Original Research Article

ASSESSMENT OF THE GENETIC DIVERSITY OF THE TRADITIONAL TREE SPECIES *Kigelia africana (sausage tree)* USING MOLECULAR MARKERS FOR CONSERVATION GENOMICS IN KENYA

ABSTRACT:

This study is the first to explore the genetic composition of ancient *Kigelia africana* across a distribution range in Kenya. *Kigelia africana* is a native forest species of Kenya as far as we know it; it is widely planted in the central regions of the country by the Kikuyu tribe who inhabit this region for making their traditional brew Muratina. Unlike other tree species like Acacia, *Kigelia africana* has seldom been studied, although there is ample evidence of its great ecological and economic value. Because of cultural reasons, natural populations are rare in the wild. Hence the study seeks to explore the genetic diversity and composition of ancient the tree distributed across various regions in Kenya.

In this study, four ancient tree populations were investigated to explore the genetic diversity and composition of *Kigelia africana* through DArTseq technology. Thirty-two (32) Plant seed samples were obtained from various locations, their DNA extracted, libraries prepared, and sequenced using the Illumina 2500 high throughput sequencer.

A total of 8,556 SilicoDArT and 3,703 SNP markers were selected and used. The average PIC was 0.45 and 0.41 for the SilicoDArT and SNPs respectively. The population structure and average linkage hierarchical clustering based on the SNPs revealed two distinct subpopulations and a few smaller admixture groups. Both marker types identified all 32 landraces as potential duplicates with very low genetic diversity between individuals. The heterozygosity defining the genetic variation within each subpopulation was around 0.25. A mantel test showed good harmony between DArTseq and SNP marker data sets. It also showed no significant correlation between genetic diversity and the geographical coordinates of the tree samples. The results of this study provide important information and insights for decision-makers, farmers, and breeders to make the necessary actions to conserve this culturally important tree.

Keywords: genetic composition, Kenya, Kigelia Africana, polymorphism information content, SNP markers, genetic diversity.

INTRODUCTION

1.1 Background Information

Small isolated populations of species face an increased risk of losing adaptive variation due to genetic drift and inbreeding, highlighting the importance of genetic diversity in plant species for resilience against threats. The loss of genetic diversity can lead to inbreeding depression, reduced adaptation, and decreased long-term species survival. Kigelia africana, commonly known as the sausage tree, is characterized by large maroon flowers, a squat trunk, and distinctive fruits. The plant's roots, wood, and leaves contain various compounds, including naphthoquinones and flavonoids. Kigelia africana has medicinal uses, treating dysentery and venereal diseases, and exhibits analgesic and anti-inflammatory effects. Additionally, in different cultures, the tree serves various purposes, such as flavoring beer in Malawi, making canoes in Botswana and Zimbabwe, and producing an alcoholic beverage in Central Kenya. However, caution is advised as the fresh fruit is poisonous.

In terms of genetic diversity, population genetic studies play a crucial role in the conservation and breeding of tree species. Single-nucleotide polymorphisms (SNPs) have proven to be abundant in the genome, offering detailed insights into population genetics. The DArTseq technology, based on the Genotyping-by-Sequencing (GBS) principle, generates both SNP and DArTSeq markers, providing high consistency and reproducibility in diversity studies. This technology efficiently targets low-copy-number sequences through a complexity reduction method and has been successfully applied in genetic diversity studies for various species.

Understanding the genetic diversity and structure of tree populations is essential for conservation efforts and breeding programs. The DArTseq technology, with its SNP and DArTSeq markers, offers a valuable tool for such studies, aiding in the preservation and sustainable use of species like Kigelia africana (Agyare et al., 2013; Bussmann et al., 2021; Cai et al., 2020; Joffe Pitta, 2003; Liu et al., 2020; Pascual et al., 2020; Tamokou & Kuete, 2014; Wikipedia, 2021).

1.2 Statement of the Problem and Justification of the study

Studies employing NGS to address conservation genomics and subsequent conservation strategies for threatened plants are still rare (Liu et al., 2020). The *Kigelia africana* has been a part of the traditional practices of the Agikuyu and Akamba people of Kenya for decades. As such, there is a need to give it a genetic identity to be able to conserve its precious economic and cultural value. Since this species is also grown in other parts of Africa, such as Zimbabwe and Malawi, the local landraces should be genotyped to allow further classification and insight into their genetic constituents and distribution. Most of the tree species are sparsely distributed in central and eastern regions of Kenya. Genetic data generated from this allows us to determine whether there is any genetic variation within the species from different locations within the Kenyan borders.

“Advances in molecular biology and high-throughput genotyping technologies have significantly impacted the field of plant conservation, shifting from a phenotype-based

to a genotype-based characterization. Molecular markers are invaluable tools for assessing plants' genetic resources by improving our understanding of the distribution and the extent of genetic variation within and among species" (Porth and El-Kassaby, 2014). This study sets to determine the genetic diversity and composition, given the tree's economic and ecological importance. This plant has great potential to be developed as a source of medical intervention by pharmaceutical industries according to (Saini et al., 2008).

1.3 Research Objectives

The overarching goal of this research is to comprehensively evaluate the genetic composition of ancient *Kigelia africana* trees distributed across various regions in Kenya, with a primary focus on genomic conservation. The study aims to employ cutting-edge DArTseq technology to achieve two specific objectives. Firstly, it seeks to assess the genetic diversity within *Kigelia africana*, utilizing DArTseq technology to generate comprehensive insights into the species' genomic makeup. Secondly, the research aims to investigate the genetic differentiation in allelic frequencies among different populations of *Kigelia africana* within Kenya. By addressing these specific objectives, the study endeavors to contribute novel findings to the understanding of the genetic intricacies of *Kigelia africana*.

LITERATURE REVIEW

2.1 Kigelia Africana classification and Morphology

Kigelia africana, a flowering plant classified under the domain Eukaryota, kingdom Plantae, and family Bignoniaceae, exhibits a wide distribution across sub-Saharan Africa, various islands, the Americas, and parts of Asia. The tree, growing up to 25 meters in height, features opposite or clustered leaves with imparipinnate leaflets. Genetic diversity within K. africana is crucial for the livelihoods of local communities and has implications for traditional and entrepreneurial uses. Clinical trials authenticate its pharmacological properties, particularly in traditional medicine for diseases like cancer. Despite efforts to scientifically validate its uses, standardization challenges persist in commercially available products. The need for genetic characterization, ethnobotany studies, and standardization is emphasized to enhance understanding, validate traditional uses, and isolate bioactive phytochemicals. In regions like Benin, the plant is employed for wound treatment, diabetes, toothache, and skin diseases. The conservation genomics aspect aims to provide genomic information supporting the preservation of Kenyan flora and effective conservation strategies for tree species globally (Areces-Berazain, 2022; Dossou-Yovo et al., 2022; Nabatanzi et al., 2020; Wambua Mukavi et al., 2020).

