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

Charles Orek (corek.publish@gmail.com)
South Eastern Kenya University  https://orcid.org/0000-0001-7310-4052

Martina Kyallo
ILRI: International Livestock Research Institute

Shorinola Oluwaseyi
ILRI: International Livestock Research Institute

Nasser Yao
Alliance of Bioversity International and CIAT Africa Hub

Research Article

Keywords: Genotyping-by-sequencing, landraces, improved genotypes, variety identification

Posted Date: February 18th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1295398/v1

License: ☇️ ⑥ This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

Background: Future demand for cassava is expected to increase to mitigate climatic changes, sustain food security and provide raw materials for industry. To meet these demands, adoption of modern omics methods ensures reliability, precision and timely delivery of more productive and resilient varieties. Therefore the purpose of this study was to contribute towards accurate identification of cassava accessions from a mix of duplicate clones, diverse local landraces (LARs) and improved genotypes (IMGs) in farmer fields. This is vital for cassava breeding.

Results: A total of 112 germplasms sampled through a field survey in major cassava growing regions of Kenya, were genotyped using single nucleotide polymorphisms (SNPs) markers generated through genotyping-by-sequencing (GBS) approach. Of the 33672 SNPs, 88% were anchored onto chromosomes, 3% in scaffolds and 9% could not be mapped. LD pruning and identity by state matrix estimation revealed 5808 SNPs that were used for hierarchical clustering and ADMIXTURE analysis for ancestries. Considering a sub-population of 2 - 20, a 5-fold cross-validation procedure identified 14 subpopulations present in the population from which the population structure was modeled. Approximately 48% of the germplasms were classified into 17 independent clusters as identical clones or duplicates. The remaining 52% formed admixtures and hence unique or non-duplicated clones; reducing the total number of samples surveyed from 112 to 73. Of the duplicates, 10 clusters were formed from LARs, four from IMGs, and three from a mix of both LARs and IMGs. The major and minor clusters contained 8 and 2 accessions, respectively. About 71% of clusters contained accessions from the same geographical region while 29% had accessions from different regions. The results revealed genetic relationships amongst LARs and IMGs. Duplication of LARs was attributed to historical sharing or exchange of planting materials by farmers while duplicates of IMGs could be attributed to convergent evolution, selection, or sharing of common parentage. The high number of admixtures or unique clones implied minimal loss of genetic diversity. Geographical restriction of clusters adduced to the minimal movement of planting materials across the country, perhaps linked to either inefficient seed distribution system or disease-driven quarantine measures.

Conclusions: GBS was successfully used to study the genetic relatedness of cassava genetic resources and variety identification in farmer fields. This omics approach and data herein generated could be adopted by breeders and other stakeholders in designing efficient and effective cassava improvement programs which might include the development of a core set of diagnostic markers for quality assurance, disease resistance, and targeted genomic profiling in cassava.

Introduction

Cassava (Manihot esculenta Crantz) which originated around the Amazon basin (Léotard et al. 2009; Hirst 2020; Olsen 2004; Olsen and Schaal 1999) was introduced in sub-Saharan Africa (SSA) by the Portuguese traders in the 16th century (Spencer and Ezedinma 2017) and in the East Africa coast in the 18th century (Were et al. 2004). The crop is a perennial woody shrub extensively grown in the tropical and subtropical regions of the world for its edible starchy tuberous roots, which are a major food source for developing countries (New World Encyclopedia 2020; Clifton and Keogh 2016; Liu et al. 2014; Kuete 2014). The continuous rise in cassava popularity in Africa is attributed to the crop’s low input requirement, tolerance to drought stress or low water requirement, survivability in marginal soils or soils with low nutrients, and flexible harvesting window that allows the crop to be left in the soil as a food reserve (Orek et al. 2020; Rabbi et al. 2020; Shigaki 2016). These make cassava a resilient crop important for food and nutritional security in Africa, where half a billion people eat the crop daily (Amelework et al. 2021; Tize et al. 2021; Agre et al. 2018; CIAT 2019).

Despite its significance, cassava production in SSA still lags behind other parts of the world. This has largely been attributed to low investments in breeding programs and inherent genetic challenges associated with the crop. Genetic barriers such as high heterozygosity, inbreeding depression, allopolyploid, poor seed set, irregular flowering, and the polygenic and recessive nature of many desirable traits, constrain development of new or improved varieties especially via conventional breeding (Elegba et al. 2021; Ceballos et al. 2020; Makwarela and Rey 2006). These are further compounded by a mixture of diverse local landraces and improved varieties that are often cultivated by most small-scale farmers on the same piece of land. Indeed, farmers often exchange stem cuttings or planting materials with their neighbors and neighboring communities, resulting in fields with a mixture of local cassava varieties (Andersson and de Vicente 2010; Nakabonge et al. 2018). Commonly, this results in the same ethnic or local name being assigned to different cassava germplasms or the same germplasms assigned different local names. Variety naming systems in the absence of formal seed systems can be quite temporally and spatially variable, leading to inconsistencies in the names of a particular variety (Rabbi et al. 2015). All these hamper the selection of breeding lines. To overcome these limitations, molecular approaches can assist in reliable identification, characterization, and verification of genotypes or varieties and hasten selection of appropriate parental plants (OECD 2016; Otti et al. 2011; Lebot 2009), thus improve the designing and delivery of tailored breeding objectives such as high yields (Lopez-Lavalle et al. 2021).

Accurate identification of crop cultivars is crucial in assessing the impact of crop improvement research outputs and the two commonly used identification approaches; elicitation of variety names from farmer interviews and morphological plant descriptors, have inherent uncertainty levels (Rabbi et al. 2015). The major aim of variety or cultivar identification is to catalog the crop’s genetic diversity (Lopez-Lavalle et al. 2021). There are many reports on many landraces of cassava in SSA but with limited studies on the genetic relatedness between these landraces and elite or improved accessions (Turyagyenda et al. 2012). Molecular marker technologies such as RFLPs, AFLPs, SSRs, DArTs, and SNPs among others have been used to detect polymorphisms and characterize genetic variation in cassava cultivars (Lopez-Lavalle et al. 2021). Rabbi et al (2015), successfully used SNPs derived from GBS to track and identify released cassava varieties and local landraces in Ghana, West Africa. The present study, therefore, applied the GBS approach to generate SNPs that revealed genetic relatedness amongst local landraces and improved cassava genotypes sampled from various cassava growing regions in Kenya. This is a preliminary step toward the acceleration of the cassava breeding process in the country.

