Evaluation of population genetics of Moringa oleifera Provenances from Coastal Kenya using Single Nucleotide Polymorphism (SNPs) markers

Sarah Kwamboka Ondieki[1*]
Jackson C. Korir[2]
Carol W. Hunja[3]

[1,2]School of Natural and Applied Sciences, Department of Biological Sciences, Masinde Muliro University of Science and Technology, P.O Box 190-50100 Kakamega, Kenya
[3] School of Science and Computing, South Eastern Kenya University, P. O Box 170-90200 Kitui, Kenya

*Corresponding Author: Sarah Kwamboka Ondieki, email: kwambokahsarah@gmail.com

ABSTRACT

Population genetics analysis is a prerequisite to understanding how and why genotypes and allele frequencies and change over time between and within populations. Consequently, it offers insight into the process of evolutionary change and makes it possible to map variants linked to traits that differ among populations. In the present study, Single Nucleotide Polymorphisms (SNPs) markers were utilized to study the genetic characterization of 17 provenances from the Coast region of Kenya. 164 genotypes of Moringa oleifera were selected from 17 populations and genome sequencing undertaken utilizing genotyping by sequencing (GBS). Identification of polymorphisms (SNP Calling) in the selected genotypes and population genetic studies were carried out. SNP calling was done by Illumina's SNP caller algorithm in the CASAVA software. 20,921 SNPs were called with an average call rate of 0.82. Average polymorphsm content (PIC) for the SNPs was 0.24 and reproducibility was 0.98. A phenetic tree was constructed using a neighbor-joining approach using DArT R. For the population genetics analysis, F statistic (Fst) utilising the functions StAMPP package in DArT R was performed whereby Gede and Samburu exhibited the least heterozygosity/correlation with a value of 0.0003 whereas Pwani University and Samburu had the highest correlation of genes at 0.37. Euclidean was used as a measure distance, and the average distance between the populations was 33.024. The molecular variance analysis (AMOVA) described a lower, 2.55%, variation within the population and 2.73% among the populations. The high similarity between the genotypes could be attributed to the Moringa plants in the various provenances having the same ancestry. This study may help identify links between gene allelic forms and phenotypes, allowing the alleles to be connected to desired characteristics such as rapid growth and high yield (functional analysis), because of the high frequency of SNPs and their role as a source of allele variations.

Keywords: Genotypes, Polymorphisms, Population Genetics

I. INTRODUCTION

Genotyping by Sequencing (GBS) is a novel application to discover and genotype SNPs in crop genomes and populations utilizing NGS protocols for crop improvement (J. He et al., 2014). Based on genotyping-by-sequencing, DArT recently created a new genotyping platform called DArTseq which sequences the genomic DNA sample's most informative representations (Hassani et al., 2020), (Nadeem et al., 2018). Sample analysis for this study involved a high-throughput genotyping method using DArTseqTM technology from Diversity Arrays Technology Pty Ltd (Canberra, Australia). For next-generation sequencing, HiSeq2500 and DArT-SeqTM technology are utilized. DArTSeq, also referred to as DArT-Seq markers, is used to detect both SNPs and presence-absence sequence variations (Raman et al., 2014). New methodologies for sequence-based genotyping have been devised to put the power of NGS to work for plant breeding and genomics. Through the recent advancement of genotyping-by-sequencing (GBS) techniques, thousands of loci can be produced by sequencing a specific subset of the entire genome (Elshire et al., 2011). GBS is a feasible technology that utilizes enzyme-based complexity reduction (targeting only a small section of the genome with restriction endonucleases) with DNA barcoded adapters to create multiplex libraries of samples appropriate for NGS sequencing. This approach is known as reduced restriction sequence (RRS) (Wang et al., 2020). This technique has been shown to function across a variety of species and can generate tens of thousands to hundreds of thousands of molecular markers (Elshire et al., 2011). Given that *Moringa oleifera* is a cross-pollinated tree, significant variance in quantitatively inherited traits has been documented, and high heterogeneity across multiple character forms has been observed, leading to vast genetic diversity in both natural and cultivated accessions. These findings should be taken advantage of for focused research aimed at improving the crop. Comprehensive understanding of the gene flow pattern and population genetic structure of *M. oleifera* via molecular genetic diversity and global collection population structure holds significant potential (Lakshmidevamma et al., 2021). It is essential and necessary to characterize population structure and genome diversity across various sources for breeding programs to effectively manage germplasm (Hassani et al., 2020). Polymorphism detection at the genome level exposes the evidence of variation and polymorphism, allowing

Licensed Under Creative Commons Attribution (CC BY-NC)

for easy detection of genetic diversity within a population (Wu et al., 2010), (S.-T. He et al., 2021). Comprehending the genetic diversity within advanced breeding lines could then be used to enhance varietal selection and cross breeding (Kumar et al., 2017). Molecular markers are potent biotechnological tools that are frequently employed to evaluate genetic variations among plant species with the aim of differentiating genetically unique genotypes and create nutrient-rich, high-yielding cultivars. (Sharma et al., 2024). The laborious process of developing markers can now be avoided by using sequences to in the detection and scoring of single nucleotide polymorphisms (SNPs) thanks to recent advancements in high-throughput sequencing (Chung et al., 2017). SNPs are the most recent molecular markers and can thus be utilized to detect genetic variations in plants as they are more efficient and prominent (Morgil et al., 2020). The present study aimed to undertake the molecular characterization 164 *M. oleifera* genotypes using DArTseq genotyping platform and thus evaluate the genetic population structure of the 17 provenances. As far as we are aware, this is the very first report for utilizing of SilicoDArTs and SNPs molecular variability for *Moringa oleifera* from the Coastal region of Kenya. In the future, genome-wide association mapping projects can utilize the generated DArTseq markers in the identification of QTLs and genes for traits that are desirable. This collection of markers will be extremely valuable for breeding programs.

II. MATERIALS & METHODS

2.1 Collection of Plant Materials

The original introduction sites from Kenya's Coast and some areas of its Eastern region were used to select provenances (fig. 1). This study was carried out in South Eastern Kenya University (SEKU) located 15 km off Kwa Vonza Market, along the Kitui-Machakos main road, Yatta/Kwa Vonza location, Lower Yatta Sub-County, Kitui County. The site lies at latitudes and longitudes of 037.755460 E 01.313580S respectively. The site elevation is 1173m a.s.l. as evident in figure 4 (Festus et al., 2022).164 most productive plants were selected (9 to 10 plants per provenance as seen in table 1) using a randomized complete block design (RCBD) (Festus et al., 2022).

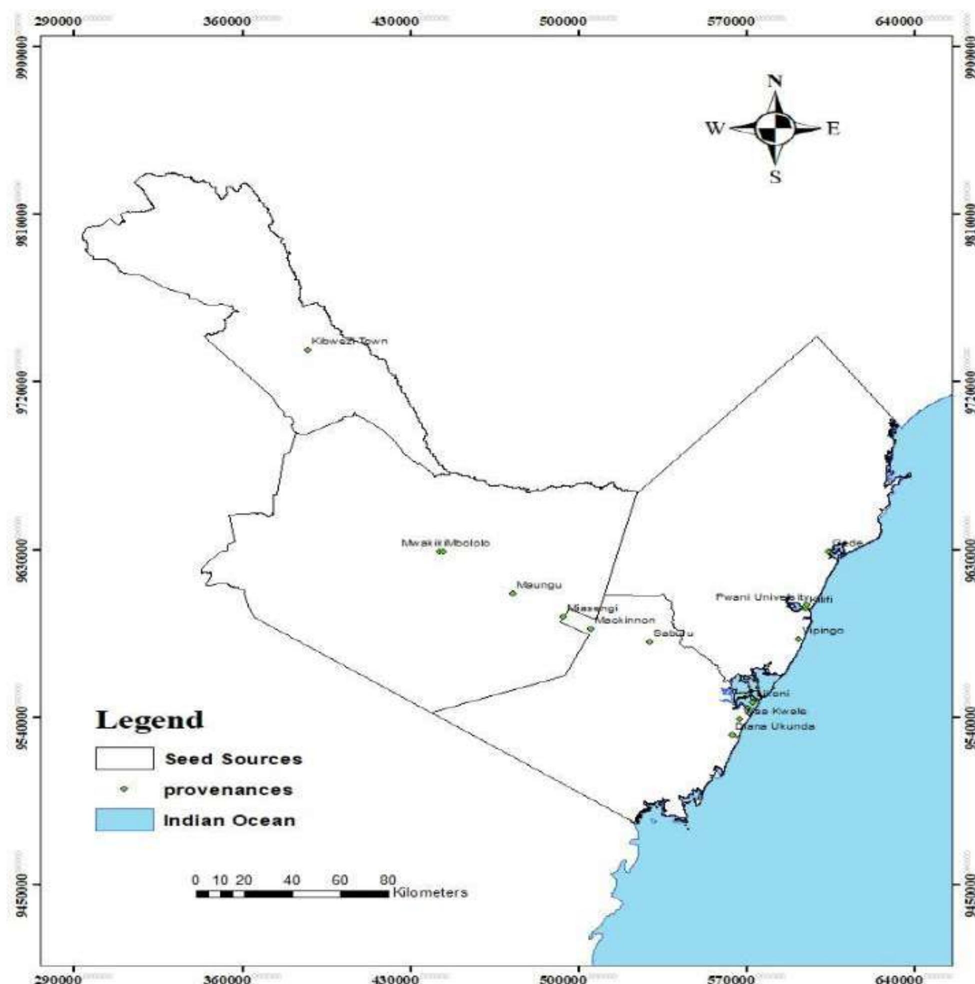

Figure 1
Location of Seed Collection Areas (Provenances) (Festus et al., 2022)