2.2 Molecular Markers used for diversity studies in Trees

"In recent times, molecular markers have proven to be invaluable tools for assessing genetic resources of tree plants by improving the understanding of the users about the distribution and the extent of genetic variation within and among the species. Knowledge of the genetic diversity of threatened tree species in any region of the world may contribute to the creation of effective strategies for their preservation, improvement, and future use"(Bedassa, 2018).

"A molecular or DNA marker is the difference in DNA nucleotide sequences between individual organisms or species, that is in proximity or closely linked to a target gene that expresses a trait. Usually, the target gene expressed trait or biological function, and the associated closely linked molecular marker are inherited together. The specific genomic location of the molecular marker within chromosomes is referred to as a locus or loci, and it may be known or unknown. The tight association of molecular markers to a trait or gene of a particular biological function, makes the markers serve as practical signs or flags that signal a particular gene locus and aid the detection or identification of the associated traits whether the genes involved are known or unknown and whether the gene(s) can be detected or not. Molecular or DNA markers do not influence traits associated with the expression or function of the linked gene or genes. DNA markers are useful for telling the individual genotypic differences (polymorphisms) in similar or different species. These differences are due to varied types of mutations of the DNA creating nucleotide sequence variations" (Amiteye, 2017).

The mutations causing these differences could be single nucleotide substitutions, rearrangements involving insertions or deletions, DNA section duplication, translocations, and inversions as well as mistakes in the replication of DNA that are tandemly repeated. Molecular marker signals that are used to reveal genotypic differences between individuals due to marker sequence differences are called polymorphic markers. On the other hand, DNA markers that cannot be used to

differentiate between or among genotypes are referred to as monomorphic markers. The characteristics of a good and very useful DNA marker are that the marker is ubiquitous and evenly distributed throughout the genome, easy to assay, replicable, cost effective, multiplexed, and can be automated. An ideal molecular marker must also be highly polymorphic, and co-dominant in expression to enable effective discrimination between homozygotes and heterozygotes, should be highly reproducible and possible to share data generated among laboratories. Also, a very good molecular DNA marker creates no detrimental effect on phenotype, is genome-specific in nature, and is multi-functional. DNA markers are categorized into various classes depending on the detection method: hybridization, polymerase chain reaction (PCR), and DNA sequence dependent molecular markers (Amiteye, 2017).

2.3 DArTSeq Technology

"A good example of sequence dependent molecular markers is the DArT (Diversity Array Technology Pty Ltd) markers. DArT markers were developed as one of the ultra-high-throughput, no prior sequence data-independent, cost effective, whole-genome genotyping techniques with a large number of markers that cover the entire genome. DArT markers have been applied successfully in genomic studies in many species including those with large and complex genomes such as barley, sugarcane, wheat, oat, and strawberry. The DArTseq method has been used in discriminating different species for population studies, diversity studies, characterization of germplasm, and studies involving genome-wide association" (Badu-Apraku et al., n.d.).

"DArT markers are developed through the use of combinations of restriction enzyme digestions to reduce genome complexity, followed by next-generation sequencing of complexity reduced representations or fragments to identify DNA polymorphisms and SNPs leading to the production of thousands of polymorphic loci in a single assay. The DArT platform generates two variants of markers, the SilicoDArT and DArTSeq SNP markers. SilicoDArT markers are dominant and are mostly scored for the absence (0) or presence (1) of a single allele while DArTSeq SNPs are co-dominant markers" (Adu et al., 2021).

A good quality genomic DNA of 50–100 ng amount is enough for purposes of DArT analysis. DArT overcomes many of the limitations of currently available marker technologies (Amiteye, 2017).

MATERIAL AND METHODS

3.1 Plant Materials

Most forests have been exposed to severe disturbance as a result of human activities, and the *K. africana* species is now found in patches in villages and national forest parks. To avoid materials from unknown sources, only ancient trees with a DBH (diameter at breast height) greater than 100 cm were selected for this study. Since this is a qualitative study, a total of 32 individuals were randomly collected from various regions in Kenya based on human interactions with the local people, especially those who brew the traditional *Muratina* beer. The formula used is: $Sample\ Size = \frac{Z^2 * p(1-p)}{C^2}$. Where Z is the confidence level, p is the expected proportion in population based on (Charan & Biswas, 2013) and expressed as decimal, and c is the confidence interval, expressed as decimal.

The name, geographic location, altitude, for each sample is recorded and described in table 1 below.

Table 1: Origin, collection sites and geographical coordinates of Kigelia africana landraces from Kenya used in this study.

Area name	County	Co-ordinates	Genotype	Quantity
Ruaka	Kiambu	-1.200527, 36.776289	Mur1,Mur2	2
Ruiru	Kiambu	-1.143403, 37.027777	Mur3	1
Juja	Kiambu	-1.2734316,36.7280686	Mur4-6	3
Witeithie	Kiambu	-1.062939, 36.995229	Mur7	1
Gatundu	Kiambu	-1.2734316,36.7280688	Mur8-9	2
Kangundo	Machakos	-1.2734316,36.7280689	Mur10-12	3
Matuu	Machakos	-1.2734316,36.7280690	Mur13-15	3
Katumani	Machakos	-1.612352, 37.203988	Mur16-18	3
Kieni	Nyeri	-0.318396, 36.753943	Mur19	1
Kanyariri	Embu	-1.2734316,36.7280693	Mur20	1
Siakago	Embu	-0.581557, 37.635987	Mur21-22	2
Maua	Meru	0.252559, 37.929558	Mur23	1
Kahuho	Kiambu	-1.195837, 36.674395	Mur24	1
Kandara	Muranga	-0.896207, 36.999131	Mur25-26	2
Kianjiruini, Maragua	Muranga	-0.795372, 37.117579	Mur27-29	3
Mida	Kilifi	-3.352570, 39.915182	Mur30	1
Dumbule	Kwale	-4.151469, 39.402422	Mur31-32	2

