



# Collection and characterization of cassava germplasm in Comoros

M. E. Ferguson<sup>1</sup> · P. Ntawuruhunga · M. O. Fouad · O. K. M. A. Hassane · M. A. Charmila · A. A. Hamza · N. A. Mohamed · I. Dieng · E. Njukwe · T. M. Shah

Received: 20 March 2023 / Accepted: 31 May 2023  
© The Author(s) 2023

**Abstract** In Comoros, cassava plays a major food security role, however yields are low as few modern cultivars are grown. Prior to the introduction of new cultivars, and as a germplasm resource for breeders, germplasm collection missions were undertaken in the three largest islands; Ngazidja, Ndzouani and Mwali; and associated farmer knowledge documented. Cassava landraces were collected from 34 farms, and 17 key informant interviews conducted. Stakes of 79 collected landraces were planted for agro-morphological characterization. All landraces

were genotyped using DaRTSeq technology and data analysed to identify duplicates. Genetic fingerprints of 46 unique landraces were co-analysed with 402 previously genotyped landraces and improved cultivars from Tanzania. From this set only one match was made with a very old cultivar, Aipin Valenca, from the Northern Zone in Tanzania. According to SNP data, germplasm from the three islands of Comoros were similarly related to one another, and more distantly related to germplasm from Tanzania. They were most closely related to germplasm from the Northern Zone in Tanzania, suggesting a possible historical introduction of germplasm from this area. Lower levels of diversity were observed on these

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10722-023-01626-4>.

M. E. Ferguson (✉) · T. M. Shah  
International Institute of Tropical Agriculture (IITA), P.O.  
Box 30709, 00100 Nairobi, Kenya  
e-mail: m.ferguson@cgiar.org

T. M. Shah  
e-mail: tm.shah@cgiar.org

P. Ntawuruhunga  
IITA, Lusaka, Zambia  
e-mail: p.ntawuruhunga@cgiar.org

M. O. Fouad · O. K. M. A. Hassane · M. A. Charmila ·  
A. A. Hamza · N. A. Mohamed  
Institut National de Recherche pour l'Agriculture, la Pêche  
et l'Environnement (INRAPE), Ngazidja, Comoros  
e-mail: fouadmohamedoussouf@gmail.com

O. K. M. A. Hassane  
e-mail: medal.umm@gmail.com

M. A. Charmila  
e-mail: charm-i2m@hotmail.fr

A. A. Hamza  
e-mail: abdouazalihamza19@gmail.com

N. A. Mohamed  
e-mail: nobataine06@yahoo.fr

I. Dieng  
IITA, Ibadan, Nigeria  
e-mail: I.Dieng@cgiar.org

E. Njukwe  
IITA, Kigali, Rwanda  
e-mail: E.Njukwe@cgiar.org

*Present Address:*  
E. Njukwe  
CORAF, Dakar, Senegal

islands, as well as the islands of Pemba and Zanzibar. This implies limited introduction and movement of cassava germplasm into the islands. Introductions of improved germplasm to Comoros is recommended with the simultaneous conservation of collected unique landraces. Two landraces with high market demand and reported tolerance to diseases were identified for further evaluation with a view to multiplication and distribution and incorporation into the breeding program.

**Keywords** Farmer knowledge · Morphological descriptors · Genotyping · Diversity

## Introduction

Africa depends on the climate-resilient, starchy root crop, cassava (*Manihot esculenta* Crantz), to feed a substantial proportion of its' population. Annual production in sub-Saharan Africa is estimated at 178 million tons, which is 61% of global production (FAOSTAT 2022). Cassava serves as a food security crop because of its concentrated source of carbohydrate, ability to produce a yield under diverse climatic conditions, ability to remain in the soil without deteriorating until needed, and its amenability to staggered harvesting (De Vries and Toenniessen 2001). It can also be used for animal feed and industrial applications for income generation (Li et al. 2017). Nevertheless, on-farm yields in Africa are low; 8.62 Mt/ha in 2020 compared to 14.6Mt/ha in Latin America in the same year (FAOSTAT 2022). This is largely due to farmers growing traditional landraces, which are often inherently low yielding and susceptible to widespread virus diseases, including cassava mosaic disease (CMD) and cassava brown streak disease (CBSD) (Legg et al. 2014). Landraces of cassava, often referred to as farmer-cultivars or heirloom cultivars, have been grown for many years, being subject to natural and artificial selection by farmers, and represent an important source of genetic variation for plant breeding. It is important that this diversity is collected, evaluated, and conserved as some of these landraces may be vulnerable to extinction through susceptibility to viruses or replacement by modern cultivars.

In Comoros, cassava is the main food crop after banana, a major source of calories and the most

resilient of local food crops. The Comoros is a group of islands located at the northern end of the Mozambique Channel of the Indian Ocean, between Madagascar and the southeast African mainland, about 290 km off the eastern coast of Africa. Agriculture contributes 35.6% to the country's GDP and employs 68% of the country's labor force and accounts for 90% of exports (IFAD 2016). Cassava germplasm in these islands comprise both landraces and improved cultivars that could have been introduced from neighboring countries, or when cassava was first introduced to the East African coast in the 17<sup>th</sup> or eighteenth centuries (Carter et al. 1992). A first step in improving cassava production in Comoros is to collect and conserve existing landraces and farmer knowledge. By being cognisant of this information, breeders are more likely to breed or introduce appropriate improved cassava cultivars that respond to the intricacies of farmers' preferences with implications for positive varietal adoption. After collection of landraces, it is important to identify duplicates in the collection as farmers often give different names to the same cultivar and thus define a unique set of germplasm for conservation and for breeders to utilize. In addition, germplasm must be evaluated, to rapidly identify any obvious, interesting characteristics, and to determine any genetic structure, or relationship to germplasm from neighboring regions.

Although traditional agro-morphological characterization is useful to assess general features of the germplasm, and identify any landraces with striking characters, molecular characterization provides a far greater number of markers that are more stable and less influenced by the environment for diversity assessment (Tiago et al. 2016). Several molecular marker technologies have been used to characterize cassava germplasm to identify genetic variability. These include Random Amplified Polymorphic DNA (Asante and Offei 2003), Amplified Fragment Length Polymorphism (Roa et al. 1997), microsatellites (Hurtado et al. 2008) and more recently, Single nucleotide polymorphic (SNP) markers (Kawuki et al. 2009; Ferguson et al. 2019, 2021; Pierre et al. 2022). The objectives of the present study were to (1) document farmers' knowledge relating to the genetic material collected, (2) assess the cassava genetic variability and diversity available in Comoros using morphological and molecular characters and (3) identify

a unique set of cassava germplasm for conservation, further evaluation and utilization by breeders.

## Materials and methods

### Collection of the germplasm

Cassava germplasm collections were conducted by the National Research Institute for Agriculture, Fisheries and the Environment (INRAPE) of Comoros (from whom permission to collect was obtained), in collaboration with the International Institute of Tropical Agriculture (IITA) in July 2019. At this time the crop was nearing maturity at approximately 10 months after planting. All the collections were carried out under the terms of the International Treaty of Plant Genetic Resources for Food and Agriculture (ITPGRFA) (<http://www.fao.org/plant-treaty/en/>). Three of the largest islands of Comoros; Ngazidja, Ndzuani and Mwali were targeted, with the objective of collecting as many different cultivars as possible covering the different agro-ecological production zones and to document related farmer knowledge. The initial choice of municipalities and regions was determined according to the following criteria: (1) whether the locality was known for its diversity in cassava; (2) whether cassava was traditionally grown, (3) whether it was subject to high or increasing disease pressure, (4) whether the landraces were somehow endangered, such as through replacement by improved cultivars or (5) environments with distinct agro-ecological characteristics where specific adaptation may have occurred.

Key informants (farmers), who had a reputation for growing or conserving landraces were identified with the help of extension agents or village elders. At each farm, three questionnaires were conducted according to Cox et al. (2014). These were designed to capture farmer knowledge and preferences: (1) a 'Key Informant Interview' to profile each farmer, their farm and identify cultivars that they grew, (2) an 'Individual Landrace' form documented information on the history of the landrace, farmer's experiences and preferences of the landrace, both of agronomy, response to biotic and abiotic stresses and culinary properties, and (3) a germplasm collection form, to document morphological and agronomic descriptors of individual landraces, based

on Fukuda et al. (2010), and describe the collection environment. Yield potential was assessed by enquiring about the number of plants that would need to be harvested to fill a 25 kg bag. Collection and documentation of farmers knowledge followed the ITPGRFA and the terms of the Convention on Biological Diversity including the Nagoya Protocol. Prior verbal informed consent was obtained from all participants for inclusion in the study.

Decision to collect a stem cutting of a landrace was made according to the following criteria; (1) a local cultivar bearing a name which had not been collected before, (2) the local cultivar name had been collected before, but in a different region, or (3) the collectors identified unusual or interesting characteristics. After selection, a collection number was assigned to uniquely identify the landrace, and the three collection forms were completed. Two to three stems from a single plant of each collected cultivar were then sampled and labelled. The label included at least the collection number, the local cultivar name and the town. Photos of different parts of each cultivar collected were also taken. On returning to the station, stakes of each cultivar collected were planted in a field genebank at INRAPE.

A database with all information documented during the collection missions was constructed according to that outlined in Ferguson et al. (2021). The database was queried to summarize information relating to the collected landraces and farmer profiles. Coordinates defining sampling sites were plotted on a map using GADM (<https://gadm.org/license.html>) (GADM 2023).

### Agro-morphological evaluation

Five stem cuttings per plot of each landrace sampled during the collection mission were planted in a randomized complete block design with two replicates at the INRAPE station for morphological characterization according to Fukuda et al. (2010) in July 2019. Plants were scored on an individual basis and an average taken for quantitative measurements, and the most frequent score for qualitative measures. A list of the traits scored, with abbreviation, equivalent Term ID from Cassavabase ([www.cassavabase.org](http://www.cassavabase.org)) (Fernandez-Pozo et al. 2015) and data type is provided in Table 1.

