ms Chepkoech, Jackline, a laboratory technician at the Integrated Genotyping Service and Support (IGSS) of the Bioscience Eastern and Central Africa supported the molecular analysis work.

References

1. FAOSTAT. Food and Agriculture Organization of the United Nations, Rome, Italy, 2019. http://www.fao.org/faostat/en/#data/QC/visualize.
2. Aloys N, Hui Ming Z. Traditional cassava foods in Burundi - A review. Food reviews international 2006; 22: 1-27. https://doi.org/10.1080/87559120500379761.
3. Acquaah G. Breeding Clonally Propagated Species. In Principles of Plant Genetics and Breeding 2012; 374-381. https://doi.org/10.1002/9781118313718.ch19.
4. Ceballos H, Kawuki RS, Gracen VE, Yencho GC, Hershey CH. Conventional breeding, marker-assisted selection, genomic selection and inbreeding in clonally propagated crops: a case study for cassava. Theoretical and Applied Genetics 2015; 128: 1647-1667. https://doi.org/10.1007/s00122-015-2555-4.
5. Fukuda WM, de Oliveira S, Iglesias C. Cassava breeding. Crop Breeding and Applied Biotechnology 2002; 2: 617-638.
6. Ceballos H, Iglesias CA, Pérez JC, Dixon AG. Cassava breeding: opportunities and challenges. Plant molecular biology 2004; 56: 503-516. https://doi.org/10.1007/s11103-004-5010-5.

7. Ceballos H, Pérez JC, Joaqui Barandica O, Lenis JI, Morante N, Calle F, et al. Cassava breeding I: The value of breeding value. Frontiers in Plant Science 2016; 7:1227. https://doi.org/10.3389/fpls.2016.01227.
8. Borromeo TH. Importance of plant genetic resources in sustainable development: global challenges, and solutions being developed in the Philippines. Journal of Developments in Sustainable Agriculture 2012; 7: 23-32.
9. Afuape SO, Okocha PI, Njoku D. Multivariate assessment of the agromorphological variability and yield components among sweet potato (*Ipomoea batatas* (L.) Lam) landraces. African Journal of Plant Science 2011; 5: 123-132. https://doi.org/10.5897/AJPS.9000176.
10. Chiveu CJ, Dangasuk OG, Omunyin ME, Wachira FN. Quantitative variation among Kenyan populations of *Acacia senegal* (L.) Willd. for gum production, seed and growth traits. New Forests 2009; 38: 1. https://doi.org/10.1007/s11056-008-9128-1.
11. Diouf M, Mbengue NB, Kante A. Caractérisation des accessions de 4 espèces de légumes-feuilles traditionnels (Hibiscus sabdariffa L., Vigna unguiculata (L.) WALP, Amaranthus L. spp et Moringa oleifera LAM) au Sénégal. 2007.
12. Moura EF, Farias Neto JT, Sampaio JE, Silva DT, Ramalho GF. Identification of duplicates of cassava accessions sampled on the North Region of Brazil using microsatellite markers. Acta Amazonica 2013; 43: 461-467. https://doi.org/10.1590/S0044-59672013000400008.
13. Nadjiam D, Sarr PS, Naïtormbaïdé M, Mbaïguinam JM, Guisse A. Agro-morphological characterization of cassava (Manihot esculenta Crantz) cultivars from Chad. Agricultural Sciences 2016; 7: 479-492. https://doi.org/10.4236/as.2016.77049.
14. Agre AP, Gueye B, Adjatin A, Dansi M, Bathacharjee R, Rabbi IY, et al. Folk taxonomy and traditional management of cassava (Manihot esculenta Crantz) diversity in

