International Journal of Applied Research
Journal HP: www.intjar.com; ISSN: 2411-6610

Molecular phylogeny and genetic diversity of *Dodonaea viscosa* populations from North Rift and Coastal Kenya using ITS1 and rbcL markers

P.K. Langat[1]*, E.W. Njenga[1], P. Jeruto[1], C. W. Lukhoba[2], F.M. Musila[3]

[1]Department of Biological Sciences, University of Eldoret, Kenya
[2]Department of Biological Sciences, University of Nairobi, Kenya
[3]Department of Biological and Life Sciences, Technical University of Kenya, Kenya

ARTICLE INFO

Article history

Received: 21 March 2025
Accepted: 27 April 2025

Keywords

Dodonaea viscosa, ITS1, rbcL, Molecular phylogeny and PCA

Corresponding Author

P.K. Langat
Email: pklangat1985@gmail.com

ABSTRACT

The genus *Dodonaea* (Sapindaceae) comprises slightly over 70 species. *Dodonaea viscosa* is a plant of Australian origin with seven subspecies. In Kenya, there are two subspecies. *D. viscosa* has been used medicinally to manage several ailments. It is considered a difficult taxonomic group due to its polymorphic nature. Owing to its uses in ethnomedicine and the taxonomic challenge of Kenyan subspecies, a molecular phylogenetic study was done using *ITS1* and *rbcL* genes from Kenyan populations. DNA was extracted, sequenced, exported to MEGA 11, PhyloSuite, and R for analysis. Aligned sequences were subjected to evolutionary analyses and PCA while phylogeny was constructed using maximum likelihood (ML) method. Nucleotide frequency and substitution rates, nucleotide disparity index, and evolutionary divergence of the genes were very low implying a short genetic distance among the Kenyan populations. ML trees showed that the highland and coastal *Dodonaea* populations were closely related as they are monophyletic and have high bootstrap values. PCA showed that *rbcL* genes contributed to most variations in *Dodonaea* populations where two clusters based on *rbcL* genes were observed. In contrast, *ITS1* genes contributed to fewer variations and grouped all *Dodonaea* populations into one cluster. This study provides more data to solve the taxonomic challenge of *D. viscosa*. There is a need for further studies targeting other gene regions to understand well the phylogenetic relationships of Kenyan populations of *Dodonaea*.

1. Introduction

The genus *Dodonaea* belongs to the family Sapindaceae and comprises over 70 species (Buerki *et al.,* 2010; Christmas *et al*., 2018). *Dodonaea viscosa* (L.) Jacq is a plant of Australian origin given the name *Dodonaea* in honor of Rembert Dodoens a 16th century botanist and *viscosa* derived from viscosus a Latin word meaning sticky (Lawal & Yunusa, 2013). It is distributed over warm-temperature, tropical, and subtropical climates in South America, Africa, Mexico, India, the Northern Mariana Islands, Virginia Islands, Florida, Arizona, and South America (Anilreddy, 2009). It is the only species of *Dodonaea* distributed outside Australia (Harrington, 2008). There are seven subspecies of *Dodonaea*, including subspecies *angustissima, cuneata, mucronata, spatulata, viscosa, burmanniana,* and *angustifolia*. Three of them viz *viscosa, burmanniana,* and *angustifolia* are distributed outside Australia with *viscosa* and *angustifolia* found in Kenya (Christmas *et al.,* 2015; Kaigongi, 2020a).

D. viscosa is a tree or shrub that grows in a wide range of habitats including rocky or salty soils, wind, and drought to a height of 7 metres (Anilreddy, 2009; Rani *et al*., 2009; West 1984). It has alternate, simple leaves without stipules, very short petiole (up to 2.5 mm long), or none at all. The blade is oblanceolate or broadly to narrowly elliptical, with a conspicuous midrib on both sides and 15-20 often indistinct pairs of lateral veins. The margins are entire, both surfaces glabrous but glandular, and the leaves are coated (especially when young) with viscid glandular exudates (Lawal & Yunusa, 2013). The inflorescent flowers are inconspicuous forming colorful capsules which are their fruits with varying numbers of wings (2-4) Khan & Ismail, 2019; Kuswanto *et al.*, 2022). Their flowers are either male or female borne on different plants hence the plant is dioecious (Orpin, 2018).

D. viscosa has been used for thatching and making hedges to protect residences and cultivated crops while young stems are used for spinning "Kothani" a type of handcraft (Kannan 2016). Its leaves are used as fodder while the dry branches are fuel (Murad *et al*., 2013). In Kenya the Pokot community use this plant also as fodder and for construction (Wigrup, 2006). The pastoralists from Baringo County uses the plant to construct houses

Cite this article: *Langat, P.K., Njenga, E.W., Jeruto1, P., Lukhoba, C. W. & Musila, F.M. (2025). Molecular phylogeny and genetic diversity of Dodonaea viscosa populations from North Rift and Coastal Kenya using ITS1 and rbcL markers. International Journal of Applied Research, 11: 34-44. DOI: 10.5281/zenodo.15742932*

and fences since it is resistant to infestation by mites (Vehrs & Heller, 2017). Medicinally, the plant has been used around the world (Al-Snafi, 2017; Patel & Coogan, 2008; Shiaka 2019) to manage a number of ailments including dermatitis, rheumatism, stomachaches, splenic pains, skin rashes, antiplasmodial and also having radical scavenging properties (Omosa *et al.,* 2016; Rao & Pratap, 2022).

D. viscosa has been considered a difficult taxonomic group owing to its polymorphic nature (Khan & Ismail, 2019). The two Kenyan taxa, which include *D. viscosa* and *D. angustifolia*, remain a major challenge taxonomically. Some taxonomists consider *D. angustifolia* as a subspecies of *D. viscosa* (Christmas *et al*., 2015; Hossain*,* 2019; Priyadi, 2021) while others treat them as varieties (Rani *et al*., 2009) and as distinct species (Beentje, 1994). This could be attributed to many taxonomic revisions that it has undergone (Kaigongi *et al*., 2020b) and the fact that its classification has been based on mainly leaf characters (Christmas *et al*., 2018).

Owing to its diverse ethnobotanical importance and the taxonomic challenge of this widely distributed species, an in-depth taxonomic study was done to provide more genetic data that could be useful in identification. This was based on DNA sequence data obtained from the *ITS1* region of nuclear ribosomal DNA and *rbcL* of the chloroplast genes. The relationships between the populations were determined by phylogenetic trees obtained.

2. Materials and Methods

2.1. Sample collection and identification

Populations of *D. viscosa* subspecies *angustifolia* leaves were collected from Baringo, Elgeyo Marakwet, Uasin Gishu counties while *D. viscosa* subspecies *viscosa* leaves from Kwale county in Kenya. They were transported to the University of Eldoret Herbarium for identification by a taxonomist. The samples were placed in ziplock bags and preserved in a cooler box waiting for DNA extraction.