**Table 1** Quantitative and qualitative traits scored with abbreviation, Term ID from the trait ontology in Cassavabase and data type

| Descriptor number <sup>a</sup>       | Trait                                 | Abbreviation | Term ID from Cassava ontology <sup>b</sup> | Data type                |
|--------------------------------------|---------------------------------------|--------------|--|--------------------------|
| <i>Descriptors scored at 3 MAP</i>   |                                       |              |  |                          |
| 1                                    | Colour of apical leaves               | CoApLF       | CO_334:0000101                             | Qualitative              |
| 2                                    | Pubescence of apical leaves           | PubALF       | CO_334:0000309                             | Qualitative              |
| <i>Descriptors scored at 6 MAP</i>   |                                       |              |  |                          |
| 3                                    | Leaf retention                        | LFRet        | CO_334:0000048                             | Qualitative              |
| 4                                    | Shape of central leaflet              | SCenLF       | CO_334:0000119                             | Qualitative              |
| 5                                    | Colour of petiole                     | CoPet        | CO_334:0000023                             | Qualitative              |
| 6                                    | Colour of leaves                      | CoLF         | CO_334:0000102                             | Qualitative              |
| 7                                    | Number of leaf lobes                  | NoLFLo       | CO_334:0000082                             | Qualitative              |
| 8                                    | Length of leaf lobe (cm)              | LeLFLo       | CO_334:0000391                             | Quantitative             |
| 9                                    | Width of leaf lobe (cm)               | WLFLo        | CO_334:0000165                             | Quantitative             |
| 10                                   | Ratio of lobe length to width         | RatioLfLoLe  |  | Quantitative             |
| 11                                   | Lobe margin                           | LoMar        | CO_334:0000392                             | Qualitative              |
| 13                                   | Colour of leaf vein                   | CoLFVein     | CO_334:0000081                             | Qualitative              |
| 14                                   | Orientation of petiole                | OrPet        | CO_334:0000259                             | Qualitative              |
| 15                                   | Flowering                             | Flow         | CO_334:0000111                             | Qualitative              |
| 16                                   | Pollen                                | Poll         | CO_334:0000260                             | Qualitative              |
| <i>Descriptors scored at 9 MAP</i>   |                                       |              |  |                          |
| 17                                   | Prominence of foliar scars            | PromLfScar   | CO_334:0000105                             | Qualitative              |
| 18                                   | Colour of stem cortex                 | CoStemCor    | CO_334:0000261                             | Qualitative              |
| 19                                   | Colour of stem epidermis              | CoStemEp     | CO_334:0000262                             | Qualitative              |
| 20                                   | Colour of stem exterior               | CoStemEx     | CO_334:0000062                             | Qualitative              |
| 21                                   | Distance between leaf scars           | DisBLFScars  | CO_334:0000397                             | Qualitative              |
| 22                                   | Growth habit of stem                  | GHStem       | CO_334:0000104                             | Qualitative              |
| 23                                   | Colour of end branches of adult plant | CoEndBAP     |  | Qualitative              |
| 24                                   | Length of stipules                    | LenStp       | CO_334:0000100                             | Qualitative              |
| 25                                   | Stipule margin                        | StpMar       | CO_334:0000265                             | Qualitative              |
| <i>Descriptors scored at harvest</i> |                                       |              |  |                          |
| 26                                   | Fruit                                 | Frt          |  | Qualitative              |
| 27                                   | Seeds                                 | Seed         |  | Qualitative (invariable) |
| 28                                   | Plant height (cm)                     | PltHt        | CO_334:0000018                             | Quantitative             |
| 29                                   | Height to first branching (cm)        | HtFB         | CO_334:0000106                             | Quantitative             |
| 30                                   | Levels of branching                   | LevBranch    | CO_334:0000296                             | Qualitative              |
| 31                                   | Branching habit                       | BH           | CO_334:0000140                             | Qualitative              |
| 33                                   | Shape of plant                        | ShPLT        | CO_334:0000264                             | Qualitative              |
| 34                                   | Number of storage roots per plant     | NoStRoot     | CO_334:0000011                             | Qualitative              |
| 35                                   | Number of commercial roots per plant  | NoComRoot    | CO_334:0000169                             | Qualitative              |
| 36                                   | Extent of root peduncle               | ExtRtPed     | CO_334:0000050                             | Qualitative              |
| 37                                   | Root constrictions                    | RtCon        | CO_334:0000446                             | Qualitative              |
| 38                                   | Root shape                            | RtSh         | CO_334:0000020                             | Qualitative              |
| 39                                   | External color of storage root        | ExCoStRt     | CO_334:0000064                             | Qualitative              |
| 40                                   | Colour of root pulp (parenchyma)      | CoRtPulp     | CO_334:0000021                             | Qualitative              |
| 41                                   | Colour of root cortex                 | CoRtCor      | CO_334:0000115                             | Qualitative              |
| 42                                   | Cortex: ease of peeling               | EasePeel     | CO_334:0000362                             | Qualitative              |
| 43                                   | Texture of root epidermis             | TexRtEpi     | CO_334:0000455                             | Qualitative              |

**Table 1** (continued)

| Descriptor number <sup>a</sup> | Trait                      | Abbreviation | Term ID from Cassava ontology <sup>b</sup> | Data type    |
|--------------------------------|----------------------------|--------------|--|--------------|
| 44                             | Root taste                 | RtTaste      | CO_334:0000085                             | Qualitative  |
| 45                             | Cortex thickness           | CorThickP    | CO_334:0002006                             | Qualitative  |
| 48                             | Harvest Index              | HI           | CO_334:0000015                             | Quantitative |
| 48a                            | Root Fresh Weight (kg)     | RootFWt      | CO_334:0000131                             | Quantitative |
| 48b                            | Above ground Weight (kg)   | Biomass      | CO_334:0000016                             | Quantitative |
| 50                             | Post-harvest deterioration | PHD          | CO_334:0000077                             | Qualitative  |

<sup>a</sup>According to Fukuda et al. (2010)

<sup>b</sup>Available in Cassavabase (<https://www.cassavabase.org>)

### Analysis of phenotyping data

Before the agro-morphological data was analysed, duplicates as identified through genetic fingerprinting (genotyping) (see below) were removed, leaving a dataset of 46 unique landraces. Initially a univariate analysis was performed. The quantitative (continuous) data were analyzed using the linear mixed model. For each trait, both clone and replication were considered as random effects. Best Linear Unbiased Predictions (BLUPs) of clone and associated standard errors derived from the linear mixed model were computed. Heritability for each trait was estimated using the method of Cullis et al. (2006) and all traits with a heritability less than 0.1 were excluded from the analysis. The relative frequency distribution of qualitative data that lists the number of occurrences for each trait level was computed and visualized.

A two-step multivariate clustering analysis was then adopted, for both traits and clone. Step 1 was a dimension reduction, i.e., reducing a mixture (high number) of qualitative and quantitative traits to a small number of quantitative synthetic variables, each of which is a linear combination of the original set of traits. A hierarchical clustering algorithm was then implemented, based on PCAMIX (Kiers 1991), a generalization of Principal Component Analysis (PCA) and Multivariate Correspondence Analysis (MCA). All the traits linked to a synthetic variable are classified in the same cluster, which is considered to be homogeneous. Step 2 was to group the clones using the small set of synthetic variables using the hierarchical agglomerative criterion of Ward. The optimal number of clusters was determined using 30 indices to identify the best clustering scheme by varying all

combinations of number of clusters, distance measures, and clustering methods.

Quantitative univariate analyses were carried out in R, Version 4.0.5 (R Core Team 2018) using ASReml-R (v4.1) (Gilmour et al. 2014) for fitting the mixed models while the relative frequency distribution was calculated for each of the qualitative traits and visualized in R using the tidyverse package (Wickham et al. 2019). Multivariate clustering analysis were performed in R (R Core Team 2018) using the ClustOfVar package (Chavent et al. 2017) for the clustering of the traits, the NbClust package (Charrad et al. 2014) for the optimal number of clusters of the clones, and the factoextra package (Kassambara and Mundt 2020) for plotting the dendrograms.

### Genotyping

DaRTSeq SNP genotyping was conducted to (1) identify duplicates within collections made in Comoros, (2) determine the identification of Comoros landraces using an existing set of references, and (3) determine the relationship of Comoros landraces to those from Tanzania (Ferguson et al. 2021). For Comoros germplasm, four 6 mm diameter leaf disks were sampled from young leaf tissue of one plant per cultivar from the field genebank at INRAPE and placed in a tube in 96 well format and dried in an oven overnight at 45 °C. Samples were shipped to Intertek, Adelaide, South Australia, for DNA extraction and forwarding to Diversity Array Technologies (DaRT), Canberra, Australia, for DaRTSeq™ SNP genotyping (Sansaloni et al. 2011). Existing genotyping data was used for the other genotypes included from

Tanzania, which were co-analysed with new data from Comoros.

#### Analysis of genotyping data

To identify duplicates in the Comoros collection and determine any matches with known landraces and breeding lines, genotyping results of 76 genotypes with prefix FMO, collected in Comoros were analysed together with 34 landraces (Online Resource 1) and 96 breeding lines (Online Resource 2) which included 21 known duplicates (7 breeding lines and 14 landraces) (a total of 214 entries). To combine data from Comoros germplasm and the reference genotypes, raw reads were combined by DArT and SNPs were called on the combined data set. All SNPs with more than 1% missing data were removed from the analysis, together with all monomorphic loci and genotypes with more than 5% missing data. Data was not filtered for minor allele frequency. Euclidean distances was calculated among clones to determine genetic distance which were then visualized as a dendrogram. This was converted to a similarity by subtracting from 1. Duplicates were identified, based on a cut-off of greater than 85% similarity. This was determined from the lowest similarity of 21 known duplicates which was Mumba (0.852). To select the most representative from a set of FMO duplicates, the one with the highest mean similarity was selected. For duplicates with two clones, the average will be the same so one was selected at random.

The correlation of the distance matrix generated from morphological data with that from genotypic data was calculated using the Mantel test.

#### *The relationship among islands according to cassava cultivars*

To determine the relationship of clones among islands a Discriminant Analysis of Principle Components (DAPC) using 50 principal components and two eigenvalues was conducted. Population structure was further described using the Weir and Cockerham (1984) overall and/or pairwise estimators of Wrights F statistics ( $F_{st}$  and  $F_{is}$ ). The fixation index  $F_{st}$  is a measure of the identity of individuals within subpopulations compared to individuals from other subpopulations within the total population or the total homozygosity due to the Wahlund effect. It is thus a

measure of differentiation between subpopulations.  $F_{st} = 1 - (H_s/H_t)$ , where  $H_s$  is average of intra-population genetic diversity and  $H_t$  is genetic diversity across populations considered as a single population (total diversity). According to Wright:  $0 < F_{st} < 0.05$  is weak differentiation;  $0.05 < F_{st} < 0.15$  is moderate differentiation;  $0.15 < F_{st} < 0.25$  is significant differentiation; and  $F_{st} > 0.25$  is very important differentiation (Wright 1965).

$F_{is}$  measures the identity (or homozygosity) of alleles within individuals within sub-populations as compared to Hardy–Weinberg expectations, it is thus a measure of deviation from random mating. Finally,  $F_{IT}$  is a measure of homozygosity of individuals in the total population and thus measures the deviation from Hardy–Weinberg due to local deviation from panmixia and Wahlund effect. In addition, diversity was described using observed heterozygosity ( $H_o$ ) which is the proportion of heterozygous individuals in a population, and expected heterozygosity ( $H_e$ ), also known as Nei's unbiased estimator of gene diversity, both calculated by population and overall.