495 southern and central Benin. International Journal of Innovation and Scientific
496 Research 2016; 20: 500-515.
497 15. Rabbi IY, Kayondo SI, Bauchet G, Yusuf M, Aghogho CI, Ogunpaimo K, et al. Genome-
498 wide association analysis reveals new insights into the genetic architecture of
499 defensive, agro-morphological and quality-related traits in cassava. Plant Molecular
500 Biology 2020; 30: 1-9. https://doi.org/10.1007/s11103-020-01038-3.
501 16. De Souza CR. Genetic and genomic studies of cassava. Genes, Genomes and Genomics
502 2007; 1: 157-166.
503 17. Fregene M, Angel F, Gómez R, Rodríguez F, Chavarriaga P, Roca W, et al. A molecular
504 genetic map of cassava (*Manihot esculenta* Crantz). Theoretical and Applied Genetics
505 1997; 95: 431-441. https://doi.org/10.1007/s001220050580.
506 18. Kizito EB, Bua A, Fregene M, Egwang T, Gullberg U, Westerbergh A. The effect of
507 cassava mosaic disease on the genetic diversity of cassava in Uganda. Euphytica 2005;
508 146: 45–54. https://doi.org/10.1007/s10681-005-2959-3.
509 19. Fregene MA, Suarez M, Mkumbira J, Kulembeka H, Ndedya E, Kulaya A, et al. Simple
510 sequence repeat marker diversity in cassava landraces: genetic diversity and
511 differentiation in an asexually propagated crop. Theoretical and Applied Genetics
512 2003; 107: 1083-1093. https://doi.org/10.1007/s00122-003-1348-3.
513 20. Montero-Rojas M, Correa AM, Siritunga D. Molecular differentiation and diversity of
514 cassava (Manihot esculenta) taken from 162 locations across Puerto Rico and
515 assessed with microsatellite markers. AoB Plants. 2011.
516 21. Elias M, Panaud O, Robert T. Assessment of genetic variability in a traditional cassava
517 (Manihot esculenta Crantz) farming system, using AFLP markers. Heredity 2000; 85:
518 219-230. https://doi.org/10.1046/j.1365-2540.2000.00749.x.

519 22. Fregene M, Bernal A, Duque M, Dixon A, Tohme J. AFLP analysis of African cassava
520 (Manihot esculenta Crantz) germplasm resistant to the cassava mosaic disease
521 (CMD). Theoretical and Applied Genetics 2000; 100: 678-685.
522 https://doi.org/10.1007/s001220051339.
523 23. Marmey P, Beeching JR, Hamon S, Charrier A. Evaluation of cassava (Manihot
524 esculenta Crantz) germplasm collections using RAPD markers. Euphytica 1993; 74:
525 203-209. https://doi.org/10.1007/BF00040402.
526 24. Karim KY, Ifie B, Dzidzienyo D, Danquah EY, Blay ET, Whyte JB, et al. Genetic
527 characterization of cassava (Manihot esculenta Crantz) genotypes using agro-
528 morphological and single nucleotide polymorphism markers. Physiology and
529 Molecular Biology of Plants 2019; 26: 317-330. https://doi.org/10.1007/s12298-019-
530 00740-x.
531 25. Mammadov J, Aggarwal R, Buyyarapu R, Kumpatla S. SNP markers and their impact on
532 plant breeding. International journal of plant genomics 2012; 2012.
533 https://doi.org/10.1155/2012/728398.
534 26. Masumba EA, Kapinga F, Mkamilo G, Salum K, Kulembeka H, Rounsley S, et al. QTL
535 associated with resistance to cassava brown streak and cassava mosaic diseases in a
536 bi-parental cross of two Tanzanian farmer varieties, Namikonga and Albert.
537 Theoretical and Applied Genetics 2017; 130: 2069-2090.
538 https://doi.org/10.1007/s00122-017-2943-z.
539 27. Nzuki I, Katari MS, Bredeson JV, Masumba E, Kapinga F, Salum K, et al. QTL mapping
540 for pest and disease resistance in cassava and coincidence of some QTL with
541 introgression regions derived from Manihot glaziovii. Frontiers in plant science 2017;
542 8:1168. https://doi.org/10.3389/fpls.2017.01168.

543 28. Ferguson ME, Shah T, Kulakow P, Ceballos H. A global overview of cassava genetic
544 diversity. PloS one 2019; 14: e0224763.
545 https://doi.org/10.1371/journal.pone.0224763.
546 29. Xia L, Peng K, Yang S, Wenzl P, De Vicente MC, Fregene M, Kilian A. DArT for high-
547 throughput genotyping of cassava (Manihot esculenta) and its wild relatives.
548 Theoretical and Applied Genetics 2005; 110: 1092-1098.
549 https://doi.org/10.1007/s00122-005-1937-4.
550 30. Wenzl P, Huttner E, Carling J, Xia L, Blois H, Caig V, et al. Diversity Arrays Technology
551 (DArT): A generic high-density genotyping platform. In7th Internatitional Safflower
552 Conference. 2008.
553 31. Fukuda WM, Guevara CL, Kawuki R, Ferguson ME. Selected morphological and
554 agronomic descriptors for the characterization of cassava. IITA. 2010.
555 32. Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H et al. Diversity arrays technology:
556 a generic genome profiling technology on open platforms. Methods in Molecular
557 Biology 2012; 888: 67-89. https://doi.org/10.1007/978-1-61779-870-2_5. PMID:
558 22665276.
559 33. Raman R, Cowley RB, Raman H, Luckett DJ. Analyses using SSR and DArT molecular
560 markers reveal that Ethiopian accessions of white lupin (*Lupinus albus* L.) represent a
561 unique gene pool. Open Journal of Genetics 2014; 4: 87-98.
562 https://doi.org/10.4236/ojgen.2014.42012.
563 34. Gruber B, Unmack P, Berry O, Georges A. Introduction to dartR. User Manual. 2019.
564 35. Adamack AT, Gruber B. PopGenReport: simplifying basic population genetic analyses
565 in R. Methods in Ecol Evol. 2014; 5: 384-387. https://doi.org/10.1111/2041-
566 210X.12158.

