

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/384803237>

Molecular characterization of cassava (*Manihot esculenta* Crantz) germplasm in Kenya

Article · October 2024

CITATIONS

0

READS

28

7 authors, including:



Charles Nyamwamu
University of Eldoret

8 PUBLICATIONS 1 CITATION

[SEE PROFILE](#)



Pascaline Jeruto
University of Eldoret

28 PUBLICATIONS 766 CITATIONS

[SEE PROFILE](#)



Elizabeth Wanjiku Njenga
University of Eldoret

28 PUBLICATIONS 189 CITATIONS

[SEE PROFILE](#)



Anne Owiti
University of Nairobi

4 PUBLICATIONS 9 CITATIONS

[SEE PROFILE](#)



Molecular characterization of cassava (*Manihot esculenta* Crantz) germplasm in Kenya

Nyamwamu Nyarang'o Charles*¹, Pascaline Jeruto¹, Elizabeth Njenga¹, Emmy Chepkoech², Anne A. Owiti³, Peter Futi Arama⁴ and Richard Mwanza Mulwa⁵

¹School of Science, Department of Biological Sciences, University of Eldoret, P. O. Box 1125 - 30100 Eldoret, Kenya.

²School of Agriculture and Biotechnology, Department of Biotechnology, University of Eldoret Kenya. P. O. Box 1125 - 30100 Eldoret, Kenya.

³Department of Biochemistry, University of Nairobi, P.O. Box 30197, Nairobi 00100, Kenya

⁴School of Science, Agriculture and Environmental Studies, Rongo University, P.O. Box 103 - 40404, Rongo, Kenya.

⁵Faculty of Agriculture, Department of Agronomy, Horticulture and Soil Science, Egerton University, P.O. Box 536 – 20201 Egerton, Kenya.

Received Date: September 14, 2024; **Accepted Date:** September 19, 2024; **Published Date:** September 24, 2024;

***Corresponding author:** Nyamwamu Nyarang'o Charles, School of Science, Department of Biological Sciences, University of Eldoret, P. O. Box 1125 - 30100 Eldoret, Kenya. Email: nyamwamucharles@gmail.com

Abstract

Globally, as a staple food crop, cassava (*Manihot esculenta* Crantz) provides millions of people with a substantial amount of carbohydrates. Selection of the appropriate parental forms for breeding programs is the most crucial decision made by plant breeders in order to maximize genetic variability and produce superior recombinant varieties. However, insufficient genetic diversity and population structure data regarding Kenyan cassava accessions hinder the appropriate breeding parent selection process. Thus, this study sought to determine the genetic diversity and population structure among 40 sampled cassava accessions grown in Kenya by use of start-codon-targeted (SCoT) molecular markers. The study utilized 15 SCoT molecular markers. A total of 119 fragments were amplified, of which 89.9% were polymorphic with an average of 7.13 polymorphic fragments per primer. The polymorphic information content (PIC) value and primer resolving power (Rp) of 0.35 and 3.44 respectively, revealed a moderate genetic diversity among the accessions. A dendrogram based on the unweighted pair group method of arithmetic means (UPGMA) grouped the 40 cassava accessions into two clusters at 0.35

genetic similarity coefficients. Bayesian structural analysis identified two subpopulations as well as a few admixed accessions. Analysis of molecular variance (AMOVA) revealed a variance of 84% within the subpopulations and 14% among the subpopulations. The moderate level of genetic variation in the cassava accessions that SCoT molecular markers were able to successfully identify, can serve as a tool for expanding the genetic base in cassava breeding initiatives. Cassava breeding and variety development may benefit from the selection and hybridization of parental lines from the various clusters and sub-clusters that have been established.

Keywords: Cassava, Start codon targeted, Genetic diversity, Germplasm, Kenya

Introduction

Cassava (*Manihot esculenta* Crantz) is not only a staple food crop in tropical regions but also a significant source of carbohydrates for millions of people across sub-Saharan Africa, Asia and South America (FAOSTAT, 2020). However, despite its importance, cassava production faces several challenges,

including diseases and the misidentification of varieties due to traditional naming systems (FAO, 2018).

In recent years, molecular characterization has become crucial in overcoming these challenges, aiding in the identification of cassava genotypes, understanding genetic diversity and enhancing breeding programs (Tumuhimise *et al.*, 2016; Orek *et al.*, 2020). Molecular markers, such as microsatellites (SSR), single nucleotide polymorphisms (SNP), and DNA sequencing technologies, have revolutionized the study of cassava germplasm. These tools allow for the accurate identification of varieties, differentiation of closely related genotypes and assessment of genetic diversity (Rabbi *et al.*, 2020).

Molecular characterization is particularly important in cassava due to the crop's vegetative propagation and the presence of morphologically similar but genetically distinct varieties. This makes traditional phenotypic characterization insufficient for accurate identification and conservation efforts (Abdulai *et al.*, 2022). Therefore, genetic markers have allowed researchers to overcome these limitations by providing precise, reliable data on genetic variation within and among cassava populations.

Moreover, molecular characterization has played a key role in cassava breeding programs aimed at developing disease-resistant and high-yielding varieties. In Kenya, efforts to combat cassava mosaic disease (CMD) and cassava brown streak disease (CBSD) have benefited from molecular tools, which have helped breeders identify and introgress resistance genes from wild relatives into cultivated varieties (Maruthi *et al.*, 2020). This approach has led to the development of improved cassava varieties with enhanced resistance to these devastating diseases, thereby ensuring food security for millions of smallholder farmers (Orek *et al.*, 2020).

Thus, molecular characterization is a powerful tool in cassava research, providing detailed insights into the genetic diversity of the crop and supporting efforts to improve its resilience and productivity. Through integration of molecular markers with traditional phenotypic assessments, researchers and breeders can ensure the conservation of genetic resources and the development of new varieties that meet the challenges of modern agriculture (Tumuhimise *et al.*, 2016; Wang *et al.*, 2020).

Assessment of genetic diversity and variety differentiation in cassava has been carried out in sub-Saharan Africa through agro-morphological traits and molecular markers (Aina *et al.*, 2019; Wang *et al.*, 2020; Rabbi *et al.*, 2020; Orek *et al.*, 2020). Despite the fact that agro-morphological traits are mostly used in the determination of genetic diversity, they do not demonstrate the true genetic relatedness of the accessions and are strongly influenced by environmental factors (Tumuhimise *et al.*, 2016; Maruthi *et al.*, 2020). Notably, molecular markers have been utilized as powerful tools for estimating genetic diversity in most plant species with great success and accuracy because they are abundant and unaffected by environmental

parameters (Rabbi *et al.*, 2020; Abdulai *et al.*, 2022). Various DNA-based markers such as restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), and simple sequence repeats (SSR) have been used in Africa for taxonomic classification, genetic diversity, association analysis and QTL mapping studies in cassava (Aina *et al.*, 2019; Rabbi *et al.*, 2020; Orek *et al.*, 2020; Abdulai *et al.*, 2022).

Start-codon-targeted (SCoT) markers are based on polymorphisms in the short, highly conserved region of plant genes surrounding the ATG of the start codon (Collard & Mackill, 2009). Therefore, they can distinguish genetic variation in specific genes linked to a specific trait. They are a simple, novel, cost-effective, highly polymorphic, and reproducible molecular marker that does not require prior sequence information. In addition, detection is carried out using agarose-gel-based techniques, making them simple and relatively cheap to use (Collard & Mackill, 2009).

