Development and Utilization of Microsatellite Markers to Assess Genetic Variation Coupled with Modelling Range Shifts of *Dodonaea Viscosa* in Isolated Taita Hills and Mount Kenya Forests

Josphat K. Saina (jksaina@wbgcas.cn)
Chinese Academy of Sciences Wuhan Botanical Garden https://orcid.org/0000-0003-3371-8521
Andrew W. Gichira
Chinese Academy of Sciences Wuhan Botanical Garden
Boniface K. Ngarega
Chinese Academy of Sciences Wuhan Botanical Garden
Zhi Zhong Li
Chinese Academy of Sciences Wuhan Botanical Garden
Robert W. Gituru
Jomo Kenyatta University of Agriculture and Technology
Guang Wan Hu
Chinese Academy of Sciences Wuhan Botanical Garden
Kuo Liao (liaokuo@wbgcas.cn)
Chinese Academy of Sciences Wuhan Botanical Garden

Research Article

Keywords: Dodonaea viscosa, Genetic diversity, Habitat fragmentation, Microsatellites, Next-generation sequencing.

Posted Date: May 17th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-509475/v1

Abstract

For the protection and maintenance of fragmented and highly disturbed habitats, understanding genetic variation is essential. The Taita Hills of Kenya is the northernmost part of the Eastern Arc Mountains and has been identified as one of the top 10 biodiversity hotspots globally. The current forests in the Taita Hills have been highly fragmented over the past century. In order to appraise the influence of anthropological disturbance and fragmentation on the genetic variation of *Dodonaea viscosa* (Sapindaceae), we studied its preliminary genetic variability and population structure using newly developed microsatellite (SSR) markers, combined with ecological niche modeling analyses. We utilized the Illumina paired-end technology to sequence the *D. viscosa's* genome and developed its microsatellite markers. In total, 646,428 sequences were analyzed and 49,836 SSRs were identified from 42,638 sequences. A total of 18 primer pairs were designed to test polymorphism among 92 individuals across eight populations. The average observed heterozygosity and expected heterozygosity ranged from 0.119 to 0.982 and from 0.227 to 0.691, respectively. Analysis of molecular variance (AMOVA) revealed 78% variance within individuals and only 20% among the eight populations. According to SDM results, *D. viscosa's* suitable habitats have been gradually reducing since the last glacial maximum (LGM), and the situation will worsen under the extreme pessimist scenario of RCP 8.5. Moreover, genetic diversity was significantly greater in larger fragments. Therefore, urgent conservation management of smaller fragmented patches is necessary to protect this disturbed region and maintain the genetic resources.

Introduction

Forest fragmentation and land destruction lead to substantial alterations in habitat qualities and biochemical conditions in the remaining forest portions [1], causing adverse effects on reproductive systems, biological interactions, seed dispersal, and individual fitness in altered landscapes. The disjointed habitats interrupt gene migration among populations and lead to the emergence of small remnant and isolated populations [2–4] which have high extinction risks as a result of genetic drift, natural catastrophes, environmental and demographic stochasticity. High rates of allelic drift, reduced genetic connectivity, and high inbreeding may result from devastating effects of small and isolated populations in the discontinuous landscape [5–7]. As a result of genetic drift increase in genetic differentiation occurs, resulting in erosion of genetic variation among the populations, gradually lowering the species fitness and populations adaptability when coping with varying environmental conditions. The gradual effects of recent anthropogenic activities and habitat fragmentation drastically affects plant species' genetic composition among the isolated forest patches [7–9].

The Taita Hills (Kenya) is among the top 10 global biodiversity hotspots that has suffered a greater forest loss during the past several years, thus categorized among the most threatened locations [10], characterized by the fragmented and disturbed ecosystem. It is dominated by major valuable plant species such as *Aningeria* spp, *Strombosia* spp, and *Dodonaea viscosa*, boasting rich diversity and endemic plants and animals [11]. The major driver for the current distribution of plants in the area is commonly thought to be climatic oscillations and geographical shifts associated with the quaternary [12]. The Taita Hills region is characterized by favorable climate and fertile soils encouraging massive clearance of natural forests by local inhabitants for farming [13]. As a result, the forest remnants have undergone fragmentation and currently surrounded by cultivated farmlands for over a century [14, 13, 15]. The existing Taita Hills landscape comprises of three distinct segregates [16], Sagala Hill separated from the rest by a valley, Mbololo Hill disconnected by a valley from the last isolate Dabida Hill which contains (Ngangao, Vuria and Yale Hills). The Taita Hills forests share vital plant species despite the relatively enormous barriers between the different mountain crests, with slopes and peaks of Mount Kenya, Nyambeni, Mau, Aberdares, Cherangani Hills, Mukinduri and Tinderet Peak [17]. Historical or recent isolation may have resulted in

high variation in the Taita Hills shaping its flora when compared with other mountains [13, 15]. In recent times, the seed dispersal efficiency of different plants in these destroyed and isolated forest fragments has declined [18, 15], and these species may be facing difficulties in keeping steady populations in entire natural forest remnants. This signifies that some forest patches are at high risk of extinction [19, 20].

The East African Rift-Valley system plays a vital role in disrupting gene flow, although only limited studies have examined the genetic variation of plant species that are widely distributed [21]. In two *Acacia* species, the genetic diversity was found to be high. However, the genetic differentiation among populations was lower in *A. senegal* than *A. mellifera*, suggesting that in *A. senegal* [22], gene flow barriers have less impact on shaping gene flow. In *Prunus africana*, the East African Rift System played a major role as a gene flow barrier [21], similar to the study on *A. mellifera* in tropical Africa [23, 24]. In Kenya, *D. viscosa* is distributed from coastal parts like Taita hills (Voi) to northern areas: Rift valley, Nairobi, Nanyuki, and Kajiado [25]. This species has dense or erect, multi-stemmed shrub with variable leaf shapes and the fruit is a capsule with 3–4 wings and bears the spreading ability. Moreover, it is found in different habitats such as desert gullies, arid shrub lands and temperate woodlands [26]. Therefore, morphological differences and ecological valence in this crucial species make it a perfect model to examine its adaptability to the surrounding conditions and divergence [27, 26].

Our current knowledge of the genetic effects of forest fragmentation and land destruction in Kenya's isolated forests is limited. The Taita Hills being a top 10 global biodiversity hotspot area and center of species endemism offers a unique opportunity to study how climatic oscillations influence patterns of differentiation and diversity. Furthermore, *Dodonaea viscosa* species yielded large amounts of the most biomass nitrogen in this region, therefore can be use in reforestation, reclamation of marshy and degraded lands, and as a soil stabilizer. Thus we utilized *D. viscosa* to fill this knowledge gap. Besides, due to the absence of reliable molecular markers in *D. viscosa*, only its genomic resources are available in public database, we firstly developed the SSR markers for this species using Next-generation sequencing (NGS) technology. Secondly, we used some of these newly developed polymorphic loci to test the genetic variation of *D. viscosa* in the remaining forest patches. Lastly, we inferred the potential global climate change influence on the distribution of *D. viscosa* in the Last Glacial Maximum (LGM), present and the future. Our results will help understand adaptive variation, population structure, gene flow, genetic diversity, and potential refugia in the LGM and provide information for designing effective conservation measures in various fragmented mountain massifs found in east Africa.

Materials And Methods

Plant sampling

A total of 92 individuals were sampled from eight populations in two different distribution zones in Kenya, including Taita hills in Voi and Mt. Kenya (Table 1; Fig. 1). Global positioning system (GPS) was used to record the geographical locations of each population. Leaf samples were collected randomly in the field, dried in silica gel, and later transported to the Chinese Academy of Sciences (Wuhan, China). A copy of voucher specimens were deposited in the National museums of Kenya and Herbarium of Wuhan Botanical Garden.

