


RESEARCH

Open Access



Unlocking the geography of Azobé timber (*Lophira alata*): revealing spatial genetic structure beyond species boundaries

Barbara Rocha Venancio Meyer-Sand^{1*} , Laura E. Boeschoten^{1,2}, Gaël U.D. Bouka³, Jannici C.U. Ciliane-Madikou³, G. Arjen de Groot⁴, Nathalie de Vries⁵, Nestor L. Engone Obiang⁶, Danny Esselink⁵, Mesly Guieshon-Engongoro³, Olivier J. Hardy⁷, Simon Jansen⁸, Joël J. Loumeto³, Dieu-merci M.F. Mbika³, Cynel G. Moundounga⁹, Dyana Ndiade-Bourobou¹⁰, Rita M.D. Ndangani³, Marinus J. M. Smulders⁵, Steve N. Tassiamba¹¹, Martin T. Tchamba¹¹, Bijoux B.L. Toumba-Paka³, Herman T. Zanguim¹¹, Pascaline T. Zemtsa¹¹ and Pieter A. Zuidema¹

Abstract

Background The illegal trade of tropical timber constitutes a major and persistent environmental problem. Since the detection of fraud in trade documents remains challenging, forensic tools that can independently trace timber origin are needed. In this study, we evaluated the potential of the chloroplast genome (plastome) as a genetic tool to verify the claimed species and geographic origin of timber from Azobé (*Lophira alata*), an intensively exploited and threatened tropical tree species.

Results We sampled 480 trees from *Lophira alata* and the congeneric species *L. lanceolata* across nine countries in Central and West Africa. Sampling included *L. alata* trees from 15 logging concessions in Cameroon, Gabon and the Republic of the Congo. DNA was isolated from the cambium or leaf tissue, and complete plastid genomes were assembled. A total of 228 SNPs from 436 trees were retained, which formed 35 pDNA haplotypes (with a length of 179 SNPs). The two *Lophira* species shared one plastid haplotype and contained several closely related plastid haplotypes. For the exploited *L. alata*, we detected a moderately strong correlation between genetic and spatial distances. Two haplotypes were widely spread across the core of Central Africa, while several others were more spatially constrained or endemic, for example, in West Gabon (potentially a *L. alata* cryptic species) and Northern Congo.

Conclusions The distribution of haplotypes revealed a clear spatial structure. Some widely spread haplotypes potentially hamper site distinction of Azobé wood samples, but still reveal their wider region of origin. In regions where endemic haplotypes are present, differentiation may be successful at finer scales. Thus, the potential spatial resolution for timber tracing may vary across regions. We assembled the first reference database of plastome-wide SNP datasets for Azobé timber, with a focus on the major logging areas. Our work represents a step towards plastome-based timber tracing for this species, but also reveals limited potential of this method for species

*Correspondence:
Barbara Rocha Venancio Meyer-Sand
barbara.rvms@gmail.com

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

differentiation. To validate the potential of the plastid genome for timber tracing, further steps, including assignment and blind sample tests, will be needed.

Keywords Spatial genetic structure, Plastid genome, Chloroplast genome, SNPs, *Lophira*, Azobé, Origin differentiation.

Introduction

Illegal timber exploitation is a major and persistent threat to forest ecosystems, resulting in significant environmental, economic, and social problems. This is especially the case for timber-producing tropical countries, where the proportion of illegally sourced timber can reach 50 to 90% [1]. National and international legislation to curb the illegal timber trade exist (e.g., the Regulation on Deforestation Free Products, EUDR [2], but their enforcement through verifying authenticity of trade documents remains challenging [3, 4]. Hence, forensic methods that can independently determine the geographic origin or the taxonomic identity of traded timber based on intrinsic wood properties are crucially needed [4]. These timber tracing methods include wood anatomical, chemical and genetic methods [4]. Genetic methods are frequently employed to check the taxonomic identity of samples, but their use for tracing geographic origin is still limited [4]. One of the reasons for this situation is that genetic timber tracing requires population-specific genetic markers that show variation in regions where timber fraud occurs, as well as robust reference databases representative of the species distribution and regions with high exploitation pressure.

Another major technical challenges in using genetic methods in timber tracing is the generally low quality of DNA in the traded parts of wood (i.e., heartwood [5, 6]). The low DNA quality seriously hampers the application of various types of genetic markers, including even those based on relatively short fragments such as microsatellite markers, which are relatively cheap and easy obtain and analyze [7, 8]. This has prompted the development of markers based on variation at only a single DNA nucleotide (i.e., single nucleotide polymorphisms, SNPs). Assessing the variation for SNPs based on the sequencing of short DNA fragments may enhance success rates compared to PCR amplification of fragments of > 100 nucleotides, such as commonly used microsatellite markers. However, when using shotgun sequencing of the nuclear genome, the resulting low coverage per individual fragment may also hamper reliable SNP calling. A possible solution is the use of the plastid genome (or ‘plastome’), which occurs in multiple copies per cell, as it is much shorter and may be less prone to degradation than the nuclear genome [9–11]. Consequently, the likelihood of obtaining DNA sequences and successfully genotyping SNPs from small amounts of degraded DNA extracted from wood is greater for plastid genomes than

for nuclear genomes. For several tropical timbers, Mascarello et al. (2021) [12] noted that intraspecific variation in highly variable parts of the plastome offers potential for genetic tracing. Yet, thus far most genetic tracing studies have used polymorphisms in the nuclear genome [4, 13]. Current bioinformatic tools allow the assembly of whole plastomes based on the multitude of short sequence reads obtained by shotgun sequencing and map the SNP positions onto the plastome, which together define a ‘plastid haplotype’ (as the plastome is inherited clonally and does not recombine). The identified haplotypes compose the reference database. The reference haplotype database can subsequently be extended to include new SNP positions when data from new, genetically different samples are added.

A second challenge in developing genetic tracing methods is the scarcity of reference samples [4]. Reference samples must represent major timber logging areas, including illegal logging hotspots as well as protected regions to capture a representative diversity of SNPs and haplotypes frequencies. Additionally, sufficient sampling intensity per site is needed to allow population genetic analyses during assignment tests. Thus, a critical initial step in developing genetic tracing methods is the creation of informative SNP sets through intensive field sampling across the timber production range.

Here, we perform the first step for a major African timber species: Azobé (*Lophira alata* Banks ex C.F. Gaertn, Ochnaceae). Azobé timber is commonly applied in hydraulic engineering (sluices, bridges) and is among the most valuable and most widely traded hardwood species from the Congo Basin. Together with six other flagship species, it accounts for 50% of the timber production in the Congo Basin [14]. Azobé species is currently listed as vulnerable on the Red List of the IUCN due to high logging pressure [14–16], which causes a depletion of mature individuals and fragmentation of populations. Azobé is part of the genus *Lophira* (Ochnaceae), which comprises two main recognized species across Africa: *Lophira alata* and *L. lanceolata*. These two species are morphologically similar and widely distributed, co-occurring in sympatry in contact zones along a 3000 km stretch between the rainforest and the savannah [16, 17]. *Lophira alata* is a monoecious, deciduous and wind-dispersed species [18], which occurs mainly in wet tropical forests from Guinea to the Democratic Republic of the Congo [19], while the closely related relative, *L. lanceolata*, occurs in woodlands and dry forests [17, 20].

Lophira alata produces small, inconspicuous flowers that are considered to be primarily insect pollinated [21] and large, woody, round nuts known as “Bongossi nut”, which are animal dispersed. To assess the potential for species identification, we included the closely related species *L. lanceolata*. Moreover, using nuclear microsatellite loci, Ewédjè et al. (2020) [17] were able to differentiate the two previously known species, and they also detected a cryptic species within *L. alata*, endemic to western Gabon, which has not yet been formally described but which we refer to as *L. alata*-WG.

A first approach to employ spatial genetic structure of Azobé was done by Blanc-Jolivet et al. (2021) [22] by developing a set of single nucleotide polymorphism (SNP) markers for *L. alata* from West and Central Africa, including 75 nuclear, 20 chloroplast, and 28 mitochondrial SNPs. By employing the SNP set, a theoretical accuracy of 86% was achieved when differentiating the origin between West and Central Africa [22]. Building on this work, we propose the employment of plastome-wide SNPs for phylogeographic and spatial structure assessment, with aim on forensic analyses (genetic tracing) [23], similar to the chloroplast super-barcodes proposed by Li et al. (2015) [24]. As mentioned before, targeting only the plastome can significantly reduce laboratory challenges, sequencing depth requirements for reliable SNP genotyping [23], and, hence, lowering the associated costs. To cover the whole *Lophira* complex and recently discovered genetic substructure, we also included samples of the congeneric species *Lophira lanceolata*, which potentially hybridizes with *L. alata* in contact regions [17, 20]. We also included a potential cryptic species in West Gabon (*L. alata*-WG) recently described and also traded as Azobé timber [17]. We obtained reference samples from across the distribution range of Azobé, with high sampling efforts (in terms of number of sites and number of samples per site) in managed forest concessions in three main Azobé-trading countries: Cameroon, Congo and Gabon [25].