#### *The relationship of Comoros landraces to those from Tanzania*

For the purpose of determining the relationship of Comoros landraces to those from Tanzania, SNP data from 402 Tanzanian genotypes, published by Ferguson et al. (2021) was used together with the dataset of 76 genotypes from Comoros, which included the duplicates. This resulted in a dataset of 478 genotypes. Again, to combine datasets, raw reads were aligned by DArT and SNPs were called on the combined data set. All data filtering was as described above.

Genotypes were assigned to one of 11 regions, depending on where they were collected; these were; Mwali (16 samples), Ndzuani (19) and Ngazidja (41) Islands of the Comoros, the Lake (58), Western (30), Central (61), Southern Highlands (79), Coastal (53), Northern (65) Zones of Tanzania and the Islands of Zanzibar (33) and Pemba (20) in Tanzania. As described above, the relationship between clones was illustrated using a dendrogram based on euclidean distance. In addition, a Discriminant Analysis of Principle Components (DAPC) was conducted to visualize the relationship between populations. Nei's distance (Nei 1972, 1987) was calculated based on

geographical location and a dendrogram constructed. Analysis of genotypic data was carried out in R, Version 4.0.5 (R Core Team 2018). The relationship between unique clones was explored using a dendrogram based on euclidean distance using DaRTR (Gruber et al. 2018) package with the tree constructed using hclust in phyclus package (Chen 2010; Chen et al. 2013) and plotted using the package ggtree (Yu et al. 2017). The Discriminant Analysis of Principle Components (DAPC) was performed using adegenet package (Jombart 2008) and the population structure described using hierfstat package (De Meeûs and Goudet 2007; Goudet and Jombart 2021). Nei's distance (Nei 1972, 1987) was calculated based on geographical location using heirfsat package (De Meeûs and Goudet 2007). The correlation of the distance matrix generated from morphological data with that

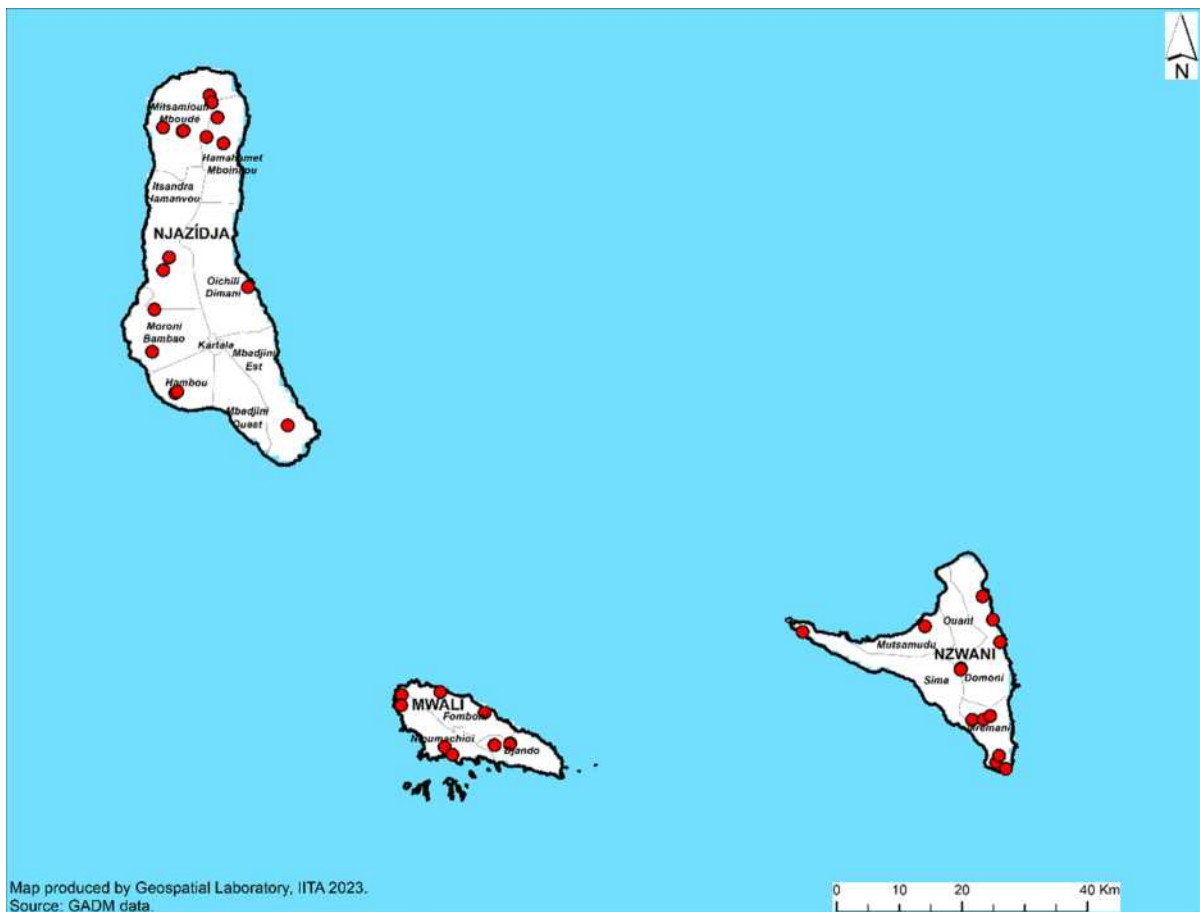
from genotypic data was calculated using the Mantel test using the ade4 library (Bougard and Dray 2018).

## Results

### Collection missions

#### *Collection sites*

Cassava landraces were collected from 34 production sites across the three islands of Ngazidja, Ndzouani and Mwali (Fig. 1). The average altitude of farms was 471 m asl in Ngazidja, 375 m asl in Ndzouani and 161 m asl in Mwali. A full report on the collection mission (in French) can be found in Online Resource 3.



**Fig. 1** Location of collection sites on three Islands in the Comoros, Ngazidja, Ngazidja and Mwali

### *Profiles of surveyed farmers and their farms*

Seventeen cassava farmers (key informants) were surveyed across the country, including nine from the island of Ngazidja, five from Ndzouani and three from Mwali (Fig. 1). Men represented 88% of farmers surveyed in Ngazidja, 100% in Ndzouani and 33% in Mwali. The age of key informants varied between 30 and 80 years old. The majority of people interviewed (82%) were heads of households who made the decision on which of each cassava cultivar to grow, and the quantity. They were selected by the extension personal based on their knowledge on the history of cassava cultivation in their respective areas.

Forty-seven percent of farmers had grown cassava for between 5 and 19 years, the same proportion had grown the crop from between 20 and 49 years and six percent had grown cassava for more than 50 years. Forty-one percent of key informants had no level of education and depended on agriculture as their main source of income, 23.5% of farmers had a primary education level and also depended on agriculture as a main source of income. An equal number of farmers had secondary education and also depended on agriculture as their main source of income and only 12% of the farmers interviewed had a higher level of education and a formal job that provided them with a main income. All of the farmers surveyed had less than one hectare of agricultural land and cultivated between 50 and 100 cuttings of different cultivars of cassava. All farmers practiced intercropping of cassava with others crops such as corn, peanuts, taro and bananas.

### *The local cultivars collected*

Among the 79 local cultivars collected, 41 were from the island of Ngazidja, 21 from Ndzouani and 17 from Mwali (Online Resource 4). In general, the cultivars collected had local names. The translation in French of some of them indicated certain characteristics of the cultivar such as "Mdja" which means "comes early", "Mhogomoudou" which means "black manioc", "Katsayidoudou" which means "dry and bitter". These local cultivars were named by the farmers according to their morphological or agronomic traits and, in some cases, in according to the person who introduced them to the region or according to the place where they originated. Some of the local

cultivars had more than one name even within the same region. Ten of the 79 cultivars collected did not have a name, so were assigned a name according to the name of the farmer who supplied them, for example Djaf 1 and Djaf 2. It was estimated that 20% of the collected cultivars arrived in the Comoros before 2010, 34% were estimated to have arrived in the various regions after 2010 and the remaining 46% had an unknown time of arrival in Comoros.

### *Increase, decrease or rejection of the cultivation of cassava cultivars*

Over the past five years, the cultivation of 36 cultivars among the 79 cassava cultivars collected (45.5%) had increased in the different regions of the country. Reasons for this included higher yield, earliness in production, improved taste quality and possibly resistance to diseases. In addition, 38 cassava cultivars saw their production stagnate or decreased in the different regions because of their low yield, their high sensitivity to diseases and their long production period. The production trend of five cultivars among the 79 collected during the last five years was not known by the farmers surveyed.

In almost all the regions of the Comoros archipelago, the different cultivars of cassava were grown for sale and personal consumption from the leaves and storage roots. No other use was mentioned by the farmers surveyed. It is reported that 44.6% of the cultivars collected were sweet types that are often consumed directly after boiling in water, and the same number were intermediate in taste. The farmers surveyed reported the existence of three bitter cultivars (Katsa yidoudou, Yidoudou and Maweni Med 1) which were processed by drying in the sun and making into flour before consumption.

### *The yield potential of local cultivars*

Yield potential was assessed in terms of the number of plants that would need to be harvested to fill a 25 kg bag. Thirty-nine percent of cultivars were in the highest yielding category, in which two to eight cassava plants would be needed to fill a 25 kg bag, while 20% of the cultivars collected needed 10 to 19 plants to be harvested. Fourteen percent of the cultivars collected had a low yield, where it was necessary to harvest between 20 and 40 plants of cassava



to fill a 25 kg bag. The yield was very poor for 3% of cultivars collected where it was necessary to harvest more than 40 plants to fill a 25 kg bag. For 24% of the cultivars collected, farmers have no estimate of their yield. Duration of storage of cassava in the soil after maturation varied from between 6 months and 5 years, however the majority of farmers indicated that the taste decreased with increasing storage period in the soil.

#### *The sensitivity of landraces to biotic and abiotic stresses*

It was indicated that 83.5% of cultivars collected performed well under conditions of drought, and yield did not change substantially. For the remaining cultivars, farmers had no idea about their ability to tolerate drought. In terms of maturity, farmers classified the cultivars collected into three groups: 41 were considered early cultivars, whose maturity period varied between 6 and 8 months; 16 were considered medium maturity taking 9–12 months to reach maturity, and 14 were considered late maturing, taking 13–24 months to reach maturity. For the remaining eight cultivars, the farmers had no idea about their maturity period.

At the time of collection, according to observations of a participating cassava breeder, 59% of cassava landraces did not show CMD symptoms, while 72% of landraces did not show CBSD symptoms. The incidence of CMD was higher on the island of Ngazidja, than on Ndouani and Mwali, with CBSD being observed exclusively on Ngazidja. The majority of farmers were not able to easily recognize the symptoms of these two diseases. Farmers had noticed that the sensitivity to CMD and CBSD was more pronounced at lower elevations. They also reported the presence of parasites such as green mites and mealybugs.

