**Diversity Study of Acacia Trees in Kenya using DArTseq Sequencing Technology**

# **ABSTRACT**

The proposed project aims to conduct a comprehensive diversity study of Acacia trees (Acacia spp.) in Kenya using Diversity Arrays Technology sequencing (DArTseq). Acacia trees play a vital role in the ecological and economic landscape of Kenya, but limited information exists regarding their genetic diversity and population structure. This study will employ cutting-edge DArTseq technology to analyze 30 Acacia tree samples collected from various regions in Kenya. The outcomes of this research will provide valuable insights into the genetic makeup, relatedness, and conservation strategies for Acacia trees in the region.

# **Abbreviations**

ASAL - arid and semi-arid lands

PCR - Polymerase chain reaction

DArT - Diversity Arrays Technology

DArTseq - Diversity Array Technology Sequence

GBS - Genotyping by Synthesis

PCA - Principal Component analysis

HWE - Hardy-Weinburg Equilibrium

SNP - Single Nucleotide Polymorphism

UPGMA - Unweighted Pair Group Method with Arithmetic Mean

PIC - Polymorphism Information Content

AMOVA - Analysis of Molecular Variance DNA - Deoxyribonucleic acid

# **INTRODUCTION**

Acacia trees are prominent components of ecosystems in Kenya, playing crucial roles in soil enrichment, biodiversity maintenance, and supporting numerous wildlife species. The acacia classification is of the kingdom: Plantae, order: Fabales, family: Fabaceae, genus: Vachellia, and species: V. *nilotica*. The Family Fabaceae or Leguminosae includes plants like Pea, bean or legume family, is the third largest Angiosperm (flowering plants) family with over 700 genera and about 20 000 species.

This tall tree reaches a height of up to 30 meters and is characterized by its smooth, yellow/green photosynthetic bark. The small bipinnate leaves feature paired straight stipules that are white and spinescent. The numerous bisexual flowers form round yellow spikes, each exhibiting regularity. The flowers contain exserted stamens and pistils with a superior ovary and extending style. The fruit is a non-sickle-shaped, flattish pod, tardily dehiscent, measuring up to 13cm in length. The tree has a dense spherical crown, with stems and branchlets often dark to black in color, fissured bark, and a greyish-pinkish slash that releases reddish low-quality gum. Young trees display thin, straight, light-gray spines in axillary pairs, typically ranging from 3 to 12 pairs long and 5 to 7.5 cm in length. Mature trees, on the other hand, usually lack thorns. The bipinnate leaves have 3-6 pairs of pinnulae and 10-30 pairs of leaflets, each leaflet measuring 4-5 mm long and exhibiting +/- tomentose characteristics. The rachis bears a gland at the bottom of the last pair of pinnulae. The flowers are arranged in globulous heads, 1.2-1.5 cm in diameter, of a bright golden-yellow color, either axillary or whorly on peduncles 2-3 cm long located at the end of the branches. The pods are grey, thick, softly tomentose, straight or slightly curved, measuring 5 to 15 cm long on a pedicel, and 0.5 to 1.2 cm wide (David Becking, 2023).

In Kenya, the most dominant species are the *Acacia Senegal, Acacia xanthophloea, Acacia nilotica,* and the *Acacia brevispica*. The arid and semi-arid land (ASALs) of Africa is mostly degraded due to human interference, and climate change. They barely receive adequate rainfall annually (less than 400 mm). Hence the Acacia tree species has been detrimental in ensuring both agroecosystems restoration, land reclamation through nitrogen fixation, and providing local communities with survival income (Omondi et al., 2010).

Acacia trees play a vital role as a valuable natural resource for rural communities inhabiting arid regions worldwide. These trees serve multiple purposes, including providing livestock fodder, medicinal resources, timber, poles, charcoal, and fuel wood. Acacia pollination is by insects, and they later develop fruits after 4 to 6 months. Additionally, Acacia plants contribute to sustaining various life forms while offering pollen and nectar for honey production. In the Arid and Semi-arid Lands of Kenya, specific Acacia species serve as crucial livelihood sources. In Kitui County, Kenya, efforts have been made to explore wild silk production, but the primary significance of Acacia woodlands lies in the generation of high-quality honey. The honey, renowned for its exceptional quality, experiences strong demand both locally and nationally, making honey production a significant source of livelihood for the communities in the area. (Warui et al., 2018). Acacia xanthophloea bark tannin could be a potential new source of vegetable tannin agent (Cheloti et al., 2023).

In India, there are more than 1500 medicinal plants and half of these are being effectively used in curing different diseases. And the leaves of the *Acacia nilotica* have been tested to have high levels of total phenolic content, higher antioxidant activity, and higher protein content, compared to the pods and bark. Hence the A. *nilotica* is being used in the control and cure of diabetis melitus as it contains anti-diabetic properties (Sadalage et al., 2023).

Understanding the genetic diversity and population structure of Acacia trees is essential for effective conservation and sustainable management. According to (Omondi et al., 2010), few studies have been done on the genetic diversity of the acacia trees of Kenya. One study of *A. senegal* using random amplified polymorphic DNA (RAPD) and inter-specific simple sequence repeat (ISSR), resulted in a moderate level of diversity (H = 0.283) of the tree species.