567 36. Jombart T, Ahmed I. Adegenet 1.3-1: new tools for the analysis of genome-wide SNP
568 data. Bioinformatics 2011; 27: 3070-3071.
569 https://doi.org/10.1093/bioinformatics/btr521.
570 37. Jombart T. Adegenet: a R package for the multivariate analysis of genetic markers.
571 Bioinformatics 2008; 24:1403–1405.
572 https://doi.org/10.1093/bioinformatics/btn129.
573 38. El Mousadik A, Petit RJ. High level of genetic differentiation for allelic richness among
574 populations of the argan tree [Argania spinosa (L.) Skeels] endemic to Morocco.
575 Theoretical and applied genetics 1996; 92: 832-839.
576 https://doi.org/10.1007/BF00221895.
577 39. Hoque MN, Rahman L. Estimation of Euclidean distance for different morpho-
578 physiological characters in some wild and cultivated rice genotypes (Oryza sativa
579 L.). Journal of Biological Sciences 2007; 7, 86-88.
580 https://doi.org/10.3923/jbs.2007.86.88.
581 40. Perrier X, Jacquemoud-Collet JP. DARwin software. 2006. http://darwin.cirad.fr/
582 41. Nkansah GO, Ofosu-Budu KG, Ayarna AW. Genetic diversity among local and
583 introduced avocado germplasm based on morpho-agronomic traits. International
584 Journal of Plant Breeding and Genetics 2013; 7: 76-91.
585 https://dx.doi.org/10.3923/ijpbg.2013.76.91.
586 42. Asare PA, Galyuon IK, Sarfo JK, Tetteh JP. Morphological and molecular based diversity
587 studies of some cassava (Manihot esculenta Crantz) germplasm in Ghana. African
588 Journal of Biotechnology 2011; 10:13900-13908. https://doi.org/10.5897/AJB11.929.
589 43. Phosaengsri W, Banterng P, Vorasoot N, Jogloy S, Theerakulpisut P. Leaf performances
590 of cassava genotypes in different seasons and its relationship with biomass. Turkish
591 Journal of Field Crops 2019; 24: 54-64. https://doi.org/10.17557/tjfc.564116.

592 44. Khumaida N, Maharani S, Ardie SW. The leaf color performance on several lines of
593 cassava and its relation with tuber yield as early reference. Procedia Environmental
594 Sciences 2015; 24: 39-46. https://doi.org/10.1016/j.proenv.2015.03.007.
595 45. Ehleringer JR, Mooney HA. Leaf hairs: effects on physiological activity and adaptive
596 value to a desert shrub. Oecologia 1978; 37: 183-200.
597 https://doi.org/10.1007/BF00344990.
598 46. Eze SC, Aba SC, Amuji CF. Phenotypic diversity in agro morphological features of six
599 new cassava (Manihot esculenta Crantz) genotypes evaluated in Nsukka, Southeast,
600 Nigeria. Nigerian Journal of Crop Science 2016; 3: 31- 49.
601 47. Kosh-Komba E, Akpavi S, Yao AW, Atato A, Duval MF, Dourma M, et al. Diversité
602 agromorphologique de Manihot esculenta Crantz.(Euphorbiaceae) cultivée dans trois
603 zones agroclimatiques en République centrafricaine (RCA). European Scientific
604 Journal ; 2014.
605 48. Adu MO, Asare PA, Asare-Bediako E, Amenorpe G, Ackah FK, Afutu E, et al.
606 Characterizing shoot and root system trait variability and contribution to genotypic
607 variability in juvenile cassava (Manihot esculenta Crantz) plants. Heliyon 2018; 4:
608 e00665. https://doi.org/10.1016/j.heliyon.2018.e00665. PMID: 30003159
609 49. Banoc DM, Yamauchi A, Iijima M, Kono Y. Root system development of cassava and
610 sweetpotato during early growth stage as affected by high root zone temperature.
611 Plant production science 1999; 2: 247-251. https://doi.org/10.1626/pps.2.247.
612 50. Njenga P, Edema R, Kamau J. Combining ability for beta-carotene and important
613 quantitative traits in a cassava F1 population. Journal of Plant Breeding and Crop
614 Science 2014; 6: 24-30. https://doi.org/10.5897/JPBCS12.069.
615 51. Chiwona-Karltun L, Brimer L, Kalenga Saka JD, Mhone AR, Mkumbira J, Johansson L, et
616 al. Bitter taste in cassava roots correlates with cyanogenic glucoside levels. Journal of