DNA isolation, library preparation, sequencing, and primer design

DNA was extracted from silica gel dried leaf material of *D. viscosa* using the cetyltrimethylammonium bromide (CTAB) method [28]. The Qubit DNA Assay kit the Qubit 2.0 Fluorometer (Life Technologies, San Diego, CA) was used to verify the purity and concentration of the DNA. The quality and concentration of the DNA were checked using

Qubit DNA Assay kit in Qubit 2.0 Fluorometer (Life Technologies, San Diego, CA). Three samples with high-quality DNA were selected for Illumina paired-end library construction. About 4 ug DNA per sample were sheared using a Covaris S220 (Massachusetts, USA) to generate > 750 bp fragments, following the Illumina TruSeq DNA Library preparation Kit (Illumina, California, USA) and using a multiplex identifier index. Library sequencing with 150 bp paired-end reads was conducted using Illumina HiSeq 2000 Platform at Novogene Bioinformatics Technology Co. Ltd (Beijing, China). The generated raw reads were subjected to a stringent filtering process. The following read sequences were discarded prior to assembly; reads with base pairs exceeding 10% with low-quality score Q < 20, ambiguous paired reads with N content exceeding 10 % and other contaminants. De novo assembly was conducted using the de Bruijin graph-based Velvet 2.0 [29] software. MIcroSAtellite identification tool (MISA; http://pgrc.ipk-gatersleben.de/misa/) was used to scrutinize the generated sequences for the availability of Simple Sequence repeats. Minimum of 10, 5, 4, iterations were considered for Mono, Di- and Trinucleotides, respectively, whereas a minimum of 3 iterations was considered for both Penta and Hexa-nucleotides. We used Primer3 [30] software at default settings to design primers.

Primer validation

Using genomic DNA isolated from six individuals, a total of 25 SSR primers were chosen at random and tested for amplification and specificity. Polymerase Chain Reaction (PCR) was performed in a 20 µL reaction mixture containing; 11.3 µL ddH_2O, 0.2 µL Taq Polymerase enzyme, 0.5 µL dNTPs, 1 µL Forward primer and 1 µL Reverse primer, 2.5 µL Taq Buffer, and 2 µL of Genomic DNA. PCR amplification conditions were as follows: an initial denaturation step of 4 minutes at 94°C, followed by 30 cycles of denaturation for 40 sec at 94°C, annealing for 30 s at 52°C – 54°C, and extension of 40 s at 72°C; then a final extension step of 10 min at 72°C. 2 % Agarose gel was used to quantify the PCR products. We selected 18 primers that produced clearly defined bands, labeled the 5’ end of the forward sequence with the 6-FAM fluorescent dye, and then performed genotyping using those markers. The STR sequences were analyzed using GeneMarker software (Soft Genetics).

Data analysis

The deviations from Hardy-Weinberg equilibrium (HWE) with Bonferroni corrections and the polymorphic information content (PIC) for each locus and were checked in Cervus 3.0 [31]. For each SSR locus analyzed, the average number of alleles per locus N_a, number of effective alleles N_e, observed heterozygosity (H_o), expected heterozygosity (H_e), Shannon’s diversity index (I), inbreeding coefficient (F_{IS}) was calculated using GenAlEx v6.5 [32].

STRUCTURE v2.3.4 [33] based on the Bayesian approach, was run with different (k) values from 1 to 8 to investigate the genetic structure. We performed 10 independent runs per k value, with a burn-in period length of 100,000 and 500,000 replicates of Monte Chain Markov Carlo (MCMC). The iterations were carried out with the correlated allele frequency and admixture model. Optimal value of k was determined by examining the Δk statistic proposed by Evanno [34]. Basing on the Greedy algorithm, 10,000 repeats were independently run using CLUMP v.1.1.4 [35] to average all data sets and determine the value of k. The graphical illustrations were examine using DISTRUCT version 1.1 [36]. Nei's genetic distance [37] matrix was applied to estimate the genetic relationships between populations using the Molecular Evolutionary Genetics Analysis (MEGA) v.6.06 program [38] using Unweighted Pair Group Method with Arithmetic Mean (UPGMA). To understand more about population relationships and quantify the genetic variation, Principal Coordinate Analysis (PCoA), and analysis of molecular variance (AMOVA) were executed in GenAlEx v.6.5 [32].

Lastly, bottleneck events like populations decrease, heterozygosity excess were detected using Bottleneck v.1.2.02 [39]. A two-phase model (TPM) was set up using 70% SMM and 30% IAM) steps set as recommended [39] with10,

000 simulation iterations under the infinite alleles model (IAM), the stepwise mutation model (SMM) and Wilcoxon signed-rank test.

Species distribution modeling

MAXENT v.3.3.3 was used to perform species distribution modeling of *D. viscosa* to predict its potential distribution in East Africa [40]. We obtained 19 bioclimatic variables in a resolution of 2.5 arcsec (ca. 5-km2) for the current and past (Last Glacial Maximum, LGM) scenarios from WorldClim1.4 (http://www.worldclim.org/, [41]). An additional elevation raster was also obtained from the Worldclim database and included as an environmental variable. Since collinearity may cause uncertainties and lower the statistical power of models [42, 43], a spatial correlation test was implemented on the bioclimatic variables. The test was performed to remove one variable in a pair of correlated variables at a threshold of 0.8 [44], and was implemented in ENMTools package in R, using the function *raster.cor.matrix* [45]. Eventually, seven variables were selected as representative of climate factors (Table 7).

To assess the probable future distribution in *D. viscosa* ranges, we utilized two representative concentration pathway (RCP) scenarios from the Community Climate System Model see 4 (CCSM4) model [46], For the pessimistic scenario, RCP8.5, the increase in global mean surface temperatures will likely be 2.6–4.8°C, while the concentration of CO2 will approximately be 1,350 ppm by 2100. Under the intermediate scenarios, the global mean surface temperatures are projected to rise by 1.5–3.2°C, and the CO2 concentrations to 850 ppm in RCP 4.5 [47]. The model performance was evaluated by the area under the curve (AUC) values of the receiving operator curve (ROC) [48].

The models of the LGM, present, and future distribution were created using 67 locations obtained from the Global Biodiversity Information Facility (GBIF) and 8 actual occurrences from our field study that were linked with an accurate GPS mapping. Maxent models were run with default parameters except the 75% random test percentage, 5000 background iterations, and 10 subsample replicates. We selected the logistic output format for the three periods, current, LGM, and future, and using this format, a map of continuous probability values ranging from 0 to 1 was generated. ArcMap v10.5 (ESRI, Redlands, CA) was used to visualize the output logistic predictions.

Results

Simple sequence repeats in the *Dodonaea viscosa*

To evaluate the assembly quality and develop new molecular markers, the generated 646,428 sequences were examined for potential SSRs. A total of 42,638 sequences were found to contain 49,836 microsatellites. 4,927 sequences contained multiple SSRs. Among the 49,836 SSRs consisting various repeat types (Table 2), mononucleotide repeats were the most abundant accounting for 40.77% (20,320) followed by Di-, 31.39 % (15,644), and tri-, 25.98 % (12,949), while the rest, tetra, penta- and hexa-nucleotide repeat units accounted for 1.14 % (568), 0.21 % (102) and 0.51 % (253) of all the SSRs respectively. Most SSRs repeat units ranged from 5 to 13, with five repeat units 14.77 % (7,363) six repeat units 18.88 % (9,407) and ten repeat units 27.70 % (13,804) as the most repeat types (Table 2). In the di-nucleotide repeat SSRs, AG/CT was the most abundant 49.53 % (7,749), followed by AT/AT 30% (4,693) and AC/GT 20.37% (3,186). Di-nucleotide GC/CG rich repeats were extremely scarce accounting for only 0.1% (16) (Fig. 2A). AAG/CTT 47.14% (6,104) was the most frequent followed by ATC/ATG 15.04% (1948) and AAT/ATT 9.84% (1,274) of the Tri-nucleotide repeats (Fig. 2B). ACAT/ATGT and AAAT/ATTT accounted for 32.04% (182) and 27.82 (158) of all the tetra repeats (Fig. 2C). Both the penta and Hexa-nucleotide repeat motifs had low abundance of 0.21% (102) and 0.51% (253) of all SSRs, respectively.