#### *Culinary qualities and market demand*

In terms of the perceptions of cooking quality, 68% of collected landraces were considered to be good, with a light colour and a short cooking period. Ten cultivars (13%) were classified by farmers as hard, requiring a very long time to cook. They were frequently dried and stored for specific preparations

with coconut milk. The remaining 15 cultivars (19%) could not be assessed in terms of cooking quality.

A market demand analysis was carried out based on knowledge of selling different cultivars of cassava at the market. The scale varied from “cultivars in high demand” to “demand unknown”. After consolidation of cultivars based on genotyping results which identified identical clones, six cultivars were identified as being in particularly high demand, based on high demand of all synonyms. These were Trewe\_FMO73 (synonyms FMO32, -41, -7 and -75, collected on Mwali and Ngazidja. HS52\_FMO7 was identified as being in exceptional demand), Mze Hamidou\_FMO37 (synonyms FMO27, -44, -56 collected on Ngazidja and Ndouani), Mdja\_FMO1 (synonyms FMO19, -63, -50, collected on all three islands), H57\_FMO13 (synonym FMO21, from Ngazidja), Mlaregno 16\_FMO40 (synonym FMO47 from Ngazidja and Ndouani) and Mlaregno\_FMO49 (synonym FMO48 from Ndouani) (Online Resource 5). In addition, Namanga\_FMO8 with no synonyms was in exceptional demand in Ngazidja, and an additional 10 cultivars with no synonyms in high demand (Online Resource 5) across other islands. The presence or absence of disease symptoms at the time of collection is indicated in Online Resource 5. Interestingly CBSD was only observed on Ngazidja, and not in the other two inlands. Of note was H57\_FMO13 (synonym FMO21, both from Ngazidja) which was in high demand and recognized as being tolerant to both CMD and CBSD. In addition, Mlaregno 16\_FMO40 (synonym FMO47 from Ngazidja and Ndouani) was consistently scored tolerant to both CMD and CBSD. The most popular cultivars were sweet and softer such as H57\_FMO13 and Mlaregno 16\_FMO40 (and their synonyms), whereas the least requested were hard and bitter cultivars such as Maweni Med 2\_FMO22.

#### *Conservation practices for germplasm*

To ensure the conservation of local cultivars, farmers grow cassava in several plots, and plant at several times throughout the year. This method of conservation is supplemented by sharing or exchange of genetic material between farmers in the same region and sometimes even different regions or islands. Other farmers conserve cassava cultivars by delaying the harvest until the next planting season. Exchanging cassava cuttings as a gift is a common practice,

especially in Ngazidja, however this does promote the spread of disease when exchanged cuttings are infected. On the other two islands planting material is mostly purchased. This offers the possibility of being able to choose the cultivars which are free from disease and parasites.

### Analysis of agro-morphological / phenotypic data

Establishment of cuttings at the INRAPE station in M'Be ranged from zero to 100%, with an average of 88%.

### Qualitative traits

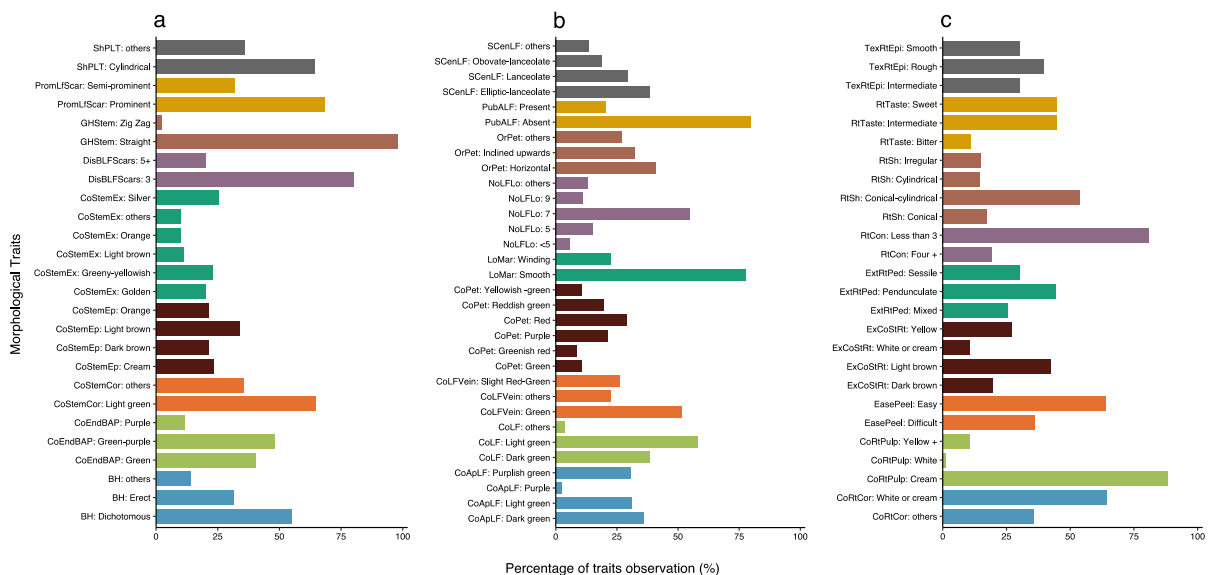
The percentage distribution of scores in qualitative trait categories is given in Fig. 2a–c for stems, leaves and roots respectively. The most frequent scores for each leaf trait were elliptic-lanceolate light green leaves with seven lobes, green veins and smooth margins. Petioles were most frequently horizontal and red and apical leaves dark green without pubescence.

Stems were most frequently straight, cylindrical with prominent leaf scars which were fairly close to one another. Stems were silver with a light brown epidermis and a light green cortex. End branches were most frequently green-purple in colour. The branching habit of the plants was mostly dichotomous, followed by erect.

Most storage roots were conical-cylindrical in shape, light brown in colour and pedunculate with few or no constrictions. The texture of the storage root epidermis was mostly rough. The storage roots were mostly easy to peel, with a cream root pulp (parenchyma) and white or cream cortex with a sweet or intermediate taste in equal proportion.

### Quantitative traits

Broad-sense heritability, mean and standard error of 11 quantitative traits are given in Table 2.



**Fig. 2** The percentage distribution of scores in qualitative trait categories for **a** stems, **b** leaves and **c** roots. Abbreviations for the traits scored are according to Table 1. **a** ShPLT, shape of plant; PromLScar, prominence of leaf scar; GHStem, growth habit of stem; DisBLFScars, distance between leaf scars; CoStemEx, colour of stem exterior; CoStemEp, colour of stem epidermis; CoStemCor, colour of stem cortex; CoEndBAP, colour of end branches of adult plant; BH, branching habit. **b** SCenLF, shape of central leaflet; PubALF, pubescence

of apical leaf; OrPet, orientation of petiole; NoLFLo, number of leaf lobes; LoMar, lobe margin; CoPet, colour of petiole; CoLFVein, colour of leaf vein; CoLF, colour of leaf; CoApLF, colour of apical leaf. **c** TexRtEpi, texture of root epidermis; RtTaste, taste of storage root; RtSh, root shape; RiCon, root constrictions; ExRtPed, extent of root peduncle; ExCoStRt, exterior colour of the storage root; EasePeel, ease of peeling; CoRtPulp, colour of the root pulp; CoRtCor, colour of the root cortex

**Table 2** Heritability, mean and standard errors of the quantitative traits

| Clone | Biomass | CorThickD | CorThickM | CorThickP | HI   | HtFB | LeLFLo | WLFLo | RatioLfLoLe | PltHt | RootFWt |
|-------|---------|-----------|-----------|-----------|------|------|--------|-------|-------------|-------|---------|
| H2    | 0.489   | 0.20      | 0.00      | 0.03      | 0.31 | 0.00 | 0.52   | 0.58  | 0.73        | 0.47  | 0.16    |
| Mean  | 1.41    | 3.86      | 3.74      | 2.15      | 0.31 | 1.35 | 16.01  | 3.69  | 4.72        | 2.19  | 0.80    |
| SE    | 0.28    | 0.39      | 0.13      | 0.19      | 0.05 | 0.56 | 1.83   | 0.45  | 0.89        | 0.40  | 0.18    |

Where CorThickD=Root cortex thickness (distal end); CorThickM=Root cortex thickness (middle or root); CorThickP=Root cortex thickness (proximal end); HI=Harvest index; HtFB=Height of first branch; LeLFLo=Length of leaf lobe; WLFLo=Width of leaf lobe; RatioLfLoLe=Ratio of lobe length to width; PltHt=Plant height; RootFWt=Root fresh weight

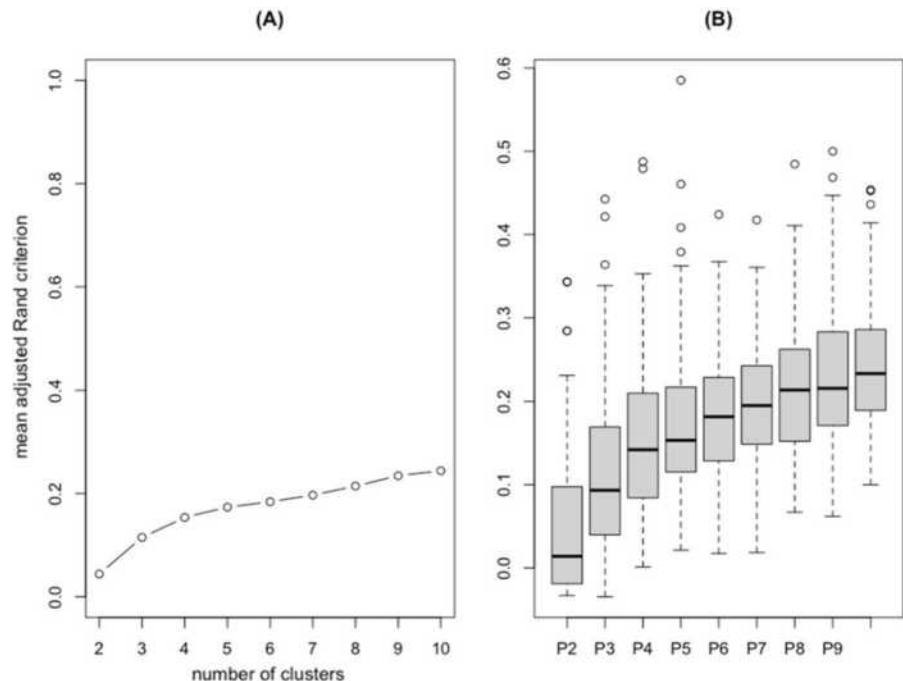
### Qualitative and quantitative data combined

