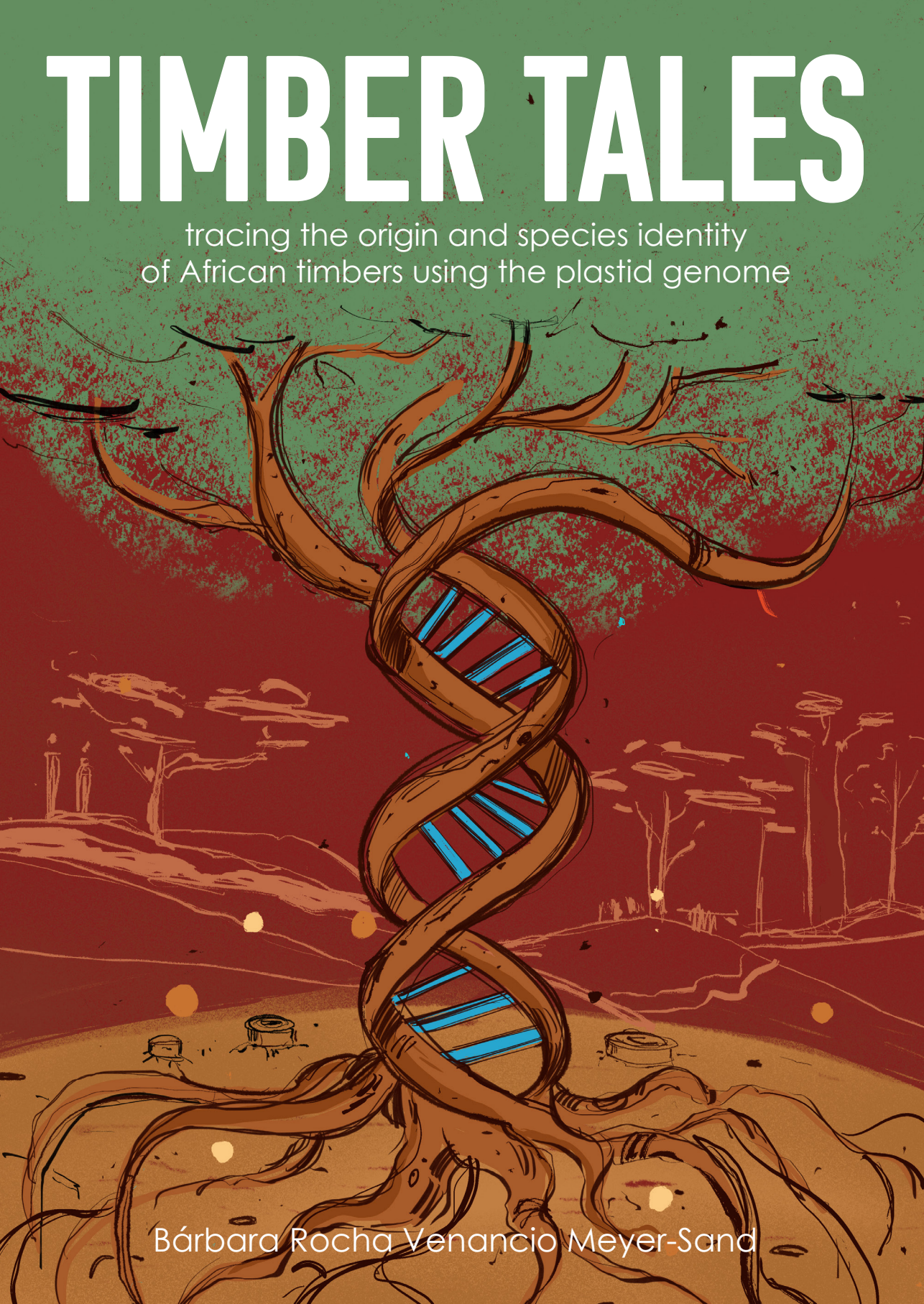
TIMBER TALES

tracing the origin and species identity
of African timbers using the plastid genome



Bárbara Rocha Venancio Meyer-Sand

Propositions

1. The spatial resolution of timber tracing with plastomes is determined by the geographic distribution and frequency of common and rare haplotypes. (this thesis)

2. The most significant limitation in developing plastome tracing lies in the effectiveness of statistical tools. (this thesis)

3. Decolonizing science for equitable, sustainable, and locally relevant knowledge creation demands embracing diverse insights and actively engaging partners from the start.

4. To become independent researchers, we must first learn how to rely on our peers.

5. Racism and discrimination thrive in silence.

6. The resilience of systems stems from their diversity.

Propositions belonging to the thesis, entitled

Timber tales: tracing the origin and species identity of African timbers using plastid genomes

Barbara Rocha Venancio Meyer-Sand

Wageningen, 18 June 2024

Timber tales:

tracing the origin and species identity of African timbers using the plastid genome

Bárbara Rocha Venancio Meyer-Sand

Thesis committee

Promotor

Prof. Dr P.A. Zuidema

Personal chair at the Forest Ecology and Forest Management Group

Wageningen University & Research

Co-promotors

Dr M.J.M. Smulders

Business Unit Manager, Plant Breeding

Wageningen University & Research

Dr G.A. de Groot

Senior researcher, Animal Ecology

Wageningen University & Research

Other member

Prof. Dr B.J. Zwaan, Wageningen University & Research

Dr E. Honorio Coronado, Kew Gardens, London, United Kingdom

Dr J. Duminil, Institut de Recherche pour le Développement (IRD), Montpellier, France

Dr T. Peelen, Douane Laboratorium/Customs Laboratory, Amsterdam

This research was conducted under the auspices of the Graduate School Production Ecology & Resource Conservation (PE&RC)

Timber tales:

tracing the origin and species identity of African timbers using the plastid genome

Bárbara Rocha Venancio Meyer-Sand

Thesis

submitted in fulfillment of the requirements for the degree of doctor

at Wageningen University

by the authority of the Rector Magnificus,

Prof. Dr C Kroeze,

in the presence of the

Thesis Committee appointed by the Academic Board

to be defended in public

on Tuesday 18 June 2024

at 11 a.m. in the Omnia Auditorium.

Bárbara Rocha Venancio Meyer-Sand

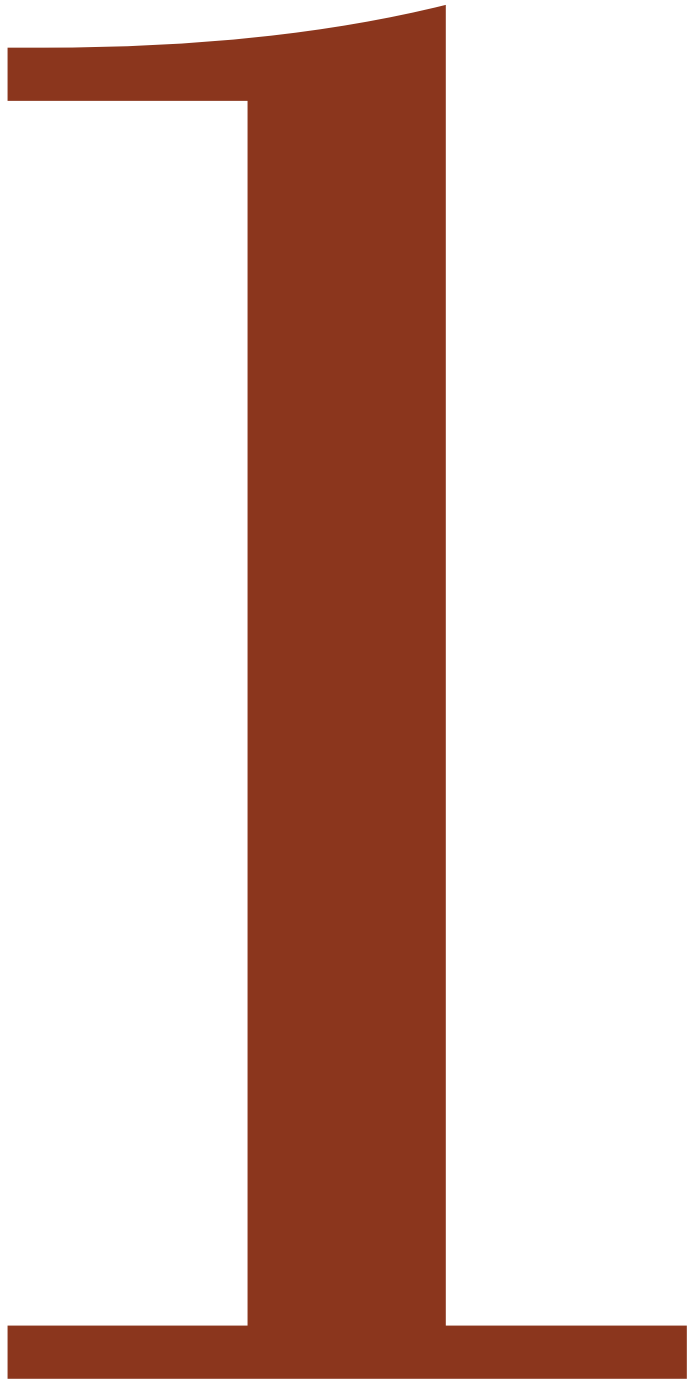
Timber tales: tracing the origin and species identity of African timbers using the plastid genome

193 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2024)
With references, with summary in English

DOI : https://doi.org/10.18174/658494

Table of contents



Chapter 1

General Introduction

General Introduction

1.1 ILLEGAL TIMBER TRADE

The illegal logging and trade of timber significantly hampers sustainable forest management and conservation and undermines legal timber exploitation. Illegal timber accounts for over half of the timber export of major timber-producing countries in Asia and Africa (Hoare, 2015). Hoare (2015) estimated a volume of 60 million m^3 of illegal timber from tropical countries to the 10 countries that import and consume the most. With a value of USD 20 billion annually, illegal timber products represent a major sector within the organized crime (INTERPOL, 2023). Effective policies and enforcement strategies to combat this global crime are urgently needed.

The most commonly reported illegal activities along the timber supply chain include the logging of trees or the transport, processing, export and sale of timber without the required permits. This can concern either the absence of permits, the use of permits that cover a different region than from where timber was actually sourced, or the use of falsified permits (i.e. that claim a false species and/or origin). Trade routes of illegal timber are often complex, involving corruption and laundering during the value chain, e.g., by inserting illegal timber into batches of legally sourced timber (INTERPOL, 2020). The complexity of the timber trade routes and the admixture of timber from different species or geographic origins into wooden products provide opportunities to use and sell illegal timber and represent substantial hurdles to verifying claims of species and origin claims.

1.2 CONVENTIONS AND REGULATIONS TO FIGHT ILLEGAL TIMBER TRADE

Various regulations aim to prevent the entry of illegally sourced wood into markets. Important examples include the US Lacey Act (2008), the European Union Timber Regulation (EUTR, 2010), and the forthcoming European Union Regulation on deforestation-free products (EUDR)(The European Parliament And The Council Of The European Union, 2023), set to replace the EUTR in 2025. The EUTR and EUDR mandate transparency in the timber value

chain, requiring operators and traders to perform due diligence. This process includes collecting data, assessing, and mitigating risks, necessitating real-time product information such as detailed descriptions, quantities, and – in the case of the EUDR – the geolocation (polygons of origin) of the production site. The EUDR further requires evidence confirming that products are deforestation-free and comply with the origin country's laws. The EUDR does not only apply to wooden products, but also to cocoa, coffee, soy, rubber, beef and palm oil (The European Parliament And The Council Of The European Union, 2023). Another key international legislation to halt illegal timber trade is the Convention on International Trade in Endangered Species of Wild Flora and Fauna (1975, CITES), which prohibits the trade of timber from threatened tree species.

So far, these legislative initiatives to curb trade in illegal timber have not achieved the intended goals. This lack of success is partly due to a significant gap between the intentions of legislation and their enforcement. The verification of timber names and geographic sources as claimed in trade documents is problematic due to fraudulent declarations and document forgery. This problem is persistent because of the difficulties in identifying forged or fraudulently declared documents (Kleinschmit et al., 2016).

Consequently, there is a pressing need for innovative methods that allow authentication of traded timber. Developing and adopting independent tools that leverage the intrinsic properties of timber could significantly bridge the regulatory gap, facilitating better governance and law enforcement (The European Parliament And The Council Of The European Union, 2023; Low et al., 2022). To help bridge this regulatory gap, the EUDR mentions the application of forensic techniques, such as genetic and chemical profiling. Similarly, the United Nations Office on Drugs and Crime (UNODC) has recognized the global need for forensic timber identification tools (Dormontt et al., 2015). In recent years, researchers across the globe are addressing this challenge, by developing and testing a range of methods to improve the traceability of timber products. This thesis contributes to reaching that goal.

1.3 FORENSIC METHODS APPLIED TO TIMBER LEGALITY VERIFICATION

Morphological, chemical and genetic methods may help establishing the species and geographic provenance of timber based on intrinsic properties of the wood. Since these are

intrinsic wood characteristics, they cannot be tampered with (Dormontt et al., 2015; Low et al., 2022). Some of these methods are well established as forensic tools for some major agricultural commodities or traded wildlife, or in human forensic contexts.

A range of forensic methods based on intrinsic properties of the wood exists, as illustrated in Figure 1.1. These methods use visual (wood anatomy, machine vision) (Akhmetzyanov et al., 2020b), chemical (NIRS, mass spectrometry, stable isotopes) (Vlam et al., 2018; Boeschoten et al., 2023), and genetic properties (for a review on forensic methods see Dormontt et al., 2015). Yet, to date these methods have rarely been used to verify timber legality claims in practice because of lacking or incomplete reference databases (Low et al., 2022). Although the visual, chemical and/or genetic properties of a piece of timber can be measured precisely, the forensic evidence is derived from a comparison with reference databases of these characteristics for the species concerned across its distributional area. Thus, for forensic tests to yield powerful evidence, comprehensive reference databases are required.

For the identification of the species or genus of traded timber, wood anatomy and DNA barcoding (genetics) are the most applied methods at this moment. Reference data for these techniques are available for 100% and 86% of the 322 most commonly traded tree taxa, respectively (Low et al., 2022). By contrast, when considering techniques to verify the geographic origin of the timber, chemical and genetic reference data are available for only a quarter of these 322 taxa, with spatial coverage of these reference datasets often being incomplete (Low et al., 2022). The common employment of genetic methods (for taxon and origin identification), and anatomical methods (for taxon identification) (Low et al., 2022) is likely in part due to the fact that they can be used for both species and origin detection goals simultaneously (e.g., Honorio Coronado et al., (2020)).

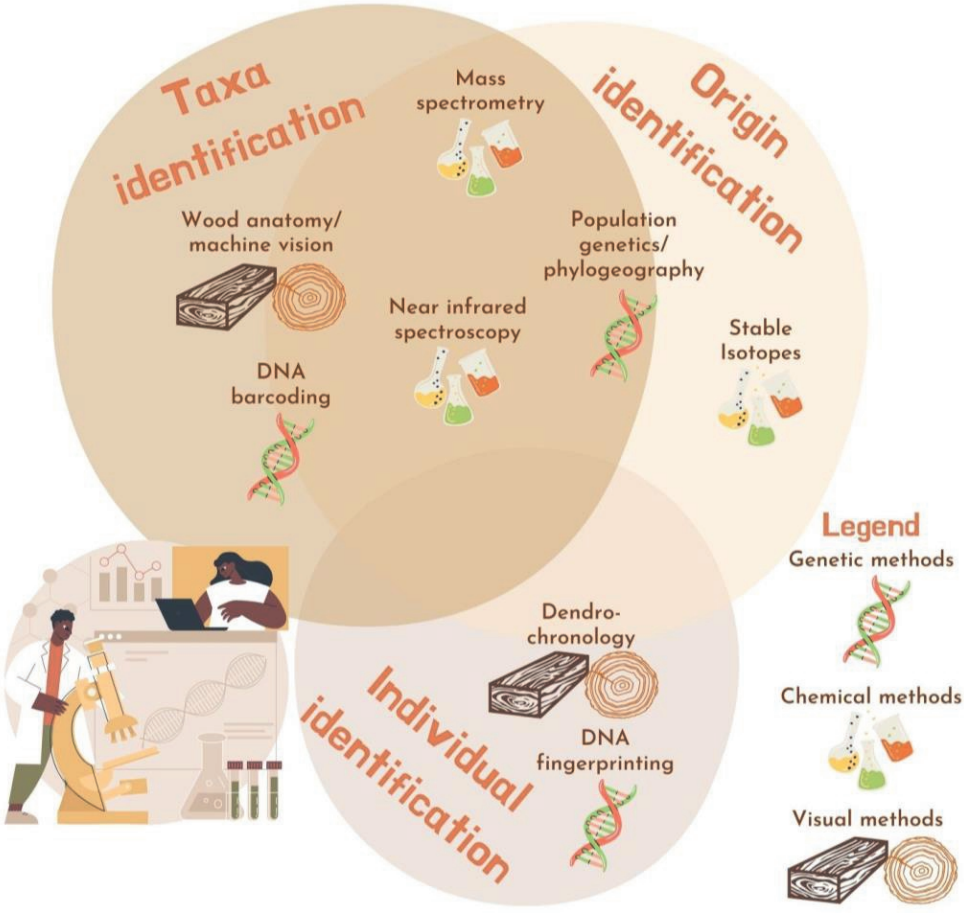


Figure 1.1 : Overview of Timber Tracing Methods that can be applied in forensic tests of claimed timber origin. Methods are grouped by aims (taxon identification, origin identification, and individual identification). Genetic methods include DNA barcoding, population genetics/phylogeography; chemical methods include mass spectrometry, near-infrared spectroscopy, stable isotopes; visual methods include wood anatomy, machine vision, and dendrochronology. Figure based on (Table 2, Dormontt et al., 2015)

1.4 TIMBER DETECTIVES

To overcome barriers in the development and operationalization of timber tracing techniques reference collections need to be expanded and techniques need to be further developed, rigorously tested and compared for performance. Increasingly, researchers are joining forces

in international efforts to realize this. Collaborative initiatives include the Global Timber Tracking Network ("Global Timber Tracking Network," 2016), World Forest ID (World Forest ID, 2021), expertise centers such as the Germany's Thünen Centre of Competence on the Origin of Timber, and projects such as the Timtrace project to which this PhD thesis belongs (www.timtrace.nl). These initiatives often involve international collaborations among scientists of different disciplines, NGOs and authorities. This collective momentum underscores a growing consensus: no single method suffices to address the complexities of timber frauds across species and continents (Dormontt et al., 2015; Low et al., 2022). A synergistic approach combining various disciplines is crucial for building strong evidence against illegal logging.

The Timtrace project based at Wageningen University in the Netherlands aims to develop and test techniques to trace back the geographic origin and species of tropical timbers. The project contributed to overcoming the following main barriers in the operationalization of timber tracing: the low spatial resolution of chemical tracing techniques, the poor DNA quality obtained from wood, unclarity about the power of organelle DNA markers, the lack of standardized statistical methods, and the absence of international reference databases. It targeted the following timbers: Tali (*Erythrophleum* spp.), Azobé (*Lophira alata*) from Central Africa, and Meranti (*Shorea* spp.) from South-East Asia (not included in this thesis). The project conducted feasibility studies, evaluated statistical approaches, and developed new tracing techniques. Two PhD theses were produced as part of this project: Laura Boeschoten developed and tested chemical tracing techniques, including stable isotopes and multi-element analyses, resulting in a PhD thesis entitled "Chemical methods for timber tracing" (Boeschoten, 2023). This thesis forms the second PhD thesis of the Timtrace project, with a focus on developing genetic timber tracing tools. The Timtrace project also aimed to compare and integrate chemical and genetic techniques, which was done in a joint chapter, published in both PhD theses.

Crucial for the results of the Timtrace project has been the collaboration with universities and research institutes in Cameroon, Gabon, Congo Republic and Indonesia. In addition, key stakeholders in the Netherlands were also involved, including timber industry, environmental NGOs, certifying bodies, and competent authorities.

1.5 GENETIC METHODS

DNA genotyping is the general term that refers to mapping (any) genetic differences among individuals and species. This allows to determine population structure, quantify relatedness among individuals and identify parent-offspring relationships, among others. It typically involves calculating pairwise genetic similarities or their complement, pairwise genetic distances (Lowe et al., 2010; Jolivet and Degen, 2012; Hung et al., 2017; Dormontt et al., 2020). The technique is very powerful for the individuals included in the dataset or for comparing one new genotype with those genotyped before. Assigning a species name or a distance to a population can only be done if sufficient individuals have been genotyped to make a representative and meaningful genetic comparison. Hence, using genotyping with the goal to determine the correct species (Ogden et al., 2009; Höltken et al., 2012; Ng et al., 2016) or assign wood samples to their geographical origin (Smulders et al., 2008; Tnah et al., 2009; Lowe et al., 2010; Höltken et al., 2012; Jolivet and Degen, 2012; Koopman et al., 2012; Degen et al., 2013; Ng et al., 2017; Vlam et al., 2018; Low et al., 2022) require appropriate genetic markers and the existence of reference databases. Phylogenetics and population genetics analyses are then applied to distinguish individuals' origin at different scales; these techniques are currently tested in several studies (Dormontt et al., 2015; for a review see Low et al., 2022).

A specific form of genotyping that may be applied for genetic timber tracing is DNA barcoding. It focusses specifically on genetic differences in a small set of predefined DNA regions. These regions are conserved across taxa, which enables building very complete reference databases for these stretches of DNA relatively fast. DNA barcoding is applied to differentiate species ("International Barcode of Life Illuminate Biodiversity," n.d.) and it is currently the most established forensic technique in plants, used to identify the species or genus. However, it has limited application to differentiate within species or closely related taxa (Dormontt et al., 2015; Jiao et al., 2019; Low et al., 2022) as it does not contain sufficient information

Levels of differentiation among populations or clusters of plants within a species are evaluated based on genetic differences. A genetic difference used in such a way is called a genetic marker. A marker may be common or rare. Rare markers may be present in some populations or species and not in others. These so called private alleles can be used as a

diagnostic marker. This may seem ideal, but determining whether a marker is unique is not trivial (as it may occur in a tree that was not sampled, even in populations with proper coverage in the reference dataset). Common markers are also informative, as differences in their frequencies are used to determine genetic distances.

The type and number of markers, as well as the strategy to select them, affect the spatial resolution as well as the statistical power of assignment systems. Two main types of genetic markers are being used routinely. Microsatellite markers, also known as Simple Sequence Repeats (SSRs), are short, repetitive DNA sequences widely distributed throughout genomes. These markers were the most widely used in genetic studies in the past decades, including applications for forensic purposes (Smulders et al., 2008; Ng et al., 2016; Vlam et al., 2018; Low et al., 2022). Their extensive use in the past is attributed to their high variability among individuals, making them invaluable for genetic fingerprinting, population genetics, and breeding research. However, detection of microsatellites requires DNA fragments longer than approximately 150 bp (basepairs).

DNA isolated from wood is often highly degraded and available in low quantities, leading to low amplification rates. Consequently, some samples may entirely fail in the DNA amplification step, significantly hindering the yield of DNA from wood. This challenge is a critical issue in genetic analysis of timber, as highlighted in studies by Degen and Fladung (2007), and Rachmayanti et al. (2009), and Tnah et al. (2012), and But et al. (2023). Thus, methods to improve DNA extraction and amplification from wood samples are needed. The second type of markers are Single Nucleotide Polymorphisms (SNPs). SNPs require smaller DNA fragments for detection: as short as 35 bp, depending on the method of detection used. Sets of SNPs can be selected for certain purposes, e.g., to be species-specific or because they occur in certain regions of the genome.

As an alternative, since SNPs occur in a high frequency in the genome and sequencing costs have strongly declined, one may detect the SNPs that occur in DNA sequences without any preselection. Such a SNP analysis may include hundreds or thousands of SNP loci spanning the entirety of the genome(s)(Glover et al., 2010) , potentially resulting in higher genetic power to distinguish closely related individuals or to detect low levels of population differentiation.

Next-generation sequencing (NGS) technologies are developing quickly, which means that

SNP detection in DNA sequences will become more cost-efficient over time. To support matches with SNPs in other individuals and species, an increasing number of public DNA sequence databases is available, e.g. National Center for Biotechnology Information - NCBI and European Variant Archive – EVA (Glover et al., 2010; Dadzie et al., 2013; Meyer-Sand et al., 2018; Blanc-Jolivet et al., 2018; Blanc-Jolivet et al., 2018; Dormontt et al., 2020).

If sequencing costs continue to drop in the future, an analysis of complete genomes, rather than only a subset, may become affordable, in which case the field would move from a few DNA regions to genome-wide analyses (e.g. Li et al., 2015; Mascarello et al., 2021). However, at present , plant nuclear genomes are often (far) too large for a whole-genome approach to be feasibly applied on all individuals.

1.6 USING PLASTID GENOMES TO CONDUCT TIMBER TRACING

In view of the recent developments in genetic analyses, in this thesis I explored the potential of utilizing SNPs in the plastid (more specifically: chloroplast) genome (cpDNA). I used this approach to verify the species and the origin of timber for two tropical African timber species. Historically, genetic methods for verifying the origins of timber have primarily utilized polymorphisms within the nuclear genome (Blanc-Jolivet et al., 2018; Low et al., 2022). The advantage of targeting plastid genomes is that they are small and are present in multiple (hundreds or sometime thousands) copies per cell. Thus, they are more likely than nuclear DNA to be retrieved from wood tissue in sufficient quality and quantity, given the low amounts and low quality of DNA in wood (Mader et al., 2018a; Mascarello et al., 2021).

Studying nuclear or plastid genomes may result in different patterns. Nuclear genomes are inherited from both parents, undergo recombination, and harbor more sequence variants compared to plastid genomes, also per unit length. The reason for this difference is that plastids are uniparentally inherited (in plants generally maternally) and do not recombine. Hence, they have a smaller effective population size, thus fewer random mutations will occur and accumulate in time. As they are not inherited through pollen, their dispersal distance is much smaller and the population structure detected will be different from that based on nuclear genome polymorphisms. The fact that they do not recombine means that the genome inherits as one unit, so all SNPs together form one haplotype. This causes some limitations in the data analysis, as population-genetic software generally assumes that SNPs inherit

independently.

Genetic differentiation is influenced by natural barriers (topography), mating system, pollen and seed dispersal mechanisms (gene flow), and demographic processes through time and space. Understanding how genetic variation within and between populations of a tree species are spatially organized, i.e. establishing the spatial genetic structure (SGS), helps to differentiate populations genetically. Establishing the SGS for a given timber species forms the first step in assessing the potential of using genetic techniques for tracing. To understand and interpret SGS, information on historical biogeography and phylogeography is valuable, because it reflects how past environmental changes and past demographic events have structured the genetic setup of current populations (Lowe et al., 2010; Iwanaga et al., 2012; Hardy et al., 2013; Duminil et al., 2015). The spatial structure of tree species as measured by plastid markers may reflect the recolonization from different refugia, as happened in Europe since the end of the last Ice Age (Blanc-Jolivet and Liesebach, 2015), or the existence of persistent genetic barriers such as the Dahomey Gap in West Africa (Duminil et al 2015) or within the Lower Guinean region in Central Africa (Hardy et al., 2013; Duminil et al., 2015).

1.7 FOUR BARRIERS FOR GENETIC TIMBER TRACING

To deliver a genetic forensic tool to solve tropical timber legality questions, four barriers must be overcome (Figure 1.2). (a) A lack of databases to provide the basis to verify timber legality (Low et al., 2022): there is a pressing need to establish extensive sample collections. (b) Appropriate genetic markers need to be developed. (c) Factors influencing assignment test success, such as overestimation of assignment power of reference databases (Anderson, 2010) need to be verified. (d) The spatial resolution of genetic tracing methods within the distribution area of species needs to be increased, which may be done by integrating other types of information

1.8 HOW BARRIERS ARE TACKLED IN THIS THESIS

The PhD thesis aimed to develop SNP marker sets capable of detecting genetic variation to distinguish between two timber species and to provide information on their geographic origins. To this end, we undertook extensive fieldwork to collect large sets of reference

samples of two timber species (addressing barrier a). Using these samples, we developed whole plastomes that can be used in its entirety or in specific parts for SNP detection and determining the plastid haplotypes (addressing barrier b). Next, we evaluated the spatial distribution of haplotypes and the spatial genetic structure, applied DNA barcoding to blind samples to verify timber claims for the two species, and compared assignment tests with the plastome-wide SNP sets (addressing barrier c). Finally, we explored the integration of genetic tracing with multi-element and isotopic analysis (addressing barrier d). In this way, the thesis addresses critical challenges in the field of tropical timber tracing and contributes to the fight against illegal logging.

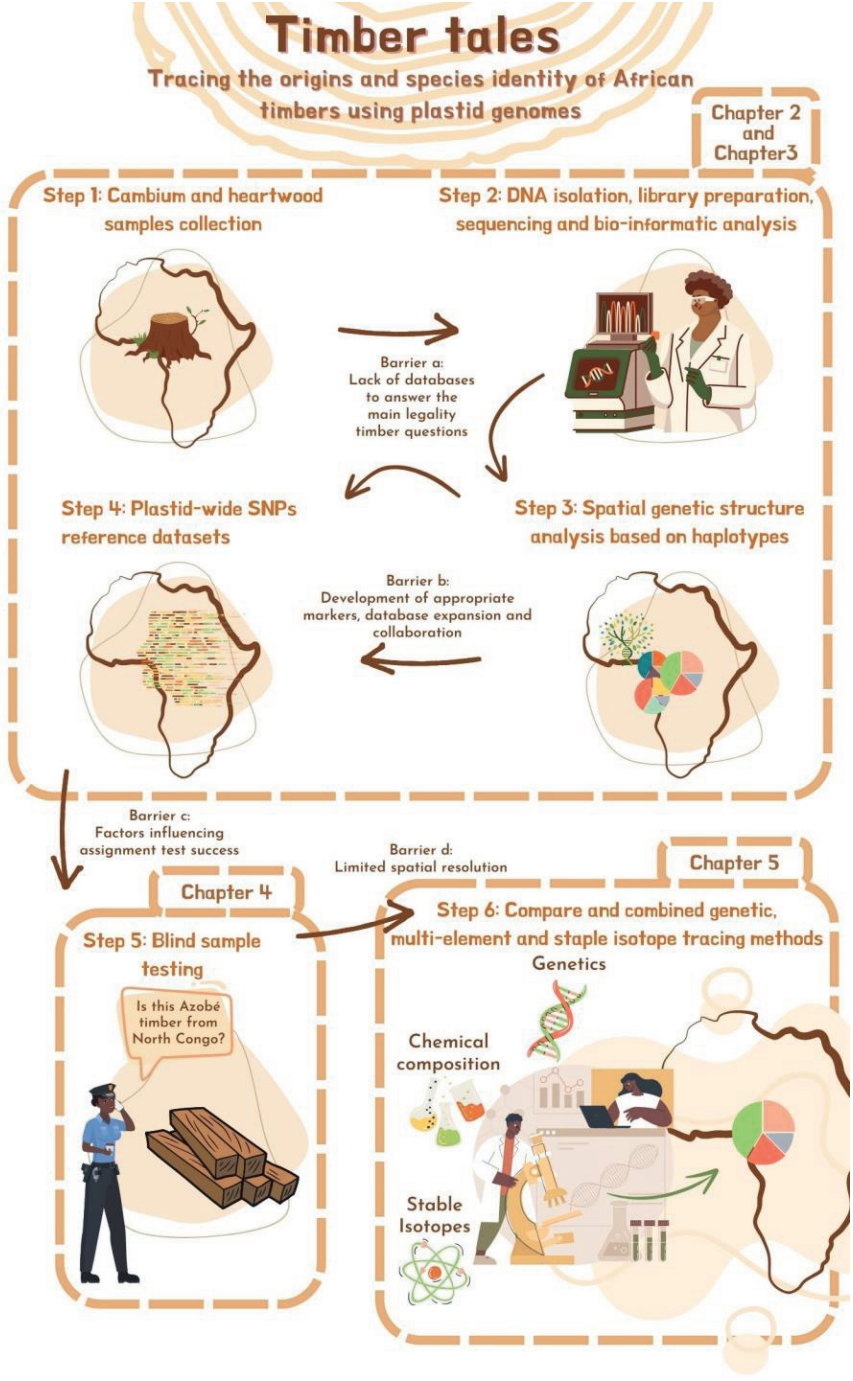


Figure 1.2: Conceptual Scheme of the thesis. This figure illustrates the six steps and four barriers in the development, testing, comparison, and integration of genetic tracing tools for timber species and origin claims verification.

1.9 THE TWO TIMBERS STUDIED

The two timber groups focused on in this study are Azobé (*Lophira alata*) and Tali (*Erythrophleum suaveolens* and *E. ivorense*), which are important tropical timbers in the Dutch and European markets. For material, we organized over 10 sampling campaigns that collected samples of 700 trees from 15 different forest concessions in the main producing countries of these timbers, Cameroon, Congo Republic and Gabon. We also included previously collected samples such as DNA extracts from existing sapwood samples collected at two concessions in Cameroon in 2015, described in Vlam et al. (2018), and leaf and cambium samples collected during several field expeditions by the Free University of Brussels in nine African countries, covering the full distributional range of both timbers.

Tali and Azobé are two highly valued timber species with distinct characteristics and challenges in identification and tracing. Tali timber, belonging to the genus *Erythrophleum* within the Fabaceae-Caesalpinioideae family, consists of *Erythrophleum suaveolens* (Guill. Et Perr. Brenan) and *E. ivorense* (A. Chev.). Utilized in heavy construction, these species are morphologically similar and traded under the same name, complicating their distinction. *E. ivorense*, favored for its commercial properties, thrives in coastal wet forests of the Guineo-Congolian center of endemism, whereas *E. suaveolens* spans the Guineo-Congolian/Sudanian Transition zone (Hardy et al., 2013; Duminil et al., 2015; Gorel, 2019). Both species are hermaphroditic, insect-pollinated, and employ mixed seed dispersal strategies involving ballochory and zoochory with primates as key dispersers (Poulsen et al., 2001; Segers, 2018; Hardy et al., 2019). The two *Erythrophleum* species are also difficult to distinguish in the field. Despite haplotype sharing between the two species in contact zone, the plastid genome has shown potential to differentiate the two species (Duminil et al., 2010; Hardy et al., 2013). Phylogeographical signals and spatially structured haplotypes have also been noted, indicating genetic diversity and spatial genetic structure (Duminil et al., 2013).