Licensed Under Creative Commons Attribution (CC BY-NC)

Table 1
No of Trees sampled per provenance

	Provenances	No. of trees sampled
1	Mwakiki	10
2	Pwani University	10
3	Shika Adabu	10
4	Miasenyi	10
5	Diani-Ukunda	10
6	Maungu	10
7	India	10
8	Kilifi	10
9	Likoni	10
10	Kibwezi Town	10
11	Mbololo	10
12	Gede	9
13	Arabuko Sokoke	9
14	Mackinon	9
15	Waa Kwale	9
16	Vipingo	9
17	Samburu	9
	Total	**164**

2.2 Sample preparation

Approximately 10 young and fresh *M. oleifera* leaves from the 164 sampled trees were selected and placed in a microtube. The samples were lypholized overnight at Seqart Africa Laboratories situated at International Livestock and Research Institute (ILRI), Kenya. The dried *M. oleifera* samples were then grinded using a Geno Grinder at 1300 rpm for 3 minutes. The resultant samples were utilized in DNA isolation (Ondieki et al., 2023).

2.3 Genotyping by Sequencing

DNA extraction and purification

Genomic DNA isolation was performed using NucleoMag® 96 Plant according to manufacturer's protocol. Quantification of the isolated DNA was done using 0.8% Agarose gel to visualize bands.

2.4 Library Preparation

A genomic representation of the samples was produced by combining two restriction enzymes (Pst1-CTGCAG-, Mse I) with adapters connected by barcode to identify each sample to function within a single lane on the Illumina HiSeq2500 device. The ends of the DNA fragments of interest were modified by adding DNA barcoded adapters and ligase to facilitate ligation with primers for amplification. MSE 1 was used as a common adapter and PST 1 enzyme as a barcode adapter. Effective PCR amplification was achieved at the MSE I site. 1.2% Agarose gel electrophoresis was used for quality control of the GBS library. The amplified fragments were combined in equimolar proportions using Tecan, an equipment designed for combining big samples. After applying the pooled samples to the Illumina c-Bot bridging PCR, a fully automated clonal cluster generation process, the Illumina HiSeq2500 system was used for sequencing. Sequencing libraries were prepared in compliance with SEQART Africa optimized protocols. The sequence DNA libraries were sequenced using an Illumina Hiseq 2500 (Ondieki et al., 2023).

III. DATA ANALYSIS

The analytical program developed and patented by DArT Pvt. Ltd., Australia, was used to generate two types of data, scores for ''presence/absence'' markers (dominant), called SilicoDArTs and SNP markers (Spinoso-Castillo et al., 2020). Several quality parameters for DArTseq and SNP markers including call rate, polymorphic information content (PIC), and reproducibility, were computed by the DArTsoft v.7.4.7 (Hassani et al., 2020). This information was provided by DArT Pty Ltd to produce DArTSNP (table 4) and SilicoDArT (table 5). High quality markers were identified for this particular investigation by filtering the FASTQ data based on quality parameters, (Ondieki et al., 2023). Data resulting from the 164 genotypes was analyzed using KDCompute, a platform provided by Seqart Africa for various analyses and Diversity Arrays Technology (DArT) R, an R stats package developed by DArT. Reproducible

Licensed Under Creative Commons Attribution (CC BY-NC)

and well-marked bands were scored for the presence (1) and absence (0) of single nucleotide polymorphisms (SNPs) using Gel Analyzer 2010a to form a binary data matrix (Spinoso-Castillo et al., 2020). Polymorphism information content (PIC) was calculated using silico DArT. Analysis of molecular variance (AMOVA) was computed based on the genetic distance matrix. For the population genetics analysis, F statistic (Fst) utilising the functions StAMPP package in DArT R was performed. Euclidean was used as a measure of distance. A phenetic tree was constructed using a neighbor-joining approach using DArT R. Analysis of molecular variance (AMOVA) was utilized to carry out population genetic analysis using KDCompute Software (silicoDArT) and DArT R.

IV. RESULTS

4.1 To Identify Polymophisms in the Selected Genotypes

A total of 37,853 markers (SilicoDArTs; table 5 and DArTSNPs; table 4) were scored using DArTsoft v.7.4.7 as binary data (1 present, 0 absent, for a marker in the genomic DNA of each accession). Several quality parameters for DArTseq and SNP markers including call rate, polymorphic information content (PIC), and reproducibility, were computed by the DArTsoft v.7.4.7. There were16,932 SilicoDArTs and 20,921 DArTSNPs markers.

The heterozygosity of the samples (FreqHet) ranged from 0 to 0.698 and exhibited an average heterozygosity of 0.11. The proportion of samples, which score as homozygous for the SNP allele (FreqHomSnp), ranged from 0 to 0.99 with an average of 0.2 (table 4). The proportion of samples for which the genotype score is “1” in the SNP allele row (OneRatioSnp) ranged from 0.006 to 1 with an average of 0.3. The average count of the reference, which is the sum of the tag read counts for all samples divided by the number of samples with non-zero tag read counts for the reference allele (AvgCountRef) ranged between 2.57 to 261.87 whereas its average was 24.96.

The average count of the SNPs, which is the sum of the tag read count for all samples divided by the number of samples with non-zero tag read counts for the SNP allele (AvgCountSnp) ranged between 2.5 to 181.97 and the average was 16.77. The proportion of technical replicate assay pairs for which the marker score is consistent (RepAvg) was 1. The proportion of samples for which the genotype score is "1" (One Ratio) was 1. The polymorphism information content (PIC) for the SNP allele row (PICSnp) ranged from 0 to 0.5 and exhibited an average of 0.24 whereas the average of the PIC of the Reference and SNP allele rows exhibited a range of 0.006 to 0.5 with average of 0.2 (table 2).

Table 2
DArT SNP marker quality and diversity

Marker quality and diversity	Mean	Min	Max
Call rate	0.82	0.396	1
Reproducibility	0.98	0.006	1
One ratio reference allele	0.8	0.006	1
SNP allele frequency	0.3	0.006	1
Heterozygosity	0.11	0	0.698
PIC for SNP	0.24	0	0.5

Licensed Under Creative Commons Attribution (CC BY-NC)