Microsatellites Validation, polymorphism assessment and genetic diversity and structure

Eighteen primers were used to screen for polymorphism among 92 individuals (Table 3). Of the 18 loci, 12 displayed polymorphism with the allele number varying from 2 to 11 (average of 3.31) and six yielded monomorphic products. Taking no account of the six monomorphic markers and two markers with low PIC values, ten markers were used to assess genetic diversity (Table 4). The 10 SSR markers successfully produced clear amplification products generating a total of 295 alleles across 92 individuals of eight *D. viscosa* populations. Among the 10 loci, the number of observed alleles (N_a) per locus ranged from 2.250 to 5.250 with a mean value of 3.688. Dodo 001 and Dodo 026 loci had the lowest mean values of (N_a) of 2.250 and highest at locus Dodo 032. The mean effective alleles per locus had an average of 2.420 alleles per SSR, ranging from 1.360 to 3.564. The mean value of Shannon index ranged from 0.402 to 1.375, with an average of 0.929. The average expected heterozygosity and observed heterozygosity were 0.519 and 0.508 respectively. Average H_o values varied from Dodo 001 (0.119) to Dodo 004 (0.982) and H_e ranged from Dodo 001 (0.227) to Dodo 032 (0.691) (Table 4).

Of the ten polymorphic loci, Dodo 001 had the highest index of genetic variation (F_{ST} = 0.528), whereas Dodo 004 had the lowest (F_{ST} = 0.109), with the overall F_{ST} value of 0.221. The locus with the highest gene flow was Dodo 026 (N_m = 2.200), lowest in Dodo 001 (N_m = 0.223), and the average gene flow of the 10 loci was 1.174 (Table 4). The average PIC value was 0.605 ranging from 0.383 to 0.805. Dodo 020, Dodo 032, Dodo 010, and Dodo 004 SSRs were the most useful with high PIC values of 0.805, 0.799, 0.741, and 0.718 respectively. Significant deviations from the HWE were observed at all the other loci (P< 0.001) except for Dodo 002, Dodo 026, and Dodo 032.

Variation parameters between the populations were presented in (Table 1). The number of alleles per locus (N_a) varied between 2.700 (Vuria 1 population) to 4.600 (Yale population) averaging 3.69 alleles per population. The mean number of effective alleles per locus (N_e) for all populations ranged from 1.973 to 2.806 with Yale population (N_e = 2.806) showing the highest number of effective alleles and lowest in Vuria 1 population (N_e = 1.973). The mean observed (H_o) and expected (H_e) heterozygosities for all populations were 0.519 and 0.508, respectively, Observed heterozygosity (H_o) ranged from 0.440 (Ngangao 2 population) to 0.578 (Mbololo population) while the expected heterozygosity values (H_e) values varied between 0.443 (Vuria 1 population) to 0.582 (Vuria 2 population). Significant inbreeding coefficients (F_{IS}) were detected in all the populations. Varying from 0.747 to 1.105.

The AMOVA (Table 5) of eight populations showed that 78% of total variance was found within the populations while 20% was found among the populations and only 2% among the individuals. STRUCTURE analysis identified three groups among the eight populations. (Fig. 3A, Δk = 3). The two populations from Mt. Kenya (Mt. Kenya 1 and Mt. Kenya 2) and population (Mbololo) showed a close genetic similarity forming one cluster, Ngangao 1, Vuria 1 and Yale populations clustered together, while some individuals in Vuria 2 and Ngangao 2 formed the third cluster. Ngangao 2 population showed a high level of admixture, Vuria 2 and Yale also had some of individuals in different clusters. Population genetic distance cluster analysis based on Unweighted Pair Group Method with Arithmetic Mean (UPGMA) by MEGA software analysis revealed a congruent result with STRUCTURE analysis (Fig. 3C). PCoA supported STRUCTURE and UPGMA cluster analyses indicating that these populations could be separated into three groups (data not shown).

Bottleneck tests showed that three populations, Ngangao 1, Vuria 1 and Mt. Kenya 2 displayed shifted modes (Table 6.). Vuria 1 population revealed significant bottleneck IAM (P < 0.05) However, TPM and SMM (P < 0.05) on Wilcoxon

signed rank test displayed no significant population bottlenecks.

Species distribution modeling

Maxent models revealed high levels of predictive performance under all scenarios (i.e. Current = 0.905, LGM = 0.889, RCP 4.5 = 0.923, RCP 8.5 = 0.921) (Table 7). The most important variables limiting the distribution of *D. viscosa* during the LGM were: elevation (relative contribution: 42.7%), temperature seasonality (Bio4 - relative contribution: 17.8%), mean temperature of the wettest quarter (Bio8 - relative contribution: 12.0%), and precipitation of driest month (Bio14 - relative contribution: 11.2%). Besides, for the current and future models, elevation, Bio4, Bio8, and Bio18 (precipitation of warmest quarter), were the most important factors limiting the distribution of *D. viscosa* (Table 8). The past and future range simulations were fairly consistent with the current potential distribution and covered areas in Kenya and Tanzania, with central and western Kenya, northern and southwestern Tanzania having the most favorable conditions (Fig. 4).Similar high habitat suitability was also observed in Kenya and Tanzania along the Indian Ocean Coast. The LGM conditions observed high population expansion and a significant increase in high suitability areas for *D. viscosa* compared to the future. Finally, future climatic projections for *D. viscosa* revealed that the pessimistic scenario RCP85 was more optimistic and highlighted the potential range expansion by 2070 compared to the intermediate pessimistic scenario RCP4.5. High suitable ranges in Kenya and Tanzania were significantly reduced during RCP 4.5 than RCP8.5. Precisely, areas in Southern Tanzania and Eastern Uganda revealed suitable habitats during the pessimistic scenario RCP8.5.

Discussion

In this study, frequencies of perfect SSRs composing of 1–6 bp long motifs were calculated. A total of 49,836 microsatellites with an average density of 280.24 SSRs/Mb sequence data were identified. The SSR density detected in this study is nearly similar with Jujube (321.56 SSRs/Mb) [49]. Plants exhibit variation of SSRs Repeat units within their genomes, for example, in *Paleosuchus trigonatus*, di-nucleotide repeat motif predominated [50], whereas tetra-nucleotide repeats were the most frequent in *Coffea canephora* [51], and tri-nucleotide repeat motifs being most frequent in *Hyalessa fuscata* [52]. The mono-nucleotide repeat motifs predominated in our study, which is similar to *Alibertia edulis* [53]. Among the di-nucleotide repeats, AG/CT was the most abundant, followed by AT/AT and AC/GT while GC/CG repeats were extremely rare in all the Di-nucleotide repeats. AAG/CTT was the most frequent, followed by ATC/ATG and AAT/ATT of the Tri-nucleotide repeats. In addition, microsatellites in *D. viscosa* exhibited AG/CT preference.

The sampled natural populations of *D. viscos*a exhibited high genetic diversity (H_o = 0.519, I = 0.929). In this study, the Mbololo population located in the Northeast of Voi (Kenya) showed the highest genetic diversity (H_o = 0.578) while the Ngangao 2 population located in the North West of Voi Kenya, showed the lowest genetic diversity (H_o = 0.440). Our study species' genetic diversity was much higher than the previously recorded in *Blighia sapida* (Sapindaceae) using SSR markers [54]. Of the 10 SSR loci, four of them had a negative inbreeding coefficient. Furthermore, Dodo 002, Dodo 026 and Dodo 032 deviated significantly from Hardy-Weinberg equilibrium, suggesting excess heterozygosity. High variation was found within populations (78%) compared to the variation among populations (20%) as revealed by AMOVA, showing that individuals within populations are genetically different, hence in accordance with the life history of this species. Our results were congruent with the *Aesculus* (Sapindaceae), which showed high variation within populations (86.06–89.90%) among populations 7.69- 13% [55].

In the present study, the value of differentiation index F_{ST} (0.221) showed a moderate genetic differentiation, while the 20% variation between populations revealed by AMOVA, could have resulted due to high gene flow between *D. viscosa* populations (Nm = 1.174). This result indicates that this species' seeds could have been transported from one point to another by either wind or animals, hence reducing the genetic differentiation among the populations. Principle Coordinates Analysis (PCoA) grouped together Mt. Kenya 1 & 2, Ngangao 2 and Mbololo individuals, and the second and third indiscriminately grouping the remaining samples into distinct genetic clusters. These results suggest that individuals from Mt. Kenya and Mbololo share many genetic units that differentiate them from other populations.