2.2 DNA extraction

A modified CTAB method of Rathod *et al.*, 2018, was used in the extraction of DNA. A pre-chilled mortar and pestle were used to crush 0.5 g of young leaf tissue from various *Dodonaea* populations. They were crushed using 1 ml of lysis buffer, which contained 250 mM NaCl, 25 mM EDTA (pH 8.2), 0.5% SDS, 200 mM Tris base, and 0.1 g of polyvinylpyrrolidone (PVP) for a 1000 ml extraction buffer solution. The crushed extracts were moved to another autoclaved microcentrifuge tube. Every tube was handled in the same way, as described herein. Using a float, the microcentrifuge tubes containing the powdered homogenate were incubated for 1 hour at 65°C. The tubes were centrifuged using a cooling centrifuge (5415 R) at 14000 rpm for two minutes at 4°C. Using a micropipette, the supernatant was transferred to newly labeled tubes, and the samples were mixed with an equal volume of chloroform: isoamyl alcohol (24:1). To guarantee optimal pigment extraction in the chloroform layer, the tubes were carefully mixed by slowly inverting them fifteen to twenty times. The tubes were centrifuged for 10 minutes at 4°C at 14000 rpm. Using a micropipette, the supernatant was carefully transferred to new tubes. Each supernatant was then given 12 µl of RNase (1 mg/ml stock) and gently mixed by inverting the tubes five to ten times.

The tubes were incubated in a water bath for thirty minutes at 37°C. The samples were mixed gently with an equal volume of chloroform, and isoamyl alcohol (24:1) by inverting the tubes 4-6 times. The samples were then centrifuged at 14000 rpm for 10 minutes at 4°C. The aqueous layer was moved to a fresh tube, and chilled ethanol was added to it and incubated for 12 hours. After being thawed on ice, the microcentrifuge tubes were centrifuged for five minutes at 4°C at 14000 rpm. After discarding the supernatant, 750 µl of cooled 70% ethanol was added to each tube's pellet, and the tubes were gently inverted 2 times to wash the DNA pellet. The tubes were centrifuged once more for five minutes at 4°C at 14000 rpm, and the supernatant was disposed of with a micropipette. The DNA pellets were dried for 1 hour and suspended in Nuclease-free water.

2.3 PCR and sequencing

ITS1 primer was used for PCR amplification of the *ITS1* section of nrDNA while *rbcL* was used for amplification of the chloroplast gene (Patwardhan, 2014; Korall & Kenrick, 2002; Omelchenko *et al*., 2022). The two primers are presented in Table 1.

Table 1: The primers used in PCR

Primer	Direction	Primer sequence	Reference(s)
ITS1	F	5'-GGAAGGAGAAGTCGTAACAAGG-3'	(Omelchenko *et al*., 2022).
	R	5'-AGATATCCGTTGCCGAGAGT-3'	
rbcL	F	5'-ATGTCACCACAAACAGAAAC-3'	(Algarni, 2022; Shamso & Fouad, 2019).
	R	5'-TCGCATGTACCTGCAGTAGC-3'	

The primers used in this study were obtained from Macrogen, Europe. For each population, a Master mix for PCR was prepared by the addition of 5µL flexi DNA polymerase, 2µL Magnesium chloride ($MgCl_{2)}$, 14.3µL Nuclease-free water, primers (0.5µL of forward and 0.5µL of reverse), 0.4µL dNTPs (dGTP, dATP, dCTP, dTTP), Taq DNA polymerase and 2µL DNA. A final concentration of 25µL of the master mix was constituted and placed in each PCR tube. A Thermal Cycler (PTC-100) was used for the amplification reactions. It was set up to include a three-minute initial denaturation at 94°C, one-minute annealing at 58 °C, one-minute extension at 72 °C, and a final 7-minute extension at 72 °C after 35 cycles. A gel electrophoresis was finally done to confirm whether amplification was done. The products were sent to Macrogen, Europe for purification and subjected to a DNA Cycle sequencing Kit to sequence the DNA based on Sanger method (Kluesner *et al*., 2018; Men *et al*., 2008).

2.4. Phylogeny construction and Evolutionary analysis

Basic evolutionary analyses and phylogenetic reconstructions were performed on the obtained DNA sequences as recommended by (Tamura *et al*., 2021) and (Kumar *et al*., 2008) in MEGA 11 and Phylosuite (Zhang, *et al*., 2020). *ITS1* and *rbcL* forward sequences of *D. viscosa* were manually edited at the ends to remove incorrect bases that could compromise quality of DNA. Then, the Multiple Sequence Alignment (MSA) program MAFFT was used to align the DNA sequences (Katoh *et al*., 2009). Evolutionary analyses included measures of nucleotide substitution rates, base composition disparity index, and evolutionary divergence between *Dodonaea* subspecies. Codon positions included in the analyses were 1st+2nd+3rd+Noncoding. During the phylogenetic analysis of the two genes, Hasegawa-Kishino-Yano substitution model was followed (Takahashi & Nei, 2000). The model assumed the rates and patterns of substitution were uniform among the four nucleotide sites. The method of tree construction was maximum likelihood (ML) whereas for the ML tree, the heuristic approach was the Subtree Pruning Regrafting (SPG). The test for phylogeny was bootstrap resampling where the number of bootstrap replications were set at 5000 for each ML tree constructed. The trees with the highest log likelihood were selected as the best trees to depict the phylogenetic relationships of coastal and highland *Dodonaea* populations under study. An ML tree was also constructed based on concatenated alignments from the two genes. PartitionFinder2 was used to partition the concatenated dataset to choose the model and partitioning scheme that best suits the concatenated sequences. (Lanfear *et al*., 2017). ML trees were rooted by including *Sinoradlkofera minor* as an outgroup to show the hypothetical last common ancestor of the *D. viscosa* under study. *Sinoradlkofera* is a genus within the Sapindaceae family just like *Dodonaea.* Sequences for the outgroup were obtained from NCBI GenBank (KX527258.1 and OR199385.1 for *rbcL* and *ITS1* genes respectively) and then aggregated with the *Dodonaea* sequences of the populations under study during the generation of the ML phylogenetic tree. Finally, ML trees were viewed and updated in ITOL (Interactive Tree of Life), an online platform for managing, displaying, and annotating phylogenetic trees (Letunic & Bork, 2024; Yang *et al*., 2023).

2.5. Principal Component Analysis (PCA)

Sequence data were subjected to PCA. Principal components are a few linear combinations of the original variables that maximally explain the variance of all the variables. PCA can be used to show relationships between samples and variables (Bro & Smilde, 2014). In our case, PCA can show which genes are distantly related from the others, or the ones which have a lot of variations compared to the rest. PCA requires numerical data, hence the sequence data recorded by a set of five digits (A, T, G, C, and a gap (-) was replaced by boolean vector that is presented by 0 and 1 forming a sequence matrix that was later converted into numerical data. This transformation of DNA or protein data into numerical data has merits that no

information is lost and so it is completely reversible. This transformation was done in R statistical program following R script provided by (Konishi, 2019). Obtained numerical data was then subjected to PCA in R using FactoMineR and Factoextra R packages to identify principal components and quality of representation of the principal components.