Azobé (*Lophira alata,* Banks ex C.F.Gaertn.), found across the Guineo-Congolian region, is known for its dense wood and ecological significance. The species is insect-pollinated and produces "Bongossi nuts" dispersed by animals (Ouédraogo et al., 2020). The inclusion of the congeneric species *Lophira lanceolata*, which occupies different ecological niches, aids in species identification efforts (Ewédjè et al., 2020; Fahey et al., 2021). While *L. lanceolata* prefers woodlands and dry forests, *L. alata* is predominant in wet tropical forests, with both

species coexisting in transitional zones (Piñeiro et al., 2015; Ewédjè et al., 2020). Recent genetic studies by Ewédjè et al. (2020) have differentiated *L. alata* from a cryptic species within it, endemic to western Gabon (*L. alata* WG), using nuclear microsatellite loci. *Lophira alata* was described as having a low degree of genetic diversity and could benefit from the use of whole plastome sequences for its genetic characterization, as pointed out by Mascarello et al. (2021).

1.10 THESIS OUTLINE

In Chapter 2, I investigated the use of plastid genome polymorphisms for identifying Tali species and determining the source of Tali timber.

In Chapter 3, I studied plastid genome polymorphisms of Azobé timber and determined the association between genetic and spatial distances .

After obtaining plastid-wide SNP reference datasets for Azobé (Chapter 2) and Tali (Chapter 3), Chapter 4 contains tests of their robustness for species and origin identification, including tests with blind samples.

Chapter 5 combines genetic and chemical methods for timber tracing. We assessed individual chemical and genetic tracing method accuracy and reliability, and the potential benefits of combining the methodologies to enhance the fine-scale tracing of timber.

Chapter 6 presents the general discussion, in which I summarize and integrate results of the experimental chapters, compare them to literature and discuss their merits. Here I also added a discussion on the ethics of tracing studies and present an outline on the future of timber tracing.

Chapter 2

Using whole chloroplast genome sequence variation to identify species and assess the spatial genetic structure of a valuable African timber species

Using whole chloroplast genome sequence variation to identify species and assess the spatial genetic structure of a valuable African timber species

Bárbara Rocha Venâncio Meyer-Sand, Marinus J. M. Smulders, G. Arjen de Groot, Laura E. Boeschoten, Danny Esselink, Nathalie de Vries, Steve N. Tassiamba, Martin T. Tchamba, Herman T. Zanguim, Pascaline T. Zemtsa, Dyana Ndiade Bourobou, Joël J. Loumeto, Gaël U. Dipelet Bouka, Olivier J. Hardy, and Pieter A. Zuidema

Abstract

Genetic tracing is frequently applied in wildlife forensic cases, but methodology to infer timber legality is still a matter of active research. In this study we assessed the potential of polymorphisms in the plastid (chloroplast) genome for species identification and source inference for the case of Tali (*Erythrophleum suaveolens* and *E. ivorense*), a highly valuable timber from Africa. DNA was isolated from leaf, cambium or sapwood material of 235 trees of Tali (*E. suaveolens* and *E. ivorense*) sampled in situ in Central and West Africa. Complete plastid genomes were assembled from whole genome shotgun sequencing. We observed a clear genetic differentiation between the two timber-producing species *E. suaveolens* and *E. ivorense*, with haplotypes assigned to one of the species. A within-species genetic structure was evident for both species, but it was stronger in *E. suaveolens* with two groups in East Cameroon occurring in different parts of the sampled area. This study represents a first step to further evaluate the potential of using the plastid genome for taxonomic and geographic verification of Tali timber. Such analyses need to take into account that wide geographical distribution of certain haplotypes may put restrictions on geographic tracing in certain regions.

Keywords: timber tracing; plastid genome; chloroplast genome; SNPs; *Erythrophleum*; Tali.

2.1. INTRODUCTION

Illegal logging undermines sustainable forest management and contributes to exacerbating biodiversity loss in forests. Illegal timber trade is particularly problematic in tropical Asia and Africa where it represents over half of the traded volume of various countries (Hoare, 2015). To reduce illegal trade, legislation and international treaties have been developed, including the US Lacey Act (2008), the European Union Timber Regulation (2010), European Union Regulation on deforestation-free products (2023), and International Trade in Endangered Species of Wild Flora and Fauna (CITES). Despite these efforts, tropical timber is still at high risk of illegality because of poor governance and the difficulty to authenticate documents relating to species identity and/or geographic origin of the traded timber, which hinder law enforcement (Kleinschmit et al., 2016). Thus, the development and establishment of reliable forensic methods are needed to verify trade claims on geographical provenance and species identity (The European Parliament And The Council Of The European Union, 2023; Low et al., 2022). Various forensic timber tracing methods exist. Morphological, chemical and genetic methods can provide answers on species, provenance and chain of custody integrity based on intrinsic properties of the wood (Dormontt et al., 2015). Some of these methods are well established as forensic tools for some commodities (Gonzálvez et al., 2009; Ogden and Linacre, 2015). While morphological and chemical methods are usually limited to tracing either the species or geographic origin (Dormontt et al., 2015), genetic methods can be used for both goals simultaneously (e.g., Honorio Coronado et al., 2020). Taxonomic classification and phylogeography using genetics are well-established fields (e.g., Duminil et al., 2010; Höltken et al., 2012; Daïnou et al., 2016; He et al., 2019; Bouka et al., 2022), supported by the initiative Barcode of Life Consortia ("International Barcode of Life – Illuminate Biodiversity," n.d.). However, despite the potential advantages of genetic analysis for achieving both goals, the application of this method has been limited by incomplete reference databases (Dormontt et al., 2015; Low et al., 2022). So far, the focus of genetic tracing has been on species identification, covering now 86% of 322 timber taxa with high priority for timber tracing against 23% for origin identification (Low et al., 2022), hampering the potential for source tracking for a majority of timber species. This contrast is primarily due to the substantial sampling efforts required to establish a reference dataset, which has to be representative of the genetic diversity of a species. To sufficiently capture

the genetic diversity of a species, a reference dataset must include several individuals per site, from multiple sites spanning the natural distribution of the species. This can result in the need to collect hundreds to thousands of trees, which presents significant logistical and financial challenges.

In addition to extensive sampling efforts, the quality of the genetic data is crucial to establish a robust reference database. A robust reference database must be transferable to heartwood tissue, which is known to yield low concentrations of highly fragmented DNA (Rachmayanti et al., 2006; Degen and Fladung, 2007; Taberlet et al., 2007). The genetic data is strongly influenced by the quality of the DNA isolates, the sequencing strategy and the type and number of genetic markers used. The development of next-generation sequencing (NGS) techniques for detecting Single Nucleotide Polymorphisms (SNPs) allows to overcome the limitations for genetically tracing wood tissues (Blanc-Jolivet and Liesebach, 2015). Firstly, SNPs are the most common type of polymorphism (Singh et al., 2013). A whole genome can harbor thousands to tens of thousands of loci (Glover et al., 2010). So, an analysis of a subset of the genome already can yield high differentiation power for species distinction and analyses of the genetic structure of a species. Secondly, the analysis of the SNPs set can be automated, decreasing errors and enabling streamlined standard analysis among laboratories, ultimately facilitating the use of the tracing tool (Blanc-Jolivet and Liesebach, 2015; Singhal et al., 2011). Thirdly, SNP detection may be done on small DNA fragments. So far, only a few studies have addressed origin tracing and species identification simultaneously, usually using microsatellite markers (Blanc-Jolivet and Liesebach, 2015), even less studies have used SNPs for these purposes (e.g. Honorio Coronado et al., 2020).

Development of SNP sets for timber tracing has been mostly focused on the nuclear genome (nDNA), with marginal attention paid to plastid DNA (Blanc-Jolivet and Liesebach, 2015). The plastid genome, also known as plastome (pDNA), is much smaller than the nuclear genome, offering a lower level of genetic variation. Platomes are usually inherited maternally, and their low variation is due to the rare or absent recombination, as they evolve clonally (Blanc-Jolivet and Liesebach, 2015). Recurrent exchange of pDNA has been reported in some taxa (e.g. Quercus [Blanc-Jolivet and Liesebach, 2015]), what may hamper species identification solely based on pDNA. However, plastid loci are successfully used for DNA barcoding to differentiate species (Taberlet et al., 2007) and detection of invasive plant species (Van de Wiel et al., 2009; Ghahramanzadeh et al., 2013).

For wood tracing purposes, pDNA offers several advantages over nDNA. Plastid genomes occur in high copy numbers in each cell, compared to only two copies of the nuclear genome, and the plastid organelle structure may be less prone to degradation (Mader et al., 2018; Fahey et al., 2021). For instance, plant DNA extracted from leaves may yield up to 40 –50% reads representing chloroplast DNA (Shi et al., 2012), even though the plastid genome is much smaller than the nuclear genome. Hence, the probability to detect SNPs on the plastid genome from small amounts of degraded DNA extracted from wood is higher than doing so on the nuclear genome.

Here we use whole-genome sequencing to develop a SNP set in the plastid genomes of the African timber species 'Tali' to identify the two botanical species producing this timber (*Erythrophleum suaveolens* and *E. ivorense*), and to infer spatial genetic structure. We thus conduct the first necessary step in developing and verifying the potential for genetic tracing of |Tali timber. We answer the following research questions: a) What are the major genetic lineages observed and do they correspond to botanical species? and b) What is the spatial distribution of observed haplotypes, and at what spatial scales are haplotypes spatially separated? We discuss the potential of our results for using cpDNA in timber identification, as well as future studies needed to assess the potential for timber tracing applications.

2.2 MATERIALS AND METHODS

2.2.1 Studied species

Tali is a highly valued timber group that belongs to the genus Erythrophleum (Fabaceae-Caesalpinioideae) and composed by two botanical species: *Erythrophleum suaveolens* (Guill. Et Perr.) Brenan and *E. ivorense* (A. Chev.). The wood of Tali is used for heavy structures such as bridges, railway sleepers, terraces and industrial floors (Gorel, 2019). The two Tali producing species are difficult to distinguish in the field due to their morphological similarities and therefore are harvested and traded under the same common name, even though *E. ivorense* is reported to present more desirable commercial features (Gorel, 2019). The two botanical species mostly occur in different areas: *E. ivorense* is distributed in the coastal wet forest of the Guineo-Congolian Center of endemism (White, 1983), while *E. suaveolens* is present from the Guineo-Congolian/Sudanian Transition zone through the

Guineo-Congolian zone (Duminil et al., 2015; Gorel, 2019). *E. ivorense* and *E. suaveolens* are both hermaphroditic and likely pollinated by insects (Segers, 2018). Primary seed dispersal is ballochoric and secondary dispersal zoochoric with many primates identified as mean seeds dispersers agents in wet forests (Poulsen et al., 2001; Koné et al., 2008; Hardy et al., 2019).

The two *Erythrophleum* species are difficult to distinguish in the field, however Duminil et al. (2010, 2013) have showed the potential of genetic-based approach to differentiate between *E. ivorense* and *E. suaveolens*. The species differentiation was successful based on the chloroplast DNA (Duminil et al., 2010), however one exception of haplotype sharing were detected in contact zones (using cpDNA sequences and nSSR) (Duminil et al., 2013). Duminil et al. (2010) also reported phylogeographical signals for one of the two the chloroplast sequences tested, the inter-genic fragment trnC-petN1R, with individuals in the same population being genetically closer than between populations. The haplotypes detected for the same pDNA fragment were spatially structure, although with wide distribution (Duminil et al., 2013).

2.2.2 Sampling area and strategy

For the present study Tali samples from three sources were used: (a) 217 newly collected cambium samples in 2019 and 2020 in 10 concessions in Cameroon, (b) 26 DNA extracts from existing sapwood samples collected at two concessions in Cameroon in 2015, described in Vlam et al. (2018), and (c) 45 herbarium leaf and cambium samples collected during several field expeditions by the Free University of Brussels in nine African countries covering the full distributional range of both botanical species. Newly collected samples (type a) were obtained from sites at distances ranging from 20 km up to 640 km apart (Figure 2.2a). For these samples, sampled trees within each site were at least 100 meters apart and at most 5000 m. Sampled trees were geo-referenced, and had their diameter at breast hight (DBH) recorded. All sample types were dried with silica gel.

2.2.3 DNA Isolation, library preparation and sequencing

DNA was isolated from leaf, cambium or sapwood material. Leaf and cambium tissue (source a and c) of 262 trees were used for DNA isolation with an optimised cetyltrimethyl

ammonium bromide (CTAB) protocol as described by Dumolin et al. (1995) with additional cleaning steps (Supporting note S2.1). The 26 DNA extracts (source b) from sapwood tissue were extracted and purified as described by Vlam et al. (2018). DNA purity of all 288 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 288 DNA isolates were used to prepare three paired-ended (300 bp or more insert size) libraries with the RIPTIDE High Throughput Rapid Library Prep Kit (Twist Bioscience, South San Francisco, USA). The libraries were sequenced, paired end libraries (2 × 150 bp), on an Illumina Novaseq6000 (Novogene, Cambridge, United Kingdom).

2.2.4 Data analysis

2.2.4.1 Bioinformatic analysis

Bowtie 2.2.5 (Langmead and Salzberg, 2012) was employed to map reads from each of the 287 samples against the *Erythrophleum suaveolens* Yi14880-KUN plastid genome (Genbank: MN709858.1) using default parameters. Picard 2.23.0 was utilized to mark duplicate reads. Raw variants were called using NGSEP 4.0.1 (Tello et al., 2019) MultisampleVariantsDetector with parameter '-ploidy 2'. A variant call file was generated with a custom Python script. BCFtools 1.15.1 was applied to perform SNP filtering, in which only biallelic SNPs were retained. Downstream analyses were carried out using R version 4.1.0. The detected variants were filtered for a minimum sequencing depth of five reads and a maximum depth of 200, genotypes that did not fulfil the criteria were considered as missing data (SNPfiltR package [DeRaad, 2023]). Finally, we removed individuals with data missing for >50% of the SNPs and subsequently SNPs with data missing for >10% of the individuals (SNPfiltR package [DeRaad, 2023]).

2.2.4.2 Species differentiation

We assessed the number of major lineages and verified if these corresponded to the two botanical species. The number of main lineages was detected with a principal component

analysis (PCA, missing data was replaced with mean of allelic frequency) (Ade4 Package v1.7-22), supported by Neighbor-joining bootstrapped dendrogram (Ape package v5.6-2, [Paradis and Schliep, 2019)) and k-means clustering based on values of Bayesian Information Criterion (BIC) (Adegenet package v2.1.8, (Jombart, 2008)]). A discriminant analysis of principle components (DAPC) (Adegenet package) was conducted to assign trees to the identified lineages. The assignment of samples was used to verify tree identification and correct potentially misidentified trees, as *E. ivorense* and *E. suaveolens* are known to be difficult to differentiate in the field. We opted for the utilization of PCA, and subsequently DAPC, which are widely employed tools for analyzing nuclear DNA. This choice was made due to the consistent clustering patterns observed between PCA results and the neighbor-joining tree. As a result of this consistency, we employ the terms "clusters" and "lineages" interchangeably to convey our findings. The validation and species (re-) assignment of species identity was done using the DAPC clusters, and how the five botanically vouchered specimens were clustered (two *E. suaveolens* and three *E. ivorense* specimens, Supporting table S2.2). For samples with possible incongruency between the identification in the field and the genetic assignment, we visually compared the locations of the individuals with the natural distribution of each of the *Erythrophleum* species, and to those of samples in the genetic clusters identified by Duminil et al. (2013). Only specimens confidently identified (consensus amongst the natural species distribution or known genetic clusters, and clustered with vouchered specimens) as belonging to one of the target species were retained in the reference database for species identification purposes.

2.2.4.3 Haplotypes, their spatial distribution and the potential of haplotypes for tracing purposes

The Pegas package (Paradis, 2010) was used to define the haplotypes, the number and frequency of haplotypes. Haplotypes were defined in a non-restrictive manner with inclusion of sequences of varying lengths within the same haplotype, while ambiguities resulted in certain sequences being assigned to different haplotype (Pegas package [(Paradis, 2010)]). To visualize the relationship amongst haplotypes, we constructed a median joining network (Bandelt et al., 1999) using all SNPs for both species, implemented in the Pegas package in R. To assess the tracing potential, spatial-genetic differences, and patterns, we mapped the distribution of haplotypes onto a sampling area map using QGIS.

2.3. RESULTS

2.3.1 Reference chloroplast genome

The chloroplast genome of *Erythrophleum suaveolens* Yi14880-KUN (GenBank: MN709858.1, 160344 bp) served as reference for mapping of the 287 trees. After mapping, a variant call was conducted and a total of 7597 putative SNPs were detected. After quality filtering (minimum five and maximum of 200 reads per SNP, maximum 50% missing values per tree, and 10% missing per SNP), 323 biallelic SNPs and 235 trees were retained for downstream analysis.

2.3.2 Species differentiation

In the PCA used to evaluate the genetic structure of the data, the first PCA axis explained a large percentage (71%) of variation, indicating that the samples are obtained from two genetically distinct groups (Figure 2.1a). Based on the PCA results, Neighbor-joining tree and after exploring the BIC values of in k-means clustering for different cluster numbers (Supporting figure S2.3), a DAPC cluster analysis was conducted for K=2 groups. All five vouchered specimens (two *E. suaveolens* and three *E. ivorense*) were consistently matched to the clusters/lineages composed predominantly by trees of the same species. The individuals' clustering was manually verified by comparing the group to which vouchered samples were matched to, with the natural distribution of each of the *Erythrophleum* species and with previous identified genetic clusters (Duminil et al., 2013). The posterior membership probabilities from DAPC clustering for all 235 trees were very close to 1 for K=2 (Figure 2.1b). This species classification allowed to assign 55 trees to species level (10: *E. suaveolens,* 45: *E. ivorense*), re-assign 8 individuals (out of 149) from *E. suaveolens* to *E. ivorense*, and 6 (out of 33) from *E. ivorense* to *E. suaveolens*.

Two individuals (L1610179 and LK0031) were assigned to *E. ivorense* in the DAPC analysis, however they are seemingly closer to *E. suaveolens* in the PCA. Both trees occur in regions where *E. suaveolens* occur, and match with previously described gene pools of *E. suaveolens* West (SW) and *E. suaveolens* Centre North (SCn)(Duminil et al., 2013). Due to this inconsistency their assignment might not be reliable. Nei's distance (Nei, 1972) between the

lineages was 0.45. Together, we conclude that our method can separate tree samples as *E. ivorense* and *E. suaveolens*, however not potential hybrids between the species.

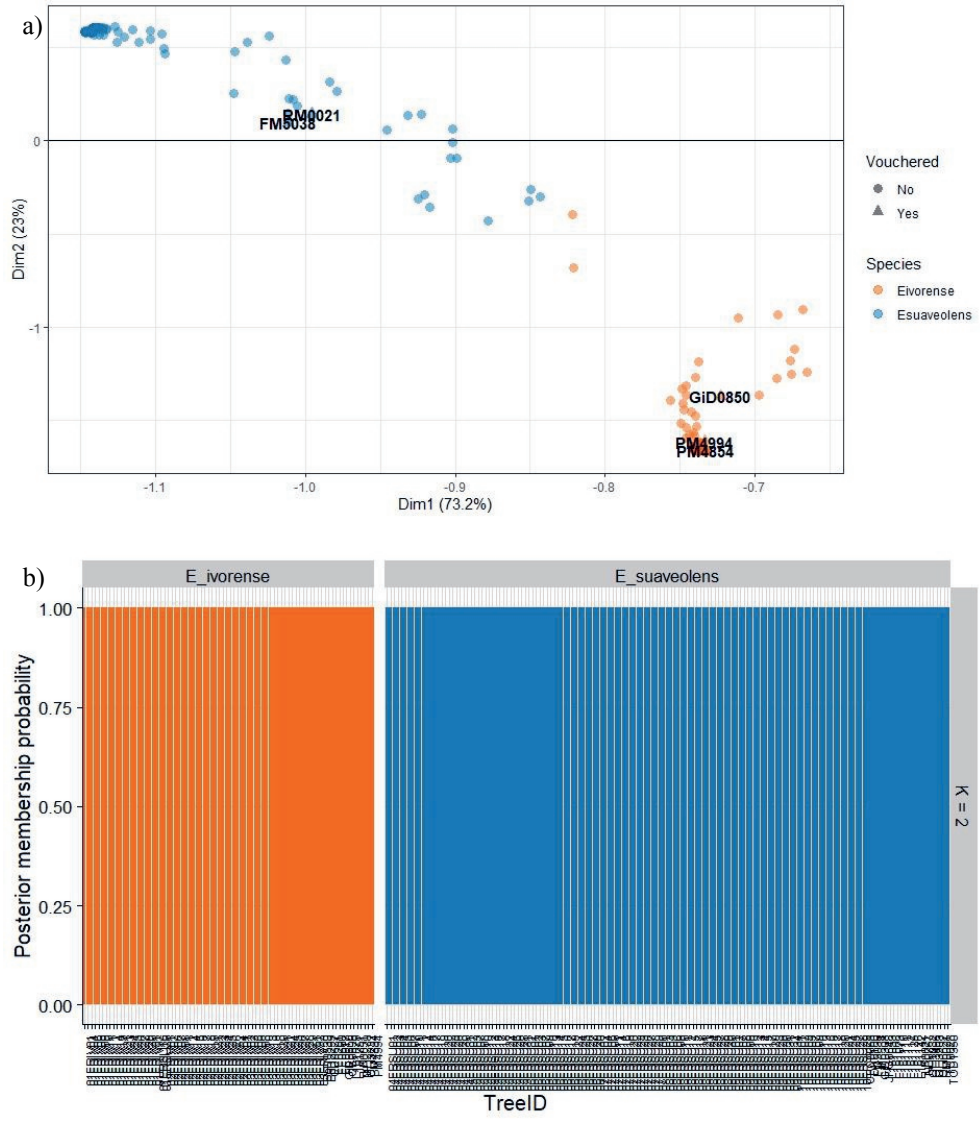


Figure 2.1 : Differentiation of two botanical species within Tali. a) – A principal component analysis (PCA) shows two distinct genetic clusters along the first axis. PCA is based on 311 biallelic SNPs in 235 samples 156 *E. suaveolens* and 80 *E. ivorense*). Labelled trees are herbarium vouchered specimens. b) Bar plots of the posterior probabilities of group assignment for each of the 235 sampled trees for two clusters (K=2) using the Discriminant Analysis of Principal Components (DAPC).

2.3.3 Haplotypes and their spatial distribution

The dataset contained 323 segregating sites (SNPs), of which 105 were found within *E. ivorense*, and 97 within *E. suaveolens*. A total of 44 haplotypes were detected, of which 18 belonged to trees identified as *E. ivorense* and 26 to trees identified as *E. suaveolens*. Among the 18 haplotypes observed in *E. ivorense*, eight were found in Cameroon, while the remaining 10 were distributed from Liberia to Gabon (Figure 2.2a and b). For *E. suaveolens*, we observed a total of eight haplotypes in Cameroon and the remaining 18 were distributed from Ivory Coast to the Democratic Republic of the Congo (Figure 2.2a and c).

The haplotypes distribution showed distinct patterns at different geographic distance ranges. Haplotypes found in West Africa are distinct from those found in Central Africa with site-private haplotypes in most sites (light salmon for *E. ivorense* and light blue for *E. suaveolens*, Figure 2.2). *E. ivorense* haplotypes are geographically consistent, with haplotypes being regionally (West to Central Africa, and/or site> 200 km apart) constrain, for example with H7 in Benin and Togo, and H8 occurring in Liberia with 8 mutations from the closest (incomplete) haplotype. For *E. suaveolens* regional patterns are also revealing, for example very distinct haplotypes such H42 restricted to the Gabonese and Congolese (Rep. of Congo) coast, and H32 and H41 to DRC.

At the local scale (within Cameroon) genetic differentiation is mostly determined by species distribution, as *E. ivorense* is restricted to coastal areas while *E. suaveolens* occurs in the Centre and East regions (Figure 2.2a). Common haplotypes in both species, such as H4 and H3 for *E. ivorense* and H23 and H25 for *E. suaveolens*, occurred in more than one site making the site differentiation difficult. In Cameroon *E. suaveolens* haplotypic patterns indicated two clear lineages: one consisting of sites RECODEM, and 10051, and 10053 the second of sites COBACAM, 10010, 10012, and 10026. Although the grouping indicates a more likely origin, as a few haplotypes (H23 and H25) are shared amongst some of the sites, it is difficult to distinguish at the site level based on their haplotypic composition. Overall, the haplotype networks for both species show that haplotypes occurring in nearby locations are also genetically similar, thus limiting the ability to genetically distinguish close-by sites that share haplotypes.

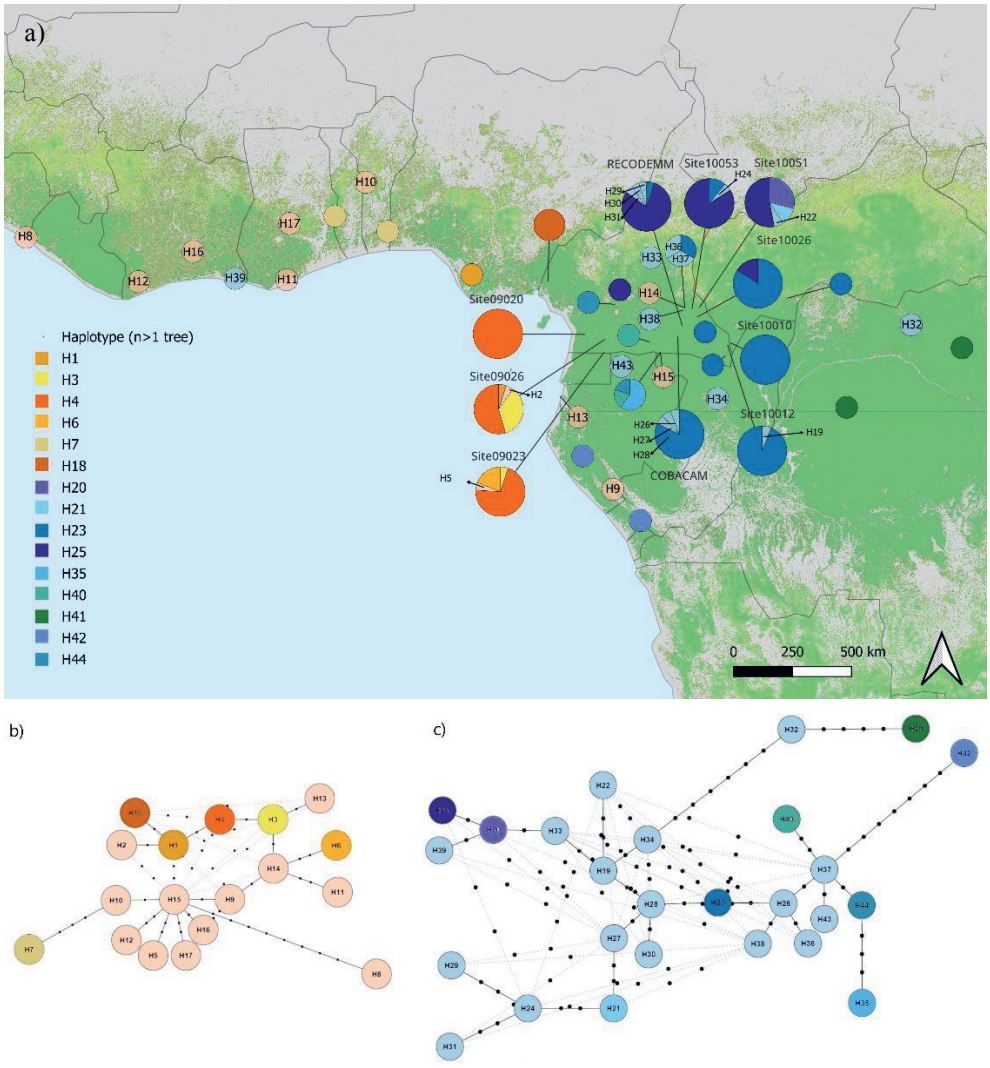


Figure 2.2 : a) Distribution of haplotypes in the whole sampling area, with focus in sites with sampling sizes above 17 in Cameroon. Haplotype network based on 323 SNPs of the chloroplast genome b) E. ivorense and c) E. suaveolens (solid line: most likely links in the network, dashed line: alternative links of the network). Haplotypes colored in different colors are present in more than one individual; light salmon indicates unique *E. ivorense* haplotypes, and light blue indicates unique *E. suaveolens* haplotypes.

2.4. DISCUSSION

We assessed the ability of genetic polymorphisms across the plastid genome to differentiate species and to determine the degree of spatial genetic structure of the widely traded timber Tali (*Erythrophleum* spp.). Using this approach, we were able to distinguish the two botanical species, and at the same time we found clear differences in the spatial distribution of the haplotypes.

2.4.1 Major lineages and species identification

The whole plastid genome SNPs set displayed a clear difference between the two botanical species *E. suaveolens* and *E. ivorense*. The finding is supported by distinct clades in the phylogenetic tree (Supporting Figure S2.4). Our findings are in line with Duminil et al. (2013) study that characterized nuclear SSR and cpDNA haplotypes in a large area (West and Central Africa), and found that species do not share cpDNA haplotypes, except for one individual case detected in a contact zone. Our results are also consistent with the results reported by Duminil and colleagues (2010) in Cameroon using one plastid marker (trnC-petN1R intergenic region), who found one main haplotype for *E. ivorense* and two main haplotypes for *E. suaveolens*. In our study, the individuals identified as *E. ivorense* were restricted to the coast (except for L161079 in East Cameroon), whereas *E. suaveolens* individuals were restricted to the inner lands (except for LK0021 at the coast in Ivory Coast), congruent with the known distribution of the species (White 1983). L161079 do not occur within the expected species distribution, however LK0021's location coincides with the occurrence area of the previously described gene pool of *E. suaveolens* West (SW)(Duminil et al., 2013). Their haplotypes were incomplete (about 35% and 60% of SNPs, respectively), what possibly explains this result. The clustering analysis (Figure 2.1b) confirmed a sharp distinction between the two species. The strong distinction between species may be partially attributed to the species' parapatric distribution, linked to their ecological niches, and therefore less likely to change chloroplast. The finding that the haplotypes were assigned to one of the two Tali species, enables species identification of this wood based on haplotypes (pDNA). Indeed, pDNA barcoding was shown to be effective in identifying species, including Tali (Duminil et al., 2010; Fahey et al., 2021). However, caution is recommended for taxa known for frequent chloroplast captures (e.g. Quercus [Blanc-Jolivet and Liesebach, 2015]), and

additional sampling of closely related co-occurring species (Fahey et al., 2021) is advised to avoid misinterpreting limited genetic information.

Identification of species through plastid DNA barcode sequences is a well-established method and is expected to be reliable if there is no recurrent chloroplast captures among related species. However, the low mutation rate of the plastid genome often leads to insufficient taxa resolution (Taberlet et al., 2007) and may therefore turn out to be extremely difficult for some taxa (Matiz-Ceron et al., 2022). This insufficient resolution can be linked to complex demographic history (chloroplast sharing) and/or the small number of plastid genome regions employed in DNA barcoding to detect polymorphisms. For the latter, as we use the whole plastid genome SNPs, the risk of not capturing polymorphisms was strongly reduced. Indeed, 40% of the total polymorphic loci was observed between the two species.

DNA barcodes for species identification are fragments of 400-800 bp that are generally amplified by PCR. It may be difficult or impossible to do this on degraded DNA fragments (Matiz-Ceron et al., 2022), such as can be extracted from wood. Therefore, the proposed SNP approach based on whole genome sequencing increases the possibility to obtain genetic information even for samples with degraded DNA.

2.4.2 Spatial genetic patterns and origin differentiation

Spatial structure was clear at large and regional scales, but was less clear at finer scales (sites level). Our results thus suggest that regional scale genetic differentiation is straightforward since haplotypes found in West Africa are distinct from those found in Central Africa. At the local scale (within Cameroon) genetic differentiation is mostly determined by species distribution, as *E. ivorense* is restricted to coastal areas while *E. suaveolens* occurs in the Centre and East regions (Figure 2.2a). At the fine scale , we had expected to find a more fine-grained genetic structure than what was previously reported by Duminil et al. (2013), based on 12 nSSR markers and three pDNA markers. However, two characteristics limit the ability to genetically differentiate closely located sites for both Tali species. The first is the high frequency of three haplotypes that were present in multiple sites (one for *E. ivorense*- H4, and two for *E. suaveolens* – H23 and H25). Second, there is a low number of site-private haplotypes, making close by sites genetically less distinct.

For *E. ivorense* one out of three sites, with more than 17 sampled trees, presented only one

haplotype (H4), whereas the eastern sites (Site09023 and Site09026) presented a mix of six haplotypes. The haplotype diversity found in eastern sites could be driven by its proximity to a putative refugia described by Maley (1996), potentially harboring more lineages, hence higher haplotype diversity. Another possible explanation would be the proximity to the frontier of its main habitat, the evergreen forest (Shi and Chen, 2012; Duminil et al., 2013). As for *E. suaveolens* two main groups were observed: Es1 (Site10010, Site10012, Site10026, and COBACAM) and Es2 (Site10053, Site10051, and RECODEMM). Between the two groups two haplotypes were shared, H8 common in Es2 was present in one site of Es1 (Site10053), and H25 common in Es1 was present in one site of Es2. Our findings are in line with Duminil et al. (2013), who showed similar patterns of substructure for both *E. suaveolens* and *E. ivorense* in Cameroon. The observed split in the haplotype composition appears to be somewhat latitudinal, but also longitudinal across the climatic range. The effect of latitude on the sites' haplotypic composition could be linked to a possible reproductive isolation between Es1 and Es2 *E. suaveolens* groups. The reproductive isolation between same populations species may be induced by flowering asynchrony from west (Central Cameroon) to east (Southern Central African Republic), as reported for *E. suaveolens* by Ouedraogo et al. (2020). Phenology asynchrony amongst populations of the same species may be led by climate factors, for example the effect rainfall gradients on the distribution of *E. suaveolens* gene pools (Gorel et al., 2019). Usually, these events are susceptible to limit pollination across trees and are recorded more by nuclear markers than plastid markers. Past populations fragmentation, barriers to gene flow and/or limited seed dispersal between populations can lead to sub-structure, as observed within *E. suaveolens* along the climate gradients in its natural distribution, especially in East Cameroon. Our *E. suaveolens* tree sampling resulted in a slightly higher resolution due to a higher number of endemic haplotypes when compared to *E. ivorense*. The higher haplotypic diversity for *E. suaveolens* is likely linked to the larger sample size, but may in part also be caused by the fragmented distribution of *E. suaveolens* in geological history, mainly during the series of glacial and interglacial periods during the Pleistocene (Duminil et al. 2013, 2015). A similar pattern of increasing gradient of West-East haplotypic diversity in the same study area in Cameroon was also observed between *B. toxisperma* populations using chloroplast microsatellites markers (Ndiade-Bourobou et al., 2020). As observed within many tree species studies with different plastid markers (microsatellites and SNPs), Cameroonian East-West haplotype genetic diversity gradient may

probably result from past tree species populations fragmentation occurring before or during the Last Glacial Maximum (LGM) due to significant drops in temperature and rainfalls (Maley 1997; Anhuf et al. 2006).