The UPGMA analysis on genetic distance placed the eight populations into two main groups, leaving one population ungrouped. We expected the individuals from the same sites to group together, however, Ngangao 2 and Mbololo in Taita hills clustered with Mt. Kenya populations while Ngangao1, Vuria1, Vuria2 and Yale were grouped into a distinct cluster. Mount Kenya 1 did not cluster with any other population, showing a more distant relationship to the other populations. This suggests that Mount Kenya 1 could be the Centre of diversity for this species in the region due to increased genetic differentiation, however, more data is needed to confirm this assumption. STRUCTURE analysis placed all the eight populations into three groups (k = 3), with individuals from Mt. Kenya, Mbololo and Ngangao 2 population clustering together, while the remaining populations were found in cluster 2 and 3. Vuria 2 population had nearly equal membership in the two major clusters 1 & 2 and were assigned to multiple parental clusters, while Ngangao 2 population showed mixed proportions with a larger proportion of individuals clustering with Mbololo and Mt. Kenya populations (Fig. 3B).

Rich genetic diversity was seen in the Taita hills and Mt. Kenya populations, especially in larger fragmented localities as projected by conservation genetics theory [56]. Our finding of high genetic diversity in *D. viscosa* populations is comparable with studies of other plant species with small populations and restricted distribution [57, 58] or the one from fragmented and excessively exploited medicinal shrub *plumbago* [59]. Mbololo being the largest fragment (220 ha) has the best preserved vegetation and richest flora, whereas Ngangao, which is the second largest (206 ha), has only 120 ha occupied by native forest while the remaining parts are introduced plantations and rocky surfaces [13, 60]. Vuria, a fragment located on a steep ridge, suffered massive land clearance and natural forest fires some years back, while Yale fragment has 16 ha of natural forest. This finding is congruent with [61, 7] studies which revealed that decline in population sizes after forest fragmentation might reduce the genetic diversity gradually after several generations. Similarly, based on fragment size and isolation study done by [61], larger forest remnants had significantly richer genetic diversity than smaller, more disturbed and isolated hills for the overall data set. This is clearly evident, as shown by Mt Kenya and Mbololo populations revealing high diversity. Moreover, our finding supported the previous studies that Mbololo forest has high diversity richness as a result of low level disturbance compared with Ngangao forest [16, 15] and, this might be the reason as to why the Ngangao 2 forest and Vuria 2 were assigned multiple clusters. With the severe deforestation and fragmentation processes, the signs of bottlenecks were detected in more disturbed and fragmented forest patches, suggesting that *D. viscosa* experienced a recent bottleneck.

The SDM analyses inferred that *D. viscosa* had expanded its ranges during the LGM. During this period, Africa observed relatively cold and dry climatic conditions that may have persisted until the end of LGM [62]. As a result of *D. viscosa's* ability to tolerate different habitats and spread quickly [26], the species likely extended its ranges and population extents throughout this period. Conversely, the species ranges are observed to reduce in future scenarios. The anticipated future global warming will likely alter the persistence of different biodiversity [63].

Conclusions

Based on Illumina paired-end sequences, we validated eighteen microsatellite markers for this species, of which twelve were highly polymorphic with an average PIC value of 0.605 and an average allele number of 2.420 comparable to that of white spruce and black spruce [64]. Despite forest fragmentation, high genetic diversity is still maintained in populations with large patches compared to those with reduced forest areas. These results suggest that protection of smaller and isolated fragments should be put in place to avoid further destruction of existing habitats and more emphasis on conservation awareness among local residents is of great urgency. Lastly, our new findings from the SDMs of *D. viscosa* could throw light on species with similar geographical distributions and other taxa in this region which are threatened or endangered.

Declarations

Acknowledgment

This research was funded by the Sino-Africa Joint Research Center (Nos.Y323771W07, SAJC201322).

Author Contributions

Kuo Liao conceived and designed the experiment, Josphat K. Saina performed the experiments, analyzed data and wrote the first draft of the manuscript. Zhi-Zhong Li, Andrew W. Gichira and Boniface K. Ngarega assisted in data analyses. Andrew W. Gichira, Boniface K. Ngarega, Zhi-Zhong Li, and Robert W. Gituru commented on previous versions of the manuscript. Josphat K. Saina and Guangwan Hu collected the plant materials. All authors read and approved the final version of the manuscript.

Compliance with ethical standards

Conflict of interest

The authors declare that they have no conflict of interest

References

1. Wilson MC, Chen XY, Corlett RT, Didham RK, Ding P, Holt RD, Holyoak M, Hu G, Hughes AC, Jiang L, Laurance WF, Liu JJ, Pimm SL, Robinson SK, Russo SE, Si XF, Wilcove DS, Wu JG, Yu MJ (2016) Habitat fragmentation and biodiversity conservation: key findings and future challenges. Landscape Ecol 31:219-227. doi:10.1007/s10980-015-0312-3
2. Honnay O, Jacquemyn H, Bossuyt B, Hermy M (2005) Forest fragmentation effects on patch occupancy and population viability of herbaceous plant species. New Phytol 166:723-736. doi:10.1111/j.1469-8137.2005.01352.x
3. Ouborg NJ, Vergeer P, Mix C (2006) The rough edges of the conservation genetics paradigm for plants. J Ecol 94:1233-1248. doi:10.1111/j.1365-2745.2006.01167.x
4. Yineger H, Schmidt DJ, Hughes JM (2014) Genetic structuring of remnant forest patches in an endangered medicinal tree in North-western Ethiopia. BMC Genet 15. doi:10.1186/1471-2156-15-31
5. Angeloni F, Ouborg NJ, Leimu R (2011) Meta-analysis on the association of population size and life history with inbreeding depression in plants. Biol Conserv 144:35-43. doi:10.1016/j.biocon.2010.08.016

6. Schlaepfer DR, Braschler B, Rusterholz H-P, Baur B (2018) Genetic effects of anthropogenic habitat fragmentation on remnant animal and plant populations: a meta-analysis. Ecosphere 9. doi:10.1002/ecs2.2488
7. Lowe AJ, Boshier D, Ward M, Bacles CFE, Navarro C (2005) Genetic resource impacts of habitat loss and degradation, reconciling empirical evidence and predicted theory for neotropical trees. Heredity 95:255-273. doi:10.1038/sj.hdy.6800725
8. Leonardi S, Piovani P, Scalfi M, Piotti A, Giannini R, Menozzi P (2012) Effect of Habitat Fragmentation on the Genetic Diversity and Structure of Peripheral Populations of Beech in Central Italy. J Hered 103:408-417. doi:10.1093/jhered/ess004
9. Dool SE, Puechmaille SJ, Kelleher C, McAney K, Teeling EC (2016) The effects of human-mediated habitat fragmentation on a sedentary woodland-associated species (*Rhinolophus hipposideros*) at its range margin. Acta Chiropterologica 18:377-393. doi:10.3161/15081109acc2016.18.2.006
10. Lovett J (1993) Eastern Arc moist forest flora. In: Biogeography and ecology of the rain forests of eastern Africa. Cambridge University Press,
11. Beentje H, Ndiang'ui N (1988) An ecological and floristic study of the forests of the Taita Hills, Kenya. UTAFITI, Occas Pap Nation Mus Kenya 1:23-66
12. Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences 359:183-195
13. Pellikka PK, Lötjönen M, Siljander M, Lens L (2009) Airborne remote sensing of spatiotemporal change (1955–2004) in indigenous and exotic forest cover in the Taita Hills, Kenya. International Journal of Applied Earth Observation and Geoinformation 11:221-232
14. Newmark W (1998) Forest area, fragmentation, and loss in the Eastern Arc Mountains: implications for the conservation of biological diversity. Journal of East African Natural History 87:29-36
15. Aerts R, Thijs KW, Lehouck V, Beentje H, Bytebier B, Matthysen E, Gulinck H, Lens L, Muys B (2011) Woody plant communities of isolated Afromontane cloud forests in Taita Hills, Kenya. Plant Ecol 212:639-649
16. Brooks T, Lens L, Barnes J, Barnes R, Kihuria JK, Wilder C (1998) The conservation status of the forest birds of the Taita Hills, Kenya. Bird Conservation International 8:119-139
17. Beentje H, Ihlenfeldt H (1990) The forests of Kenya. Mitteilungen aus dem Institut für Allgemeine Botanik Hamburg 23:265-286
18. Cordeiro NJ, Ndangalasi HJ, McEntee JP, Howe HF (2009) Disperser limitation and recruitment of an endemic African tree in a fragmented landscape. Ecology 90:1030-1041
19. Tilman D, May RM, Lehman CL, Nowak MA (1994) Habitat destruction and the extinction debt.
20. Wu X, Ruhsam M, Wen Y, Thomas P, Worth JRP, Lin X, Wang M, Li X, Chen L, Lamxay V, Nam Le C, Coffman G (2020) The last primary forests of the Tertiary relict *Glyptostrobus pensilis* contain the highest genetic diversity. Forestry 93:359-375. doi:10.1093/forestry/cpz063
21. Kadu C, Schueler S, Konrad H, Muluvi G, EYOG-MATIG O, Muchugi A, Williams V, Ramamonjisoa L, Kapinga C, Foahom B (2011) Phylogeography of the Afromontane *Prunus africana* reveals a former migration corridor between East and West African highlands. Molecular Ecology 20:165-178
22. Omondi SF, Kireger E, Dangasuk OG, Chikamai B, Odee DW, Cavers S, Khasa DP (2010) Genetic diversity and population structure of *Acacia senegal* (L) Willd. in Kenya. Tropical Plant Biology 3:59-70
23. Davies F, Verdcourt B (1998) Flora of Tropical East Africa: Sapindaceae. Rotterdam: AA Balkema 108p-illus ISBN 471979210