Optimal partitioning of all agro-morphological traits into clusters suggested seven clusters. This is illustrated in Fig. 3a which shows an optimal partition with increasing clusters number based on aggregation distances. This is confirmed with the dispersion of the adjusted rand index (Fig. 3b) with a lower dispersion from seven clusters. Agro-morphological traits clustered according to this grouping is illustrated in Fig. 4. This dendrogram can be “cut” with a horizontal line at a height with maximum distance up and down, indicating seven groups of traits. From this grouping, seven synthetic variables were generated. As expected, root weight and harvest index were closely associated as were number of storage roots and number of commercial roots, shape of plant and

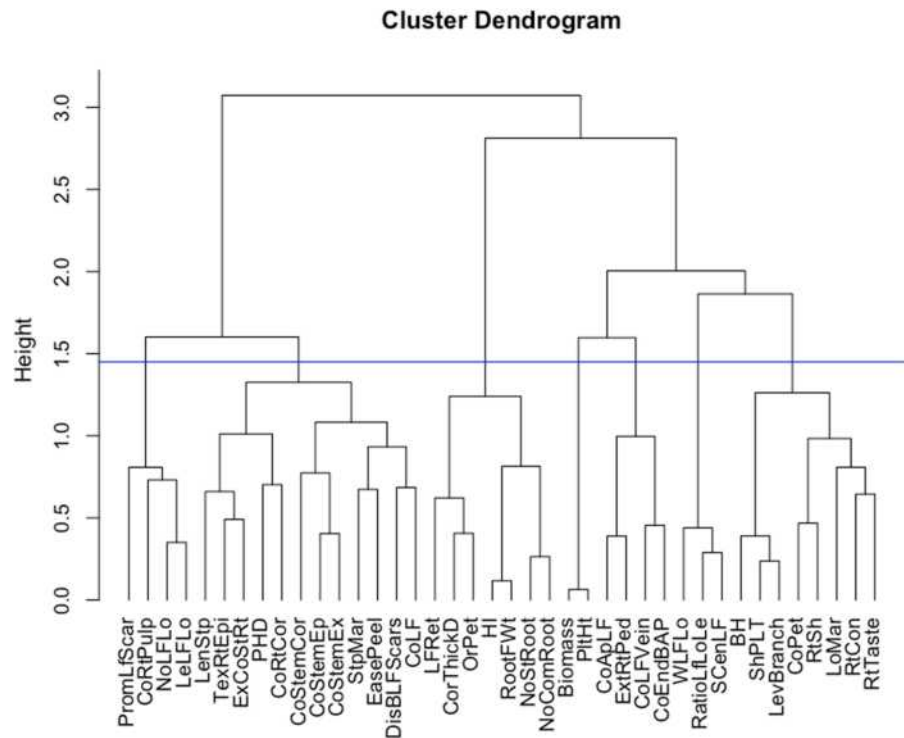
levels of branching, number and length of leaf lobes, shape of central leaflet and the ratio of the leaflet width and length, plant height and biomass.

The seven synthetic variables were used to group the clones into four clusters (Fig. 5). The indices used for cluster validation equally suggested two, four or 10 clusters of clones. Four groups were considered using the same principle of a horizontal line on the dendrogram that have a maximum distance up and down at four clusters. Of the genotypes from Ngazidja, 12% were from Cluster 1, 63% from Cluster 2, 17% from Cluster 3 and 8% from Cluster 4. Of the genotypes from Ndzouani, 10% were from Cluster 1, 30% from Cluster 2, 20% from Cluster 3 and 40% from Cluster 4. Of the genotypes from Mwali, none were from Cluster 1, 42% from Cluster 2, 8% from Cluster 3 and 50% from Cluster 4. Cluster 2 had

**Fig. 3** Stability of the partitions, mean adjusted Rand index from 100 bootstrap (A) and dispersion of the adjusted Rand index (B)



**Fig. 4** Relationship among quantitative and qualitative traits, from which seven synthetic variables were derived



a good representation of genotypes from each island, whereas Cluster 4 had good representation from Ndzuani and Mwali.

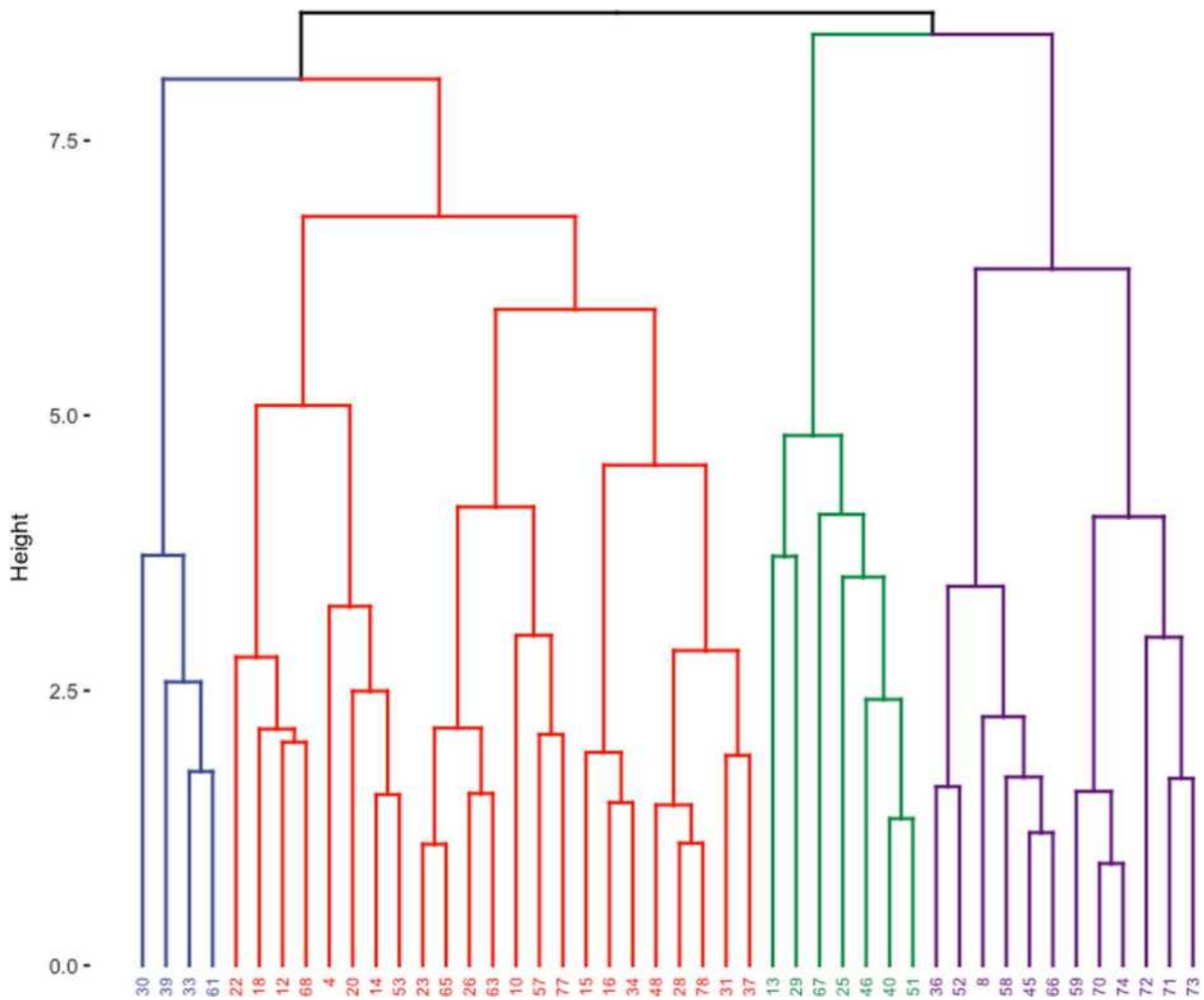
### Genotyping

#### *Identification of duplicates and identity based on known landraces and breeding lines*

Genetic fingerprints were obtained from genotyping data of 76 of the 79 cassava samples collected. Initially calls at 42,804 SNP positions were obtained, but after data cleaning, all the 76 entries remained (Maximum missing value within any genotype was 1.3%) and 14,071 SNPs. The lowest similarity value of 21 known duplicates included in the analysis was of two Mumba entries with a similarity of 0.852, so a cut-off of  $\geq 0.85$  was used to define duplicates. Based on this, a set of 46 unique genotypes was defined within the 76 FMO genotypes from Comoros, and the only landrace or breeding line included in a panel of known clones, that matched with landraces from

Comoros, as a duplicate was Aipin Valenca (a very old cultivar, and variant of Namikonga, with wild species characteristics, and possibly some CBD resistance) (Cluster 1 containing FMO1, 19, 50 and 63). Duplicates are listed in Online Resource 6 together with the most representative clone of a duplicate group. The relationship between unique FMO clones is illustrated in the dendrogram, based on Euclidean distance, in Fig. 6. The distance matrix of morphological data has a positive and significant relationship with that of genotypic data (Mantel statistic: 0.19,  $p$  value = 0.02), indicating that as clones become more dissimilar in morphology, they also become more dissimilar in terms of genotypic information.

Four main clusters and two smaller clusters consisting of FMO10, 12 and 25 in one, and FMO66 and 33 in the other are evident from Fig. 6, and identified by colour. All these clones are from the island of Ngazidja except FMO66 which is from Mwali. The relationship between 76 clones by island, is also given in a Discriminant Analysis of Principle Component (DAPC) plot in Fig. 7. This relationship is