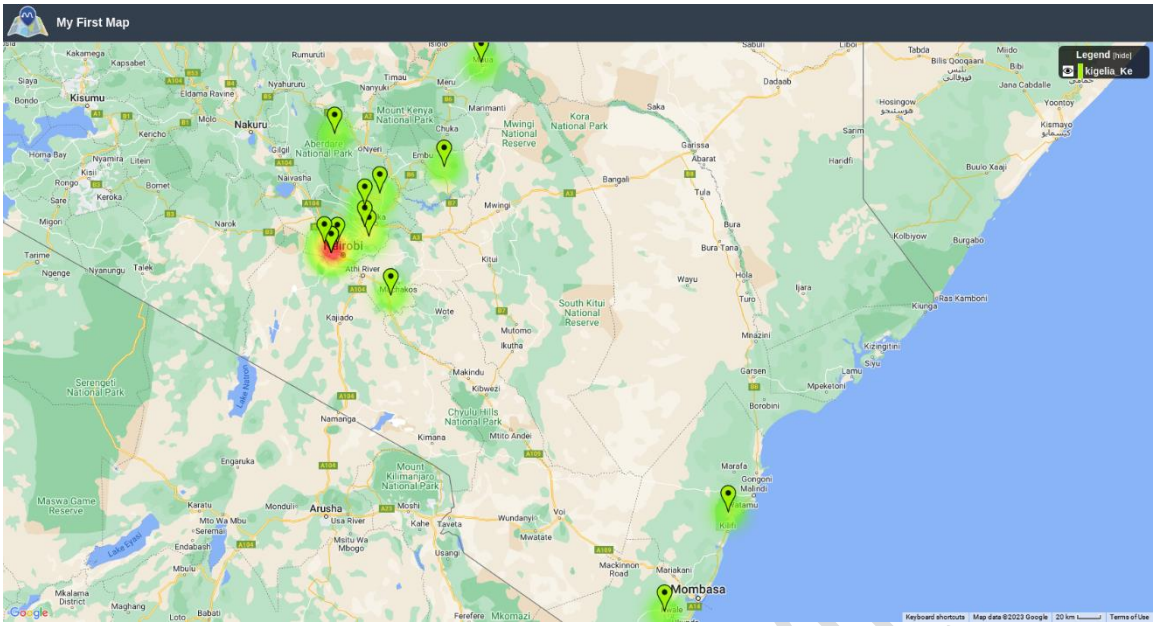

Figure 1: A geographical map of *Kigelia africana* sample collection locations in Kenya used in this study.

3.2 DNA Isolation

Leaf DNA was isolated and purified using the NucleoMag 96 Plant genomic DNA extraction kit (Macherey–Nagel, Du¨ren, Germany), following the manufacturer's instructions. Concentration of the extracted DNA were normalized within the range of 50–100 ng/ul. The quality and quantity of the DNA samples was then checked on 0.8% agarose gel.

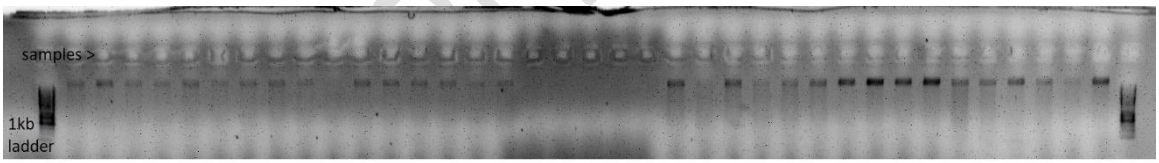

Figure 2: DNA bands on 0.8% Agarose Gel for the 32 *K. africana* samples

3.3 Library Construction and Sequencing

Libraries were constructed following the protocol described in (Kilian et al., 2012). Two DArTseq complexity reduction methods had to be tested since this was the first time these tree species were being sequenced. A rare cutting restriction endonuclease enzyme PstI (50 -CTGCA|G-30) in combination with two different frequently cutting restriction enzymes HpaII (50-C|CGG-30) and MseI (50 -T|TAA-30) were tested. The PstI/HpaII combination was selected as the best performing method. For each sample, 2 ul of DNA was digested with the PstI/HpaII restriction enzyme combination. Digestion products were ligated to barcoded adapters pair annealed to the two restriction enzyme overhangs. The PstI-compatible adapters include the partial attachment sequence for the 'Read 1 End' of the Illumina flow cell, a barcode of variable length (4–8 bp), and the PstI-compatible overhang sequence. The reverse adapters include the partial sequence for the 'Read 2 End' of the Illumina flow cell and MawI compatible overhang sequence. The adapter-ligated fragments were amplified in a Polymerase Chain Reaction (PCR) using optimized settings for a total of 35 cycles.

After PCR, equimolar amounts of the amplified products from each sample were pooled together, purified, and loaded on the cBot (Illumina, Inc., San Diego, CA, USA) for clustering on an Illumina Single Read flow cell. Libraries were then sequenced in the Illumina Hiseq 2500 using the single read sequencing protocol. A proprietary automatic genotypic data analytical pipeline, DArTsoft14, developed by DArT Pty Ltd, Canberra, Australia, was used to generate allele calls for SNP and DArT markers from the sequence data generated (Kafoutchoni et al., 2021). Ameker were scored a '0', '1', 'and '- ', representing presence, absence, and no-zero count for the silico dart markers. The SNP markers were scored as '1' for the SNP allele homozygote, '0' for reference allele homozygote, and '2' for heterozygotes (Adu et al., 2021). For this study, SNP markers were used as the preferred marker of choice.

3.4 Marker Quality Parameters

SNP markers were selected for best performance based on their polymorphic information content (PIC), percentage call rate, and marker percentage reproducibility from the duplicated sample replicates. The PIC shows the diversity of the marker within the populations while showing its ability to detect polymorphism among the individuals in a population. Since DArTseq and SNP markers are based on dominance (presence/ absence), PIC ranges from zero for monomorphic markers, to 0.5 for markers present in 50% of individuals and are absent in the remaining 50%. Markers quality parameters were trimmed automatically using the DArTsoft v14

The DArT software automatically computed several quality parameters for each DArTseq and SNP marker, such as call rate, polymorphic information content (PIC), and reproducibility of both markers(Baloch et al., 2017).

3.5 Genetic diversity and population relationship analysis

Population structure and genetic diversity were calculated from each of the 32 samples' DArTSeq and SNP data. The newly developed and released dartR version 2 for conservation genetic analysis was used for the statistical analysis and visualization of the data. Diversity indices were estimated to show the clear diversity, if any, between populations. These indices include observed and expected heterozygosity (Ho, He), population inbreeding coefficient (Fis), total gene diversity (Ht), and the gene diversity among collected samples (Dst)(Mijangos et al., 2022).

To get a clear picture of the genetic structure of Kigelia africana in Kenya, STRUCTURE software was used using the Bayesian clustering algorithm. This was flexibly estimated inside the dartR package. A neighbor-joining tree was constructed using the SNP and DArTSeq, principal components analysis (PCA) based on a pairwise genetic distance matrix of the accessions, and Hierarchical analysis of molecular variance (AMOVA) was used to support the hierarchical structure analysis. The genetic differentiation between populations was analyzed by estimating the pairwise fixation index (Fst) (Wadl et al., 2018). Similarities between trees will be estimated using Dice coefficients of similarity. The genetic similarity among genotypes will be estimated from the dissimilarity (distance) matrix generated from a simple matching coefficient. The resulting dissimilarity matrix will be further analyzed using the probability that the alleles at a random locus are identical in state (IBS). Principal component analysis (PCA) was used to assess the diversity among the *Kigelia africana* accessions (Padmaja, 2009).

RESULTS

4.1 DArTseq and SNP detection

A total of 11,793 SNP markers were generated after sequencing. A final selection of 3,703 markers were selected with an >90% reproducibility, and >80% call rate. DArTseq markers were reduced to 8,556 from a total of 26,352. This was due to a lot of low call rate markers below 80%. The average call rate was observed at 0.99% while reproducibility for the markers was observed at 1 meaning a 100% consistency in the marker scoring.