In this study we addressed the following research questions: (1) What are the main genetic clusters based on plastome-wide SNPs, and do these clusters correspond to the abovementioned *Lophira* species? (2) What are the main patterns of spatial genetic structure, and how do they relate to the origin differentiation of Azobé? (3) Can haplotypes based on plastid-wide polymorphic SNPs be used for origin differentiation?

Methods

Sampling area and strategy

The *L. alata* and *L. lanceolata* samples were collected from two sources: (a) 96 leaf and cambium samples (52 *L. alata* and 28 *L. lanceolata*) collected during several field expeditions by the Université Libre de Bruxelles in nine

African countries, covering most of the distributional range of both botanical species (Fig. 3a, sites with 1–6 trees), and (b) 384 newly collected cambium samples of *L. alata* between 2019 and 2022 in 15 forest concessions in Cameroon, Gabon and the Republic of the Congo (Fig. 3a, sites with ≥ 17 trees). Type b samples were obtained from sites at distances ranging from 15 km to 1050 km apart. The sampled trees within each of these sites were at least 100 m apart and at most 5000 m apart and were at least 30 cm in diameter at breast height. All trees were georeferenced, and all sample types were dried with silica gel (see the list of samples in Additional file 1).

Type a sampled trees were identified by various technical botanists during multiple field expeditions. No herbarium specimens were collected. Identification of sampled trees type b from GAB1 and GAB2 were conducted by Raoul Niangadouma (botanist at IPHAME-TRA, Herbarier National du Gabon) and Giresse Nziengui Armand (botanist at CEB/GAB2). Identification of trees sampled in CON2 was carried out by Issac Zombo Dikele (botanist at CIB/CON2), and in CAM1, it was conducted by Bertrand Belibi (consultant botanist). For the remaining type b samples, trees were identified by the technical botanists of each respective forest concessionaire company.

Laboratory analysis

DNA was isolated from leaf or cambium tissue of 480 trees with an optimized cetyltrimethyl ammonium bromide (CTAB) protocol as described by Dumolin et al. (1995) [26] with additional cleaning steps (Additional file 2). The DNA purity of all extracts was checked with Nanodrop (Thermo Fisher Scientific, Schwerte, Germany). DNA concentrations were measured with the Qubit™ kit (Thermo Fisher Scientific, Schwerte, Germany) following the manufacturer's instructions, and 1.5% agarose gel was used to check the fragments length range. The DNA isolates were used to prepare five paired-ended libraries with 300 bp or more insert sizes with the RIPTIDE High Throughput Rapid Library Prep Kit (Twist Bioscience, South San Francisco, USA). These libraries were sequenced with Illumina Novaseq6000 PE150 (Novogene, Cambridge, United Kingdom).

Data analysis

Bioinformatics

The Illumina sequences were assembled into 480 plastid genomes by mapping them against the annotated chloroplast genome of Azobé (MZ274135.1 [12] using Bowtie2 [27]). The variant call considered all mapped reads without filtering and was performed using NGSEPcore [28]. A variant call file containing only biallelic loci was generated, in which heterozygous variants were maintained only when at least one sampled tree was homozygous for

the minor allele using R version 4.1.0 [29], the detected variants underwent further filtering. Specifically, we excluded SNPs with sequencing depths of less than five reads and greater than 250 reads and more than 10% missing data (SNPfiltR package [30]). Genotypes that did not fulfill the abovementioned criteria were considered missing data. Individual trees with more than 10% missing data across SNPs were also removed from the dataset (SNPfiltR package [30]). The effect of filtering on clustering patterns was evaluated by PCA (principal component analysis) (Adegenet package [31]), with completeness of data per SNP (across individuals) varying between 70 and 90% (Additional file 3). These analyses revealed that the number and distribution of clusters were rather robust to filtering procedures. In addition, samples shown between clusters in the PCA generally had close to 10% missing data (Additional file 3). Our filtering procedure resulted in a dataset of 228 SNP sequences from 436 trees (408 *L. alata* and 28 *L. lanceolata* individuals).

Phylogenetic analysis and data structure

We evaluated the primary structure of the data based on a Randomized Axelerated Maximum Likelihood (RAxML) phylogenetic-tree [32]. For this analysis, heterozygous SNPs were transformed into missing data, and SNP loci with more than 20% missing data were removed from the data. The SNPs subset comprised 179 SNPs, which were combined into 35 unique haplotypes presented in the 436 trees. The RAxML phylogenetic-tree was generated using rapid bootstrapping with 1000 inferences and the best scoring phylogenetic-tree was selected.

The following analyses were performed on the complete dataset of 228 SNPs for 436 trees. A Bayesian information criterion (BIC) (Adegenet package [31]) was computed to determine the number of sub-clusters of samples within the three main lineages (as perceived from the RAxML phylogenetic-tree). The subgroupings for $K=10$ (to account for the observed sub-structure of the BIC analysis) were subsequently assessed with a discriminant analysis of principal components (DAPC) (Adegenet package [31]). We chose to perform DAPC as it is a widely used tool for analyzing nuclear DNA and applied in other research fields, which makes the results of the present study comparable with those of other tracing studies (e.g., chemical or visual tracing). Post-probabilities of membership were calculated for $K=10$. The consistent sub-clustering patterns observed in both the RAxML phylogenetic-tree and DAPC analyses, except for K3f, led us to opt for the utilization of DAPC to depict the structure observed in the data. K3f sub-cluster exhibited substructure, grouping 11 individuals scattered across different branches of the phylogenetic-tree, for which we did not conduct further analysis due to the low number of samples. Given the congruence, we employ the terms

“clusters” and “lineages” interchangeably. To visualize the geographic structure of these clusters, cluster membership was plotted on a map (using QGIS- qgis.org).

Haplotypes' characterization, distribution and relationships

The SNP subset of 179 SNPs for the 436 Azobé trees (used in the phylogenetic-tree) was used to define plastid haplotypes. We computed the number of segregating sites (S), nucleotide diversity and diversity (π). We created a median joining haplotype network of unique haplotypes [33] to assess their relationships with PopArt [34], and combined this network with a haplotype distribution map (QGIS). The algorithm implemented in PopArt further removed loci with more than 5% missing data.

Genetic versus geographic distances

Genetic distances were calculated for the 15 sites at which more than 17 individuals were sampled ($n=362$ Azobé trees, 228 SNPs). Genetic distance was calculated using Nei's distance (Adegenet package [31, 35]). Geographic distances were calculated as pairwise distances between centers of sites. A Mantel test was performed for strictly *L. alata* individuals, and *L. alata*-WG individuals were not included. We conducted the Mantel test with 10,000 permutations to evaluate the significance of the relationships between genetic and geographic distances among the sites.

Results

Main genetic clusters, their distribution and correspondence to species

The results of the phylogenetic-tree analysis (Figs. 1 and 2a) revealed three main clades. The first clade, K1 (orange circles and triangles, Fig. 2a), included samples from *L. lanceolata* and *L. alata* from extreme West Africa, from the Dahomey gap to the West. The second clade, K2 (green circles, Fig. 2a), is composed exclusively of *L. alata* from two sites in West Gabon (likely the cryptic *L. alata*, here referred to as “*L. alata*-WG” [17]). The third clade, K3 (blue circles and triangles, Fig. 2a), contains samples from all other sites in Central Africa (from the Dahomey gap into Central Africa), including *L. lanceolata* (gray triangles, Fig. 2b). There were clear genetic differences between *L. lanceolata* and *L. alata* in Central Africa, as shown by gray triangles clustered separately from other K3 sub-clusters (K3d, Figs. 1 and 2b). The grouping suggested that the geographic signal (West vs. Central Africa) was more strongly dominant than the species signal, as *L. lanceolata* from West vs. Central Africa clustered separately.