The short flanking regions of the ATG start codon are highly conserved across plant species and have led to the widespread use of SCoT markers in quantitative trait locus (QTL) mapping, marker-assisted breeding, bulked segregate analysis, and genetic variation studies (Collard & Mackill, 2009).

Using SCoT markers to characterize genetic diversity in cassava germplasm has not been described in any molecular studies yet. Therefore, the objective of the current study was to determine the genetic diversity and population structure among cassava germplasm in Kenya. The information generated from this study will be a critical preliminary step towards developing strategies for future cassava conservation strategies, genetic improvement and variety development.

Materials and Methods

The Plant Materials

A total of 40 samples of the cassava accessions were proportionately taken from four distinct clusters derived from earlier morphological characterization studies (Nyamwamu *et al.*, 2023). Two months after planting, from each sample, 2 newly grown apical leaf tissues from a single stem, measuring about 6 cm were cut, packed in labelled polythene bags size 3 and then placed in a cool box for transportation to the Centre for Biotechnology and Bioinformatics (Biochemistry) Laboratory, University of Nairobi for molecular analysis.

Genomic DNA Extraction

The cassava leaf samples from the 40 samples were preserved in a freezer at -86°C. From each sample, 0.50 mg of the frozen leaves were ground using a mortar and pestle. Genomic DNA extraction protocol using the cetyltrimethyl ammonium bromide (CTAB) method (Doyle and Doyle, 1990) was adopted. The pre-heated CTAB extraction buffer, maintained at 65°C, was added to the ground leaf material. The mixture was then transferred into 1.5 ml Eppendorf tubes,

flicked to mix, and placed in a water bath set at 65°C for an hour. Then, the Eppendorf tubes were taken out of the water bath and centrifuged at 14,000 rpm for 10 minutes at room temperature (approximately 25°C). The supernatant, or the top layer, was carefully poured into clean Eppendorf tubes.

In a fume hood, chloroform: isoamyl alcohol, in a 24:1 ratio, was added to the tubes, and the mixture was gently agitated for 5 minutes. The tubes were then centrifuged again at 14,000 rpm for 15 minutes at room temperature. After centrifugation, the upper phase was carefully transferred into new, clean Eppendorf tubes. To precipitate the DNA, 750 µL of ice-cold isopropanol was added. The tubes were then centrifuged at 13,000 rpm for 10 minutes at room temperature. Once the centrifugation was complete, the supernatant was discarded.

The pellets remaining in the tubes were washed using 500 µL of 70% ethanol, cooled to -20°C, by gently flicking the tubes. After the washing, the ethanol was decanted, and any remaining liquid was carefully removed using a pipette, ensuring the pellet was not disturbed. The pellets were then dried under a vacuum for 15 minutes. Once dry, 100 µL of nuclease free water (NF-H₂O) was added to dissolve the pellets followed by treatment with RNase through the addition of 0.5 mg/mL Ribonuclease A and incubated in a water bath at 37°C for 30 min. They were then removed from incubation for DNA quality and quantity check.

DNA quality and quantity check

Weighing of 0.32 mg of agarose using a digital scale was carried out. Subsequently, 80 ml of TBEX buffer was measured with a 100 ml graduated cylinder. The weighed agarose and buffer solution were combined in a 500 ml conical flask. To this mixture, 0.5 µL of ethidium bromide, a stain, was added. The conical flask containing the mixture was placed in a fume hood to cool for 20 minutes. Once cooled, the molten agarose solution was poured into a gel tray, into which a well comb had already been inserted. The gel was allowed to solidify at room temperature for 30 minutes before the comb was carefully removed. The gel box was then filled with 1x TBEX buffer until the gel was fully submerged.

The gel box was placed into an electrophoresis tank, which was set to run. Following this, a DNA ladder was loaded into the first well. For each DNA sample, 0.7 µL was pipetted and mixed with 0.3 µL of loading dye before being loaded into the subsequent wells. The electrophoresis tank was allowed to run for one hour. The gel was removed from the tank and placed in UV spectrophotometer for the estimation of the quality and quantity of the extracted DNA samples by resolving these DNA samples. A final DNA concentration of 50 ng/µL was prepared and stored at -20 °C until use. The DNA was extracted from each of the 40 samples of the cassava accessions for subsequent SCoT analysis.

Optimization of Primer Conditions and SCoT-PCR Amplification

The SCoT-PCR analysis was carried out at the Centre for Biotechnology and Bioinformatics (Biochemistry) Laboratory, University of Nairobi according to protocols established by Collard and Mackill, (2009) with some minor modifications including optimization of the annealing temperature of the SCoT primers and the duration of PCR thermal cycling conditions. Twenty SCoT primers were tested for their ability to prime to DNA of cassava samples. The primers that either failed to amplify or produced faint bands were excluded from the study. Fifteen SCoT primers that produced consistent amplification and clear banding patterns were used for analysis of the genetic diversity of the forty sample cassavas.

PCR reactions were performed in 25 µL volume using 12.5 µL one Taq Quick-Load 2_ Master Mix with a standard PCR buffer (New England Biolabs, Hertfordshire, UK), 10 mM of the primer, 50 ng of template DNA, and the reaction mixtures were topped up to 25 µL with nuclease-free water.

PCR amplifications were performed in a Veriti thermocycler (Bio-Rad, Singapore) with the following thermocycling conditions: initial denaturation at 94 °C for 3 min followed by 35 cycles of denaturation at 94 °C for 1 min, primer annealing at 50 °C for 1 min, and extension at 72 °C for 1.5 min. The amplification process was completed with a 5 min final extension at 72 °C and the PCR products were maintained at 4 °C. The PCR reaction for each SCoT primer was performed at least twice using DNA from two individual samples of the same accession. Where the PCR amplifications and banding patterns were not consistent, a third PCR amplification was carried out. Only clear and reproducible bands were used in the data analysis.

Visualization of Amplified PCR Products and Data Analysis

The PCR products were resolved by electrophoresis on a 1.5% ethidium-bromide stained

agarose gel in 1X TBE buffer. Electrophoresis was carried out at 70 V for 60 min and PCR products were visualized using a Gel-Doc TM XR+ Imaging System (Bio-Rad, GmbH, FeldKirchen, Germany).

The molecular weights of the PCR products were estimated using a Gene Ruler 1 kb Plus DNA marker (Fischer Thermo Scientific, Waltham, MA, USA). All PCR-amplified SCoT fragments were detected on gels and scored as binary data for presence (1) or absence (0). Only reproducible and well-defined bands were scored.

Polymorphic and monomorphic bands were determined for each SCoT primer. The genetic similarity, resolving power for each primer, and genetic distances based on Nei's coefficients between pairs were analyzed using Popgene software, version 3.5 (www.ualberta.ca/fyeh/popgene.pdf, accessed on 25 August 2024).

Polymorphism information content (PIC) per locus was computed using Power Marker (version 3.25). Principal coordinate analysis (PCoA) and analysis of molecular variance (AMOVA) were performed using GENALEX 6.5 software (Peakall and Smouse, 2006) (accessed on 26 August 2024). The distance matrices were generated based on Jaccard's similarity coefficient. Similarity matrices were subjected to cluster analysis through the unweighted pair group method with arithmetic mean (UPGMA) and a dendrogram was constructed using FigTree software (Version 1.4.2; accessed on 27 August 2024).