Materials And Methods
Sample collection

A field survey was carried out in April 2018 in selected areas within major cassava growing regions of Nyanza, western, eastern, and coastal Kenya (Fig. 1). Systematic sampling was applied to identify cassava farmers or farms for cassava leaf collection (Koima et al. 2018). This involved stopping at regular predetermined intervals (~2-5 km) allowing wide coverage of the surveyed areas between farmer fields along the major motorable roads traversing each sampling location (Mware et al. 2009). The local name of the landraces and/or names of villages and GPS coordinates where the samples were collected were recorded (Table 1). Cassava leaves were harvested and pooled from five plants per landrace or genotype. The leaves were immediately transferred to falcon tubes half-filled with silica gels to preserve their integrity prior to DNA extraction.

Sequencing cassava using DArTSeq

Cassava leaf samples were sent to Integrated Genotyping Service and Support (IGSS) platform located at the Biosciences eastern and central Africa (BecA-ILRI) Hub in Nairobi, Kenya for genotyping. DNA extraction was done using TANBead Plant extraction kit. The quality and quantity of genomic DNA were determined using NanoDrop ND-1000 (Thermo Fisher Scientific) and agarose gel electrophoresis. Libraries were constructed according to Kilian et al. (2012). DArTSeq complexity reduction method through digestion of genomic DNA using a combination of _PstI_ and _MseI_ restriction enzymes and ligation of barcoded adapters followed by PCR amplification of adapter-ligated fragments. Libraries were sequenced using single read sequencing runs for 77 bases. Next generation sequencing was carried out using the Illumina Hiseq2500. DArTseq markers scoring was achieved using DArtsoft14 which is an in-house marker scoring pipeline based on algorithms. Two types of DArTseq markers were scored, SilicoDAT markers and SNP markers which were both scored as binary for presence/absence (1 and 0, respectively) of the restriction fragment with the marker sequence in genomic representation of the sample. Both SilicoDAT markers and SNP markers were aligned to the reference genomes of Cassava_v61 to identify chromosome positions.

Data analysis

The quality of the SNP data was filtered using TASSEL and SNPs anchored on scaffold or missing chromosome information were discarded. TASSEL was also used to select SNPs with >0.05 minor allele frequencies (MAF) and SNPs with no more than 20% missing genotype data. For LD pruning and IBS matrix estimating, Plink 1.9 was used to select for SNP with less than 0.5 R² LD value within each 50-SNP window size i.e. considering 50 SNPs at a time, the LD between them should be less than 0.5 LD R². Two methods were used for grouping the genotypes including hierarchical clustering using identity by state (IBS) matrix and a model-based maximum likelihood estimation of individual ancestries from multi-locus SNP genotype datasets using ADMIXTURE (Rabbi et al. 2015). IBS examines if two lines are identical based on the nucleotide (SNP alleles) that they share. Using the pruned SNPs from Plink, IBS matrix was calculated with the distance function of Plink (Purcell et al. 2007). The matrix was used for hierarchical clustering using the Ward2 method for distance estimation. The critical distance threshold used to declare two genotypes are identical was 0.05 based on the empirically determined evidence suggested by Rabbi et al (2015) from the distribution of distances between duplicated DNA of 64 cassava samples. A ward’s minimum variance hierarchical cluster dendrogram (Fig. 3) was then generated from IBS matrix using Analyses of Phylogenetics and Evolution (APE) package (Paradis et al. 2004) implemented within R software (R Core Team, 2020).

After filtering, LD pruning and IBS matrix were used to determine the LD threshold and select SNPs accordingly. The same set of LD-pruned SNPs used for the hierarchical clustering was also used for ADMIXTURE to identify ancestries of the collected cassava germplasms (Rabbi et al. 2015). The model-based clustering approach implemented in ADMIXTURE assumes linkage equilibrium among loci and Hardy-Weinberg equilibrium within ancestral populations (Frichot et al. 2014; Rabbi et al. 2015). Considering a sub-population of 2 - 20, a 5-fold cross-validation procedure was used to select the optimum number of sub-populations present in the population as 14. The population structure was then modeled with the optimum number of underlying sub-population groups (Fig. 5).

Results

Cassava germplasms

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

Filtering and selection of SNPs and optimum population identification

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

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

Geographically, majority of the clusters (12 of the 17 or ~71%) contained accessions sampled from the same region (Table 3). These included clusters II, III, IV, V, VI, VII, IX, XI, XIII, XIV, XV, and XVI. The remaining five of the 17 (~29%) clusters (I, VIII, X, XII & XVII) had accessions sampled from different regions (Table 3). For instance cluster I, VIII and XII were from regions in closer proximity (Siaya = 0°26'N; 33°58'E, and Kakamega = 0°16'N; 34°45'E) while cluster X (Kitui = 0°10'S; 37°50'E, and Kakamega = 0°16'N; 34°45'E) and XVII (Makueni = 1°48'; 37°37'E, and Kakamega = 0°16'N; 34°45'E) represented clustering of accessions from far regions (Table 3). Landraces from Kilifi (3°40'S; 39°45'E) located in coastal Kenya did not cluster with landraces or improved genotypes from other regions (see cluster XIII and XIV) (Table 3). Compared to other regions, Kitui (0°10'S; 37°50'E) had a majority (5) of different clusters (II, III, IV, V & VI).