24. Ruiz Guajardo J, Schnabel A, Ennos R, Preuss S, OTERO-ARNAIZ A, Stone G (2010) Landscape genetics of the key African acacia species *Senegalia mellifera* (Vahl)–the importance of the Kenyan Rift Valley. Molecular Ecology 19:5126-5139
25. Beentje H, Adamson J, Bhanderi D (1994) Kenya trees, shrubs, and lianas. National Museums of Kenya,
26. Christmas MJ, Biffin E, Lowe AJ (2015) Transcriptome sequencing, annotation and polymorphism detection in the hop bush, *Dodonaea viscosa*. BMC Genomics 16:1
27. West J (1984) A revision of *Dodonaea* Miller (Sapindaceae) in Australia. Aust Syst Bot 7:1-194
28. Doyle J (1991) DNA protocols for plants. In: Molecular techniques in taxonomy. Springer, pp 283-293
29. Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome research 18:821-829
30. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3—new capabilities and interfaces. Nucleic acids research 40:e115-e115
31. Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16:1099-1106
32. Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular ecology notes 6:288-295
33. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945-959
34. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611-2620
35. Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801-1806
36. Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. Mol Ecol Notes 4:137-138
37. Nei M (1972) Genetic distance between populations. American naturalist:283-292
38. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular biology and evolution 30:2725-2729
39. Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001-2014
40. Phillips SJ, Dudík M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31:161-175
41. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology: A Journal of the Royal Meteorological Society 25:1965-1978
42. Dormann CF, McPherson JM, Araujo MB, Bivand R, Bolliger J, Carl G, Davies RG, Hirzel A, Jetz W, Kissling WD, Kuehn I, Ohlemueller R, Peres-Neto PR, Reineking B, Schroeder B, Schurr FM, Wilson R (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30:609-628. doi:10.1111/j.2007.0906-7590.05171.x
43. de Marco Junior P, Nobrega CC (2018) Evaluating collinearity effects on species distribution models: An approach based on virtual species simulation. Plos One 13. doi:10.1371/journal.pone.0202403

44. Gormley AM, Forsyth DM, Griffioen P, Lindeman M, Ramsey DSL, Scroggie MP, Woodford L (2011) Using presence-only and presence-absence data to estimate the current and potential distributions of established invasive species. J Appl Ecol 48:25-34. doi:10.1111/j.1365-2664.2010.01911.x
45. Warren DL, Matzke NJ, Cardillo M, Baumgartner JB, Beaumont LJ, Turelli M, Glor RE, Huron NA, Simoes M, Iglesias TL, Piquet JC, Dinnage R (2021) ENMTools 1.0: an R package for comparative ecological biogeography. Ecography 44:504-511. doi:10.1111/ecog.05485
46. Gent PR, Danabasoglu G, Donner LJ, Holland MM, Hunke EC, Jayne SR, Lawrence DM, Neale RB, Rasch PJ, Vertenstein M, Worley PH, Yang ZL, Zhang MH (2011) The Community Climate System Model Version 4. J Clim 24:4973-4991. doi:10.1175/2011jcli4083.1
47. Collins M, Knutti R, Arblaster J, Dufresne J-L, Fichefet T, Friedlingstein P, Gao X, Gutowski WJ, Johns T, Krinner G (2013) Long-term climate change: projections, commitments and irreversibility. In: Climate Change 2013-The Physical Science Basis: Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, pp 1029-1136
48. Mas J-F, Soares Filho B, Pontius RG, Jr., Farfan Gutierrez M, Rodrigues H (2013) A Suite of Tools for ROC Analysis of Spatial Models. Isprs International Journal of Geo-Information 2:869-887. doi:10.3390/ijgi2030869
49. Fu P-c, Zhang Y-Z, Ya H-y, Gao Q-b (2016) Characterization of SSR genomic abundance and identification of SSR markers for population genetics in Chinese jujube (*Ziziphus jujuba* Mill.). PeerJ 4:e1735
50. Muniz FL, Ximenes AM, Bittencourt PS, Hernandez-Rangel SM, Campos Z, Hrbek T, Farias IP (2019) Detecting population structure of *Paleosuchus trigonatus* (Alligatoridae: Caimaninae) through microsatellites markers developed by next generation sequencing. Mol Biol Rep 46:2473-2484. doi:10.1007/s11033-019-04709-7
51. Ogutu C, Fang T, Yan L, Wang L, Huang L, Wang X, Ma B, Deng X, Owiti A, Nyende A (2016) Characterization and utilization of microsatellites in the *Coffea canephora* genome to assess genetic association between wild species in Kenya and cultivated coffee. Tree Genetics & Genomes 12:1-9
52. Nguyen HQ, Chae S, Kim E, Jang Y (2019) Characterization of polymorphic loci for two cicada species: *Cryptotympana atrata* and *Hyalessa fuscata* (Hemiptera: Cicadidae) (vol 46, pg 1555, 2019). Mol Biol Rep 46:3615-3615. doi:10.1007/s11033-019-04877-6
53. Dantas LG, Alencar L, Huettel B, Pedrosa-Harand A (2019) Development of ten microsatellite markers for **Alibertia edulis** (Rubiaceae), a Brazilian savanna tree species. Mol Biol Rep 46:4593-4597. doi:10.1007/s11033-019-04819-2
54. Ekué MR, Gailing O, Finkeldey R (2009) Transferability of simple sequence repeat (SSR) markers developed in *Litchi chinensis* to *Blighia sapida* (Sapindaceae). Plant molecular biology reporter 27:570-574
55. Thomas DT, Ahedor AR, Williams CF, Crawford DJ, Xiang QYJ (2008) Genetic Analysis of a Broad Hybrid Zone in *Aesculus* (Sapindaceae): Is There Evidence of Long-Distance Pollen Dispersal? Int J Plant Sci 169:647-657
56. Allendorf FW, Luikart G (2009) Conservation and the genetics of populations. John Wiley & Sons,
57. Turchetto C, Segatto ALA, Mader G, Rodrigues DM, Bonatto SL, Freitas LB (2016) High levels of genetic diversity and population structure in an endemic and rare species: implications for conservation. Aob Plants 8. doi:10.1093/aobpla/plw002
58. Li Z-Z, Lu M-X, Gichira AW, Islam MR, Wang Q-F, Chen J-M (2019) Genetic diversity and population structure of *Ottelia acuminata* var. jingxiensis, an endangered endemic aquatic plant from southwest China. Aquat Bot 152:20-26. doi:10.1016/j.aquabot.2018.09.004
59. Panda S, Naik D, Kamble A (2015) Population structure and genetic diversity of the perennial medicinal shrub *Plumbago*. Aob Plants 7. doi:10.1093/aobpla/plv048

60. Borghesio L, Samba D, Githiru M, Bennun L, Norris K (2010) Population estimates and habitat use by the Critically Endangered Taita *Apalis fuscigularis* in south-eastern Kenya. Bird Conservation International 20:440
61. Yineger H, Schmidt DJ, Hughes JM (2014) Genetic structuring of remnant forest patches in an endangered medicinal tree in North-western Ethiopia. BMC genetics 15:1
62. Goldewijk KK (2001) Estimating global land use change over the past 300 years: The HYDE Database. Global Biogeochemical Cycles 15:417-433. doi:10.1029/1999gb001232
63. Sintayehu DW (2018) Impact of climate change on biodiversity and associated key ecosystem services in Africa: a systematic review. Ecosystem Health and Sustainability 4:225-239. doi:10.1080/20964129.2018.1530054
64. Rajora OP, Mann IK (2021) Development and characterization of Novel EST-based single-copy genic microsatellite DNA markers in white spruce and black spruce. Mol Biol Rep. doi:10.1007/s11033-021-06231-1

Tables

Table 1: Collection information and genetic parameters for *Dodonaea viscosa* populations.