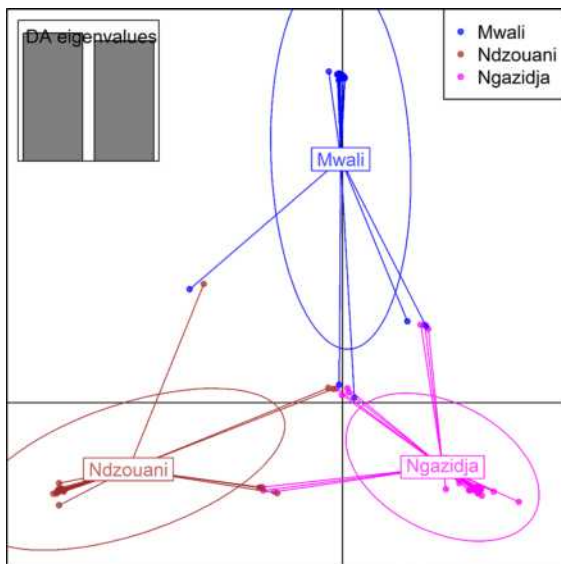
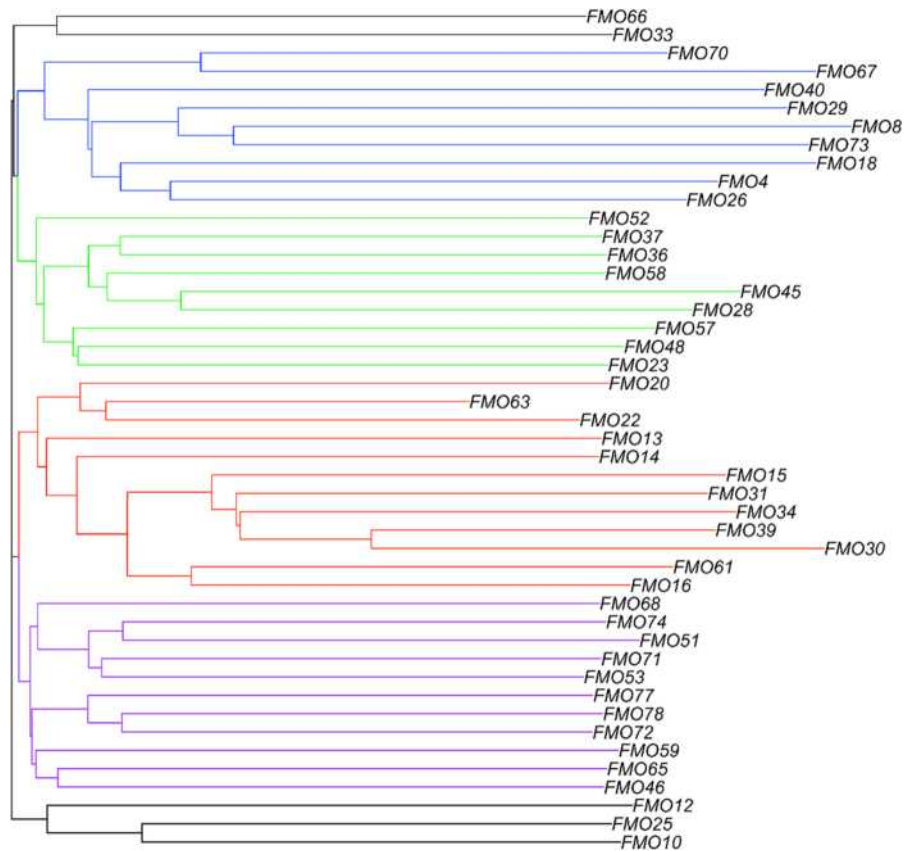
**Fig. 5** Relationship between 46 unique landraces, once duplicates had been removed, based on quantitative and qualitative agro-morphological traits. Numbers refer to FMO number

supported by the unbiased estimator of the Wright's  $F$  statistics,  $F_{st}$  (Wright 1965) in which Ngazidja is similarly related to both Mwali ( $F_{st}=0.0212$ ) and Ndouzani ( $F_{st}=0.0219$ ) with Mwali and Ndouzani slightly more distantly related ( $F_{st}=0.0234$ ). This is similar to Nei's pairwise genetic distance (Nei 1972, 1987) in which Mwali and Ngazidja are most closely related (0.0219) followed by Ngazidja and Ndouzani (0.0224) with Mwali and Ndouzani being most distantly related at 0.0234. This relationship is also

evident in the dendrogram in Fig. 8. An  $F_{st}$  less than 0.05 is considered weak differentiation.

Diversity indices among Islands is provided in Table 3, calculated in Hierfstat package (De Meeûs and Goudet 2007; Goudet and Jombart 2021) in R. The mean observed heterozygosity ( $H_o$ ) of Ndouani is slightly higher than the other two islands. The within population gene diversity ( $H_s$ ) and overall gene diversity ( $H_t$ ) are greatest in Ngazidja with the largest number of samples, compared to Ndouani and

**Fig. 6** Relationship of 46 unique or most representative cultivars from a cluster of cassava duplicate clones from Comoros, based on Euclidean distance and SNP data. Branch lengths are not representative of distance. Different colours define the six main clusters



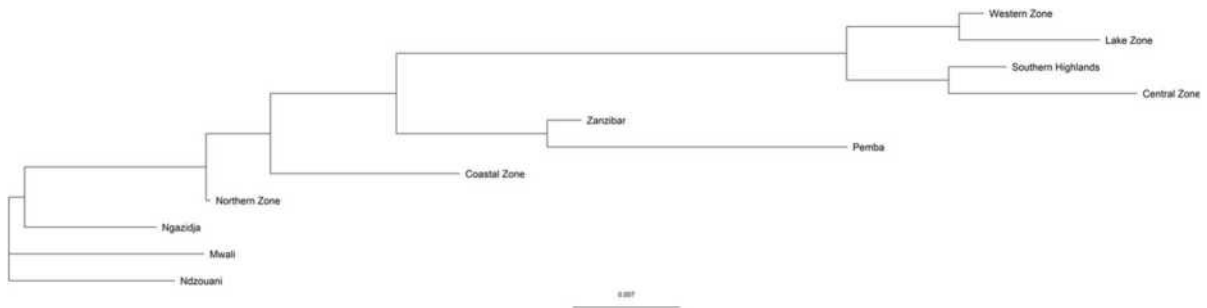
**Fig. 7** Discriminate analysis of principle coordinate analysis, showing relationship among 76 collected clones from different islands

Mwali. The amount of gene diversity among samples was also highest in Ngazidja, followed by Mwali, and was very low in Ndouani (0.002).

#### *Relationship with Tanzanian landraces*

After filtering, a total of 13,954 SNP loci remained in the dataset of 475 genotypes. The only match of Comoros genotypes with any Tanzanian genotypes was FMO63 and BKP56 which also matched with Aipin Valenca, which is an extremely old cultivar, with a distance of 0.024.

Germplasm from Tanzania and Comoros was analysed based on 11 geographical areas, according to Nei's distance, and the result given in Fig. 8, Table 4. The relationship of individual landraces, colour-coded according to region is provided in a DAPC plot in Fig. 9, and circular dendrogram in Fig. 10. From these representations, it is clear that cassava germplasm from Comoros is quite distinct from that from Tanzania. It is most closely related to germplasm from the Northern Zone (which includes the northern



**Fig. 8** Relationship of landraces from different regions of Tanzania and the Comoros

**Table 3** Diversity indices by island

| Sub-population | N  | Ho     | Hs     | Ht     | Dst    | Fst    | Fis      |
|----------------|----|--------|--------|--------|--------|--------|----------|
| Mwali          | 16 | 0.2860 | 0.2323 | 0.2410 | 0.0087 | 0.0361 | - 0.2313 |
| Ndzouani       | 19 | 0.2974 | 0.2626 | 0.2628 | 0.0002 | 0.0009 | - 0.1323 |
| Ngazidja       | 41 | 0.2848 | 0.2694 | 0.2853 | 0.0159 | 0.0558 | - 0.0574 |
| Overall        | 76 | 0.2823 | 0.2598 | 0.2638 | 0.0040 | 0.0151 | - 0.0866 |

Ho = mean observed heterozygosities

Hs = The within population gene diversity (sometimes misleadingly called expected heterozygosity):

Ht = overall gene diversity

Dst = Ht - Hs The amount of gene diversity among samples

Fst = Dst/Ht (This is not the same as Nei's Gst, Nei's Gst is an estimator of Fst based on allele frequencies only)

Fis = 1 - Ho/Hs

coastal area) and the Coastal Zone of Tanzania, rather than the islands of Zanzibar and Pemba. The distinction between that and germplasm from the Lake, Western, Central and Southern Highlands Zones of Tanzania is clear.

Interestingly, cassava landraces from the Islands of the Comoros showed lower levels of heterozygosity (Ho) (lowest 0.2079) than those from the Tanzanian mainland (highest 0.2342 in Western Zone), and similar to that from the island of Zanzibar (0.2123) (Table 5). Similarly, within population gene diversity (Hs) and overall gene diversity (Ht) was lower on the islands, including Comoros, as well as Pemba and Zanzibar.

## Discussion

Landraces of cassava have provided a reliable food source for generations of people living on the Islands of the Comoros. In anticipation of the introduction

of higher yielding, more disease resistant cultivars, and due to the threat of current viral pressure, INRAPE, in collaboration with IITA, collected 79 landraces of cassava for conservation and breeding. By far the largest number of landraces (41) were collected from Ngazidja in 13 production sites, 21 cultivars were collected in Ndzouani in 12 sites and 17 cultivars were collected in Mwali in 9 sites. Most of the cultivars had local names given according to their characteristic traits, history such as the person who introduced them to the area or depending on where they originated. Interestingly, most of the producers were men, aged between 30 and 80 years, who grew cassava roots and leaves for personal use and sale, with a strong preference for sweet cultivars. Cultivars with sweet, soft storage roots had the highest market demand. Of note were two cultivars, H57\_FMO13 and Mlaregno 16\_FMO40 (and their synonyms) which were in high market demand and consistently scored tolerant to both CMD and CBSD during collection. These cultivars should be further

**Table 4** Nei's genetic distance between 11 geographical regions in Comoros and Tanzania

|                        | Central Zone, Tz | Coastal Zone, Tz | Lake Zone, Tz | Mwali  | Ndzouani | Ngazidja | Northern Zone, Tz | Pemba, Tz | Southern Highlands, Tz | Western Zone, Tz |
|------------------------|------------------|------------------|---------------|--------|----------|----------|-------------------|-----------|------------------------|------------------|
| Central Zone, Tz       |                  |                  |               |        |          |          |                   |           |                        |                  |
| Coastal Zone, Tz       | 0.0711           |                  |               |        |          |          |                   |           |                        |                  |
| Lake Zone, Tz          | 0.0242           | 0.0764           |               |        |          |          |                   |           |                        |                  |
| Mwali                  | 0.0992           | 0.0411           | 0.0901        |        |          |          |                   |           |                        |                  |
| Ndzouani               | 0.0892           | 0.0402           | 0.0821        | 0.0236 |          |          |                   |           |                        |                  |
| Ngazidja               | 0.0764           | 0.0396           | 0.0682        | 0.0215 | 0.0214   |          |                   |           |                        |                  |
| Northern Zone, Tz      | 0.063            | 0.0151           | 0.0607        | 0.0254 | 0.0215   | 0.0222   |                   |           |                        |                  |
| Pemba, Tz              | 0.0697           | 0.0444           | 0.0732        | 0.0732 | 0.0711   | 0.0607   | 0.0416            |           |                        |                  |
| Southern Highlands, Tz | 0.0161           | 0.0645           | 0.0279        | 0.0803 | 0.0778   | 0.0605   | 0.0541            | 0.0631    |                        |                  |
| Western Zone, Tz       | 0.0295           | 0.0648           | 0.0108        | 0.074  | 0.064    | 0.0574   | 0.0504            | 0.0709    | 0.0284                 |                  |
| Zanzibar, Tz           | 0.0527           | 0.0232           | 0.0591        | 0.05   | 0.0424   | 0.0377   | 0.0214            | 0.0174    | 0.0513                 | 0.0541           |

evaluated for response to CMD and CBSD with the view to both rapid propagation and distribution to farmers and incorporation into a breeding program in Comoros. On the Islands of Ndzouani and Mwali planting material was frequently purchased, indicating that the promotion of a commercial seed system for cassava might be readily accepted.