The analysis of the sub-structure using successive k-means in Adegenet identified the number of sub-clusters as 9 or 10 (Additional file 4). We selected the greatest sub-cluster number, 10, to account for the even

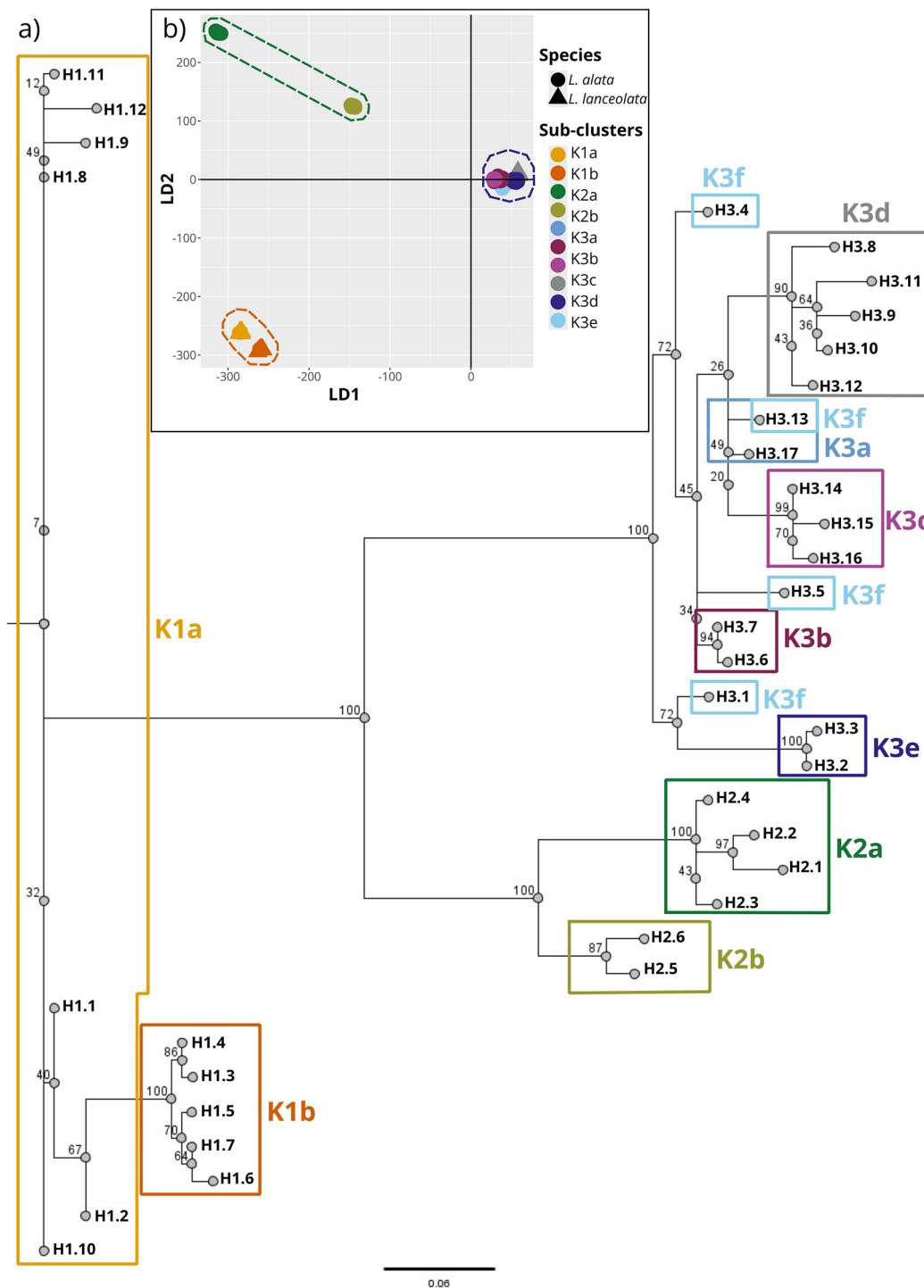


Fig. 1 Phylogenetic tree and subclusters for Azobé in West and Central Africa. **a)** Results of the Randomized Axelerated Maximum Likelihood (RAxML) phylogenetic tree based on 35 haplotypes (179 pSNPs). Numbers starting with "H" indicate haplotype, numbers next to branches represent bootstrapped support (%) of the branches, and colored contours indicate (sub-)clusters. **b)** DAPC based on 228 pSNPs showing sub-clusters and species, with colors corresponding with those of the phylogenetic tree. The orange, green and blue dashed lines encompass individuals belonging to clusters K1, K2 and K3, respectively. The data used in the DAPC contained NAs

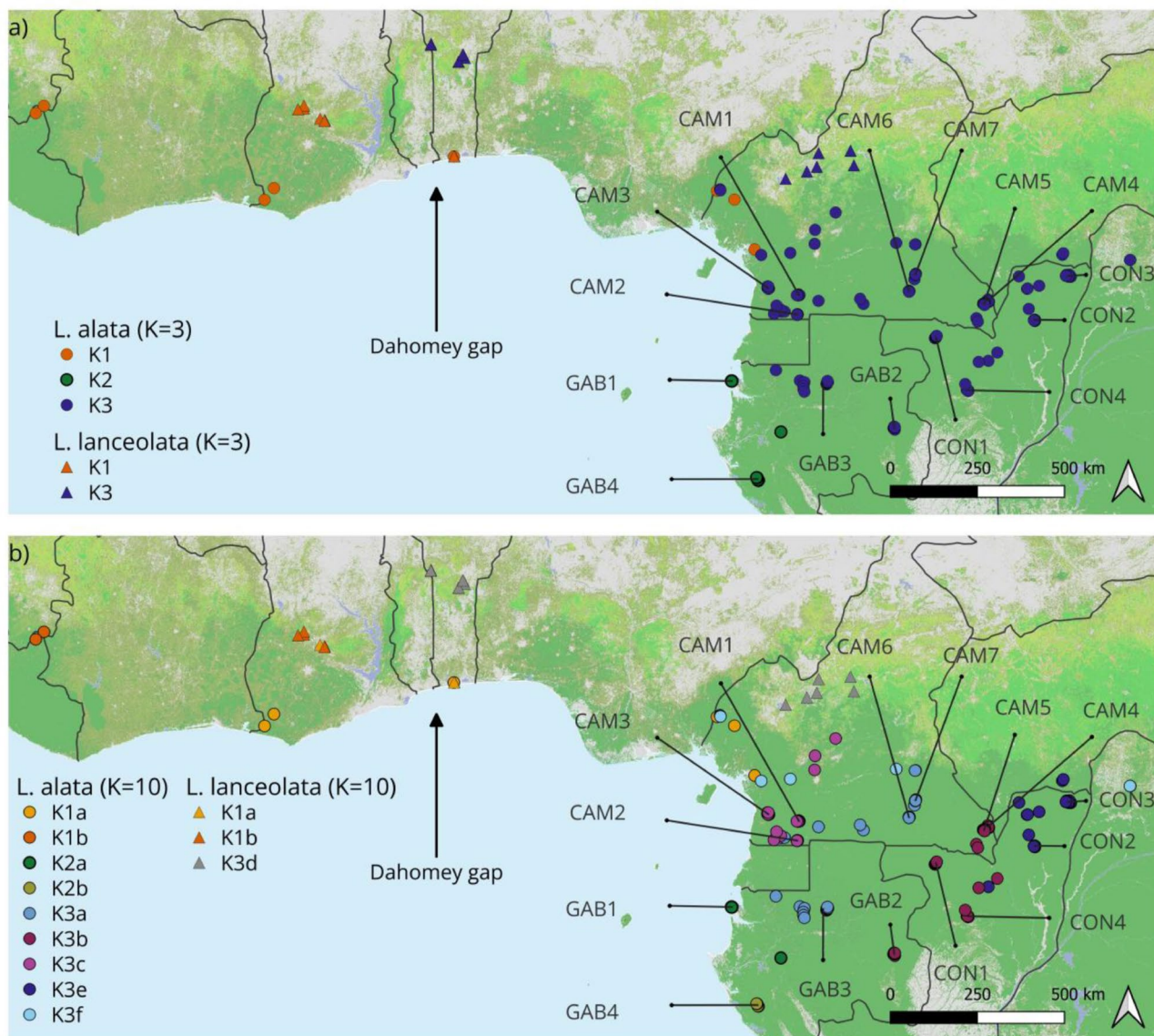


Fig. 2 Spatial distribution of genetic clusters. Spatial distribution of three main genetic clades (a) and 10 sub-clusters (b) for Azobé (*Lophira alata* and *L. lanceolata*) from West and Central Africa. Both analyses were performed with all 436 trees. Site names were indicated only for locations with samples size ≥ 17 trees (samples type b). The main three lineages were based on 179 pSNPs and were defined as perceived in the RAxML phylogenetic tree (Fig. 1). The 10 sub-lineages were based on 228 pSNPs using Discriminant Analysis of Principal Components (DAPC), and congruent with the sub-lineages observed in the RAxML phylogenetic-tree. Note that color coding in panels A and B is different

greater number of branches present on the RAxML phylogenetic-tree (Fig. 1). Assignment of samples to these 10 clusters using DAPC showed that K1 and K2 clusters were split into two sub-clusters each (K1a-b and K2a-b, respectively; Fig. 2b). The original K3 cluster is split into six sub-clusters (K3a-f). These sub-cluster patterns are supported by the clades in the RAxML phylogenetic-tree (Fig. 1), except for individuals assigned to K3f, which are present in several branches of the phylogenetic-tree.