2.4.3 cpDNA for timber tracing

Clearly, the current study is only a first step in fully assessing the potential of cpDNA to trace Tali timber. Yet, our results of species differentiation and spatial genetic structure can provide some ideas.

To verify a trade claim, a given Haplotype(s) of Unknown Origin (HUO) would have to occur or be closely related to haplotypes that occur at the declared origin. The use of haplotype frequency for tracing can then provide a positive match, however that is restricted to the haplotypes present in the reference dataset. But what if the HUO is not included in the reference data? Here, the haplotype network approach may provide additional clues. As pointed out by Mostefai et al. (2022), haplotype networks may help to elucidate the relationships between sequences within and also across populations. Hence, the use of a haplotype network approach to assess the relationship between an HUO and the reference haplotypes presented in this study, allows for potential identification of the geographic region from which the wood sample likely originates. Even if the HUO is absent from the reference dataset, the haplotype network approach can estimate the number of mutations that separate it from the closest haplotype in the dataset. Although the haplotype network method cannot precisely pinpoint the origin site, it may potentially identify the origin of trees/wood between sites or in the forest block area of where the forest plot from which the timber was taken is located, subject to a maximum probability of occurrence. Haplotype networks have been utilized in various studies, such as identifying lineages of the invasive *Drosophila suzukii* in Brazil (Ferronato et al., 2019) and of SARS-CoV-2 (Mostefai et al., 2022) in link with their occurrence worldwide. Their application for tracing purposes specifically is yet to be explored.

When further verification of the potential is conducted, such a haplotype network approach may offer several benefits for timber genetic tracing. Firstly, it provides a practical choice for studying plant tissues yielding low DNA quantity and quality, such as wood, as it is easier to obtain reliable pDNA compared to nDNA. Secondly, when dealing with fragmented DNA,

the use of genome-wide SNPs instead of long targeted DNA sequences or a predefined set of SNP (e.g., those in typically 400-800bp for DNA barcoding) increases the amount of data and reduces ascertainment bias. Thirdly, SNP analysis demonstrates high reproducibility, ensuring consistent results across different laboratories. Lastly, one of the significant advantages is the ability to assess, at the very least, the region of origin of a haplotype with unknown origins. Narrowing down of the geographic area of origin facilitates the disentanglement of trade routes and allows investigations efforts to be focused on a specific area, which is one of the major challenges faced by law enforcement.

2.5 CONCLUSION AND NEXT STEPS

Whole genome sequencing, SNP calling in the plastid genome, and a SNP haplotype network analysis approach were developed for Tali timber in West and Central Africa area. Our results show good prospects to differentiate species as well as clear spatial genetic structure. Therewith, this study is a first but promising step to verify the potential for timber tracing. To fully evaluate the potential for forensic applications, additional analyses are needed, including developing posterior assignment and exclusion probabilities (of the geographic origin), as well as testing the assignment performance with blind samples.

2.6 ACKNOWLEDGEMENTS

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. We thank all the collaborating timber companies and their field teams for facilitating the fieldwork and all involved colleagues at our partner institutes University of Dschang, Marien Ngouabi University, IRET/CENAREST and the National Herbarium of Gabon for help in arranging fieldwork.

2.7 PERMITS AND AUTHORIZATIONS

Cameroon: Research Permit No. 00000116/MINRESI/B00/C00/C10/C12 (Yaounde, 09 Sep 2019); Research Permit No. 000066/MINRESI/B00/C00/C10/C12 (Yaounde, 07 Jun 2021);

Scientific research permit No. 2144 PRBS/MINFOF/SETAT/SG/DFAP/SDVEF/SC/NGY (Yaounde, 23 Jul 2021); ABS Permit 00010/MINEPDED/CNA/NP-ABS/ABS-FP (Yaounde, 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

Supporting information

Supporting note S2.1 : Modified CTBA DNA isolation protocol with extra purification steps.

Adapted DNA-Extraction

French Protocol modified for difficult samples (V2)

For 25 ml CTAB buffer:

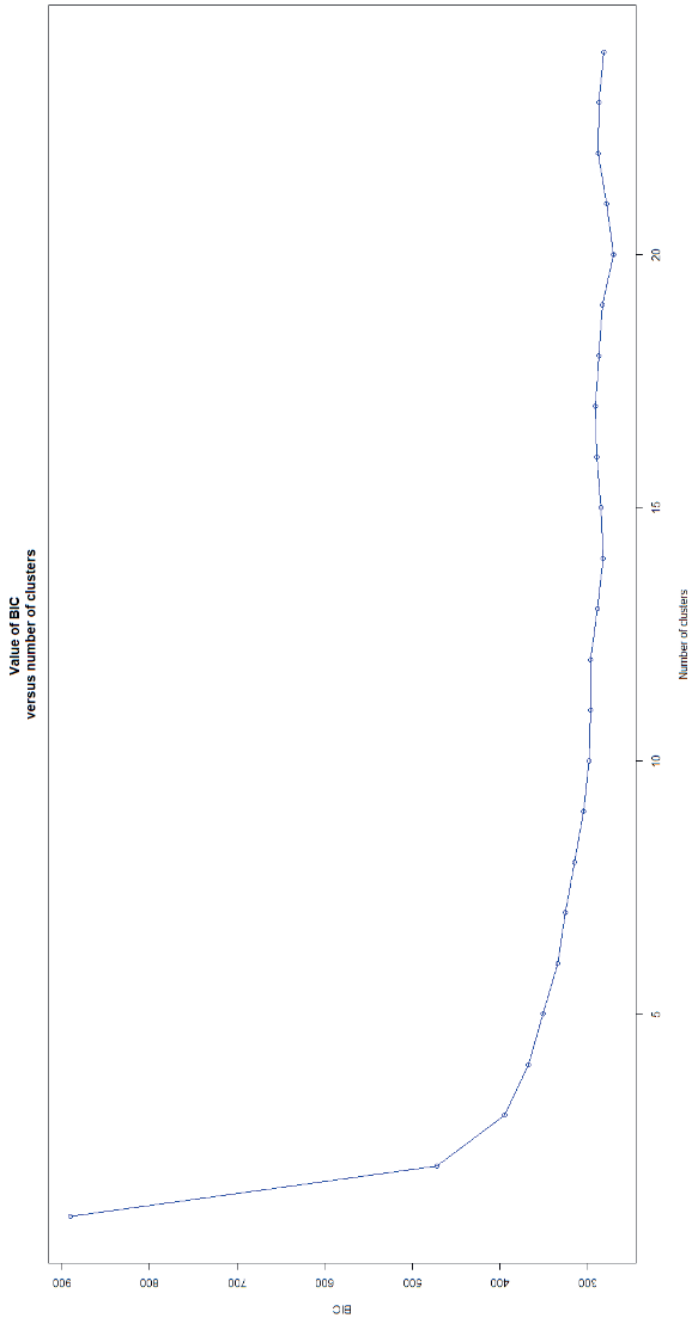
Volume/quantity to
0.5 g CTAB
7 mL of NaCl 5 M
1 mL of EDTA 0.5 M
2.5mL of 1 M Tris - HCl
1 g PVP-40

1. Grind ca.100 mg of plant material (ca.0,5 cm² of cambium flakes ; ~2 cm²eaves; ~ 10-12 needles; ~5 large and primed buds) in 2 ml-tubes in a Retsch-Mill. 2 Beads in each tube. Tip out the beads after grinding
2. Add 1 ml of preheated (65°C) CTAB-buffer and 30 µl Proteinase K, vortex until the solution is homogenous
3. Add 50 µl DTT (under the fume hood)
4. Vortex briefly
5. Incubate for 1h at 65°C. Shake steadly 1200 rpm.
6. Add 900 µl chloroform:IAA (24:1) and mix (vortex) gently about 30 s to get an emulsion
7. Centrifuge at least for 10 min, 13300 rpm
8. Retrieve the upper phase aqueous phase and transfer it into a clean 2 ml tube.
9. Add 10 µl RNase A (20 mg/ml). Incubate 1 hour at 37°C.
10. Add 500 µl chloroform:IAA (24:1), mix and centrifuge at 13300 rpm for 10 mins at 4°C.
11. In the meantime prepare new 2 ml tubes with 225 µl 3M sodium acetate, pH 5.2
12. Transfer ~ 600 µl of the upper phase to the a new sodium acetate tube. ATTENTION: If the upper phase is not clear, the chloroform:IAA (24:1) precipitation step has to be repeated in advance.
13. Add 600 µl of ice-cold isopropyl alcohol and mix
14. Incubate overnight at -20°C
15. Centrifuge 10 min, 13000 rpm, 4°C
16. Wash the pellet in 1 ml 70% ethanol

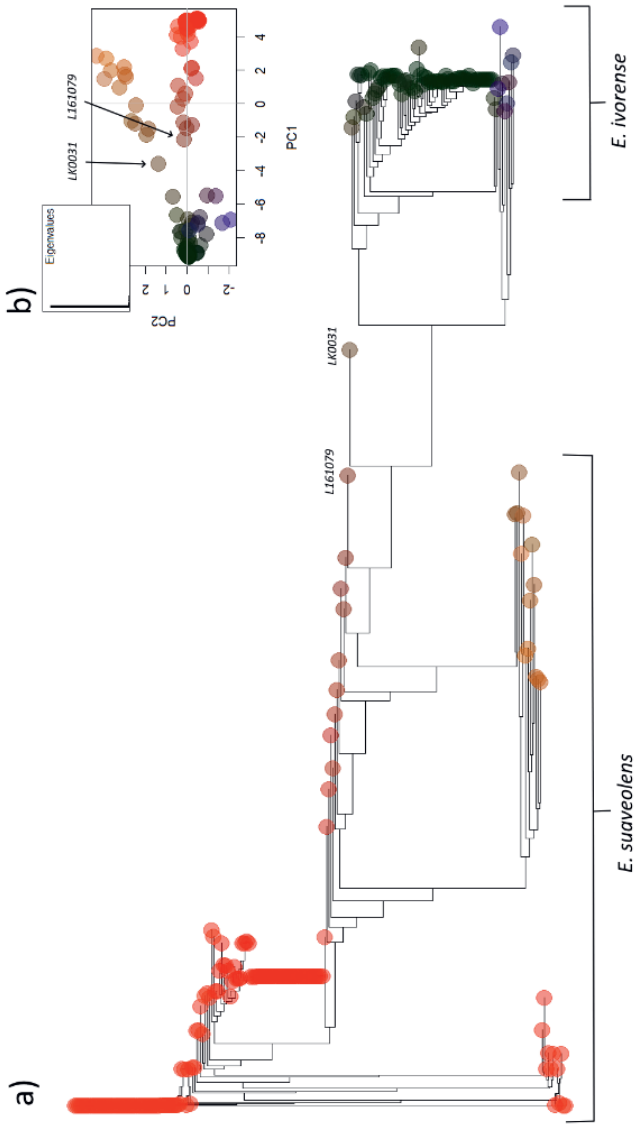
17. Centrifuge 10 min, 13000 rpm, 4°C
18. Wash the pellet in 1 ml 96% ethanol
19. Centrifuge 10 min, 13000 rpm, 4°C
20. Decant liquid, spin down again for a few seconds and pipet off the rest of the liquid
21. Dry the pellet in a heat block at 37°C for 20 min (lids open!)
22. Add 30 µl AE (Quiagen kit) resuspend the pellet
23. Incubate the samples at 4°C overnight, so that the pellet resuspends completely.

Supporting table S2.2 Botanic vouchered samples of *E. suaveolens* and *E. ivorense* specimens.

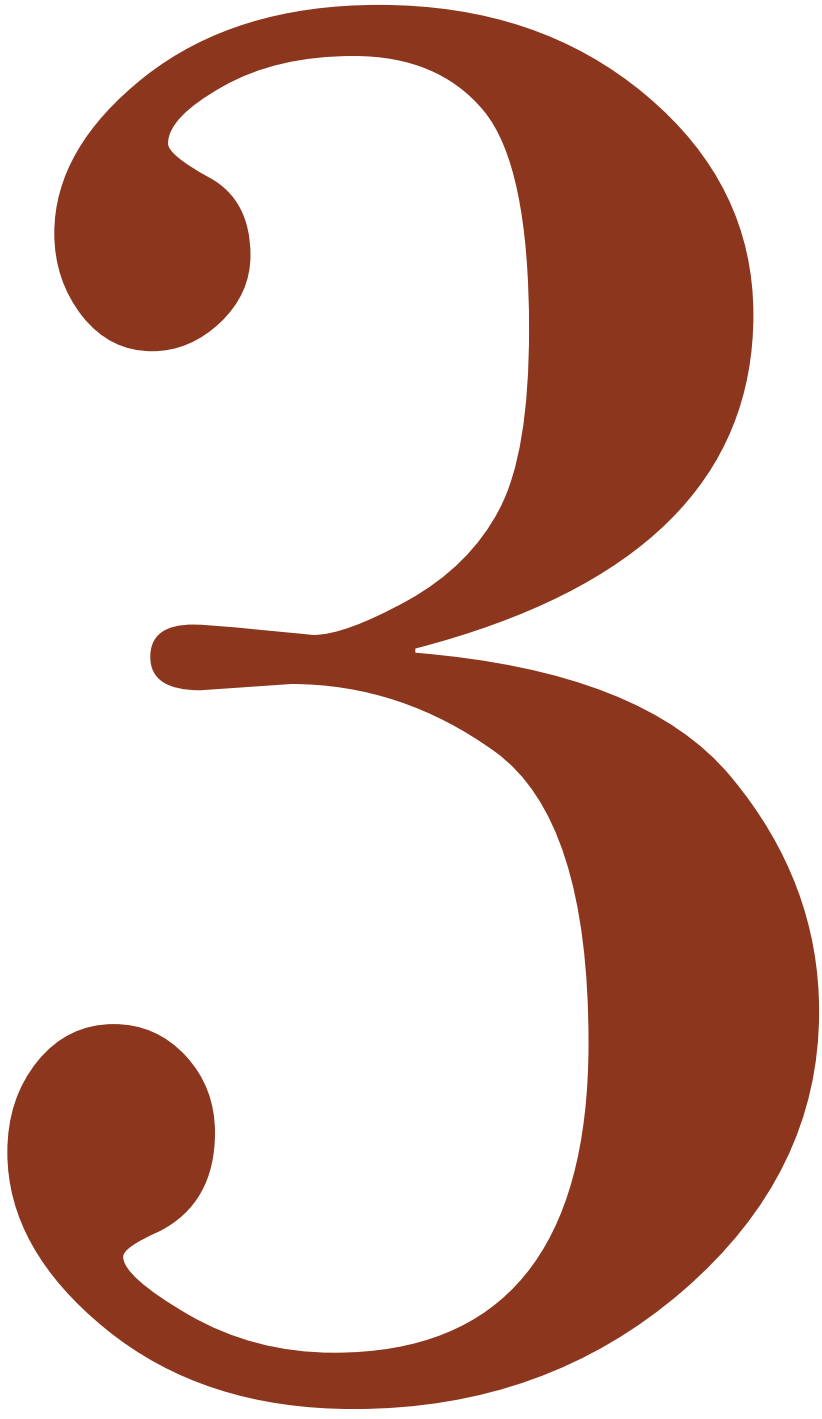
Tree ID	Species	Country	Concession	Latitude	Longitude	N° herbarium
FM5038	*E. suaveolens*	Cameroon	Nké, Cameroon	4.72774	11.64765	Nké, Cameroon
RM0021	*E. suaveolens*	Gabon	Oyem (TTIB)	1.6521	12.15868	MR 610
GiD0850	*E. ivorense*	Gabon	Classified forest of Mondah	0.5787	9.3346	DN 902
PM4854	*E. ivorense*	Cameroon	Korup National Park	5.062	8.854	PM4854
PM4994	*E. ivorense*	Cameroon	Korup National Park	5.062	8.854	PM4994



Supporting figure S2.3 : Bayesian Information Criterion (BIC) versus the number of clusters including all *Erythrophleum* individuals. The value a BIC sharply drops until K=3, than the slope softens between 3 and 4, and BIC values steadily dropped until K=11.



Supporting figure S2.4 : The colour correspondence between the Neighbour-joining tree and PCA analysis. a) Neighbour-joining tree including all *Erythrophleum* individuals, and indicating the species and the trees with unreliable identification (L161079 and LK0031). b) PCA of all samples indicating the trees with unreliable identification (L161079 and LK0031). Data used in both plots contained NAs.



Chapter 3

Unlocking the geography of Azobé timber (*Lophira alata*): revealing genetic geographic signals beyond species boundaries through plastome-wide SNPs

Unlocking the geography of Azobé timber (*Lophira alata*): revealing genetic geographic signals beyond species boundaries through plastome-wide SNPs

Bárbara Rocha Venâncio Meyer-Sand, Laura E. Boeschoten, Gaël U.D. Bouka, Jannici C.U. Ciliane-Madikou, Nestor L. Engone Obiang, Mesly Guieshon-Engongoro, G. Arjen de Groot, Olivier J. Hardy, Simon Jansen, Joël J. Loumeto, Dieu-merci M.F. Mbika, Cynel G. Moundounga, Rita M.D. Ndangani, Dyana Ndiade Bourobou, Nathalie de Vries, Steve N. Tassiamba, Martin T. Tchamba, Bijoux B.L. Toumba-Paka, Herman T. Zanguim, Pascaline T. Zemtsa, Marinus J. M. Smulders, and Pieter A. Zuidema

Abstract

More than half of the timber produced by major tropical exporting countries is illegally sourced. Despite worldwide efforts to combat this illicit trade, it continues to pose a significant environmental challenge. The complexity of the trade routes and the multi-source and/or species product composition represents a substantial hurdle to verify claims of species and origin claims. In this study, we evaluate whether the plastid (chloroplast) genome can possibly be applied to verify the species and the origin of Azobé timber (*Lophira alata*), listed as vulnerable on IUCN's Red List due to the high logging pressure. DNA was isolated from the cambium or leaf tissue of 480 trees (*Lophira alata* and the congeneric species *L. lanceolata)* from Central and West Africa. Complete plastid genomes were assembled, and a total of 228 SNPs from 436 trees were retained, which formed 35 cpDNA haplotypes. Our analysis revealed no clear genetic differentiation of the plastid genomes between the two *Lophira* species, as both included some shared and several very closely related plastid haplotypes. This genetic similarity may be due to chloroplast capture across these species.

The distribution of haplotypes did reveal a clear spatial structure. Within the commercially important *Lophira alata* we detected a moderately strong correlation between genetic and spatial distances. Our results provide a first step towards using chloroplast haplotypes based on polymorphic SNP datasets to verify claims of geographic origin of Azobé timber. Yet, they cannot be applied to differentiate the two *Lophira* species. To verify the potential of the plastid genome for timber , further steps are needed, including assignment tests and blind sample tests.

Keywords: timber tracing; plastid genome; chloroplast genome; SNPs; *Lophira*; Azobé.

3.1 INTRODUCTION

Illegal timber exploitation is one of the main drivers of the degradation of tropical forests. Trade of illegally logged timber represents an estimated global value of $51 - $152 billion USD annually, accounting for 50-90% of the timber produced by many tropical countries (Hoare, 2015). A number of intensively traded tropical timber species is particularly vulnerable to illegal logging, resulting in serious risks of (economic) extinction (Dormontt et al., 2015; Low et al., 2022). In recent decades, a number of regulations and conventions have been put into place to curb illegal timber trade (e.g. CITES worldwide, EUTR and EUDR in Europe, Lacey Act in USA). Despite these efforts, illegal timber trade remains a large environmental problem, in part due to ineffective legislation and law enforcement. The main challenge to control timber legality is to verify the authenticity of trade documents or declarations of species and geographic origin (Dormontt et al., 2015). Other difficulties often reported by authorities and timber operators are the complex and long trade routes of tropical timber and the frequent mixing of timber batches from different sources. Such mixing may involve timber derived from different geographic origins, timber of different botanic species, and illegal timber with timber with a legal status (Nellemann and INTERPOL Environmental Crime Programme, 2012).

To verify trade declarations and/or to support legal timber trade the use of forensic methods has been frequently proposed (Low et al., 2022; The European Parliament And The Council Of The European Union, 2023). Forensic verification of timber origin is based on intrinsic properties of the wood and includes visual (wood anatomy, machine vision)(Akhmetzyanov et al., 2020b), chemical (NIRS, mass spectrometry, stable isotopes) (Vlam et al., 2018;

Boeschoten et al., 2022), and/or genetic methods. At this moment, genetic methods are most advanced in timber forensics (Lowe et al., 2010; Degen et al., 2013; Blanc-Jolivet et al., 2018; Akhmetzyanov et al., 2020a; Dormontt et al., 2020; for review see Low et al., 2022).

Most genetic approaches proposed so far have used polymorphisms in the nuclear genome to verify timber origin (Blanc-Jolivet et al., 2018;.Blanc-Jolivet and Liesbach, 2015; Low et al., 2022). However, Mascarello et al. (2021) pointed out the potential of the intraspecific variation in highly variable parts of the plastid genome (also known as "plastome") for several African genera. Plastid genomes occur many more copies per cell, compared to the nuclear genome, and may be less prone to degradation (Fahey et al., 2021; Henry, 2005, chap. 4; Mader et al., 2018). Consequently, the likelihood of obtaining DNA sequences and successfully genotyping SNPs (Single Nucleotide Polymorphisms) from small amounts of degraded DNA extracted from wood may be higher when using plastid genomes (and its haplotypes) compared to nuclear genomes. Finally, next generation sequencing methods now allow extracting reads and assembling the whole plastome, instead of amplifying a few predefined regions with limited genetic variation. In the present study we investigate the potential of the plastid genome for species and origin identification of the valuable African timber Azobé (*Lophira alata*).

Azobé (*Lophira alata* Banks ex C.F. Gaertn, Ochnaceae) is commonly applied in hydraulic constructions. It is amongst 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 (Eba'a Atyi and Lescuyer, 2022). The species is currently listed as vulnerable on IUCN's Red List due to the high logging pressure (Plouvier, 1997; Piñeiro et al., 2015; Eba'a Atyi and Lescuyer, 2022), which causes a depletion of mature individuals and fragmentation of populations. The congeneric species *Lophira lanceolata,* occurs in the same region (Aubreville 1959; Ewédjè et al., 2020), and is known to hybridize with *L. alata.*

In the present study we investigate the potential of using the whole plastid DNA for species and origin identification of Azobé. We address the following question: 1) What are the main genetic clusters and do these correspond to the above-mentioned *Lophira* species? 2) Are the clusters spatially structured? And 3) can plastid haplotypes be informative for forensic purposes?

3.2 METHODS

3.2.1 Studied species

Azobé (*Lophira alata* Banks ex C.F. Gaertn, Ochnaceae) is a deciduous tree that occurs in the Guineo-Congolian centre of endemism, spanning from Guinea to the Democratic Republic of Congo (GBIF.org, 2021). *Lophira alata* produces small, inconspicuous flowers that are considered to be primarily insect-pollinated (Ouédraogo et al., 2020) large, woody, round nuts known as a "Bongossi nut", which are animal dispersed. To be able to assess the potential of species identification, we also included the closely related species *L. lanceolata*. *Lophira alata* occurs mainly in wet tropical forests, while the closely related relative, *Lophira lanceolata,* occurs in woodlands and dry forests (Aubreville 1959; Ewédjè et al., 2020). The congeneric species can be found in sympatry in contact zones between rainforest and savannah (Piñeiro et al., 2015; Ewédjè et al., 2020). Using nuclear microsatellite loci, Ewédjè et al. (2020) 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 will refer to as *L. alata* WG. However, their study did not assess the potential to differentiate geographic origin. Blanc-Jolivet et al. (2021) developed a single nucleotide polymorphism (SNP) set for *L. alata* from West and Central Africa, including 75 nuclear, 20 chloroplast and 28 mitochondrial SNP markers, and were able to detect the origin with a theoretical performance of 86%, based on leave-one-out test.

3.2.2 Sampling area and strategy

The *L. alata* and *L. lanceolata* samples were collected from two sources: (a) 96 herbarium 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, and (b) 384 newly collected cambium samples of L. alata between 2019 and 2022 in 15 concessions in Cameroon, Gabon and Republic of Congo. Type-b samples were obtained from sites at distances ranging from 15 km up to 1050 km apart (Figure 3.2). Sampled trees within each of these sites were at least 100 meters apart and at most 5000 m, and were at least 30 cm in diameter at breast height. All trees were georeferenced, and all sample types were dried with silica gel.

3.2.3 Laboratory analysis

DNA was isolated from leaf or cambium material of 480 trees with an optimized cetyltrimethyl ammonium bromide (CTAB) protocol as described by Dumolin et al. (1995) with additional cleaning steps (APPENDIX 3.S1). 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, 300 bp or more insert size, 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).

3.2.4 Data analysis

3.2.4.1 Bioinformatics

The Illumina sequences of the 480 trees were mapped to the annotated chloroplast genome of Azobé (MZ274135.1, Mascarello et al., 2021) using Bowtie2 (Langmead & Salzberg, 2012). The variant call considered all mapped reads without filtering and was performed using NGSEPcore (Tello et al., 2019). A variant call file containing only biallelic loci was generated, in which heterozygous were maintained only when at least one sampled tree was homozygous for the minor allele. Using R version 4.1.0 (R Core Team, 2021) the detected variants underwent further filtering. Specifically, we excluded SNPs with sequencing depth of below five reads and above 250 reads, and more than 10% of missing data (SNPfiltR package, DeRaad 2023). Genotypes that did not fulfil the criteria mentioned were considered as missing data. Individual trees with more than 10% missing data across SNPs were also removed from the dataset (SNPfiltR package, DeRaad 2023). The effect of filtering on clustering patterns was evaluated by PCA analyses (see description below). These analyses revealed that the number and distribution of clusters was rather robust to filtering procedures. In addition, samples shown in between clusters in the PCA usually had more than 10% missing data (Supporting Figure S3.1). Our filtering procedure resulted in a dataset of 228 SNP sequences of 436 trees (408 *L. alata* and 28 *L. lanceolata* individuals).

3.2.4.2 Phylogenetic analysis and data structure

We evaluated the primary structure of the data based on a Randomized Axelerated Maximum Likelihood (RAxML) phylogenetic tree (Stamatakis, 2006). For this analysis heterozygotes 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 for the 436 trees. A Bayesian information criterion (BIC) (Adegenet package) was computed to determine an optimal number of sub-clusters of samples within the three main lineages (as perceived from the phylogenetic tree). The sub-groupings for K=10 (to account for observed substructure of the BIC analysis) were assessed with a discriminant analysis of principle components (DAPC; Adegenet package). We chose to perform a discriminant analysis of principle components (DAPC) as they are widely recognized tools for analyzing nuclear DNA and applied in other research fields, which makes results of the present study comparable with other tracing studies (e.g. chemical or visual tracing). The DAPC (Adegenet package v2.1.8, (Jombart, 2008)]) was performed on the full dataset of 436 trees and 228 SNPs. Post-probabilities of membership were calculated for K=10. The consistent sub-clustering patterns observed in both the phylogenetic tree and DAPC analyses led us to opt for the utilization of the DAPC to depict the structure observed in the data. Given the congruence, we will employ the terms "clusters" and "lineages" interchangeably. To visualize the geographic structure of these clusters, cluster membership was plotted on a map (QGIS).

3.2.4.3 Haplotypes' characterization, distribution and relationship

The SNPs subset of 179 SNPs for the 436 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 (Bandelt et al., 1999) to assess their relationship (PopArt, https://popart.maths.otago.ac.nz), and combined this with the haplotype distribution map (QGIS).

3.2.4.4 Genetic versus geographic distances

Genetic distances were calculated for the 15 sites at which more than 17 trees were sampled (n= 362 trees, 228 SNPs). Genetic distance was defined as the proportion of unshared alleles

per population pair, defined as 1-(Proportion of Shared Alleles) [PopGenReport package, (Gruber and Adamack), 2015]. Geographic distances were calculated as pairwise distances between centers of sites. We conducted a Mantel test with 10,000 permutations to evaluate the significance of the relationship between genetic and geographic distances among the sites.

3.3 RESULTS

3.3.1 Main genetic clusters, their distribution and correspondence to species

The results of the phylogenetic tree analysis (inset in Figure 3.1a) shows a division into three main clades. The first clade, K1 (orange Figure 3.1a) includes samples from *L. lanceolata* and *L. alata* from extreme West Africa, from the Dahomey gap to the West. The second clade, K2, is composed exclusively of *L. alata* from two sites in West Gabon (likely the cryptic *L. alata*, here referred to as "*L. alata*-WG" [Ewédjè et al.,2020]). The third clade, K3, contains samples from all other sites in Central Africa (from the Dahomey gap into Central Africa), including *L. lanceolata* (grey triangles, Figure 3.1b). There were clear genetic differences between *L. lanceolata* and *L. alata* in Central Africa, shown by grey triangles clustered separately from other K3 sub-clusters (K3d, Figure 3.1b and on Phylogenetic tree). The grouping suggests a stronger dominant geographic signal (West vs Central Africa) than species signal, as *L. lanceolata* from West vs Central Africa clustered separately.

The analysis of sub-structure using successive k-means in Adegenet identified an optimal number of 10 clusters (Supporting Figure S3.2). 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; Figure 3.1b). 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 (Supporting Figure S3.3).

The 10 sub-clusters were spatially structured at varying scales, and/or species-specific (Figure 3.1b). The sub-cluster K1a encompassed *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 occurred in the region where the cryptic *L. alata*-WG species occurs (Ewédjè et al., 2020). The sub-cluster K2a

exclusively consisted of *L. alata* specimens from two sites (GAB1 and GAB7), and K2b comprised *L. alata* solely from the GAB4 site. Five of the K3 sub-clusters (except K3d) were restricted to Central Africa, whereas K3d spanned from Central Africa into the Dahomey Gap. With the exception of K3d, all other sub-clusters exclusively consisted of *L. alata* trees. K3a was mostly present in Cameroon and Gabon, while the K3b cluster presented a wider distribution, in Cameroon, Gabon and Congo. Cluster K3c solely consists of trees from nine sites in Cameroon. The cluster K3d was composed of *L. lanceolata* from Benin (three sites) and Cameroon (six sites). Cluster K3e was composed of only *L. alata* from Congo, and four individuals from neighboring sites in Cameroon and Central African Republic. Cluster K3f comprised scattered individuals from Cameroon and DRC.

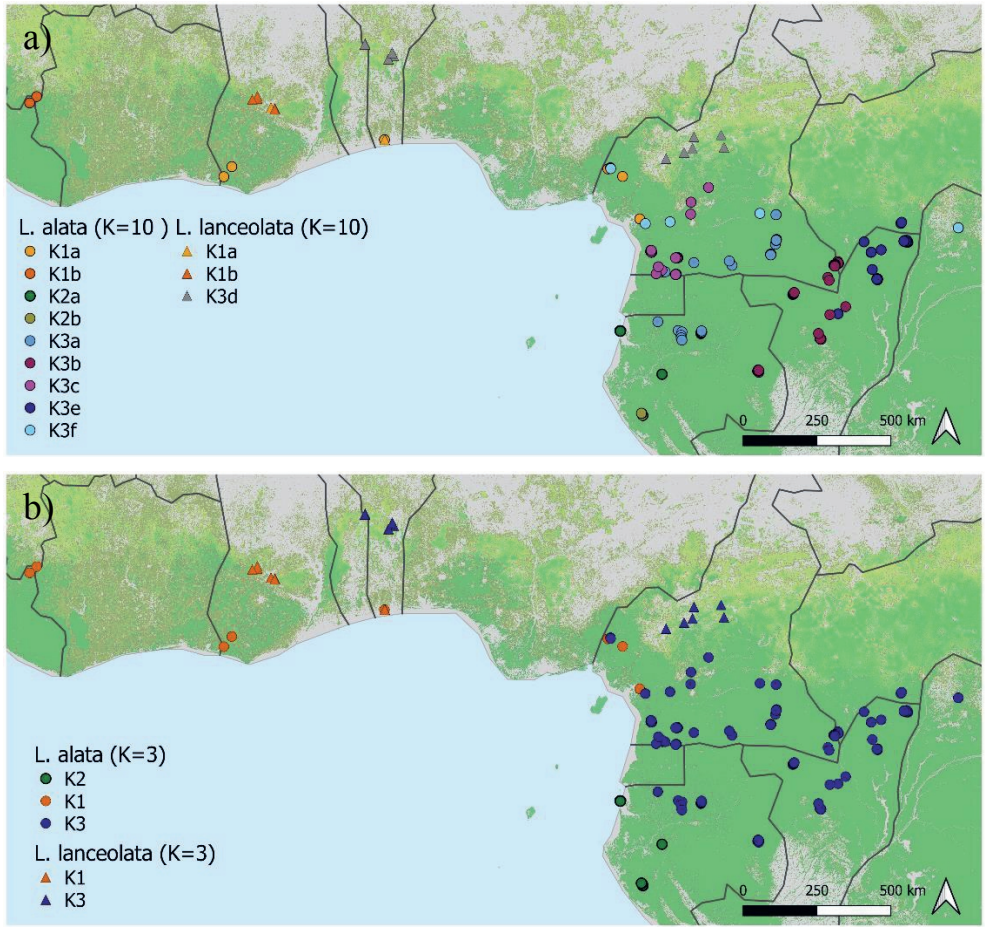


Figure. 3.1: Spatial distribution of three genetic clades (a) and 10 sub-clades (a) for Azobé (*Lophira alata* and *L. Lanceolata*) from West and Central Africa. Both analyses were performed with all 436 trees. The main three lineages were based on 179 cpSNPs and were defined as perceived in the RAxML phylogenetic tree. The 10 sub-lineages were based on 228 cpSNPs 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.