Table 3

Shows the SNP position

AlleleID	AlleleSequence	TrimmedSequence	SNP	SnpPosition
100081123	TGCAGAACAGCTACTCTTCCCTTGATATCCATCTTGGGCAGTCATATTCTGTCTTGGTCACAGCTAACC	TGCAGAACAGCTACTCTTCCCTTGATATCCATCTTGGGCAGTCATATTCTGTCTTGGTCACAGCTAACC	33:T>C	33
100081125	TGCAGATAATTCGTCGTGCGAATGAGAATGGGGAAGATCCATTACATCTAAGCAACAGATTTTGTGAAG	TGCAGATAATTCGTCGTGCGAATGAGAATGGGGAAGATCCATTACATCTAAGCAACAGATTTTGTGAAG	19:G>A	19
100081139	TGCAGCTCTGATTCGTAAAGTTAGAGGCACATCATTCTTCTCTAGAGTTGCTTCCTGAGTTTCTGAAAG	TGCAGCTCTGATTCGTAAAGTTAGAGGCACATCATTCTTCTCTAGAGTTGCTTCCTGAGTTTCTGAAAG	20:T>C	20
100081142	TGCAGCTTGAATTTGAAAGCAAGGGCTTCAGGTTTATGTCACCTGCATATGTTTCTGGGACTTCTTGTA	TGCAGCTTGAATTTGAAAGCAAGGGCTTCAGGTTTATGTCACCTGCATATGTTTCTGGGACTTCTTGTA	26:T>C	26
100081143	TGCAGCTTTCGTTTTTGTAAAGTCAGCTGAACCACTAGGGACACATTTTCCATGTGAAAAAAACAACTC	TGCAGCTTTCGTTTTTGTAAAGTCAGCTGAACCACTAGGGACACATTTTCCATGTGAAAAAAACAACTC	23:C>T	23
100081157	TGCAGTAACATCAAACCAAGTATCTTGGTTGGTGGCTTAGCTGCATATATGGTCAGAGAAGACATTGGA	TGCAGTAACATCAAACCAAGTATCTTGGTTGGTGGCTTAGCTGCATATATGGTCAGAGAAGACATTGGA	26:G>A	26
100081166	TGCAGTTATTATGGAAACCAGGGAGGTGGAGCACTTCCTAAAGATACTTCTGTAGCTGTCTGCACAATA	TGCAGTTATTATGGAAACCAGGGAGGTGGAGCACTTCCTAAAGATACTTCTGTAGCTGTCTGCACAATA	59:C>T	59
100081174	TGCAGTTTTGAACCAAGTGTCAGCATCAGTGAATTGCAGACAACCTACATTTTAGATTATAGACCATAA	TGCAGTTTTGAACCAAGTGTCAGCATCAGTGAATTGCAGACAACCTACATTTTAGATTATAGACCATAA	66:T>G	66
100081180	TGCAGAAAAGAAAAGCAGTTGCTGGTCATTCGAAGGATAACCCTGAACATGCAGTAAAACTGTTTCTTC	TGCAGAAAAGAAAAGCAGTTGCTGGTCATTCGAAGGATAACCCTGAACATGCAGTAAAACTGTTTCTTC	31:G>C	31
100081190	TGCAGAATAAGTTGCTGCTAACAGCAATGCAACAAATGCTGGAGGGAGTAAGTGGGGCTCAAAGGAAGA	TGCAGAATAAGTTGCTGCTAACAGCAATGCAACAAATGCTGGAGGGAGTAAGTGGGGCTCAAAGGAAGA	16:G>A	16
100081191	TGCAGAATGAAACAAATACAAATGATGTTATTGCAGTTAGAAAAAGATAAATTTTTTTCCGAGAAATAG	TGCAGAATGAAACAAATACAAATGATGTTATTGCAGTTAGAAAAAGATAAATTTTTTTCCGAGAAATAG	35:G>A	35
100081193	TGCAGAATGAGTAAAAGGGAAGAAAAAGTGGGGGGAATTTGGGAAAGGAGGGAAATTGAGGAGAATAAA	TGCAGAATGAGTAAAAGGGAAGAAAAAGTGGGGGGAATTTGGGAAAGGAGGGAAATTGAGGAGAATAAA	66:A>T	66
100081198	TGCAGAGAGGAGTGTAATTTCTGGAGACAATCAAGGTGGGAAGTTCTGTTTCAATTCTGGCAGCATCTA	TGCAGAGAGGAGTGTAATTTCTGGAGACAATCAAGGTGGGAAGTTCTGTTTCAATTCTGGCAGCATCTA	45:C>T	45
100081202	TGCAGAGGCTCAGATGATCATGGGAAATTGGCAACAATATACAAGTCAGAGAAGAGAGATCAGATGGTA	TGCAGAGGCTCAGATGATCATGGGAAATTGGCAACAATATACAAGTCAGAGAAGAGAGATCAGATGGTA	54:A>G	54
100081203	TGCAGAGGTGACGGTTGGTAGTAAGAACCGCATGTATGAGCGGCCTGACGCATCTGGAGATCTGCTTGT	TGCAGAGGTGACGGTTGGTAGTAAGAACCGCATGTATGAGCGGCCTGACGCATCTGGAGATCTGCTTGT	43:C>T	43
100081208	TGCAGATACACTAACTGTAAAAATGAGTGTCAAGTAAAGGTAAGTGCAAATTGCCTATGCCCCTTTACA	TGCAGATACACTAACTGTAAAAATGAGTGTCAAGTAAAGGTAAGTGCAAATTGCCTATGCCCCTTTACA	50:T>G	50
100081217	TGCAGCAACAGCAGCAACTCTCCCAGCTGCAGCAACAACTGCCACAGCAGCAGCAAATGGTGGGGTCAG	TGCAGCAACAGCAGCAACTCTCCCAGCTGCAGCAACAACTGCCACAGCAGCAGCAAATGGTGGGGTCAG	40:G>A	40
100081223	TGCAGCAGGATCTAAAATTGTCTTTGATTCAGTTTCCGCATGAAGGAGTAAGCAGAGAAGATCAACCTA	TGCAGCAGGATCTAAAATTGTCTTTGATTCAGTTTCCGCATGAAGGAGTAAGCAGAGAAGATCAACCTA	68:A>T	68
100081235	TGCAGCCTCTTTCTGCCAAACTCCGTCTTCCAGCACTTCTCCTGGTTCTCTCCTTCACCATAGCCATGG	TGCAGCCTCTTTCTGCCAAACTCCGTCTTCCAGCACTTCTCCTGGTTCTCTCCTTCACCATAGCCATGG	20:C>T	20
100081239	TGCAGCGTCAGGAATAGATGACTAATAAAAACGTCTTGAAGACACTATTTGACTCCGAAAAACCTGACG	TGCAGCGTCAGGAATAGATGACTAATAAAAACGTCTTGAAGACACTATTTGACTCCGAAAAACCTGACG	31:C>T	31
100081246	TGCAGCTGGTGATCTTGAAGCATGCACTCAATGTGAAAGGAAATGTACGTGGAATTTCTGCTTAGTCTC	TGCAGCTGGTGATCTTGAAGCATGCACTCAATGTGAAAGGAAATGTACGTGGAATTTCTGCTTAGTCTC	54:T>A	54
100081253	TGCAGCTTTAGCAGCCTGCACCCCAGCTAGTGCATCTTCAATAACAATACACTGAGAGAAATAGACAAA	TGCAGCTTTAGCAGCCTGCACCCCAGCTAGTGCATCTTCAATAACAATACACTGAGAGAAATAGACAAA	57:G>A	57
100081263	TGCAGGCTGATAACTATGAAAAGGTCTTCTCCCACATGAAACCTAACAGCATTCTTGGGCTTTCTCATG	TGCAGGCTGATAACTATGAAAAGGTCTTCTCCCACATGAAACCTAACAGCATTCTTGGGCTTTCTCATG	66:A>G	66

Licensed Under Creative Commons Attribution (CC BY-NC)