3. Results and Discussion

3.1 Gel electrophoresis, PCR & Sequencing

Gel electrophoresis was performed on eleven (11) populations of *D. viscosa* collected from various sites in Kenya (Figure 1). This was done to ascertain the presence of genomic DNA. DNA amplification was done at the molecular laboratory at the Center for Biotechnology and Bioinformatics (CEBIB), in the University of Nairobi. The DNA sequences obtained were purified and Sanger sequenced at Macrogen Europe, Netherlands. Eleven *rbcL* and eleven *ITS1* genes obtained were later deposited in the NCBI Genbank and their accession numbers are shown in Table 2.

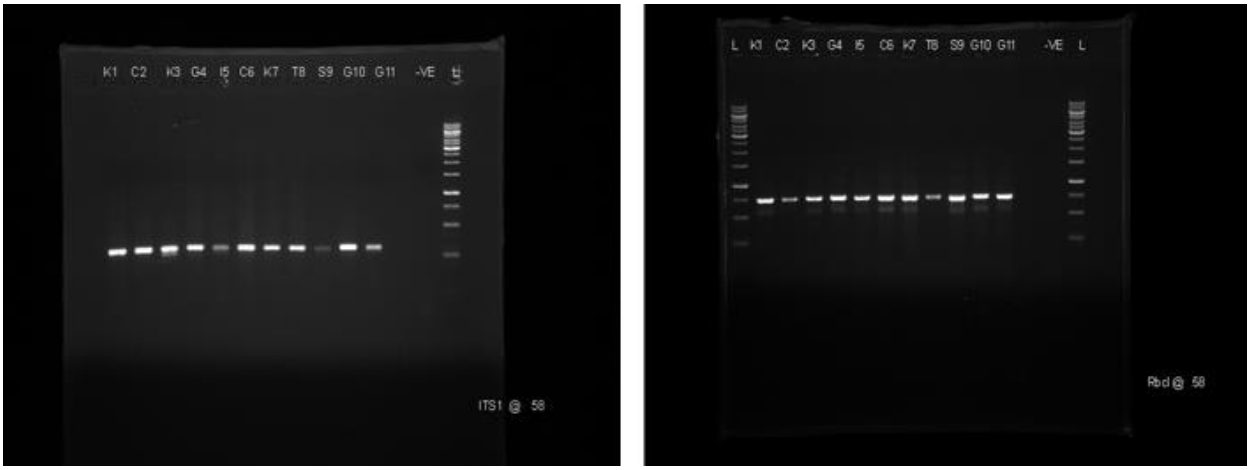

Figure 1: Gel electrophoresis image of the *ITS1* and *rbcL* gene from 11 populations with white bands showing genomic DNA. K1: Kabarnet, C2: Cheploch, K3: Koriema, G4: Gazi 3B, I5: Iten, C6: Chepyogot, K7: Kapchemutwa, T8: Turbo, S9: Sergoit, G10: Gazi M5 and G11: Gazi 1C

Table 2: Kenyan *Dodonaea* populations Genbank accession numbers

Dodonaea viscosa populations	Gene	Genbank Accession Number	Gene	Genbank Accession Number
Cheploch; Kenya, Rift Valley	*rbcL*	PQ901279	*ITS1*	PV056202
Chepyogot; Kenya, Rift Valley	*rbcL*	PQ901280	*ITS1*	PV056203
Gazi M5; Kenya, Mombasa/Coast	*rbcL*	PQ901282	*ITS1*	PV056205
Gazi 1C; Kenya, Mombasa/Coast	*rbcL*	PQ901283	*ITS1*	PV056206
Gazi 3B; Kenya, Mombasa/Coast	*rbcL*	PQ901281	*ITS1*	PV056204
Iten; Kenya, Rift Valley	*rbcL*	PQ901284	*ITS1*	PV056212
Koriema; Kenya, Rift Valley	*rbcL*	PQ901286	*ITS1*	PV056207
Kapchemutwa; Kenya, Rift Valley	*rbcL*	PQ901287	*ITS1*	PV056208
Kabarnet; Kenya, Rift Valley	*rbcL*	PQ901285	*ITS1*	PV056209
Sergoit; Kenya, Rift Valley	*rbcL*	PQ901288	*ITS1*	PV056210
Turbo; Kenya, Rift Valley	*rbcL*	PQ901289	*ITS1*	PV056211

3.2. Nucleotide frequency and substitution rates

The best model for explaining the pattern of nucleotide substitution during evolution has the lowest Bayesian Information Criterion (BIC) score (Tamura *et al*., 2021). From this analysis, the best nucleotide evolutionary model of both *rbcL* and *ITS1* sequences was Hasegawa-Kishino-Yano (HKY) model (Takahashi & Nei, 2000) with the lowest BIC value of 13737.355 and 52590.554 for *rbcL* and *ITS1* genes respectively. According to the HKY model, the frequency of each nucleotide of adenine (A), thymine (T), cytosine (C), and guanine (G) was 0.27, 0.25, 0.19, and 0.27,

respectively and the rate of nucleotide substitution in the *rbcL* sequences ranged from the lowest value of 0.04 in AC and GC to the highest value of 0.17 in GA and AG. Similarly, the frequency of nucleotides in the *ITS1* sequences adenine (A), thymine(T), cytosine (C), and guanine (G) was 0.25, 0.21, 0.23, and 0.29, respectively while the rate of nucleotide substitution in the *ITS1* sequences ranged from lowest value of 0.05 in AC, AT, GT, and GC to the highest value of 0.14 in GA (Figure 2).