3.3.2 Haplotype distribution and network

A total of 35 haplotypes were detected, 23 for *L. alata*, 11 for *L. lanceolata* and one shared between the two species (H1.10) (Table 3.1; Figure 3.2). We observed S=128 segregating sites, and π=0.21 nucleotide diversity. The *L. lanceolata* haplotypes were present in clades K1 and K3 (indicated with dotted lines), and closer to *L. alata* haplotypes from the same

geographic region of origin than to each other, consistent with the intertwingled grouping patterns observed in cluster analysis and on the phylogenetic tree. *L. lanceolata* presented one common haplotype (H3.12) present in eight trees from Benin. The algorithm implemented in PopArt further removed loci with more than 5% missing data, therefore the shared haplotype H1.10 (indicated by * in Figure 3.2b) was combined with H1.8 in the haplotype network. The shared haplotype H1.10 was closely related to haplotypes of both species.

Within *L. alata*, the two most common haplotypes were H3.7 and H3.17, which were seven SNPs apart. They represented over half of the trees in the dataset, with 118 and 117 trees, respectively (Figure 3.2). H3.7 was central in the network and occurs in East-Cameroon and East-Gabon, and West Congo, whereas H3.17 occurs in Central Gabon and Cameroon. Other common haplotypes (H3.3, H2.6, H3.15, H3.16, and H3.2) were present in 14-39 trees each (Table 3.1). Haplotype H3.2 was restricted to Congo, and (with H3.3) was the most genetically distinct haplotype in the Clade K3 with 13 mutations apart from the closest haplotype. H3.15 and H3.16 exclusively occurred in sites in West Cameroon (CAM1, CAM2 and CAM3). Haplotype H2.6 is part of the most genetically distant branch of the network, that encompass other very genetically distant haplotypes (H2.1-H2.6, shades of green) and more than 25 mutations away from the closest haplotype (H3.4). All haplotypes in the K2 branch are confined to West Gabon. On the other hand, 12 haplotypes occurred in only 2-7 individuals (also colored on Figure 3.2a and b) and the remaining 4 haplotypes were only present in one individual each.

A few sites displayed very characteristic haplotypic compositions. For instance, while most sites in Cameroon presented individuals belonging to H3.17 and H3.7, the site CAM1 and CAM3 were mostly composed of individuals presenting haplotype H3.14-16. Similarly in Gabon, the sites GAB1 and GAB4 had a distinct haplotypic composition from other Gabonese sites. The haplotype network branch arrangement was in line with the distribution of clades K1-K3 distribution in the phylogenetic tree, and with the spatial structure of the haplotypes.

3.3.3 Spatial genetic structure

Genetic distances amongst the 15 sites increased with geographic distances. In Figure 3.3, two point clouds were observed. In the upper cloud, the proportion of different alleles spans between 25-29%. Here, every pairwise comparison is between one of the two distinct sub-cluster sites in West Gabon (K2a - site GAB1 and K2b- site GAB4) with all other sites. The

lower cloud presents the proportion of different alleles spanning from 0 to 12%, with the highest genetic distance between GAB1 and GAB4, followed by pairwise distances including one of the two North Congolese sites (K3e, sites CON2 and CON3) ranging from 9 to 10%. The gap between 12 and 25% of the proportion of different alleles indicates that different biological processes are in place in GAB1 and GAB4, most likely speciation. This inference is supported by the fact that the sites are located where *L. alata*-WG has been reported to occur (Ewédjè et al. 2020). Therefore, we performed a Mantel test only for the lower point cloud, restricting the analysis to comparisons between sites with individuals of the same species. Genetic distances increased with increased geographic distances, with a moderate to strong Mantel correlation coefficient ($r=0.54$, $p<0.001$, 10000 replicates).

Table 3.1. Haplotype frequencies per species. * H1.10 is shared between *L. alata* and *L. lanceolata*.

Clade	Haplotypes	Haplotype frequencies		
		L.alata	*L.alata* -WG	*L.lanceolata*
K1	H1.5	1	0	0
	H1.6	1	0	0
	H1.8	1	0	0
	H1.9	4	0	0
	H1.11	2	0	0
	H1.10	1	0	4
	H1.1	0	0	1
	H1.12	0	0	2
	H1.2	0	0	1
	H1.3	0	0	1
	H1.4	0	0	3
	H1.7	0	0	2
K2	H2.1	0	1	0
	H2.2	0	4	0
	H2.3	0	6	0
	H2.4	0	7	0
	H2.5	0	5	0
	H2.6	0	15	0
K3	H3.1	3	0	0
	H3.2	39	0	0
	H3.3	14	0	0
	H3.4	2	0	0
	H3.5	2	0	0
	H3.6	5	0	0
	H3.7	118	0	0
	H3.13	3	0	0
	H3.14	5	0	0
	H3.15	22	0	0
	H3.16	30	0	0
	H3.17	117	0	0
	H3.8	0	0	1
	H3.9	0	0	1
	H3.10	0	0	1
	H3.11	0	0	3
	H3.12	0	0	8

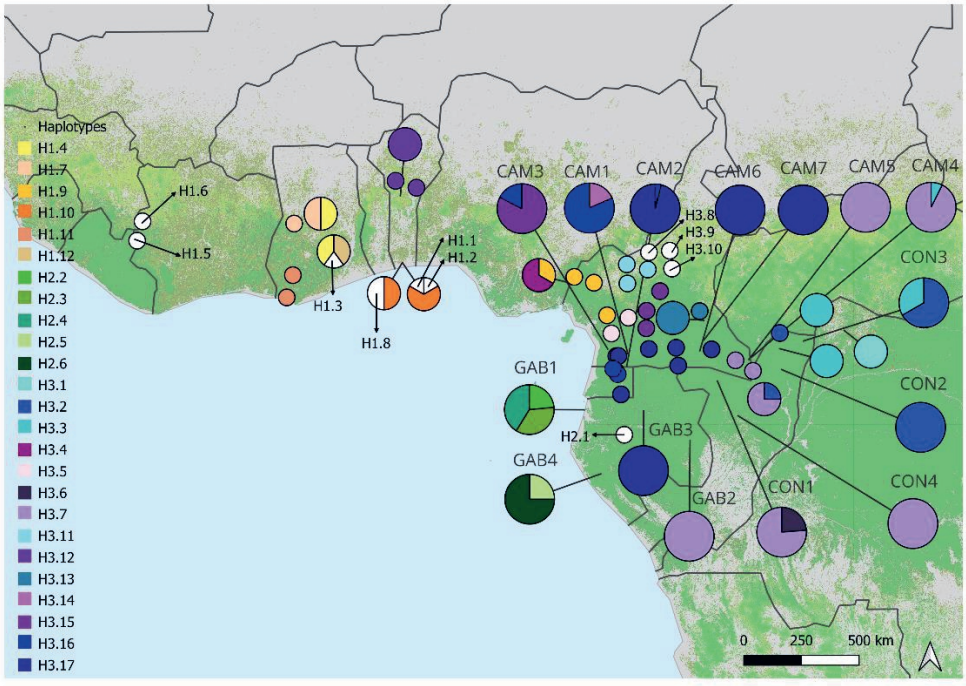


Figure 3.2 (part 1 of 2) : Haplotypes identified for Lophira from West and Central Africa. (a) Geographic distribution of haplotypes in the study area. The size of the circles reflects the sample size in the sites, with for large circles n≥ 17 trees, for intermediately sized circles between 2-6 trees, and for small circles n=1. (b) Haplotype network based on 228 SNPs of the chloroplast 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 indicated haplotypes found in *L. alata-WG*. Haplotype H1.10 is indicated with "*", and it is shared between *L. alata* and *L. lanceolata*. Only haplotypes present in two or more individuals were colored.

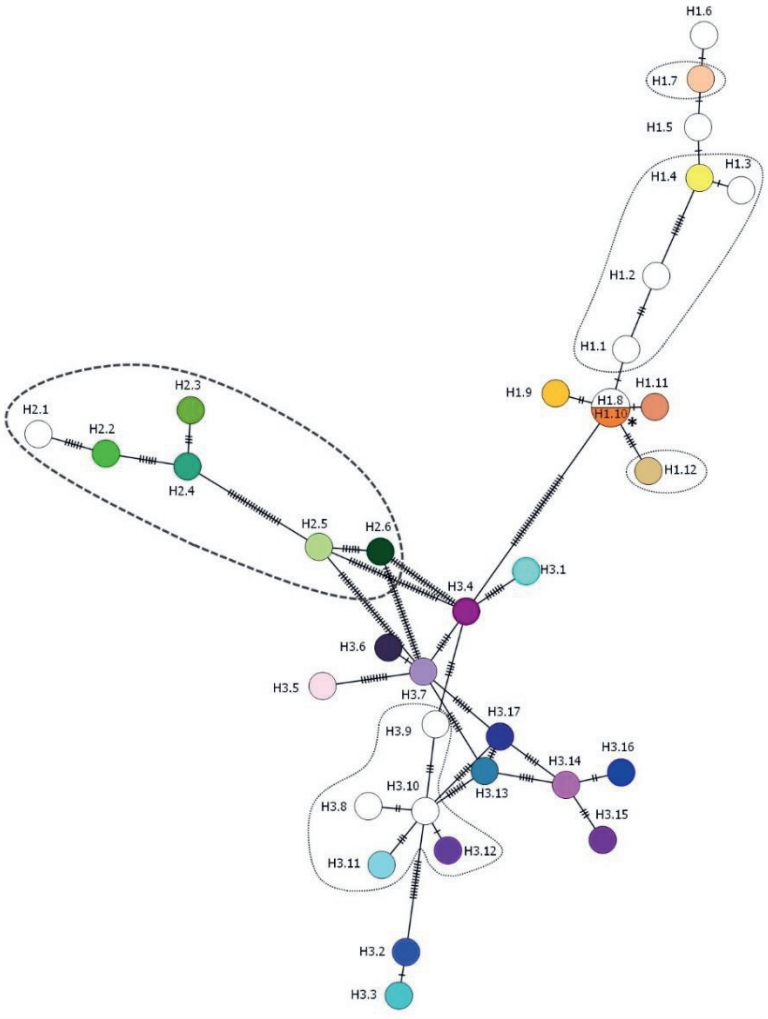


Figure 3.2 (part 2 of 2) : Haplotypes identified for Lophira from West and Central Africa. (a) Geographic distribution of haplotypes in the study area. The size of the circles reflects the sample size in the sites, with for large circles n≥ 17 trees, for intermediately sized circles between 2-6 trees, and for small circles n=1. (b) Haplotype network based on 228 SNPs of the chloroplast 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 indicated haplotypes found in *L. alata-WG*. Haplotype H1.10 is indicated with "*", and it is shared between *L. alata* and *L. lanceolata*. Only haplotypes present in two or more individuals were colored.

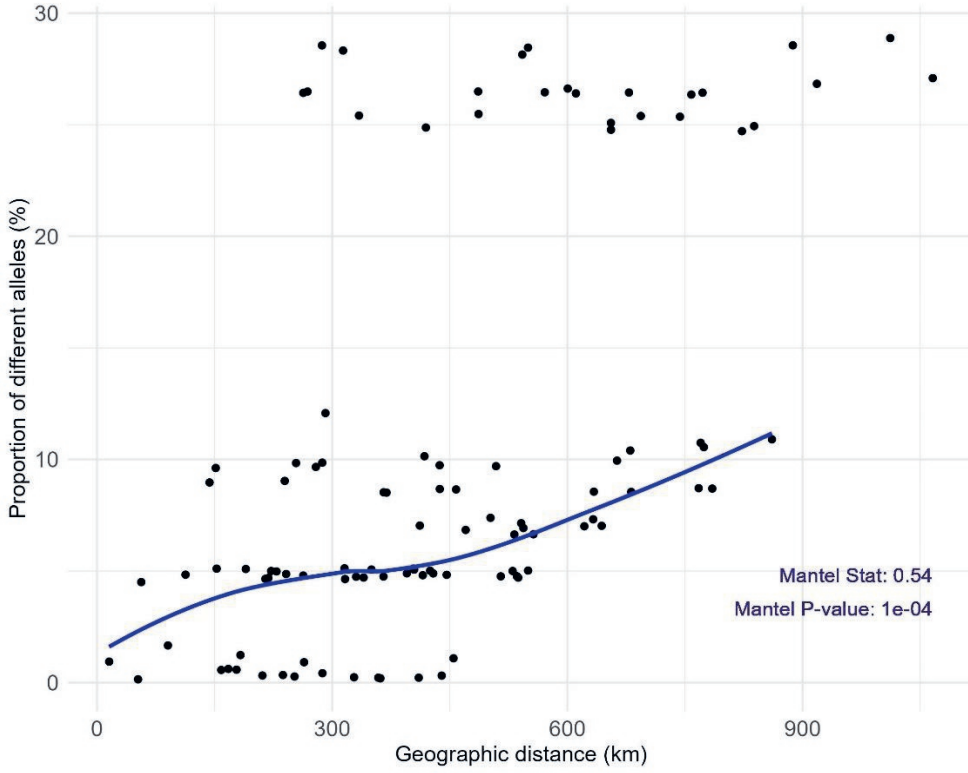


Figure 3.3. Figure Relationship between genetic and spatial distances based on the proportion of distinct alleles. Each point represents a comparison of two out of the 15 sites with >17 sampled trees each (CAM1, CAM2, CAM3, CAM4, CAM5, CAM6, CAM7, CON1, CON2, CON3, CON4, GAB1, GAB2, GAB3 and GAB4). The Mantel test and Loess regression are only conducted for the lower point cloud, which excludes all pairwise comparisons with the genetically very distinct sites GAB1 and GAB4 (these comparisons form the points at the top of the Figure).

3.4 DISCUSSION

To contribute to the development of forensic tracing methods for Azobé, we sequenced the plastid genome of 408 *L. alata* and 28 *L. lanceolata* individuals from nine African countries, covering the largest part of their distributions. We mapped the reads against the reference genome of *L. alata* (MZ274135.1) and detected 228 polymorphic SNPs that could be combined into 35 haplotypes of 179 SNPs in length. Remarkably, half of the sampled *L. alata*

trees had one of two highly common haplotypes. In the remaining trees, many different haplotypes occurred, all with low frequencies.

3.4.1 Species differentiation

The results of our study clearly identified three main clusters, but these 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 distinct geographic regions (Figure 3.1a). The first cluster, K1, is composed of *L. alata* and *L. lanceolata* trees from West Africa. The second cluster, K2, is restricted to the coastal areas and encompasses *L. alata* from West Gabon exclusively. The third cluster, K3, with both species, is mostly present in the core of the Central African region. This cluster also contains some individuals from the inland in Benin. Further analysis revealed that some of the genetic variation was shared between individuals of *L. alata* and of *L. lanceolata* (e.g. cluster K1a composed of *L. alata* and *L. lanceolata* individuals from Cameroon, Benin and Ghana). The complex relationships amongst the haplotypes of both species were clear in the combined network (Figure 3.2b), that did not reveal a clear species separation. The topology of the RAxML phylogenetic tree also indicates that the study species (*L. alata*, *L. alata*-WG and *L. lanceolata*) are not monophyletic from the perspective of their plastomes (Figure S3.3). Hu et al. (2015, 2016) indicated that plastome markers may be of limited use for species delimitation when plastid capture/introgression among species occurred. Indeed, a lack of species-specific polymorphic variation has been reported in the *Lophira* genus (Mascarello et al., 2021). Such sharing of the same (of very similar) haplotypes across co-generic species was also found for other tropical tree genera, including *Carapa* (Duminil et al., 2012), *Greenwayodendron* (Migliore et al., 2019), and *Brachystegia* (Boom et al., 2021). The shared genetic variation between *L. alata* and *L. lanceolata* may be indicative of historical hybridization events (chloroplast capture via introgression), and/or incomplete lineage sorting (Xu et al., 2012; Duminil et al., 2013; Hu et al., 2016; Migliore et al., 2019; Ewédjè et al., 2020). The complex demographic history and low substitution rates in pDNA likely led to the retention of ancestral genetic variants within the two species' gene pools.

We observed a clear distinction of the trees of the two sampled sites in West Gabon (GAB1 and GAB4) from all other sites. This observation aligns with the earlier findings of a cryptic

L. alata species in West Gabon which was identified using nuclear markers (Ewédjè et al. 2020). Both Gabonese sites clustered separately from other *L. alata* trees at the two clustering levels (K=3 and K=10), and they were also separated on the RAxML phylogenetic tree (Figure S3.3, K2a and K2b). The clustering pattern was highlighted on the genetic versus geographic distances plot, which shows a lack of genetic distance gradient in every pairwise comparison that includes one of the two West Gabonese sites as the distance if far bigger than among the other sites. Overall, our analyses suggest the existence of ancient lineages in West Gabon as well as a high degree of isolation between the *L. alata*-WG and *L. alata*/*L. lanceolata*.

In spite of the fact that most haplotypes are species-specific, species distinction based on haplotypes may be prone to errors. Firstly, we identified one shared haplotype between *L. lanceolata* and *L. alata*. The shared haplotype was present in one individual identified as *L. alata* and 4 trees identified as *L. lanceolata* (Figure 3.2b, Table 3.1). Secondly, the very complex relationship observed between the haplotypes that were found in trees of the two species (Figure 3.2b) reenforces the difficulty of separating the *Lophira* species based on pDNA, even more so in contact zones where hybridization and chloroplast capture may occur. A possible approach to tackle this specific issue could be to include nuclear markers to help species identification, as pointed out by Hu et al. (2016). In the case of Azobé, the markers developed by Ewédjè et al. (2020) could be used for this purpose.

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*.

3.4.2 Spatial structure in *Lophira alata*

The cluster and haplotype distribution revealed a clear genetic spatial structure, especially within *L. alata*, as indicated by the moderately strong positive correlation found with the proportions of different alleles increasing with spatial distance (Figure 3.3). The K1 sub-clusters were the most widespread ones, occurring from West Africa sites and along the coast region in the Central Africa, followed by K3, which occurred in both the inner land of Benin and in Central Africa. The distinct sub-cluster K3e was not previously detected using nuclear microsatellite markers (Ewédjè et al., 2020), emphasizing the potential power of plastid

markers in revealing population structure and differentiation (Mascarello et al., 2023). Six out of the 10 sub-clusters were spatially contained at moderate to small areas (<550 km) (Figure 3.1b). The K2 cluster had the most confined spread, as it was found exclusively in West Gabon.

Similar to the cluster analysis, *L. alata* haplotypes were spatially structured with varying geographic distribution. Half (234) of the sampled trees had one of the two very common haplotypes, H3.7 and H3.17 (118 and 117 trees, respectively). Trees with such haplotypes occurred in different areas, and therefore these two haplotypes show discriminatory potential between the regions (but not within them). This differentiation is consistent with genetic discontinuities that have been reported for several tree species between Northern and Southern Central Africa (Hardy et al., 2013). The genetic discontinuities can be related to climatic seasonality. In this case, the rainfall seasonality may cause a phenological delay in flowering of *L. alata* (Ouédraogo et al., 2020), that may create a pre-zygotic barrier for gene flow by pollen.

Interestingly, the other 173 *L. alata* trees contained a range of 21 more or less infrequently occurring haplotypes. Less common haplotypes are less likely to be shared between population and their occurrence in populations creates very characteristic haplotypic compositions in certain sites/regions, even if the common haplotypes are also present (Figure 3.2). The characteristic haplotypic composition suggests potential for fine-scale distinction in certain regions. In the region where the common haplotypes are present, we have within Cameroon the sites CAM1 and CAM2 composed of individuals presenting H.14-16. Similarly in Congo, sites CON2 and CON3, and Gabon, sites GAB1 and GAB4, with distinct haplotypic composition from other sites in their respective countries.

3.4.3 Possible use of plastic genome for timber tracing of Azobé

The moderately strong correlation between genetic and geographic distances, and the spatial structure of both clusters and haplotypes, provide support to the possible use of pDNA to trace Azobé timber . The genetic spatial structure observed is in line with the high spatial structure reported in studies that used cpDNA markers in tree species in the sub-Saharan region (Duminil et al., 2013; Lompo et al., 2018). Because haplotypes are generally much older than the current populations, it is to be expected that over small geographical distances

haplotypes may occur from lineages that are genetically quite distinct from each other. An example for the higher genetic variation amongst populations in cpDNA than nSSR (Deng et al., 2020) is subcluster K3e in Congo, which was not detected using nSSR by Ewédjè et al. (2020).

Even though as much as half of the trees belong to just two haplotypes, in many instances individuals or sites that are geographically close are still very distinct genetically (e.g., within Gabon, North Congo).

Our results provide a first step in the evaluation of the potential to use haplotypes for tracing of Azobé timber. A full evaluation of that potential requires further steps, including conducting self-assignment tests and tests with blind samples (i.e., with trees of a priori unknown origin). Several assignment methods exist, including the Training, Holdout & Leave-one-out (THL, [Anderson, 2010]) that may reduce the upward bias assessment of correct assignments, Random Forest analyses including the THL principles, and/or the Nearest Neighbour Approach that can deal with sub-optimal datasets (Degen et al., 2017).

3.5. CONCLUSION

We showed that species identification of the two *Lophira* species solely based on pDNA haplotypes is hampered, possibly due to hybridization (chloroplast capture) in contact zones as well as because of the complex evolutionary relationship among the haplotypes. Yet, a third, cryptic *Lophira* species could be reliably distinguished genetically. Across samples from all *Lophira* species, we detected a moderately strong correlation between genetic and spatial distance, with haplotypes being spatial structured at varying spatial scales. Our results suggest that haplotypes resulting from polymorphic SNP datasets can possibly be used for origin verification of Azobé timber, yet further tests are needed to fully assess this potential.

3.6 ACKNOWLEDGEMENTS

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 from WorldForestID. We thank all the collaborating timber companies and their field teams for facilitating the fieldwork and all involved colleagues at our partner

institutes University of Dschang, Marien Ngouabi University, IRET/CENAREST and the National Herbarium of Gabon for help in arranging fieldwork.

3.7 PERMITS AND AUTHORIZATIONS

Cameroon: Research Permit No. 00000116/MINRESI/B00/C00/C10/C12 (Yaounde, 09 Sep 2019); Research Permit No. 000066/MINRESI/B00/C00/C10/C12 (Yaounde, 07 Jun 2021); Scientific research permit No. 2144 PRBS/MINFOF/SETAT/SG/DFAP/SDVEF/SC/NGY (Yaounde, 23 Jul 2021); ABS Permit 00010/MINEPDED/CNA/NP-ABS/ABS-FP (Yaounde, 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

Supporting information

Supporting note S3.1. Optimized cetyltrimethyl ammonium bromide (CTAB) with additional cleaning steps.

Extraction of DNA with CTAB

Necessary buffers:

- NaCl 5 M: 29,22 g for 100 mL
- EDTA 0.5 M pH8.0: 18,61 g EDTA + 2 g NaOH per 100 mL (to adjust pH)
- Tris - HCl 1 M pH8.0: 8.88 g TrisHCl + 5.3 g Tris Base for 100 mL (to adjust pH)
- Cold Isopropanol (-20 ° C)
- Cold 70% ethanol (-20 ° C)
- Ethanol 95% cold (-20 ° C)
- Chloroform : isoamylalcohol (24 : 1)
- RNase A 20 mg/mL
- TE pH 8.0: 1 mL Tris 1 M pH8.0 + 200 µl of EDTA 0.5 M pH8.0 to 100 mL
- H2O milliQ

For 25 mL CTAB buffer:

Volume/quantity to	Final concentration
0.5 g CTAB	2% CTAB
7 mL of NaCl 5 M	1.4 M NaCl
1 mL of EDTA 0.5 M	20 mM EDTA pH 8.0
2.5mL of 1 M Tris - HCl	100 mM Tris - HCl pH 8.0
Storage at room temperature	
1 mL DTT	2% DTT
1 g PVP	4% PVP
0.5 mL Proteinase K (20 mg/mL)	0.4 mg/mL Proteinase K

Protocol:

Weigh approximately 100 mg of powder in a 2 mL tube.

1. Prepare a master mix with 22.560 mL CTBA buffer (with PVP) + 960 mL DTT + 480 µL Proteinase K (20 mg/mL). OR 40µl of DTT and 20µl of Proteinase K per sample/tube.
2. Add 900 µl of CTAB buffer pre-heated at 60 °

C. Incubate at 60 °C for 1 hour stirring occasionally (every 10 min).

3. Add chloroform : isoamylalcohol (24 : 1) 1V or 1 mL. **(900 µl)** Mix 1 min by inverting the tubes. Centrifuge 10 min at 14 000 x g. Retrieve the upper aqueous phase and transfer it into a clean 2 mL tube.

4. Add 10 µL of RNAse A (20 mg/ml). Incubate 1 hour h at 37 ° C.

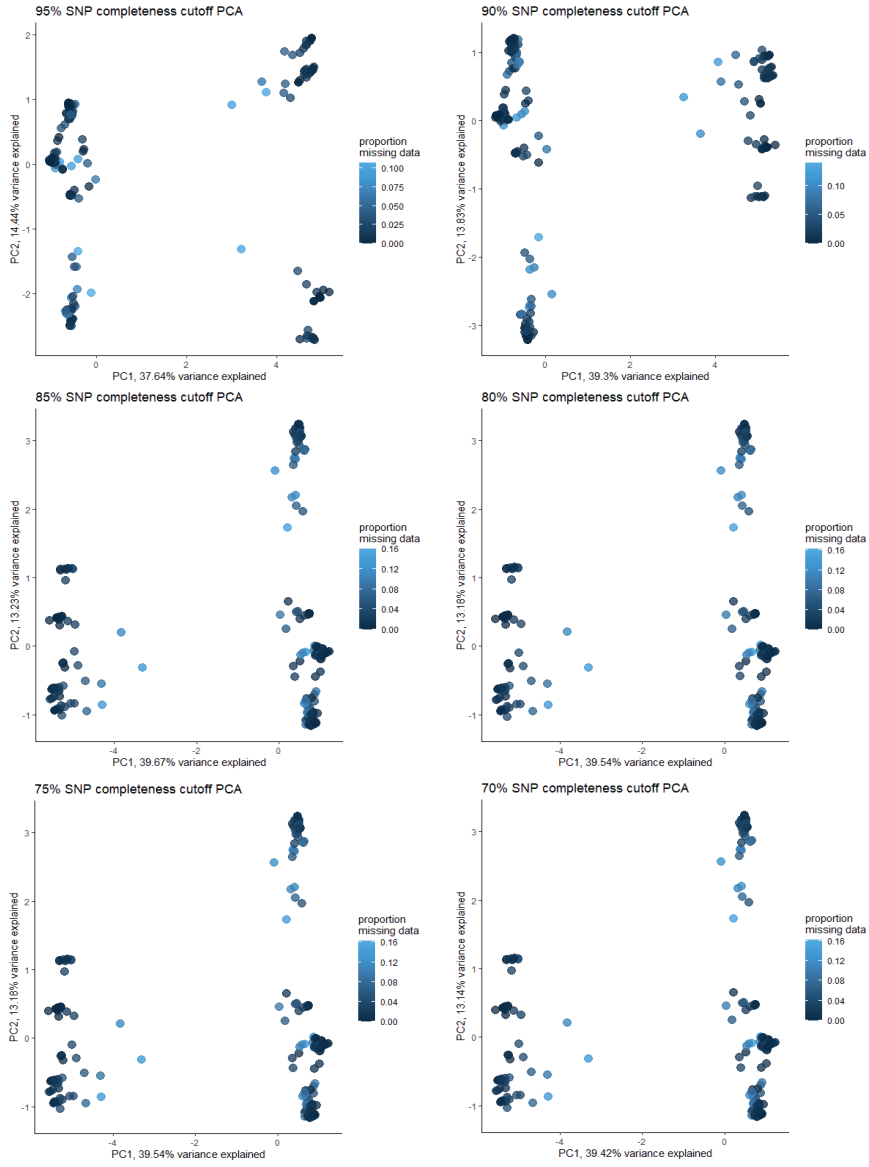
5. Add chloroform : isoamylalcohol (24 : 1) 1V or 1 mL. Mix 1 min by inverting the tubes. Centrifuge 10 min 14 000 x g. retrieve the upper aqueous phase and transfer it into a clean tube of 2 mL.

6. Estimate the amount recovered. Add 2/3 volume of cold isopropanol. Leave ½ hour minimum at -20ºC. Centrifuge at max speed max for 15 min at 4ºC.

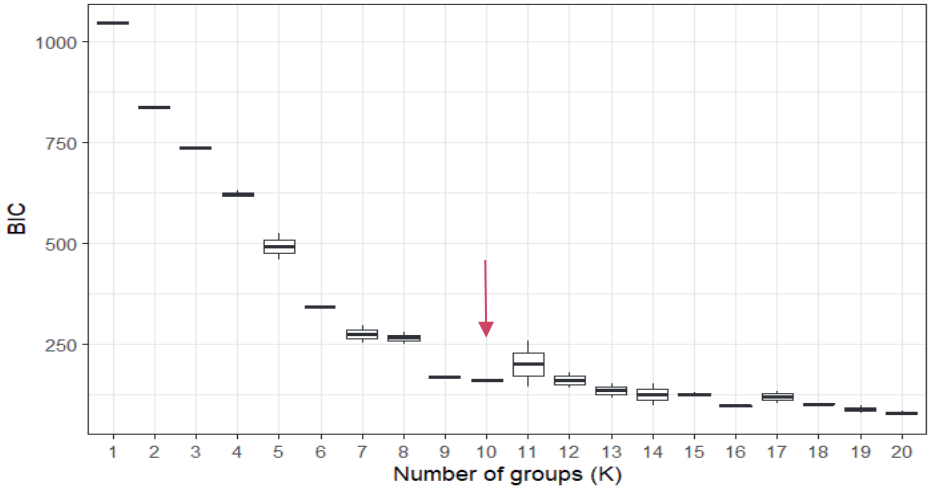
7. Remove the supernatant. Add 700 µL of EtOH 70% cold. Invert the tube several times to loosen the pellet. Centrifuge at max speed for 5 min at 4°C.

8. Remove the supernatant. Add 700 µL of EtOH 95% cold. Invert the tube several times to loosen the pellet. Centrifuge at max speed for 5 min at 4°C.

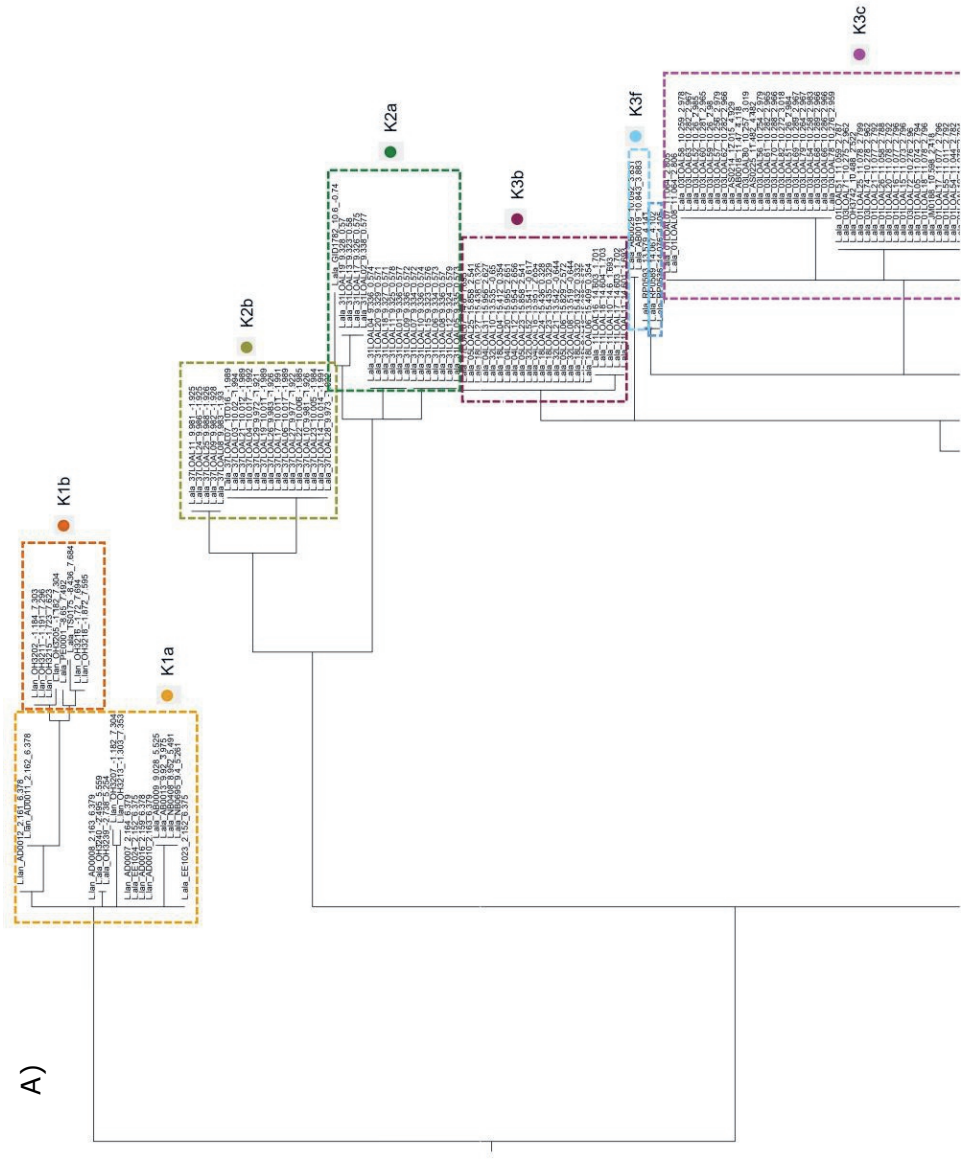
9. Dry the pellet. Resume in 100 µL H_2O milliQ (or TE^{-4}).