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

Discussion

Most of the sampled materials (approximately 63%) were local landraces compared to improved cassava genotypes that constituted 37%. This implied cultivation of more local cassava varieties or landraces which have been attributed to farmer preferred characteristics such as culinary attributes and cooking quality, sweet or bitter tastes, early maturity, pests and disease resistance, high yield, root storability in the ground, drought tolerance among other traits (Nakabonge et al. 2018; Bentley et al. 2017; Woyengo 2011). Farmers often hold several generations of knowledge concerning the attributes of landraces and sometimes have specific reasons why they retain particular cultivars (Ferguson et al. 2021). On the reverse, the results implied minimal adoption and cultivation of the improved varieties in Kenya, a potential drawback for the management of cassava diseases as most of the improved genotypes had been bred and introduced for resistance or tolerance to CMD and CBSD. This was corroborated by earlier studies on low dissemination, adoption, and production of improved cassava varieties in Africa, a situation that was linked to lack of involvement of farmers and end-users in designing, planning, and execution of breeding strategies and objectives (Nakabonge et al. 2018; Bechoff et al. 2018; Woyengo et al. 2014; Kamau et al. 2011). Farmer preferences and varietal attributes influence the adoption of new cassava varieties (Okuku et al. 2018; Ndumumuremyi et al. 2016; Kamau et al. 2016; Khonje et al. 2015). It is however noted that farmer preferences or attributes of the genotyped landraces and improved varieties were not assessed in the present study.

The SNPs marker data generated using GBS was successfully used to determine genetic relatedness among sampled cassava germplasms. From a total of 33672 SNPs identified, 5808 SNPs (~17%) obtained after LD pruning and IBS matrix estimation were used for hierarchical clustering and ADMIXTURE analysis to identify ancestries. This enabled the identification of germplasms that clustered together as well as unique or non-duplicated germplasms. Thus, a large number of SNPs may not be needed to achieve accurate identification of cassava varieties, whether in farmers’ fields or formal germplasms collections (Lopez-Lavalle et al. 2021; Rabbi et al. 2015; Ferguson et al. 2012). A further study could be initiated to identify these SNPs and design KASP markers for varietal identification. Knowledge of the existence of duplicates in the field is important during the collection of variability and evaluation and selection of parents for cassava improvement or breeding purposes. Similarly, genomic or SNPs markers have been used to confirm that particular cassava accessions are not identical, and others are possible duplicates (Mbanjo et al. 2021; Ferguson et al. 2012). They have also been used to track local landraces and assess the adoption of improved varieties (Rabbi et al. 2015; Assfaw et al. 2017; Turyagyenda et al. 2012). Accurate identification of crop cultivars is crucial in assessing the impact of crop improvement research outputs (Rabbi et al. 2015). Generally, the genomic approach contributes to further characterization of cassava genetic resources, an important step in improving cassava production in Kenya.

Further results from the present study showed that the majority of the duplicated clones were local landraces while geographically, most of the duplicated accessions were sampled either from the same region or from different regions of closer proximity. These redundancies were previously attributed to the historical sharing of cassava accessions or the same germplasms exchange between farmers with different genotype names (Albuquerque et al. 2019). Farmers often exchange planting materials with their neighbors or different neighboring communities, resulting in fields with a mixture of local cassava
varieties (Andersson and de Vicente 2010; Nakabonge et al. 2018). Thus the same ethnic or local name could be assigned to different cassava germplasms or the same germplasms assigned different local names. Variety naming systems in the absence of formal seed systems can be quite temporally and spatially variable leading to inconsistencies in the names of a particular variety (Rabbi et al. 2015). The informal farmer-farmer seed distribution system is often inefficient, denying farmers in far flung areas access or a share of alternative planting materials.

Ferguson et al (2021) reported that individual cassava landraces were not widely distributed across Tanzania with limited farmer-to-farmer diffusion with implications for seed systems. Indeed, smallholder farmers recycle stem cuttings of traditional landrace cultivars (Nweke et al. 2002) and there is a flow of seed within and outside the villages, with little introduction of new cultivars (Mtunguja et al. 2014). The absence of an effective seed distribution system (Kyamanywa et al. 2011) has limited farmers' access to planting materials from improved genotypes. Additionally, elicitation of cassava variety names from farmer interviews during surveys and/or use of morphological plant descriptors have had inherent uncertainty levels (Rabbi et al. 2015). Morphological descriptors are also greatly influenced by the environment and show continuous variation and high plasticity, with most of them only scorable at maturity (Ndung'u et al. 2014). Restrictions of clusters to the same geographical areas where accessions were sampled could also be attributed to quarantine measures that restricted the movement of planting materials in order to stop the spread of virus diseases such as CMD and CBSD.

Similarities in cassava accessions can also arise due to convergent evolution, selection, or sharing of common parentage (Ndung'u et al. 2014). This was probably the case in Kitui region where the majority of duplicates were improved genotypes that had shared the same parents during their breeding for resistance to cassava brown streak disease (Koima and Orek 2018). Crops gradually lose their genetic variability through domestication and breeding, resulting in more uniform cultivars and reducing their recombination rates (Rufo et al. 2019). This could perhaps be used to explain clusters that included both improved genotypes and local landraces. It is however noted that no recent evidence has shown loss of genetic variation from genetic drift during the introduction of cassava to Africa (Ferguson et al. 2019). The relatively low levels of diversity reported in the previous study were only observed in IITA breeders’ germplasms and may represent rather a genetic bottleneck (Ferguson et al. 2019). For future breeding programs involving hybridization or selection, de Oliveira et al (2015) recommended the introduction of new genetic variability into commercial cultivars to avoid low genetic variation and to improve the quality of cassava roots. The unique or non-duplicated landraces and improved genotypes in the present study represented a more expanded cassava genetic pool from which variability can be derived for future breeding purposes. It might also be important to build the core collection of the 73 unique genotypes studied in this study for further efficient conservation and cassava breeding. High genetic diversity drives better crop adaptation to emerging environmental cues.

**Conclusion**

Molecular markers have an important role to play as farmers frequently give different names to the same cultivar or landraces, making identification difficult, particularly as cassava varieties are not easy to distinguish morphologically (Mbanjo et al. 2021). This enables the correct assessment of adoption rates, which in turn, influences breeding priorities and agricultural policies (Kretzschmar et al. 2018). Knowledge on the extent of genetic diversity among cassava landraces and improved genotypes in Kenya using GBS-derived SNP markers may promote their conservation and/or efficient selection and utilization as parental lines for breeding for biotic and abiotic stress tolerance. Although local landraces may be low-yielding, they may have high genetic diversity that could promote gene flow through hybridization (Turyagyenda et al. 2012), enabling crop improvement and adaptability of species to changing climatic conditions, new pests, and diseases (Prempeh et al. 2020).