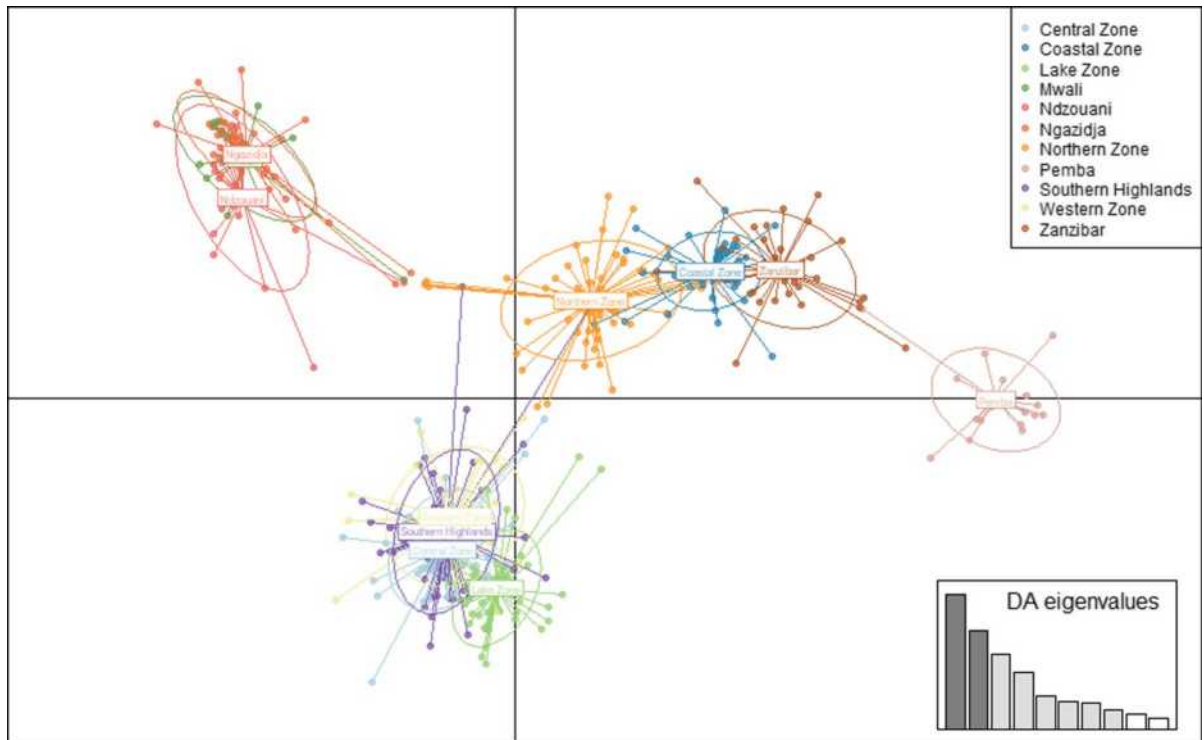
Interestingly the incidence of both CMD and CBSD was higher on the island of Ngazidja, than on Ndzouani and Mwali, despite the fact that the average altitude of collection was higher (471 m), than in Ndzouani (375 m) and Mwali (161 m). Increasing altitude has generally been found to be associated with decreased disease incidence, particularly for CBSD (Nichols 1950; Hillocks et al. 2002; Hillocks and Jennings 2003), although it has been noted that below 1000 m, the general pattern of higher incidences of CBSD at lower altitude is not necessarily true for specific sites. It is influenced by the source of planting material, whether from infected cuttings and/or the sensitivity of the cultivars grown (Hillocks and Jennings 2003). In fact, CBSD was only observed, by a cassava breeder participating in the study, on Ngazidja, and not on the other two inlands. It may be only a matter of time until it is introduced to Mwali and Ndzouani.

All agro-morphological characters were clustered into seven synthetic variables, which were then used to group the clones into four clusters. As expected, traits such as harvest index (HI) and root fresh weight (RootFWt) were closely related, as were plant height (PltHt) and biomass which consisted of one synthetic variable, and levels of branching (LevBranch) and the shape of the plant (ShPLT).

Based on genotyping data, none of the Tanzanian landraces shared identity with any of the Comoros landraces, and only one cultivar, known as Aipin Valenca, matched with four cultivars from Comoros, FMO1,19,50 and 63. Aipin Valenca is a very old cultivar, originally imported from Brazil, for use in the CMD and CBSD resistance breeding program that operated at Amani in northern Tanganyika (now Tanzania) in 1937 (Nichols 1947; Jennings 1957). It was found to be a promising parent particularly for CBSD resistance and was used as a parent in the interspecific backcrossing programme (Hillocks and Jennings 2003).

The relationship among islands is to be expected with Ngazidja forming an intermediary position

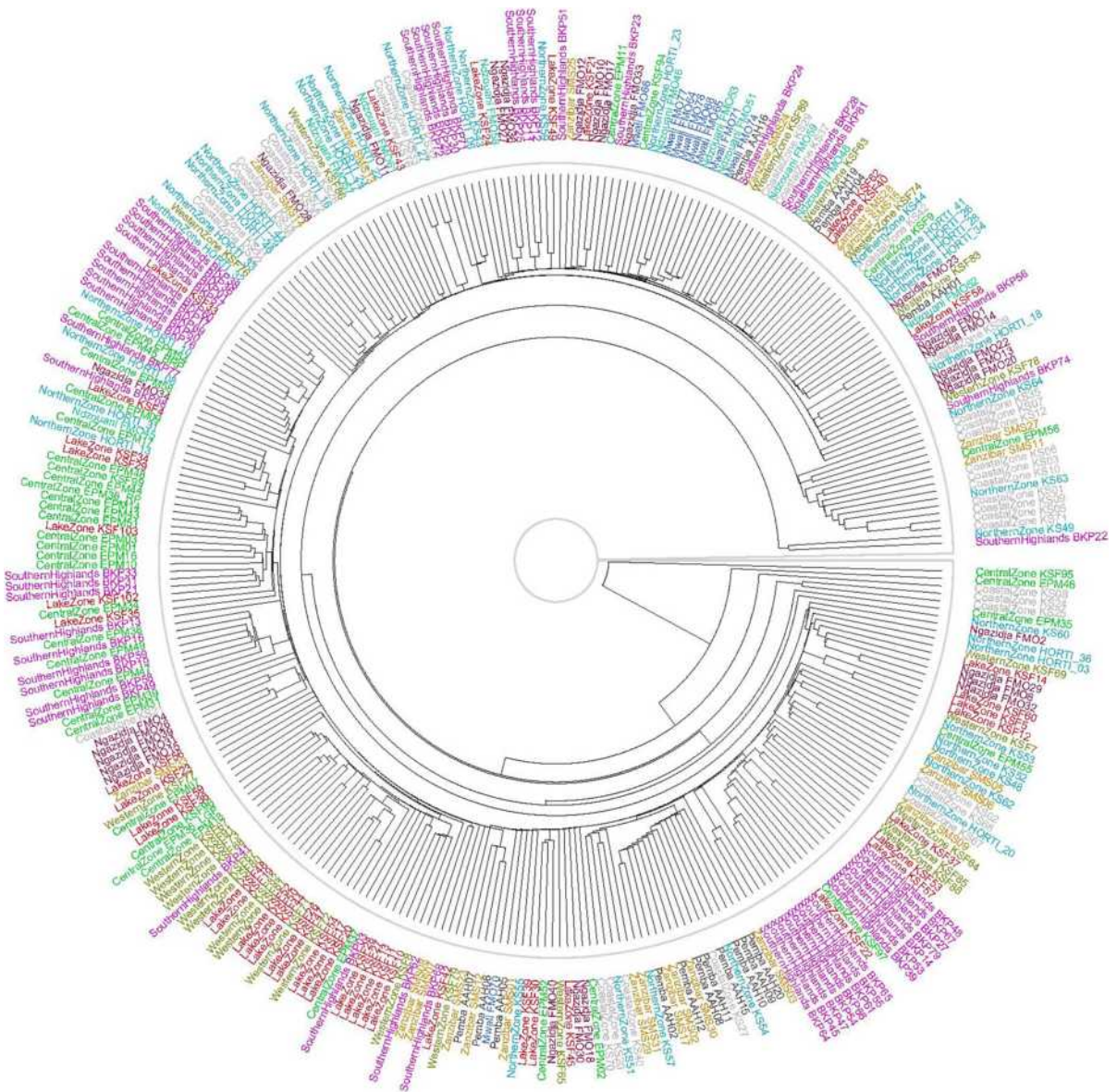




**Fig. 9** Discriminant analysis of principle components (DAPC) analysis showing relationships of landraces from different regions of Tanzania and the Comoros

between Mwali and Ndzhouani according to both Weir and Cockerham (1984) and Nei (1972). Ngazidja is the island where most landraces were collected (41 of the 76 landraces, based on genotyping data). Germplasm from all three islands was most closely related to each other than to germplasm from different Zones in Tanzania, including Pemba and Zanzibar. Relative to these Zones, cassava germplasm from the Comoros of all three islands was most closely related to that from the Northern Zone. Fascinatingly this was the Zone hosting the Amani Breeding Program and could suggest an historical transfer of germplasm from that program to the Comoros, particularly as the old cultivar Aipin Valenca was found in the Comoros. Following from the Northern Zone, germplasm from Comoros was next closely related to that from the Tanzanian Coastal Zone, and that from Zanzibar. Interestingly germplasm from Pemba was more distantly related, as was germplasm from the Lake Zone, Southern Zone, Western Zone and Central Zones of Tanzania.

Cassava landraces from the Comoros appear to demonstrate some of the characteristics in terms of diversity that we would expect from being isolated on islands with restricted germplasm movement. Observed heterozygosity ( $H_o$ ) was the lowest of all 11 regions for Ngazidja, followed by Zanzibar, then Ndzhouani and Mwali (Table 5). This indicates increased selfing, or more likely in cassava, cross-pollination amongst clonal individuals, which is effectively selfing, has increased homozygosity. In addition, the within population gene diversity ( $H_s$ ) was lowest for all the islands; Pemba (0.1727), followed by Ndzhouani and Mwali (0.1883), Zanzibar (0.1894) and Ngazidja (0.1958). Total gene diversity ( $H_t$ ) was very similar with Ndzhouani and Mwali having the lowest (0.1883), followed by Pemba, Ngazidja and Zanzibar. These low levels of diversity are likely due to restricted movement of germplasm due to the isolation of being on an island. This suggests that the Comoros should pro-actively import germplasm to enhance diversity, and the availability of improved germplasm.