The data from the 15 polymorphic SCoT markers were subjected to population structure analysis based on the admixture model clustering method in the software package STRUCTURE 2.3.4 (Pritchard et al., 2000). This model was run by varying the number of assumed population (K) from 1 to 10 (K = group numbers formed according to the STRUCTURE software (Version 2.3.4; accessed on 29 August 2024). A burn-in period of 10,000 and Markov Chain Monte Carlo (MCMC) replications of 100,000 after each burn-in was used. The optimum population (K) which best estimated the structure of the 40 cassava accessions was predicted using the Evanno's method (Evanno et al., 2005) through the online-based software STRUCTURE HARVESTER (Version A.2) (Earl and VonHoldt, 2012) (accessed on 29 August 2024). The model was repeated for the K at maximum DK with a burn-in period of 100,000 and an MCMC of 100,000 after each burn-in with ten alterations.

Results

SCoT Marker Analysis

Out of the 20 SCoT markers tested for their ability to amplify cassava DNA samples, 15 markers which showed polymorphic amplification fingerprint patterns were used to analyze the genetic diversity of the 40 samples of the cassava accessions. The 15 SCoT markers generated a total of 119 amplified DNA fragments with an average of 8 bands per marker (Table 1).

Out of the 119 amplified fragments, 107 (90%) were polymorphic. Similarly, out of the 15 SCoT markers, two markers namely SCoT12 and SCoT15 were the most polymorphic with 10 and 13 polymorphic bands respectively. The lowest numbers of polymorphic bands (3) were amplified using SCoT11 (Table 1). The polymorphic information content (PIC) value of all the SCoT markers ranged from 0.27 (SCoT11) to 0.37 (SCoT16, SCoT21, SCoT23 and SCoT29) with an average of 0.35 (Table 1). Nei's gene diversity (heterozygosity) varied from 0.28 (SCoT11) to 0.50 (SCoT16 and SCoT 21) with an average of 0.45 while Shannon information index ranged from 0.10 to 0.35 with an average value of 0.25. The highest and lowest Rp were 5.75 (for SCoT15) and 1.20 (for SCoT11), respectively (Table 1).

Table 1: Total and number of polymorphic bands, gene diversity and resolving power per SCoT marker used for the analysis of the 40 cassava accessions

No	SCoT Marker	Code of SCoT marker	NAB	NPB	%P	PIC	H	I	Rp	D	E	MI
1	SCoT2	ST2	10	9	90	0.36	0.48	0.217	4.2	0.84	4	0.004
2	SCoT 6	ST6	5	5	100	0.34	0.44	0.217	2.35	0.55	3.4	0.007
3	SCoT 9	ST9	5	5	100	0.36	0.47	0.306	3.05	0.86	1.9	0.004
4	SCoT11	ST11	6	3	50	0.24	0.28	0.104	1.2	0.31	5	0.006
5	SCoT12	ST12	10	10	100	0.35	0.44	0.219	3.85	0.89	3.3	0.004
6	SCoT15	ST15	13	13	100	0.33	0.42	0.256	5.75	0.91	3.9	0.003
7	SCoT16	ST16	7	6	85.7	0.37	0.50	0.330	3.2	0.74	3.6	0.006
8	SCoT 21	ST21	9	9	100	0.37	0.50	0.259	5	0.74	4.6	0.006
9	SCoT 23	ST23	10	8	80	0.37	0.50	0.274	5	0.72	5.3	0.006
10	SCoT26	ST26	7	6	85.7	0.33	0.42	0.196	2.6	0.51	4.9	0.007
11	SCoT29	ST29	7	6	85.7	0.37	0.50	0.206	1.9	0.74	3.6	0.006
12	SCoT32	ST32	9	8	88.8	0.35	0.46	0.239	3	0.87	3.3	0.004
13	SCoT33	ST33	6	6	100	0.36	0.46	0.332	4	0.87	2.2	0.004
14	SCoT34	ST34	6	6	100	0.36	0.46	0.352	3.9	0.60	3.8	0.007
15	SCoT35	ST35	9	7	77.7	0.36	0.47	0.200	2.65	0.86	3.4	0.005
		Total	119	107	89.57			3.705	51.65			
		Mean	7.93	7.13		0.35	0.45	0.247	3.44	0.734	3.75	0.005

NAB=Number of amplified bands; **NPB**=Number of polymorphic bands; **%P**=Percentage polymorphism; **PIC**=Polymorphic information content; **H**=Nei's gene diversity (Heterozygosity)= $I =$ Shannon information index; **Rp**=Resolving power; **D**=Discriminating power; **E**=effective multiple ratio; **MI**=Marker index

Genetic Diversity and Cluster Analysis of the Cassava Accessions

The genetic diversity and relationships among the studied cassava accessions were determined by Jaccard's similarity coefficient. Based on Jaccard's (J) similarity coefficient and using the UPGMA method, the dendrogram (Figure 1) divided the 40 cassava accessions into two major clusters (I and II) at a Jaccard's similarity coefficient of 0.35. The two major clusters namely I and II each was composed of 20 cassava accessions respectively. Cluster I comprised of 20 accessions that were divided into sub-clusters, sub-cluster I(a) with 19 cassava

accessions and sub-cluster I(b) with only 1 accession. All the 19 cassava accessions (75, 109, 112,113, 87,98, 88,114, 113, 68, 73,63, 89, 99, 105, 96, 125, 64 and 104 in sub-cluster I(a) were obtained from Migori, Homa Bay, Busia and Kilifi counties (Nyanza/western and Coastal zones) while the 1 cassava accession (74) in sub-cluster I(b) was obtained from Migori county. Cluster 2 comprises of 20 accessions (4, 6, 40, 7, 9, 14, 15, 16, 25, 18, 20, 24, 35, 36, 44, 58, 39, 61 62 and 10) and were also obtained from Makueni, Kilifi, Homa bay and Migori counties (Central/eastern, Coastal and Nyanza/western zones) as shown in Figure 1.