The 10 sub-clusters were spatially structured at varying scales and/or species-specific (Fig. 2b). The sub-cluster K1a included *L. alata* individuals from Benin, Cameroon

(four sites) and Ghana (two sites), as well as *L. lanceolata* from Benin and Ghana. The sub-cluster K1b only covered West African individuals and comprised *L. lanceolata* from Ghana, along with *L. alata* from Guinea and Liberia. The sub-clusters K2a and K2b were also quite distinct from each other, and were found in the region where the cryptic *L. alata*-WG species occurs [17]. Sub-cluster K2a consisted of individuals present at two sites, while K2b was found only in the GAB4 concession. All K3 sub-clusters except for K3d were restricted to Central Africa and consisted of *L. alata* trees, whereas K3d spanned from Central Africa into the Dahomey Gap and was composed

of *L. lanceolata*. Among the K3 sub-clusters, K3a was mostly present in Cameroon and Gabon, the K3b cluster presented a wider distribution (Cameroon, Gabon and Congo), K3c represented trees from nine sites in Cameroon, and K3e was mostly composed of trees from the Republic of the Congo. Finally, cluster K3f comprised scattered individuals from Cameroon and the DRC.

Haplotype distribution and network

A total of 35 haplotypes were detected, 23 for *L. alata* (6 for *L. alata*-WG), 11 for *L. lanceolata* and one shared between the two species (H1.10) (Additional file 5; Figs. 1 and 3). We observed $S = 128$ segregating sites and $\pi = 0.21$ nucleotide diversity. The *L. lanceolata* haplotypes were present in clades K1 and K3 (indicated with dotted lines, Fig. 3b) and closer to *L. alata* haplotypes from the same geographic region of origin than to each other, consistent with the intertwined grouping patterns observed in cluster analysis and on the phylogenetic-tree (Fig. 1). *L. lanceolata* presented one common haplotype (H3.12) in eight individuals from Benin. As the algorithm implemented in PopArt removed loci with more than 5% missing data, the shared haplotype H1.10 (indicated by * in Fig. 3b) was combined with H1.8 in the haplotype

network. The shared haplotype H1.10 was closely related to the haplotypes of both species.

Within *L. alata*, the two most common haplotypes were H3.7 and H3.17, which were seven SNPs apart. They were present in more than half of the individuals in the dataset, with 118 and 117 trees, respectively (Fig. 3). H3.7 was central in the network and occurred in East Cameroon, East Gabon, and the West Congo, whereas H3.17 occurred in Central Gabon and Cameroon (Fig. 3a). Other common haplotypes (H3.3, H2.6, H3.15, H3.16, and H3.2) were present in 14–39 individuals each (Additional file 5). Haplotype H3.2 was restricted to the Congo, and (with H3.3) it was the most genetically distinct haplotype in clade K3, with 13 mutations away from the closest haplotype. H3.15 and H3.16 exclusively occurred at sites in West Cameroon (CAM1, CAM2 and CAM3). Haplotype H2.6 is part of the most genetically distant branch of the network and encompasses other very genetically distant haplotypes (H2.1–H2.6, shades of green) and more than 25 mutations away from the closest haplotype (H3.4). Within clade K2, a split between H2.1–4 and H2.5–6 haplotypes was clear, in line with the clustering patterns observed in the DAPC results. All the haplotypes in the K2 branch were confined to West Gabon. The remaining 12 haplotypes occurred in only

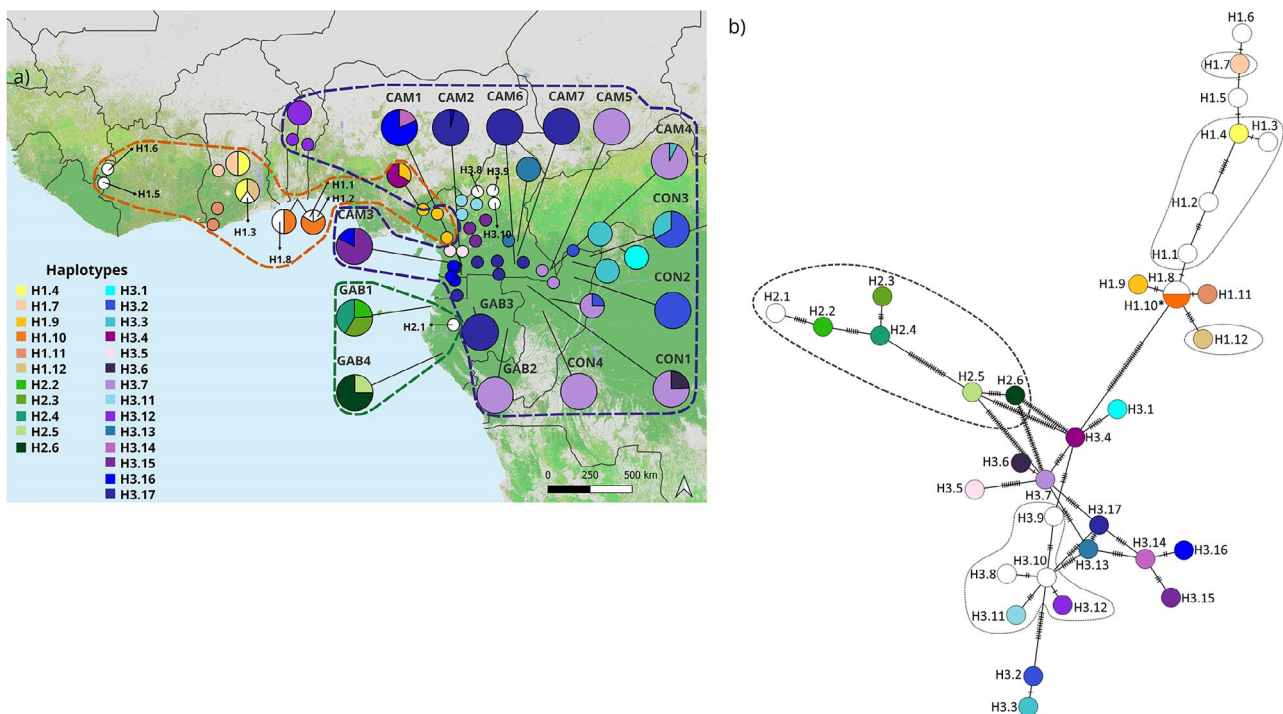


Fig. 3 Haplotypes identified for *Lophira* sp. from West and Central Africa. (a) Geographic distribution of haplotypes in the study area. The size of the circles reflects the sample size per sites (large: $n \geq 17$ trees, intermediate: 2–6 trees, small $n = 1$). Rare haplotypes ($n < 3$ trees) are shown in white, and their haplotype code is mentioned. Orange, green and blue dashed lines group sites with individuals of cluster K1, K2 and K3 respectively. Site names are labeled only for locations with $n \geq 17$ trees. (b) Haplotype network of the plastid genome, with each haplotype represented by a circle and perpendicular short lines on the branches indicating the number of mutations between the haplotypes. Dotted lines indicate haplotypes exclusively found in *L. lanceolata*. Dashed lines indicate haplotypes found in *L. alata*-WG. Haplotype H1.10 is indicated with “*”, and it is shared between *L. alata* and *L. lanceolata*