Table 4

Extrapolation showing part of the DArT SNPs

AlleleID	AlleleSeq	TrimmedS	Chrom_G	ChromPos	AlnCnt_G	AlnEvalue	SNP	SnpPositi	CallRate	OneRatioRef	OneRatioSnp	FreqHomRef	FreqHomSnp	FreqHets	PICRef	PICSnp	AvgPIC	AvgCount	AvgCount
10008112	TGCAGAA	TGCAGAA	JAJFZO01	1081492	1	2.57E-29	33:T>C	33	0.932927	1	0.006536	0.993464	0	0.006536	0	0.012986	0.006493	10.81053	4
10008112	TGCAGAT	TGCAGAT	JAJFZO01	1134086	1	2.57E-29	19:G>A	19	0.987805	0.919753	0.358025	0.641975	0.080247	0.277778	0.147615	0.459686	0.30365	32.05376	23.88732
10008113	TGCAGCT	TGCAGCT	JAJFZO01	8071886	1	2.57E-29	20:T>C	20	0.908537	0.993289	0.006711	0.993289	0.006711	0	0.013333	0.013333	0.013333	5.97753	3
10008114	TGCAGCT	TGCAGCT	JAJFZO01	1380373	1	2.57E-29	26:T>C	26	0.957317	0.993631	0.025478	0.974522	0.006369	0.019108	0.012658	0.049657	0.031157	8.20635	5
10008114	TGCAGCT	TGCAGCT	JAJFZO01	4803571	1	2.57E-29	23:C>T	23	0.987805	1	0.018519	0.981481	0	0.018519	0	0.036351	0.018176	14.75622	7
10008115	TGCAGTA	TGCAGTA	JAJFZO01	1287261	1	2.57E-29	26:G>A	26	0.97561	0.9875	0.1125	0.8875	0.0125	0.1	0.024687	0.199688	0.112187	23.70918	13
10008116	TGCAGTT	TGCAGTT	JAJFZO01	12952433	1	2.57E-29	59:C>T	59	0.987805	1	0.006173	0.993827	0	0.006173	0	0.012269	0.006135	14.60891	14
100081174	TGCAGTT	TGCAGTT	JAJFZO01	9636080	1	2.57E-29	66:T>G	66	0.926829	1	0.006579	0.993421	0	0.006579	0	0.013071	0.006536	5.97354	3
10008118	TGCAGAA	TGCAGAA	JAJFZO01	1587065	1	2.57E-29	31:G>C	31	0.963415	1	0.012658	0.987342	0	0.012658	0	0.024996	0.012498	19.62694	10.5
10008119	TGCAGAA	TGCAGAA	JAJFZO01	5471128	1	2.57E-29	16:G>A	16	0.95122	1	0.00641	0.99359	0	0.00641	0	0.012738	0.006369	10.77604	3
100081191	TGCAGAA	TGCAGAA	JAJFZO01	3841676	1	4.29E-27	35:G>A	35	0.823171	1	0.022222	0.977778	0	0.022222	0	0.043457	0.021728	23.40964	8.33333
10008119	TGCAGAA	TGCAGAATGAGTAA		0	0	999	66:A>T	66	0.920732	0.993377	0.006623	0.993377	0.006623	0	0.013157	0.013157	0.013157	29.06011	6
10008119	TGCAGAG	TGCAGAG	JAJFZO01	3413525	1	2.57E-29	45:C>T	45	0.945122	0.96129	0.109677	0.890323	0.03871	0.070968	0.074422	0.195297	0.13486	11.21667	4.63636
10008120	TGCAGAG	TGCAGAG	JAJFZO01	2564438	1	2.57E-29	54:A>G	54	0.737805	1	0.008264	0.991736	0	0.008264	0	0.016392	0.008196	10.44595	6
10008120	TGCAGAG	TGCAGAG	JAJFZO01	6652954	1	2.57E-29	43:C>T	43	0.890244	0.931507	0.157534	0.842466	0.068493	0.089041	0.127604	0.265434	0.196519	19.8	18.07143
10008120	TGCAGAT	TGCAGAT	JAJFZO01	4479332	1	1.19E-27	50:T>G	50	0.957317	1	0.012739	0.987261	0	0.012739	0	0.025153	0.012577	23.35385	8
10008121	TGCAGCA	TGCAGCAACAGCAG		0	0	999	40:G>A	40	0.792683	0.923077	0.223077	0.776923	0.076923	0.146154	0.142012	0.346627	0.24432	10.23129	6.85714
10008122	TGCAGCA	TGCAGCA	JAJFZO01	4203361	1	2.57E-29	68:A>T	68	0.646341	0.990566	0.009434	0.990566	0.009434	0	0.01869	0.01869	0.01869	10.96774	10
10008123	TGCAGCC	TGCAGCC	JAJFZO01	9486987	1	2.57E-29	20:C>T	20	0.689024	0.955752	0.061947	0.938053	0.044248	0.017699	0.08458	0.116219	0.100399	9.02308	7.42857
10008123	TGCAGCG	TGCAGCG	JAJFZO01	1403201	1	1.19E-27	31:C>T	31	0.95122	0.961538	0.121795	0.878205	0.038462	0.083333	0.073964	0.213922	0.143943	14.22162	9.55
10008124	TGCAGCT	TGCAGCT	JAJFZO01	142616	1	2.57E-29	54:T>A	54	0.707317	0.991379	0.025862	0.974138	0.008621	0.017241	0.017093	0.050386	0.03374	11.85507	6.33333
10008125	TGCAGCT	TGCAGCT	JAJFZO01	496515	1	2.57E-29	57:G>A	57	0.835366	0.963504	0.043796	0.956204	0.036496	0.007299	0.070329	0.083755	0.077042	4.46951	3.14286
10008126	TGCAGGC	TGCAGGC	JAJFZO01	4947313	1	2.57E-29	66:A>G	66	0.884146	1	0.006897	0.993103	0	0.006897	0	0.013698	0.006849	7.22807	4
100081264	TGCAGGC	TGCAGGC	JAJFZO01	3849388	1	2.57E-29	65:C>T	65	0.987805	0.993827	0.006173	0.993827	0.006173	0	0.012269	0.012269	0.012269	10.90955	8
10008127	TGCAGGG	TGCAGGG	JAJFZO01	8530295	1	1.19E-27	55:T>G	55	0.878049	0.013889	0.986111	0.013889	0.986111	0	0.027392	0.027392	0.027392	5	3.41916
100081274	TGCAGGT	TGCAGGTTATGCTAC		0	0	999	65:T>A	65	0.640244	0.914286	0.161905	0.838095	0.085714	0.07619	0.156735	0.271383	0.214059	24.48276	14.19048
10008128	TGCAGTA	TGCAGTACATACAGA		0	0	999	54:T>A	54	0.603659	0.838384	0.262626	0.737374	0.161616	0.10101	0.270993	0.387307	0.32915	7.20952	7.07143

Licensed Under Creative Commons Attribution (CC BY-NC)

Table 5
SilicoDArT data

CloneID	AlleleSeq	TrimmedSeq	Chrom_G	ChromPos_	AlnCnt_G	AlnEvalue_	CallRate	OneRatio	PIC	AvgReadDepth	StDevReadDepth	Qpmr	Reproducibility
1E+08	TGCAGCA	TGCAGCAAA	JAJFZO01	7182074	1	2.00E-25	0.987805	0.5	0.5	18.7767	12.18142	1.34569	1
1E+08	TGCAGGC	TGCAGGCATC	JAJFZO01	7234507	1	2.32E-18	0.95122	0.5	0.5	91	50.20822	1.88081	1
1E+08	TGCAGGT	TGCAGGTGA	JAJFZO01	2234104	1	2.28E-10	0.926829	0.5	0.5	43.76106	24.93328	1.82621	0.96875
1E+08	TGCAGGT	TGCAGGTTTGGTTTTTGG		0	0	999	0.97561	0.5	0.5	16.35577	11.29663	1.33205	1
1E+08	TGCAGAA	TGCAGAAAACCCTTTTCA		0	0	999	0.926829	0.5	0.5	52.76577	31.15204	1.68763	1
1E+08	TGCAGTT	TGCAGTTCCA	JAJFZO01	1745915	1	2.57E-29	0.97561	0.5	0.5	12.48485	8.21844	1.36118	1
1E+08	TGCAGCC	TGCAGCCAAT	JAJFZO01	3364433	1	2.57E-29	0.97561	0.5	0.5	55.0297	37.0123	1.41039	0.972222
1E+08	TGCAGTT	TGCAGTTCACATTATGTA		0	0	999	0.987805	0.5	0.5	17.20192	11.0178	1.31335	1
1E+08	TGCAGTT	TGCAGTTTGA	JAJFZO01	3545927	1	1.19E-27	0.97561	0.5	0.5	31.69524	19.18885	1.63721	0.967742
1E+08	TGCAGTG	TGCAGTGAA	JAJFZO01	728694	1	1.21E-17	0.981707	0.496894	0.499981	14.27778	9.7978	1.3863	0.971429
1E+08	TGCAGAA	TGCAGAAGC	JAJFZO01	1189910	1	3.32E-28	0.981707	0.496894	0.499981	23.06	14.7308	1.39655	1
1E+08	TGCAGTG	TGCAGTGACA	JAJFZO01	1927057	1	3.27E-22	0.981707	0.503106	0.499981	76.65094	45.20249	1.51127	1
1E+08	TGCAGAT	TGCAGATGATGACGAAT		0	0	999	0.981707	0.503106	0.499981	47.79381	32.58389	1.30261	1
1E+08	TGCAGTT	TGCAGTTGTA	JAJFZO01	3339549	1	2.57E-29	0.981707	0.496894	0.499981	47.09524	32.57885	1.42585	0.972222
1E+08	TGCAGCC	TGCAGCCATCTTCCACCT		0	0	999	0.969512	0.496855	0.49998	25.63462	15.48762	1.62878	1
1E+08	TGCAGTT	TGCAGTTGAG	JAJFZO01	3953646	1	2.57E-29	0.969512	0.496855	0.49998	18.52679	12.48997	1.43282	1
1E+08	TGCAGAA	TGCAGAAAGT	JAJFZO01	1059215	1	5.55E-26	0.957317	0.503185	0.49998	28.26852	16.4354	1.80234	1
1E+08	TGCAGGC	TGCAGGCAA	JAJFZO01	5178365	1	1.19E-27	0.957317	0.496815	0.49998	33.43396	20.50173	1.63188	1
1E+08	TGCAGTG	TGCAGTGCAC	JAJFZO01	5647666	1	5.55E-26	0.969512	0.503145	0.49998	46.29126	27.30896	1.56535	1
1E+08	TGCAGTA	TGCAGTACTG	JAJFZO01	336292	1	1.01E-16	0.969512	0.496855	0.49998	121.25	64.25313	2.02043	0.967742
1E+08	TGCAGTA	TGCAGTAAG	JAJFZO01	414810	1	5.55E-26	0.957317	0.496815	0.49998	42.83929	25.23056	1.46251	1
1E+08	TGCAGTT	TGCAGTTCTA	JAJFZO01	904378	1	1.20E-22	0.932927	0.496732	0.499979	7.92523	5.24128	1.48091	0.971429
1E+08	TGCAGCC	TGCAGCCCAA	JAJFZO01	5589654	1	9.29E-24	0.945122	0.496774	0.499979	56.2973	32.72368	1.69628	1
1E+08	TGCAGAT	TGCAGATTGA	JAJFZO01	1184248	1	3.98E-21	0.932927	0.496732	0.499979	62.34906	35.35428	1.78743	0.96875
1E+08	TGCAGAA	TGCAGAAATC	JAJFZO01	8307354	1	2.58E-24	0.932927	0.503268	0.499979	82.93	41.43604	2.06084	1
1E+08	TGCAGCG	TGCAGCGGA	JAJFZO01	986015	1	5.55E-26	0.945122	0.496774	0.499979	43.26923	25.2637	1.85436	1
1E+08	TGCAGTG	TGCAGTGGTG	JAJFZO01	3256843	1	2.57E-29	0.945122	0.503226	0.499979	29.83654	17.29891	1.63022	0.970588
1E+08	TGCAGTG	TGCAGTGCAA	JAJFZO01	6440321	1	2.57E-29	0.920732	0.496689	0.499978	11.32743	7.61048	1.66402	0.967742