**Declarations**

**Acknowledgements:**

The authors thank Dr. Nasser Yao and Dr. Oluwaseyi Shorinola and Martina Kyallo for their valuable comments, inputs, corrections and technical assistance during development of the manuscript

**Author contribution:**

OC conceptualized the research; KM and OC carried out data collection and methodology; OS and OC carried out data curation and analysis: NY and KM helped with funding acquisition and resources; OC prepared the original draft and editing; KM, OS and NY provided supervision and project administration; all authors read and approved the final manuscript. Where OC = Orek Charles; KM = Kyallo Martina; OS = Oluwaseyi Shorinola; NY = Nasser Yao.

**Funding:**

This research was funded by the BecA-ILRI Hub through the Africa Biosciences Challenge Fund (ABCF) program. The ABCF Program is funded by the Australian Department for Foreign Affairs and Trade (DFAT) through the Biosciences eastern and central Africa - The Commonwealth Scientific and Industrial Research Organization (BecA-CSIRO) partnership; the Syngenta Foundation for Sustainable Agriculture (SFSA); the Bill and Melinda Gates Foundation (BMGF); the UK Department for International Development (DFID) and the Swedish International Development Cooperation Agency (Sida).

**Availability of data and materials:**

All data are provided in the manuscript. To access materials researched on, please contact the corresponding author.
Code availability:

Not applicable.

Author Declarations:

Ethics approval and consent to participate:

Not applicable.

Consent for publication:

Not applicable.

Conflict of interest:

The authors declare no competing interests

References

1. Agre AP, Bhattacharjee R, Rabbi IY, et al (2018) Classification of elite cassava varieties (Manihot esculenta Crantz) cultivated in Benin Republic using farmers’ knowledge, morphological traits and simple sequence repeat markers. Genet Res. Crop Evol 65:513–525.

2. Albuquerque HYG, de Oliveira EJ, Brito AC, et al (2019) Identification of duplicates in cassava germplasm banks based on single nucleotide polymorphisms (SNPs). Sci. agric. (Piracicaba, Braz.) 76 (4) https://doi.org/10.1590/1678-992X-2017-0389

3. Amelework AB, Bairu MW, Obakeng M, Venter SL, Laing M (2021) Adoption and Promotion of Resilient Crops for Climate Risk Mitigation and Import Substitution: A Case Analysis of Cassava for South African Agriculture. Front. in Sust. Food Syst 5:105

4. Andersson MS, de Vicente MC (2010) Cassava, manioc, yuca”, Chapter 6, in: Andersson, M.S. and M.C. de Vicente (eds.), Gene Flow Between Crops and Their Wild Relatives, Johns Hopkins University Press, Baltimore, Maryland, 125-146.

5. Assfaw WT, Girma TG, Abdoulaye T, Rabbi IY, Olanrewaju A, Bentley J, Manyong V  (2017)  The cassava monitoring survey in Nigeria nal report. IITA, Ibadan, Nigeria. ISBN 978-978-8444-81-7.66

6. Bechoff A, Tomlins K, Fliedel G, Becerra Lopez-lavalle LA, Westby A, Hershey C, et al (2018) Cassava traits and end-user preference: Relating traits to consumer liking, sensory perception, and genetics. Crit. Rev. Food Sci. Nutr. 58:547–567.

7. Bentley J, Olanrewaju A, Madu T, Olaosebikan O, Abdoulaye T, Wossen V, et al (2017) Cassava farmers’ preferences for varieties and seed dissemination system in Nigeria: Gender and regional perspectives. IITA Monograph, IITA, Ibadan, Nigeria.

8. CIAT (2019) Cassava: https://ciat.cgiar.org/what-we-do/breeding-better-crops/rooting-for-cassava/

9. Ceballos H, Rojanaripiched C, Phumichai C, Becerra LA, et al (2020) Excellence in Cassava Breeding: Perspectives for the Future. Crop Breed Genet Genom.2 (2):e200008. https://doi.org/10.20900/cbgg20200008

10. Clifton P, Keogh J (2016) Starch: Encyclopedia of Food and Health, pp. 146-151

11. de Oliveira EJ, Alves F, Alves SL, de Oliveira V, da Silva S (2015) Genetic variation of traits related to quality of cassava roots using affinity propagation algorithm. Genetics & Plant Breeding, Sci. Agric.72 (1) https://doi.org/10.1590/0103-9016-2014-0043

12. Elegba W, McCallum EJ, Wilhelm G, Vanderschuren H (2021) Genetic transformation and regeneration of a farmer-preferred cassava cultivar from Ghana. Fron. in Plant Sci. 12:909

13. 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(8): e0255326. https://doi.org/10.1371/journal.pone.0255326

14. Ferguson ME, Shah T, Kulakow P, Ceballos H (2019) A global overview of cassava genetic diversity. PloS one, 14(11), e0224763. https://doi.org/10.1371/journal.pone.0224763

15. Ferguson ME, Hearne SJ, Close TJ, et al (2012) Identification, validation and high-throughput genotyping of transcribed gene SNPs in cassava. Theor. Appl. Genet. 124:685–695 https://doi.org/10.1007/s00122-011-1739-9

16. Frichot E, Mathieu F, Trouillon T, Bouchard G, Francois O (2014) Fast and efficient estimation of individual ancestry coefficients. Genetics, 196(4):973–83

17. Hirst KK (2020) The History and Domestication of Cassava.” ThoughtCo, Aug. 28, 2020,thoughtco.com/cassava-manioc-domestication-170321.

18. Kamau J, Melis R, Laing M, Derera J, Shanahan P, Ngugi K, Mgwiga Y (2016) Farmers’ Perceptions of Production Constraints and Preferences in Cassava Grown in Semi-Arid Areas of Kenya. Int.J.Curr.Microbiol.App.Sci, Vol. 5(3): 844-859.