In this project, however, we were inclined to utilizing the latest, highly reproducible, and powerful DArTseq sequencing technology to assess the genetic variation among 30 Acacia tree samples collected from diverse regions in Kenya. DArTseq uses the genotyping-by-sequencing (GBS) technology to sequence and generate data from novel non-referenced genomes. This then generates single-nucleotide polymorphisms (SNP) and DArTseq markers called silicoDArTs. This technology has proved over many studies to be robust, and of high quality in genomics studies across various species and applications (Cai et al., 2020).

Thus, the objective of this research is to; (1) assess the genetic diversity of Acacia trees in Kenya using DArTseq technology; (2) identify the population structure and relatedness of Acacia tree species across different regions in Kenya, and (3) provide valuable data for the formulation of effective conservation strategies and further research for Acacia trees in the region.

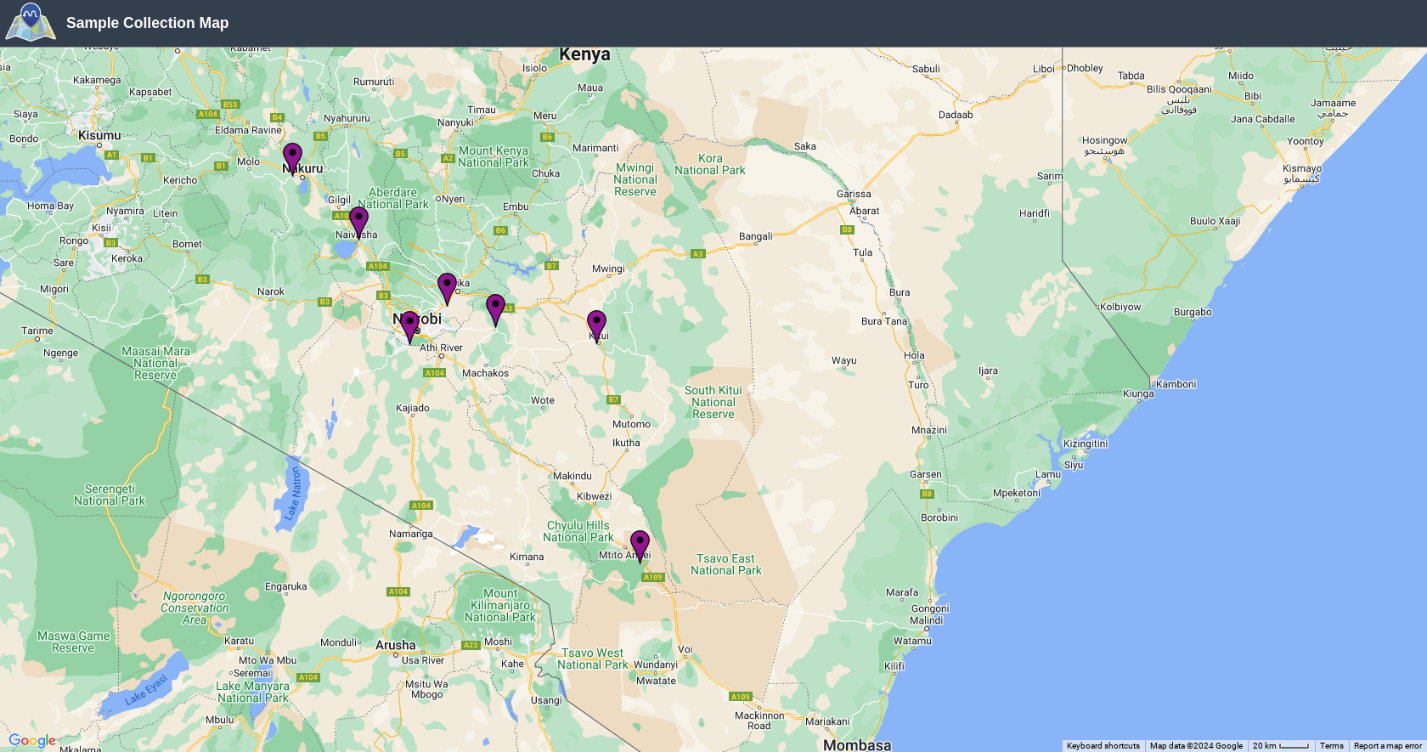
# **2. MATERIALS AND** **METHODS**

## **2.1 Plant Material:**

A total of 7 Acacia tree leaf samples were randomly collected from different ecological zones across Kenya, ensuring representation from distinct habitats. Their geographical position co-ordinates were recorded. The low number is due to the wild and unchartered nature of some of these tree locations including wildlife game reserves. To ensure collection of ancient samples, the trees had to have a DBH (diameter at breast height) greater than 90 cm.