4.2 Genetic diversity and population structure

All markers had a PIC ranging between 0.39 to 0.45 and an average of 0.41 which is very informative. Overall polymorphism information content (PIC) of the DArTseq markers was 0.45 and 0.41for the SNP markers. The average expected heterozygosity (He) in the population varied from 0.30 for DArTseq and 0.41for SNPs (Table 2). The mean observed (Ho) and expected (He) heterozygosity (Table 2) collaborates with the high PIC values above.

Table 2. Basic statistics and genetic diversity of *K. africana* based on SNP and SilicoDArT markers.

	Ho	He	Hs	Ht	Dst	Htp	Dstp	Fst	Fstp	Fis	Dest
SNP	0.33	0.41	0.50	0.50	0.00	0.50	0.00	-0.01	-0.01	0.35	-0.01
silicoDArT	0.39	0.31	0.38	0.38	0.00	0.38	0.00	0.00	0.00	-0.33	0.00

The minor allele frequencies by locus for SNP data scored a minimum of 0.23 and a mean of 0.44. MAF for DArTseq dominant markers was not calculated.

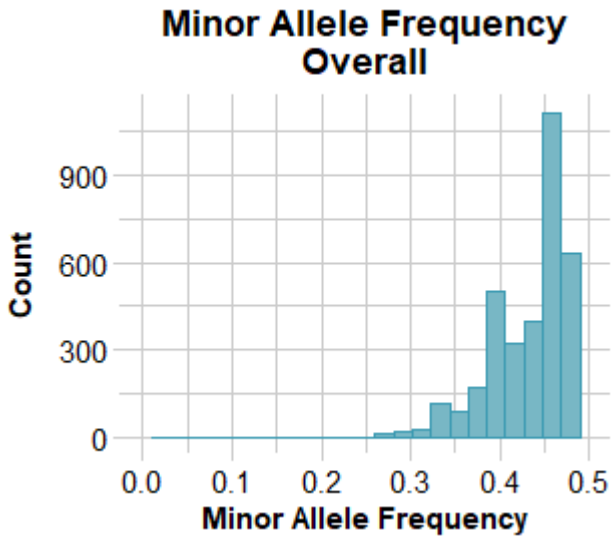

Figure *3*. The mean minor allele frequency (MAF) based on SNPs

4.3 Population structure analysis

Genetic similarities among the *K. africana* individuals were assessed using the SNP markers and the results revealed 3 clusters, which was also supported by the Delta-K plot. With more individuals in one cluster than the other two clusters of Kilifi and Nyeri populations, which had one sample each. A neighbor joining tree was

constructed and showed similar clustering based on the SNP and silicoDArT data (figure).

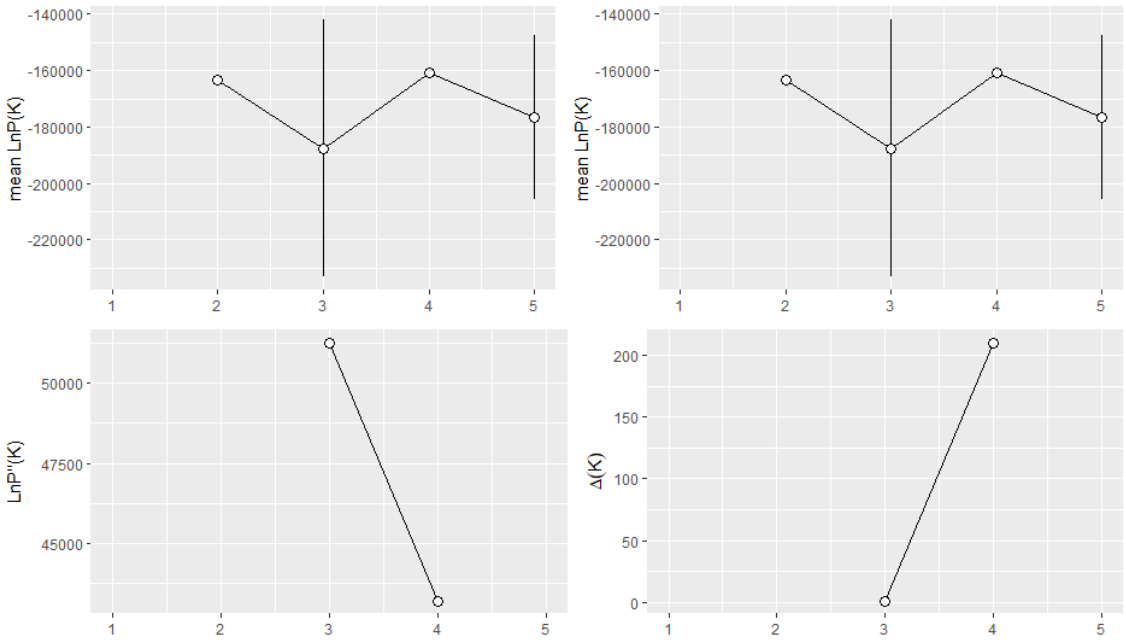

Figure 4: Mean LnP(K), LnP(K), and Delta-K(ΔK) observed in Structure analyses for K values of 1-5 in the K. africana populations.

A Neighbor joining tree was constructed from the Euclidean distances calculated from the DArTSeq and SNP data. The samples were grouped into 3 clusters based on location as seen in the figure below.