19. Kamau J, Melis R, Laing M, Derera J, Shanahan P, Ngugi ECK (2011) Farmers’ participatory selection for early bulking cassava genotypes in semi-arid Eastern Kenya. Journal of Plant Breeding and Crop Science 3(3):44-52

20. Khonje M, Mkandawire P, Manda J, Alene DA (2015) Analysis of adoption and impacts of improved cassava varieties in Zambia. International Conference of Agricultural Economics, 1 – 28
21. Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, et al. (2012) Diversity arrays technology: a generic genome profiling technology on open platforms. Methods Mol. Biol. 88: 67–89. https://doi.org/10.1007/978-1-61779-870-2_5

22. Koima IN, Orek CO (2018) Response to Cassava Brown Streak Disease infections in local and improved cassava genotypes under field and greenhouse assays in lower eastern Kenya. Int. J. of Path. Res. 1(3):1-14 https://doi.org/10.9734/ijip/2018/v1i329616

23. Koima IN, Orek CO, Nguluu SN (2018) Distribution of Cassava Mosaic and Cassava Brown Streak Diseases in agro-ecological zones of lower Eastern Kenya. UJSRT, 3(1):391 – 399.

24. Kretzschmar T, Mbanjo EG N, Magalit GA, Dwiyanti MS, Habib MA, Diaz MG, et al (2018) DNA fingerprinting at farm level maps rice biodiversity across Bangladesh and reveals regional varietal preferences. Sci. Rep. 8:14920.

25. Kuete V (2014) Physical, Hematological and Histopathological Signs of Toxicity Induced by African Medicinal Plants. Toxicological Survey of African Medicinal Plants, 635 – 657

26. Kyamanywa S, Kashaija IN, Getu E, Amata R, Senkeshia N, Kullaya A (2011) Enhancing Food Security through Improved Seed Systems of Appropriate Varieties of Cassava, Potato and Sweetpotato Resilient to Climate Change in Eastern Africa. Nairobi, Kenya, ILRI.

27. Lebot V (2009) Tropical Root and Tuber Crops: Cassava, Sweet Potato, Yams, and Aroids, Crop Production Science in Horticulture Series, Volume 17, CAB, Wallingford, United Kingdom.

28. Léotard G, Duputié A, Kjellberg F, Douzery E, Debain C, Granville JJ, McKey D (2009) Phylogeography and the origin of cassava: New insights from the northern rim of the Amazonian basin. Molecular phylogenetics and evolution, 53

29. Liu Q, Liu J, Zhang P, He S (2014) Root and tuber crops, in: Encyclopedia of Agriculture and Food Systems, https://doi.org/10.1016/B978-0-444-52512-3.00151-0

30. Lopez-Lavalle LAB, Bohorquez-Chaux A, Zhang X (2021) Identification of Cassava Varieties in Ex-Situ Collections and Global Farmer's Fields: An Update from 1990 to 2020. IntechOpen, 1-30

31. Makwarela M, Rey C (2006) Cassava Biotechnology, a southern African Perspective. Biotechnology and Molecular Biology Review, 1(1):2-11

32. Mbanjo EGN, Rabbi IY, Ferguson ME, Kayondo SI, Hwa EN, Tripathi L, Kulakow P, Egesi C (2021) Technological Innovations for Improving Cassava Production in Sub-Saharan Africa. Frontiers in Genetics, 11:1829 https://doi.org/10.3389/fgene.2020.623736

33. Mtunguqa MK, Laswai HS, Muzanila YC, Ndunguru J (2014) Farmers knowledge on selection and conservation of cassava (Manihot esculenta) genetic resources in Tanzania. J. Biol. Agric. Healthcare 4:120–129.

34. Mware B, Narla R, Amata R, Olubayo F, Songa J, Kyamanywa S, Ateka EM (2009) Efficiency of cassava brown streak virus transmission by two whitefly species in coastal Kenya. J. Gen. Mol. Virol. 1: 40-45.

35. Nakabonge G, Samukoya C, Baguma Y (2018) Local varieties of cassava: Conservation, Cultivation and use in Uganda. Environ. Dev. Sust. 20: 2427–2445.

36. New World Encyclopedia (2020) Cassava: Retrieved on Sept. 6, 2021: https://www.newworldencyclopedia.org/p/index.php?title=Cassava&oldid=1030031

37. Ndung’u JN, Wachira FN, Kinyua MG, Lelgut DK, Njau P, Okwaro H, Obiero H (2014) Genetic diversity study of Kenyan cassava germplasm using simple sequence repeats. African J. of Biot.13 (8):926-935

38. Nduwumuremyi A, Melis R, Shanahan P, Asiimwe T (2016) Participatory appraisal of preferred traits, production constraints and postharvest challenges for cassava farmers in Rwanda. Food Sec. 8:375–388 https://doi.org/10.1007/s12571-016-0556-z

39. Nweke FI, Spencer DSC, Lynam JK (2002) The Cassava Transformation: Africa’s Best-Kept Secret, Michigan State University Press; 2002.

40. OECD (2016) Cassava (Manihot esculenta Crantz), in Safety Assessment of Transgenic Organisms in the Environment, 6: OECD Consensus Documents, OECD Publishing, and Paris. https://doi.org/10.1787/9789264253421-6-en

41. Olsen KM (2004) SNPs, SSRs and inferences on cassava’s origin. Plant Mol. Biol.56:517–526

42. Olsen KM, Schaal B (1999) Evidence on the origin of cassava: Phylogeography of cassava Manihot esculenta Crantz. Proc. Natl. Acad. Sci. USA, 96:5586–5591.

43. Okuku IO, Nyikal RA, Otieno DJ (2018) An assessment of the effect of varietal attributes on the adoption of improved cassava in HomaBay County, Kenya. MSc. Thesis, University of Nairobi

44. Orek C, Gruijsem W, Ferguson M, Vanderschuren H (2020) Morpho-physiological and molecular evaluation of drought tolerance in cassava (Manihot esculenta Crantz). Field Crops Research, 255, http://doi.org/10.1016/j.fcr.2020.107861

45. Otte I, Fakoya A, Andrew I, Gedil M (2011) Development of genomic tools for verification of hybrids and selfed progenies in cassava (Manihot esculenta Crantz). African J. of Bio. 10(76): 17400-17408

46. Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. Bioinformatics, 20(2):289–90.