Population	Altitude (m)	Longitude	Latitude	N	N_a	N_e	H_o	H_e	F_{IS}
Ngangao 1	1830	038 ° 20'26" E	03 ° 22'16" S	10	3.600	2.345	0.560	0.468	0.886
Vuria 1	1809	038 ° 20'07" E	03 ° 22'56" S	7	2.700	1.973	0.514	0.443	0.747
Vuria 2	1804	038 ° 19'39" E	03 ° 24'14" S	8	4.100	2.979	0.457	0.582	1.105
Yale	2039	038 ° 19'44" E	03 ° 24'03" S	18	4.600	2.806	0.500	0.550	1.060
Mt. Kenya 1	2165	037 ° 33'32" E	00 ° 08'33" N	21	4.100	2.448	0.576	0.534	0.972
Mbololo	1184	038 ° 25'53" E	03 ° 23'23" S	9	3.200	2.011	0.578	0.454	0.798
Mt. Kenya 2	1885	038 ° 19'44" E	03 ° 21'19" S	10	3.300	2.334	0.530	0.487	0.856
Ngangao 2	1885	038 ° 20'19" E	03 ° 24'03" S	10	3.900	2.464	0.440	0.546	1.008
Average					3.688	2.420	0.519	0.508	0.929

N, Sample size; N_a, number of different alleles; N_e, number of effective alleles; H_o, observed heterozygosity; H_e, expected heterozygosity; F_{IS}, fixation index.

Table 2: The frequency of different repeat motifs of *Dodonaea viscosa*

Repeats	Mo	Di	Tri	Tetra	Penta	Hexa	Totals	Percentage
5	-	0	6796	385	74	108	7363	14.8
6	-	6433	2846	73	10	45	9407	18.9
7	-	3412	1268	37	5	30	4752	9.5
8	-	2275	559	21	5	20	2880	5.8
9	-	1442	357	11	1	10	1821	3.7
10	12755	830	205	6	0	8	13804	27.7
11	3884	447	155	7	2	8	4503	9.0
12	1702	277	112	7	1	4	2103	4.2
13	845	182	98	2	0	5	1132	2.3
14	453	97	87	0	0	15	652	1.3
>15	681	249	466	19	4	0	1419	2.8
Total	20320	15644	12949	568	102	253	49836	100
Percentage	40.8	31.4	26.0	1.1	0.2	0.5	100	

Di-, dinucleotides; Tri-, trinucleotides; Tetra-, tetranucleotides; Penta, pentanucleotides; Hexa, hexanucleotides.

Table 3: Characterization of eighteen microsatellite markers for *Dodonaea viscosa*

Locus	Primer sequences (5 ' – 3 ')	Repeat motif	Allele size (bp)	T_m (° C)	Accession number
Dodo 001	F: GCTCGGCAATCAAGTCTTTT R: ATTCTTCTGACGGTGGGAGA	(TAA)5	216-230	54	KX685429
Dodo 002	F: ATCTGGCTTGTTGTTGGTCC R: TGGTCCCATTGATGGATTTT	(TGT)5	229-242	52	KX685430
Dodo 003	F: TGGCCAAAGAGTTGGAAAGA R: GGCCATGAATTGTTGATTGA	(AC)7	237-256	52	KX685431
Dodo 004	F: GGTGTTTCAATGGCTCTGGT R: AGATGGGATCTGTGTTTGGC	(AT)8	222-239	54	KX685432
Dodo 010	F: AGAGAGGGATGGTAAGGGGA R: TGAAAACTCCCCATCCTCTG	(GAG)5	218-260	52	KX685433
Dodo 015	F: CCATCGCAATACAAGTGGTC R: GATGCAACAAAGTTCACCCA	(AT)6	223-228	52	KX685434
Dodo 020	F: GGGAGCAGTTTCCACTCTTG R: TTTTGGCAACATTTTTGTTTG	(AC)9	173-211	52	KX685435
Dodo 023	F: TTTTCGTGAGATGTGCTTCG R: CCTCTTTCTTCAGCTTAGCCC	(TC)6	213-216	52	KX685436
Dodo 024	F: GGTTTGAATCCTTTCACCGA R: GAGACAGTGGTCCAGTGCAG	(TAGC)5	255-262	52	KX685437
Dodo 026	F: GACAAACATCCAAAGTGCCC R: GCCAGAGAAGAGATCAACGG	(TC)9	210-223	52	KX685438
Dodo 029	F: GAGACAGAGATGGGGACAGG R: GGTTCATCACCATGCCTTCT	(AGGGAT)6	222-257	54	KX685439
Dodo 032	F: CTGTCACAAGTTCGGCAAGA R: ACTGCCACCCTGAATTTGTC	(TGTT)6	273-298	53	KX685440
Dodo 008	F: CGCATAAGTCAATTTTCCCA R: GCAGCTCAGTAAGTTAGGCCA	(CT)6	272	52	KX685441
Dodo 012	F: AAAAATTGGGGCTTTTACTTCC R: TCATGCATCAATAACACACACAA	(TTG)5	263	52	KX685442
Dodo 018	F: GCCTGAAGCAAATATTGGGA	(CT)7	218	52	KX685443

	R: GCGAGCACTCTGCATCATAA				
Dodo 021	F: CAGCCCATTGATCATCCTCT R: CCTGCAGAGAAAAAGATGGC	(CAC)5	250	52	KX685444
Dodo 030	F: CAAAAGGCCGATCAAATGTT R: TGGGAAGAAAAATGTTTGGC	(AAAG)5	225	52	KX685445
Dodo 031	F: CGTTTGTTAGGCATGGGTCT R: TGCTCATGCTCTTGCTTGTT	(TG)8	245	52	KX685446

T_m, Melting temperature

Table 4: Summary statistics for genetic variation of ten SSR markers in *Dodonaea viscosa*

Locus	Repeat motif	T_m	N_a	N_e	H_o	H_e	F_{IS}	F_{ST}	N_m	I	PIC
Dodo001	(TAA)5	54	2.250	1.360	0.119	0.227	0.476	0.528	0.223	0.402	0.465
Dodo002	(TGT)5	52	3.375	1.971	0.540	0.475	-0.137	0.132	1.648	0.839	0.481
Dodo003	(AC)7	52	2.875	1.705	0.352	0.370	0.048	0.286	0.625	0.652	0.530
Dodo004	(AT)8	54	4.625	3.335	0.982	0.683	-0.436	0.109	2.048	1.292	0.718
Dodo010	(GAG)5	52	4.750	3.380	0.738	0.650	-0.136	0.179	1.148	1.240	0.741
Dodo020	(AC)9	52	5.000	2.972	0.515	0.641	0.196	0.224	0.866	1.244	0.805
Dodo023	(TC)6	52	2.625	1.834	0.282	0.409	0.312	0.267	0.686	0.679	0.472
Dodo026	(TC)9	52	2.250	1.693	0.383	0.388	0.013	0.102	2.200	0.604	0.383
Dodo029	(AGGGAT)6	54	3.875	2.384	0.533	0.543	0.019	0.236	0.807	0.964	0.652
Dodo032	(TGTT)6	53	5.250	3.564	0.750	0.691	-0.086	0.144	1.490	1.375	0.799
Average			3.688	2.420	0.519	0.508	0.027	0.221	1.174	0.929	0.605

N_a, number of alleles; N_e, effective number of alleles; H_o, observed heterozygosity; H_e expected heterozygosity; F_{IS}, fixation index; F_{ST}, genetic differentiation; I, Shannon's information index; PIC polymorphism information; N_m = [(1 / F_{ST}) - 1] / 4.