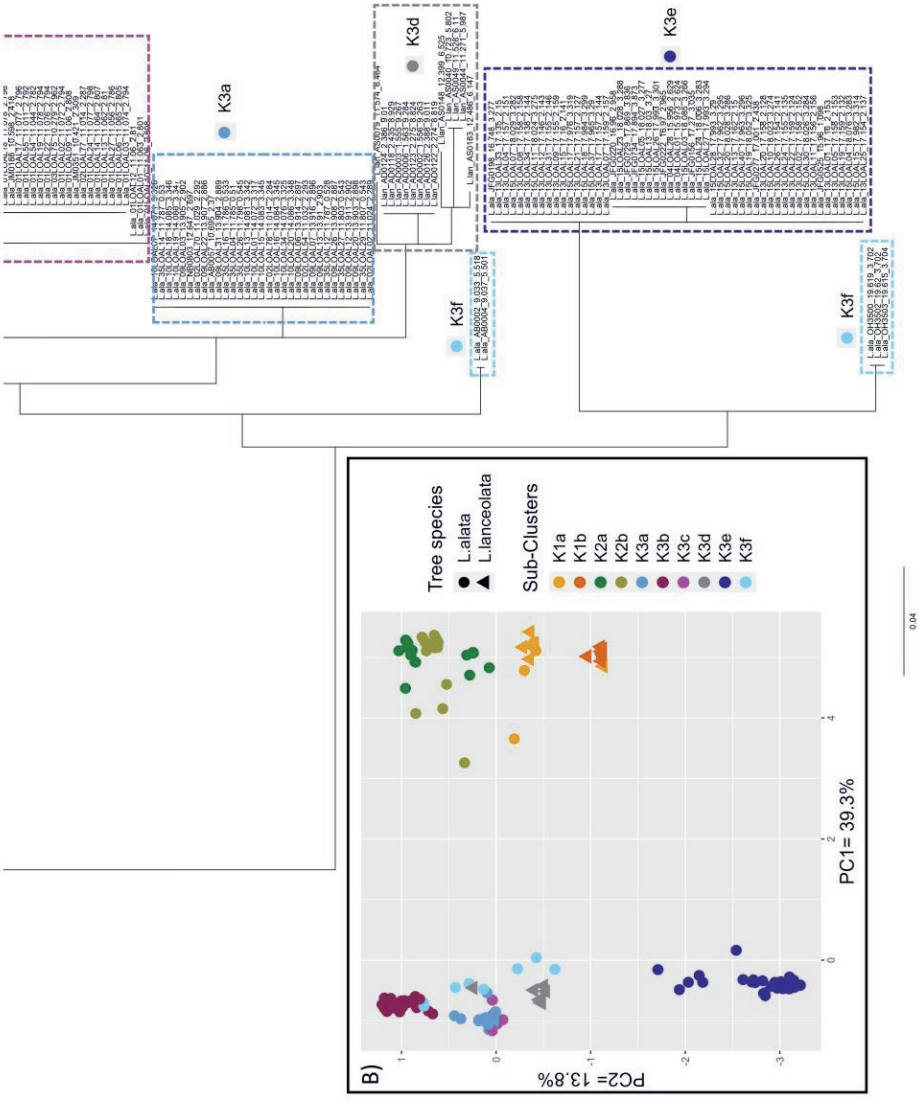
Supporting figure S3.2. Clustering patterns are robust with respect to the choice of thresholds of the SNPs missing data proportion. The threshold varied between 70 - 95% of missing data per SNPs (across individuals) shown on the PCA clusters.



Supporting figure S3.3. K-means clustering based values of Bayesian Information Criterion.



Supporting figure S3.4. (part 1 of 2) The color correspondence between the Randomized Axelerated Maximum Likelihood (RAxML) phylogenetic tree and DAPC analysis equivalence including all *Lophira* individuals. a) RAxML phylogenetic tree, with an indication of (sub-)clusters. b) DAPC with the color equivalence. Data used in the DAPC contained NAs.



Supporting figure S3.4. (part 1 of 2) The color correspondence between the Randomized Axelerated Maximum Likelihood (RAxML) phylogenetic tree and DAPC analysis equivalence including all *Lophira* individuals. a) RAxML phylogenetic tree, with an indication of (sub-)clusters. b) DAPC with the color equivalence. Data used in the DAPC contained NAs.

Chapter 4

Navigating timber trails: using plastid genomes for species and origin verification of two African timbers

Navigating timber trails: using plastid genomes for species and origin verification of two African timbers

Bárbara Rocha Venâncio Meyer-Sand, Laura E. Boeschoten, Gaël U.D. Bouka, Jannici C.U. Ciliane-Madikou, Nestor L. Engone Obiang, Mesly Guieshon-Engongoro, G. Arjen de Groot, Olivier J. Hardy, Simon Jansen, Kadir T. Özdoğan, Joël J. Loumeto, Dieu-merci M.F. Mbika, Cynel G. Moundounga, Rita M.D. Ndangani, Dyana Ndiade Bourobou, Nathalie de Vries, Steve N. Tassiamba, Martin T. Tchamba, Bijoux B.L. Toumba-Paka, Herman T. Zanguim, Pascaline T. Zemtsa, Marinus J. M. Smulders, and Pieter A. Zuidema

Abstract

More than half of the tropical timber produced by major exporting countries is illegally sourced. The complexity of the trade routes and the multi-source and/or multi-species composition of tropical timber represent a substantial hurdle to verifying claims of species and origin claims. We explored the potential of utilizing genetic variation in the plastid (chloroplast) genome to verify the species identity and the most likely geographical origin of timber of Azobé (*Lophira alata*) and Tali (*Erythrophleum* spp.), This was done using reference datasets of plastome-wide SNPs recently developed for West en Central Africa. These sets of SNPs and their geographical locations were used to train two different classification models: Random Forest Classification Model (RFCM) and GeoAssign (GA) models, which were used for self-assignment tests. RFCM assigned 36- 43% of the samples to the population of origin, while this was 5% for GA. For both species, the majority of samples (>60%) was assigned to a population within 200 km of the true origin using RFCM and even more (>80%) for Tali when using GA. To verify the assignment potential of the reference datasets, we produced plastome SNP data for a set of 90 blind samples of which 45

were claimed as Tali and 45 as Azobé. The set included 60 cambium samples, which all yielded useful SNP data, as well as 30 heartwood samples, 19 of which could not be used due to bad data quality. We first extracted data for three barcode regions (rbcL, matK and trnL) to check the claimed species identity. All remaining heartwood samples and 90% of the cambium samples were assigned to the correct taxon. For Azobé, 72% cambium and all four heartwood samples were assigned within 100 km from their true origin (GA), while this was 20% for Tali. This low percentage was likely caused by the fact that the provenance of two thirds of Tali samples was from outside the area covered by the reference dataset. For blind samples within the reference sample area, 60% of the samples were assigned within 100 km. Site assignment of blind samples using GA performed better than RFCM, probably because the latter requires data imputation. Overall, these results suggest a potential for plastid genome analysis to be developed into a timber tracing tool which can identify origin for multiple species and may do so at a within-country scale. To reach the level of a forensic application, reference databases need to be expanded to cover the full geographic areas of the tested species.

4.1 INTRODUCTION

Timber trade contributes significantly to the revenue of tropical countries. It is estimated that 50-90% of the timber exported from major tropical timber producing countries is illegal (Hoare, 2015). Globally, the illegal timber trade is estimated to have an annual volume of $51- 152 billion USD (Hoare, 2015). Illegal timber trade contributes to deforestation and forest degradation and loss of biodiversity. Moreover, it also attracts illegal activities such as corruption, tax evasion, fraud and laundering, and unfair competition with illegal sourced timber products ("Illegal Logging and Deforestation | Basic Page," 2022; INTERPOL, 2020; Nature, n.d.). Among tropical timber, the most valuable timber species are particularly vulnerable to illegal logging and trade.

To counteract illegal timber trade, several international regulations and conventions are in place (e.g., CITES as an international treaty, the Lacey Act in the US, the recent EU Regulation on Deforestation-free Products). Yet, the lack of transparency in the supply chain

and the difficulty to verify trade declarations and/or forged documents hamper enforcement of laws and regulations (Lowe et al., 2016). To fight illegal timber trade, forgery-proof tools to verify trade declaration of species and geographic based on intrinsic timber properties are being developed (Low et al., 2022).

Visual, chemical and genetic methods have been translated from other fields of expertise to verify species claims for a large number of tropical timbers (Bouka et al., 2022; Degen et al., 2017; Hung et al., 2017; Meyer-Sand et al., 2018; Celine Blanc-Jolivet et al., 2018; He et al., 2019; Schroeder et al., 2016; Ng et al., 2016, 2022; Low et al., 2022; Boeschoten et al., 2023). However, methods to verify the geographic origin of tropical timber have received much less attention. Only 24% (23% genetics) of 322 priority timber taxa have been covered, and only a handful of scientific studies have tested the accuracy of these methods, usually by conducting self-assignment tests within the reference dataset. Very few studies tested assignment success of samples from unknown origin or contested claims ("blind samples") (Low et al., 2022). While self-assignment tests evaluate the robustness and performance of reference datasets, conducting blind sample testing is a crucial step to assess applicability in forensic settings. Implementing blind samples' tests enables to test the robustness and the accuracy of the method a) in the geographical areas covered by the reference database, and evaluate the methods b) when confronted with samples from other regions and species, as well as c) the performance of the method for samples of varying DNA quality.

We have developed plastid-wide SNP (Single Nucleotide Polymorphisms) reference datasets for two valuable African timbers: Tali (*Erythrophleum* spp) and Azobé (*Lophira* spp) (Rocha Venancio Meyer-Sand et al., Chapters 2 and 3). We focused on the chloroplast as it occurs in multiple copies per cell, it is less prone to DNA degradation, and it is smaller in comparison to the nuclear genomes. Hence, when using highly degraded DNA in traded timber (heartwood), the use of chloroplast genomes it is more likely to yield the necessary minimum amount of genetic information necessary for origin verification compared to nuclear genomes (Mader et al., 2018a). The low quality DNA in wood also prompted us and others to transition from using the longer SSR markers to shorter SNPs (Morin and Mccarthy, 2007; Degen et al., 2017; Meyer-Sand et al., 2018; Blanc-Jolivet et al., 2018; Blanc-Jolivet et al., 2021). We obtained genetic information from chloroplasts for hundreds of trees sampled across the

distribution areas of our study species in West and Central Africa, with the aim to develop a genetic tool to identify species and detect the likely geographical location of timber samples.

In this study, our objective was to evaluate a method based on cpDNA for verifying both species and geographic origins. First, we performed a self-assignment test to assess the quality and the accuracy of assignment of our reference datasets (Aim 1). The reference datasets were developed using cambium samples with good DNA, to ensure high quality reference data. Subsequently, we tested "blind samples" to evaluate the accuracy of species claims verification by first checking the species identity (DNA barcoding using public databases) (Aim 2). When the species claim was confirmed, we conducted geographic assignment tests using the cpSNPs sets for high-quality (cambium) and low-quality (heartwood) DNA samples, aiming to extend our findings to the traded parts of wood (Aim 3). In analyses for reference samples and blind samples we compared the performance of two statistical assignment methods that differ in assumptions: GeoAssign (GA) (Degen et al., 2017) for genetic data and Random Forest Classification Model (RFCM), a classification machine learning tool used for a wide range of types of data (Wright and Ziegler, 2017) (Aim 4).

4.2 MATERIALS &METHODS

4.2.1 Study species

We studied two valuable and widely traded African timbers: Azobé (*Lophira alata*, Ochnaceae) and Tali (*Erythrophleum suaveolens* and *E. ivorense*). Azobé, a deciduous tree native to the Guieneo-Congolian center of endemism, is known for its inconspicuous insect-pollinated flowers and distinctive Bongossi nuts, dispersed by animals. To assess species identification potential, the closely related *L. lanceolata*, which has distinct habitat preferences, was also included in the sampling. Ewédjè et al. (2020) revealed a cryptic species, *L. alata* WG, endemic to western Gabon. Blanc-Jolivet et al. (2021) further differentiated *L. alata* origins with an SNP set, emphasizing the significance of such distinctions in informing sustainable timber trade practices. The second studied timber, Tali (Fabaceae-Caesalpinioideae), comprises two botanical species: *E. suaveolens* and *E.*

ivorense, both used for heavy structures such as bridges, railway sleepers, terraces and industrial floors (Gorel, 2019; Gorel et al., 2019). Timber of *E. ivorense* is often preferred commercially. Both species are hermaphroditic, likely insect-pollinated (Segers, 2018) and occur from the Guineo-Congolian/Sudanian Transition zone through the Guineo-Congolian zone (Duminil et al., 2015; Gorel, 2019). Morphological similarities between these species hamper identification in the field. Genetic studies by Duminil et al. (2010, 2013) highlighted the potential of chloroplast DNA to differentiate the Tali species.

4.2.2 Reference samples

The reference datasets for Azobé and Tali were developed before, as described by Rocha Venancio Meyer-Sand et al. (Chapters 2 and 3). The initial Azobé reference dataset consisted of 436 trees sampled in forest concessions across West and Central Africa, and the initial Tali reference dataset of 237 trees sampled in forest concessions across West and Central Africa, with a focus in Cameroon. DNA isolation, library preparation and sequencing were performed as described in Chapter 2 and 3, for Tali and Azobé respectively. The plastid genomes were assembled and aligned. The initial filtering produced 184 biallelic SNPs for Azobé and 324 biallelic SNPs for Tali. As several samples had low coverage, we here employed an extra filtering step against potentially unreliable SNPs by checking for homozygous individuals for both SNP alleles, which removed 44 SNPs from the Azobé dataset and 7 SNPs from the Tali dataset.

4.2.3 Blind samples

4.2.3.1 Types and strategy for blind samples selection

Blind samples are samples of a given (claimed) species and (a priori) unknown geographic origin. We used three types of blind samples: (i) trees from within the concessions where we sampled the reference samples (aim a), but not included in our reference database; (ii) trees from within the distribution area sampled, but from sites where no samples were collected for the reference set (aim a); and (iii) trees from outside of the distribution area of the reference dataset and/or from related species not sample before (aim b). For each of the targeted timbers

30 cambium and 15 heartwood samples were used for DNA isolation (aim c). The correct species and origin were only known to one co-author (M.V.) until the assignments had been finalized.

4.2.3.2 DNA extraction and sequencing of the blind samples

DNA isolation was performed using the Qiagen DNeasy Plant Mini Kit with additional cleaning steps. DNA concentrations were quantified using the Qubit™ kit (Thermo Fisher Scientific, Schwerte, Germany). The isolated DNA served as the basis for constructing two paired-ended libraries, one containing cambium and another containing heartwood DNA isolates, each with an insert size of 300 bp or greater. The RIPTIDE High Throughput Rapid Library Prep Kit from Twist Bioscience (South San Francisco, USA) was employed for library preparation. Paired-end sequencing (2x150 bp) was carried out using a Illumina Novaseq6000 (Novogene, Cambridge, United Kingdom).

4.2.3.3 Bioinformatics analyses of the blind samples

For each of the 45 blind samples per timber type, reads were mapped against their respective reference plastomes (chloroplast genomes)—Azobé: GenBank accession number MZ274135.1 and Tali: GenBank accession number MN709858.1—using Bowtie 2.2.5 (Langmead and Salzberg, 2012) with default settings. Duplicate reads were marked using Picard 2.23.0. Then, based on the species claims of the blind samples, plastome-wide SNPs (derived from their respective species' reference set) were called using NGSEP 4.0.1 (Tello et al., 2019) and the MultisampleVariantsDetector, applying parameters '-ploidy 2, -knownVariants'.

4.2.4 Assignment tests

4.2.4.1 Self-assignment tests

We performed the self-assignment using Random Forest Classification Mode (RFCM) and GeoAssign (GA). RFCM as implemented in the Ranger package (Wright and Ziegler, 2017) constructs decision trees using a subset of the features and a subset of the training data.

Subsequently, it combines the results of the individual trees to make a final prediction. The Ranger package for RFCM cannot handle missing data, therefore data imputation was required. For this, missing data in the SNP dataset were replaced with the most common allele for the given SNP within each concession. In case the two most common alleles were equally frequent in a given concession these samples were removed. The final datasets, also used as reference for the blind sample assignment, contained 389 trees with 184 SNPs each for Azobé, and 203 trees with 324 SNPs each for Tali.

To assess the uncertainty associated with RFCM, the reference dataset was split into 25 random subsets of training (70%) and test (30%) trees without overlap. These RFCMs were each run 25 times. For every model, distances between the true and assigned origins were computed and averaged across the 25 replicate RFCMs. The accuracy of RFCM was then calculated as the percentage of correctly assigned test trees at the site level.

The GA tool implements the nearest neighbor approach, which uses genetic distance to assign individual trees to a geographic origin (Degen et al., 2017). The algorithm can handle missing data and thus does not require data imputation, but it requires at least five individuals per concession, hence some of the sampled concessions had to be removed from the dataset. The final datasets for GA (also used for the blind samples) contained 378 trees with 184 SNPs each for Azobé, and 201 trees with 324 SNPs each for Tali. The self-assignment was performed on the reference dataset of each timber against itself (overlapping trees). GA computes the most likely site of origin for each tree based on 5th percentile of most similar individuals and an 80% minimum proportion of complete loci. For each assigned tree, we reported the results of the most likely site of origin, with values of the assignment indices in the different reference sites. The index is the proportion of nearest neighbors, with the weighting based on the size of the respective reference group. We then computed the distances between assigned and correct sites.

4.2.4.2 Blind samples tests

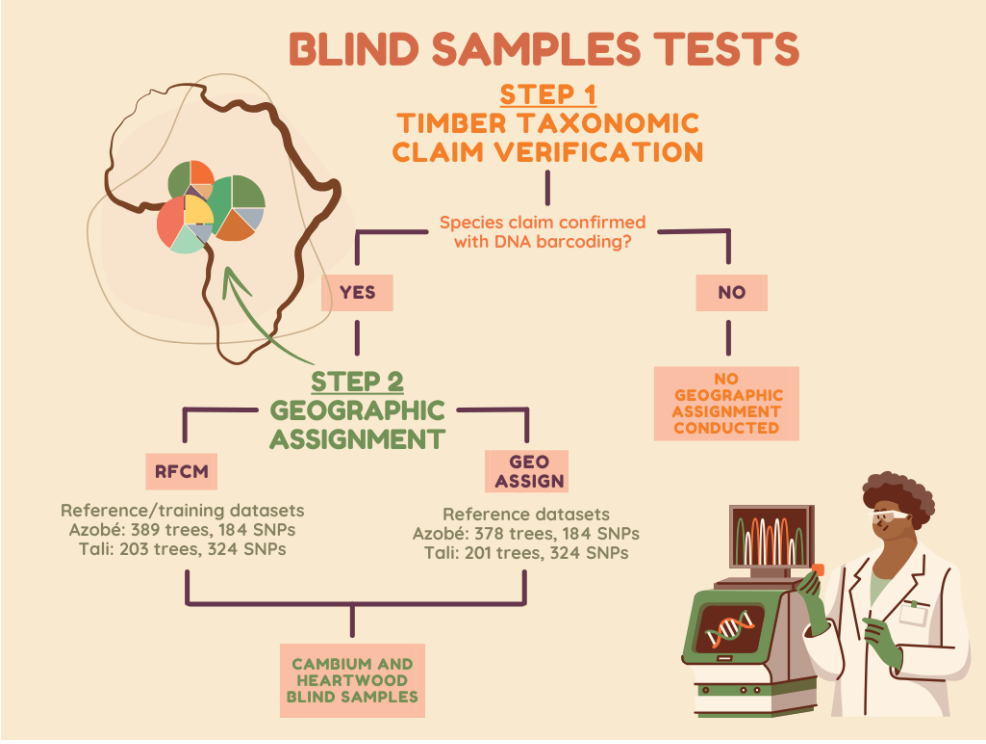


Figure 4.1. Analytical Workflow for Blind Sample Testing. This schematic illustrates the sequential steps undertaken to validate and geographically assign 30 cambium and 15 heartwood blind samples per timber claim, focusing on two specific timber identities: Azobé and Tali. Initially, each sample undergoes taxonomic verification against its claimed timber type. Upon confirmation, samples are subjected to geographic origin assignment through two statistical approaches: RFCM and GeoAssign (GA). The figure also details the reference individual count utilized for each geographic assignment method.

The blind sample analysis was conducted using the following steps (Figure 4.1).

Step 1: Timber taxonomic claim verification (Figure 4.1). To verify the claim of timber (*E. ivorense* and *E. suaveolens*, or *L. alata*) we evaluated the genetic similarity across identical positions in very closely related species. To build the closely related species reference dataset (SPP BLAST dataset) we selected sequences for three known barcode regions: rbcL, matK and trnL and created a local reference database using the makeblastdb function of BLAST+

(Camacho et al., 2009). The SPP BLAST dataset included both species complexes, Tali and Azobé, as the percentage of identical positions for the other species group served as the assessment for number of mismatches expected for non-related species. In addition to Tali we included all co-generic species, *Erythrophleum ivorense*, *E. suaveolens*, *E. africanum*, *E. fordii*, *E. teysmannii*, and *E. chlorostachys*, with the first three species exclusively occurring in Africa and the last three exclusively in Asia and Oceania. In addition to Azobé we included *L. lanceolata* and species of the sister clade Elvasia DC (Schneider et al., 2021): *Elvasia calophyllea*, *E. elvasioides*, *E. capixaba*, *E. essequibensis*, and *E. macrostipularis*. An overview of the DNA barcoded included in the SPP BLAST dataset is given in Table 4.1. The full list of NCBI accession numbers of the barcode sequences are given in Supporting table S4.1.

Before aligning the sequencing data to the local reference database, we performed low complexity filtering (Dust=1), quality trimming ($q \geq 30$), fragment length filtering (>30 bp) and duplicate removal (Kjær et al., 2022). We aligned the pre-processed data to the reference cpDNA sequences according to the species claim by using blastn function of BLAST+ (Camacho et al., 2009). Once aligned, we extracted the three regions (rbcL, matK and trnL) and performed a BLAST to the SPP BLAST dataset. We ranked BLAST output first by read length, and then by percentage of identical reads. We reported the decision made based on the highest identity percentage(s) (above 98%) that presented the longest read (minimum of 70 bp) for each of the three targeted barcode regions that hits were obtained for the individual. The trnL region was only present for *L. alata* in the reference SPP BLAST dataset. The timber assignment was done at the genus level.

Step 2. Geographic assignment (Figure 4.1). The reference datasets employed in the blind sample assignment were the same as used in the self-assignment tests. Blind samples with >20% missing data were removed from the analyses. For GA, pops with >5 trees were kept, and for RFCM, all individuals for which data could be imputed were retained.

Table 4.1. Overview of DNA Barcoded Species in the SPP BLAST Dataset: three known barcode regions (rbcL, matK, and trnL) used to construct a local reference database for closely related species. It includes species complexes for Tali (*Erythrophleum ivorense*, *E. suaveolens*, *E. africanum*, *E. fordii*, *E. teysmannii*, *E. chlorostachys*) and Azobé (*Lophira alata*), along with *L. lanceolata* and species from the sister clade *Elvasia* DC. RuBisCO: Ribulose 1,5-bisphosphate carboxylase/oxygenase.

Related species	Relationship	matK	rbcL	trnL	RuBisCO
Erythrophleum chlorostachys	Related to Tali	1	1	0	0
Erythrophleum ivorense	Tali	8	5	1	0
Erythrophleum suaveolens	Tali	9	2	1	0
Erythrophleum fordii	Related to Tali	4	4	0	1
Erythrophleum africanum	Related to Tali	2	1	0	0
Erythrophleum teysmannii	Related to Tali	2	1	0	0
Lophira lanceolata	Related to Azobé	1	2	0	0
Lophira alata	Azobé	1	1	1	0
Elvasia calophyllea	Related to Azobé	2	1	2	0
Elvasia elvasioides	Related to Azobé	0	1	1	0
Elvasia capixaba	Related to Azobé	0	0	1	0
Elvasia essequibensis	Related to Azobé	0	1	1	0
Elvasia macrostipularis	Related to Azobé	0	1	1	0

4.3 RESULTS

4.3.1 Self-assignment tests

Across methods and timbers, self-assignments accuracy at the site level (<10km) varied from 5% to 43%. Overall, the percentage of correct site assignments was higher for RFCM (36-43%) than for GA (5%), for both timbers (Figure 4.3). Yet, when trees were not assigned to their site of origin, the distances to the true origin were shorter when using GA (90% within 300 km) than for RFCM (90% within 500; Figure 4.3). When zooming in on correct assignments within 100 km, the differences between methods depend on the timber. For Azobé, correct assignment within 100 km was more likely using RFCM (50%) than GA (39%) (Figure 4.3a and 4.3b). For Tali, however, GA performed better with 83% of the trees assigned correctly within 100 km, versus 70% with RFCM (Figure 4.3c and 4.3d). Except for samples from one site (CAM39), the vast majority of the samples were assigned to

neighboring sites known to have shared or genetically similar haplotypes (Chapter 2).

4.3.2 Verification of timber claims for blind samples

4.3.2.1 Step1: Timber taxonomic claim verification

Samples that were correctly identified at the genus level were considered as having their timber claim confirmed (*Erythrophleum* = Tali and *Lophira* = Azobé), because there are no other traded timbers in these genera. For Azobé, we also considered the sample as correctly identified when the family level was confirmed because the other genus within the Ochnaceae, Elvasia DC [Schneider et al., 2021], included in the SPP BLAST dataset (Tab. 4.1), does not contain timber species either. For Tali, among the *Erythrophleum* species included in the SPP BLAST dataset (Table 4.1), only *E. africanum* is also a timber species (the Ordeal tree (En) or Mucaráti (Po)), but it does not occur in the Congo basin .

The BLAST analysis of the cambium samples yielded 90% correct genus level identification for both the Azobé and Tali samples (Figure 4.2). For the remaining blind samples claimed as Azobé, two were correctly identified at the family level, and one was correctly excluded (as not Azobé). For the remaining Tali blind samples, one was correctly excluded (as not Tali) while two samples failed as no information was retrieved for the three barcode regions.

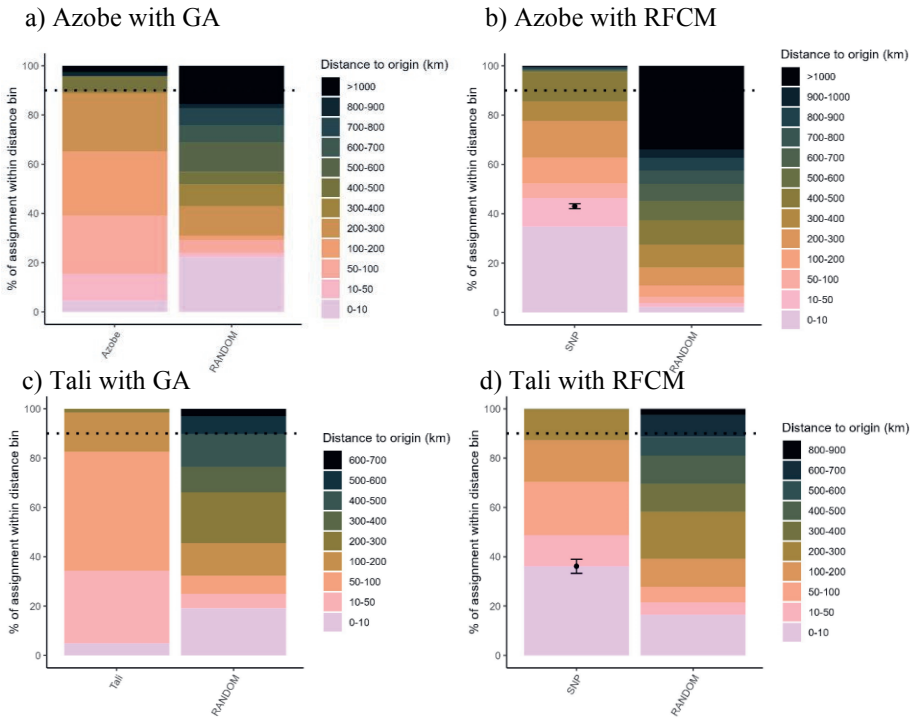


Figure 4.3 : Distance to predicted origin (km), as a percentage of the total number of tested trees per distance category (bin) per method, "RANDOM" denotes the distribution of pairwise distances in the reference dataset, symbolizing random assignments, while the dotted line signifies 90% accuracy. Self-assignment results were obtained with GeoAssign (GA) for a) Azobé and for c) Tali, and with Random Forest Classification Model (RFCM) for b) Azobé and for d) Tali. For the RFCM models we also provide the average correct assignment accuracy, depicted by a dot with an error bar indicating the variation (st. dev.) in assignment accuracy.

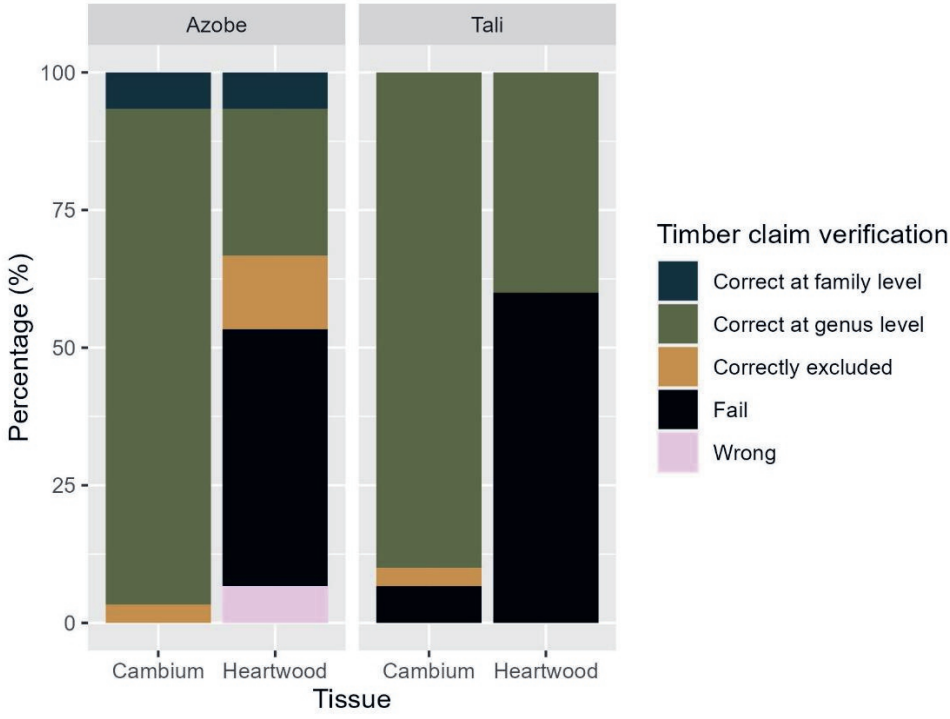


Figure 4.4 : Timber claim verification through Taxonomic Assignment. The bar graphs indicate the success rate of correctly identifying blind samples to their claimed timber identities (Azobé or Tali), at both genus and family taxonomic levels. It categorizes outcomes into three distinct groups: 'correct' (accurate taxonomic assignment), 'fail' (insufficient data to determine barcode region), and 'wrong' (incorrect identification), for two tissue types (cambium and heartwood).

While timber taxonomic claim verification for cambium samples of both Azobé and Tali was successful, heartwood samples posed serious challenges, with a 53% failure rate attributed to the lower retrieval rate of DNA barcode regions (Figure 4.2). However, among the heartwood samples from which sufficient data were obtained, a significant proportion (93%) was accurately identified at either the genus or family level, or correctly excluded, with only a single sample misidentified. For the 15 heartwood samples per timber, we were able to

retrieve the barcode regions for 53% Azobé and 40% Tali samples (Figure 4.2). Of the Azobé heartwood samples, 40% were identified at the genus level, one was identified at the family level, and one Azobé sample was wrongly identified as Tali (*Erythrophleum* spp). The six Tali heartwood samples were correctly identified at the genus level. Among the blind samples claimed as Tali that failed, two did not belong to the *Erythrophleum* genus, but were *Entandrophragma utile* and *Lophira alata*. These failed samples had a very low DNA coverage (~0.69x).

4.3.2.2 Step 2: Geographic assignment

Across species and assignment methods, cambium samples yielded higher data quality and quantity compared to heartwood samples (Figure 4.3 and Figure 4.4). All Azobé cambium blind samples were geographically assigned, compared to only 30% of the heartwood samples (Figure 4.3a). For Tali, one cambium sample failed (4%) while this was 62% for heartwood samples (Figure 4.3b).

The geographic assignment of Azobé was successful within 100 km of the true origin for 72% of cambium and 31% of heartwood samples when using GA (all samples that did not fail; Figure 4.3a and Figure 4.4a). When employing RFCM, this was considerably lower: only 28% of cambium samples were assigned within 100 km of true origin, and the remaining cambium samples were assigned to up 820 km from the true origin (Figure 4.3a and 4.4b). For heartwood blind samples, RFCM also performed poorly with 31% of the heartwood samples (all four samples that did not fail) assigned at least 300 km from the true origin (Figure 4.3a).

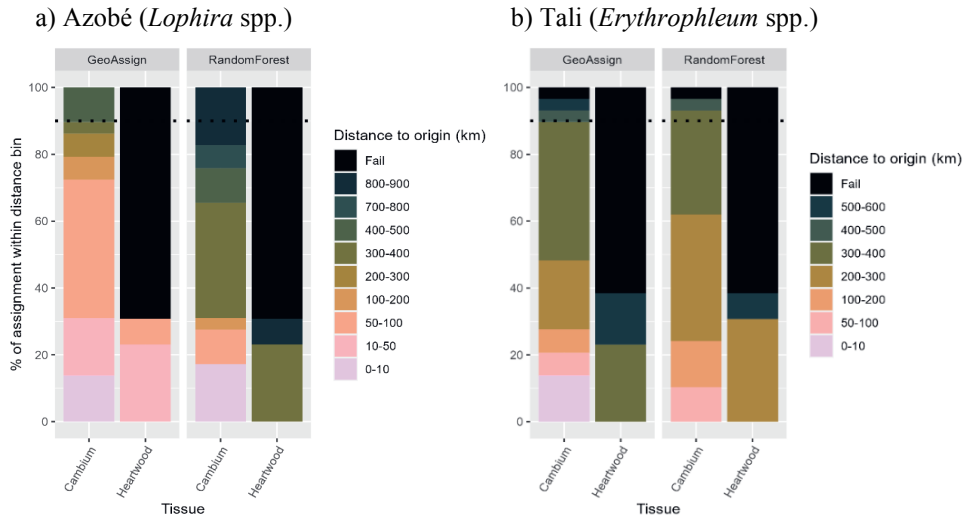


Figure 4.5 : Geographic Assignment Accuracy of Azobé and Tali Blind Samples. This figure presents the effectiveness of geographic assignment for blind samples of Azobé and Tali, using both GA and RFCM. The bar graphs show the percentage of blind sample trees within specified distance categories (km) to their predicted origin, segmented by method. The analysis is split into two panels: a) for Azobé and b) for Tali samples. A dotted line marks the threshold for 90% of the samples, serving as a benchmark for the precision of each method in predicting the correct geographic origin.