47. Prempah WNA, Manu-Aduening JA, Quain MD, Asante IK, Ofsei SK, Danquah EY (2020) Assessment of genetic diversity among cassava landraces using single nucleotide polymorphic markers. Afr. J. Biot, Vol. 19(6):383-391

48. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575. https://doi.org/10.1016/S0002-9297(07)01108-4

49. Rabbi IY, Kayondo SI, Bauchet G, Yusuf M, Aghogho CI, Ogunpaimo K, et al (2020) Genome-wide association analysis reveals new insights into the genetic architecture of defensive, agro-morphological and quality-related traits in cassava. Plant Mol Biol. https://doi.org/10.1007/s11103-020-01038-3
51. Rabbi IV, Kulakow PA, Manu-Aduening JA, Dankyi AA, Asibuo JY, Parkes EY, et al (2015) Tracking crop varieties using genotyping-by-sequencing markers: a case study using cassava (Manihot esculenta Crantz). BMC Genetics 6:115 https://doi.org/10.1186/s12863-015-0273-1
52. Rufo R, Alvaro F, Royo C, Soriano JM (2019) From landraces to improved cultivars: Assessment of genetic diversity and population structure of Mediterranean wheat using SNP markers. PLoS ONE 14(7): e0219867 https://doi.org/10.1371/journal.pone.0219867
53. Shigaki T (2016) Cassava: The Nature and Uses. Encyclopedia of Food and Health, 687-693
54. Spencer D, Ezedinma C (2017) Cassava cultivation in sub-Saharan Africa. https://doi.org/10.19103/AS.2016.0014.06
55. Tize I, Fotso AK, Nukene EN, Masso C, Ngome FA, Suh C, Lendzemo VW, Nchoutnji I, et al (2021) New cassava germplasm for food and nutritional security in Central Africa. Sci Rep 11:7394
56. Turyagyenda LF, Kizito EB, Ferguson ME, Baguma Y, Harvey JW, Gibson P, et al (2012) Genetic diversity among farmer-preferred cassava landraces in Uganda. Afr. Crop Sci. J. 20: 15–30.
57. Were HK, Winter S, Maiss E (2004) Viruses infecting cassava in Kenya. Plant Disease, 88:17-22
58. Woyengo WV, Melis R, Shanahan P, Odongo OM (2014) Participatory evaluation methods of cassava varieties preferred in mild-altitude tropical climate conditions of western Kenya. African J. of Agric. Res. 9(7):1326-1333.
59. Woyengo WV (2011) Cassava breeding through complementary conventional and participatory approaches in western Kenya. Doctoral thesis, University of KwaZulu-Natal, South Africa

Tables

Table 1

Cassava landraces and genotypes sampled during field surveys from different cassava growing regions of Kenya.
| Local ID | Code | Region / GPS | Origin / Attributes | Local ID | Code | Region / GPS | Origin / Attributes | Local ID | Code | Region |
|----------|------|--------------|---------------------|----------|------|--------------|---------------------|----------|------|--------|
| Shavirotsi | KK1  | 0°16'N;34°45'E | Landrace / no information | Nya-Yenga | SYA8 | 0°26'N;33°58'E | Landrace / no information | Kitwa_Il | MK4  | 1°48'S; |
| Bwichina | KK2  | " | Landrace / no information | Nya-Gang | SYA9 | " | Landrace / no information | Kitwa_Il | MK5  | " |
| Lunyalala | KK3  | " | Landrace / no information | Nyal-Kada | SYA10 | " | Landrace / no information | Masokani_I | MK6  | " |
| Shanina | KK4  | " | Landrace / no information | Nya-Udai | SYA11 | " | Landrace / CMD susceptible | Masokani_Il | MK7  | " |
| Mukulusu | KK5  | " | Landrace / no information | AdhiamboLera | SYA12 | " | Landrace / CMD susceptible | Kaliluni | MK8  | " |
| Itenyi | KK6  | " | Landrace / no information | Nya-Bungoma | SYA13 | " | Landrace / no information | Muvila | MK9  | " |
| Shisembe | KK7  | " | Landrace / no information | Lady Gay | SYA14 | " | Improved genotype | TC14 | MK10 | " |
| Inzakula | KK8  | " | Landrace / no information | Kiboko297 | SEK1 | 0°10'S;37°50'E | KALRO / CBSD resistant | TC4-Katune | MK11 | " |
| Shitaho | KK9  | " | Landrace / no information | Thika272 | SEK2 | " | KALRO / CBSD resistant | 99/0056 | MK12 | " |
| Lugala | KK10 | " | Landrace / no information | Thika273 | SEK3 | " | KALRO / CBSD resistant | Kalimbini_I | MK13 | " |
| Lugusisti | KK11 | " | Landrace / no information | Kiboko275 | SEK4 | " | KALRO / CBSD resistant | Kalimbini_Il | MK14 | " |
| Banasa | KK12 | " | Landrace / no information | Kiboko274 | SEK5 | " | KALRO / CBSD resistant | Kalimbini_III | MK15 | " |
| Isambe | KK13 | " | Landrace / no information | Thika280 | SEK6 | " | KALRO / CBSD resistant | Kalimbini_IV | MK16 | " |
| Isulu | KK14 | " | Landrace / no information | Kiboko300 | SEK7 | " | KALRO / CBSD resistant | Katsuhanza | MK17 | " |
| Ikholi | KK15 | " | Landrace / no information | Kiboko271 | SEK8 | " | KALRO / CBSD resistant | Kasukari (990127) | MK18 | " |
| Ingotse | KK16 | " | Landrace / no information | Thika279 | SEK9 | " | KALRO / CBSD resistant | Kitivo | MK19 | " |
| Shikoti | KK17 | " | Landrace / no information | Thika289 | SEK10 | " | KALRO / CBSD resistant | Kimutwa | MK20 | " |
| Shipalo | KK18 | " | Landrace / no information | Kiboko295 | SEK11 | " | KALRO / CBSD resistant | Mumbuni | MK21 | " |
| Shamiloli | KK19 | " | Landrace / no information | Kiboko277 | SEK12 | " | KALRO / CBSD resistant | Halu | KF1  | 3°40'S; |
| Madioli | KK20 | " | Landrace / no information | Kiboko276 | SEK13 | " | KALRO / CBSD resistant | Kibandameno | KF2 | " |
| Shiswa | KK21 | " | Landrace / no information | Thika278 | SEK14 | " | KALRO / CBSD resistant | Agriculture | KF3 | " |
| MM96/1871 | KK22 | IITA / CMD resistant | Kiboko281 | SEK15 | KALRO / CBSD resistant | Tajarika/KME-0802 | KF4 |
| MM97/0293 | KK23 | KALRO / CMD resistant | Thika5 | SEK16 | Landrace / CMD resistant | Kaleso | KF5 |
| Magana | KK24 | Landrace / CBSD resistant | Serere | SEK17 | CIAT / CBSD susceptible | Soyosoyo | KF6 |
| CK9 | KK25 | Landrace / no information | Kiboko9 | SEK18 | KALRO / CBSD resistant | Sokoke_I | KF7 |
| Matuja | KK26 | Landrace / CMD susceptible | Kiboko10 | SEK19 | KALRO / CBSD resistant | Sokoke_II | KF8 |
| Fumbachai | KK27 | Landrace / no information | Kiboko11 | SEK20 | KALRO / CBSD resistant | Kakanjuni_I | KF9 |
| MM98/1313-HS | KK28 | KALRO / Improved | Kiboko159 | SEK21 | KALRO / CBSD resistant | Kakanjuni_II | KF10 |
| MH95/0183 | KK29 | IITA / CMD resistant | Kiboko257 | SEK22 | KALRO / CBSD resistant | Kakanjuni_III | KF11 |
| MM98/0686 | KK30 | IITA / Improved genotype | Kiboko258 | SEK23 | KALRO / CBSD resistant | Mkongo_I | KF12 |
| MM96/0686 | KK31 | KALRO / CMD resistant | Kiboko259 | SEK24 | KALRO / CBSD resistant | Mkongo_II | KF13 |
| Aruaro | SYA1 | 0°26’N;33°58’E | Landrace / no information | Kiboko267 | SEK25 | KALRO / CBSD resistant | Cha-Vyango_I | KF14 |
| Othigo-Diep | SYA2 | Landrace / no information | Kiboko268 | SEK26 | KALRO / CBSD resistant | Cha-Vyango_II | KF15 |
| Nyakatanegi_I | SYA3 | Landrace / no information | Kiboko269 | SEK27 | KALRO / CBSD resistant | Chumani | KF16 |
| Nyakatanegi_II | SYA4 | Landrace / no information | Kiboko270 | SEK28 | KALRO / CBSD resistant | Matano-Manne | KF17 |
| Nya-Uyoma | SYA5 | Landrace / no information | Kasioni | MK1 | 1°48’S;37°37’E | Landrace / no information | KALRO | KF18 |
| Kamis | SYA6 | Landrace / CMD susceptible | Kisimba | MK2 | Landrace / no information |
| Nya-Uganda | SYA7 | Landrace / CMD susceptible | Kitwa_I | MK3 | Landrace / no information |