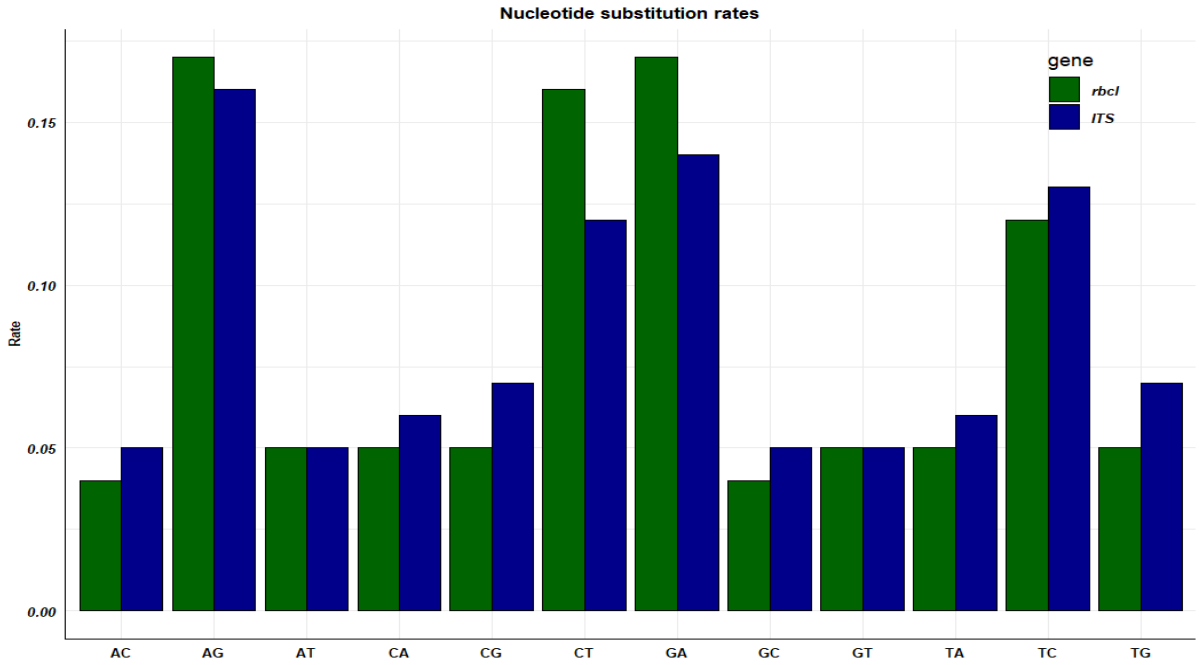

Figure 2: Nucleotide substitution rates in both *rbcL* and *ITS1* genes

3.3. Nucleotide Disparity Indices

Base disparity index measures homogeneity of substitution patterns between molecular sequences. In comparative sequence analysis, it is frequently assumed that the sequences have undergone the same nucleotide substitution trend over time. Violation of this assumption is known to significantly affect the accuracy of phylogenetic inference and tests of evolutionary hypotheses. Figure 3, displays the base composition bias disparity values for each site between sequencing pairings in the populations under study. Larger variations in base composition biases are shown by values greater than zero, which explains the heterogeneity in the substitution rates (Al-Atiyat & Aljumaah, 2014). A lower nucleotide disparity index indicates a closer relationship. Based on the *rbcL* sequences, the highest value was observed between Turbo group and the rest of the groups where the disparity index between Turbo *rbcL* was >1 in all comparisons. Other higher disparity indices were observed between Sergoit *rbcL* and Cheploch *rbcL* and finally between Karbarnet *rbcL* and Cheploch *rbcL* genes. The other comparison between the other groups gave low disparity indices of zero or near zero implying close relatedness. Hence, distantly related groups include Kabarnet, Sergoit and Turbo based on nucleotide disparity indices. Based on the *ITS1* genes higher nucleotide disparity indices of 3.55 were observed in Gazi populations when compared to the other populations. In addition, a higher disparity index was also observed between Koriema and the other groups, indeed a disparity index of 4.32 was observed between Koriema and Cheploch groups. It is clear that disparity indices clearly delineate the coastal populations represented by Gazi group from the highland population represented by the other groups. Although Koriema belongs to the highland group, *ITS1* nucleotide substitution pattern is different from the other highland groups but also different from the coastal groups.