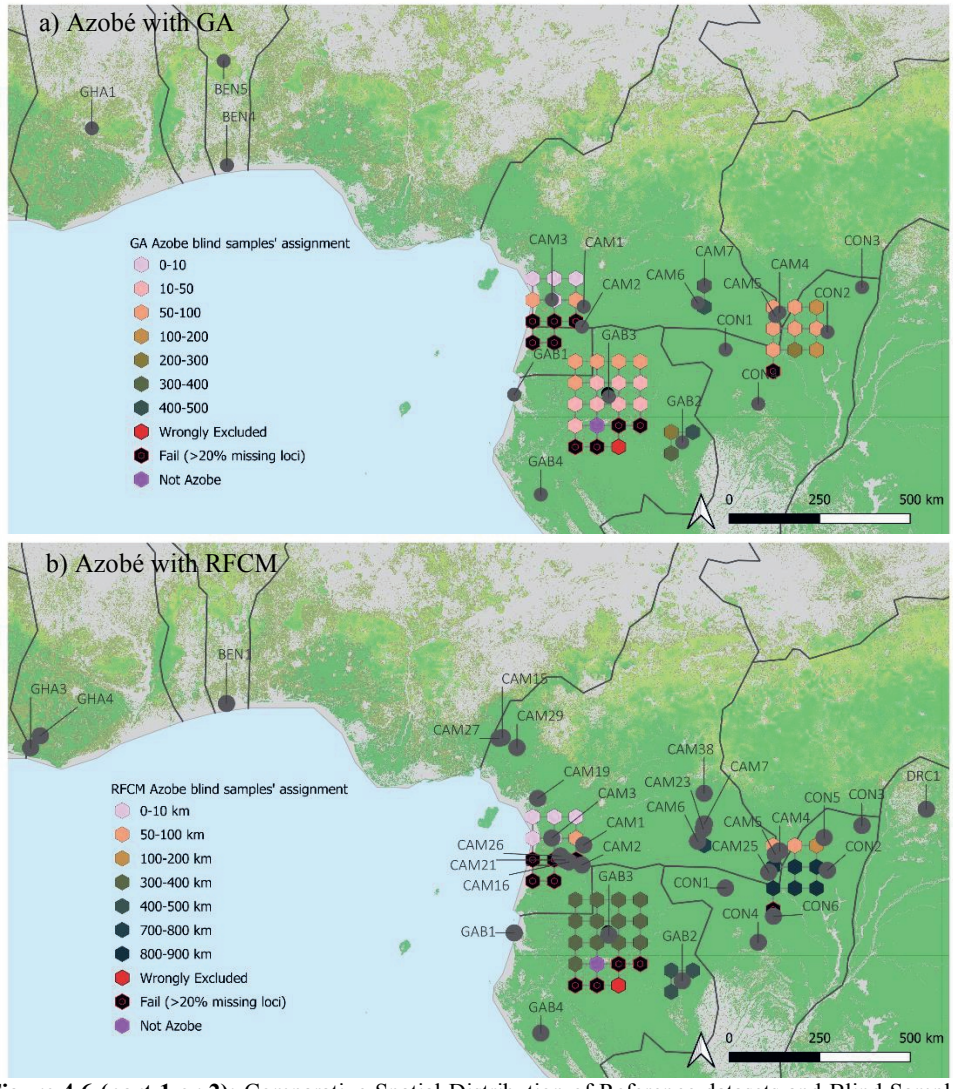


Figure 4.6 (part 1 or 2): Comparative Spatial Distribution of Reference datasets and Blind Samples Assignment Results Across Methods and Species. This figure maps the geographic locations of reference and blind samples for Azobé and Tali, analysed with GA and RFCM. a) Azobé GeoAssign, b) Azobé RFCM, c) Tali GeoAssign, and d) Tali RFCM. Blind samples are color-coded based on distance categories, if they were correctly excluded (purple for Azobé and orange for Tali), wrongly excluded (red), or if they failed (black).

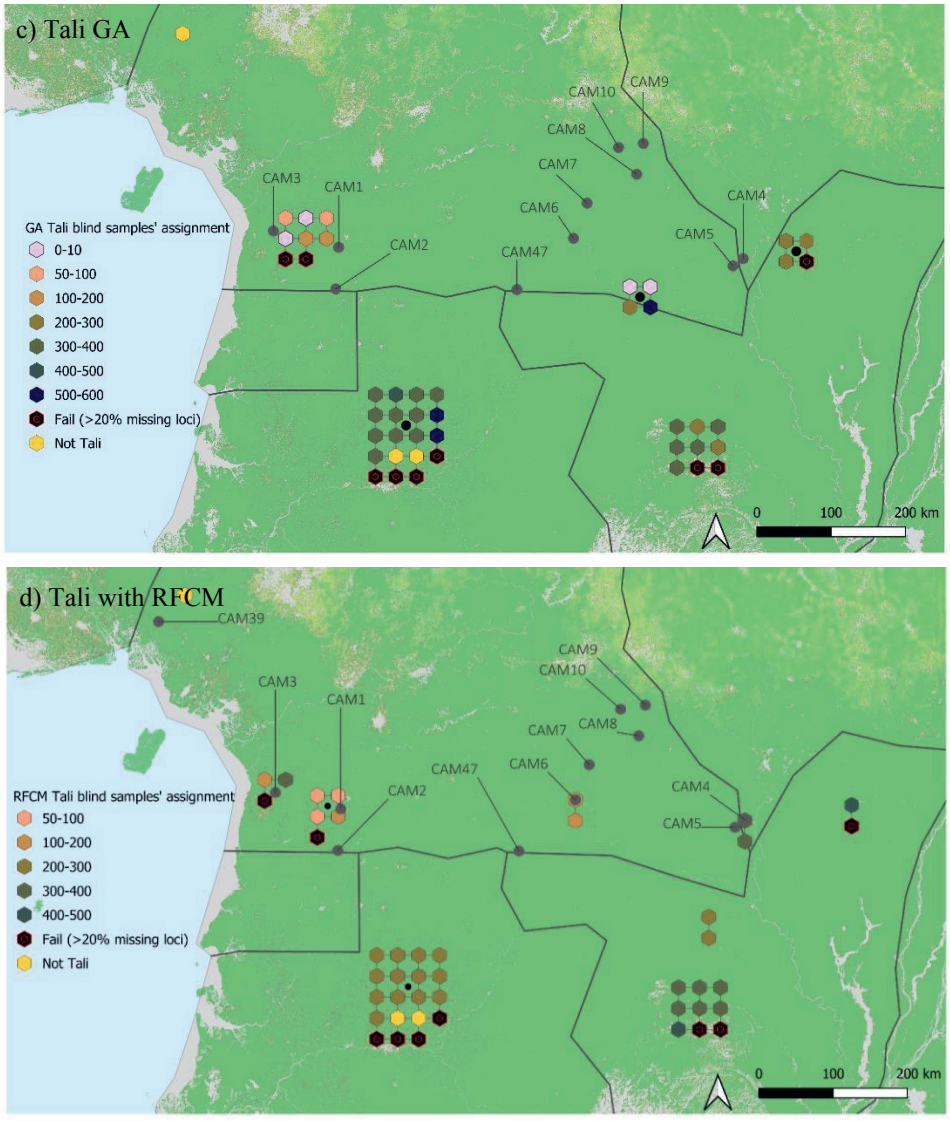


Figure 4.7 (part 2 or 2):: Comparative Spatial Distribution of Reference datasets and Blind Samples Assignment Results Across Methods and Species. This figure maps the geographic locations of reference and blind samples for Azobé and Tali, analysed with GA and RFCM. a) Azobé GeoAssign, b) Azobé RFCM, c) Tali GeoAssign, and d) Tali RFCM. Blind samples are color-coded based on distance categories, if they were correctly excluded (purple for Azobé and orange for Tali), wrongly excluded (red), or if they failed (black).

For Tali, the geographic assignment at the site level of cambium blind samples was overall lower (Figure 4.3b, 4.4c and d). Only 10 and 21% of the cambium blind samples, RFCM and GA respectively, were assigned within 100 km of the true origin (Figure 4.3b, Figure 4.4c and d). Both methods performed similarly at medium to large scales, with GA assigning 90% of the cambium samples within 400 km of the true origin RFCM 93% (Figure 4.3b).

Not unexpectedly, there was a strong effect of the coverage of sites in the reference datasets. Across species, all 28% of the cambium samples that were assigned within 118 km came from the region that was covered by the reference datasets: Cameroon. Of the remaining 22 (72%) cambium samples, 19 were assigned between 200-455 km (Figure 4.3b, Figure 4.4c and d). These samples were from Gabon and Congo, both countries that were not included in the reference dataset (Figure 4.3b, Figure 4.4c and d). For Tali, 71% (30 out of 42) of the blind samples were obtained from outside the area covered in the reference datasets (Figure 4.4c and d). When only Cameroonian blind samples were considered (all cambium), accuracy improved to 60% of samples assigned between within 100 km, and 100% within 230 km using GA (Supporting Figure S4.1, Figure 4.4c and d). Random Forest performed poorly compared to GA, with only 30% of cambium samples assigned within 100 km. The remaining 70% varied between 102-338 km from the true origin (Supporting Figure S4.1, Figure 4.4c and d). For the heartwood samples that could be assigned, performance was much lower, and there was no difference between the assignment methods (Figure 4.4). All Cameroonian heartwood samples failed, not reaching 80% of complete loci. GA assigned the other samples to populations between 300-600 km from the true origin, for RFCM this was between 200-600 km (Figure. 4.3b).

4.4 DISCUSSION

The present study tested 1) the robustness for the reference dataset by self-assigning of the samples in the reference datasets , 2) testing the species claim using DNA barcoding for three known barcode regions and 3) the accuracy of assignment using blind samples, performed for cambium and heartwood samples with two statistical methods.

4.4.1 Species claim check

To address the question, "Is it truly Azobé/Tali?" we employed DNA barcoding to verify the timber identity of blind samples. Samples identified at the genus level were considered to have their timber claims confirmed (*Erythrophleum* for Tali and *Lophira* for Azobé), given the absence of other traded timbers within these genera. We could verify the claimed taxonomic identity of the blind samples with a high accuracy, also when the claimed timber was incorrect. Therefore, we conclude that the sequence data generated through the strategy employed here effectively allows verification of taxonomic claims. Correctly excluding blind samples from other species was identified as a barrier for the application of multi-element analysis for timber tracing of Tali and Azobé (Boeschoten et al., 2023). Our results show that using DNA barcoding, these barriers can be overcome, after which multi-element analyses could be performed to verify geographic origin.

The ability to verify species claims using DNA barcoding can also be used to evaluate compliance with CITES. For instance, in a study on CITES listed species *Gonystylus bancanus*, genetic taxonomic species identification based on two cpDNA (trnH-psbA and trnL) barcode regions was successful in 90% of the cases (Ng et al., 2016). Our analysis was less successful in separating closely related species. For instance: Azobé (*L. alata*) could not be separated from *L. lanceolata*, likely due to cross-species chloroplast capture. Chloroplast capture occurs in both Azobé and Tali (chapters 2 and 3) and this may hamper species identification solely based on plastomes (Fahey et al., 2021). The identification at the genus level (verifying the timber claim) for the two timbers studied here suffices, as most co-generic species either do not occur or are not traded as timber under the same common names. Yet, for other traded tropical timber species complexes, when taxonomic identification to the species level is needed, only plastid SNP markers might not suffice. In that case, a combination of nuclear and chloroplast SNPs may increase the success rate, as shown for instance for *Dipteryx* species (Honorio Coronado et al., 2020).

4.4.2 DNA quality

Wood DNA undergoes degradation and exists in limited quantities due to various factors including age, chemical composition, or degradation related to industrial processing (Rachmayanti et al., 2006; Degen and Fladung, 2007; Rachmayanti et al., 2009b; Lowe and Cross, 2011; Hung et al., 2017; He et al., 2019). In our blind samples we indeed observed marked differences in DNA quality. Cambium blind samples yielded higher quality DNA compared to heartwood samples. Almost all cambium blind samples (98%) yielded sufficient data (>80% complete SNP information across the plastid genome) to be geographically assigned, while only 34% of the heartwood samples reached that level. The success rate for heartwood samples was slightly higher (47%) when we focused only on the three barcode regions in the plastid genome. Using a similar Qiagen kit-based protocol, Tnah et al. (2012) obtained a comparable difference, achieving twice the amplification success for cambium and sapwood than that for heartwood tissue samples. Poor DNA quality and quantity from heartwood represents a major barrier in the practical application of genetic analyses in timber tracing. Recent efforts have tried to overcome this barrier. For instance, But et al. (2023) compared three DNA isolation methods especially developed for wood and wooden products. Their findings indicated that SDS-based and CCDB methods outperformed the Qiagen kit-based method, suggesting that the main protocol used here was not the most suitable one. This demonstrates that there is ample scope to refine and optimize current methodologies in order to obtain DNA of sufficient quantity and quality for timber tracing. Considering that methods for single-cell sequencing of plant material have been developed and are being applied to gene expression (Cuperus, 2022), it is likely that issues with DNA integrity (fragmentation in the wood tissue and extra fragmentation during the extraction) may be more difficult to overcome than the low amount of DNA that can be extracted.

4.4.3 Blind sample assignment and the spatial coverage of reference data

For Tali, over two thirds of the blind samples were obtained from Congo and Gabon, while the reference samples were mostly from Cameroon. This spatial mismatch resulted in long distances between true and assigned origins. In case of a poor genetic match between blind

and reference samples, both assignment methods (GA and RFCM) will assign the blind sample to the most likely reference site. In this case, GA assigns a blind sample to the site with the highest share of genetically resembling individuals (Figure 4.3, Figure 4.4 a and c). For samples outside the area covered by the Tali reference dataset, the RFCM assignment performed similarly to GeoAssign, with 90% of the cambium samples assigned within 400 km (Figure 4.2). The data replacement with the most common allele may have reduced the distances between assigned and true origin for samples for samples from outside the reference area (from Congo and Gabon), as the imputation would make any unknown haplotype genetically closer to the haplotypes present in the Tali reference dataset. Therefore, some trees may have been assigned to similar haplotypes even though they were not identical. Nonetheless, the lack of reference populations of Tali in Congo and Gabon increased the mean distance of predicted to correct sites. Indeed, when limiting the blind sample set for Tali to those that were from regions covered by the reference dataset (Supporting Figure S4.1), the success rate improved: 60% of the Cameroonian samples were assigned within 100 km of the true origin. Not many studies have dealt with assignment of blind samples from outside the region covered by reference samples (Boeschoten et al., 2023). In general, and evident for Tali, the low representation of certain areas in reference databases for timber tracing poses a problem in their use for origin verification and assignment (Low et al., 2022). This highlights the importance of a comprehensive genetic reference dataset that encompasses all areas prone to (illegal) logging as well as protected regions.

4.4.4 Comparing success rate of blind samples in timber tracing

When blind samples were assigned from areas covered by the reference database, 40% of the Tali cambium samples were assigned to their correct site of origin (<10 km). The fine scale tracing (<100 km) assignment tests were also quite accurate, with 60-72% of samples for both Tali and Azobé being assigned within 100 kilometers of their true origin. Assignment tests conducted in most genetic studies are self-assignment, often based on leave-one-out-cross-validation methods (e.g. Ng et al., 2016; Degen et al., 2017; Honorio Coronado et al., 2020; Blanc-Jolivet et al., 2021), and usually do not extend the tests to blind samples from outside of the sampling area (Low et al., 2022). Nevertheless, we can put our blind sample results

into perspective by comparing them to self-assignment results as reported in several earlier studies. At larger scales, comparable results were obtained for Azobé by Blanc-Jolivet et al. (2021). Using 75 nuclear SNPs, 20 cpSNPs and 28 mtSNPs Blanc-Jolivet et al. (2021) obtained a theoretical correct assignment between West and Central Africa of 86% (self-assignment test with leave-one-Out procedure using the Bayesian criteria of Rannala and Mountain 1997). The self-assignment tests conducted for Cumaru (*Dipteryx* spp.) using nuclear, chloroplast and mitochondrial SNPs assigned using GeoAssign achieved theoretical success rates of 91–100% to the correct genetic cluster and 69–92% to the country of origin (Honorio Coronado et al., 2020). We achieved accuracies of assignment at similar spatial scales for blind samples obtained from inside the area covered by the reference datasets which are comparable to those obtained by Honorio Coronado et al. (2020) using a combination of nuclear, mitochondrial, and chloroplast SNPs. All our assignments for Tali and Azobé were accurately placed within 400 and 500 km, respectively. This underscores the effectiveness of employing a genome-wide SNP set, also for the plastome of the studied timber groups, which is noted for its lower genetic diversity compared to nuclear genomes (Mader et al., 2018a). In their study on *Gonystylus bancanus*, Ng et al. (2016) employed sixteen nuclear STR markers and they were able to successfully differentiate among populations, genetic clusters, and regions. Self-assignments at the site level (55% success rate) yielded a level comparable to our results of ~40% success (<10 km) for Tali samples from the area covered by the reference database. Similarly, at finer spatial scale (<100 km) than the genetic cluster detected (hundreds of kilometres), Ng et al. (2016) obtained 100% success for *G. bancanus* while we obtained 72% success for Azobé cambium and all four heartwood blind samples assigned within 100 km (intra-regional level). The totality of our cambium blind samples were assigned within 500 and 300 km for Azobé and Tali respectively, which is somewhat shorter than results from Ng et al. (2016) who self-assigned 100% of the samples to genetic clusters within 600-850 km.

A blind sample geographic assignment study using multi-element composition was also conducted for Tali (*Erythrophleum* spp.) and Azobé (*Lophira* spp.) by Boeschoten et al. (2023). Our blind sample results partially align with those of Boeschoten et al. (2023), who also obtained 70-72% correct origin identification across species. The shared chemical profile across species, found by Boeschoten et al. (2023), can enable the use of chemical reference

data across species, overcoming the extensive sampling burden that is associated with genetic profiles (see below).

Overall, our blind sample results are in line with self-assignment results obtained in other genetic tracing studies. Yet, the level of successful assignment of heartwood is still low compared to cambium. Improvement of these success rates requires better DNA extraction protocols, more extensive reference databases and more effective statistical tests.

4.4.5 The impact of different assignment methods on success rates

The two assignment methods applied here varied in their efficacy to determine the origin of timber samples. GA outperformed RFCM for blind samples from areas covered by the reference datasets, yielding 60-72% success rate within 100 km across species, compared to a 10-28% when using RFCMs.

As RFCMs cannot handle missing data, imputation is required, particularly for timber samples with many missing data. Imputing data seems to have negatively impacted assignment success, possibly contributing to the lower success of RFCM compared to GA. Both methods performed equally poor for samples coming from outside the region covered by reference samples. However, GA provides probabilities associated with the site assigned and other possible sites, which may facilitate sample exclusion when previous knowledge of the spatial genetic structure is present for the marker used. On the other hand, more samples of individual trees were kept in the RFCM reference dataset, e.g. herbarium samples from West Africa included in the study, as GA requires a minimum of 5 individuals per site.

GA had higher fine spatial accuracy (<100km), and whenever samples were miss-assigned, the overall distance to the mis-predicted site was smaller than obtained with RFCM, also for samples from outside the sampling range. GA also identified other likely sites of origin when a haplotype is present in multiple locations and cannot be differentiated. The method will assign the samples to the site with the highest share of similar individuals. If a particular haplotype is also present in other sites they cannot be excluded as probable origin. This adds a layer of information on other possible sites of origin.

Spatial accuracy of genetic tracing techniques based on plastid genomes is intricately linked

to the spatial distribution of haplotypes within species. This relationship highlights potential limitations in tracing precision due to the natural genetic variation and geographic spread of haplotypes. Understanding this distribution is crucial for enhancing tracing efforts, indicating the need for detailed haplotype mapping to improve origin determination accuracy. An unexplored area of potential to improve haplotype analyses is the combination of genetic data with geographic information related to possible genetic barriers. This may assist in predicting the probabilities of haplotypes and thus help to refine tracing methodologies and achieve more accurate identification of timber origins in regions with low haplotype diversity or low sampling coverage.

The GA method, a nearest neighbor approach (Degen et al., 2017), has demonstrated superior performance over the Bayesian assignment approach reported by Honorio Coronado et al. (2020), a finding that aligns with the higher performance of GA over RFCM for our blind samples. Furthermore, GA offers insights into other potential sites of origin, allowing for the consideration of a broader area, especially for haplotypes with extensive distribution. This capability makes GeoAssign the preferred method in our study for samples of unknown origins. However, further comparative analysis of assignment statistical methods across different taxa is necessary to evaluate the consistency of this finding and get to generic recommendations on the most appropriate assignment method.

4.5 CONCLUSION

Timber taxonomic claim verification at the genus level were successful using cpDNA, but taxonomic identification at the species level was not possible. We also found that timber origin verification using cpSNP haplotypes relies on the spatial distribution and commonness of haplotypes. Blind samples with commonly found haplotypes were assigned to the most likely site, but sites where these haplotypes were also present could not be excluded. Although at the site level (<10 km) self-assignment success was higher for RFCM than GeoAssign, the reverse was true for blind sample assignment. DNA quality from heartwood samples was low, and continues to form a major technical barrier to implement genetic timber tracing. Next steps to implement cpDNA tracing are to achieve a sampling coverage that encompass the

(illegal) logging hotspots as wells as protected areas, the development of adapted statistical tools, and the forensic developmental validation of the cpSNPs reference and timber species claim verification sets for Tali and Azobé following the Scientific Working Group on DNA Analysis Methods (SWGDAM, 2012).

4.6 ACKNOWLEDGEMENTS

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 from WorldForestID. We thank all the collaborating timber companies and their field teams for facilitating the fieldwork and all involved colleagues at our partner institutes University of Dschang, Marien Ngouabi University, IRET/CENAREST and the National Herbarium of Gabon for help in arranging fieldwork.

4.7 PERMITS AND AUTHORIZATIONS

Cameroon: Research Permit No. 00000116/MINRESI/B00/C00/C10/C12 (Yaounde, 09 Sep 2019); Research Permit No. 000066/MINRESI/B00/C00/C10/C12 (Yaounde, 07 Jun 2021); Scientific research permit No. 2144 PRBS/MINFOF/SETAT/SG/DFAP/SDVEF/SC/NGY (Yaounde, 23 Jul 2021); ABS Permit 00010/MINEPDED/CNA/NP-ABS/ABS-FP (Yaounde, 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

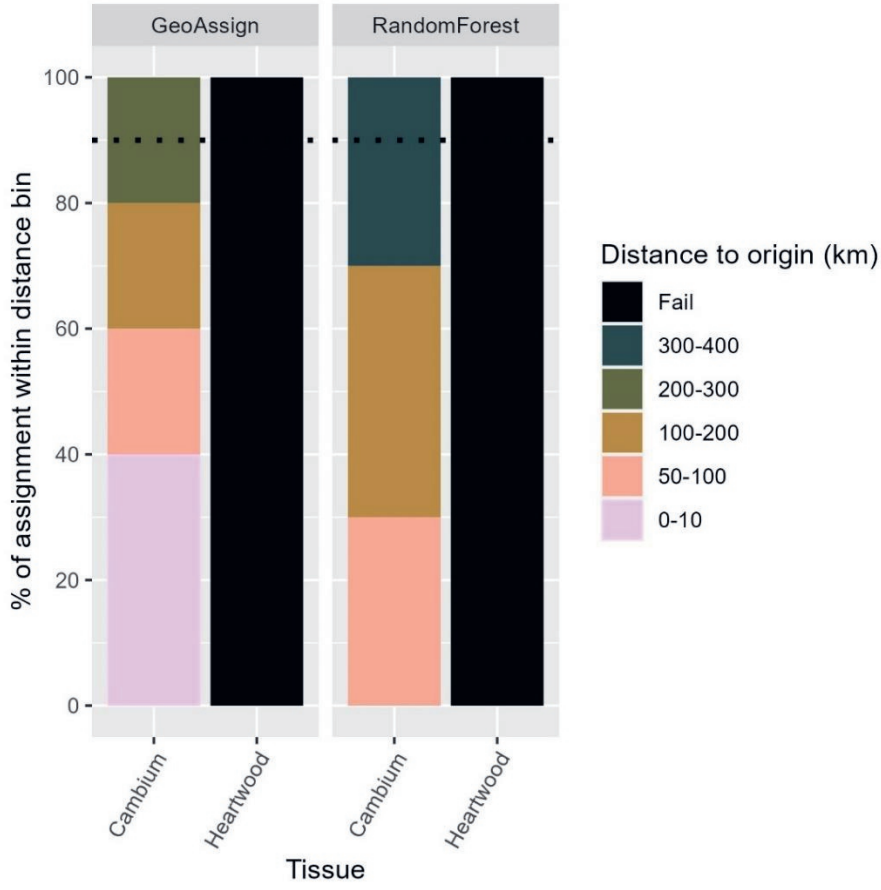
Supporting information

Supporting table S4.1. Studied timber related species, DNA barcode regions, and their NCBI accession numbers.

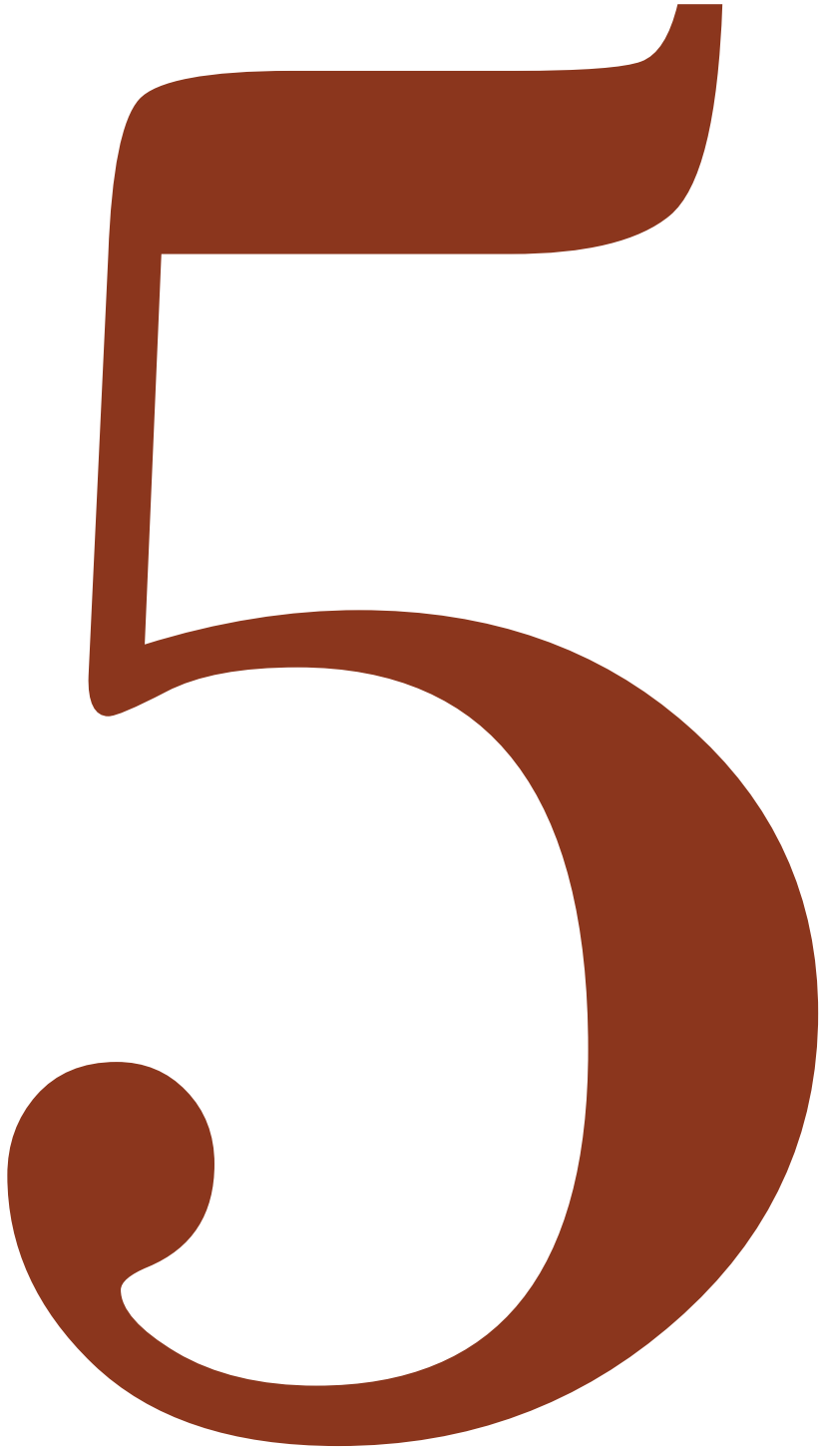
Related species	*DNA Barcode region*	*Accession numbers*
Erythrophleum chlorostachys	rbcL	KF496399.1
Erythrophleum chlorostachys	matK	JX840178.1
Erythrophleum ivorense	rbcL	MN591448.1
Erythrophleum ivorense	matK	MN579899.1
Erythrophleum ivorense	trnL	AF365102.1
Erythrophleum ivorense	matK	JX840188.1
Erythrophleum ivorense	matK	JX840185.1
Erythrophleum ivorense	matK	JX840185.1
Erythrophleum ivorense	matK	JX840182.1
Erythrophleum ivorense	rbcL	KC628491.1
Erythrophleum ivorense	rbcL	KC628237.1
Erythrophleum ivorense	rbcL	KC628051.1
Erythrophleum ivorense	matK	KC627808.1
Erythrophleum ivorense	matK	KC627480.1
Erythrophleum ivorense	rbcL	U74205.1
Erythrophleum ivorense	matK	GU321974.1
Erythrophleum suaveolens	matK	MK898790.1
Erythrophleum suaveolens	rbcL	MK895580.1
Erythrophleum suaveolens	matK	KX162080.1
Erythrophleum suaveolens	matK	EU361949.1
Erythrophleum suaveolens	trnL	AF365103.1
Erythrophleum suaveolens	matK	JX840187.1
Erythrophleum suaveolens	matK	JX840186.1
Erythrophleum suaveolens	matK	JX840184.1
Erythrophleum suaveolens	matK	JX840183.1

Erythrophleum suaveolens	matK	JX840181.1
Erythrophleum suaveolens	rbcL	JX572569.1
Erythrophleum suaveolens	matK	JX517934.1
Erythrophleum fordii	rbcL	MT933820.1
Erythrophleum fordii	rbcL	MT933819.1
Erythrophleum fordii	matK	MW044137.1
Erythrophleum fordii	matK	MW044136.1
Erythrophleum fordii	matK	HQ415268.1
Erythrophleum fordii	RuBisCO	HQ415085.1
Erythrophleum fordii	rbcL	KP094601.1
Erythrophleum fordii	rbcL	KP094600.1
Erythrophleum fordii	matK	JX840179.1
Erythrophleum africanum	matK	JX840180.1
Erythrophleum africanum	rbcL	JX572568.1
Erythrophleum africanum	matK	JX517525.1
Erythrophleum teysmannii	rbcL	AB925785.1
Erythrophleum teysmannii	matK	AB925154.1
Erythrophleum teysmannii	matK	LC080894.1
Lophira lanceolata	rbcL	MT385763.1
Lophira lanceolata	matK	FJ670029.1
Lophira lanceolata	rbcL	FJ670172.1
Lophira alata	trnL	KF263437.1
Lophira alata	rbcL	KF263367.1
Lophira alata	matK	KF263272.1
Elvasia calophyllea	matK	FJ670028.1
Elvasia calophyllea	trnL	KF263415.1
Elvasia calophyllea	matK	KF263228.1
Elvasia calophyllea	rbcL	FJ670171.1
Elvasia calophyllea	trnL	AY763233.1
Elvasia elvasioides	trnL	KF263424.1

Elvasia elvasioides	rbcL	KF263350.1
Elvasia capixaba	trnL	KF263423.1
Elvasia essequibensis	trnL	KP196903.1
Elvasia essequibensis	rbcL	KP196864.1
Elvasia macrostipularis	trnL	KP196904.1
Elvasia macrostipularis	rbcL	KP196865.1



Supporting Figure S4.1. Distance to predicted origin (km) as % of the total number of test trees per distance category (bin) per method obtained for Tali, only with Cameroonian blind samples (covered by reference datasets) with GeoAssign and Random Forest methods. The dotted line indicates 90% accuracy.



Chapter 5

Combining genetic and chemical methods boosts accuracy of timber tracing at small spatial scales

Combining genetic and chemical methods boosts accuracy of timber tracing at small spatial scales

Laura E. Boeschoten*, Bárbara Rocha Venâncio Meyer-Sand*, Arnoud Boom, Gäel U.D. Bouka, Jannici C.U. Ciliane-Madikou, Nestor Laurier Engone Obiang, Mesly Guieshon Engongoro, G. Arjen de Groot, Jöel J. Loumeto, Dieu-merci M.F. Mbika, Cynel G. Moundounga, Rita M.D. Ndangani, Dyana Ndiade Bourobou, Ute Sass-Klaassen, Marinus J. M. Smulders, Steve N. Tassiamba, Martin T. Tchamba, Bijoux B.L. Toumba-Paka, Mart Vlam, Herman T. Zanguim, Pascaline T. Zemtsa, and Pieter A. Zuidema

* L.E. Boeschoten and B. Rocha Venâncio Meyer-Sand contributed equally to this paper.

Abstract

Improving transparency in the timber supply chain is crucial to halt illegal timber trade and promote sustainable timber trade. To support law enforcement, independent verification of timber origin is needed, which can be based on genetic and chemical wood properties. For forensic cases, these methods need to achieve a high tracing accuracy (above c. 90%) at small scales (below c. 100 km). So far, no single method has yielded both these goals for large geographic areas. Combining tracing methods may improve verification accuracy, but this has never been tested. In this study, we collected wood samples of Azobé (*Lophira alata*), an economically important timber, from 13 georeferenced sites in three Central African countries. For a varying number of trees per site, we (1) detected genome-wide chloroplast Single Nucleotide Polymorphisms (cpSNPs), (2) measured three stable isotope ratios ($\delta^{18}O$, $\delta^{2}H$ and $\delta^{34}S$) and (3) analyzed the multi-elemental composition. Based on the three methods, we developed Random Forest models for site classification and tested their individual as well as combined assignment accuracy. The accuracy with which samples were assigned to their site of origin varied between 43% and 73% when applying single methods. Combining

cpSNPs and multi-element analysis boosted assignment success to 83% at site level and yielded correct assignment of 90% within 100 km. Combining all three methods further increased accuracy to 94% within 100 km, and even yielded a 90% accuracy at 50 km distance. Methods complemented each other, such that a low accuracy of one was compensated by a high accuracy of another. Our study shows that by combining tracing methods unprecedented accuracy levels in assigning timber origin can be attained at small spatial scales across three countries. It further demonstrates the potential of combining methods to boost tracing accuracy and resolution for timber, and likely can also be applied to illegal wildlife and agricultural commodities.