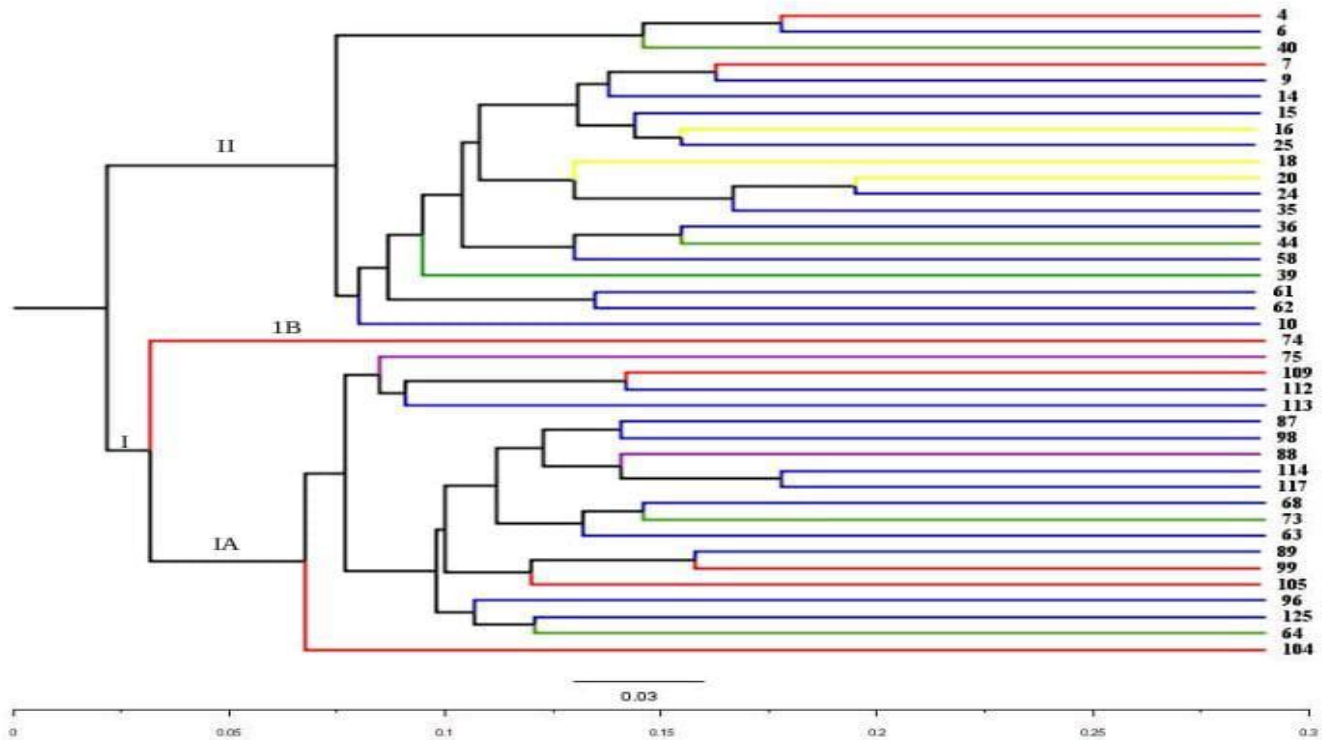


Figure 1: Dendrogram based on UPGMA showing the relationships among the 40 cassava accessions using SCoT marker data based on Jaccard's similarity index. Cassava accessions represented by similar colors in the dendrogram were obtained from the same county.

The similarity coefficient among the 40 accessions ranged from 0.35 to 0.78 with an average of 0.57. The highest similarity coefficient of 0.78 was recorded between accessions 4 & 7 and as well as between 114 &117 while the lowest value of 0.35 was obtained between 61 and 74 (Table 2).

Table 2: Pairwise genetic similarity among the 40 cassava accessions revealed by 15 SCoT Primers

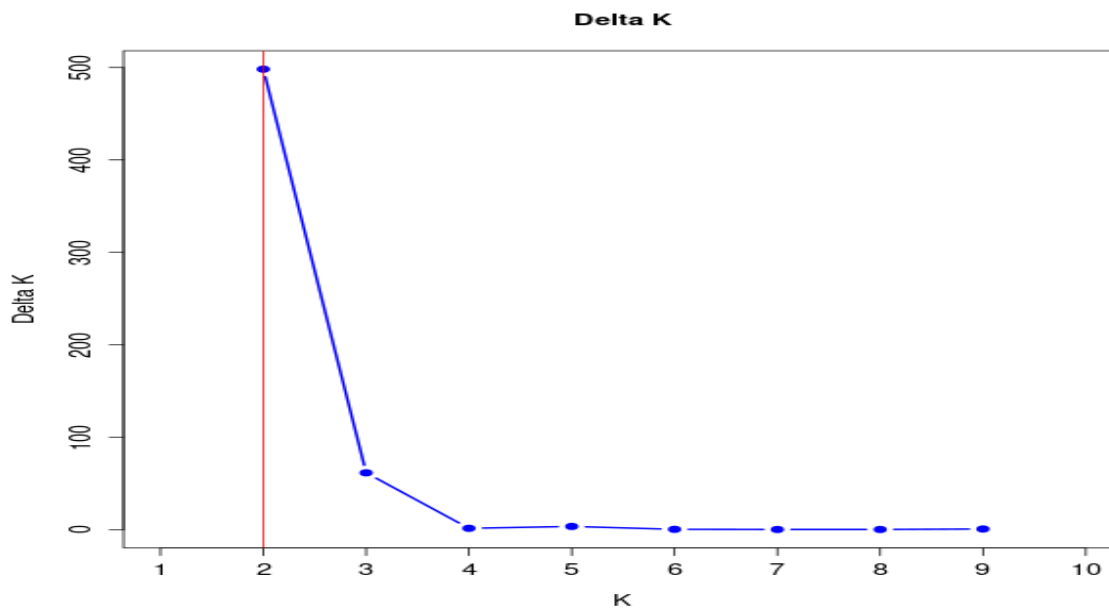


Figure 2: The estimated membership fraction using LnP(D)-derived delta K (DK) with cluster (K) ranged from 1 to 10 for K = 2.

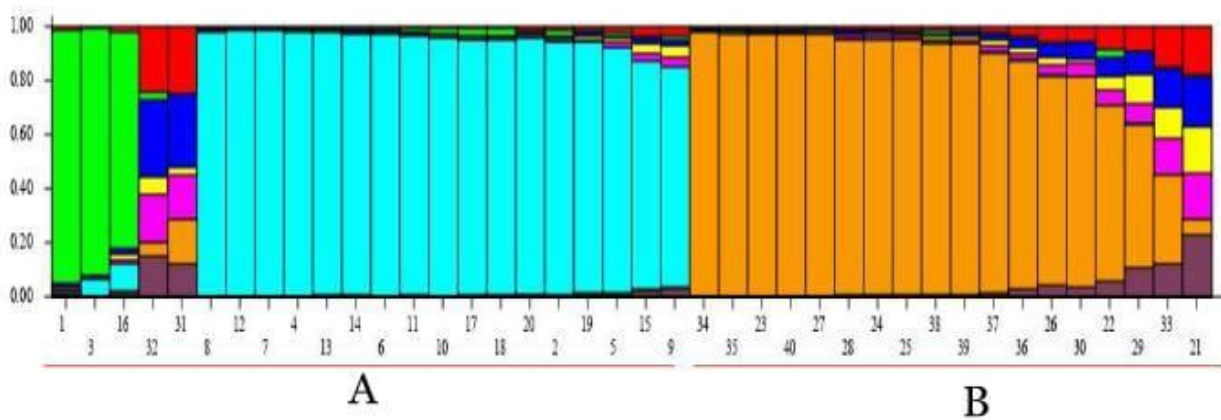


Figure 3: Population structure of 40 cassava accessions inferred STRUCTURE analysis based on SCoT marker data. Each color represents a single genetic group, namely A and B. Each solid bar represents a single accession, while each color represents a genetic group. The numbers 1–40 represents the different cassava samples.

Analysis of Molecular Variance (AMOVA)

Analysis of molecular variance (AMOVA) was used to evaluate the population differentiation among and within the

cassava accessions. The SCoT markers' data revealed 16% of the genetic variation among the population and 84% of the variation within the population (Table 4).

Table 4: Analysis of molecular variance (AMOVA) based on SCoT markers for the 40 cassava accessions

	Df	SSD	MSD	Variance component	Percentage of Variation	P-Value
Among Population	4	132.1	33.0	3.0	16%	0.002
Within Population	35	540.9	15.5	15.5	84%	
Total	39	673.0		18.4	100%	

*Significant at $p \leq 0.002$.