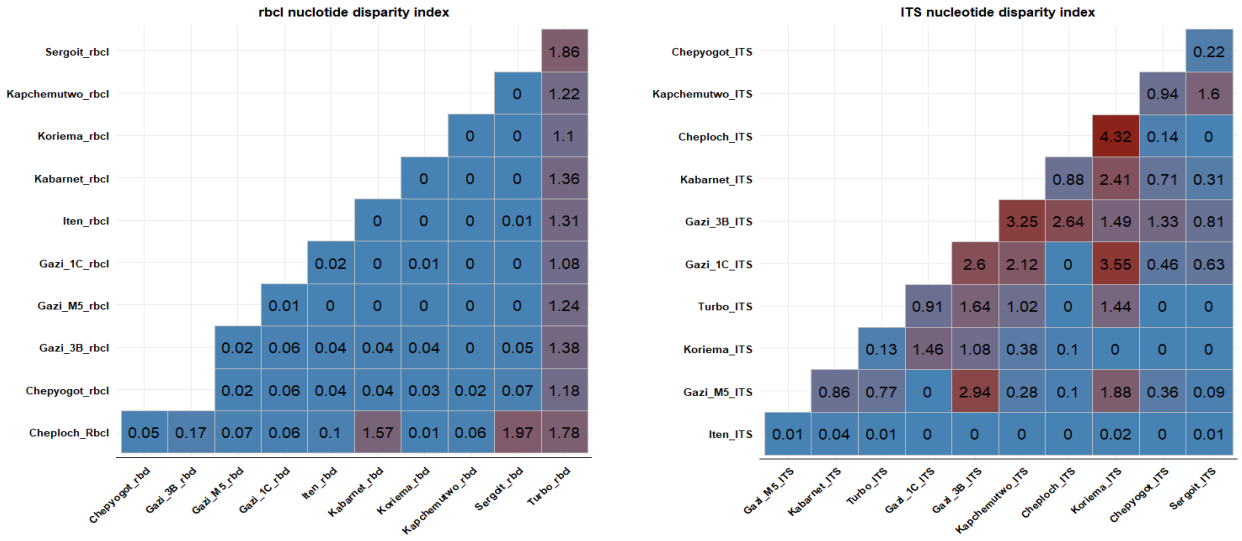

Figure 3: Nucleotide Disparity Indices

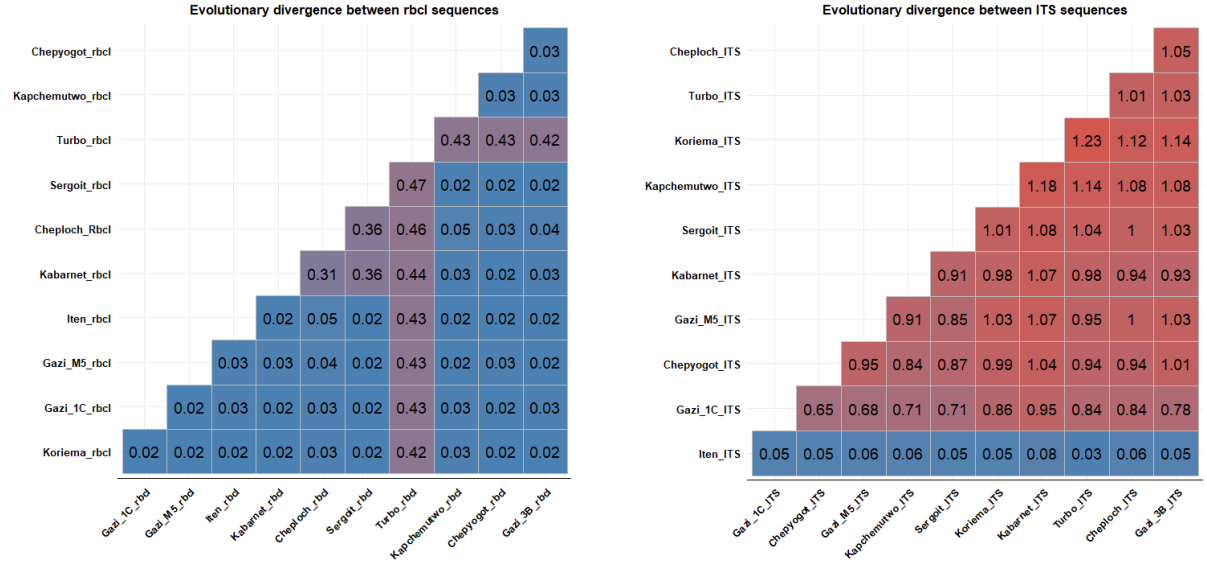

Figure 4: Evolutionary divergence between sequences

3.4. Evolutionary divergence between rbcL and ITS1 genes

In phylogenetic reconstruction and molecular evolution research, determining evolutionary divergence is crucial. The number of nucleotide substitutions between two sequences is the most accurate technique to calculate their distance from one another. Distance matrix methods such as Unweighted Pair Group Method with Arithmetic Mean (UPGMA) and Neighbor-Joining (NJ) are based on sequence divergence. Figure 4 shows an evolutionary divergence between *Dodonaea* populations based on the two genes. Higher divergence suggests more substitutions between sequences and hence larger differences between sequences. In the *rbcL* sequences, most genetic distances are close to zero implying the *Dodonaea* populations are closely related. However, slightly higher distances ranging from 0.31 to 0.47 can be seen in Turbo, Kabanet, Sergoit, and Cheploch groups when compared to the other groups. Although *rbcL* indicates close relationship among the populations, Turbo, Kabarnet, Sergoit, and Cheploch groups could have some slight genetic differences. With regard to the *ITS1* genes, the majority of the comparison gave a distance of 0.65 to 1.23, implying that there were more substitutions/differences among the *ITS1* genes compared to the *rbcL* genes. However, the Iten

ITS1 shows lower genetic distances (0.03-0.08) compared to the rest implying that this group is closely related to the groups under study.

3.4.1. Maximum likelihood Trees

ML trees are shown in Figure 5. Initial trees for the ML heuristic search were attained through the application of Neighbor Joining algorithms to pairwise edit distances generated from all the sequences from the two genes based HKY model of evolution. The final tree selected was the tree topology with the highest log likelihood value. Bootstrap resampling was selected as a test of phylogeny which checked the significance of the trees constructed (Musila *et al*., 2017). A monophyletic clade is formed by species that emanate from the same node, while species that are adjacent to one another on the tree indicate a close genetic relationship. Bootstrap values at the node indicate how well the tree topology/monophyletic clade is supported. A bootstrap value of more than 70 indicates that the branching pattern is strongly supported, a value of about 50 indicates moderate support for the tree topology, and a value less than 50 indicates poor support (Chebet *et al.*, 2022).

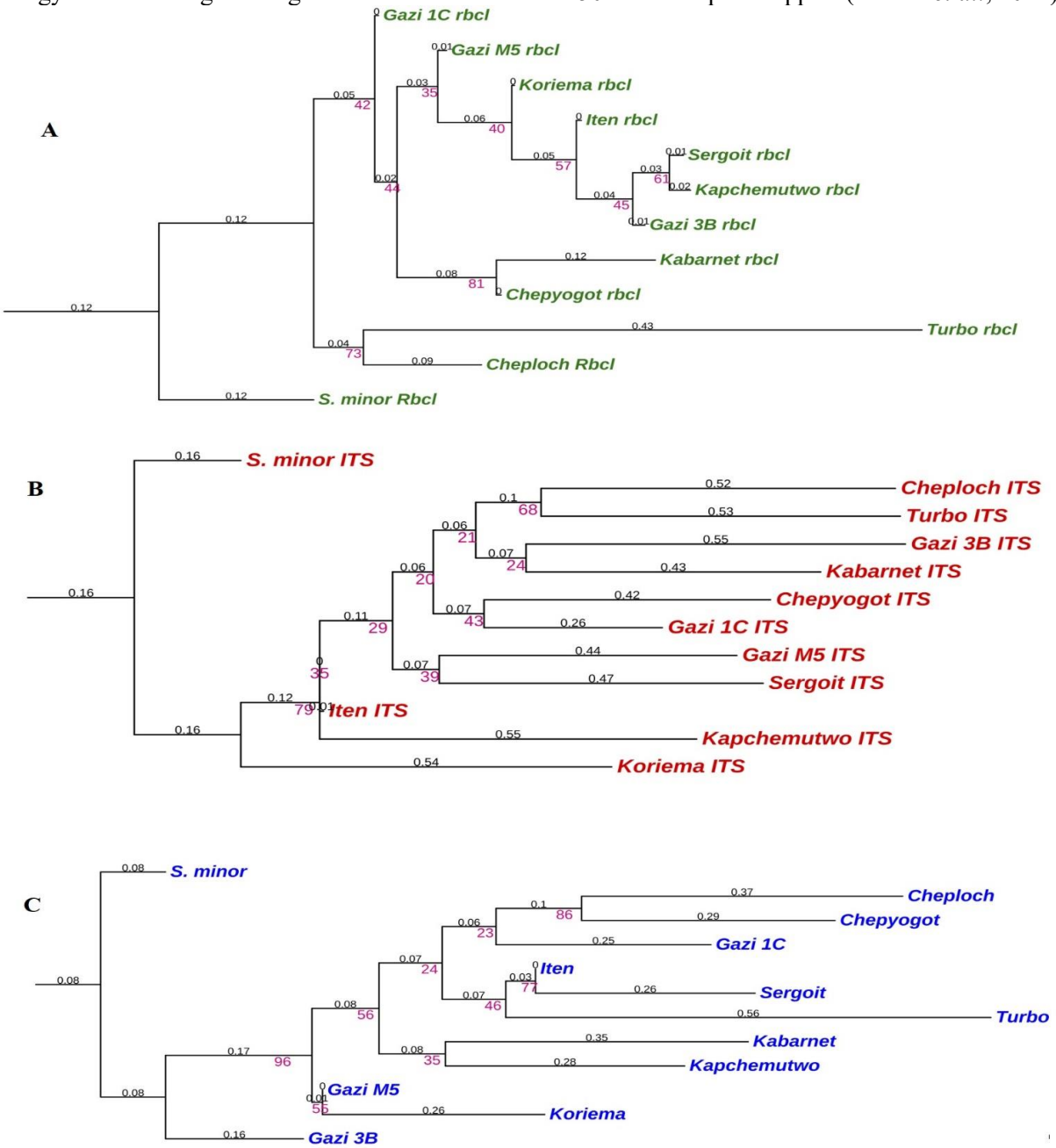

Figure 5: B; *rbcL* ML tree, **B**; *ITS1* ML tree, **C;** *rbcL* and *ITS1* concatenated ML tree

From the *rbcL* ML tree (Figure 5A), Cheploch and turbo populations are well supported with a bootstrap value of 73, implying that these two groups are closely related. Similarly close relation can be deduced between Chepyogot and Kabarnet groups with a bootstrap value of 81. The other groups are poorly supported by a bootstrap value of 42. Generally, the *rbcL* ML tree has grouped the coastal *Dodonaea* populations which are represented by the Gazi group and some highland groups in one cluster though poorly supported by a bootstrap value of 42 and another cluster well supported by a bootstrap support value of 73 containing high populations represented by Turbo and Cheploch distantly related to the coastal cluster. With regard to *ITS1* genes (Figure 5B), the ML tree shows that the study populations are monophyletic and the clade is supported by a bootstrap value of 79, only Koriema group is seen to be distantly related from the rest. A concatenated tree arose when all the *rbcL* and *ITS1* genes from each species were combined and then used in phylogeny reconstruction. From the concatenated tree (Figure 5C), monophyletic nature of the populations is also evidenced where the clade is supported by a bootstrap value of 96. Similarly, the coastal *Dodonaea* populations especially those represented by Gazi 3B and Gazi M5 have been shown to be distantly related to the *Dodonaea* populations represented by the other groups.