Licensed Under Creative Commons Attribution (CC BY-NC)

4.2 Population Genetics

Bar plot on the Number of Individuals per Population

Diani-Ukunda, India, Kibwezi Town, Kilifi, Likoni, Maungu, Mbololo, Miasenyi, Mwakiki, Pwani University and Shika Adabu provenances had 10 genotypes each whereas Arabuko Sokoke, Gede, Mackinon, Samburu, Vipingo and Waa Kwale had provenances had 9 genotypes as shown in fig. 2.

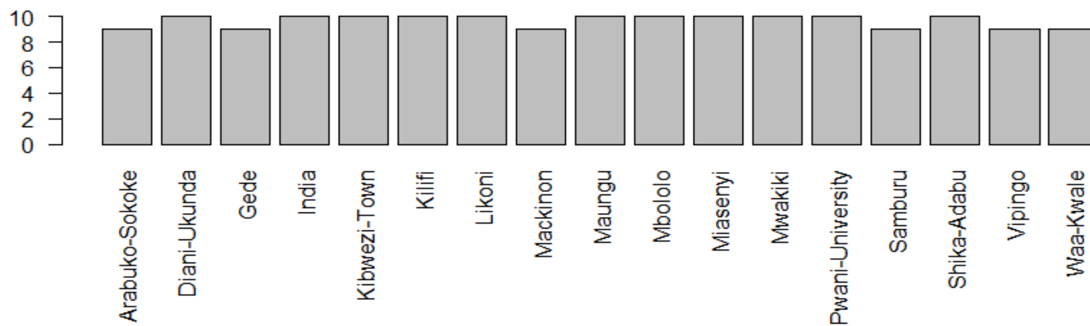

Figure 2
Bar plot showing populations/ provenances

F Statistics

For Fst, the functions StAMPP package were used as they enable parallel computing hence faster. Fst nboots = 1, percentage was 95 and nclusters = 1. The genetic distances were in centiMorgans (cM)

Licensed Under Creative Commons Attribution (CC BY-NC)

Table 6
Fst table showing the distances in centiMorgans

	Arabuko-Sokoke	Diani-Ukunda	Gede	India	Kibwezi-Town	Kilifi	Likoni	Mackinon	Maungu	Mbololo	Miasenyi	Mwakiki	Pwani-University	Shika-Adabu	Samburu	Vipingo	Waa-Kwale
Arabuko-Sokoke	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Diani-Ukunda	0.08839573	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Gede	0.08350471	0.05436086	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
India	0.09123628	0.10682290	0.0831651876	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Kibwezi-Town	0.09667251	0.10330296	0.1358056494	0.13164604	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Kilifi	0.05980152	0.04185168	0.0809777092	0.09218634	0.06229153	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Likoni	0.23878713	0.16166061	0.1723103958	0.19987526	0.23928502	0.16178895	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Mackinon	0.27382156	0.24313803	0.2851214155	0.28918113	0.25815672	0.21890244	0.40816105	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Maungu	0.08556645	0.06754122	0.1125005719	0.12296296	0.08263517	0.01545695	0.18599915	0.2446184	NA	NA	NA	NA	NA	NA	NA	NA	NA
Mbololo	0.12637178	0.08676262	0.0908647419	0.11953297	0.15208738	0.08184686	0.02263307	0.3232576	0.10975701	NA	NA	NA	NA	NA	NA	NA	NA
Miasenyi	0.22036031	0.17774202	0.1937571500	0.20481694	0.23564218	0.17685166	0.19439592	0.4105880	0.20447853	0.17187817	NA	NA	NA	NA	NA	NA	NA
Mwakiki	0.22037932	0.13525650	0.1419753162	0.18225293	0.21543672	0.15565820	0.09504356	0.4012120	0.18731091	0.07154605	0.1881584	NA	NA	NA	NA	NA	NA
Pwani-University	0.19971295	0.22030874	0.2620468680	0.24810477	0.21604694	0.14690566	0.37524804	0.3183340	0.19077273	0.28724940	0.3614970	0.3717897	NA	NA	NA	NA	NA
Shika-Adabu	0.11499504	0.11537405	0.1323495056	0.13495842	0.06140674	0.09905247	0.22643222	0.2960261	0.12646844	0.14104922	0.2338500	0.1948355	0.2572970	NA	NA	NA	NA
Samburu	0.06033521	0.04156700	0.0003597574	0.06309578	0.10437766	0.02906750	0.13491462	0.2425008	0.07033899	0.06255422	0.1729000	0.1281419	0.2160161	0.1113208	NA	NA	NA
Vipingo	0.08982599	0.04312460	0.0648701125	0.10214033	0.09899893	0.02812665	0.09184752	0.2474152	0.05503756	0.03094450	0.1659348	0.1012815	0.2093105	0.1174521	0.03704638	NA	NA
Waa-Kwale	0.11575017	0.09217460	0.0829989068	0.12887380	0.14181266	0.08259242	0.14365688	0.2877884	0.10662040	0.07174563	0.2027366	0.1505751	0.2641509	0.1388378	0.06071000	0.05671634	NA

Licensed Under Creative Commons Attribution (CC BY-NC)

Table 7

Population genetic distances using Euclidean as a distance measure

```
Distance measure: euclidean
  No. of populations = 17
  Average no. of individuals per population = 9.647059
  No. of loci = 20921
  Miniumum Distance:  19.32
  Maximum Distance:  48.45
  Average Distance:  33.024

Completed: gl.dist.pop
```

	Arabuko-Sokoke	Diani-Ukunda	Gede	India	Kibwezi-Town	Kilifi	Likoni	Mackinon	Maungu	Mbololo
Diani-Ukunda	31.22686									
Gede	33.60015	28.17896								
India	31.70141	30.38961	30.71674							
Kibwezi-Town	30.58207	28.96980	33.57191	31.67714						
Kilifi	27.95676	23.67949	29.62305	28.59167	24.74401					
Likoni	38.27676	31.25721	32.96569	34.65025	37.02305	31.45459				
Mackinon	44.88385	40.78608	45.22790	45.02551	40.99392	39.02065	47.81913			
Maungu	29.98550	26.02887	31.98760	31.13192	26.46278	20.04084	32.81997	40.22005		
Mbololo	33.50289	28.30668	30.45371	31.23737	32.84962	27.40933	19.32103	45.07161	29.52481	
Miasenyi	37.11695	32.64182	34.71259	35.31060	36.91502	32.88203	30.36716	48.45087	34.47238	31.28294
Mwakiki	37.76208	29.93681	31.63940	33.98191	35.71957	31.66914	23.32104	47.81128	33.60243	24.16349
Pwani-University	37.65138	37.59373	42.18504	40.48078	36.77890	31.72450	46.01462	43.07642	34.80793	41.99274
Samburu	30.90221	26.11825	25.45931	28.17443	30.49295	23.96683	30.11192	42.13376	27.73024	27.67423
Shika-Adabu	31.72473	29.76132	32.88128	31.63799	24.06153	27.99242	35.38716	43.29907	30.11396	31.49249
Vipingo	32.16829	25.28038	29.84218	30.68687	29.10405	22.95677	25.90581	41.36523	25.36065	24.02726
Waa-Kwale	35.33271	30.69813	32.56732	34.02535	33.65759	29.34749	30.16808	44.18125	30.99582	28.37661

	Miasenyi	Mwakiki	Pwani-University	Samburu	Shika-Adabu	Vipingo
Diani-Ukunda						
Gede						
India						
Kibwezi-Town						
Kilifi						
Likoni						
Mackinon						
Maungu						
Mbololo						
Miasenyi						
Mwakiki	30.39267					
Pwani-University	45.22906	46.18995				
Samburu	33.20795	30.52589	38.62910			
Shika-Adabu	36.20826	33.64458	39.53715	30.79956		
Vipingo	31.88543	27.56765	36.96960	26.59748	30.34971	
Waa-Kwale	34.77817	31.58159	41.52538	29.99404	32.83590	28.69758

Licensed Under Creative Commons Attribution (CC BY-NC)

4.3 Population Euclidean Distances

The genetic distance between the 17 populations of *M. oleifera* genotypes were calculated and computed using DArT R package, based on Euclidean as a measure of distance. It was evident that the Euclidean distances ranged from 19.32 and 48.45 with an average distance of 33.024. The least distance was that of Mbololo and Likoni while the highest distance was between Miasenyi and Mackinon.