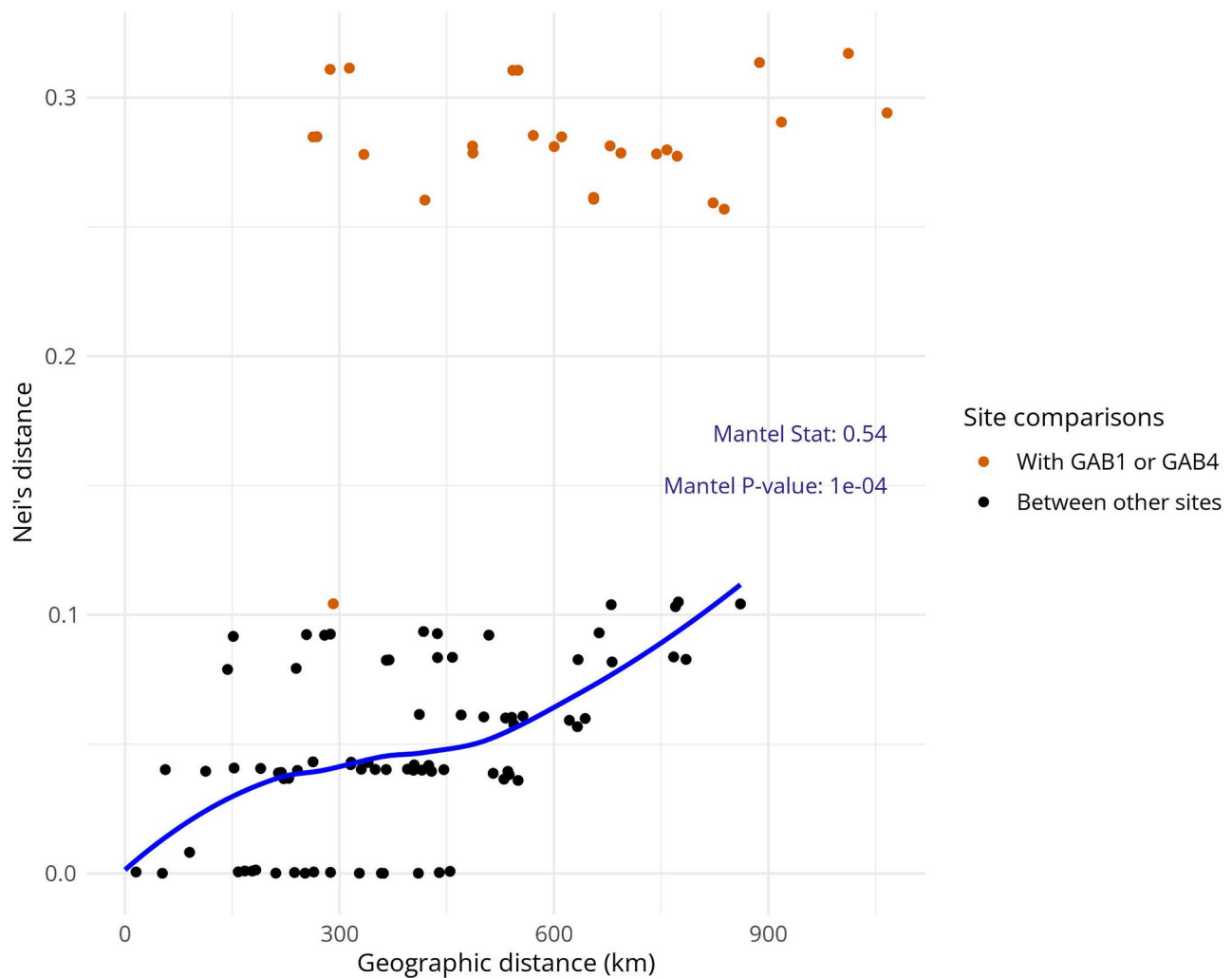


Fig. 4 Relationship between Nei's genetic distance and spatial distances. Each point represents a comparison of two out of the 15 sites with > 17 sampled trees each, either including one of the two genetically very distinct sites (GAB1 and GAB4, orange) or excluding those sites (black). The Mantel test and Loess regression are only conducted for the site comparisons without GAB1 and GAB4. The orange point with a genetic distance of approximately 0.1 represents genetic and geographic distances between the GAB1 and GAB 4 sites

2–7 individuals (also colored in Fig. 3a and b), and 4 haplotypes were present in only one individual each.

A few sites displayed very characteristic haplotypic compositions. For instance, while most sites in Cameroon contained individuals with haplotypes H3.17 and H3.7, sites CAM1 and CAM3 were mostly composed of individuals with haplotype H3.14–16. In Gabon, the GAB1 and GAB4 sites had haplotypic compositions distinct from those of the other Gabonese sites. Similarly, in the North Congo, H3.2 and H3.3 are endemic to the area and are very distinct from other *L. alata* haplotypes (Figs. 1a 2b, 3a and b, cluster K3e). The haplotype network branch arrangement was in line with the distribution of clades K1–K3 in the phylogenetic-tree and with the spatial structure of the haplotypes.

Spatial genetic structure

Plotting the pairwise genetic distances against geographic distances revealed two subsets of points (Fig. 4). The upper subset (orange) contains comparisons that include one of the West Gabonese sites (GAB1 and GAB4), which represents the genetically distinct clusters (K2a and K2b) in regions where the cryptic *L. alata*-WG is found [17]. These comparisons are characterized by particularly high genetic distances (0.26–0.32), which may be indicative of speciation. The lower subset shows considerably smaller genetic distances (a maximum of 0.10 for comparisons including trees from the North Congo, and between GAB1 and GAB4). In the subset excluding the two Gabonese sites, the genetic distances among the sites increased with geographic distances with a moderate to strong Mantel correlation coefficient ($r=0.54$, $p<0.001$, 10,000 replicates).

Discussion

Species differentiation

The results of our genetic study of plastid DNA identified three main clusters that did not fully correspond to species or geographical regions. Contrary to our hypothesis, two of the three main clades (K1 and K3) corresponded to clusters that contained both *Lophira* species but from trees in distinct geographic regions (Fig. 2a). Further analysis revealed that some of the genetic variation was shared between individuals of *L. alata* and of *L. lanceolata* (e.g., cluster K1a was composed of *L. alata* and *L. lanceolata* individuals from Cameroon, Benin and Ghana). The combined haplotype network showed a complex relationship among the plastid haplotypes of both species (Fig. 3b) and did not reveal a clear species separation between *L. alata* and *L. lanceolata*. The topology of the RAxML phylogenetic-tree also indicated that the plastomes of the studied species (*L. alata*, *L. alata*-WG and *L. lanceolata*) are not monophyletic (Fig. 1a). According to Hu et al. (2015, 2016) [36, 37], plastome markers may be of limited use for species delimitation when plastid capture/introgression occurs among species. Indeed, a lack of species-specific polymorphic variation has been reported in the *Lophira* genus [12, 38]. Sharing of the same (or very similar) haplotypes across co-generic species was also found in other tropical tree genera, including *Carapa* [39], *Greenwayodendron* [40], and *Brachystegia* [41]. The shared genetic variation between *L. alata* and *L. lanceolata*, and the diversification detected in K1 and K3 may be indicative of historical hybridization events (chloroplast capture via introgression) and/or incomplete lineage sorting, similarly as observed for co-occurring *Eucalyptus* species, leading to geographic patterns that reach across species boundaries [11, 17, 37, 40, 42, 43].

In contrast, we observed a clear distinction between the trees of the two sampled sites in West Gabon (GAB1 and GAB4) and those of all other sites. This observation aligns with the earlier findings of a cryptic *L. alata* species in West Gabon, whose existence was inferred from nuclear markers [17]. The haplotypes exclusively found in these populations formed a separate branch in the haplotype network. Indeed, both Gabonese sites clustered separately from the other trees at the two clustering levels (K=3 and K=10), and they were separated in the RAxML phylogenetic-tree (Fig. 1, K2a and K2b). These sites also exhibited much higher genetic distances from other sites. Our results are consistent with the description of a cryptic *L. alata*-WG by Ewédjè et al. (2020) [17], which appears to be endemic. Overall, the high degree of isolation between *L. alata*-WG and *L. alata*/*L. lanceolata* revealed by our analyses, combined with the low mutation rates expected in plastid genome, suggest the existence of ancient lineages in West Gabon. The striking

genetic difference observed between *L. alata*-WG (K2) and *L. alata*, which occur in parapatry, could be attributed to ancient forest fragmentation [44]. This genetic divergence does not appear to be associated with any obvious biogeographical barriers, although, Ewédjè et al. (2020) suggested it might result from reproductive isolation driven by flowering asynchrony [17].

The small distributional area of this cryptic species overlaps with one of the main logging hotspots of Azobé. This situation, combined with a general lack of sustainable timber extraction in Azobé exploitation causes concerns about long-term viability of this cryptic species. Studies on the demographic and evolutionary effects of logging on *Lophira alata* and *L. alata* WG are needed, similar to those conducted for *Distemonanthus benthamianus* and *Erythrophloeum suaveolens* [45, 46]. Such studies may guide policy makers to establish species-specific management and conservation strategies for this major African timber.

Within the large group of *L. alata* and *L. lanceolata* samples, one haplotype was shared between the two species. Even though the other haplotypes were found only in individuals of one species in our sample set, attempting species distinction based on haplotypes may be prone to errors, as the haplotype network shows that closely related haplotypes were found in trees of the two species, and there is no reason to assume that our sampling was exhaustive in terms of occurrence of (rare) haplotypes in both species. Thus, employing the plastome-wide SNPs set as a “SuperBarcode” for identifying Azobé at the species level, similar to the approach suggested by Li et al. (2015) [24], is complicated by the complex sharing and similarity of plastid haplotypes. A possible approach could be to include nuclear markers to help species identification, as pointed out by Hu et al. (2016) [37]. Further investigation into the genomic dynamics and evolutionary history is needed to understand haplotype sharing and genetic interactions between *L. alata*, *L. alata*-WG and *L. lanceolata*. In the case of Azobé, the present method using genome wide pSNPs set combined with the nuclear markers developed by Ewédjè et al. (2020) [17] could be used for this purpose.