3.4.2. Principal component analysis

PCA led to the identification of two principal components (PC1 and PC2) among the *rbcL* and *ITS1* sequences which account for most of the variations seen in the *Dodonaea* populations. Moreover, it was also possible to identify which groups for *Dodonaea* contributed more to most of the variations (Quality of representation (Cos2) to the principal components (Figure 6). Cos2 shows that Gazi *rbcL*, Kapchemutwa *rbcL*, Iten *rbcL*, Chepyogot *rbcL,* and Koriema *rbcL* contributed to most of the variation observed in the study populations. This variation could be attributed to more genetic diversity in the *rbcL* genes.

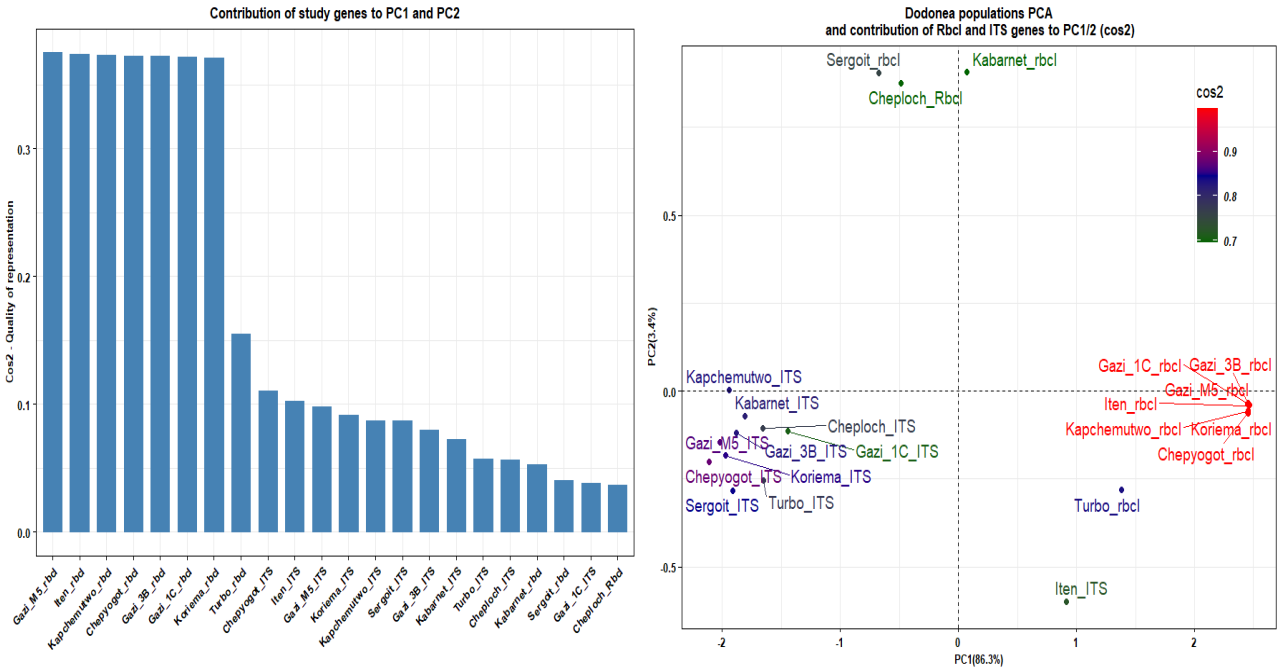

Figure 6: Quality of representation (Cos2) to PC1 and PC2

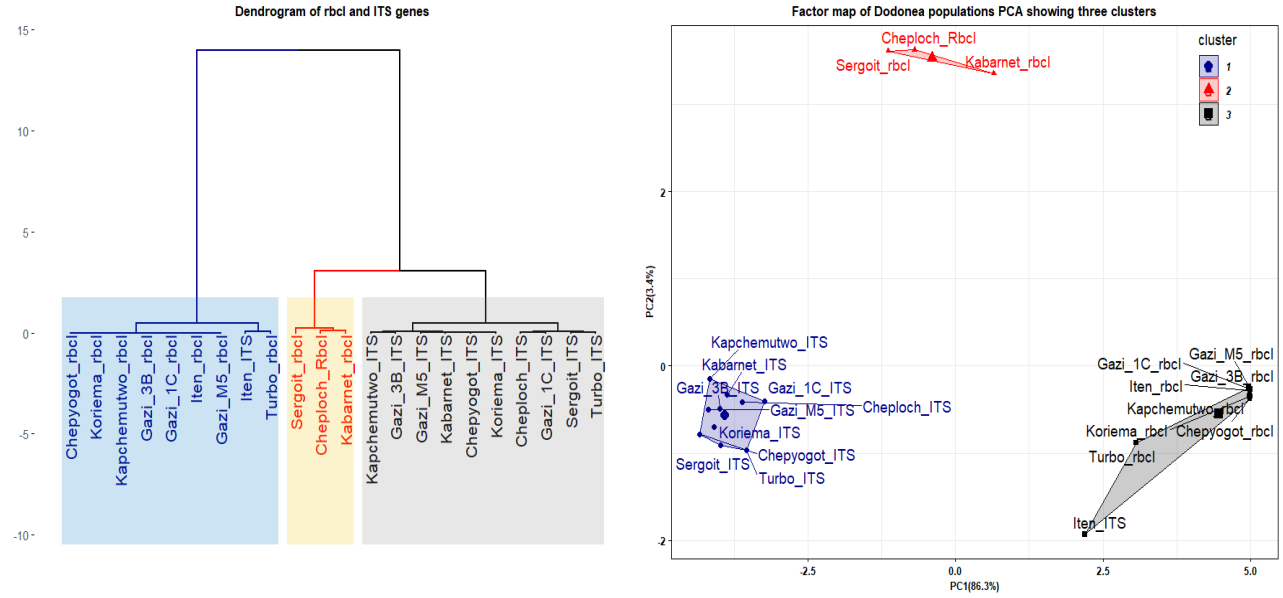

Figure 7: Dendrogram and Factor map of *Dodonaea* populations

PC1 accounted for 86.1% of the overall variation seen in *Dodonaea* populations and factoring the *rbcL* genes, PCA was able to group Sergoit, Cheploch, and Kabarnet groups in one cluster from the other *rbcL* groups. This implying Sergoit, Cheploch, and Kabarnet which represent the highland *Dodonaea* populations are distantly related to the other groups some of which represent the coastal *Dodonaea* populations which converge in a different cluster. These groupings are presented by the dendrogram in Figure 7, which shows that *rbcL* genes separates *Dodonaea* subspecies into two clusters where Sergoit, Cheploch, and Kabarnet groups have been separated from the rest of the groups based on *rbcL* genes. *ITS1* genes have low representation on the overall variation and PCA analysis of *ITS1* genes has grouped all *Dodonaea* populations in one cluster.