Table 5: Analysis of molecular variance (AMOVA) of *Dodonaea viscosa*

Source	d.f.	SS	MS	CV	%
Among Populations	7	123.726	17.675	0.665	20%
Among Individuals	84	231.573	2.757	0.063	2%
Within populations	92	242.000	2.630	2.630	78%
Total	183	597.299		3.359	100%

d.f., Degrees of freedom; SS, Sum of squares; MS, mean sum of squares; CV, Variance component estimates; %, total percentage of total variation.

Table 6: Results of Bottleneck tests among populations of *Dodonaea viscosa* based on the Wilcoxon test method and mode-shift test.

Population	I.A.M	TPM	SMM	MODE-SHIFT
Ngangao 1	0.496	0.82	0.164	shifted mode
Vuria 1	0.027*	0.129	0.359	shifted mode
Vuria 2	0.105	0.193	0.77	normal L-shaped
Yale	0.084	0.846	0.232	normal L-shaped
Mt. Kenya 1	0.084	0.625	0.625	normal L-shaped
Mbololo	0.922	0.375	0.625	normal L-shaped
Mt. Kenya 2	0.16	0.769	0.625	shifted mode
Ngangao 2	0.275	0.922	0.625	normal L-shaped

IAM, infinite allele model; TPM, two-phrased model of mutation; SMM, stepwise mutation model

Table 7: Model performance

Period	**AUC**	**MTP**
Current	0.905	0.1805
LGM	0.889	0.1643
RCP 4.5	0.923	0.2011
RCP 8.5	0.921	0.1770

AUC, Area Under Curve; MTP, Minimum Training Presence

Table 8: Percent contributions of the variables to the distribution of *Dodonaea viscosa* according to Maxent Modeling (bold values are the most important variables)

Variable		LGM	Current	RCP4.5	RCP8.5
Bio2	mean diurnal range	2.6	1.3	1.1	1.0
Bio4	temperature seasonality	**17.8**	**17.8**	**17.1**	**16.2**
Bio8	mean temperature of wettest quarter	**12.0**	**15.8**	**18.2**	**17.0**
Bio13	precipitation of wettest month	2.3	3.2	2.8	3.4
Bio14	precipitation of driest month	**11.2**	12.1	8.5	8.8
Bio18	precipitation of warmest quarter	2.9	**13.3**	**18.7**	**20.0**
Bio19	precipitation of coldest quarter	8.5	1.7	4.6	6.0
Elevation		**42.7**	**34.7**	**29.0**	**27.5**

Figures

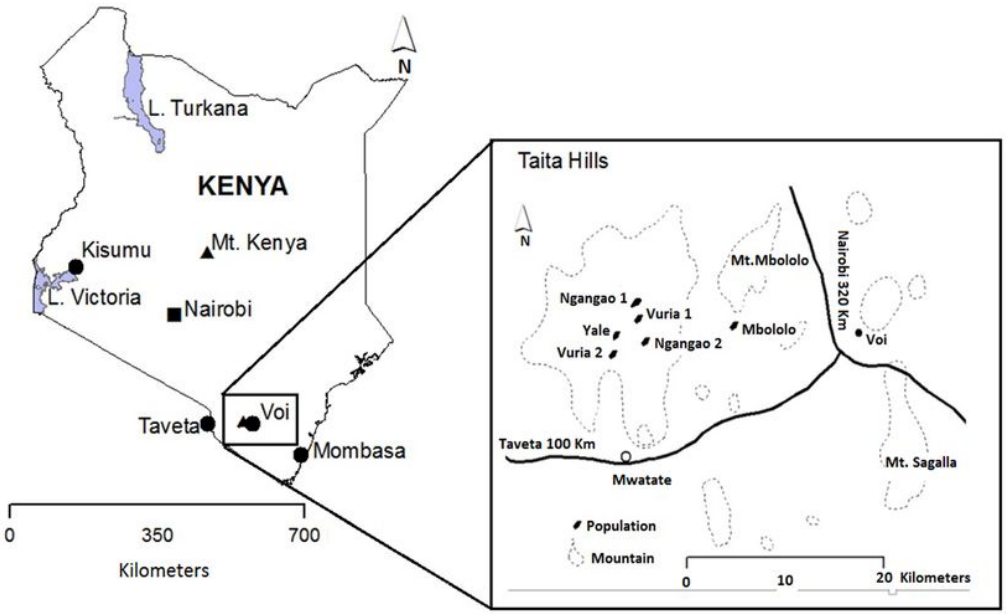

Figure 1

Geographical map indicating sampling locations of Dodonaea viscosa. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

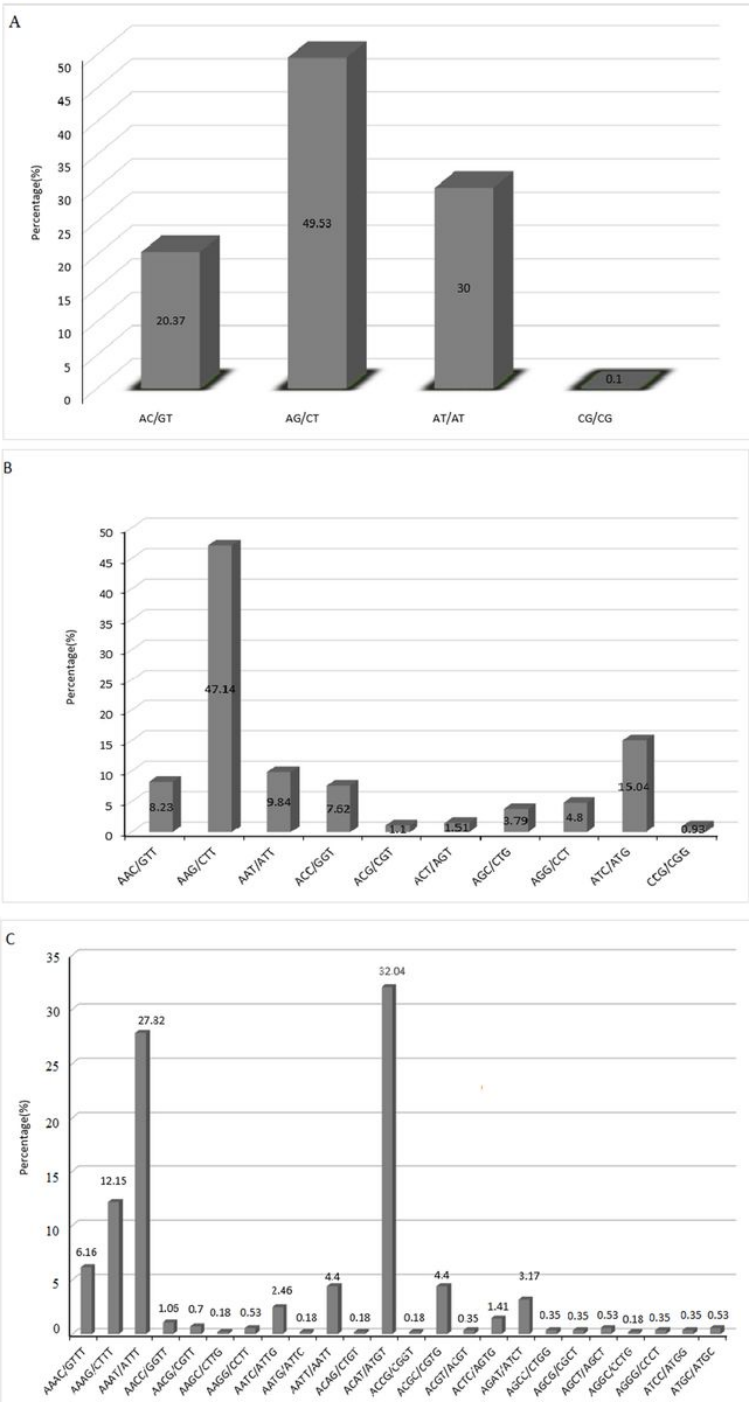

Figure 2

Characterization and frequency of different motifs among dinucleotide repeats (A), trinucleotide repeats (B) and the tetranucleotide repeats (C).