Principal Coordinates Analysis (PCoA)

In order to ascertain the genetic relationships between and within the populations and also assess the consistency of differentiation between the populations using cluster analysis, the principal coordinate analysis was carried out. The 40 cassava accessions were classified similarly by the PCoA analysis and the cluster analysis (Figure 3). The percentage of

the total variation explained by the first three dimensions of the PCoA axis was 35.69% (first axis = 22.99%, second axis = 6.54%, and third axis = 6.16%). Accessions from Busia, Homa bay, Makueni, and Migori counties (Western, Nyanza and eastern zones) were clustered together (Figure 4).

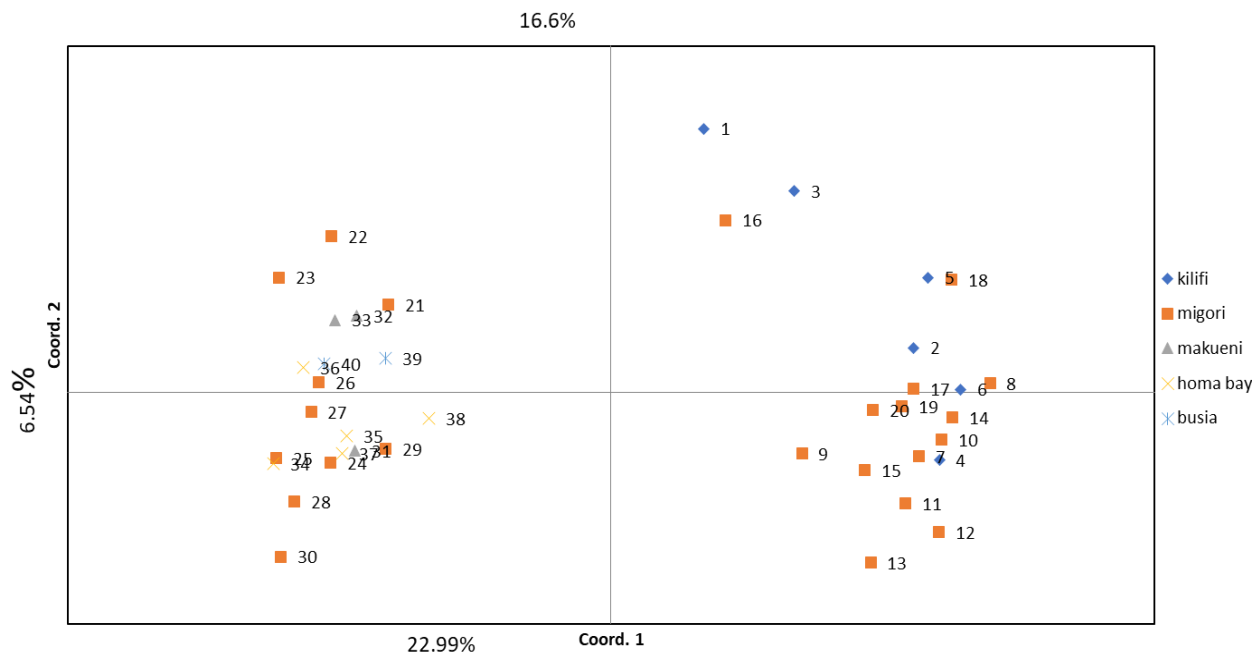


Figure 4. Bi-plots derived from principal coordinate analysis of 40 cassava accessions using SCoT data. The numbers plotted represents individual cassava samples. The colors represent county/site of collection of the accessions: red= Migori, Blue= Kilifi, Green= Makueni, Light blue= Busia, Purple= Homa bay.

Discussion

The genetic variation in cassava (*Manihot esculenta Crantz*) should be considered when developing conservation and utilization strategies as well as expediting breeding programs. Molecular markers are efficient and accurate tools to reveal and estimate genetic diversity and to determine the population structure of most plant species (Okogbenin *et al.*, 2019; Rabbi *et al.*, 2020). There are a variety of gene-based markers that have been developed to aid in the investigation of genetic diversity and population structure analyses in crop plants including SCoT molecular markers (Sánchez *et al.*, 2018; Sheat *et al.*, 2021).

The SCoT marker has been used in genetic diversity analysis, phylogenetic relationships and DNA fingerprinting of economically important food crops and medicinal plants including yam (*Dioscorea spp.*) (Owiti *et al.*, 2023), mango (Chen *et al.*, 2014), grape (Gupta *et al.*, 2019), orchid (Sun *et al.*, 2020), durum wheat (Jia *et al.*, 2018), rose (Yin *et al.*, 2020), Diospyros (Liu *et al.*, 2019), *Elymus sibiricus* (Yu *et al.*, 2020), *Vigna unguiculata* (Ndiaye *et al.*, 2018), *Taxus media* (Hu *et al.*, 2020), *Dendrobium* (Shen *et al.*, 2019),

Chrysanthemum morifolium (Zhao *et al.*, 2021), coconut (Wang *et al.*, 2018) and *Physalis species* (Pandey *et al.*, 2021).

Studies have indicated that SCoT markers have good capabilities in genetic research due to their ability to reveal polymorphisms in conserved regions and their high reliability as compared with other molecular systems (Chen & Gao, 2018; Gupta & Kumari, 2020). Notably, this is the first report on evaluating the genetic diversity and population structure of various cassava accessions and establishing phylogenetic relationships among them using SCoT markers. The 15 SCoT markers amplified 119 bands, out of which 107 (89.9%) were polymorphic and this high polymorphism indicates the informative nature of the SCoT markers used. The high level of polymorphism (89.9%) demonstrated a relatively high genetic diversity among genotypes of the cassava accessions. In other related studies, Zhao *et al.*, (2017) highlighted the high reliability and ability of SCoT markers to detect polymorphisms in conserved regions, making them suitable for assessing genetic diversity in various plant species. Their study on *Gossypium hirsutum* L. demonstrated that SCoT markers revealed a high level of polymorphism, with over 80% polymorphic bands, similar to the 90% polymorphism observed in cassava

accessions (Xiong *et al.*, 2011). In line with this study's findings, Owiti *et al.*, (2023) observed a relatively high level of genetic diversity (95%) using SCoT markers to detect polymorphisms in yams (*Dioscorea spp.*). Additionally, Guo *et al.*, (2012) reported a comparable polymorphism rate of 88% in *Carthamus tinctorius* L., emphasizing the utility of SCoT markers in revealing genetic variation and population structure in plant species. The consistent high polymorphism rates across different studies affirm the robustness of SCoT markers in genetic diversity studies.

In contrast, other molecular markers like Random Amplified Polymorphic DNA (RAPD) and Inter-Simple Sequence Repeats (ISSR) have shown varied levels of polymorphism, generally lower than SCoT markers. For example, Rodrigues *et al.*, (2019) reported a polymorphism level of 76% in cassava using RAPD markers, which is significantly lower than the 89.9% observed with SCoT markers in the current study. Similarly, Akinwale *et al.*, (2010) found that ISSR markers exhibited 72% polymorphism in their study on cassava, further supporting the assertion that SCoT markers provide more reliable and higher resolution in detecting genetic diversity among cassava genotypes. However, Que *et al.*, (2014) revealed 92.85% polymorphism using SCoT markers in determining the genetic diversity of sugarcane accessions within a local sugarcane germplasm collection from China. The high polymorphism revealed by SCoT markers in the present study not only aligns with but also exceeds the polymorphism rates reported in these previous studies, demonstrating their superior efficacy in genetic research and diversity studies.