4. Conclusion and Recommendations

Analysis of *rbcL* sequences revealed distinct divergence in the Turbo population. *ITS1* disparity index clearly delineated populations along the coast represented by Gazi from the other highland populations. There were low nucleotide frequencies in the *rbcL* and *ITS1* sequences with nucleotide frequency rates ranging from 0.04-0.17 in *rbcL* sequences and 0.05-0.14 in *ITS1* sequences. From the evolutionary divergence between sequences, it was evident that most genetic distances were close to zero in the *rbcL* whereas in *ITS1* there were higher genetic distances. PCA separated the populations into two main clusters based on *rbcL* sequences while *ITS1* placed all populations in one cluster. ML trees shows that the all populations are monophyletic.

In conclusion, nucleotide substitution rates, base disparity indices and evolutionary distances between the *rbcL* and *ITS1* genes of Kenyan *Dodonaea* populations are very low which indicating a close evolutionary relationship between the coastal and highland populations. These findings resolves the existing taxonomic challenge on this plant with lots of ethnobotanical uses. Further studies are recommended targeting other gene regions with the use of different primers to delimit the highland and coastal populations of *D. viscosa.*

5. Acknowledgement

We wish to acknowledge the University of Eldoret and the University of Nairobi for their support in this research.

Author Contributions: The following authors contributed to this study: conceptualization, P.K.L. and C.W.L., methodology, E.W.N. and P.J., data analysis, P.K.L. and F.M.M., writing-original draft preparation P.K.L., writing- review and editing, P.K.L. and F.M.M. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: We wish to confirm that there are no conflicts of interest associated with this publication.

6. References

Al-Atiyat, R. M., & Aljumaah, R. S. (2014). Genetic distances and phylogenetic trees of different Awassi sheep populations based on DNA sequencing. *Genetics and Molecular Research*, *13*(3), 6557-6568. http://dx.doi.org/10.4238/2014.August.26.6

Algarni, A. A. (2022). Molecular identification and phylogenetic analysis of *Aloe shadensis* from Saudi Arabia based on *matK, rbcL* and *ITS* DNA barcode sequence. *Saudi Journal of Biological Sciences*, *29*(2), 1125–1133. https://doi.org/10.1016/j.sjbs.2021.09.053

Al-Snafi, A. E. (2017). A review on *Dodonaea viscosa*: A potential medicinal plant. *IOSR Journal of Pharmacy, 7*(2), 10-21.

Anilreddy, B. (2009). Preparation, characterization and biological evaluation of some overview of *Dodonaea viscosa* Linn. *J Pharm Sci Technol*, *1*(1), 1-9.

Beentje, H. (1994). Kenya trees shrubs and lianas. National Museums of Kenya, Nairobi.

Bro, R., & Smilde, A. K. (2014). Principal component analysis. *Analytical methods, 6(9),* 2812-2831. *https://doi.org/10.1039/C3AY41907J*

Buerki, S., Lowry Ii, P. P., Alvarez, N., Razafimandimbison, S. G., Küpfer, P., & Callmander, M. W. (2010). Phylogeny and circumscription of Sapindaceae revisited: Molecular sequence data, morphology and biogeography support recognition of a new family, Xanthoceraceae. *Plant Ecology and Evolution*, *143*(2), 148–159. https://doi.org/10.5091/plecevo.2010.437

Chebet, D., Musila, F. M., Kituyi, S. N., Muthike, G. M., & Kaigongi, M. M. (2022). Molecular Phylogeny of Selected Kenyan Eucalyptus Species Inferred from *MatK, rbcL* and *TrnL-F* Genes and Their Suitability for Power Transmission Poles. *Diversity*, *14*(7), 563. https://doi.org/10.3390/d14070563

Christmas, M. J., Biffin, E., & Lowe, A. J. (2015). Transcriptome sequencing, annotation and polymorphism detection in the hop bush, *Dodonaea viscosa*. *BMC genomics*, 16, 1-13. https://doi.org/10.1186/s12864-015-1987-1.

Christmas, M. J., Biffin, E., & Lowe, A. J. (2018). Measuring genome-wide genetic variation to reassess subspecies classifications in *Dodonaea viscosa* (Sapindaceae). *Australian Journal of Botany*, *66*(4), 287. https://doi.org/10.1071/BT17046

Harrington, M. (2008). Phylogeny and evolutionary history of Sapindaceae and *Dodonaea* (Doctoral dissertation, James Cook University). https://doi.org/10.25903/7jxf-1f72

Hossain, M. A. (2019). Biological and phytochemicals review of Omani medicinal plant *Dodonaea viscosa*. *Journal of King Saud University-Science*, *31*(4), 1089-1094. https://doi.org/10.1016/j.jksus.2018.09.012

Kaigongi, M. M. (2020a). Chemosystematics, Antimicrobial and Antiproliferative Activities of the Kenyan Zanthoxylum Species and *Dodonaea viscosa* Populations *(Doctoral dissertation, University of Nairobi)*.

Kaigongi, M. M., Lukhoba, C. W., Ochieng, P. J., Taylor, M., Yenesew, A., & Makunga, N. P. (2020b). LC-MS-Based metabolomics for the chemosystematics of Kenyan *Dodonaea viscosa* jacq (sapindaceae) populations. *Molecules, 25*(18), 4130. https://doi.org/10.3390/molecules25184130

Kannan, M., Senthil Kumar, T., & Rao, M. (2016). Utilization of plant resources for non-medicinal purposes by Malayali tribes of Kalrayan hills of Salem District, Tamil Nadu, India. *International Journal of Herbal Medicine, 4*(1), 47-58.

Katoh, K., Asimenos, G., & Toh, H. (2009). Multiple alignment of DNA sequences with MAFFT. *Bioinformatics for DNA sequence analysis*, 537, 39-64.

Khan, D., & Ismail, S. (2019). Fruit types, brood-size, germination and seedling morphology of hopbush [(*Dodonaea viscosa* (L.) Jacq.), Family Sapindaceae]. *International Journal of Biology and Biotehnology*, *16*, 811-833.