the Science of Food and Agriculture 2004; 84: 581-590. https://doi.org/10.1002/jsfa.1699.

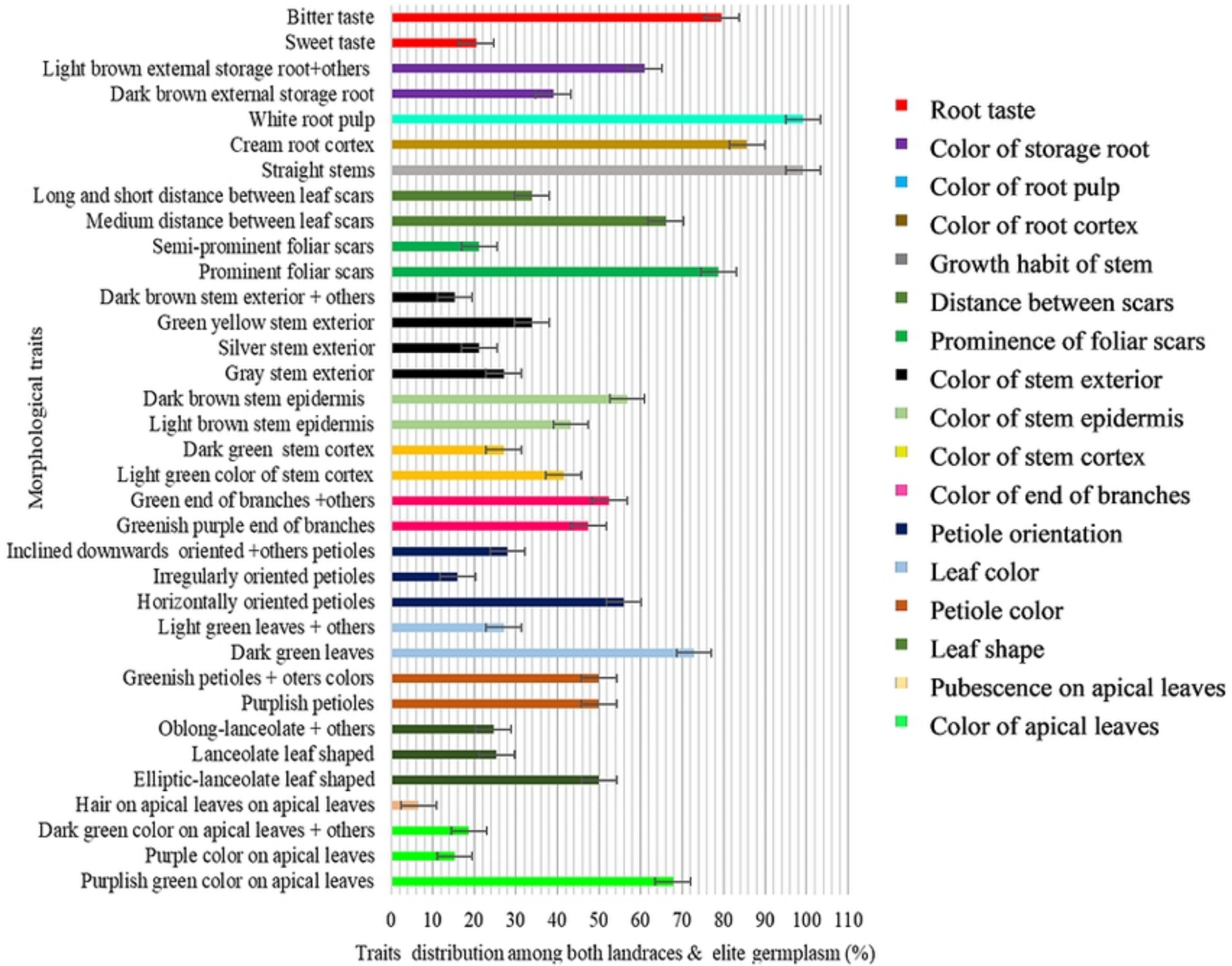
52. Sujii PS, Fernandes ET, Azevedo VC, Ciampi AY, Martins K, de O Wadt LH. Morphological and molecular characteristics do not confirm popular classification of the Brazil nut tree in Acre, Brazil. Genet Mol Res. 2013; 12: 4018-4027. http://dx.doi.org/10.4238/2013.September.27.3.
53. Feldberg K, Váňa J, Schulze C, Bombosch A, Heinrichs J. Morphologically similar but genetically distinct: on the differentiation of Syzygiella concreta and S. perfoliata (Adelanthaceae subfam. Jamesonielloideae). The Bryologist 2011; 114: 686-695. https://doi.org/10.1639/0007-2745-114.4.686.
54. Darkwa K, Agre P, Olasanmi B, Iseki K, Matsumoto R, Powell A, et al. Comparative assessment of genetic diversity matrices and clustering methods in white Guinea yam (Dioscorea rotundata) based on morphological and molecular markers. Sci Rep. 2020; 10: 1-4. https://doi.org/10.1038/s41598-020-69925-9.
55. Albuquerque HY, Oliveira EJ, Brito AC, Andrade LR, Carmo CD, Morgante CV, et al. Identification of duplicates in cassava germplasm banks based on single-nucleotide polymorphisms (SNPs). Scientia Agricola 2019; 76: 328-336. https://doi.org/10.1590/1678-992X-2017-0389.
56. Arnaud-Haond S, Duarte CM, Alberto F, Serrao EA. Standardizing methods to address clonality in population studies. Mol Ecol. 2007; 16: 5115-5139. https://doi.org/10.1111/j.1365-294X.2007.03535.x.