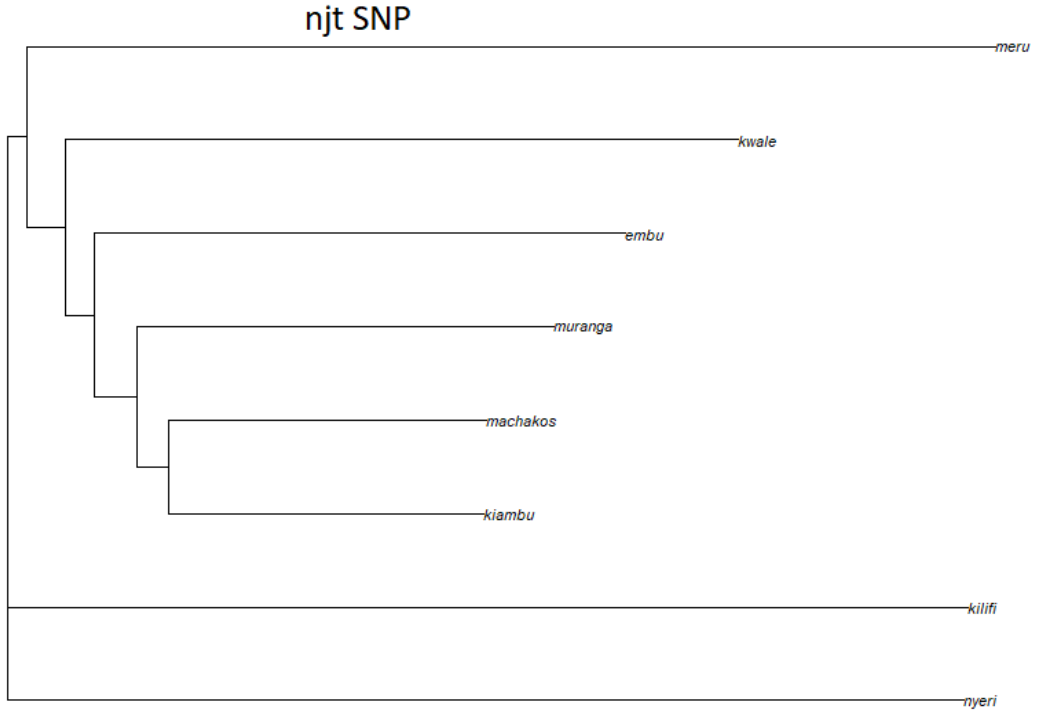

Figure 5: A neighbor joining tree of K. africana SNP data

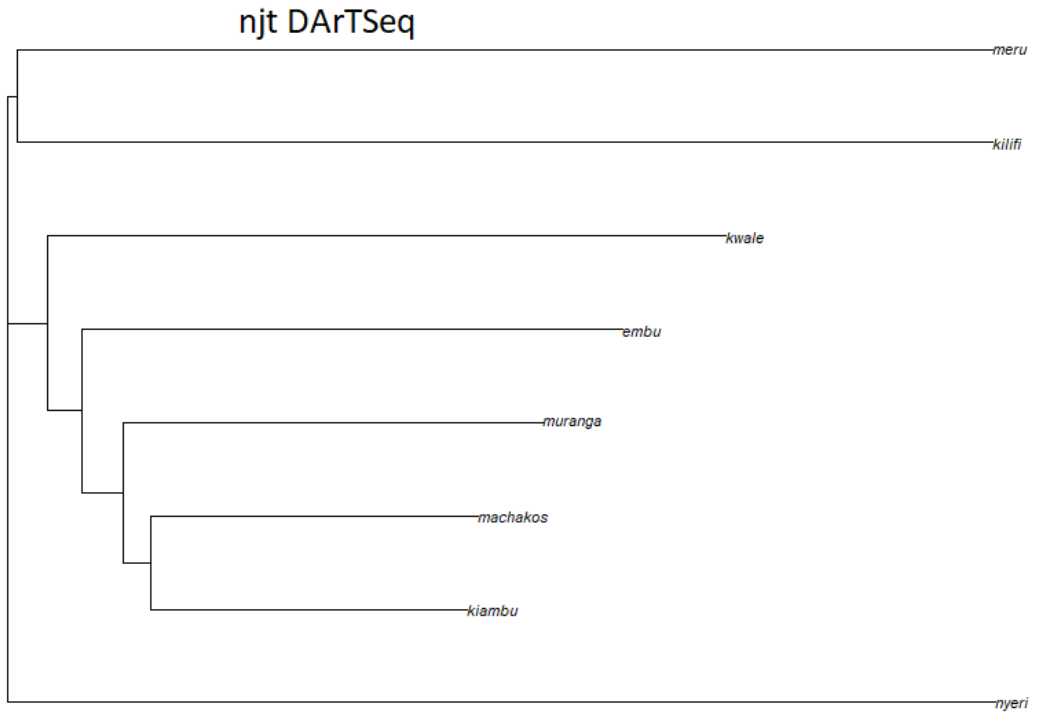

Figure 6: A neighbor joining tree of K. africana silicoDArT data

Based on (Sherwin et al., 2021), the diversity summary of the provided *K. africana* samples was calculated including the allelic richness (q = 0), Shannon information (q = 1), and heterozygosity (q = 2).

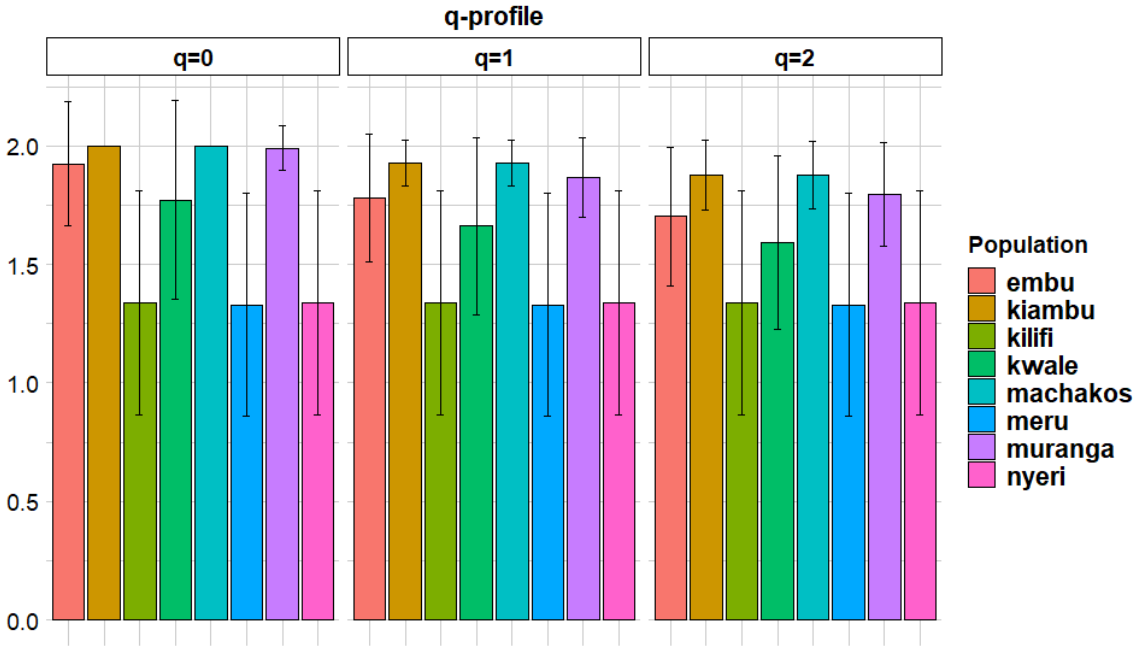

Figure 7: Population Diversity Summary based on SNP data

Individual genetic diversity was analyzed by principal coordinate analysis (PCA) as shown in figure 9 and 10 below. The PCA analysis showed very low average variance of 3.5% for silicoDArT, and 4.7% for SNP data.