It has been reported that Nei's gene diversity and Shannon's information index are important in the study of genetic diversity in plant species (Hodge *et al.*, 2017). In this study, the average Nei's gene diversity (expected heterozygosity) and Shannon information index were 0.45 and 0.25, respectively, indicating moderate genetic diversity for the cassava accessions. The results showed that the germplasm investigated in this study exhibited moderate genetic variation that can be exploited for cassava improvement. Similarly, Tumuhimbise *et al.*, (2018) reported moderate genetic diversity in Ugandan cassava germplasm with Nei's gene diversity values of 0.43 and a Shannon information index of 0.25. Their study emphasized that such genetic diversity is essential for maintaining the adaptability and resilience of cassava populations, especially in the face of pests and diseases like cassava mosaic disease. Another study by Ezenwaka *et al.*, (2017) on cassava accessions in Nigeria found Nei's gene diversity ranging from 0.42 to 0.48 and a Shannon information index of 0.24. These findings are consistent with the moderate genetic diversity observed in this current study, further affirming the importance of such diversity for cassava breeding programs aimed at enhancing yield, disease resistance, and climate adaptability. In contrast, Singh *et al.*, (2020) reported slightly higher genetic diversity in cassava accessions from India, with Nei's gene diversity values averaging 0.50 and a Shannon information index of 0.28. The slightly higher diversity in this case was attributed to the broader genetic base of the Indian cassava germplasm, which includes both indigenous and introduced varieties.

The PIC value represents the informativeness of a marker in detecting polymorphism and is usually used to reveal the differences among crop accessions based on genetic relationships. PIC values of less than 0.25 indicate a low level of polymorphism, values between 0.25 and 0.50 indicate intermediate polymorphism, and values more than 0.50 indicate high polymorphism (Botstein *et al.*, 1980; Ge, 2013). In this study, the PIC values ranged from 0.27 (SCoT11) to 0.37 (SCoT16, SCoT21, SCoT23 and SCoT29) with an average of 0.35, which confirms that the SCoT markers used in this study were highly informative and thus possess high discriminatory power based on the description by Botstein *et al.*, (1980). In addition, these results indicated that the SCoT markers used in this study are very informative and efficient and thus can be used for species authentication by developing species-specific sequence-characterized amplified region (SCAR) markers. The genetic relationships between the tested cassava accessions were estimated by calculating the similarity coefficients and the accessions were grouped based on the similarities using the UPGMA method. The similarity coefficient ranged from 0.35 to 0.78 with an average of 0.57, indicating moderate genetic diversity among the 40 cassava accessions. The highest similarity coefficient of 0.78 was recorded between accessions 4 & 7 and also between 114 and 117 indicating that these accessions are highly genetically similar. The lowest similarity index of 0.35 was obtained between 61 and 74, suggesting that these accessions were more genetically diverse and are appropriate for use for cassava improvement. The dendrogram constructed using the UPGMA method separated the 40 cassava accessions into two major clusters/groups at the similarity coefficient of 0.35.

The population structure was analyzed using STRUCTURE software to perform clustering, and the results showed that $K = 2$ had the best delta K (DK) interference. According to Pittard *et al.*, (2000), the Bayesian clustering technique helps identify population structure and distribute people or portions of genetic material among multiple clusters. STRUCTURE program assigns distinct accessions to various populations based on the discovered allele frequencies. The 40 cassava accessions used in this study were divided into two populations and an admixture group using population structure analysis.

The results of the principal coordinate analysis (PCoA) were comparable to those of the cluster analysis and STRUCTURE methods. Analysis of molecular variance, (AMOVA) showed that there was a significant level of genetic variability (84%), indicating that there was potential for effective selection from a mini-core collection of the tested population because all samples had a variety of genetic backgrounds. High genetic variability within the population indicates that different cassava accessions have diverged within a same population, suggesting that genetic resources are present at each collecting county. In line with these findings, a study by Rabbi *et al.*, (2020) on the genetic diversity of cassava in Nigeria found that 82% of the total genetic variance was attributed to within-population variability, which also indicated a high potential for effective selection from the genetic resources available. Similarly, Kawuki *et al.*, (2018) observed that 85% of the genetic variance in Ugandan cassava

germplasm was within populations, suggesting that significant genetic diversity exists even within small, localized populations. This trend is not unique to cassava; for instance, Ramu *et al.*, (2017) reported that in sorghum, the majority of genetic diversity was found within populations rather than between them, emphasizing the importance of preserving within-population genetic resources for breeding programs. These studies consistently demonstrate that high genetic variability within populations is a common characteristic in crops with wide geographical distributions, such as cassava, and affirms the importance of utilizing this diversity in breeding and conservation efforts. Notably, within-population selection enables identification of genotypes adapted to specific climatic conditions or management practices to achieve higher genetic gains during crop improvement. Thus, the study revealed a considerable genetic diversity among cassava germplasm in Kenya.

The moderate level of genetic variation in the cassava accessions that SCoT molecular markers were able to successfully identify could be a useful tool for expanding the genetic base of cassava breeding initiatives. In order to increase the breeding value of cassava accessions and enable crosses between accessions that are farther apart to provide hybrids that are likely to have higher genetic potential than their parents, parental selection can benefit from the results of genetic diversity and genetic distance. The accessions in this study were grouped into two major genetic clusters that included a large number of intermediates. Since these two genetic clusters would be treated as separate evolutionary units, choosing parents should be based on the larger inter-cluster distance.

The UPGMA dendrogram and PCoA revealed no relationship between the locations of the cassava accessions' growing zones and the clusters. When compared to samples from other counties, the UPGMA dendrogram revealed that most cassava accessions from the same cassava growing zone had a higher level of genetic similarity. As a result, crosses between the various geographical zones' cassava accessions should be made. The knowledge these results offer on Kenyan cassava accessions is critical for the preservation of genetic resources and the start of germplasm enhancement initiatives.

Conclusion

The SCoT markers used in this study are efficient, reliable, and informative in differentiating the studied cassava accessions. SCoT markers revealed the presence of a moderate genetic diversity among the 40 sampled cassava accessions in Kenya. The low level of PIC observed at some of the loci suggested a moderate level of genetic variability among the cassava accessions.

The highly polymorphic SCoT markers identified in this study could be used in generating useful molecular descriptors for fingerprinting cassava accessions from Kenya. The cluster analysis grouped the accessions into two clusters and sub-clusters irrespective of their collection zones/counties.

The moderate genetic diversity should be utilized for breeding strategies for the genetic improvement of cassava cultivars. The information generated in this study is of great interest for the design of future collections and the management of cassava germplasm.

In addition, some of the accessions used in this study could be a source of desirable genes and can be used by plant breeders to develop varieties with a wide genetic base that can respond better to climate change challenges.

Recommendations

Through utilization of the genetic variability identified through SCoT markers, breeders can create cassava cultivars with a broader genetic base, which is essential for resilience against biotic and abiotic stresses.

Such a database would facilitate the accurate identification and classification of cassava accessions, ensuring that valuable genetic resources are conserved and effectively utilized in breeding programs.

Furthermore, the establishment of a standardized protocol for SCoT marker analysis across different research institutions would promote consistency in the characterization of cassava germplasm, aiding in the global effort to improve cassava production.

Advanced molecular techniques such as gene mapping and genome-wide association studies (GWAS) should be employed to pinpoint these genes, facilitating their incorporation into breeding programs aimed at developing improved cassava varieties. This targeted approach will ensure that the genetic improvements are not only effective but also sustainable, contributing to the long-term viability of cassava production in Kenya and beyond.