4.4 Analysis of Molecular Variance (AMOVA)

Analysis of molecular variance (AMOVA) was utilized to carry out population genetic analysis using KDCompute Software (silicoDArT) and DArT R. The standard deviation and variance for individual alleles were calculated. The sum of the tag read counts for all samples, divided by the number of samples with non-zero tag read counts (AvgReadDepth) had a range of 255.26 to 5.05 with an average of 39.8. The standard deviation of the number of tag reads for all samples with non-zero tag read counts (StDevReadDepth) ranged from 0 to 124.14 exhibiting an average of 18. 96. The average of the normalized non-zero tag read counts divided by the standard deviation of the normalized non-zero tag read counts which is the variance (Qpmr), exhibited a range of 1.25 and 2262.8 with an average Qpmr of 2.58. Reproducibility ranged from 0.95 to 1 and had an average reproducibility of 0.99. From these data, it is evident that the average of the sum of the tag read counts for all the SNPs called was 39.8, the standard deviation of the SNPs was 18.96 whereas the variance of the SNPs was 2.58. From DArT R, the table below summarizes the values of analysis of molecular variance.

4.5 Neighbor-Joining Tree

It constructs a phenetic tree using a neighbor-joining approach or UPGMA using DArT R and thus enables the visualization of the populations' genetic similarity. Euclidean distances were utilized. A phylogenetic tree with 17 tips and 15 internal nodes was resultant. It is evident from the phenetic tree (figure 3) that there were 4 clusters of the *M. oleifera* provenances. The 1st cluster had 4 populations. Pwani University and Mackinon populations were sub grouped together, indicating their close relationship. Maungu and Kilifi were also sub clustered together. In the 2nd cluster were Shika Adabu and Kibwezi Town whereas in cluster 3 only Arabuko Sokoke was evident. Lastly, cluster 4 had 10 populations clustered together. Diani-Ukunda, India, Samburu, Gede, Waa Kwale, Vipingo, Miasenyi, Mwakiki, Mbololo and Likoni provenances were all clustered here. This shows their close genetic similarity.

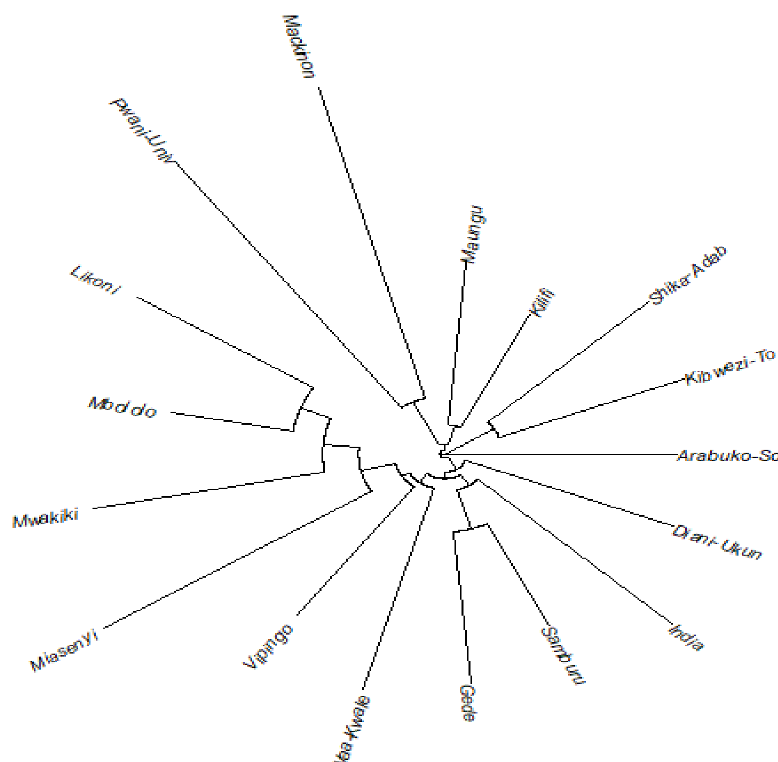

Figure 3
Phenetic tree for the population using Neighbor-joining approach from DArT

Licensed Under Creative Commons Attribution (CC BY-NC)

V. DISCUSSION

5.1 To Identify Polymorphisms in the Selected Varieties

As evident in Table 2, the allele with allele ID 100081123|F|0-33:T>C-33:T>C whose TGCAGAACAGCTACTCTTCCCTTGATATCCATCTTGGGCAGTCATATTCTGTCTTGGTCACAGCTAACC had a mutation or base pair change at position 33. The nucleotide change was from T to C hence the SNP was at position 33.

Another allele with Allele ID 100081125|F|0-19:G>A-19:G>A with the sequence TGCAGATAATTCGTCGTGCGAATGAGAATGGGGAAGATCCATTACATCTAAGCAACAGATTTTGTGAAG had a single nucleotide polymorphism at position 19 from G to A.

The call rate for the 20,921 SNPs ranged from 0.396 to 1 with an average call rate of 0.82. The proportion of samples for which the genotype score is "1" was averagely 0.30 with a minimum score of 0.006 and a maximum score of 1. The SNP exhibited a reproducibility of 0.98. There was an evidently high reproducibility and consistent marker score.

PIC for the SNPs was averagely 0.24 with a maximum PIC being 0.5. PIC is divided into three categories with regard to the polymorphism value, high (PIC value higher than 0.5), medium (value between 0.25 and 0.5) and low (lower than 0.25) (Vaiman et al., 1994; Xie et al., 2010)**.** In a study conducted by (Rajalakshmi et al., 2019), (Genetic Diversity, Population Structure and Correlation Study in *M. oleifera* Lam. Using ISSR and SRAP) Markers, fifteen inter-simple sequence repeat (ISSR) primers produced the average PIC value of 0.53 while the genetic variability study using sequence-related amplified polymorphism (SRAP) primer combinations exhibited an average PIC of 0.44. (Muslihatin et al., 2022) used ISSR markers to study the genetic diversity of *M. oleifera* where the average PIC value in the results of their study was 0.46. (kumar Ganesan et al., 2014) carried out Genetic diversity and population structure study of drumstick by utilizing morphological and SSR markers whereby the PIC value for SSR primers exhibited an average of 0.15 whereas the gene diversity ranged exhibited an average of 0.18. this was evidently lower than the current study. For the characterization of *M. oleifera*, (Saini et al., 2013) utilized RAPD, ISSR, and cytochrome P450 makers and reported high PIC values of 0.72, 0.81, and 0.68 for RAPD, ISSR, and cytochrome P450, respectively. It is evident that the average PIC for the SNP markers was lower. (kumar Ganesan et al., 2014) postulated that the reason for lower PIC could be due to the fact that SSR markers generated on average 1.84 bands per primer followed by minimum number of unique bands generated in the genotypes studied in comparison with (Muluvi et al., 1999) who reported 59 bands per primer with AFLP marker. In our study, this could also be attributed to SNP markers being changes on single nucleotide bases. (Sousa et al., 2017) found a mean PIC value of 0.35 with 11,187 SNP markers while studying *C. Arabica* and concluded that the low PIC value evidences the narrow genetic base of *C. Arabica*. Since PIC measures a marker's capacity to detect polymorphism, it can be assumed that the SNPs' low PIC average was brought on by the marker's specificity and selectivity. The heterozygosity of the loci ranged from 0 to 0.698 and exhibited an average heterozygosity of 0.11. This heterozygosity average was similar to that of (kumar Ganesan et al., 2014) average heterozygosity across 20 loci was 0.10.

The sum of the tag read counts for all samples, divided by the number of samples with non-zero tag read counts (AvgReadDepth) had a range of 255.26 to 5.05 with an average of 39.8. The standard deviation of the number of tag reads for all samples with non-zero tag read counts (StDevReadDepth) ranged from 0 to 124.14 exhibiting an average of 18. 96. The average of the normalized non-zero tag read counts divided by the standard deviation of the normalized non-zero tag read counts which is the variance (Qpmr), exhibited a range of 1.25 and 2262.8 with an average Qpmr of 2.58. Reproducibility ranged from 0.95 to 1 and had an average reproducibility of 0.99.