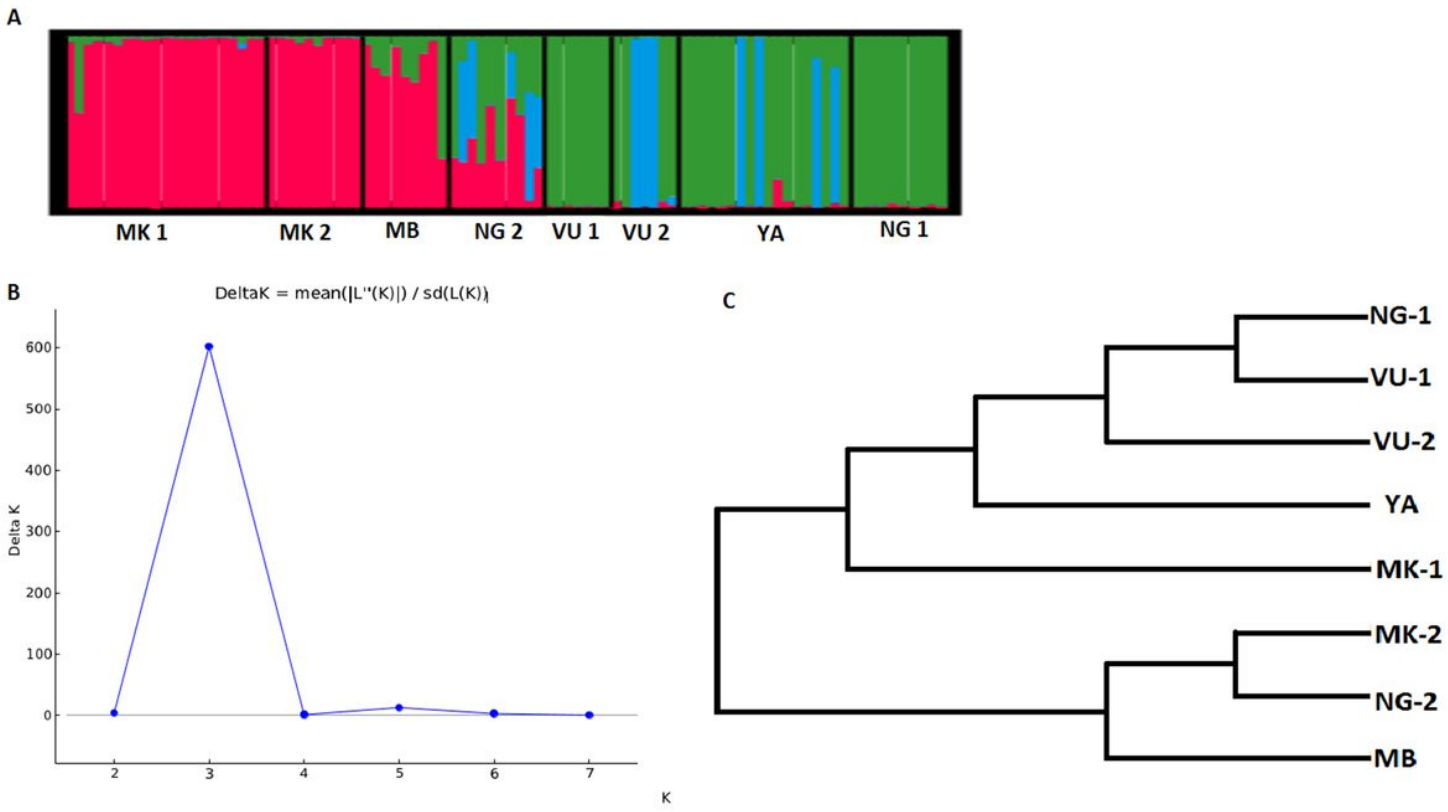

Figure 3

Genetic relationship among 92 individuals of Dodonaea viscosa tested in this study. (A) K=3 peak. (B). Population structure derived from the analysis of Ten SSRs. MK-1, MK-2 = Mt. Kenya 1&2 and MB = Mbololo, NG-1 = Ngangao 1, VU-1, VU-2 = Vuria 1 &2 and YA- Yale population. (C). UPGMA dendrogram of genetic relationships among eight collected populations based on Nei's (1972) genetic distance using SSR data.

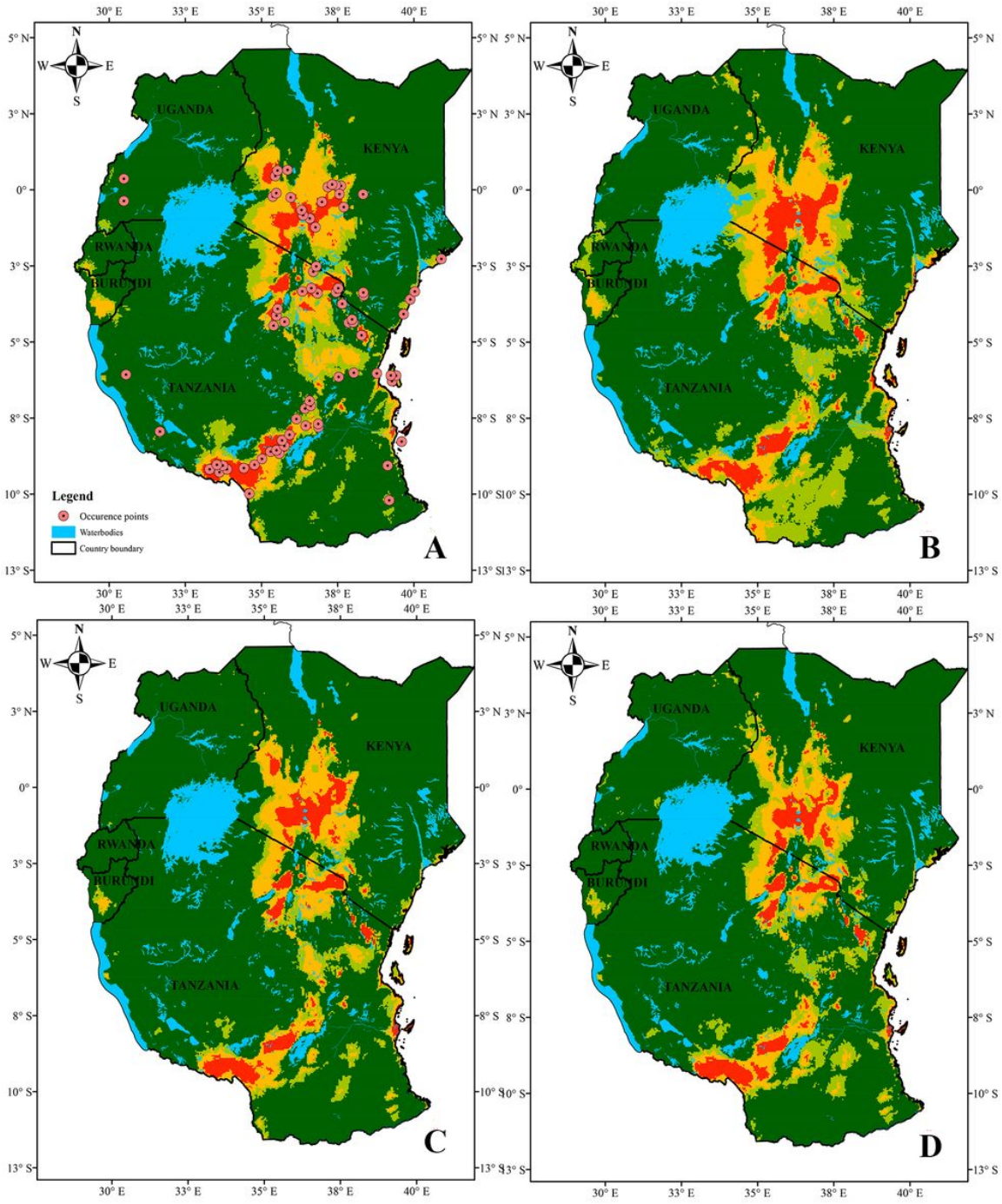

Figure 4

Predicted probability of occurrence for Dodonaea viscosa across East Africa based on species distributional modelling outcomes for A) current, B) Last Glacial Maximum, C) future (2070) RCP4.5 of intermediate greenhouse gas emissions scenario, and D) future (2070) RCP 8.5 of extreme greenhouse gas emissions scenario. Habitat suitability is classified as follows, high- in red, medium- in yellow, low- in light green, and unsuitable- in dark green. Circular pink dots denote occurrence localities used in developing the models in the present study obtained from GBIF Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.