**Table 1: Origin, collection sites and geographical coordinates of Acacia samples from Kenya used in this study**

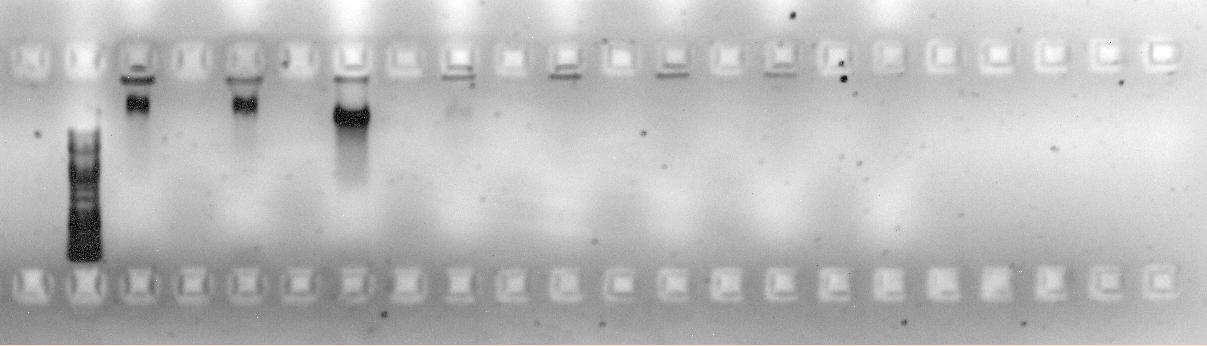
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sample | Sample type | County | Area Name | Co-ordinates |
| SBK1 | Leaf | Kitui | Kitui | -1.379637, 37.987279 |
| SBK2 | Leaf | Taita taveta | Tsavo | -2.800206, 38.262899 |
| SBK3 | Leaf | Nakuru | Kaptembwa | -0.295316, 36.019097 |
| SAK1 | Leaf | Nakuru | Naivasha | -0.709782, 36.446686 |
| SAK2 | Bark | Machakos | Tala | -1.272193, 37.329637 |
| SAK3 | Bark | Kiambu | Juja | -1.135752, 37.020516 |
| SAK4 | Bark | Kajiado | Ongata Rongai | -1.384197, 36.775701 |



**Figure 1: The geographical map of Acacia sample collection locations in Kenya used in this study**

## **2.2 DNA Extraction:**

High-quality genomic DNA were extracted from the leaves of each collected sample using using the NucleoMag 96 Plant genomic DNA extraction kit (Macherey–Nagel, Du¨ren, Germany), following the manufacturer’s instructions. The quantification of the extracted DNA was then checked manually using 0.7% agarose



**Figure 2: DNA Quality Check showing bands on 0.7% Agarose Gel for the 7 Acacia samples**

**Table 2: DNA quality and purity was measurements recorded using a DeNovix spectrophotometer.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sample | Sample type | Concentration (ng/ul) | 260/280 | 260/230 |
| SBK1 | Leaf | 9.760 | 1.79 | 1.08 |
| SBK2 | Leaf | 202.150 | 1.58 | 0.56 |
| SBK3 | Leaf | 42.529 | 1.17 | 0.87 |
| SAK1 | Leaf | 31.660 | 1.59 | 0.50 |
| SAK2 | Bark | 28.149 | 1.114 | 0.24 |
| SAK3 | Bark | 3.894 | 1.37 | 0.42 |
| SAK4 | Bark | 7.532 | 1.38 | 0.36 |

## **2.3 Library Preparation and Sequencing:**

The genomic DNA samples underwent DArTseq library preparation, which is a genome representation method that targets specific genomic regions, mostly exons due to methylation sensitivity. This technology allows for the simultaneous screening of thousands of loci, providing extensive coverage of the genome. 2ul of DNA, in full replication was digested using a combination of the rare cutting endonuclease enzyme PstI (5`-CTGCA|G-3`) in combination with the frequently cutting restriction enzyme MseI (5`-T|TAA-3`). The resulting digested DNA fragments were ligated to other pieces of fragments like the barcoded adaptors (4-9 bp) and sequencing adapter overhangs. The new fragments were then amplified in a Polymerase Chain Reaction (PCR) machine using optimized settings for 35 cycles. The PCR products were then checked for quality and correct fragment size by running them in a 1.2% agarose gel.



**Figure 3: Acacia library quality control check in agarose showing failed sample SAK3**

Sample SAK3, failed quality check, due to poor DNA concentration and purity as shown in the quality scores highlighted above.

The samples that passed the quality check were then pooled together, purified and loaded on to the cBot (Illumina, Inc., San Diego, CA, USA) for cluster generation via bridge amplification on a single read flowcell. The sequencing was then carried out on an Illumina HiSeq2500. Allele calling was done using the proprietary analytical pipeline called DArTsoft14, developed by DArT Pty Ltd, Canberra, Australia. SNP and DArT markers are then generated where DArTseq markers scored a ‘0’, ‘1’, ‘and ‘- ‘, representing presence, absence, and no-zero count. The SNP markers were scored as ‘1’ for the SNP allele homozygote, ‘0’ for reference allele homozygote, and ‘2’ for heterozygotes presence.