Acknowledgements

The author would like to acknowledge the input, support and guidance accorded by the project study supervisors, University of Eldoret (Seed Grant-cohort 8), the Biological Sciences Department of University of Eldoret, Rongo University, Department of Agriculture and Environmental Studies and the biometricians. Great appreciation goes to Ann Owiti, a biotechnologist in the Centre of Biotechnology and Bioinformatics laboratory (University of Nairobi) who guided and offered much support in laboratory work.

Conflict of Interest

“The author(s) declare(s) that there is no conflict of interest.”

References

- Abdulai, A.-L., Asante, I.K., Manu-Aduening, J.A., Gracen, V., Offei, S.K., & Akromah, R. (2022). Genetic diversity assessment of cassava genotypes using simple sequence repeat (SSR) markers. *Journal of Plant Breeding and Crop Science*, 14(2), 73-82. <https://doi.org/10.5897/JPBCS2022.0910>.
- Aina, O.O., Dixon, A.G.O., & Akinrinde, E.A. (2019). Genetic diversity in cassava genotypes assessed using simple sequence repeat (SSR) markers. *African Journal of Biotechnology*, 18(10), 207-217. <https://doi.org/10.5897/AJB2019.16874>.
- Akinwale, M. G., Akinyele, B. O., Dixon, A. G. O., & Odiyi, A. C. (2010). Genetic variability among forty-three cassava genotypes in three agro-ecological zones of Nigeria. *Journal of Plant Breeding and Crop Science*, 2(5), 104-109.
- Botstein, D.; White, R.L.; Skolnick, M.; Davis, R.W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum Genet.*, 32, 314–331.
- Chen, Y., Li, M., & Gao, X. (2018). Assessing genetic diversity in cultivated barley (*Hordeum vulgare* L.) using SCoT markers. *Journal of Integrative Agriculture*, 17(3), 518-527. [https://doi.org/10.1016/S2095-3119\(17\)61712-6](https://doi.org/10.1016/S2095-3119(17)61712-6).
- Chen, C., Bock, C.H., & Hwang, C.-F. (2014). Characterization of mango germplasm by SCoT analysis. *Scientia Horticulturae*, 174, 20-25. <https://doi.org/10.1016/j.scienta.2014.04.008>.
- Collard, B.C.Y., & Mackill, D.J. (2009). Start codon targeted (SCoT) polymorphism: A simple, novel DNA marker technique for generating gene-targeted markers in plants. *Plant Molecular Biology Reporter*, 27(1), 86-93. <https://doi.org/10.1007/s11105-008-0060-5>.
- Doyle, J.J. and Doyle, J.L. (1990). Isolation of plant DNA from fresh tissue. *Focus*, 12, 13–15.
- Earl, D.A. (2012). VonHoldt, B.M. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour*, 4, 359–361.
- Evanno, G.; Regnaut, S.; Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* 2005, 14, 2611–2620.
- Ezenwaka, L. C., Anaduaka, E. G., Agbagwa, I. O., & Onyeka, T. J. (2017). Genetic diversity and structure of Nigerian cassava germplasm as revealed by SSR markers. *Journal of Plant Breeding and Crop Science*, 9(11), 230-239. <https://doi.org/10.5897/JPBCS2017.0655>.
- FAO. (2018). Save and Grow Cassava: A Guide to Sustainable Production Intensification. Rome: Food and Agriculture Organization of the United Nations. Website: www.fao.org.
- FAOSTAT. (2020). Production and Trade. Food and Agriculture Organization of the United Nations. Rome:FAO. [2020.http://www.fao.org/faostat/en/#data/QC/visualize](http://www.fao.org/faostat/en/#data/QC/visualize).
- Ge, H.; Liu, Y.; Jiang, M.; Zhang, J.; Han, H.; Chen, H. (2013). Analysis of genetic diversity and structure of eggplant populations (*Solanum melongena* L.) in China using simple sequence repeat marker. *Sci. Hortic*; 162, 71–75.
- Guo, D., Zhang, J., Liu, C., Wang, L., & Wang, Y. (2012). Genetic diversity and relationships of the safflower (*Carthamus tinctorius* L.) analyzed by SCoT and ISSR markers. *Plant Systematics and Evolution*, 298(10), 1737-1743.
- Gupta, M., Thind, S., & Boora, K. (2019). Genetic diversity analysis in grapevine (*Vitis vinifera* L.) using SCoT markers. *Vegetos*, 32(1), 154-160. <https://doi.org/10.1007/s42535-019-00138-7>.
- Gupta, P., & Kumari, S. (2020). SCoT marker-based assessment of genetic diversity and population structure in rice (*Oryza sativa* L.) landraces from eastern India. *Genetika*, 52(2), 441-458. <https://doi.org/10.2298/GENSR2002441G>.
- Hodge, T. R., Katan, J., & Wong, T. (2017). Comparative analysis of genetic diversity using Nei's gene diversity and Shannon's information index in horticultural crops. *Journal of Horticultural Science & Biotechnology*, 92(5), 469-479. <https://doi.org/10.1080/14620316.2017.1352575>.
- Hu, J., Zhang, Z., Guo, Y., & Tang, J. (2020). Genetic diversity and differentiation of *Taxus media* germplasm based on SCoT markers. *Molecular Biology Reports*, 47(2), 1395-1402. <https://doi.org/10.1007/s11033-020-05208-w>.
- Jia, Y., Li, X., Zhang, X., & Liu, G. (2018). Evaluation of genetic diversity of durum wheat (*Triticum turgidum* L.) germplasm using SCoT markers. *Genetic Resources and Crop Evolution*, 65(6), 1691-1701. <https://doi.org/10.1007/s10722-018-0656-2>.
- Kawuki, R. S., Tumuhimbise, R., Kanju, E., Masumba, E., Wanjala, B. W., Nzuki, I. & Ferguson, M. E. (2018). Genetic variation among cassava clones selected for drought tolerance in East Africa. *Crop Science*, 58(1), 1-12. <https://doi.org/10.2135/cropsci2017.05.0290>.
- Liu, C., Yang, J., Yu, Z., & Zou, W. (2019). Genetic diversity analysis of *Diospyros* species using SCoT markers. *Plant Systematics and Evolution*, 305(3),