Spatial structure of *Lophira alata*

The cluster and haplotype distributions revealed a clear genetic spatial structure in the *Lophira* genus. *Lophira alata*, the commercially important species, presented a clear isolation-by-distance pattern, as indicated by the moderately strong positive correlation between genetic and spatial distances (Fig. 4).

The K1 and K3 clusters exhibited the widest distribution, extending from West to Central Africa. Nonetheless, K1 was more prominent in West Africa, whereas K3 was predominantly restricted to Central Africa. This

distribution pattern suggests that the savannah corridor between Ghana and Benin (Dahomey Gap), may have played a role in shaping the distribution of these clusters. The distinct sub-cluster K3e in the North Congo was not previously detected using nuclear microsatellite markers [17], illustrating the potential power of plastid markers in revealing additional population structure and differentiation [38]. Six out of the 10 sub-clusters were spatially contained in small to moderate sized areas (<550 km) (Fig. 2b). The K2 cluster had the most confined spread, as it was found exclusively in West Gabon.

Half of the sampled individuals ($N=235$) belonged to one of the two very common haplotypes, H3.7 and H3.17 (118 and 117 trees, respectively). Trees bearing these haplotypes are found in distinct regions, with H3.7 spanning far east Cameroon, east Gabon, and the central-western Congo, whereas H3.17 is observed in central-south Cameroon and central-north Gabon (Fig. 3a). Therefore, these two haplotypes show discriminatory potential between these regions but not within them. The latitudinal genetic discontinuities in *L. alata* may be caused by climatic seasonality: rainfall seasonality may cause a phenological delay in flowering [21], which may create a prezygotic barrier for gene flow through pollen. This differentiation is also consistent with genetic discontinuities observed for some tree species in Central Africa, namely, *Greenwayodendron suaveolens* subsp. *suaveolens* var. *suaveolens* and *Irvingia gabonensis*, which are potentially linked to lower Guinean refugia areas during climatic oscillations [44].

Interestingly, the other 173 *L. alata* trees contained a range of 21 haplotypes that occurred at considerably lower frequencies, some of which were rare. Less common haplotypes are less likely to be shared between populations, and their occurrence in populations creates very characteristic haplotypic compositions at certain sites/regions, even when common haplotypes are also present (Fig. 3a). Rare and spatially confined haplotypes, such as H2.2-2.6 in West Gabon and H3.1-3.3 in North Congo, contribute to the genetic mosaic by creating genetic differentiation among spatially proximate individuals. This generates sufficiently large genetic differences to distinguish populations at finer spatial scales in certain regions.

Possible implications for Azobé timber tracing

What does our study imply for the potential of Azobé timber tracing with plastome markers? First, we found a moderately strong correlation between genetic and geographic distances, as well as a clear spatial structure of genetic clusters and haplotypes. The genetic spatial structure observed is in line with the high spatial structure reported in studies that used plastid DNA markers in tree species in the sub-Saharan region [43, 47]. These results suggest that this SNP set fulfills a first and basic

requirement of timber tracing: that of a clear spatial structure based on genetic differences.

Second, the varying spatial scale of haplotypic differences needs to be considered in further development of plastome-based tracing. Some plastid haplotypes were widespread, while others were rare and/or geographically confined. The implication of this is that differentiation varies across haplotypes and regions. The potential for differentiating the sites of origin (tracing) strongly relies on the spatial distribution of the specific haplotype encountered in the query sample. For instance, in a region with five widely distributed haplotypes, tracing may be more challenging than in a region where a single spatially contained haplotype predominates. We give an example: the goal of a forensic query may be to verify the origin claim in trade documentation that timber comes from Gabon. If the forensic sample presents one of the H2.1-H2.6 haplotypes from west Gabon, this implies that the sample indeed is derived from Gabon, because these haplotypes do not occur elsewhere and are many mutations apart from other haplotypes. However, if the forensic samples present one of the two wide-spread haplotypes (H3.7 and H3.17), the answer would be that they could “possibly” be provenant from Gabon, in parts of Cameroon and of the Republic of the Congo. In regions with individuals with broadly spread haplotypes, site differentiation will be limited, depending partly on the composition of haplotypes. In this case, forensic timber samples can be linked to the wider region of origin, but not the specific forest concession or regions within a country. This limited fine-scale population differentiation is attributed to low level of polymorphism in the chloroplast genome compared to the nuclear genome [13].

Third, although the sampling of this study covers a big part of logging hotspots in Central Africa, the reference database for Azobé may need to be extended to other key logging regions, including central Gabon, the Democratic Republic of the Congo, and protected areas. The extension of the genome-wide SNPs set approach, proposed here, allows for the inclusion of new individuals, and therefore allows for the detection of new genetic profiles, without assortment bias on preselecting region(s) in the genome. These advancements not only refine genetic spatial patterns for timber tracing by broadening pSNP detection and revealing new genetic profiles in illegal logging hotspots, but also encompass the genetic diversity and profiles of areas where logging should be restricted or prohibited. To achieve comprehensive sample collection, collaborative efforts in sample collection and analysis should be prioritized. Ongoing efforts by the World Forest ID [48] to extend and centralize sample collections and make these available to users will help realize this step.

Next steps

The good performance of cambium samples turned out critical for reference dataset development and for tests of the tracing method principle. However, testing the performance of heartwood samples (which are commercialized tissue) will be a first necessity, especially because the DNA from heartwood is known to be of poor quality and highly degraded. Furthermore, a full evaluation and validation of the potential of plastome tracing requires conducting (self-) assignment tests, e.g., using random forest analyses [49, 50] or a Nearest Neighbour Approach [51], as well as assignment tests of blind samples (i.e., trees of a priori unknown origin).

The use of plastid-wide SNP haplotypes

We showed that informative SNP sets can be obtained from focused analysis of the plastid genome which can be employed for studies on spatial genetic structure of tropical timber species. The use of such approach includes origin differentiation, evolution (e.g., insights on the evolution of East Asian forests obtained from plastomes of *Perseeae* [52]) and conservation. Although our study primarily generated nuclear DNA sequencing data (data not shown), the sequencing depth for nuclear DNA was generally insufficient to reliably call SNPs due to low coverage. In contrast, plastid genomes, present in multiple copies per cell [10], achieved extensive coverage, consistent with previous studies showing higher success rates for chloroplast DNA than for nuclear DNA in wooden tissues [5, 6, 53]. The greater number of plastomes, which are much shorter than nuclear genomes, results in a lower sequencing depth required to reach sufficient sequence coverage (i.e., number of read copies per SNP), although this advantage was relatively small in our study, possibly because the number of plastids per cambium cell is lower than the number of chloroplasts in photosynthetically active leaf cells. Nonetheless, as a lower sequencing depth is needed, sequencing costs are lower (in the present study, ~30 euros per sample). Cost efficiency is particularly important for countries most affected by the illegal timber trade, reducing financial burden of such analyses in an operational setting. Moreover, working with whole plastome sequences enables the incorporation of new reference samples into the dataset, allowing the inclusion of new reference samples and the detection of new variants. In this way, the development of markers will continue alongside the growth of databases and the use of existing (wood or leaf) samples, as we have done in this study. This broadens reference area coverage and potentially reveals new genetic spatial (sub)-structures of the studied species.

Conclusion

We developed an informative SNP set that revealed clear spatial structure based on samples collected across most of the logging range. We detected a moderately strong correlation between genetic and spatial distance, with haplotypes being spatially structured at varying spatial scales. However, species identification based solely on the plastome can be inconclusive, as the pDNA haplotypes of *Lophira alata* were distinct from those of a cryptic species but not from those of *L. lanceolata*, possibly because of hybridization and chloroplast capture in contact zones.

In conclusion, our results provide a first step toward using plastome haplotypes based on polymorphic SNP datasets to verify claims of geographic origin for Azobé timber, though not for species-level differentiation. This approach is practical and efficient, thanks to its low sequencing costs and the potential for easily expanding the reference database through collaboration with other laboratories.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-025-06287-2>.

Supplementary Material 1: List of samples

Supplementary Material 2: Optimized cetyltrimethyl ammonium bromide (CTAB) with additional cleaning steps

Supplementary Material 3: Clustering patterns are robust with respect to the threshold for missing SNP data. The lower completeness limit varied between 70 and 90% completeness of data per SNP (across individuals) shown on the PCA clusters

Supplementary Material 4: K-means clustering based values of Bayesian Information Criterion

Supplementary Material 5: Haplotype frequencies per species. * H1.10 is shared between *L. alata* and *L. lanceolata*

Acknowledgements

We gratefully acknowledge the support provided by the Dutch Research Council (NWO-TTW-OTP-16427), the Alberta Mennega Foundation, FSC International, and WorldForestID. We extend our thanks to the collaborating timber companies and their field teams, botanists, as well as the local communities for their reception and assistance with the fieldwork. Special thanks go to our colleagues at partner institutes, including the University of Dschang, Marien Ngouabi University, IRAF/CENAREST, the National Herbarium of Gabon, and IRET/CENAREST for their invaluable work and engagement.