## **2.4 Genetic Diversity and Population**

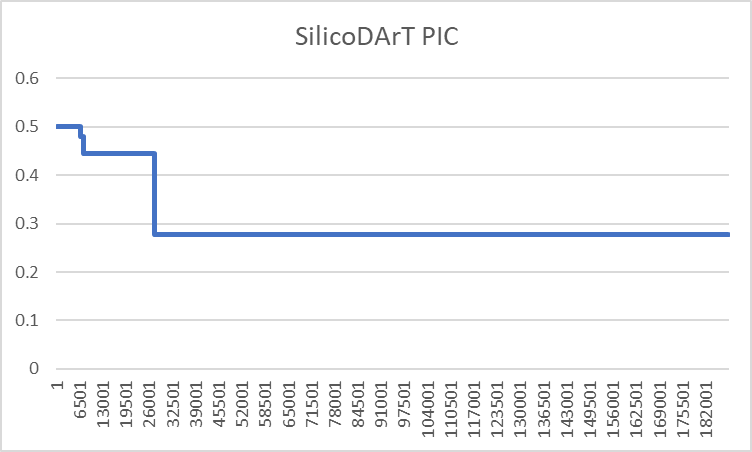
Genetic diversity and population structure was calculated from the 6 samples using both DArTseq and SNP data. Using the new version of dartR for genetic data analysis and visualization, diversity indices such as inbreeding coefficient, total gene diversity, and heterozygosity was determined. Using the SNP and DArTSeq, a neighbor-joining tree was built. The hierarchical structure analysis was supported by the use of hierarchical analysis of molecular variance (AMOVA), principal components analysis (PCA), and a pairwise genetic distance matrix of the accessions. Estimating the pairwise fixation index (Fst) allowed for the analysis of genetic divergence between the Acacia populations.

# **3. RESULTS**:

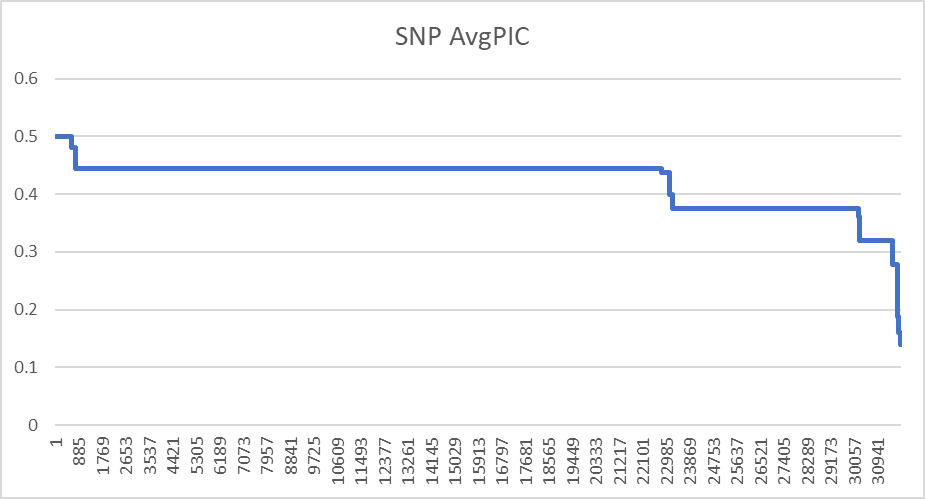
## **4.1 DArTseq and SNP Detection**

After sequencing was done in the HiSeq2500, a total of 49,678 SNP markers were generated. A final selection of 31,823 markers were selected with over 90% reproducibility, and above 90% call rate. DArTseq markers were reduced to 188,499 from a total of more than 500,000 markers. The file was too large to work with due to many low call rate markers below 70%. The average call rate was observed at 99% while reproducibility for the markers was observed at 95%. This was all done automatically from the proprietary DArTSoft14 algorithm program after sequencing is complete.

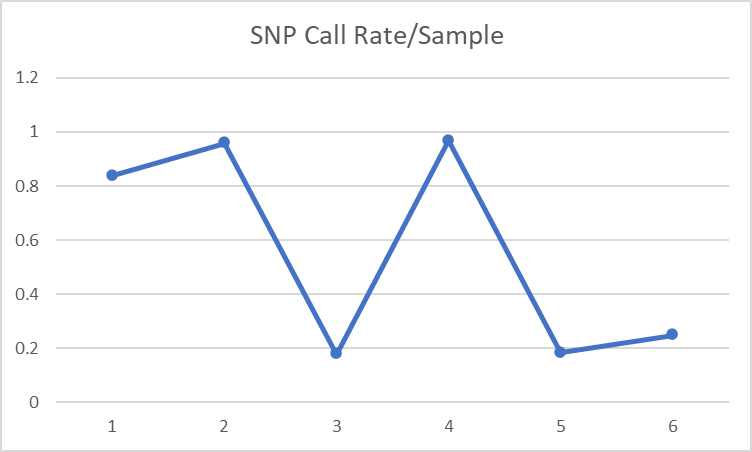
The summary of polymorphism information content (PIC) for SNP markers was between 0.5 and 0.14 with an average of 0.42. For the DArTseq markers, the PIC ranged from 0.5 to 0.27, with an average PIC of 0.30.