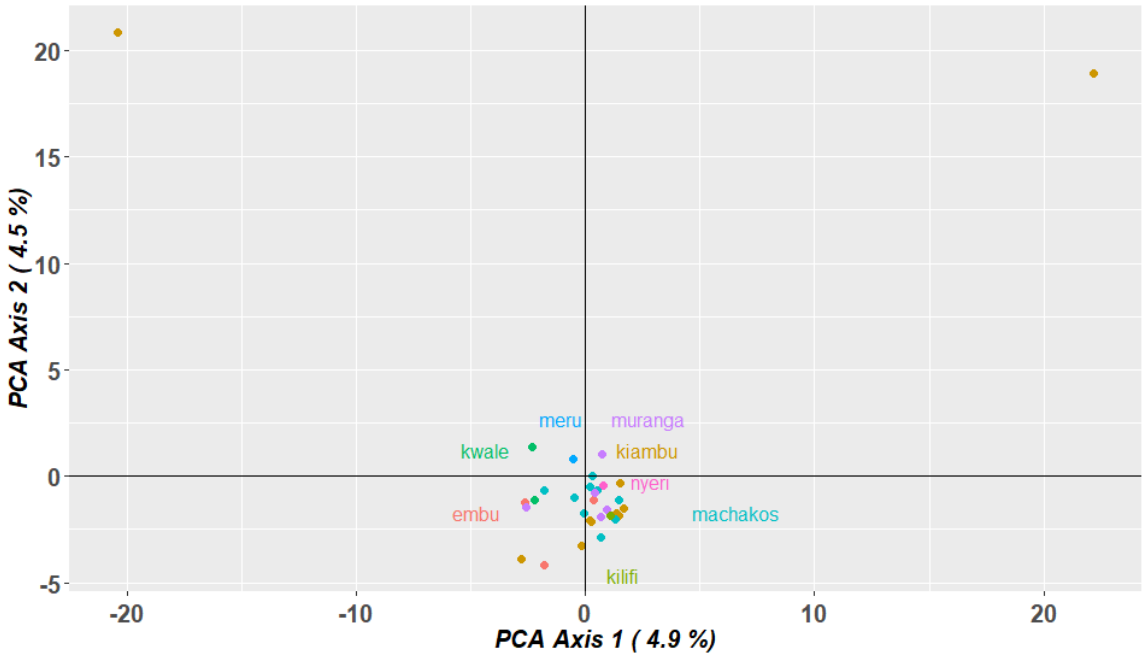

Figure 8: Principal coordinates analysis plot to infer group structure of *K. africana* based on SNP markers.

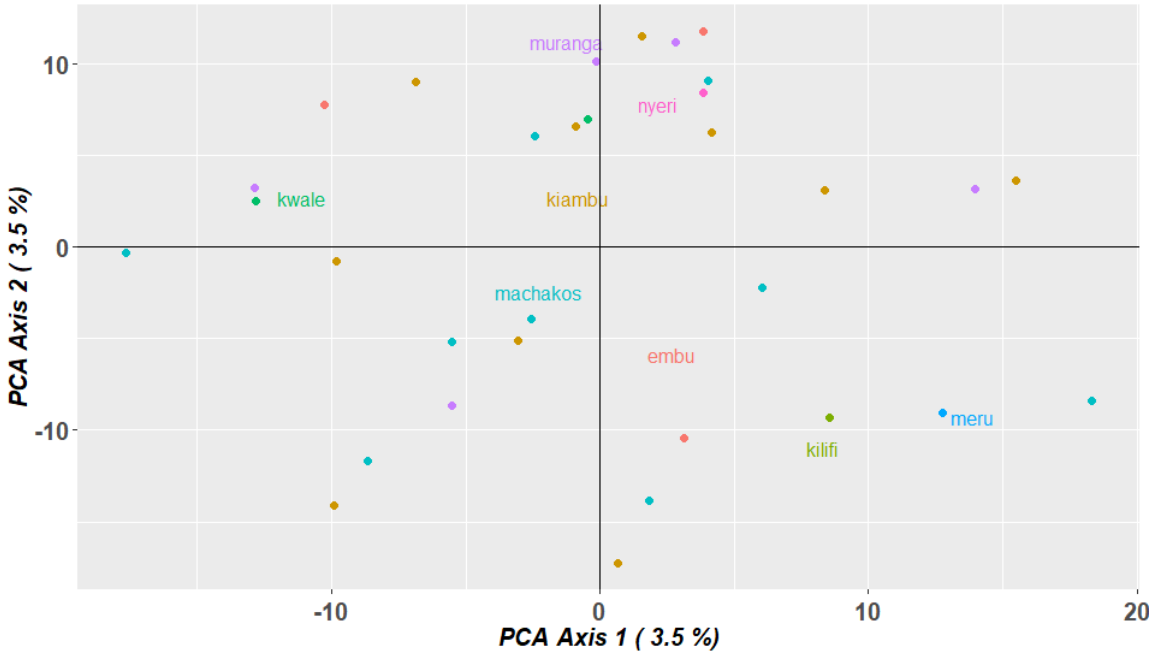

Figure 9: Principal coordinates analysis plot to infer group structure of *K. africana* based on silicoDArT markers

4.4 Sequence Similarity

Blasting all 3703 SNP and 8556 silicoDArT sequences revealed a much interesting result. Closely related matchs with e-value greater than 1.0E-11 were matching to *Sesamum indicum, Durio zibethinus, Carica papaya, Arachis duranensis, Erythranthe guttatus, Kolkwitzia amabilis, Solanum pennellii, Hevea brasiliensis, Utricularia reniformis, Hesperelaea palmer, Gossypium trilobum, Mimulus guttatus, Betula pendula, Capsicum annuum, Castilleja paramensis, Boea hygrometrica, Utricularia*

reniformis, Butomus umbellatus, Vitis vinifera, Primulina liboensis, Crescentia cujete, Lophophytum mirabile, and Tectona grandis. Most of which are tree, shrub, and herb species. This close similarity with these species suggests that the silicoDArT and SNP markers were of high quality. These blast results were from only 383 SNP markers, and 77 SilicoDArT markers from the total.

DISCUSSION, CONCLUSION AND RECOMMENDATIONS

5.1. Discussion

It's very important to understand the genetic diversity of indigenous tree species as this will shed some light on their relationships with other plant species, and important genetic and phytochemical potentials they might possess. The DArT platform proved to provide useful information on a never-before genotyped tree species, at an affordable price point. Two types of markers were used for detection, the silicoDArTs and the SNP markers. Both showed high call rate and reproducibility showed reduced genetic diversity, and strong genetic differentiation among other plant species. The high call rates and reproducibility are common among other tree species genotyped using the DArTseq technology, showing their reliability and consistency.

The results from the silicoDArT and SNP markers indicated low genetic variation in *K. africana* individual samples with potential consequences on the species' ability to recover from human population dynamics, genetic recombination, and environmental effects. Genetic diversity is measured commonly using the proportion of polymorphic loci and patterns of the observed vs expected heterozygosity. This therefore makes the PIC value ranges be described as low ranging from 0.0 to 0.10, medium as 0.10 to 0.25, high as 0.30 to 0.40, and very high as 0.40 to 0.50. The results showed both silicoDArTs and SNP had PIC ranging between 0.39 to 0.45 and an average of 0.41. This shows high to very high polymorphisms and high informativeness. Meaning the heterozygosity between the population was high.

Some tree samples were older than others, with at least 30 years of age difference, as this is the case with the Kilifi and Nyeri samples. A small insignificant genetic difference was observed between the tree species as seen by the allele frequencies below.