Author contributions

Conceptualization: B.R.V.M.-S., M.J.M.S., A.G., P.A.Z.; Data collection: B. R.V.M.-S., L.E.B., G.U.D.B., J.C.U.C.-M., N.L.E.O., M.G.-E., O.J.H., S.J., J.J.L., D.-M.M.F.M., C.G.M., R.M.D.N., D.N.B., S.N.T., M.T.T., B.B.L.T.-P., H.T.Z., S. J., P.T.Z.; Data curation: B. R.V.M.-S., D.E., N.V., O.J.H.; Formal Analysis: B. R.V.M.-S. and N.V.; Funding acquisition: B. R.V.M.-S., L.E.B., M.J.M.S., A.G., P.A.Z.; Investigation: B. R.V.M.-S.; Methodology: B. R.V.M.-S., M.J.M.S., A.G., P.A.Z., O.J.H.; Supervision: M.J.M.S., A.G., P.A.Z.; Visualization: B. R.V.M.-S.; Writing—original draft: B.R.V.M.-S.; Writing—review & editing: B.R.V.M.-S., L.E.B., G.U.D.B., J.C.U.C.-M., N.L.E.O., M.G.-E., A.G., O.J.H., S.J., J.J.L., D.-M.M.F.M., C.G.M., R.M.D.N., D.N.B., N.V., S.N.T., M.T.T., B.B.L.T.-P., H.T.Z., P.T.Z., M.J.M.S., and P.A.Z. All authors read and approved the final manuscript.

Funding

This study was supported by the Dutch Research Council (NWO-TTW-OTP-16427). Additional fieldwork support was received from the Alberta Mennega Foundation, FSC International and WorldForestID.

Data availability

The leaf and cambium samples (Additional file 1) used in this study are part of the Wood Sample Collection maintained by the Dendrochronology Laboratory of the Forest Ecology and Forest Management Group at Wageningen University & Research. The collection is housed at Droevendaalsesteeg 3, 6708 PB, Wageningen, The Netherlands. For inquiries regarding access to the samples, please contact Dr. Mathieu Decuyper at +31 (0) 317 486 195 or via email at mathieu.decuyper@wur.nl. The datasets generated during and/or analyzed during the current study are available in the Sequence Read Archive (SRA)/NCBI repository accession number: PRJNA1150388; and in the European Variation Archive (EVA) at EMBL-EBI under accession number PRJEB78866. The datasets generated during and/or analyzed during the current study are also available (except for commercial purposes) from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

All samples were obtained from forest concessions with the authorization, consent, and assistance of the respective managers. All necessary permits were obtained from the relevant authorities and are listed below. Sampling and data use authorizations by competent authorities do samples collected between 2019 and 2022: *Cameroon*: Research Permit No. 00000116/MINRESI/B00/C00/C10/C12 (Yaoundé, 09 Sep 2019); Research Permit No. 000066/MINRESI/B00/C00/C10/C12 (Yaoundé, 07 Jun 2021); Scientific research permit No. 2144 PRBS/MINFOR/SETAT/SG/DFAP/SDVEF/SC/NGY (Yaoundé, 23 Jul 2021); ABS Permit 00010/MINEPDED/CNA/NP-ABS/ABS-FP (Yaoundé, 03 Dec 2021); PIC Decision No. 00013/D/MINEPDED/CNA of 03 Dec 2021. *Gabon*: Research authorization No. AR017/21/MESRTTENCFC/CENAREST/CG/CST/CSAR. *Congo*: samples collected under the Marien Ngouabi University's research authorization: N 004/UMNG.VPRC.DCRI.SMRP and N 005/UMNG.VPRC.DCRI.SMRP. The leaf and cambium samples from Université Libre de Bruxelles used in this study were collected prior to the entry into force of the Nagoya Protocol on Access and Benefit-Sharing (ABS) on October 12, 2014. As such, these samples are not subject to the requirements outlined in the protocol.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Forest Ecology and Forest Management, Wageningen University and Research, Wageningen, The Netherlands

²Department of Ecology, Evolution and Environmental Biology, Columbia University, New York, USA

³Laboratory of Biodiversity and Ecosystems and Environmental Management, Faculty of Science and Technology, Marien Ngouabi University, Brazzaville, Congo

⁴Wageningen Environmental Research, Wageningen University and Research, Wageningen, The Netherlands

⁵Plant Breeding, Wageningen University and Research, Wageningen, The Netherlands

⁶Institute for Research in Tropical Ecology (IPHAMETRA IRET/ CENAREST), Herbar National du Gabon, Institut de Pharmacopée et de Médecine Traditionnelle, Libreville, Gabon

⁷Evolutionary Biology and Ecology Unit, CP 160/12, Faculté des Sciences, Université Libre de Bruxelles, Brussels, Belgium

⁸Department of Ecosystem Management, Climate and Biodiversity - Institute of Silviculture, University of Natural Resources and Life Sciences (BOKU), Vienna, Austria

⁹Institute for Research in Tropical Ecology (IRET/CENAREST), Libreville, Gabon

¹⁰Institute for Agronomic and Forestry Researches (IRAF/CENAREST), Libreville, Gabon

¹¹Laboratory of Environmental Geomatics, Department of Forestry, Faculty of Agronomy and Agricultural Sciences, University of Dschang, Dschang, Cameroon

Received: 12 July 2024 / Accepted: 20 February 2025

Published online: 12 March 2025

References

1. Hoare A. Tackling Illegal Logging and the Related Trade - What Progress and Where Next? [Internet]. 2015 [cited 2023 Apr 13]. Available from: <https://www.chathamhouse.org/sites/default/files/publications/research/20150715IllegalLoggingHoareFinal.pdf>
2. The European Parliament And The Council Of The European Union. Regulation (EU) 2023/1115 Of The European Parliament And Of The Council of the 31 May 2023 on the making available on the Union market and the export from the Union of certain commodities and products associated with deforestation and forest degradation and repealing Regulation (EU) No 995/2010 [Internet]. Official Journal of the European Union. 2023 [cited 2023 Aug 21]. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32023R1115>
3. Dormontt EE, Boner M, Braun B, Breulmann G, Degen B, Espinoza E, et al. Forensic timber identification: it's time to integrate disciplines to combat illegal logging. *Biol Conserv*. 2015;191:790–8.
4. Low MC, Schmitz N, Boeschoten LE, Cabezas JA, Cramm M, Haag V, et al. Tracing the world's timber: the status of scientific verification technologies for species and origin identification. *IAWA J*. 2022;44(1):63–84.
5. Rachmayanti Y, Leinemann L, Gailing O, Finkeldey R. Extraction. Amplification and characterization of wood DNA from Dipterocarpaceae. *Plant Mol Biol Rep*. 2006;24(1):45–55.
6. Rachmayanti Y, Leinemann L, Gailing O, Finkeldey R. DNA from processed and unprocessed wood: factors influencing the isolation success. *Forensic Sci International: Genet*. 2009;3(3):185–92.
7. Tsumura Y, Kado T, Yoshida K, Abe H, Ohtani M, Taguchi Y, et al. Molecular database for classifying Shorea species (Dipterocarpaceae) and techniques for checking the legitimacy of timber and wood products. *J Plant Res*. 2011;124(1):35–48.
8. Akhmetzyanov L, Copini P, Sass-Klaassen U, Schroeder H, de Groot GA, Laros I, et al. DNA of centuries-old timber can reveal its origin. *Sci Rep*. 2020;10(1):20316.
9. Henry RJ, editor. Plant diversity and evolution: genotypic and phenotypic variation in higher plants [Internet]. 1st ed. UK: CABI Publishing; 2005 [cited 2023 Aug 8]. Available from: <http://www.cabidigitallibrary.org/doi/book/http://doi.org/10.1079/9780851999043.0000>
10. Mader M, Pakull B, Blanc-Jolivet C, Paulini-Drewes M, Bouda ZHN, Degen B, et al. Complete Chloroplast genome sequences of four Meliaceae species and comparative analyses. *Int J Mol Sci*. 2018;19(3):701.
11. Fahey PS, Fowler RM, Udovicic F, Cantrill DJ, Bayly MJ. Use of plastid genome sequences in phylogeographic studies of tree species can be misleading without comprehensive sampling of co-occurring, related species. *Tree Genet Genomes*. 2021;17(6).
12. Mascarello M, Amalfi M, Asselman P, Smets E, Hardy OJ, Beeckman H et al. Genome skimming reveals novel plastid markers for the molecular identification of illegally logged African timber species. *PLoS ONE*. 2021;16(6 June 2021).
13. Blanc-Jolivet C, Liesebach M. Tracing the origin and species identity of Quercus robur and Quercus petraea in Europe: a review. *Silvae Genetica*. 2015;64(4):182–93.
14. Eba'a Atyi R, Hiol Hiol F, Lescuyer G, Mayaux P, Defourny P, Bayol N et al. The Forests of the Congo Basin: State of the Forests 2021 [Internet]. Center for International Forestry Research (CIFOR); 2022 [cited 2023 Aug 7]. Available from: <https://www.cifor.org/knowledge/publication/8700>
15. Plouvier D. Short overview of the situation of tropical moist forests and forest management in central Africa and markets for African timber. *J Trop Forests Forestry Sustainable Dev*. 1997;16(3):42–9.
16. Piñeiro R, Staquet A, Hardy OJ. Isolation of nuclear microsatellites in the African timber tree lophira Alata (Ochnaceae) and cross-amplification in L. lanceolata. *Appl Plant Sci*. 2015;3(10):1500056.
17. Ewédjé EEBK, Jansen S, Koffi GK, Staquet A, Piñeiro R, Essaba RA, et al. Species delimitation in the African tree genus lophira (Ochnaceae) reveals cryptic genetic variation. *Conserv Genet*. 2020;21(3):501–14.