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

### Table 2

The distribution of the SNPs across the cassava genome
| #  | Chromosome | No. of SNPs |
|----|------------|------------|
| #1 | 01         | 495        |
| #2 | 02         | 431        |
| #3 | 03         | 396        |
| #4 | 04         | 367        |
| #5 | 05         | 335        |
| #6 | 06         | 416        |
| #7 | 07         | 254        |
| #8 | 08         | 307        |
| #9 | 09         | 258        |
| #10| 10         | 363        |
| #11| 11         | 392        |
| #12| 12         | 262        |
| #13| 13         | 215        |
| #14| 14         | 315        |
| #15| 15         | 350        |
| #16| 16         | 249        |
| #17| 17         | 199        |
| #18| 18         | 204        |
| **Total** |       | **5,808** |

Table 3

Classification of cassava accessions into clusters based on underlying sub-population groups derived from Figure 5
| Local ID   | Type | Class      | Cluster # | Region / GPS       | Local ID   | Type | Class   | Cluster # | Region / GPS       |
|-----------|------|------------|-----------|---------------------|-----------|------|---------|-----------|---------------------|
| Matuja    | LAR  | All        | I         | 0°16'N;34°45'E     | Shavirotsi| LAR  | Unique  | NCL       | 0°16'N;34°45'E     |
| Othigo-Diep| "    | All        | I         | 0°26'N;33°58'E     | Bwichina  | LAR  | "       | "         | "                   |
| Aruaro    | "    | Identical  |           | "                   | Lunyalala | LAR  | "       | "         | "                   |
| Nya-Udai  | "    | "          |           | "                   | Mukulusu  | LAR  | "       | "         | "                   |
| Kiboko276 | IMG  | All        | II        | 0°10'S;37°50'E     | Shisembe  | LAR  | "       | "         | "                   |
| Kiboko297 | "    | "          |           | "                   | Shitaho   | LAR  | "       | "         | "                   |
| Kiboko274 | IMG  | All        | III       | 0°10'S;37°50'E     | Lugusisti | LAR  | "       | "         | "                   |
| Thika278  | "    | Identical  |           | "                   | Banasa    | LAR  | "       | "         | "                   |
| Kiboko271 | IMG  | All        | IV        | 0°10'S;37°50'E     | Ingotse   | LAR  | "       | "         | "                   |
| Thika289  | "    | Identical  |           | "                   | Shiswa    | LAR  | "       | "         | "                   |
| Kiboko300 | IMG  | All        | V         | 0°10'S;37°50'E     | MM96/1871 | IMG  | "       | "         | "                   |
| Thika273  | "    | "          |           | "                   | MM97/0293 | IMG  | "       | "         | "                   |
| Thika5    | LAR  | All        |           | "                   | Magana    | LAR  | "       | "         | "                   |
| Kiboko281 | IMG  | All        | VI        | 0°10'S;37°50'E     | CK9       | LAR  | "       | "         | "                   |
| Thika280  | "    | "          |           | "                   | MM98/1313-HS | IMG  | "       | "         | "                   |
| Itenyi    | LAR  | All        | VII       | 0°16'N;34°45'E     | MM08/2206 | IMG  | "       | "         | "                   |
| Inzakula  | "    | "          |           | "                   | MM96/0686 | IMG  | "       | "         | "                   |
| Lady Gay  | LAR  | All        | VIII      | 0°26'N;33°58'E     | Nyakatanegi-II | LAR  | "       | "         | 0°26'N;33°58'E     |
| Shanina   | "    | "          |           | "                   | Nya-Uyoma | LAR  | "       | "         | "                   |
| MH95/0183 | IMG  | All        |           | "                   | Kamis     | LAR  | "       | "         | "                   |
| Kalimbini-I| LAR  | All        | I         | 1°48'S;37°37'E     | Nya-Uganda| LAR  | "       | "         | "                   |
| Kasioni   | "    | "          |           | "                   | AdhiamboLera | LAR  | "       | "         | "                   |
| Kitwa-II  | "    | "          |           | "                   | Nya-Bungoma| LAR  | "       | "         | "                   |
| Kitwa-III | "    | All        |           | "                   | Thika272  | IMG  | "       | "         | 0°10'S;37°50'E     |
| Kitivo    | "    | Identical  | IX        | "                   | Kiboko275 | IMG  | "       | "         | "                   |
| Kitwa-I   | "    | "          |           | "                   | Thika279  | IMG  | "       | "         | "                   |
| Kimutwa   | "    | "          |           | "                   | Kiboko295 | IMG  | "       | "         | "                   |
| Mumbuni   | "    | "          |           | "                   | Kiboko277 | IMG  | "       | "         | "                   |