Kluesner, M. G., Nedveck, D. A., Lahr, W. S., Garbe, J. R., Abrahante, J. E., Webber, B. R., & Moriarity, B. S. (2018). EditR: A Method to Quantify Base Editing from Sanger Sequencing. *The CRISPR Journal*, *1*(3), 239–250. https://doi.org/10.1089/crispr.2018.0014

Konishi, T., Matsukuma, S., Fuji, H., Nakamura, D., Satou, N., & Okano, K. (2019). Principal component analysis applied directly to sequence matrix. *Scientific Reports, 9*(1), 19297. https://doi.org/10.1038/s41598-019-55253-0

Korall, P., & Kenrick, P. (2002). Phylogenetic relationships in Selaginellaceae based on *rbcL* sequences. *American Journal of Botany*, *89*(3), 506–517. https://doi.org/10.3732/ajb.89.3.506

Kumar, S., Nei, M., Dudley, J., & Tamura, K. (2008). MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in Bioinformatics*, *9*(4), 299–306. https://doi.org/10.1093/bib/bbn017

Kuswanto, L., Chusna, N. A., Purnomo, E., Krisantini, & Ahmad, M. U. (2022). Identification and documentation of wild plant species with ornamental potentials at Mount Prau, Central Java, Indonesia. *Ornamental Horticulture*, *28*(1), 110–119. https://doi.org/10.1590/2447-536X.v28i1.2418

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2017). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular biology and*

evolution, 34(3), 772-773. https://doi.org/10.1093/molbev/msw260

Lawal, D., & Yunusa, I. (2013). *Dodonea viscosa* Linn: *ITS1* medicinal, pharmacological and phytochemical properties. *International Journal of Innovation and Applied Studies, 2*(4), 476-482. http://www.issr-journals.org/ijias/

Letunic, I., & Bork, P. (2024). Interactive Tree of Life (iTOL) v6: Recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Research*, *52*(W1), W78–W82. https://doi.org/10.1093/nar/gkae268

Men, A. E., Wilson, P., Siemering, K., & Forrest, S. (2008). Sanger DNA sequencing. Next Generation Genome Sequencing: *Towards Personalized Medicine*, 1-11. https://doi.org/10.1002/9783527625130

Murad, W., Azizullah, A., Adnan, M., Tariq, A., Khan, K. U., Waheed, S., & Ahmad, A. (2013). Ethnobotanical assessment of plant resources of Banda Daud Shah, District Karak, Pakistan. *Journal of Ethnobiology and Ethnomedicine*, *9*(1), 77. http://www.ethnobiomed.com/content/9/1/77

Musila, F. M., Lukhoba, C. W., Nguta, J. M., & Dossaji, S. F. (2017). Phylogeny of ten Kenyan Plectranthus species in the Coleus clade inferred from leaf micromorphology, *rbcL* and MatK genes. *Journal of Botany, 2017*(1), 4369029. https://doi.org/10.1155/2017/4369029

Omelchenko, D. O., Krinitsina, A. A., Kasianov, A. S., Speranskaya, A. S., Chesnokova, O. V., Polevova, S. V., & Severova, E. E. (2022). Assessment of *ITS1*, *ITS2*, 5′-*ETS*, and *trnL*-F DNA Barcodes for Metabarcoding of Poaceae Pollen. *Diversity*, *14*(3), 191. https://doi.org/10.3390/d14030191

Omosa, L. K., Akala, H., Kenanda, E. O., & Ndunda, B. (2016). Variability of surface exudates of *Dodonaea angustifolia* Lf, antioxidant and antiplasmodial activities of the compounds. *Journal of Natural Sciences Research, 6*(10), *78-85.*

Orpin, J. B., Bem, A. A., & Inuwa, A. (2018). Antibacterial effect of *Dodonea viscosa* (hop bush) leaf extract in some enterobacteriaceae. *Adv Plants Agric Res, 8*(2), 213-217.

Patel, M., & Coogan, M. M. (2008). Antifungal activity of the plant *Dodonaea viscosa* var. Angustifolia on Candida albicans from HIV-infected patients. *Journal of Ethnopharmacology*, *118*(1), 173–176. https://doi.org/10.1016/j.jep.2008.03.009

Patwardhan, A., Ray, S., & Roy, A. (2014). Molecular markers in phylogenetic studies-a review. *Journal of Phylogenetics & Evolutionary Biology, 2*(2), 131. https://doi.org/10.4172/2329-9002.1000131

Priyadi, A. (2021). Species distribution models for two subspecies of *Dodonaea viscosa* (Sapindaceae) in Indonesia. *IOP Conference Series: Earth and Environmental Science*, *743*(1), 012027. https://iopscience.iop.org/article/10.1088/1755-1315/743/1/012027

Rani, M. S., Pippalla, R. S., & Mohan, K. (2009). *Dodonaea viscosa* Linn.-an overview. *Asian Journal of Pharmaceutical Research and Health Care*, *1*(1), 97-112. https://core.ac.uk/download/pdf/235570967.pdf

Rao, C. G., & Pratap, D. N. (2022). A review of the potential drug plant *Dodonaea viscose*. 7(9).

Rathod, J. P., M. Gade, R., & R. Rathod, D. (2018). Rapid Procedure for Genomic DNA Extraction from Pigeon Pea. *International Journal of Current Microbiology and Applied Sciences*, *7*(12), 863–866. https://doi.org/10.20546/ijcmas.2018.712.107

Shamso, E., & Fouad, A. (2019). *Dicliptera aegyptiaca* (Acanthaceae), A New Species from Egypt Supported by Morphological Characters and *rbcL*-based DNA Barcoding. *Egyptian Journal of Botany*, 59(2), 475-482. https://dx.doi.org/10.21608/ejbo.2019.5371.1220

Shiaka, G. P., Balarabe, M., & Adepoju, M. (2019). Antibacterial activity of *Dodoneae viscosa* leaf extract against selected enteric bacteria. *Fudma journal of sciences*, *3*(1), 434-440.

Takahashi, K., & Nei, M. (2000). Efficiencies of Fast Algorithms of Phylogenetic Inference Under the Criteria of Maximum Parsimony, Minimum Evolution, and Maximum Likelihood When a Large Number of Sequences Are Used. *Molecular Biology and Evolution*, *17*(8), 1251–1258. https://doi.org/10.1093/oxfordjournals.molbev.a026408

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, *38*(7), 3022–3027. https://doi.org/10.1093/molbev/msab120

Vehrs, H. P., & Heller, G. R. (2017). Fauna, fire, and farming: landscape formation over the past 200 years in pastoral East Pokot, Kenya. *Human Ecology*, *45*, 613-625.

West, J. G. (1984). A revision of *Dodonaea* Miller (Sapindaceae) in Australia. *Brunonia*, *7*(1), 1-194. https://doi.org/10.1071/BRU9840001

Wigrup, I. (2006). The role of indigenous knowledge in forest management. https://stud.epsilon.slu.se/12403/

Yang, L., Wang, R., Zhang, Y., Wang, Y., & Rong, H. (2023). The complete mitochondrial genome and phylogenetic analysis of *Polythlipta liquidalis* Leech, 1889 (Crambidae: Spilomelinae). *Mitochondrial DNA Part B*, *8*(10), 1063–1066. https://doi.org/10.1080/23802359.2023.2264548

Zhang, D., Gao, F., Jakovlić, I., Zou, H., Zhang, J., Li, W. X., & Wang, G. T. (2020). PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Molecular ecology resources*, *20*(1), 348-355. https://doi.org/10.1111/1755-0998.13096