**Fig. 10** Circular dendrogram showing relationships of landraces from different regions of Tanzania and the Comoros

In conclusion, cassava plays a vital role in the food security of the largely rural population of the relatively isolated Comoros Islands even though few modern cultivars are grown. This, together with the narrow genetic base of landraces, prompts the recommendation that introductions of improved

germplasm be made with a focus on cultivars with preferable sweet tasting roots. Two landraces with high market demand and reported tolerance to diseases were identified for further evaluation with a view to multiplication and distribution and incorporation into the breeding program. Cassava

**Table 5** Diversity indices of cassava landraces according to 11 Zones of Comoros and Tanzania

| Sub-population     | N   | Ho     | Hs     | Ht     | Dst    | Fst    | Fis      |
|--------------------|-----|--------|--------|--------|--------|--------|----------|
| Central Zone       | 61  | 0.2324 | 0.2120 | 0.2120 | 0.0000 | 0.0000 | - 0.0965 |
| Coastal Zone       | 53  | 0.2226 | 0.2061 | 0.2126 | 0.0065 | 0.0306 | - 0.0801 |
| Lake Zone          | 58  | 0.2277 | 0.2124 | 0.2124 | 0.0000 | 0.0000 | - 0.0721 |
| Mwali              | 16  | 0.2145 | 0.1883 | 0.1883 | 0.0000 | 0.0000 | - 0.1386 |
| Ndzouani           | 19  | 0.2144 | 0.1883 | 0.1883 | 0.0000 | 0.0000 | - 0.1386 |
| Ngazidja           | 41  | 0.2079 | 0.1958 | 0.1958 | 0.0000 | 0.0000 | - 0.0616 |
| Northern Zone      | 65  | 0.2317 | 0.2030 | 0.2130 | 0.0100 | 0.0470 | - 0.1413 |
| Pemba              | 20  | 0.2260 | 0.1727 | 0.1893 | 0.0165 | 0.0872 | - 0.3083 |
| Southern Highlands | 79  | 0.2214 | 0.2076 | 0.2076 | 0.0000 | 0.0000 | - 0.0664 |
| Western Zone       | 30  | 0.2342 | 0.2139 | 0.2139 | 0.0000 | 0.0000 | - 0.0945 |
| Zanzibar           | 33  | 0.2123 | 0.1894 | 0.2034 | 0.0141 | 0.0691 | - 0.1213 |
| Overall            | 475 | 0.2202 | 0.2031 | 0.2131 | 0.0100 | 0.0470 | - 0.0839 |

Ho = mean observed heterozygosities

Hs = The within population gene diversity (sometimes misleadingly called expected heterozygosity):

Ht = overall gene diversity

Dst = Ht - Hs The amount of gene diversity among samples

Fst = Dst/Ht (This is not the same as Nei's Gst, Nei's Gst is an estimator of Fst based on allele frequencies only)

Fis = 1 - Ho/Hs

landraces from Comoros warrant urgent conservation in IITA's international genebank due to their uniqueness which is likely a result of limited germplasm exchange due to their relative isolation on the islands.

**Funding** This work was funded by the International Fund for Agricultural Development (IFAD) under the Farm productivity Project; Family Farming and Resilience (PREFER).

#### Declarations

**Conflict of interest** The authors have no relevant financial or non-financial interests to disclose.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

#### References

- Asante IK, Offei SK (2003) RAPD-based genetic diversity study of fifty cassava (*Manihot esculenta* Crantz) genotypes. *Euphytica* 131:113–119. <https://doi.org/10.1023/A:1023056313776>
- Bougard S, Dray S (2018) Supervised multiblock analysis in R with the ade4 package. *J Stat Softw* 86:1–17. <https://doi.org/10.18637/jss.v086.i01>
- Carter S, Fresco L, Jones P, Fairbairn J (1992) An Atlas of Cassava in Africa: historical, agroecological and demographic aspects of crop distribution. CIAT, Cali
- Charrad M, Ghazzali N, Boiteau V, Niknafs A (2014) NbClust: an R package for determining the relevant number of clusters in a data set. *J Statistical Softw* 61:1–36
- Chavent M, Kuentz V, Liqueur B, Saracco J, Chavent MM (2017) PCAmixdata I. Package 'ClustOfVar'. Saracco J Clust Var
- Chen WC (2010) A quick guide for the phyclus package. Iowa State University, Ames
- Chen WC, Dorman K, Chen MWC (2013) Package 'phyclus'
- Cox TP, Nakabonge G, Ferguson M (2014) Manual for collection of cassava germplasm and associated farmer knowledge in eastern and southern Africa. IITAA, Ibadan. <https://doi.org/10.25502/ty62-gd79/p>
- Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. *JABES* 11:381–393. <https://doi.org/10.1198/108571106X154443>
- De Meeüs T, Goudet J (2007) A step-by-step tutorial to use HierFstat to analyse populations hierarchically structured at multiple levels. *Infect Genet Evol* 7:731–735

- De Vries J, Toenniessen G (2001) Securing the harvest: biotechnology, breeding and seed systems for African crops. CABI, Oxon
- Edwards JD, Saha S, Teclé IY, Strickler SR et al (2015) The Sol Genomics Network (SGN)—from genotype to phenotype to breeding. *Nucleic Acids Res* 43(Database issue):D1036–D1041
- FAOSTAT (2022) Online statistical database: Food balance. FAOSTAT. Accessed Jan 2022
- Ferguson ME, Shah T, Kulakow P, Ceballos H (2019) A global overview of cassava genetic diversity. *PLoS ONE* 14:e0224763. <https://doi.org/10.1371/journal.pone.0224763>
- Ferguson ME, Tumwegamire S, Chidzanga C, Shah T, Mtunda K, Kulembeka H et al (2021) Collection, genotyping and virus elimination of cassava landraces from Tanzania and documentation of farmer knowledge. *PLoS ONE* 16:e0255326. <https://doi.org/10.1371/journal.pone.0255326>
- Fernandez-Pozo N, Menda N, Edwards JD, Saha S, Teclé IY, Strickler SR, Bombarely A, Fisher-York T, Pujar A, Foerster H, Yan A, Mueller LA (2015) The Sol Genomics Network (SGN)—from genotype to phenotype to breeding. *Nucleic Acids Res* 43:D1036–D1041
- Fukuda W, Guevara C, Kawuki R, Ferguson M (2010) Selected morphological and agronomic descriptors for the characterization of cassava. IITA, Ibadan. <https://doi.org/10.25502/pfbm-9233/p>
- GADM (2023) <https://gadm.org/>. Accessed 20th Jan 2023
- Gilmour AR, Gogel BJ, Cullis BR, Welham SJ, Thompson R (2014) ASReml user guide release 4.1 functional specification. VSN International Ltd, Indore
- Goudet J, Jombart T (2021) hierfstat: estimation and tests of hierarchical F-statistics. R package version 0.5-10. <https://CRAN.R-project.org/package=hierfstat>
- Gruber B, Unmack PJ, Berry OF, Georges A (2018) Dartr: an R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Mol Ecol Resour* 18:691–699
- Hillocks RJ, Jennings DL (2003) Cassava brown streak disease: a review of present knowledge and research needs. *Int J Pest Manag* 49:225–234
- Hillocks RJ, Thresh JM, Tomas J, Botao M, Macia R, Zaviera R (2002) Cassava brown streak disease in northern Mozambique. *Int J Pest Manag* 48:178–181. <https://doi.org/10.1080/09670870110087376>
- Hurtado P, Olsen KM, Buitrago C, Ospina C, Marin J, Duque M et al (2008) Comparison of simple sequence repeat (SSR) and diversity array technology (DArT) markers for assessing genetic diversity in cassava (*Manihot esculenta* Crantz). *Plant Genetic Resour* 6:208–214
- IFAD (2016) Union des Comores, Note de stratégie de pays (NSP). Report No: 4149-KM
- Jennings DL (1957) Further studies in breeding cassava for virus resistance. *East Afr Agric J* 22(4):213–219
- Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btr521>
- Kassambara A, Mundt F (2020) Factoextra: extract and visualize the results of multivariate data analyses. R package version 1(7)
- Kawuki RS, Ferguson M, Labuschagne M, Kim D-J (2009) Identification, characterisation and application of single nucleotide polymorphisms for diversity assessment in cassava (*Manihot esculenta* Crantz). *Mol Breed* 23:669–684. <https://doi.org/10.1007/s11032-009-9264-0>
- Kiers HA (1991) Simple structure in component analysis techniques for mixtures of qualitative and quantitative variables. *Psychometrika* 56:197–212
- Legg J, Somado EA, Barker I, Beach L, Ceballos H, Cuellar W et al (2014) A global alliance declaring war on cassava viruses in Africa. *Food Secur* 6:231–248
- Li S, Cui Y, Zhou Y, Luo Z, Liu J, Zhao M (2017) The industrial applications of cassava: current status, opportunities and prospects. *J Sci Food Agric* 97:2282–2290
- Nei M (1972) Genetic distance between populations. *Am Nat* 106:283–292. <https://doi.org/10.1086/282771>
- Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York
- Nichols RF (1947) Breeding cassava for virus resistance. *East Afr Agric J* 12:184–194
- Nichols RF (1950) The brown streak disease of cassava: distribution, climatic effects and diagnostic symptoms. *East Afr Agric J* 15:154–160
- Pierre N, Wamalwa LN, Muiru WM, Simon B, Kanju E, Ferguson ME, Ndavi MM, Tumwegamire S (2022) Genetic diversity of local and introduced cassava germplasm in Burundi using DArTseq molecular analyses. *PLoS ONE* 17(1):e0256002. <https://doi.org/10.1371/journal.pone.0256002>
- R Core Team (2018) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org>
- Roa A, Maya M, Duque M, Tohme J, Allem AC, Bonierbale MW (1997) AFLP analysis of relationships among cassava and other *Manihot* species. *Theor Appl Genet* 95:741–750. <https://doi.org/10.1007/s001220050620>
- Sansaloni C, Petroli C, Jaccoud D, Carling J, Detering F, Grattapaglia D, Kilian A (2011) Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. *BMC Proc* 5:1–2
- Tiago AV, Rossi AAB, Tiago PV, Carpejani AA, Silva BM, Hoogerheide ESS et al (2016) Genetic diversity in cassava landraces grown on farms in Alta Floresta-MT, Brazil. *Genet Mol Res* 15:gmr.15038615. <https://doi.org/10.4238/gmr.15038615>
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358
- Wickham H, Averick M, Bryan J, Chang W, McGowan LDA, François R et al (2019) Welcome to the tidyverse. *J Open Source Softw* 4:1686
- Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* 19:395–420. <https://doi.org/10.2307/2406450>
- Yu G, Smith DK, Zhu H, Guan Y, Lam TTY (2017) Ggtree: an R package for visualization and annotation of

phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* 8:28–36

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.