Keywords: Origin verification; machine learning; illegal logging; cpDNA; stable isotope ratio; wood chemistry.

5.1 INTRODUCTION

Illegal timber trade is a global problem that contributes to deforestation, biodiversity loss and climate change (Hoare et al., 2020; Kleinschmit et al., 2016). To address this, various regulations and certification schemes have been implemented in order to reduce illegal trade Lowe et al. (2016); May (2017). However, verifying the claimed origin of wood products remains a challenge and fraudulent practices are still widespread, especially in tropical countries (Hoare & Uehara, 2022). Forensic methods that can verify timber origin are therefore crucial to ensure the legality and sustainability of timber products.

Multiple forensic methods are currently under development to enable independent verification of the origin of wood. These methods use intrinsic wood properties to verify the origin, such as genetic or chemical variations within species (Dormontt et al., 2015; Low et al., 2022). Three promising tracing methods for tropical timber include genetic differentiation, stable isotope ratios and multi-element analysis. Genetic methods remain the most applied method for origin testing of timber to date (Low et al., 2022). The population genetic approach is based on the principle that individuals that are geographically closer are also genetically more closely related. This relationship arises from historical demographic processes, which have led to spatial variation and consequently to the link between genetic

differentiation and geographical distance. Second, stable isotope ratios also have a long history of origin identification research in wood as well as in other commodities (Ehleringer et al., 2000; Gori et al., 2018). For timber, the bio-elements $\delta^{18}O$, $\delta^{2}H$, $\delta^{13}C$, $\delta^{15}N$ and $\delta^{34}S$ are most used for tracing (Ch 4+5, this thesis, Watkinson et al. (2020, 2022a); Gori et al. (2018)). The drivers of geographical variation in these isotope ratios in wood are mostly related to climate, geology and/or deposition (West et al., 2010; Cernusak et al., 2022; Wynn et al., 2014; Lehmann et al., 2022; Allen et al., 2022; van der Sleen et al., 2017). The third method is based on multi-element analysis (Boeschoten et al., 2022, 2023). A large number of elements (such as Mg, Ca, La) is measured simultaneously using mass spectrometry and based on this elemental composition, an origin-specific chemical fingerprint is defined. Geographical variation in elemental composition is caused by variation in physical and chemical soil properties (Boeschoten et al., 2022).

While each of these methods holds great potential for identifying wood origin (Vlam et al., 2018b; Ng et al., 2017; Watkinson et al., 2020; Boeschoten et al., 2023; Dormontt et al., 2020), their effectiveness in tracing origins at short distances remains challenging. The accuracy of tracing wood origins at small spatial scales (<100 km) has not consistently reached a reliable accuracy to be used as a tracing method at those distances. At short distances, high tracing accuracies based on genetic or isotopic methods have been limited to studies with a low number of sites, or to some specific species and/or some regions (Vlam et al., 2018b; Watkinson et al., 2022a), whereas in other studies only large geographical regions could be distinguished (Ng et al. (2017); Watkinson et al. (2020), Boeschoten (Chapter 4, 2023). This was also the case for multi- elemental analysis: some sites that were 50 km apart could be distinguished well, but a few sites that were >200 km apart showed a non-discriminatory chemical composition (Boeschoten et al., 2023).

As different mechanisms underly the variation used to distinguish between different origins, each method has its own strengths and weaknesses. As a result, regions where one of the methods performs poorly may coincide with those where one of the others yields high tracing accuracy. Therefore, tracing accuracy is expected to improve when combining multiple methods (Dormontt et al., 2015; Low et al., 2022), but this has never been tested for tropical timber. In archaeological wood research, combining multiple methods (anatomical proxies as

well as strontium isotope analysis) indeed yielded higher tracing precision at small spatial scales (Akhmetzyanov et al., 2019; D'Andrea et al., 2023). In other commodities, the combination of stable isotopes and multi-element analysis also yielded good tracing results (Wang et al., 2020; Drivelos & Georgiou, 2012). Furthermore, the combination of stable isotopes with genetic analysis has been applied to assign migratory birds to their breeding grounds and the combination also greatly reduced the area to which birds were assigned (Chabot et al., 2012).

The objective of this study was to test the potential for origin verification of timber of three established forensic methods (genetics, stable isotope ratios and multi-element analysis) individually as well as combined. We did so by answering the following questions: 1) what is the natural variation underlying these forensic methods and is there a clear spatial structure as a basis for tracing purposes? 2) what is the assignment accuracy at a site level of individual as well as combined forensic methods? and 3) what method or combination of methods reaches the highest accuracy at small spatial scales (<100km)? To answer these questions we focused on an economically important tropical timber species from Central Africa, traded as Azobé (*Lophira alata*). We applied all three methods to the same trees at 13 locations across the main timber exporting countries in the region (Cameroon, Gabon, and Republic of the Congo). To specifically test small-scale origin assignment, we tested tracing performance at the site level, as some sites were as close as 10 km. With this research we aimed to improve the understanding of how these tracing methods performed individually when applied to the same trees, and we aimed to test their potential for combined use.

5.2 METHODS

5.2.1 Sample collection

The study was conducted on the commercial Central African timber species Azobé (*Lophira alata* Banks ex C.F. Gaertn, Ochnaceae), Azobé occurs from Guinea to the Democratic Republic of Congo (GBIF.org, 2021), but almost all timber export is originating from Cameroon, Gabon, Republic of the Congo and Democratic Republic of Congo. Sampling was

conducted in 13 logging concessions in Cameroon, Gabon and Republic of the Congo between September 2019 and April 2022 (see Figure 5.1). All sites were natural forest concessions. Sampling was conducted in collaboration with the operating forestry companies.

At each site, we sampled heartwood as well as bark of 20-30 trees per timber species. Target trees within one site were located between 100 m and 5 km apart. Sampled trees were either standing or recently felled and were of at least 30 cm diameter at breast height (DBH). We sampled trees of different sizes at all sites, to allow conclusions independent of ontogenetic stage. The size variation of the sampled trees was comparable across all sites. The heartwood sample was collected from each tree as an increment core (Haglöf Increment borer 350 mm x 5,15 mm; n = 27), with a FAMAG plug cutter of 15 mm diameter (n = 322), as a wood chunk (n = 20) or as a wood powder sample obtained with an electrical drill (n = 133). All samples were taken at least 14 cm into the tree. The heartwood samples were stored in plastic straws or paper envelopes and properly ventilated to prevent mold. Additionally, three cambium samples per tree were taken with punches of 2.5 cm diameter. The bark samples were stored in plastic bags with silica. The silica was refreshed up to 1 week after sampling to ensure the samples were dried fully. Additionally, GPS-coordinates and DBH were recorded.

5.2.2 Lab methodology

5.2.2.1 Genetic analysis

DNA was isolated from leaf, cambium or sapwood material from between 17 and 28 trees per site. Genomic material was isolated with an optimized cetyltrimethyl ammonium bromide (CTAB) protocol (Dumolin et al., 1995) with additional cleaning steps (Supporting Note S5.1). 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 three libraries, 300 bp or more insert size, with the 'RIPTIDE High Throughput Rapid Library Prep Kit' (Twist Bioscience, South San Francisco, USA). The libraries were

sequenced with Illumina Novaseq6000 PE150 (Novogene, Cambridge, United Kingdom). The Illumina sequences of the trees were mapped to the annotated chloroplast genome (MZ274135.1, Mascarello et al. (2021)) using Bowtie2 (Langmead & Salzberg, 2012)). The variant call considered all mapped reads without filtering and was performed using NGSEPcore (Tello et al., 2023). A variant call file containing only biallelic loci was generated, and further analyses were carried out using R ver- sion 4.1.0 (R Core Team, 2021). The detected variants underwent further filtering, including a minimum sequencing depth of three reads, a maximum depth of 250, and individuals with more than 50% missing data as well as SNPs with more than 25% missing data were removed (SNPfiltR package, DeRaad (2022)). This resulted in a genetic dataset of 322 trees. Haplotypes were defined in a non-restrictive manner with inclusion of sequences of varying lengths within the same haplotype, while ambiguities resulted in certain sequences being assigned to different haplotype (Pegas package, Paradis, 2010).

5.2.2.2 Stable isotope analysis

Between four and 10 trees per site were selected for stable isotope analysis, depending on the isotope: we measured $\delta^{34}S$ in three to four trees per species per site (total of 51), $\delta^{2}H$ in six or 10 (total of 101) and $\delta^{18}O$ in six or 10 (total of 105) trees per site. Trees were selected for isotopic measurements such that multiple isotopes were measured for the same set of trees as much as possible. This resulted in a geolocated database of 295 Azobé trees in total, between 1 and 3 isotopes measured per tree. $\delta^{13}C$ was measured in 12 sites, but was not found to add to the assignment accuracy. As measurements were missing at one site, this isotope was not included in the final model.

Stable isotope measurements are explained in detail in Chapters 5 of Boeschoten (2023). In short, a subsample of heartwood from every tree was cut in radial direction including at least 3-5 cm to include wood formed during multiple years. $\delta^{18}O$ and $\delta^{2}H$ were measured in cellulose, following (Vlam et al., 2018b) for cellulose extractions. $\delta^{34}S$ was measured in whole wood. The stable isotopes ratios were determined by IRMS, expressed in per mill (‰) relative to an international reference standard (V-SMOW for $\delta^{18}O$ and $\delta^{2}H$ and CDT for $\delta^{34}S$).

5.2.2.3 Multi-element analysis

Twenty trees per site were selected for multi-element analysis. The wood chemical composition was measured following Boeschoten et al. (2022). In short, a 1.0 g subsample was cut from 3-5 cm of heartwood and dissolved in 70% HNO3 by heating in a microwave digestion system (Mars 6, CEM Cooperation, USA). The lowest detection limit per element was calculated as three times the intensity of that element in a blank standard. If elements were found in quantities below the detection limit in more than 100 samples (the equivalent of about a quarter of the trees), they were excluded. This resulted in a multi-elemental composition of 41 elements, measured in 398 Azobé trees.

5.2.3 Statistical analysis

All other statistical analyses were performed in R version 4.2.3 (R core team, 2023). To test the variation between sites for each of the forensic methods, a db-RDA was performed based on 1-proportion of shared alleles for the cpDNA (PopGenReport package, Gruber & Adamack (2022)) and on Chord distances for stable isotope ratios and multi-element analysis (vegan package, Oksanen et al., 2020)).

5.2.3.1 Site classification

Site classification models were developed based on Random Forest analysis (ranger package, (Wright & Ziegler, 2017)). Random Forest models cannot be developed with missing values in the dataset, therefore imputation steps were performed to fill the gaps in the dataset. The cpSNPs dataset (haplotypes) contained the highest number of trees per site (up to 28). Missing data in the SNP dataset were randomly distributed across cpDNA and individual trees. Therefore, the most common allele was used at the population of origin to impute missing data, which may have resulted in a slight over-estimation of the accuracy of the Random Forest assignment: due to the imputation, some trees may have ended up with the same haplotype while in fact they were very similar but not identical, but the effect on site assignment was regarded to be minimal.

In the dataset of stable isotope ratios, less individuals per site were measured than for the

genetic analysis (between 4 and 10 per isotope), because it was expected that local variation was smaller in the isotope ratios. To match the size of the genetic dataset, isotopic data was imputed for the non-measured trees. We first randomly selected a subset of trees for which all measurements were performed for δ^2H and δ^{18}O and we excluded those from the dataset as test set, to minimize the effect of the imputation on further assessment of the assignment models (50% of all fully measured trees, so 44 test trees, spread across all 13 sites). For the remaining dataset, missing values were imputed with the site means.

For the multi-element analysis, 20 trees were measured per site. Elemental concentrations for the non-measured trees were imputed with site means to match the genetic dataset, like the stable isotope ratios. Again, only trees that were measured directly were used as test trees to calculate assignment accuracy.

We then developed seven types of Random Forest models based on different reference datasets: one per forensic method (SNP, ISO, EL), one for each pairwise combination of method (SNP- ISO, SNP-EL, ISO-EL), and one with all three methods included (SNP-ISO-EL). We tested their assignment accuracy as % of correctly assigned test trees. Test trees did not contain any imputed data. The model development was repeated 25 times per Random Forest model type (with different random subset of test trees), resulting in 7 x 25 assignment models. For each model, the sites to which test trees were assigned were saved and distances between actual and assigned origin were calculated. Assignment accuracy and mismatches were visualized in donut charts, based on the 25 x 44 test trees per Random Forest model category.

5.3 RESULTS

5.3.1 Genetic analysis

After filtering, a total of 238 cpSNPs for 322 individual trees composed the genetic dataset. Three main genetic clusters were observed based on proportion of shared haplotypes among sites (Figure S5.1). The first main cluster consisted of 10 sites, spanning from West Cameroon, to North-West Congo and Centre-East Gabon. This cluster showed a clear

substructure with sites CAM1 and CAM3 in West Cameroon being the most genetically distant from the other nine sites (Figure 5.1A). The second cluster, composed of GAB1, was the most genetically distinct cluster. The clear distinction of GAB1 indicates a possible species effect, as the site is in the region where a cryptic *L. alata* species occurs (*L. alata*1 - West Gabon, Ewédjè et al. (2020)). To a lesser extent, the third cluster was also distinct from the other two clusters. It was composed of two sites in North Congo (CON2 and CON4), and such a cluster was not previously detected using nuclear microsatellite markers (Ewédjè et al., 2020).

We used a Random Forest classification model to assign trees to their most likely site of origin based on the 238 cpSNPs. Assignment success to the correct site was 50.4% ± 3.7% (Figure 5.1A, 5.2), but it showed large variation across sites. Whereas no trees were correctly assigned at some sites (CAM2 and CAM4, Figure 5.1A), at other sites correct assignment was 100% (GAB1, CAM5 and CON4). Furthermore, the genetic groups corresponded to a clear spatial structure: incorrect assignments often occurred between groups of sites that were geographically close, such as between CON2 and CON4. This was also clear from the distances to predicted origin: 72.4% of the trees were assigned within 100 km of the sample site, 89.2% within 300 km, and no trees were assigned to sites more than 500 km away (Figure 5.2). This reflects a high precision.

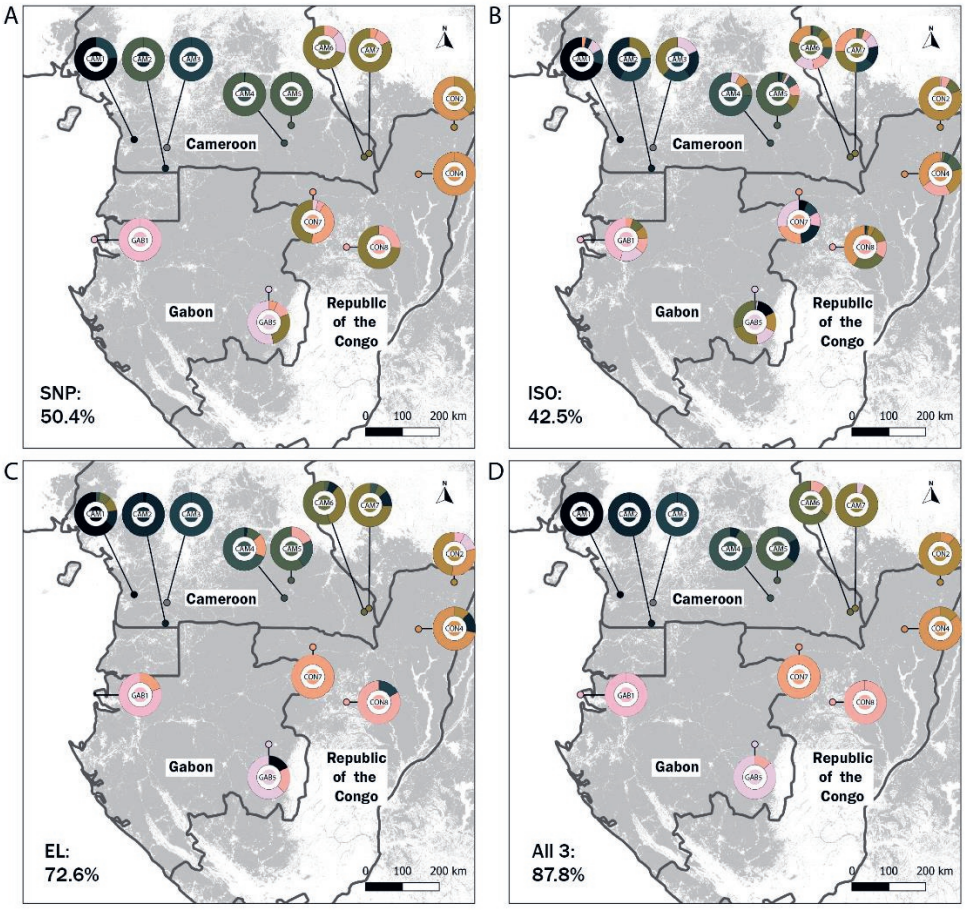


Figure 5.1: Confusion charts of the assignments of trees to their respective origin based on the different forensic methods: A) cpSNPs (SNP), B) three stable isotope ratios (ISO), C) multi- element analysis (EL) and D) all three combined. Mean assignment accuracy is indicated in the Figure. Colors in the inner circle indicate the color of the unique site. Colors in the outer circle indicate to which site the trees of that location were assigned. Primary tropical forest extent from Global Forest Watch is indicated in light grey (Turubanova et al., 2018).

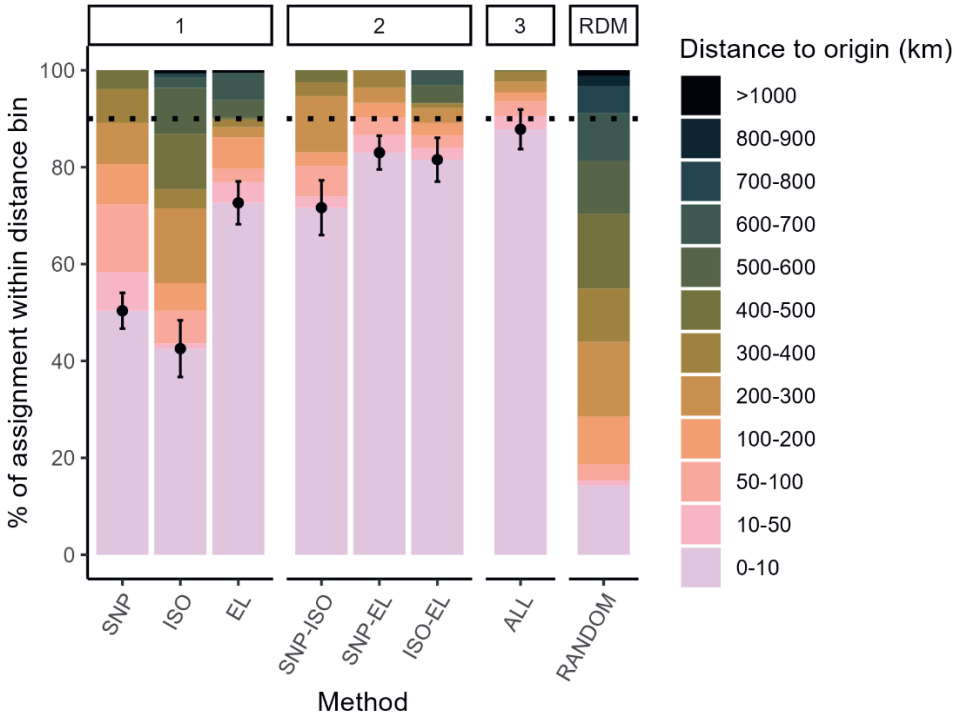


Figure 5.2: Distance to predicted origin (km), as % of the total number of test trees per distance category (bin) per method. The lowest bar represents correct site origin assignment, which is the assignment accuracy, also represented by a dot with an error bar that indicates the variation (st. dev.) in assignment accuracy, estimated using different test and training datasets of the Random Forest models. SNP = cpSNPs, ISO = stable isotope ratios, EL = multi-element analysis, ALL = all three methods, RDM = the occurrence of pairwise distances in the reference dataset, representing the result of random assignments. The dotted line indicates 90% accuracy.

5.3.2 Stable isotope analysis

Overall, stable isotope ratios ranged from 24.7 - 30.4 ‰ for $\delta^{18}O$, -9.5 - 36.55 ‰ for $\delta^{2}H$ and 4.6 - 10.9 ‰ for $\delta^{34}S$. Local variation in isotopic ratios was high, which resulted in a standard deviation within the sites that was similar to the variation in isotope ratios across all sites (Supplementry table S5.1). However, even with this high isotopic overlap among sites, site was significant in the db-RDA ($p<0.001$, Figure S5.2). Especially $\delta^{2}H$ was a strong predictor for the first axis, which explained 60.6% of the total variation.

Despite the significant differences among sites, assignment accuracy was rather low. Site-level Random Forest classification success was 43% ± 6% (Figure 5.1B, 5.2). Even though assignment success for some individual sites was around 75% (CAM1, CAM4 and CON2), it was much lower for most other sites. Furthermore, there was no clear spatial structure in the assignments. Trees were assigned to many different sites (up to eight out of 12), only 50% of the trees were assigned within 100 km and misassignments were as far as 1000 km from the original origin (low precision, Figure 5.2).

5.3.3 Multi-element analysis

Concentrations of the 41 measured elements varied between 0.001 µg/kg (Yb) and 4.0 g/kg (K) in the wood samples. Differences between sites were clear, as site was a significant predictor in the db-RDA ($p<0.001$, Figure S5.3). The Random Forest classification model based on the multi-elemental concentration performed better than that of the other methods: 73% ± 4% of trees was assigned to the correct site of origin (Figure 5.1C, 5.2) and 80% of the trees was assigned within 100 km of the sample site. Important elements for the classification based on variable importance were W, Ba, Mo, K and Cr. Sites with a high assignment accuracy were CAM2, CAM3 and CON7, where all trees were assigned correctly. However, across sites, trees that were not assigned to the correct site could be assigned to sites as far as 1000 km away (low precision, Figure 5.2). This contrasts with the stronger spatial structure in the cpSNPs data, for which distances to actual origin of mis-assigned individuals were much lower (high precision, Figure 5.2).

5.3.4 Combining methods

All the pairwise combinations of forensic methods increased assignment accuracy compared to the single methods (Figure 5.1D, 5.2), indicating that the methods complement each other. Especially combinations including multi-element analysis resulted in a high accuracy (83% and 82%). These were also the most consistent methods across different Random Forest models, indicated by a low standard deviation (Figure 5.2).

The combination of all three methods improved site assignment accuracy even further, to a

correct assignment of 88% ± 4%. It also yielded good precision: 91% of all trees was assigned within 50 km of the sample site and 94% of within 100 km. Furthermore, no trees were assigned more than 500 km away (Figure 5.2).

The complementarity of the methods was reflected in the individual site confusion charts (Figure 5.1), as the sites with the highest accuracy were not the same across the three methods. For example, three in west Cameroon (CAM1, CAM2 and CAM3) showed 100% correct site assignments based on the combination of methods, whereas none of the individual methods could distinguish all three well.

5.4 DISCUSSION

This is the first study that compared and combined genetic and chemical tracing methods to assign the origin of a tropical timber, Azobé. The combination of three methods (genetic differentiation based on chloroplast SNPs, stable isotope ratios, and multi-element analysis) resulted in a high accuracy for origin identification at small spatial scales: 88% of the trees was assigned to the correct site of origin and 93.6% was assigned within 100 km of its origin. This was considerably higher than the assignment accuracy of the individual methods (which ranged from 43% to 73%).

5.4.1 Combining methods improves tracing accuracy at small spatial scales

There was clear chemical and genetic variation within the study region (Figures S5.1-S5.3). As a result, site assignments of all three tracing methods performed better than random (Figure 5.2). Multi-elemental analysis performed best at these 13 locations (73%), followed by assignments based on cpSNPs (50%) and stable isotope ratios (43%). However, the genetic profiles were more region-specific than the other two methods: no trees were assigned more than 500 km away based on cpSNPs (Figure 5.2). In comparison, some trees (6%) were assigned more than 600 km away based on multi-element analysis, which would result in assignment to another country in this context.

Most importantly, the three methods showed distinct spatial patterns: the uniquely distin

guished sites differed between the three methods (Figure 5.1). It was clear that this was thanks to the distinct clusters and/or sites that were observed: the genetic clusters enhanced the spatial resolution of the chemical tracing for example. As a result, combining the three methods led to increased performance (88% correct at the site level, 94% within 100 km) and a reduction of the misassignment distances to less than 500 km (Figure 5.2). This complementarity can be explained by the different drivers of each method, thus resulting in a mosaic of spatial variation.

The reported combined assignment accuracy is comparable to the highest reported accuracies in other studies applying genetic analysis. Within Cameroon, blind samples of Tali were assigned to three sites as close as 14 km with 92% accuracy (Vlam et al., 2018b) and a batch of *E. cylindricum* samples was correctly verified to a concession of 2.500 km^2 with 86% accuracy (Jolivet & Degen, 2012). Yet, not all genetic studies found such differences at small spatial scales. Population assignment of *Shorea platyclados* in Malaysia yielded only 61% correct assignment at distances between 10 and 300 km (Ng et al., 2017) and for *Cedrela odorata* in Bolivia the assignment accuracy was only 66% at distances between 268 and 501 km (Paredes- Villanueva et al., 2019).

Compared to previous studies on individual chemical methods, the accuracy of the combined assignment exceeds any previously reported assignment at small spatial scales. As a comparison, *Aucoumea klaineana* timber from two locations in Gabon that were roughly 400 km apart differed in isotopic composition (δ^2H, δ^{18}O, δ^{13}C, δ^{15}N, δ^{34}S), but no assignment tests were performed (Watkinson et al., 2022a). Contrastingly, Tali timber (*Erythrophleum suaveolens* and *E. ivorense*) from three sites in Cameroon that were between 14 and 216 km apart could not be distinguished isotopically (δ^{18}O, δ^{13}C, δ^{15}N, Vlam et al. (2018b)). Overall, isotopic tracing has proven particularly successful across large environmental gradients, such as those caused by elevation (tested for *Picea abies*, Gori et al. (2018)) and over larger distances (tested for *Quercus* spp., Watkinson et al. (2020), chapter 5 in Boeschoten (2023)). However, region of origin could not even be distinguished based on δ^{13}C and δ^{18}O in four *Cedrela* species across a 500 m elevation gradient, spanning 400 km in the Bolivian Andes (Paredes-Villanueva et al., 2022). For multi-element analysis, the only existing other study was conducted on Tali for an overlapping set of sites, yielding similar assignment accuracy

as for Azobé here (69% vs 73%, Boeschoten et al. (2023)).

In contrast to previous studies, that mainly achieved high accuracy within a limited spatial extent or across countries, our study showcases the tracing potential with high accuracy across three countries while focusing on small spatial scales. To further improve small scale timber tracing for Azobé in Central Africa, we suggest improvements of the individual methods. Specifically for the genetic analysis, the development of a few nuclear DNA markers with high individual discriminating power in addition to the chloroplast DNA may add more genetic sub-structure to differentiate at the site level (Duminil et al., 2013; Blanc-Jolivet & Liesebach, 2015). Additionally, for all three methods it will be essential to cover the sampling gaps within the study region. Even though the current sampling was extensive, large parts of exploited forests remain unmeasured, such as in central Gabon or in the Democratic Republic of the Congo. Therefore, genetic and chemical profiles in these regions remain unknown.

5.4.2 Combining methods by stacking predictive maps

When sampling effort increases and more sites are added to the database, we also propose a statistical improvement for combining these different data types into a standardized approach for origin verification, illustrated in Figure 5.3. This would move tracing methods forward from the level of site verification to identifying a region of origin with a certain level of certainty, which strongly reduces the large dependency on reference datasets from all potential areas of origin (Truszkowski et al., 2023). The idea is based on prediction maps of isotopic composition, known as isoscapes (West et al. (2010), and chapters 4 and 5 in Boeschoten (2023). Based on the same principle, interpolated maps could also be made for each individual element based on relevant covariables such as soil pH and clay concentrations (Boeschoten et al., 2022). Furthermore, the frequency at which different genotypes occur at the different sites can be mapped based on Kriging statistics (Bucci et al., 2000; Manel et al., 2003). By stacking all these different maps, the potential area of origin can then be defined for any blind sample with a specific accuracy. This also requires the development an uncertainty map along with each prediction map, therewith explicitly including uncertainty in the assignment. This approach does not require every method to be applied to the same individual trees as maps are developed separately per method, making combining methods

more flexible.

To develop these stacked maps, the first requirement is to include more sites for all methods. Traditional statistical interpolation techniques such as kriging require a minimum of at least 30- 50 datapoints and these should be scattered throughout the study area, so that the distance at which spatial autocorrelation occurs can be estimated (Webster & Oliver, 2007). In the context of our study in Central Africa, this translates to at least 30-50 sampled sites that can accurately capture the genetic and chemical variations across the region. Interpolating the chemical data is relatively straightforward due to the continuous nature of the data, facilitating interpolation, as well as due to the availability of known gridded covariables that can enhance the model (West et al., 2010). However, the categorical genetic data requires the calculation of haplotype frequencies per site for the development of an interpolated map. This requires a higher number of measured trees per site. While the strictness of the minimum site requirement may vary for other interpolation methods like Quantile Regression Forests, the spatial autocorrelation between sampling points needs to be estimated in all cases to determine prediction uncertainty (Veronesi & Schillaci, 2019). Therefore, additional sampling should cover a variety of distances, to be able to have a good indication of the spatial structure found for each of the methods. The additional sites should cover the areas in which Azobé is harvested most, so that the resulting maps are relevant to be used in practice.

The main advantage of this approach lies in the potential of combining databases that don't fully overlap. The reason is that maps are produced for individual methods and only combined when assigning a (blind) sample. Therefore, the selection and distribution of sites can vary across tracing methods, depending on needs, limitations, strengths and costs of each individual method. This allows much more flexibility in the combination of methods compared to the method used in this study, which required all data to be collected at the same sites. This approach may also allow combining maps produced by different tracing projects or organizations.

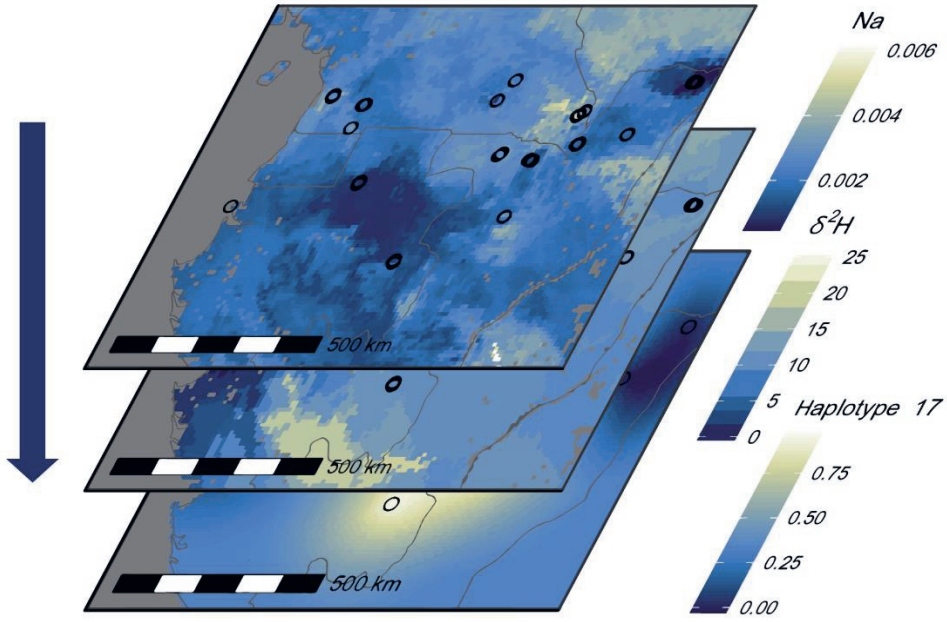


Figure 5.3: Examples of spatially interpolated maps of the study region. Individual maps based on properties of sampled trees may be integrated with maps of gridded covariables such as soil properties (for element maps) or environmental variables (for isotope maps). Prediction maps can be based on machine learning methods such as Quantile Regression Forest (chapters 4 and 5 in Boeschoten (2023)), or more classical spatial interpolation statistical techniques such as Kriging (Takoutsing & Heuvelink, 2022).

5.4.3 Implementing combined forensic tracing methods for other species

Our study reveals promising results of combining tracing methods for one species in one region. Evidently, the combination of tracing methods should be tested for other species as well as in other geographic regions. We anticipate that the complementarity of these tracing methods will be highest where the environmental gradients driving variation in wood genetic and chemical properties are smallest. This likely also explains why in our study region the combination of methods worked well. In Central Africa, no strong geographical barriers that would genetically separate tree populations are present, environmental variation driving wood isotopic signatures is limited (chapters 4 and 5 in Boeschoten (2023)), and variations in bedrock or soil type are minimal in some areas as well (Boeschoten et al., 2023). These factors

limit the potential for the individual tracing methods, but offer opportunities to reach the required accuracy levels by combining methods. Contrastingly, in regions where significant physical barriers or environmental gradients are present, such as mountain ranges, a single method may already be sufficiently reliable to address tracing issues with high accuracy, making the investment in multiple forensic methods unnecessary. Clearly, these hypotheses on the possible roles of geographic barriers and environmental gradients will need to be tested further.

5.5 CONCLUSION

The combination of three forensic tracing methods (population genetics, here with cpSNPs, stable isotope ratios and multi-element analysis) yielded highly reliable results to trace tropical timber origin at small spatial scales in Central Africa. The combined methods attained an unprecedented accuracy level in assigning the origin of timber at small spatial scales, and did so across three countries. This shows that the combination of methods is reliable and highlights its potential to be applied to address timber legality questions. Especially in areas with limited geographical barriers and environmental variation, such as in Central Africa, the combination of methods can reach highly accurate tracing results.

5.6 ACKNOWLEDGEMENTS

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 from WorldForestID. We thank all the collaborating timber companies and their field teams for facilitating the fieldwork and all involved colleagues at our partner institutes University of Dschang, Marien Ngouabi University, IRET/CENAREST and the National Herbarium of Gabon for help in arranging fieldwork.