**Figure 4. Observed PIC of silicoDArT data**



**Figure 5. Observed PIC of SNP data**



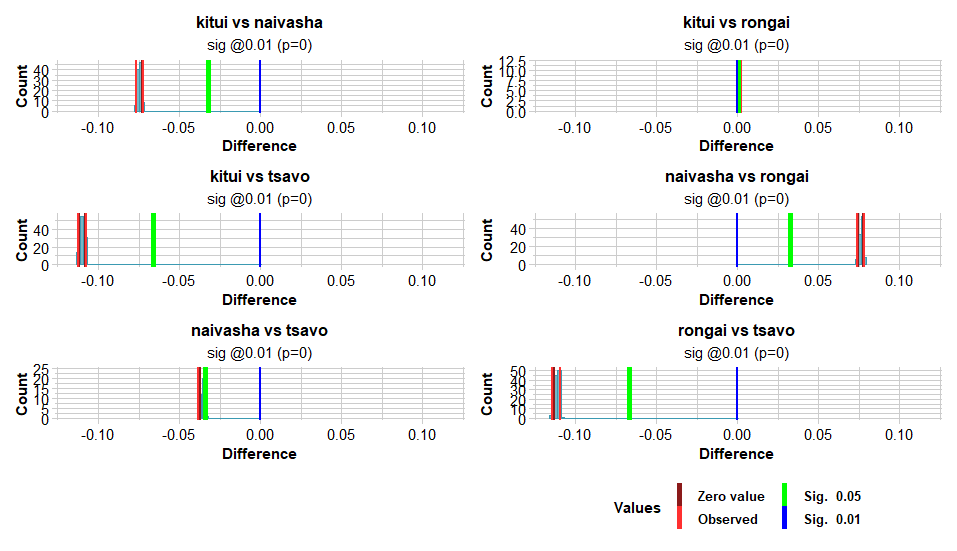
**Figure 6. Observed SNP call rate per sample**

## **4.2 Genetic Diversity**

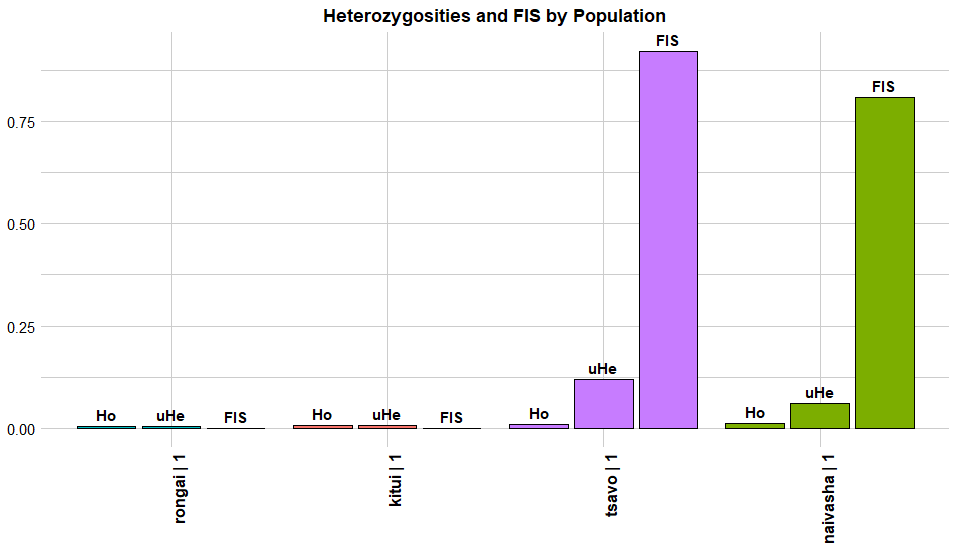
The expected heterozygosity (He) for SNP markers in the population ranged from 0.0029 to 0.068. The mean observed heterozygosity (Ho) and the He, were in synchronization with the high PIC values described above.

**Table 3. SNP observed and expected heterozygosity based on population**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| pop | nLoc | Ho | HoSD | He | HeSD | uHe | uHeSD | FIS |
| kitui | 26676 | 0.005848 | 0.076249 | 0.002924 | 0.038125 | 0.005848 | 0.076249 | 0 |
| naivasha | 31148 | 0.011558 | 0.095125 | 0.034802 | 0.125478 | 0.060584 | 0.218437 | 0.809228 |
| rongai | 7900 | 0.00443 | 0.066418 | 0.002215 | 0.033209 | 0.00443 | 0.066418 | 0 |
| tsavo | 30965 | 0.009414 | 0.093294 | 0.068719 | 0.171818 | 0.120287 | 0.300756 | 0.921739 |



**Figure 7.SNP data Heterozygosity variation between populations**



**Figure 8. SNP observed heterozygosity per population**

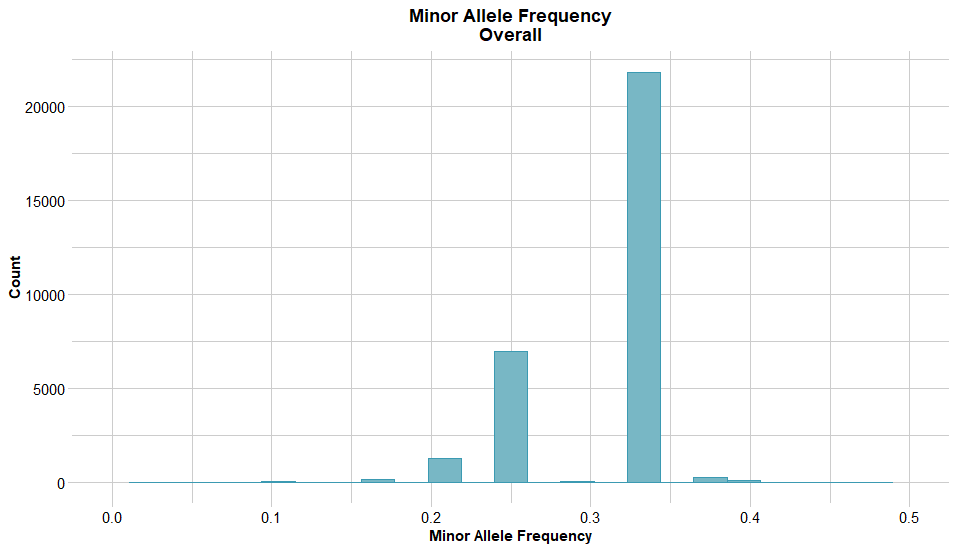
## **4.3 Population Structure Analysis**