M. oleifera being a cross-pollinated tree, high heterogeneity in many character forms was observed resulting in vast genetic diversity. *M. oleifera* seems to be a very polymorphic plant since there were several single nucleotide polymorphisms observed. With an average call rate of 95%, 116,190 SilicoDArT markers were called on the 105 Napier grass accessions. The greatest anticipated heterozygosity (He) and polymorphism information content (PIC) values for biallelic markers were both 0.5. The expected heterozygosity (He) varied from 0 to 0.5, and the PIC ranged from 0 to 0.38. He and PIC values were generally 0.24 and 0.19, respectively. With an average call rate of 87%, a total of 85,452 SNP markers were called on the accessions. According to (Muktar et al., 2019), the PIC values ranged from 0 to 0.38 with an average of 0.11 and the He values ranged from 0 to 0.5 with an average of 0.13. This were higher than observed in the current study. The average PIC of DArT SNP for safflower was 0.29 (Hassani et.al,2020). This value was almost similar to the current study, whoseaverage PIC was 0.24.

Licensed Under Creative Commons Attribution (CC BY-NC)

5.2 Population Genetics Analysis/ Study

In-depth knowledge and understanding of the gene flow pattern and population genetic structure in *M. oleifera* through molecular genetic diversity and population genetic structure of worldwide collections is of immense promise. Documentation of genetic diversity and conservation of germplasm is a necessity to develop elite varieties.

5.2.1 Population Euclidean Distances

The genetic distance between the 17 populations of *M. oleifera* genotypes were calculated and computed using DArT R package, based on Euclidean as a measure of distance (table 7). It was evident that the Euclidean distances ranged from 19.32 and 48.45 with an average distance of 33.024. The least distance was that of Mbololo and Likoni while the highest distance was between Miasenyi and Mackinon. The genetic distances exhibited between the populations was relatively low and could thus be concluded that there was least variability in the populations. This could be attributed to the spread of seeds as most of the provenances are geographically close and the source of planting material arising from a common ancestry. According to (Mgendi et al., 2010; Muluvi et al., 1999),the core collection of *M. oleifera* populations extant in Africa consists of a relatively small number of accessions imported from India, which has led to low levels of genetic diversity.

5.2 F Statistics

F statistics is used to test variances of populations (Gruber et al., 2018). Samburu and Gede exhibited the least variance 0.0003 whereas Pwani University and Mwakiki showed the highest variance with 0.37. A moderate Fst value would indicate minimal significant genetic divergence between the populations of *M. oleifera*, according to (Rufai et al., 2013). This study exhibited low Fst values between the populations signaling very low genetic divergence (table 6).

5.3 Neighbor Joining Tree

The 1st cluster had 4 populations (figure 2). Pwani University and Mackinon populations were sub grouped together, indicating their close relationship. Maungu and Kilifi were also sub clustered together. In the 2nd cluster was Shika Adabu and Kibwezi Town whereas in cluster 3 only Arabuko Sokoke was evident. Lastly, cluster 4 had 10 populations clustered together. Diani-Ukunda, India, Samburu, Gede, Waa Kwale, Vipingo, Miasenyi, Mwakiki, Mbololo and Likoni provenances were all clustered here. This shows their close genetic similarity (Fig. 2).

Mackinon and Pwani University cluster together yet they are geographically distant. This observation is also true for Maungu and Kilifi, Samburu and Gede ad also Mbololo and Likoni. Hence, the conclusion that no geographic isolation was evident and could be resultant from the accessions having a common ancestry. This could also be explained from the perspective that no evolution has taken place to guarantee change in genetic makeup of *M. oleifera* since it was introduced less than a century ago by the Indians.

In the groups analyzed, no obvious geographic isolation was observed. (Rufai et al., 2013) and (Mgendi et al., 2010) reported a similar finding based on RAPD markers. This further demonstrates that population genetic divergence cannot be explained solely by geographic distance (Verma et al., 2015). Similar findings in other plant species have been reported by numerous studies (Singhet al., 2014; (Bronzini de Caraffa et al., 2002), but some of these studies were able to distinguish genotypes depending on geographic origin (Desplanque et al., 1999); (Muluvi et al., 1999) (Li & Nelson, 2002).

5.4 Analysis of Molecular Variance (AMOVA)

AMOVA (analysis of molecular variance) was carried out to detect population differentiation utilizing molecular markers. The standard deviation and variance for individual alleles were calculated. The sum of the tag read counts for all samples, divided by the number of samples with non-zero tag read counts (AvgReadDepth) had a range of 255.26 to 5.05 with an average of 39.8. The standard deviation of the number of tag reads for all samples with non-zero tag read counts (StDevReadDepth) ranged from 0 to 124.14 exhibiting an average of 18. 96. The average of the normalized non-zero tag read counts divided by the standard deviation of the normalized non-zero tag read counts which is the variance (Qpmr), exhibited a range of 1.25 and 2262.8 with an average Qpmr of 2.58. Reproducibility ranged from 0.95 to 1 and had an average reproducibility of 0.99. From these data, it is evident that the average of the sum of the tag read counts for all the SNPs called was 39.8, the standard deviation of the SNPs was 18.96 whereas the variance of the SNPs was 2.58.

The molecular variance analysis (AMOVA) described a 2.55%, variation within the population and 2.73% among the populations. In this study the higher variation was among population. When AMOVA was used to analyze the SSR data, it was discovered that there was 2% variation between regions, 3% variation within populations, and 95% of the variation within populations (Singh et al., 2014). The majority of variation within populations is anticipated to be

Licensed Under Creative Commons Attribution (CC BY-NC)

maintained by woody, perennial, out-crossed species (Hamrick et al., 1989). In contrast to our finding that the greatest variation was seen between populations, (Muluvi et al., 1999) studied four pairs of AFLP primer combinations and found significant variation between regions and populations. However, (kumar Ganesan et al., 2014; Rufai et al., 2013), reported 95% of the total genetic variation within the population. The high similarity between the genotypes could be attributed to the *M. oleifera* plants in the various provenances having the same ancestry. It could also be resultant from the samples having been introduced from different sources.

VI. CONCLUSION

M. oleifera seems to be very polymorphic as several SNPs were observed. This research sought to evaluate the population genetics of 17 provenances of *Moringa oleifera* from Coastal Kenya through molecular markers analysis, utilizing SNPs and genotyping.

It was also evident that the genotypes in the populations exhibited less variance. Population study showed that, no geographical isolation exists between genotypes from the Coastal region of Kenya. The less variability among the populations indicates that the populations are not significantly different genetically.

This could imply that spread of planting materials has taken place in the form of cuttings, seed and/or high rates of gene flow between the adjacent populations. This could also be explained from the perspective that no evolution has taken place to guarantee change in genetic makeup of *M. oleifera* since it was introduced less than a century ago by the Indians.

This study contributes significantly to the body of knowledge on *M. oleifera* genetics, paving the way for future research endeavors and practical applications aimed at harnessing the full potential of this remarkable plant. The management, conservation, association mapping, and marker-assisted selection of the drumstick for future improvement will all benefit significantly from this study.

Recommendations

Continued research into functional genomics of *M. oleifera* will deepen our understanding of its biological mechanisms and facilitate development of innovative biotechnological applications.

More studies could be carried out using different molecular markers to ascertain the genetic divergence of Moringa in Kenya

Compliance with ethical standards

Disclosure of conflict of interest

The authors declare that there is no conflict of interest in relation to this manuscript.

Funding

This research was funded by National Research Fund (NRF), Kenya. The authors acknowledge SEQART Africa, based at International Livestock and Research Institute (ILRI), Kenya for providing the laboratory and equipment for conducting this study.

REFERENCES

Bronzini de Caraffa, V., Giannettini, J., Gambotti, C., & Maury, J. (2002). Genetic relationships between cultivated and wild olives of Corsica and Sardinia using RAPD markers. *Euphytica*, *123*(2), 263–271.

Chung, Y. S., Choi, S. C., Jun, T.-H., & Kim, C. (2017). Genotyping-by-sequencing: A promising tool for plant genetics research and breeding. *Horticulture, Environment, and Biotechnology*, *58*(5), 425–431.

Desplanque, B., Boudry, P., Broomberg, K., Saumitou-Laprade, P., Cuguen, J., & Van Dijk, H. (1999). Genetic diversity and gene flow between wild, cultivated and weedy forms of Beta vulgaris L.(Chenopodiaceae), assessed by RFLP and microsatellite markers. *Theoretical and Applied Genetics*, *98*(8), 1194–1201.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One*, *6*(5), e19379.