18. Engone Obiang NL, Ngomanda A, White L, Jeffery KJ, Chézeaux E, Picard N. Un modèle de croissance pour L'azobé, lophira Alata, Au Gabon. *Bois Trop*. 2012;314(314):65.
19. Global Biodiversity Information Facility. Global Biodiversity Information Facility. 2024 [cited 2024 Jul 10]. www.gbif.org. Available from: https://www.gbif.org/occurrence/map?taxon_key=3695610
20. Aubréville A. La Flore Forestière de La Côte D'Ivoire. Volume 3, 15th ed. Paris: Centre Technique Forestier Tropical; 1959. p. 334.
21. Ouédraogo DY, Hardy OJ, Doucet JL, Janssens SB, Wieringa JJ, Stoffelen P, et al. Latitudinal shift in the timing of flowering of tree species across tropical Africa: insights from field observations and herbarium collections. *J Trop Ecol*. 2020;36(4):159–73.
22. Blanc-Jolivet C, Mader M, Bouda HN, Massot M, Dainou K, Yene G, et al. Development of new SNP and INDEL loci for the valuable African timber species lophira Alata. *Conserv Genet Resour*. 2021;13(1):85–7.
23. Besnard G, Hernández P, Khadari B, Dorado G, Savolainen V. Genomic profiling of plastid DNA variation in the mediterranean Olive tree. *BMC Plant Biol*. 2011;11(1):80.
24. Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S. Plant DNA barcoding: from gene to genome. *Biol Rev*. 2015;90(1):157–66.
25. Biwolé A, Bourland N, Dainou K, Doucet JL. Définition du profil écologique de L'azobé, lophira Alata, Une espèce ligneuse Africaine de Grande importance: synthèse bibliographique et perspectives pour des recherches futures. *Biotechnol Agron Soc Environ*. 2012;16:217–28.
26. Dumolin S, Demesure B, Petit RJ. Inheritance of Chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. *Theor Appl Genet*. 1995;91:1253–6.
27. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods*. 2012;9(4):357–9.
28. Tello D, Gil J, Loaiza CD, Riascos JJ, Cardozo N, Duitama J. NGSEP3: accurate variant calling across species and sequencing protocols. *Bioinformatics*. 2019;35(22):4716–23.
29. R Core Team. The R Project for Statistical Computing [Internet]. 2021 [cited 2021 Jan 1]. Available from: <https://www.r-project.org/>
30. DeRaad D. SNPfiltR [Internet]. 2023 [cited 2024 Apr 29]. Available from: <https://cran.r-project.org/web/packages/SNPfiltR/index.html>
31. Jombart T. Adegnet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*. 2008;24(11):1403–5.
32. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006;22(21):2688–90.
33. Bandelt HJ, Forster P, Rohl A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*. 1999;16(1):37–48.
34. University of Otago Popart [Internet]. [cited 2024 Jul 12]. Available from: <https://popart.maths.otago.ac.nz/>
35. Nei M. Genetic distance between populations. *Am Nat*. 1972;106(949):283–92.
36. Hu H, Al-Shehbaz IA, Sun Y, Hao G, Wang Q, Liu J. Species delimitation in *Orychophragmus* (Brassicaceae) based on Chloroplast and nuclear DNA barcodes. *Taxon*. 2015;64(4):714–26.
37. Hu H, Hu Q, Al-Shehbaz IA, Luo X, Zeng T, Guo X, et al. Species delimitation and interspecific relationships of the genus *Orychophragmus* (Brassicaceae) inferred from whole Chloroplast genomes. *Front Plant Sci*. 2016;7:1826.
38. Mascarello M, Lachenaud O, Amalfi M, Smets E, Hardy OJ, Beeckman H et al. BT Šiler editor 2023 Genetic characterization of a group of commercial African timber species: from genomics to barcoding. *PLoS ONE* 18 4 e0284732.
39. Duminil J, Kenfack D, Viscosi V, Grumiau L, Hardy OJ, Meliaceae. *Mol Phylogenet Evol*. 2012;62(1):275–85.
40. Migliore J, Kaymak E, Mariac C, Couvreur TLP, Lissambou BJ, Piñeiro R, et al. Pre-Pleistocene origin of phylogeographical breaks in African rain forest trees: new insights from Greenwayodendron (Annonaceae) phylogenomics. *J Biogeogr*. 2019;46(1):212–23.
41. Boom AF, Migliore J, Kaymak E, Meerts P, Hardy OJ. Plastid introgression and evolution of African Miombo Woodlands: new insights from the plastome-based phylogeny of brachystegia trees. *J Biogeogr*. 2021;48(4):933–46.
42. Xu B, Wu N, Gao XF, Zhang LB. Analysis of DNA sequences of six Chloroplast and nuclear genes suggests incongruence, introgression, and incomplete lineage sorting in the evolution of Lespedeza (Fabaceae). *Mol Phylogenet Evol*. 2012;62(1):346–58.
43. Duminil J, Brown RP, Ewédjè EEB, Mardulyn P, Doucet JL, Hardy OJ. Large-scale pattern of genetic differentiation within African rainforest trees: insights on the roles of ecological gradients and past climate changes on the evolution of *Erythrophleum* spp (Fabaceae). *BMC Evol Biol*. 2013;13(1).
44. Hardy OJ, Born C, Budde K, Dainou K, Dauby G, Duminil J, et al. Comparative phylogeography of African rain forest trees: A review of genetic signatures of vegetation history in the Guineo-Congolian region. *CR Geosci*. 2013;345(7):284–96.
45. Duminil J, Dainou K, Kaviriri DK, Gillet P, Loo J, Doucet JL, et al. Relationships between population density, fine-scale genetic structure, mating system and pollen dispersal in a timber tree from African rainforests. *Heredity*. 2016;116(3):295–303.
46. Hardy OJ, Delaide B, Hainaut H, Gillet JF, Gillet P, Kaymak E, et al. Seed and pollen dispersal distances in two African legume timber trees and their reproductive potential under selective logging. *Mol Ecol*. 2019;28(12):3119–34.
47. Lompo D, Vinceti B, Konrad H, Gaisberger H, Geburek T. Phylogeography of African locust bean (*Parkia biglobosa*) reveals genetic divergence and spatially structured populations in West and central Africa. *J Hered*. 2018;109(7):811–24.
48. World Forest ID, World Forest ID. 2021 [cited 2024 Feb 28]. World Forest ID - WFID. Available from: <https://worldforestid.org>.
49. Wright MN, Ziegler A. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *J Stat Softw* [Internet]. 2017 [cited 2024 Jan 3];77(1). Available from: <http://www.jstatsoft.org/v77/i01/>
50. Sylvester EVA, Bentzen P, Bradbury IR, Clément M, Pearce J, Horne J, et al. Applications of random forest feature selection for fine-scale genetic population assignment. *Evol Appl*. 2018;11(2):153–65.
51. Degen B, Blanc-Jolivet C, Stierand K, Gillet E. A nearest neighbour approach by genetic distance to the assignment of individual trees to geographic origin. *Forensic Sci International: Genet*. 2017;27:132–41.
52. Xiao TW, Yan HF, Ge XJ. Plastid phylogenomics of tribe perseeae (Lauraceae) yields insights into the evolution of East Asian subtropical evergreen broad-leaved forests. *BMC Plant Biol*. 2022;22(1):32.
53. Tnah LH, Lee SL, Ng KKS, Bhassu S, Othman RY. DNA extraction from dry wood of neobalanocarpus heimii (Dipterocarpaceae) for forensic DNA profiling and timber tracking. *Wood Sci Technol*. 2012;46(5):813–25.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.