Similarities in the genetic makeup among the Acacia samples were assessed using the SNP markers exposing 5 clusters. This is being strictly supported by the delta-k plot. A neighbor joining tree was developed showing similar clustering between the SNP and SilicoDArT data. The phylogenetic tree grouped the samples into three groups, with two internal nodes (Figure nj SNP Silicos). The locus minor allele frequency (MAF) for the SNP data set scores between 0.08 and 0.5, with a mean of 0.3. Allelic richness (q =0) of the acacia samples was calculated, including Shannon information (q = 1), and heterozygosity (q = 2). Population Structure analysis using the STRUCTURE software, revealed 3 variated group structures within SNP marker data as shown below.

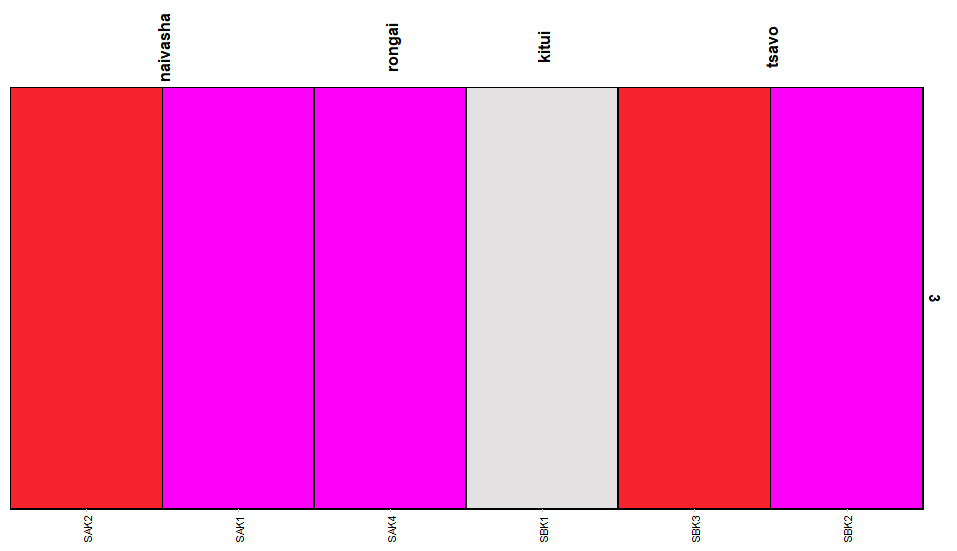
Genetic diversity of individual samples was backed up by principal coordinate analysis (PCA) despite individuals per population being too little, with the minimum required being ten individuals per population. These showed a substantial variance in diversity gap between the various populations based on combined SNP and SilicoDArT data.

**Table 4. Basic SNP statistics for Acacia samples based on SNP data**

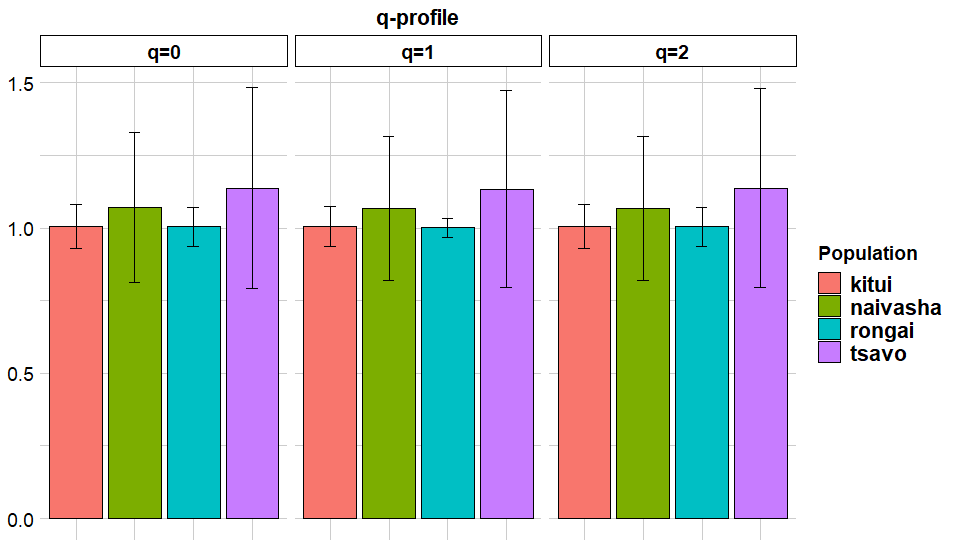
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Ho** | **Hs** | **Ht** | **Dst** | **Htp** | **Dstp** | **Fst** | **Fstp** | **Fis** | **Dest** |
| 0.008 | 0.5631 | 0.5459 | -0.0173 | 0.5308 | -0.0323 | -0.0316 | -0.0609 | 0.9858 | -0.0739 |