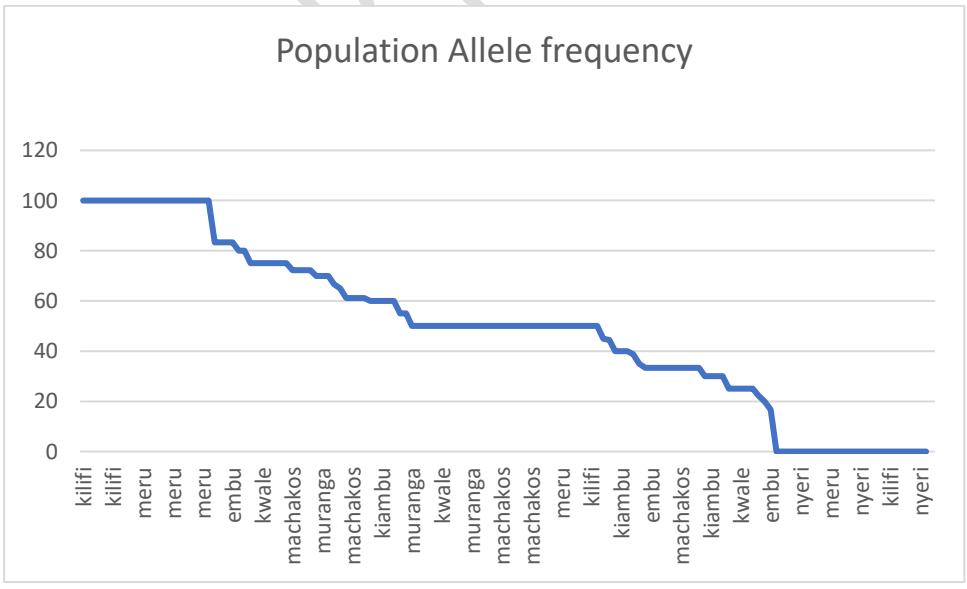

Figure 10: Allele Frequencies of various *K. africana* populations

The high PIC values observed and differences between Ho and He were consistent with the inbreeding coefficient (Fis), where Fis = 0.35 for silicoDArT and -0.33 for

SNPs. Positive Fis values are an indication that individuals in a population are more related than expected. And for SNP data having a -0.33 score shows the difference in detail data extracted, as SNP data is derived from SilicoDArT markers. However, these figures as compared to those from (Nantongo et al., 2022) which was above 0.5. This also shows that the species *K. africana* has not been adversely affected by anthropogenic factors during its existence. This makes sense, as the collection of these samples was often in remote locations with little human presence, hence low genetic diversity erosion. This was also backed up by the almost equal Ho and He values averaging 0.36 for both. When Ho is lower than He, this means there is a presence of inbreeding, also supported by the negative inbreeding coefficient (Fis) -0.33 for the SNP data.

Based on the cluster identified by the STRUCTURE analysis, low estimates of total genetic diversity (Ht), and genetic diversity (Dst) were observed more on the silicoDArT than in the SNP data (Table 2). Genetic differentiation (Fst) was lower in SNP data than in the silicoDArTs. There was also low inbreeding coefficient (Fis) from both data sets. The summary of the results shows low genetic variation within individuals and between populations using AMOVA analysis of the silicoDArTs (7.9 %), and SNPs (8.3%). SNP and silicoDArT data showed consistency as their association rated at 0.54 significance based on the Mantel test. All these tests were done using dartR package.

The observations from the neighbor joining tree showed that *K. africana* is moderately differentiated forming three distinct clusters. With the SNP clustering more tightly showing more variation as SNP markers are more abundant in plant genomes. This clustering was supported by the genetic differentiation values (Fst) which were below 0.01 showing low genetic differentiation according to (Nantongo et al., 2022). The genetic diversity index (Ht) was high for both SNP and silicoDArT at 0.50 and 0.38 respectively. This is an indication of high genetic diversity of the tree species due to high heterozygosity as shown by the high average MAF of 0.44%. This high heterozygosity may be due to restricted seed dispersal due to *K. africana's* reliance on animal vectors to transfer pollen from the flower to the stigma of another different individual as the trees as usually in isolated locations.

From the blast results, we observed closely related plant species that were mostly shrubs, trees, and herbs that at least have their sequence information available. This showed the potential pharmaceutical prospecting opportunities. The relatedness to these biodiverse plants shows potential high biodiversity that is beneficial to the world at large as biodiversity is essential for global food security(ENDEVR, 2023). The next step would be to isolate the genes these different species have in common and evaluate their genetic value. This will help future and aspiring scientists appreciate the potential and capacity of genetic variation as a key to human survival.

5.2. Conclusion

The study represents the first exploration of the genetic composition of ancient *Kigelia africana* populations in Kenya. The study has been able to identify low genetic variation within the Kenyan population, but there is a significantly high amount of genetic diversity within the Kenyan population. The potential diminishing

population size of the species despite its high genetic diversity is a threat to its genetic integrity. There is a need for *Kigelia africana* germplasm to be collected, characterized, and preserved from different populations across the African continent to maximize genetic variation conservation. It is clear that there is an immense deposit of pharmacological prospecting that is yet to be explored in these tree species.

The use of DArTseq technology enabled generation of high quality and reliable data for downstream analysis, while the use of the dartR package for statistical analysis proved to be a friendlier way for DArTseq and SNP data analysis using R software. To further explore the genetic potential of *K. africana*, in-depth research needs to be performed relating the genetic constituents and the therapeutical and/or pharmacological properties it's said to possess.

ABBREVIATIONS AND ACRONYMS

DArT: Diversity Arrays Technology

DArTseq: Diversity Array Technology sequence

GBS: Genotyping by Synthesis

PCA: Principal component analysis

SNP: Single Nucleotide Polymorphism

UPGMA: Unweighted pair group method with arithmetic mean.

PIC: Polymorphism Information Content

AMOVA: Analysis of Molecular Variance

References

Adu, B. G., Akromah, R., Amoah, S., Nyadanu, D., Yeboah, A., Aboagye, L. M., Amoah, R. A., & Owusu, E. G. (2021). High-density DArT-based SilicoDArT and SNP markers for genetic diversity and population structure studies in cassava (Manihot esculenta Crantz). *PLOS ONE*, *16*(7), e0255290. https://doi.org/10.1371/journal.pone.0255290

Agyare, C., Obiri, D. D., Boakye, Y. D., & Osafo, N. (2013). Anti-Inflammatory and Analgesic Activities of African Medicinal Plants. *Medicinal Plant Research in Africa: Pharmacology and Chemistry*, 725–752. https://doi.org/10.1016/B978-0-12-405927-6.00019-9

Amiteye, S. (2017). *Basic concepts and methodologies of DNA marker systems in plant molecular breeding*. https://doi.org/10.1016/j.heliyon.2021.e08093

Areces-Berazain, F. (2022). Kigelia africana (sausage tree). *CABI Compendium*, *CABI Compendium*. https://doi.org/10.1079/cabicompendium.29403

Badu-Apraku, B., Luísa Garcia-Oliveira, A., Petroli, D., Hearne, S., Adewale, S. A., & Gedil, M. (n.d.). *Genetic diversity and population structure of early and extra-early maturing maize germplasm adapted to sub-Saharan Africa*. https://doi.org/10.1186/s12870-021-02829-6