| Serere    | LAR  | All        |           | 0°10'S;37°50'E     | Kiboko9   | IMG  | "       | "         | "                   |
| Madioli   | "    | All        | X         | 0°16'N;34°45'E     | Kiboko10  | IMG  | "       | "         | "                   |
| Shikoti   | "    | Identical  |           | "                   | Kiboko11  | IMG  | "       | "         | "                   |
| Ikholi    | "    | "          |           | "                   | Kiboko159 | IMG  | "       | "         | "                   |
| Lugala    | LAR  | All        | XI        | 0°16'N;34°45'E     | Kiboko257 | IMG  | "       | "         | "                   |
| Shamiloli | "    | Identical  |           | "                   | Kiboko258 | IMG  | "       | "         | "                   |
| Shipalo   | "    | "          |           | "                   | Kiboko259 | IMG  | "       | "         | "                   |
| Fumbachai | LAR  | All        | XII       | 0°16'N;34°45'E     | Kiboko267 | IMG  | "       | "         | "                   |
| Isambe    | "    | "          |           | "                   | Kiboko268 | IMG  | "       | "         | "                   |
| Nyal-Kada | "    | "          |           | "                   | Kiboko269 | IMG  | "       | "         | "                   |
| KALRO     | IMG  | All        | XIII      | 3°40'S;39°45'E     | Kiboko270 | IMG  | "       | "         | "                   |
| Matano-Manne| LAR  | All        |           | "                   | Masokani-I | LAR  | "       | "         | 1°48'S;37°37'E     |
| Kakanjuni-II| "    | Identical  |           | "                   | Masokani-II | LAR  | "       | "         | "                   |
| Tajirika  | LAR  | All        |           | 3°40'S;39°45'E     | Muvila    | IMG  | "       | "         | "                   |
| Kaleso    | "    | "          |           | "                   | TC14      | IMG  | "       | "         | "                   |
| LAR   | IMG       | Locations        | GPS             |
|-------|-----------|------------------|-----------------|
| Cha-Vyango-II * | All Identical XIV | * | TC4-Katune IMG * | 0°16’N,34°45’E |
| Sokoke-I * | Identical | * | 99/0056 IMG * | 0°26’N,33°58’E |
| Chumani * |          | * | Kalimbini-II LAR * | 3°40’S,39°45’E |
| Kalimbini-III LAR | All Identical XV | 1°48’S,37°37’E | Kasuwanzaala IMG * | 0°26’N,33°58’E |
| Kalimbini-IV * |          | * | Kasukari (99/0127) IMG * | 3°40’S,39°45’E |
| Nya-Gang LAR |          | 0°16’N,34°45’E | Halu LAR * | 3°40’S,39°45’E |
| Nya-Yenga * | All Identical XVI | * | Kibandameno LAR * | 3°40’S,39°45’E |
| Nyakatineg-I * |          | * | Agriculture LAR * | 3°40’S,39°45’E |
| Kaliluni LAR |          | 0°16’N,34°45’E | Soyosoyo LAR * | 3°40’S,39°45’E |
| Kisimba * | All Identical XVII | * | Sokoke-II LAR * | 3°40’S,39°45’E |
| Isulu * |          | 0°16’N,34°45’E | Kankanjuni-I LAR * | 3°40’S,39°45’E |
|            |          | 0°16’N,34°45’E | Kankanjuni-III LAR * | 3°40’S,39°45’E |
|            |          | 0°16’N,34°45’E | Mkongo-I LAR * | 3°40’S,39°45’E |
|            |          | 0°16’N,34°45’E | Mkongo-II LAR * | 3°40’S,39°45’E |
|            |          | 0°16’N,34°45’E | Cha-Vyango-I LAR * | 3°40’S,39°45’E |

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

### Figures

**Figure 1**

| # | Region     | GPS             |
|---|------------|-----------------|
| 1 | Kakamega   | 0°16’N,34°45’E |
| 2 | Siaya      | 0°26’N,33°58’E |
| 3 | Kitui      | 0°10’S,37°50’E |
| 4 | Makueni    | 1°48’S,37°37’E |
| 5 | Kilifi     | 3°40’S,39°45’E |
Five (5) major cassava growing regions of Kenya where leaf samples of local landraces and improved genotypes were collected. These regions represent 100% areas within Kenya where cassava is cultivated. GPS indicates the global positioning system for the coordinates of the regions.

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

**Figure 3**
Hierarchical clustering dendrogram from identity by state (IBS) matrix estimation. The Red line represents the empirically determined distance threshold.
Figure 4

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

Figure 5

Barplot showing population structure modeled with 14 underlying sub-population groups from ADMIXTURE. The sample order of the hierarchical clustering was maintained for the ADMIXTURE plot for easy comparison of the out from the two grouping method. For the ADMIXTURE plot, the different colors represent the different sub-population while each bar represents each individual sample. Samples with just one color are pure lines from a single sub-population. Samples with more than one colors are admixture from different sub-populations.
Figure 6

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