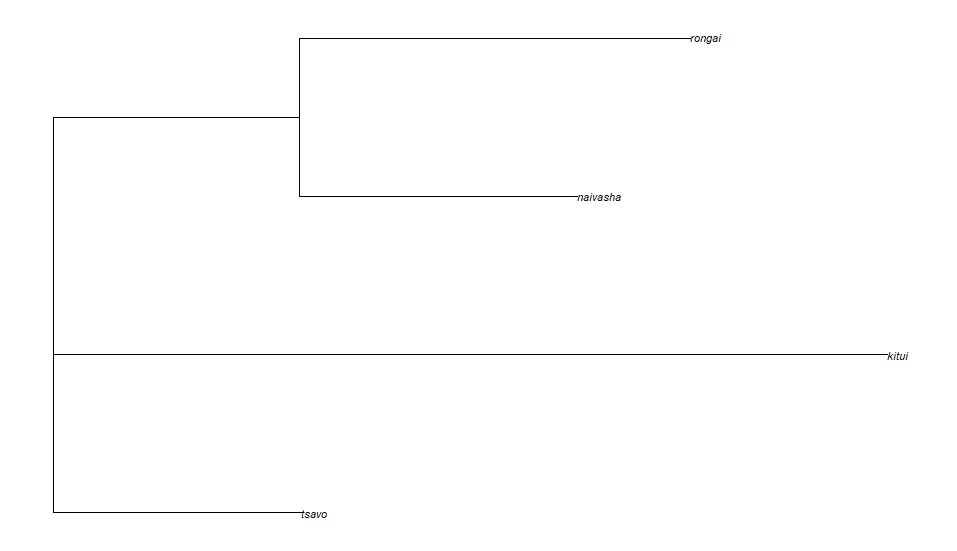
**Figure 9. The mean minor allele frequency (MAF) for all SNP loci**



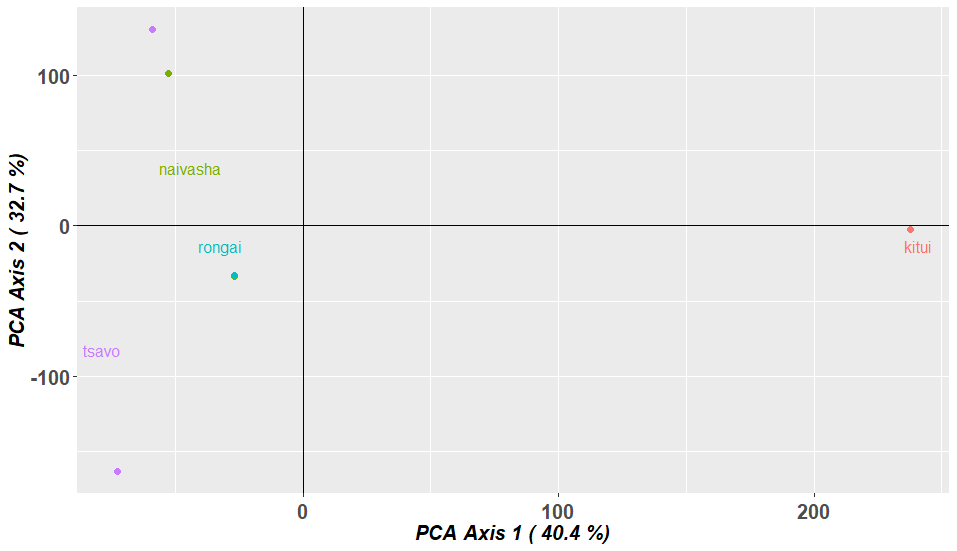
**Figure 10. SNP population structure heat map**



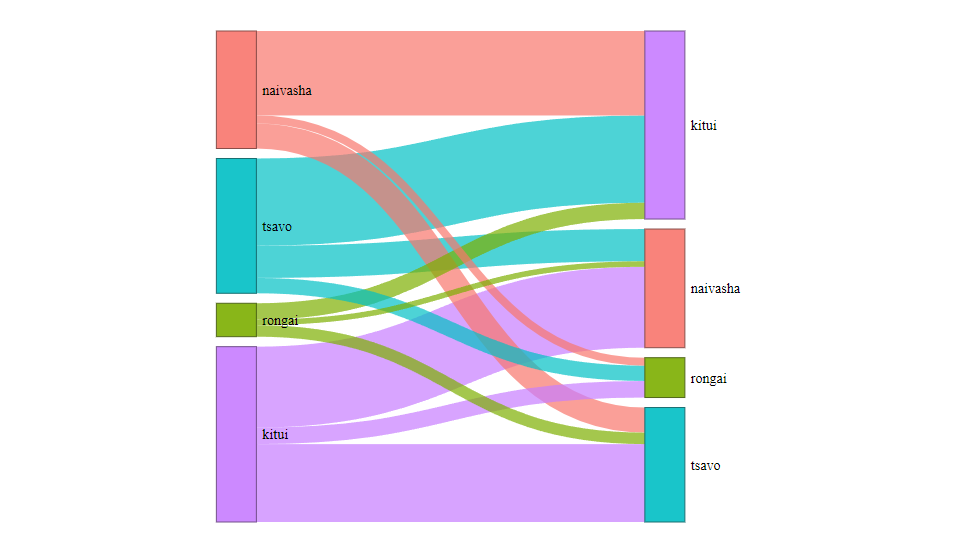
**Figure 11. SNP population diversity summary.**