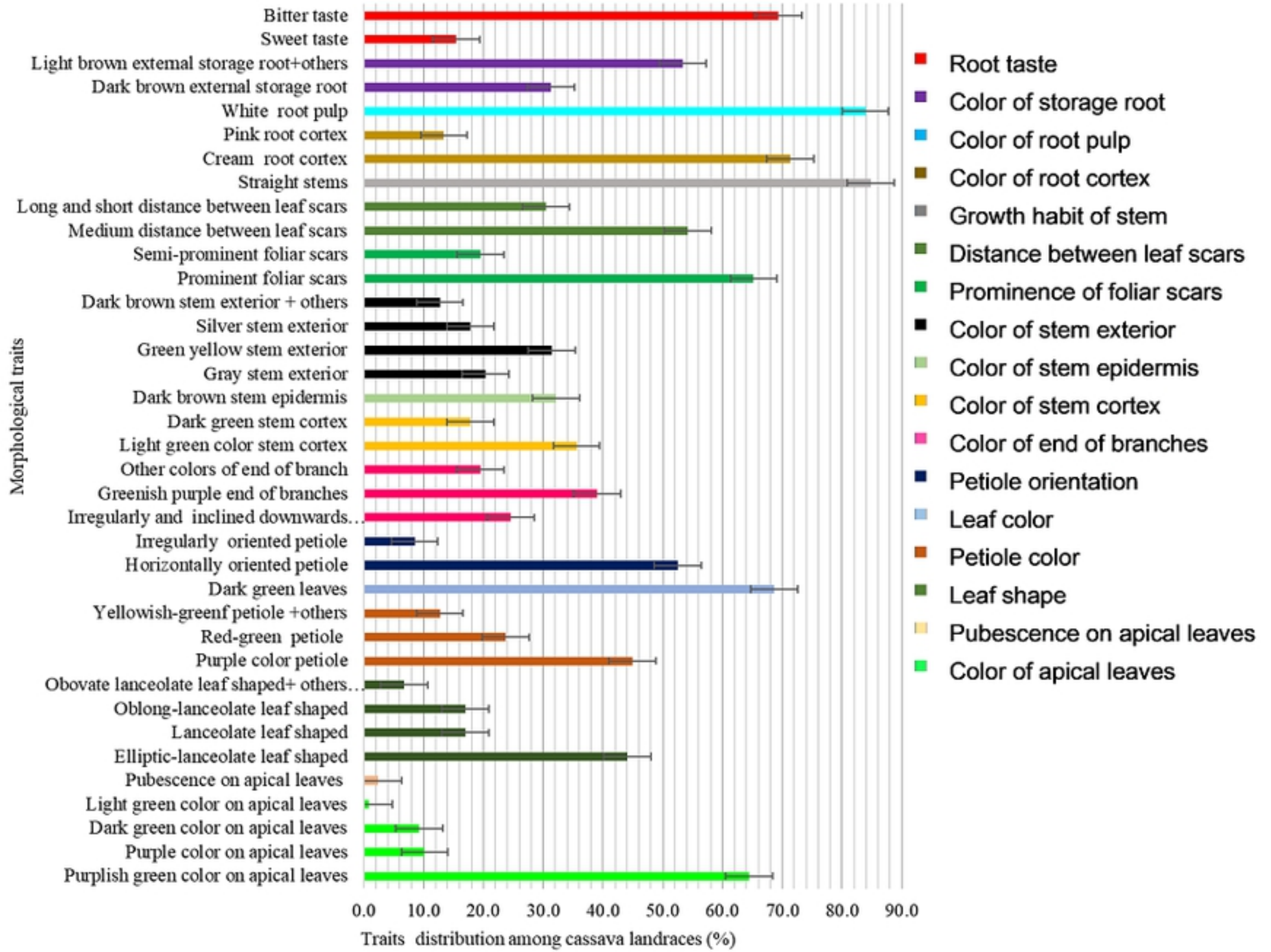
Supporting information

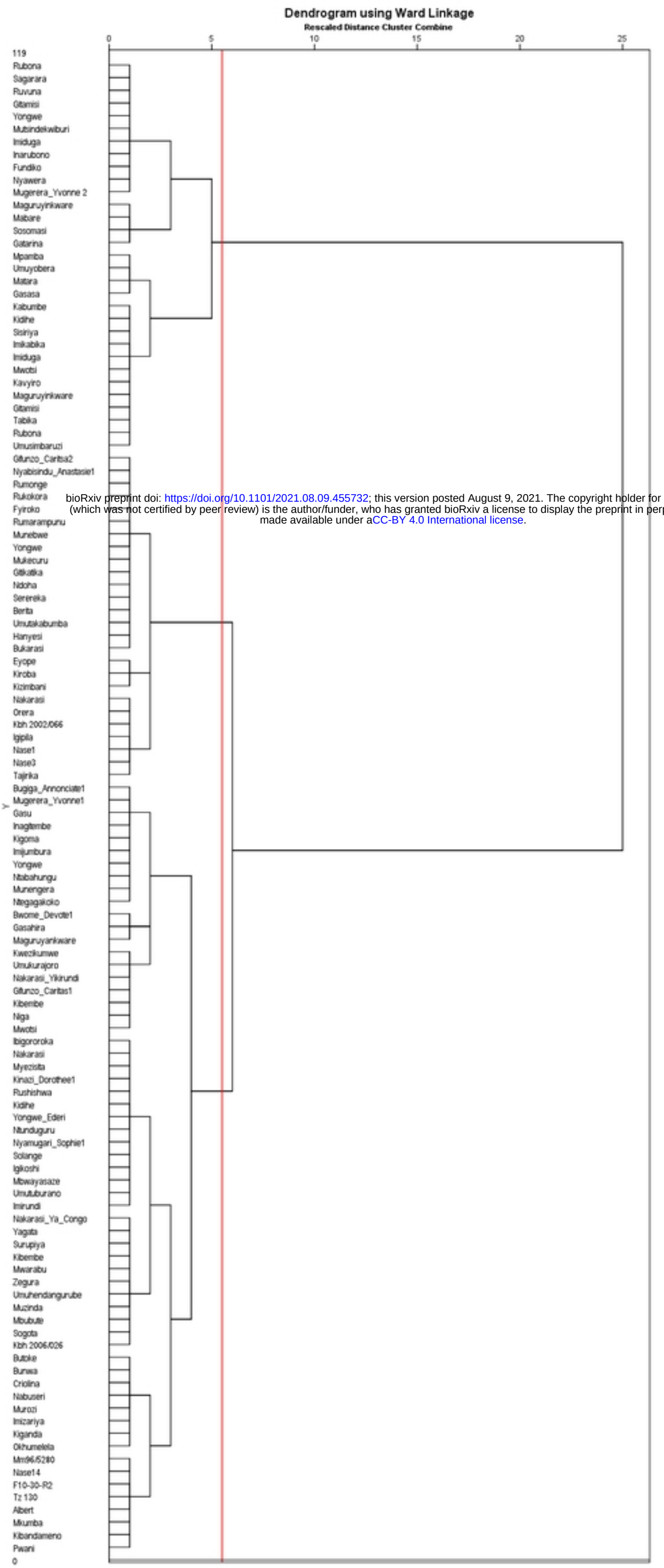
S1 Table. SNP data from DArtR Seq. 0: homozygous reference allele, 1: homozygous alternate allele, 2: heterozygous alleles, -: missing value, M1 to M18124: numbers of SNP's