Gruber, B., Unmack, P. J., Berry, O. F., & Georges, A. (2018). dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Molecular Ecology Resources*, *18*(3), 691–699.

Hamrick, J. L., Blanton, H. M., & Hamrick, K. J. (1989). Genetic structure of geographically marginal populations of ponderosa pine. *American Journal of Botany*, *76*(11), 1559–1568.

Hassani, S. M. R., Talebi, R., Pourdad, S. S., Naji, A. M., & Fayaz, F. (2020). In-depth genome diversity, population structure and linkage disequilibrium analysis of worldwide diverse safflower (Carthamus tinctorius L.)

Licensed Under Creative Commons Attribution (CC BY-NC)

accessions using NGS data generated by DArTseq technology. *Molecular Biology Reports*, *47*(3), 2123–2135. https://doi.org/10.1007/s11033-020-05312-x

He, J., Zhao, X., Laroche, A., Lu, Z.-X., Liu, H., & Li, Z. (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in Plant Science*, *5*, 484.

He, S.-T., Yang, J., Wei, J., Wu, J.-C., Zheng, Y.-X., Zhang, Y.-P., & Peng, X.-M. (2021). Genetic diversity of eleven Moringa oleifera Lam. Germplasm introduced to Yunnan, southwest China and their backgrounds during worldwide cultivation. *Genetic Resources and Crop Evolution*, 1–12.

kumar Ganesan, S., Singh, R., Choudhury, D. R., Bharadwaj, J., Gupta, V., & Singode, A. (2014). Genetic diversity and population structure study of drumstick (Moringa oleifera Lam.) using morphological and SSR markers. *Industrial Crops and Products*, *60*, 316–325.

Kumar, P., Dolkar, R., Manjunatha, G., & Pallavi, H. M. (2017). Molecular fingerprinting and assessment of genetic variations among advanced breeding lines of Moringa oleifera L. by using seed protein, RAPD and Cytochrome P450 based markers. *South African Journal of Botany*, *111*, 60–67.

Lakshmidevamma, T. N., Ugalat, J., Apoorva, K. A., Suresh, S. P., Doddamani, M., Kadam, S., Nayana, R. S., Jagadeesha, R. C., & Fakrudin, B. (2021). Genetic Diversity of Moringa (Moringa Oleifera Lam.). In *The Moringa Genome* (pp. 57–65). Springer.

Li, Z., & Nelson, R. L. (2002). RAPD marker diversity among cultivated and wild soybean accessions from four Chinese provinces. *Crop Science*, *42*(5), 1737–1744.

Mgendi, M. G., Manoko, M. L., & Nyomora, A. (2010). *Genetic diversity between cultivated and non-cultivated Moringa oleifera Lam. Provenances assessed by RAPD markers*.

Morgil, H., Gercek, Y. C., & Tulum, I. (2020). Single nucleotide polymorphisms (SNPs) in plant genetics and breeding. *The Recent Topics in Genetic Polymorphisms*, 825–400.

Muktar, M. S., Teshome, A., Hanson, J., Negawo, A. T., Habte, E., Entfellner, J.-B. D., Lee, K.-W., & Jones, C. S. (2019). Genotyping by sequencing provides new insights into the diversity of Napier grass (Cenchrus purpureus) and reveals variation in genome-wide LD patterns between collections. *Scientific Reports*, *9*(1), 1–15.

Muluvi, G. M., Sprent, J. I., Soranzo, N., Provan, J., Odee, D., Folkard, G., McNicol, J. W., & Powell, W. (1999). Amplified fragment length polymorphism (AFLP) analysis of genetic variation in Moringa oleifera Lam. *Molecular Ecology*, *8*(3), 463–470.

Muslihatin, W., Saputro, T. B., Latifah, N. I., & Himayani, C. E. S. (2022). The Genetic Diversity of Moringa Oleifera on Poteran Island-Madura Based on Petiole Colors Using ISSR (Inter Simple Sequence Repeat) Method. *7th International Conference on Biological Science (ICBS 2021)*, 39–46.

Nadeem, M. A., Habyarimana, E., Çiftçi, V., Nawaz, M. A., Karaköy, T., Comertpay, G., Shahid, M. Q., Hatipoğlu, R., Yeken, M. Z., & Ali, F. (2018). Characterization of genetic diversity in Turkish common bean gene pool using phenotypic and whole-genome DArTseq-generated silicoDArT marker information. *PloS One*, *13*(10), e0205363.

Ondieki, S. K., Hunja, C. W., Muluvi, G. M., Korir, J. C., Mutiso, F., Kitheka, J. U., Kioko, D., Kimatu, J., Ndufa, J. K., & Mutati, K. (2023). Genetic diversity of selected Moringa oleifera lam. Provenances from the coastal region of Kenya. *Magna Scientia Advanced Research and Reviews*, *8*(2), 118–128.

Rajalakshmi, R., Rajalakshmi, S., & Parida, A. (2019). Genetic diversity, population structure and correlation study in Moringa oleifera Lam. Using ISSR and SRAP markers. *Proceedings of the National Academy of Sciences, India Section B: Biological Sciences*, *89*(4), 1361–1371.

Raman, R., Cowley, R. B., Raman, H., & Luckett, D. J. (2014). Analyses using SSR and DArT molecular markers reveal that Ethiopian accessions of white lupin (Lupinus albus L.) represent a unique genepool. *Open Journal of Genetics*, *2014*. https://www.scirp.org/html/3-1370162_44675.htm

Rufai, S., Hanafi, M. M., Rafii, M. Y., Ahmad, S., Arolu, I. W., & Ferdous, J. (2013). Genetic dissection of new genotypes of drumstick tree (Moringa oleifera Lam.) using random amplified polymorphic DNA marker. *BioMed Research International*, *2013*.

Saini, R. K., Saad, K. R., Ravishankar, G. A., Giridhar, P., & Shetty, N. P. (2013). Genetic diversity of commercially grown Moringa oleifera Lam. Cultivars from India by RAPD, ISSR and cytochrome P 450-based markers. *Plant Systematics and Evolution*, *299*(7), 1205–1213.

Sharma, P., Kachhwah, S., Mahendrakar, M. D., Kothari, S. L., & Singh, R. B. (2024). Assessment of the genetic diversity and population structure in Moringa oleifera accessions using DNA markers and phenotypic descriptors. *Plant Gene*, 100462.

Singh, R., Roy Choudhury, D., Bharadwaj, J., Gupta, V., & Singode, A. (2014). *Genetic diversity and population structure study of drumstick (Moringa oleifera Lam.) using morphological and SSR markers*.

Licensed Under Creative Commons Attribution (CC BY-NC)

Sousa, T. V., Caixeta, E. T., Alkimim, E. R., de Oliveira, A. C. B., Pereira, A. A., Sakiyama, N. S., de Resende Júnior, M. F. R., & Zambolim, L. (2017). Population structure and genetic diversity of coffee progenies derived from Catuaí and Híbrido de Timor revealed by genome-wide SNP marker. *Tree Genetics & Genomes*, *13*(6), 1–16.

Vaiman, D., Mercier, D., Moazami-Goudarzi, K., Eggen, A., Ciampolini, R., Lépingle, A., Velmala, R., Kaukinen, J., Varvio, S.-L., & Martin, P. (1994). A set of 99 cattle microsatellites: Characterization, synteny mapping, and polymorphism. *Mammalian Genome*, *5*(5), 288–297.

Verma, S., Gupta, S., Bandhiwal, N., Kumar, T., Bharadwaj, C., & Bhatia, S. (2015). High-density linkage map construction and mapping of seed trait QTLs in chickpea (Cicer arietinum L.) using Genotyping-by-Sequencing (GBS). *Scientific Reports*, *5*(1), 1–14.

Wang, N., Yuan, Y., Wang, H., Yu, D., Liu, Y., Zhang, A., Gowda, M., Nair, S. K., Hao, Z., & Lu, Y. (2020). Applications of genotyping-by-sequencing (GBS) in maize genetics and breeding. *Scientific Reports*, *10*(1), 16308.

Wu, J.-C., Yang, J., Gu, Z.-J., & Zhang, Y.-P. (2010). Isolation and characterization of twenty polymorphic microsatellite loci for Moringa oleifera (Moringaceae). *HortScience*, *45*(4), 690–692.

Xie, W.-G., Zhang, X.-Q., Cai, H.-W., Liu, W., & Peng, Y. (2010). Genetic diversity analysis and transferability of cereal EST-SSR markers to orchardgrass (Dactylis glomerata L.). *Biochemical Systematics and Ecology*, *38*(4), 740–749.

Licensed Under Creative Commons Attribution (CC BY-NC)