**Figure 12. A SNP neighbour joining tree of the Acacia population**



**Figure 13. PCA plot to infer group structure of Acacia based on SNP marker data.**



**Figure 14. SNP mean absolute allele frequency differences (AFD) between pairs of populations.**

## **4.4 Sequence Similarity**

The 31,823 SNP markers generated were exposed to a nucleotide blast search. The results were further filtered based on E-value (equal to 999), and Chromosome Position (equal to 0). This left a total of 1,896 SNP marker. All blast hits had an E-value greater than 1.24E-26. The SNP markers matched to various species such as *Acacia mellifera, Acacia mangium, Dichrostachys cinerea, Lupinus angustifolius, Lophophytum mirabile, Bretschneidera sinensis, Parapiptadenia rigida, Arachis duranensis,* *Acacia koa, and Anadenanthera colubrina.*

After E-value marker filtration, 3320 SilicoDArTs were selected for nucleotide blast, the markers were matching with *Acacia harpophylla, Acacia argyrophylla, Acacia mangium, Arachis duranensis, Dichrostachys cinerea,* *Millettia pinnata,* *Parkia javanica,* *Piptadenia communis, Pararchidendron pruinosum, Bupleurum falcatum,* and *Lophophytum mirabile.*

The blast results from the SNP and SilicoDArT markers showed similar species similarities which mostly consist of shrubs, trees, and herb species.

# **4. DISCUSSION**

Indigenous tree species like the Acacia, have been in existence for thousands of years and it is important to understand their genetic structure and genetic components for conservational purposes. Their potential for pharmaceutical use needs to be understood even from a genetic perspective. Affordable sequencing the Acacia tree samples using DArTSeq technology provided major insights on the genetic diversity of the tree species across the country of Kenya. Both silicoDArT and SNP markers showed high reproducibility (above 95%) on technical replicates, and high call rate for marker scoring therefore indicating their efficiency and reliability.

Genetic diversity measured using the proportion if polymorphism in the sequence data shows low genetic variance between the different Acacia species. This also indicated low effects on the tree species from environmental and human factors. The observed versus the expected heterozygosity also showed little difference that was almost negligible (0.001). The SNP data averaged a PIC of 0.42, while silicoDArT was 0.30. This is high PIC translating to high genetic diversity within the populations due to high heterozygosity.

Older tree samples like the SBK1 from Kitui county and the SAK1 from Naivasha showed a little amount of significant difference in unique alleles as shown in figure.14 above. Both populations have the highest amount of unique SNP markers hence contributing to their variance. This is also supported by the mean allele frequency of 0.31 as seen in Figure 9.

# **CONCLUSION AND RECOMMENDATIONS**:

This project was a pioneering effort to explore the genetic diversity and population structure of Acacia trees in Kenya using DArTseq sequencing technology. The findings will have significant implications for conservation efforts, guiding policymakers and conservationists in preserving the rich genetic heritage of Acacia trees and their associated ecological importance in the region.

# **REFERENCES**

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

Cheloti, M., Fedha, P., Kiprop, A., Onyuka, A., Asava, A., Mutuku, M., Induli, M., Kundu, B., & Masenge, E. (2023). Evaluation of Mechanical Properties of Goat Leather Tanned using Acacia xanthophloea. *Textile & Leather Review*, *6*, 333–342. https://doi.org/10.31881/TLR.2023.053

David Becking. (2023, December). *The Trees of Southern Africa*. https://treesa.org/vachellia-xanthophloea/?unapproved=77441&moderation-hash=df1f45fc547293d1bc8e6f2586442c10#comment-77441

Omondi, S., Kireger, E., Dangasuk, O., Chikamai, B., Odee, D., Cavers, S., & Khasa, D. (2010). Genetic Diversity and Population Structure of Acacia senegal (L) Willd. in Kenya. *Tropical Plant Biology*, *3*, 59–70. https://doi.org/10.1007/s12042-009-9037-2

Sadalage, N., Baidya, M. M., Kondrapu, P., Ghiware, N., Shaha, A., Kalubhai, S., Nare, S., & Rajput, S. (2023). *Phytochemical evaluation, in vitro antioxidant activity and in-vivo antidiabetic activity of acacia nilotica.* https://doi.org/10.31838/ecb/2023.12.6.211

Warui, M., Gikungu, M., Bosselmann, A., & Hansted, L. (2018). *Pollination of Acacia woodlands and honey production by honey bees in Kitui, Kenya*. *6*, 40–50.