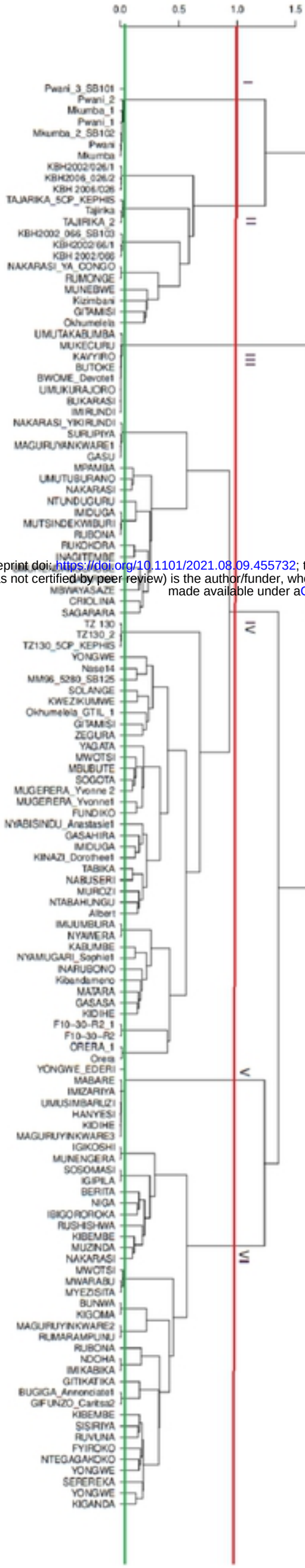
642 **S2 Table.** Hamming distance matrix.

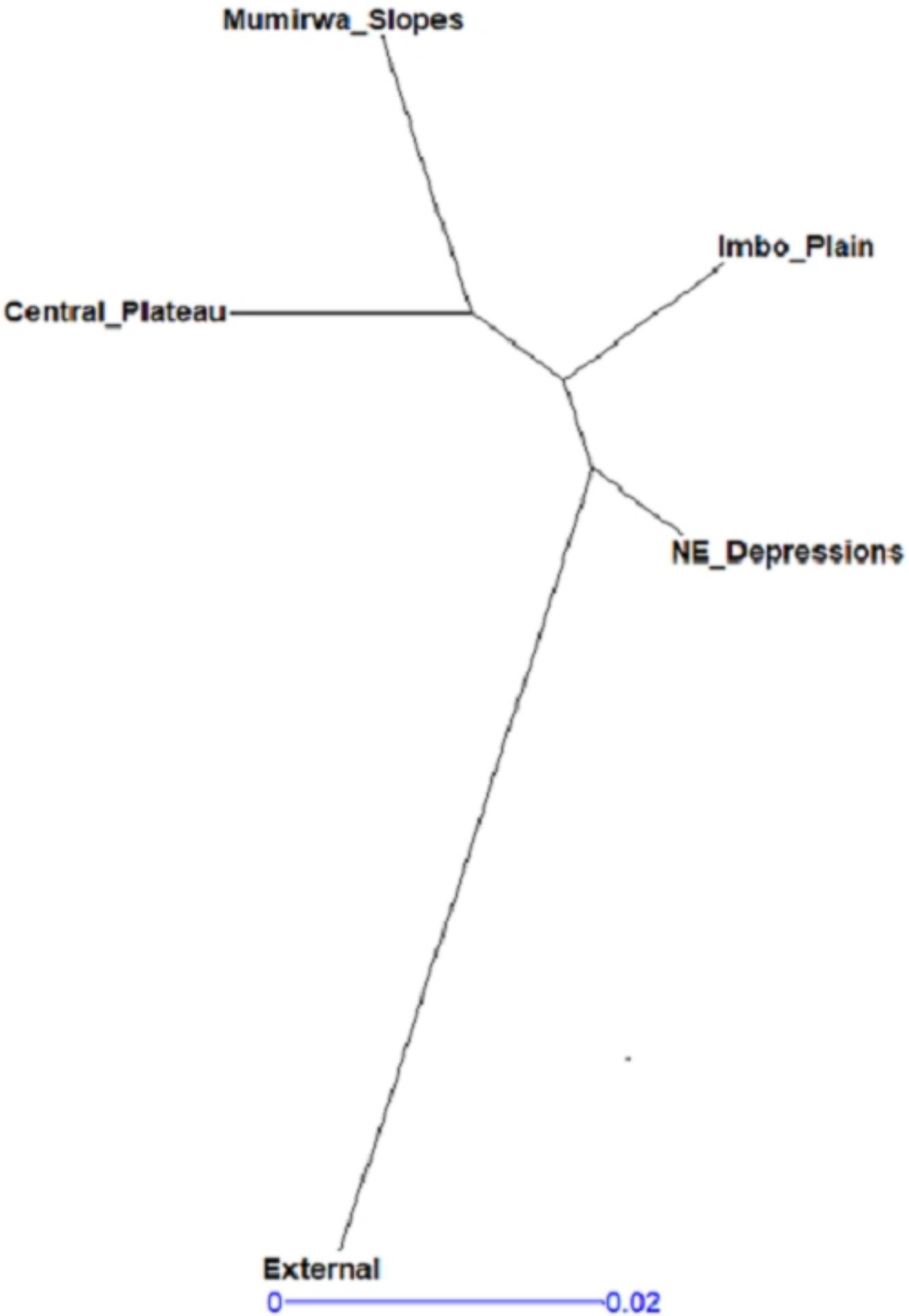
643 **S3 Table.** Morphological traits.