Baloch, F. S., Alsaleh, A., Shahid, M. Q., Çiftçi, V., E. Sáenz de Miera, L., Aasim, M., Nadeem, M. A., Aktaş, H., Özkan, H., & Hatipoğlu, R. (2017). A Whole Genome DArTseq and SNP Analysis for Genetic Diversity Assessment in Durum Wheat from Central Fertile Crescent. *PLOS ONE*, *12*(1), e0167821-. https://doi.org/10.1371/journal.pone.0167821

Bedassa, T. (2018). Molecular marker based genetic diversity in forest tree populations. *Forestry Research and Engineering: International Journal*, *2*(4), 176–182. https://doi.org/10.15406/freij.2018.02.00044

Bussmann, R. W., Paniagua-Zambrana, N. Y., & Njoroge, G. N. (2021). *Kigelia africana (Lam.) Benth. Bignoniaceae* (pp. 641–648). https://doi.org/10.1007/978-3-030-38386-2_97

Cai, M., Wen, Y., Uchiyama, K., Onuma, Y., & Tsumura, Y. (2020). Population Genetic Diversity and Structure of Ancient Tree Populations of Cryptomeria japonica var. sinensis Based on RAD-seq Data. *Forests*, *11*(11), 1192. https://doi.org/10.3390/f11111192

Charan, J., & Biswas, T. (2013). How to calculate sample size for different study designs in medical research? *Indian Journal of Psychological Medicine*, *35*(2), 121–126. https://doi.org/10.4103/0253-7176.116232

Dossou-Yovo, H. O., Vodouhè, F. G., Kindomihou, V., & Sinsin, B. (2022). Investigating the Use Profile of Kigelia africana (Lam.) Benth. through Market Survey in Benin. *Conservation*, *2*(2), 275–285. https://doi.org/10.3390/conservation2020019

ENDEVR. (2023, September). *Why Fruits Have Lost Their Vitamins*. ENDEVR Documentary. https://www.youtube.com/watch?v=2H3VhsnyCdI

Joffe Pitta. (2003, August). *Kigelia africana | PlantZAfrica*. http://pza.sanbi.org/kigelia-africana

Kafoutchoni, K., Agoyi, E., Symphorien, A., Assogbadjo, A., & Agbangla, C. (2021). Genetic diversity and population structure in a regional collection of Kersting's groundnut (Macrotyloma geocarpum (Harms) Maréchal & Baudet). *Genetic Resources and Crop Evolution*, *68*, 3285–3300. https://doi.org/10.1007/s10722-021-01187-4

Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., Caig, V., Heller-Uszynska, K., Jaccoud, D., Hopper, C., Aschenbrenner-Kilian, M., Evers, M., Peng, K., Cayla, C., Hok, P., & Uszynski, G. (2012). Diversity Arrays Technology: A Generic Genome Profiling Technology on Open Platforms. *Methods in Molecular Biology (Clifton, N.J.)*, *888*, 67–89. https://doi.org/10.1007/978-1-61779-870-2_5

Liu, D., Zhang, L., Wang, J., & Ma, Y. (2020). Conservation Genomics of a Threatened Rhododendron: Contrasting Patterns of Population Structure Revealed From Neutral and Selected SNPs. *Frontiers in Genetics*, *11*, 757. https://www.frontiersin.org/article/10.3389/fgene.2020.00757

Mijangos, J. L., Gruber, B., Berry, O., Pacioni, C., & Georges, A. (2022). dartR v2: An accessible genetic analysis platform for conservation, ecology and agriculture. *Methods in Ecology and Evolution*, *13*(10), 2150–2158. https://doi.org/10.1111/2041-210X.13918

Nabatanzi, A., Nkadimeng, S., Lall, N., Kabasa, J., & Mcgaw, L. (2020). Ethnobotany, Phytochemistry and Pharmacological Activity of Kigelia africana (Lam.) Benth. (Bignoniaceae). *Plants*, *9*, 753. https://doi.org/10.3390/plants9060753

Nantongo, J. S., Odoi, J. B., Agaba, H., & Gwali, S. (2022). SilicoDArT and SNP markers for genetic diversity and population structure analysis of Trema orientalis; a fodder species. *PLOS ONE*, *17*(8), e0267464. https://doi.org/10.1371/journal.pone.0267464

Padmaja, G. (2009). Uses and Nutritional Data of Sweetpotato. In G. Loebenstein Gad and Thottappilly (Ed.), *The Sweetpotato* (pp. 189–234). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9475-0_11

Pascual, L., Ruiz, M., López-Fernández, M., Pérez-Peña, H., Benavente, E., Vázquez, J. F., Sansaloni, C., & Giraldo, P. (2020). Genomic analysis of Spanish wheat landraces reveals their variability and potential for breeding. *BMC Genomics 2020 21:1*, *21*(1), 1–17. https://doi.org/10.1186/S12864-020-6536-X

Porth, I., & El-Kassaby, Y. (2014). Assessment of the Genetic Diversity in Forest Tree Populations Using Molecular Markers. *Diversity*, *6*, 283–295. https://doi.org/10.3390/d6020283

Saini, S., Kaur, H., Verma, B., & Singh, S. (2008). Kigelia africana (Lam.) Benth. — An overview. *Nat. Prod. Radiance*, *8*(2), 190–197.

Sherwin, W. B., Chao, A., Jost, L., & Smouse, P. E. (2021). Information theory broadens the spectrum of molecular ecology and evolution: (Trends in Ecology and Evolution 32:12, p:948–963, 2017). *Trends in Ecology & Evolution*, *36*(10), 955–956. https://doi.org/https://doi.org/10.1016/j.tree.2021.07.005

Tamokou, J.-D., & Kuete, V. (2014). Toxic Plants Used in African Traditional Medicine. In *Toxicological Survey of African Medicinal Plants* (pp. 135–180). Elsevier. https://doi.org/10.1016/B978-0-12-800018-2.00007-8

Wadl, P. A., Olukolu, B. A., Branham, S. E., Jarret, R. L., Yencho, G. C., & Jackson, D. M. (2018). Genetic Diversity and Population Structure of the USDA Sweetpotato (Ipomoea batatas) Germplasm Collections Using GBSpoly. *Frontiers in Plant Science*, *9*, 1166. https://www.frontiersin.org/article/10.3389/fpls.2018.01166

Wambua Mukavi, J., Wafula Mayeku, P., Muhoro Nyaga, J., & Naulikha Kituyi, S. (2020). In vitro anti-cancer efficacy and phyto-chemical screening of solvent extracts of Kigelia africana (Lam.) Benth. *Heliyon*, *6*(7), e04481–e04481. https://doi.org/10.1016/j.heliyon.2020.e04481

Wikipedia. (2021, September 18). *Kigelia*. Https://En.Wikipedia.Org/Wiki/Kigelia. https://en.wikipedia.org/wiki/Kigelia