- 189-197. <https://doi.org/10.1007/s00606-019-01593-2>.
23. Maruthi, M.N., Jeremiah, S.C., Mohammed, I.U., & Legg, J.P. (2020). The role of molecular characterization in cassava brown streak disease resistance breeding. *Frontiers in Plant Science*, 11, 617-624. <https://doi.org/10.3389/fpls.2020.00617>.
24. Ndiaye, A., Ngom, M., Fall, N.J., & Gueye, M. (2018). Genetic diversity analysis of cowpea (*Vigna unguiculata* L. Walp.) landraces using SCoT markers. *Journal of Plant Breeding and Crop Science*, 10(11), 291-298. <https://doi.org/10.5897/JPBCS2018.0754>.
25. Nyamwamu Nyarang'o Charles, Pascaline Jeruto, Elizabeth Njenga, Peter Futi Arama and Richard Mwanza Mulwa, (2023). Phenotypic characterization of cassava (*Manihot esculenta* Crantz) germplasm in Kenya. *ASRIC Journal on Agricultural Sciences Vol.4* (1); 152-162. Available online at www.asric.org.
26. Okogbenin, E., Egesi, C. N., Mba, C., Angel, F., Perez, J. C., Augusto, R. B. &
27. Tohme, J. (2019). Advances in cassava genomics, genetics and breeding. *Plant Molecular Biology*, 100(6), 527-550.
28. Okogbenin, E., Setter, T. L., Ferguson, M., Mutegi, R., Ceballos, H., Olasanmi, B., & Fregene, M. (2019). Phenotypic approaches to drought in cassava: Review. *Frontiers in Physiology*, 10, 392. <https://doi.org/10.3389/fphys.2019.00392>.
29. Orek, C., Ombori, O., Wasike, V. W., & Odeny, D. A. (2020). Identification and characterization of cassava (*Manihot esculenta* Crantz) germplasm in Kenya using molecular markers. *Plant Genetic Resources: Characterization and Utilization*, 18(1), 55-63. <https://doi.org/10.1017/S1479262119000345>.
30. Owiti, A. A., Bargul, J.L., Obiero, G.O., Nyaboga, E.N. (2023). Analysis of Genetic Diversity and Population Structure in Yam (*Dioscorea Species*) Germplasm Using Start Codon Targeted (SCoT) Molecular Markers. *Int. J. Plant Biol.*, 14, 299-311. <https://doi.org/10.3390/ijpb14010025>.
31. Pandey, A., Gaikwad, S.S., & Mandal, B. (2021). Molecular characterization of *Physalis* species using SCoT markers. *Scientia Horticulturae*, 287, 110268. <https://doi.org/10.1016/j.scienta.2021.110268>.
32. Peakall, R., Smouse, P. E. (2006). GenAlEx 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* 2006, 6, 288-295.
33. Pritchard, J. K., Stephens, M. and Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 2000, 155, 945-959.
34. Que, Y., Pan, Y., Lu, Y., Yang, C., Yang, Y., Huang, N., Xu, L. (2014). Genetic analysis of diversity within a Chinese local sugarcane germplasm based on Start Codon Targeted Polymorphism. *BioMed Res. Int.* 2014, 468375.
35. Rabbi, I.Y., Kayondo, S.I., Bauchet, G., Lozano, R., & Wolfe, M. (2020). Genome-wide association mapping of root yield and agronomic traits in cassava. *Euphytica*, 216(8), 124-133. <https://doi.org/10.1007/s10681-020-02662-2>.
36. Rabbi, I. Y., Udoh, L. I., Wolfe, M., Parkes, E. Y., Gedil, M. A., Dixon, A. G. O. & Kulakow, P. A. (2020). Genome-wide association analysis reveals new insights into the genetic architecture of defensive, agro-morphological, and quality traits of cassava (*Manihot esculenta* Crantz). *Plant Molecular Biology*, 104(4), 451-468.
37. Rabbi, I. Y., Hamblin, M. T., Kumar, P. L., Ikpan, A. S., & Gedil, M. A. (2020). Genomic prediction of resistance to cassava mosaic disease in diverse clones from Africa using genotyping-by-sequencing. *BMC Genomics*, 21(1), 847. <https://doi.org/10.1186/s12864-020-07237-6>.
38. Ramu, P., Esuma, W., Kawuki, R. S., Rabbi, I. Y., Egesi, C., Kulakow, P. & Bhattacharjee, R. (2017). The genetic diversity of local cassava landraces in Uganda and the effect of farming practices on their on-farm conservation. *Journal of Crop Improvement*, 31(1), 77-89. <https://doi.org/10.1080/15427528.2016.1245080>.
39. Rodrigues, M. A. T., Silva, M. F. F., Gonçalves, A. A., Moraes, M. C. S., & Veasey, E. A. (2019). Genetic diversity in cassava germplasm based on RAPD and morphological traits. *Genetics and Molecular Research*, 18(1), gmr16039952.
40. Sánchez, E., Ceballos, H., & Dufour, D. (2018). Carotenoids and cassava: The need for targeted interventions to achieve nutritional impact. *Current Opinion in Biotechnology*, 56, 108-114. <https://doi.org/10.1016/j.copbio.2018.09.009>.
41. Sheat, S., Sultan, A., Al-Kateb, H., & Alsamman, A. (2021). A comprehensive study of cassava's genetic diversity and population structure using SCoT markers. *Journal of Agricultural Science*, 13(3), 154-166. <https://doi.org/10.5539/jas.v13n3p154>.
42. Shen, Y., Zhou, X., & Liu, X. (2019). Genetic diversity analysis of *Dendrobium* germplasm using SCoT markers. *Scientia Horticulturae*, 243, 272-278. <https://doi.org/10.1016/j.scienta.2018.08.036>.
43. Singh, S., Singh, R. K., Sharma, S. K., & Chandel, P. (2020). Molecular characterization and genetic diversity analysis of cassava (*Manihot esculenta* Crantz) accessions in India using SSR markers. *Plant Genetic Resources: Characterization and Utilization*, 18(3), 173-181. <https://doi.org/10.1017/S147926212000020X>.
44. Sun, J., Fan, X., & Zhang, Z. (2020). Molecular identification and genetic diversity analysis of

- Orchidaceae* germplasm based on SCoT markers. *BMC Plant Biology*, 20(1), 118-124. <https://doi.org/10.1186/s12870-020-2321-9>.
45. Tumuhimbise, R., Melis, R., Shanahan, P. E., & Kawuki, R. S. (2018). Genetic diversity and population structure of cassava (*Manihot esculenta* Crantz) landraces and breeding lines assessed by SSR markers. *Plant Genetic Resources: Characterization and Utilization*, 16(1), 30-39. <https://doi.org/10.1017/S1479262116000397>.
46. Tumuhimbise, R., Melis, R., Shanahan, P., & Kawuki, R.S. (2016). Genetic variation and heritability of resistance to cassava brown streak disease in selected cassava genotypes in Uganda. *Euphytica*, 210(1), 97-109. <https://doi.org/10.1007/s10681-016-1706-3>.
47. Wang, B., Zhang, F., Song, Q., Shen, Q., & Lu, Y. (2018). Genetic diversity analysis of coconut (*Cocos nucifera* L.) germplasm using SCoT markers. *Agronomy*, 8(4), 55-63. <https://doi.org/10.3390/agronomy8040055>.
48. Wang, B., Zhang, F., Song, Q., Shen, Q., & Lu, Y. (2020). Genome-wide association study of starch content in cassava root using SNP markers. *Molecular Breeding*, 40(4), 36-42. <https://doi.org/10.1007/s11032-020-1115-1>.
49. Xiong, F.; Zhong, R.; Han, Z.; Jiang, J.; He, L.; Zhuang, W.; Tang, R. (2011). Start codon targeted polymorphism for evaluation of functional genetic variation and relationships in cultivated peanut (*Arachis hypogaea* L.) genotypes. *Mol. Biol. Rep*, 38, 3487–3494.
50. Zhao, X., Han, J., Hu, Y., Fu, C., Yang, Z., & Zhao, Y. (2017). Genetic diversity and population structure of *Capsicum annuum* L. revealed by SCoT and RAPD markers. *Scientia Horticulturae*, 225, 360-368. <https://doi.org/10.1016/j.scienta.2017.07.007>.

Citation: Nyamwamu CN, Jeruto P, Njenga E, Chepkoech E, Owiti AA, Arama PF, Mulwa RM (2024) Molecular characterization of cassava (*Manihot esculenta* Crantz) germplasm in Kenya. *Adv Agri Horti and Ento: AAHE-210*