Bitter taste
Sweet taste
Light brown external storage root+others
Dark brown external storage root
White root pulp
Cream root cortex
Straight stems
Long and short distance between leaf scars
Medium distance between leaf scars
Semi-prominent foliar scars
Prominent foliar scars
Dark brown stem exterior + others
Green yellow stem exterior
Silver stem exterior
Gray stem exterior
Dark brown stem epidermis
Light brown stem epidermis
Dark green stem cortex
Light green color of stem cortex
Green end of branches +others
Greenish purple end of branches
Inclined downwards oriented +others petioles
Irregularly oriented petioles
Horizontally oriented petioles
Light green leaves + others
Dark green leaves
Greenish petioles + oters colors
Purplish petioles
Oblong-lanceolate + others
Lanceolate leaf shaped
Elliptic-lanceolate leaf shaped
Hair on apical leaves on apical leaves
Dark green color on apical leaves + others
Purple color on apical leaves
Purplish green color on apical leaves
Morphological traits
0 10 20 30 40 50 60 70 80 90 100 110
Traits distribution among both landraces & elite germplasm (%)
Root taste
Color of storage root
Color of root pulp
Color of root cortex
Growth habit of stem
Distance between scars
Prominence of foliar scars
Color of stem exterior
Color of stem epidermis
Color of stem cortex
Color of end of branches
Petiole orientation
Leaf color
Petiole color
Leaf shape
Pubescence on apical leaves
Color of apical leaves


Bitter taste
Sweet taste
Light brown external storage root+others
Dark brown external storage root
White root pulp
Pink root cortex
Cream root cortex
Straight stems
Long and short distance between leaf scars
Medium distance between leaf scars
Semi-prominent foliar scars
Prominent foliar scars
Dark brown stem exterior + others
Silver stem exterior
Green yellow stem exterior
Gray stem exterior
Dark brown stem epidermis
Dark green stem cortex
Light green color stem cortex
Other colors of end of branch
Greenish purple end of branches
Irregularly and inclined downwards…
Irregularly oriented petiole
Horizontally oriented petiole
Dark green leaves
Yellowish-greenf petiole +others
Red-green petiole
Purple color petiole
Obovate lanceolate leaf shaped+ others…
Oblong-lanceolate leaf shaped
Lanceolate leaf shaped
Elliptic-lanceolate leaf shaped
Pubescence on apical leaves
Light green color on apical leaves
Dark green color on apical leaves
Purple color on apical leaves
Purplish green color on apical leaves
Morphological traits
0.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
80.0
90.0
Traits distribution among cassava landraces (%)
Root taste
Color of storage root
Color of root pulp
Color of root cortex
Growth habit of stem
Distance between leaf scars
Prominence of foliar scars
Color of stem exterior
Color of stem epidermis
Color of stem cortex
Color of end of branches
Petiole orientation
Leaf color
Petiole color
Leaf shape
Pubescence on apical leaves
Color of apical leaves






Mumirwa_Slopes
Imbo_Plain
Central_Plateau
NE_Depressions
External
0
0.02