5.7 PERMITS AND AUTHORIZATIONS

Cameroon: Research Permit No. 00000116/MINRESI/B00/C00/C10/C12 (Yaounde, 09 Sep 2019); Research Permit No. 000066/MINRESI/B00/C00/C10/C12 (Yaounde, 07 Jun 2021); Scientific research permit No. 2144 PRBS/MINFOF/SETAT/SG/DFAP/SDVEF/SC/NGY (Yaounde, 23 Jul 2021);ABS Permit 00010/MINEPDED/CNA/NP-ABS/ABS-FP (Yaounde, 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

Supporting information

Supporting Note S5.1. Modified DNA isolation protocol

Extraction of DNA with CTAB

Necessary buffers:

- NaCl 5 M: 29,22 g for 100 mL
- EDTA 0.5 M pH8.0: 18,61 g EDTA + 2 g NaOH per 100 mL (to adjust pH)
- Tris - HCl 1 M pH8.0: 8.88 g TrisHCl + 5.3 g Tris Base for 100 mL (to adjust pH)
- Cold Isopropanol (-20 ° C)
- Cold 70% ethanol (-20 ° C)
- Ethanol 95% cold (-20 ° C)
- Chloroform : isoamylalcohol (24 : 1)
- RNase A 20 mg/mL
- TE pH 8.0: 1 mL Tris 1 M pH8.0 + 200 µl of EDTA 0.5 M pH8.0 to 100 mL
- H2O milliQ

For 25 mL CTAB buffer:

Volume/quantity	Final concentration
0.5 g CTAB	2% CTAB
7 mL of NaCl 5 M	1.4 M NaCl
1 mL of EDTA 0.5 M	20 mM EDTA pH 8.0
2.5mL of 1 M Tris - HCl	100 mM Tris - HCl pH 8.0
Storage at room temperature	
1 mL DTT	2% DTT
1 g PVP	4% PVP
0.5 mL Proteinase K (20 mg/mL)	0.**4** mg/mL Proteinase K

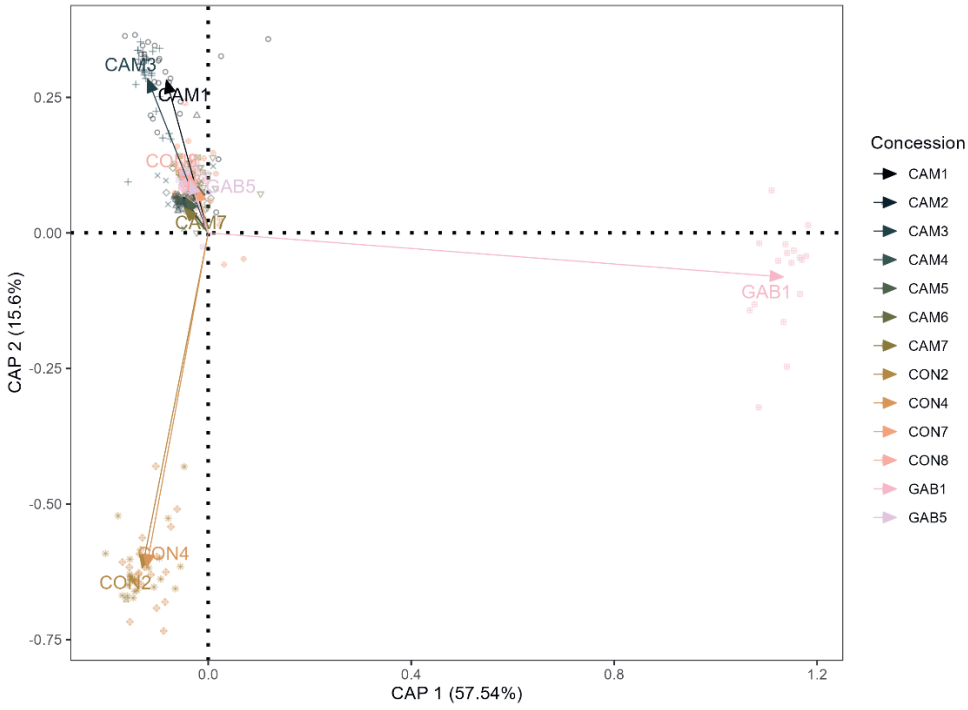
Protocol:

Weigh approximately 100 mg of powder in a 2 mL tube.

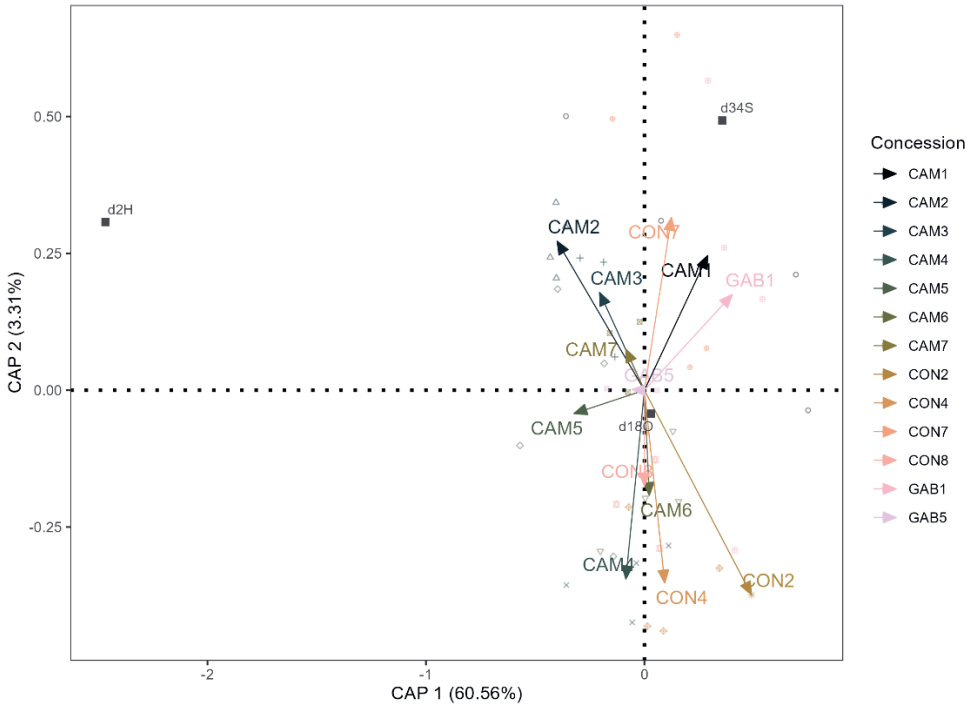
1. Prepare a master mix with 22.560 mL CTBA buffer (with PVP) + 960 mL DTT + 480 µL Proteinase K (20 mg/mL). OR 40µl of DTT and 20µl of Proteinase K per sample/tube.
2. Add 900 µl of CTAB buffer pre-heated at 60 °C. Incubate at 60 °C for 1 hour stirring occasionally (every 10 min).
3. Add chloroform : isoamylalcohol (24 : 1) 1V or 1 mL. **(900 µl)** Mix 1 min by inverting the tubes. Centrifuge 10 min at 14 000 x g. Retrieve the upper aqueous phase and transfer it into a clean 2 mL tube.
4. Add 10 µL of RNAse A (20 mg/ml). Incubate 1 hour h at 37 °C.
5. Add chloroform : isoamylalcohol (24 : 1) 1V or 1 mL. Mix 1 min by inverting the tubes. Centrifuge 10 min 14 000 x g. retrieve the upper aqueous phase and transfer it into a clean tube of 2 mL.
6. Estimate the amount recovered. Add 2/3 volume of cold isopropanol. Leave ½ hour minimum at - 20 °C. Centrifuge at max speed max for 15 min at 4 °C.
7. Remove the supernatant. Add 700 µL of EtOH 70% cold. Invert the tube several times to loosen the pellet. Centrifuge at max speed for 5 min at 4 °C.
8. Remove the supernatant. Add 700 µL of EtOH 95% cold. Invert the tube several times to loosen the pellet. Centrifuge at max speed for 5 min at 4 °C.
9. Dry the pellet. Resume in 100 µL H2 O milliQ (or TE^{-4}).

Supporting table S5.1. Mean and standard deviation of the three isotope ratios, averaged within sites as well as across sites, expressed relative to the respective international standards in ‰.

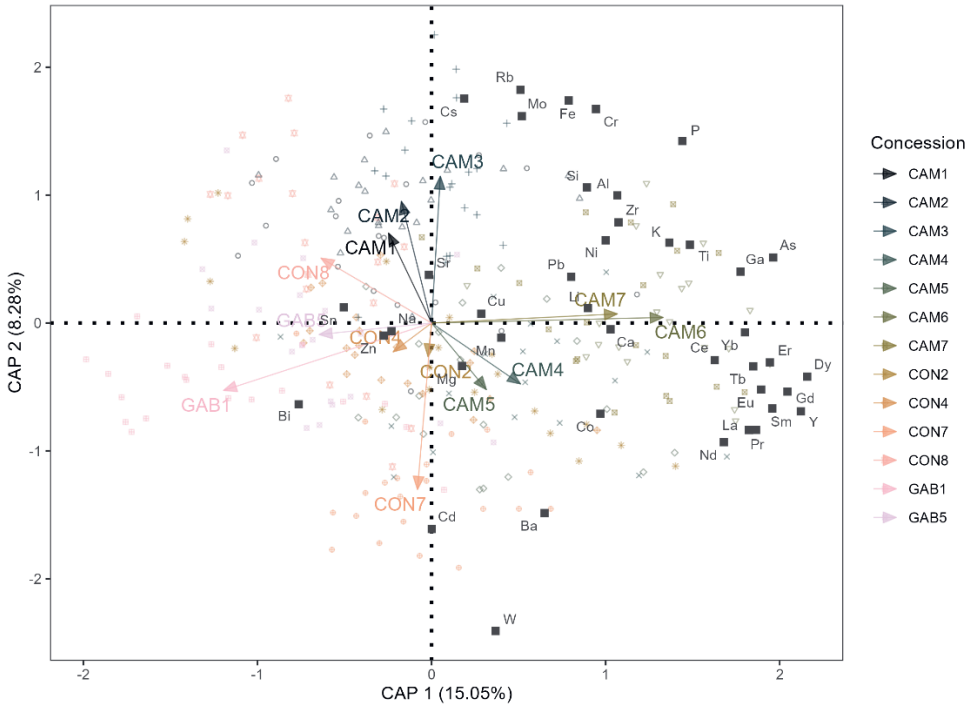
	$\delta^{18}O$	$\delta^{2}H$	$\delta^{34}S$
Mean of the averages within sites	28.0	14.4	7.2
SD within sites	0.7	7.7	0.8
Mean across all sites	27.9	12.7	7.2
SD across sites	1.1	10.7	1.4



Supporting figure S5.1. Db-RDA analysis of the genetic distances in the cpDNA, calculated based on 1- proportion of shared alleles across 13 sites, using 238 cpDNA loci for 322 trees. The colored dots indicate the trees, colored by site.



Supporting figure S5.2. Db-RDA analysis of isotopic composition of the 45 trees across 13 sites in which all three isotope ratios were measured (d18O, d2H and d34S), based on Chord distances. The coloured dots indicate the trees, coloured by site. The grey dots indicate the elements.



Supporting figure S5.3. Db-RDA analysis of the multi-elemental composition of 259 trees across 13 sites, based on Chord distances. The coloured dots indicate the trees, coloured by site. The grey dots indicate the elements.

Chapter 6

General Discussion

General Discussion

6.1 CONTEXT OF THIS PHD STUDY

My thesis aimed to advance genetic tracing tools to verify claims of species and geographic origin of tropical timber. I addressed four critical barriers: the creation of comprehensive reference databases, the development of effective genetic markers, the enhancement of assignment test success, and the improvement of spatial resolution for tracing.

To this end, extensive fieldwork was conducted to amass a large collection of reference samples (>700 trees), providing a robust basis for verification. The core of this research involved developing whole plastid genomes for SNP detection, allowing the differentiation of species and the verification of their geographic origins. This endeavor also included evaluating the spatial distribution of plastid genome haplotypes and the genetic structure of populations. In Chapters 2 and 3 I studied the plastid genome polymorphisms for Tali and Azobé timbers. In Chapter 4 I assessed the robustness of these findings for species and origin identification through blind sample tests. Plastome-wide SNP reference sets allowed us to trace 72% of Azobé blind samples within 100 km of their origin, while this was 20% for Tali.

Finally, I explored options to integrate genetic tracing with multi-element and isotopic analyses, aiming to refine the tracing accuracy further. Using plastome SNPs, stable isotope ratios, and multi-element composition, combining SNPs and elements achieved 83% accuracy at site level and 90% within 100 km. Incorporating all methods raised accuracy to 94% within 100 km and 90% at 50 km.

6.2 OBTAINING AND USING DNA FROM WOOD: CHALLENGES

One of the main challenges in genetic tracing of timber is the extraction of sufficient amounts of DNA of sufficient quality from woody tissue. Quality includes the parameters purity (including the absence of contaminants that can significantly impact the quality of sequencing libraries) and DNA fragmentation. Extensive DNA fragmentation can be intrinsic for some wood samples due to age or processing treatments, but it may also be increased by poor sample storage, mechanical shearing during extraction and purification, and enzymatic degradation (Rachmayanti et al., 2009b; Tnah et al., 2012b; But et al., 2023).

In this PhD study, I have isolated the total genomic DNA of 960 samples of Tali (*Erythrophleum* spp.) and Azobé (*Lophira* spp.) combined. I used three different protocols or kits for DNA extraction, and extracted DNA from four types of plant tissue. The DNA concentration obtained from these extractions depended on species, isolation protocol, and tissue type (Figure 6.1). Tali consistently outperformed Azobé across all tissues and protocols and in every evaluated case (cambium, heartwood, and leaf, both protocols, $p < 0.001$), indicating a strong role of species on DNA isolation procedures. This role is likely related to the production of species-specific secondary compounds and metabolites, substances which are known to complicate DNA analysis in many plants (e.g., Aboul-Maaty and Oraby, 2019; Akhmetzyanov et al., 2020; Rossi et al., 2021). Generally, young tissues (e.g. fresh leaves, sapwood) are easier than mature ones (e.g. herbarium leaves, heartwood). To account for differences among tissues and species, many standard DNA isolation protocols, such as those based on CTAB, incorporate species-specific cleaning steps (Porebski et al., n.d.; Rachmayanti et al., 2006; Tnah et al., 2012b; But et al., 2023). Indeed, also in this study, the addition of such species-specific modifications to the CTAB protocol led to enhancements in both the quantity and purity of the isolated DNA. Knowing beforehand what chemical compounds might be present in certain species complexes may help to target which cleaning steps are necessary and reduce laboratory work time and efficiency.

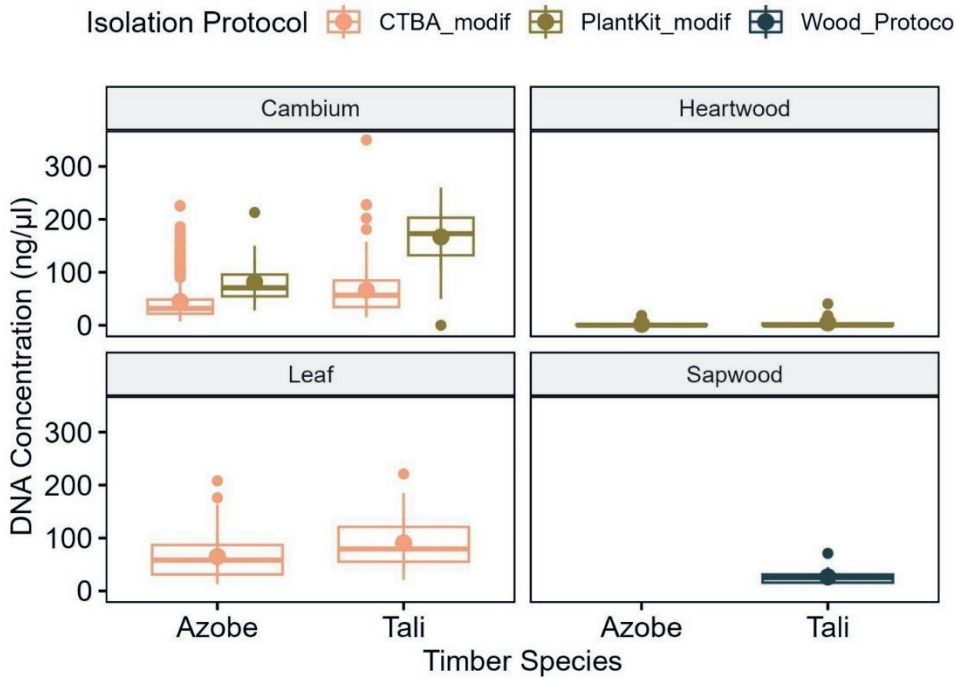


Figure 6.1. Effect of timber, tissue and DNA isolation protocol (colour) on DNA concentration. Shown are boxplots with median, inter quantile range and extreme values. DNA concentration data are derived from Chapter 2 (Tali sapwood, leaf and cambium, using CTAB protocol), Chapters 3 and 5 (Azobé leaf and cambium, CTAB protocol) and Chapter 4 (cambium and heartwood, using PlantKit protocol). No Azobé sapwood samples were included.

This study assessed the efficacy of different isolation protocols by comparing the processing of cambium samples using either the CTAB method with species-specific modifications or the Qiagen DNeasy Plant Mini Kit (final volumes 30 µl). The CTAB-based method is effective for challenging plant samples, yield good quantities of DNA, especially with small amounts of tissue, but requires hazardous chemicals and is labor-intensive, as noted by Abdel-Latif and Osman (2017). In contrast, the Qiagen DNeasy Plant Mini Kit simplifies the process and yields high-quality DNA, though it incurs higher costs and may take longer than other methods (Abdel-Latif and Osman, 2017; Dieki et al., 2022). The Plantkit significantly outperformed the CTAB protocol in DNA concentration, producing yields 1.8 and 2.5 times

greater for Azobé and Tali, respectively, supporting the findings of Tnah et al. (2012). Despite the low cost of CTAB protocol, the higher costs of the Qiagen Kit's is offset by a substantial increase in DNA yielded. Given its user-friendly process and superior DNA yield for specific species, as demonstrated in this study, the Qiagen Kit emerges as a preferred choice for if one values convenience and quality.

In assessing DNA concentrations across different plant tissue types, leaf samples were shown to be superior to cambium when subjected to the species-specific CTAB protocols. Utilizing species-specific modified CTAB protocols revealed that (herbarium) leaf samples yielded significantly higher DNA concentrations (64.87 ng/µl for Azobé and 90.72 ng/µl for Tali; $p < 0.001$) compared to cambium samples (44.70 ng/µl for Azobé and 65.88 ng/µl for Tali), similar to the findings reported by (Novaes et al., 2009). When contrasting DNA isolates from cambium and heartwood tissues, employing the Qiagen DNeasy Plant Mini Kit with extra steps (hereafter referred to as Qiagen kit), cambium samples yielded 38 to 49 times more DNA (for Tali and Azobé respectively) than heartwood samples ($p < 0.001$). The latter varied from failed isolation attempts to concentrations of 18.5 ng/µl for Azobé and 40.5 ng/µl for Tali. These results highlight the challenges of DNA extraction from heartwood, which has been repeatedly reported for tropical tree species (Rachmayanti et al., 2006, 2009b; Tnah et al., 2012b). Notably, DNA isolation from heartwood samples performed less optimally compared to cambium samples processed with the same protocol. Similar trends were shown by Tnah et al. (2012) who, using a similar Qiagen kit based protocol, achieved twice the amplification success for cambium and sapwood than for heartwood tissues samples. DNA from wood is often highly degraded and in low amounts due to age, chemical composition or industrial processing-related degradation (Rachmayanti et al., 2006; Degen and Fladung, 2007; Rachmayanti et al., 2009b; Lowe and Cross, 2011; Hung et al., 2017; He et al., 2019). But et al. (2023) recently compared three DNA isolation methods specially developed for wood and wooden products. Their findings indicated that SDS-based and CCDB methods outperformed the Qiagen-based method, which was utilized as the primary protocol in the present thesis. This highlights the continued potential for protocol enhancement in the field, suggesting that there is still ample opportunity to refine and optimize DNA isolation methodologies. DNA quality and quantity remain the main technical barrier to employ genetics for timber tracing.

Our findings indicate that genomic DNA of sufficient quality for successful sequencing-based genotyping can be obtained from various tissue types, with heartwood samples generally yielding lower DNA concentrations and thus resulting in lower sequencing success. Consequently, cambium samples have become the preferred choice for developing genetic reference data, considered a valuable source of genomic DNA and a practical alternative to leaf tissue, which is challenging to collect from tall trees (Novaes et al., 2009). As the DNA extraction success from heartwood is limited, efforts to enhance methodologies are ongoing, including Qiagen's kit update with an inhibitor removal step. Establishing genetic tracing as a reliable tool necessitates assessing different protocols' efficacy across species and understanding how species-specific chemical compositions impact DNA yield. Future research should prioritize the investigation of phylogenetic and chemical signatures influencing DNA extraction success and explore various protocols, including SDS-based and CCDB (But et al., 2023), to address challenges with heartwood samples. The protocol by Shi et al. (2012), which achieved 40–50% cpDNA purity, exemplifies how plastid-SNPs tracing tools can improve outcomes. Ultimately, selecting a method depends on research specifics, such as species focus and budgetary limits. For tracing, the selection of appropriate DNA isolation protocols for heartwood is crucial for the reliability and success of genetic tracing efforts. Enhancing DNA extraction methodologies in plant genetics research necessitates a strategic approach that includes testing and comparing isolation protocols in a ring-test setting and fostering a collaborative environment for sharing outcomes of DNA extractions. Such a collaborative effort can hasten the identification of effective methods and assist researchers in developing tailored protocols for specific species or families, thereby increasing the efficiency and automation of genetic studies.

6.3 STRATEGIES TO OBTAIN GENOMIC DATA FOR TIMBER TRACING

Genetic tracing in timber has advanced significantly with the advent of low-cost, high-throughput sequencing technologies, such as Next Generation Sequencing (NGS). In my thesis, I explored genetic tracing in timber using Whole Genome Sequencing (WGS, i.e. so-called 'shotgun sequencing') and detected Single Nucleotide Polymorphisms (SNPs) the resulting data. This aligns with the growing preference for WGS, especially for small genomes like the chloroplast, to identify species and geographic origin (Li et al., 2015; Mascarello et al., 2021). Despite the traditional use of Targeted Sequence (TS) markers, such as Simple Sequence Repeat (SSR) markers, for focused genotyping, the advent of cost-effective (Pei et al., 2023) NGS platforms like NovaSeq is shifting the focus towards WGS, and this trend is expected to continue in the future. WGS provides comprehensive genome coverage, enabling the exploration of a larger portion of genome(s) compared to TS. Studies proposed use of the whole genome for both species identification, as shown by Li et al. (2015), and for geographic tracing, as pointed out by Mascarello et al. (2021). Whole genomes can contain thousands to millions of polymorphisms, with SNPs being the most common type of polymorphism (Glover et al., 2010; Singh et al., 2013). Glover et al. (2010) demonstrated that highly informative SNP sets can outperform target markers in delineating genetic relationships at both individual and population levels.

Although our study primarily generated nuclear DNA sequencing data, the sequencing depth for nuclear DNA was generally insufficient to reliably identify SNPs, due to low coverage. By contrast, the plastid genomes, present in multiple copies per cell (Mader et al., 2018b), did achieve extensive coverage. This is consistent with previous studies that wooden tissues (unprocessed as well as processed) yielded higher success of the chloroplast DNA when compared to nuclear DNA (Rachmayanti et al., 2006, 2009b; Tnah et al., 2012b). Indeed, we obtained reliable data for the cpDNA (plastome), particularly so for cambium samples. The good performance of cambium samples was critical for the reference dataset development and for the tests of the tracing method principle. Even though most heartwood samples exhibited lower cpDNA coverage and many failed to provide data, the samples that were successfully sequenced provided ample data. Sequencing results for heartwood samples, despite only about 1% of the total generated data being plastome (1.4% for Tali and 0.62%

for Azobé), still had a depth of 196x for the cpDNA. The sequencing failure is likely linked to the low DNA amount obtained for the heartwood samples, and possibly also to higher fragmentation, although I did not study this in depth.

In this respect, further enhancement of the success rates of WGS-based SNP genotyping for heartwood samples might be possible via adjustment of the strategy for fragment size selection during library preparation. While long sequence reads may facilitate genome alignment, such reads may be too rare in the degraded DNA extracted from heartwood samples (Rachmayanti et al., 2006; Tnah et al., 2012b). Throughout all chapters in this thesis, we therefore used a protocol that selects DNA fragments with a specific length range (on average 375bp). Recent bioinformatic advances in the analysis of WGS data from ancient DNA samples, in which DNA quality may be even poorer resolution in shorter average fragment lengths (typically >100bp) (Nguyen et al., 2023), may allow focussing on shorter fragments for timber. This may then also enhance sequence coverage for heartwood samples.

The high coverage of plastomes in our study, across species and tissues, highlights the effectiveness of plastid whole-genome sequencing (WGS) in obtaining reliable genomic data. Provided that sufficient DNA is extracted from the tissue (e.g. heartwood), our results indicates that plastid WGS is both technically viable and cost effective (~30 euros per sample, library preparation plus sequencing costs). Moreover, the capability to select polymorphic SNPs from a comprehensive sequence set mitigates the risk of ascertainment bias, which could arise from pre-selecting polymorphic SNPs based on a limited number of test samples (Anderson, 2010). Working with whole plastome sequences enables the incorporation of new reference samples into the dataset, as it allows for the detection of new variants. It does broaden the reference area coverage and potentially allows detecting new genetic spatial (sub-) structure of the study species. Appropriate strategies for selecting genetic marker(s), DNA isolation protocols, genomic library preparation and sequencing are crucial to obtain reliable reference data. Transparency about these methods is also a necessary basis to improve reproducibility across molecular laboratories worldwide. As laboratory procedures become increasingly automated, it is essential to establish clear, reproducible routines. This is particularly the case for heartwood samples that often require multiple steps in isolation processes to obtain the required DNA quantity and quality. The establishment of reproducible

lab routines not only ensures data reliability but also eases the workload in the laboratory and thus paves the way towards uniform lab practices in genetic timber tracing across the world.

6.4 USING CHLOROPLAST DNA TO DIFFERENTIATE SPECIES

In this thesis, we differentiated study species using two different approaches. In Chapters 2 and 3, I evaluated species differences using the genome-wide set of chloroplast SNPs. In Chapter 4, we used three DNA barcode regions to verify species claims. The chosen method depended on the specific objective. The DNA barcode approach of Chapter 4 was conducted on blind samples that may have been derived from species or locations not covered by our reference datasets.

DNA barcoding using cpDNA is a well-established approach for taxonomic classification and phylogeography of plants, including tropical trees (e.g., Duminil et al., 2010; Höltken et al., 2012; Daïnou et al., 2016; He et al., 2019; Bouka et al., 2022). This method discriminates species based on the information of one or a few regions in the genome (i.e., the 'DNA barcodes'). The DNA barcode consists of a standardized short sequence of DNA (400–800 bp) that in principle can be amplified in all species using standardized primers and contains sufficient polymorphisms to be unique for each species. Several barcode regions exist, e.g., ITS, matK, rbcL, trnL, trnF, trnL-trnF, psbA-trnH, rpoB, and rpoC1 (Low et al., 2022). In plants the matK, rbcL, trnH-psbA, and trnL-F chloroplast regions are commonly used, often combined with the nuclear ITS sequence when the cpDNA regions do not provide sufficient discrimination power. Taxonomic classification research efforts have been encouraged by worldwide initiatives, including the 'Barcode of Life çonsortia, generating high amounts of reference barcodes which are shared via public online databases like NCBI. This necessitated a reliance on well-established reference data of DNA barcode regions, from sequences deposited at National Center for Biotechnology Information (NCBI, 2024). As a result, genetic reference data with the potential for species identification exists for 277 (86%) of the 322 taxa of tree species for which timber tracing is considered to be necessary (see Low et al.[2022] for a review on species identification status).

In Chapter 4, I successfully showed that – instead of using of the standard amplification based

barcoding-approach – sequence data for the same barcode regions could also be extracted from our plastome WGS data, allowing a comparison with the public NCBI data for these barcodes. BLAST (Camacho et al., 2009) analysis of cambium samples achieved 93-100% correct genus or family-level identification or correctly excluded for both Tali and Azobé blind samples. For samples identified as belonging to our study genera, either *Erythrophleum* (Tali) or *Lophira* (Azobé), we further leveraged whole-genome SNP data to test for species determination. While the timber claim verification was successful for cambium samples for both Azobé and Tali, heartwood samples presented challenges, with 53% of heartwood samples failing due to lower rate of DNA barcode regions retrieval. Nonetheless, for the heartwood samples that enough data was obtained, we observed a significant proportion correctly identified at the genus or family level or correctly exclude (93%). This demonstrates the potential for accurate timber verification when adequate DNA can be retrieved.

In this study, we utilized previously published and publicly available DNA barcode regions to verify timber claims at the genus level, though not at the species level. Different (combinations of) barcode markers vary in discriminatory power at the genus and species levels. Some of the commonly used combinations used are rbcL + matK (discriminate among 136 in a tropical dry evergreen forest in India [Nithaniyal et al., 2014]), trnL-trnF and ITS1 (Aquilaria species from other closely related species, and trnL (90% correct species-level identification of *Gonystylus* [Ng et al., 2016]), ITS2 + trnH-psbA (He et al., 2019), psbA-trnH + trnK (discriminate between five *Santalum* species with 100% success [Jiao et al., 2019]). Studies suggest that chloroplast DNA barcoding has advantages such as high conservation and robust amplification systems, but may face challenges in species assignment, phylogeny, and resolution due to complex speciation patterns and/or incomplete sampling (Kelchner, 2000; Duminil et al., 2010; Nock et al., 2011; Höltken et al., 2012; Zhang et al., 2013; Daniell et al., 2016; Hu et al., 2016; Daïnou et al., 2016; Yang et al., 2017; Mader et al., 2018b; Fahey et al., 2021; Mascarello et al., 2021; Jiao et al., 2019).

Although the focus on DNA barcode regions sufficed to verify timber claim of the studied timbers, our analysis confirmed that standard barcode regions only capture a minor portion of the polymorphisms that occur in the plastid genome. The standard DNA barcodes showed a limited resolution to distinguish species, particularly among closely related taxa within the

same genus, was evident for Azobé and Tali blind samples, a challenge also noted for *Pterocarpus* by Jiao et al. (2019). This challenge, alongside the decreasing costs of Next-Generation Sequencing (NGS), has prompted a shift towards Whole Genome Sequencing (WGS) for species identification. Li et al. (2015) proposed using the entire chloroplast DNA (cpDNA) as a super-barcode, shifting from genes to genomes. Genome-wide reference data are sparse and thus, this approach is feasible only for selected taxonomic groups with sufficient reference data already available. When taxonomic identification to the species level is needed, such as protected species listed in CITES ("Front | CITES," n.d.), there is a need for such expanded efforts on reference genomes assemblies and genomic databases for co-occurring related species identification (Nock et al., 2011; Li et al., 2015; Fahey et al., 2021).

In Chapter 2 and 3, we used our newly established reference datasets of plastid-wide SNPs to assess their potential as a 'super-barcode' for Tali and Azobé, respectively. While a clearer distinction among *Erythrophleum* species was achieved through plastome-wide SNP analysis (Chapter 2), the *Lophira* species discrimination remained unclear, likely due to a complex history of chloroplast capture within the genus (Chapter 3). The lack of species-specific polymorphic variation has been reported in the *Lophira* genus (Mascarello et al., 2021). The relationship between the two *Lophira* species is consistent with sharing of the same (of very similar) haplotypes across co-generic species as was also found in other tropical tree genera (Boom et al., 2021; Duminil et al., 2012; Migliore et al., 2019), and can hamper the species identification solely based on plastomes. This is linked to lower variability compared to nuclear DNA (Blanc-Jolivet and Liesebach, 2015), as it is maternally inherited and it is of non-recombinant nature.

The question that remains is at what level we need to be able to discriminate taxa to verify timber trade claims. For CITES-listed taxa, this may need to be at species level, but in other cases family or genus level may suffice (Ogden et al., 2009; Höltken et al., 2012). In this thesis (Chapter 4), the timber assignment was considered correct if the genus was verified, because other closely related species either do not occur in the region or are not used for timber. However, it is worth to further exploring recently proposed approaches (Li et al., 2015) to combine single-locus barcodes and super-barcodes. This could be done using SNP-rich genetic regions identified for African timber species (Mascarello et al., 2021).

Additionally, comparing sequenced chloroplast genomes of a large number of individuals from a set of co-occurring and related species (Fahey et al., 2021) may allow a deeper analysis of the interspecific and intraspecific variation and thus establish to what extent selected SNPs are species-specific. Such an approach could result in the creation of solid genetic reference databases for forensic analysis of internationally traded timber species (Mascarello et al., 2021).

6.5 HAPLOTYPES AS ORIGIN INDICATORS

Plastid genetic profiles – haplotypes – were used in this thesis to verify the geographic origin of timber. The frequencies and geographical distribution of haplotypes reflect the spatial genetic structure of the species. For instance, in Chapter 2 and 4 we found that the resolution was lower for Tali haplotypes that are frequent in East and West Cameroon. Similarly, in Chapters 3, 4 and 5 we found two prevalent Azobé haplotypes along the Cameroon, Gabon and Congo borders. For Tali rather a clear genetic distinction between West Africa and Central Africa was detected, as distinct haplotypes occur in both regions. However, local scale differentiation within Cameroon was primarily influenced by species distribution, with *E. ivorense* found in coastal areas and *E. suaveolens* in the Centre and East regions, as the haplotypes did not provide much within-region differentiation power. For Azobé, we identified three main genetic clusters within *Lophira* species, sharing the same geographic regions, demonstrating that the geographic genetic patterns surpassed species boundaries. If haplotypes with a wide geographic distribution are frequent, the spatial resolution to trace back a piece of wood which such haplotypes will be low. Such haplotypes will also imply that sites exclusively composed of a single haplotype are more likely to have samples assigned to them, as they present the highest frequency of the given haplotype and therefore of most closely related individuals (Degen et al., 2017). This underscores the significance of genetic diversity within a species and its impact on tracing accuracy. While widespread haplotypes challenge the precision of geographic origin identification, our research aimed to enhance this aspect in depth analysis of plastid wide SNPs, contrasting with previous studies that commonly used a combination of nuclear, mitochondrial and plastid SNPs.

Previous studies have primarily focused on population genetics and phylogeography (spatial structure) based on nuclear, plastid or mitochondrial, and occasionally exploring self-assignment within datasets (Low et al., 2022). At broader scales, Blanc-Jolivet et al. (2021) utilized a combination of nuclear, chloroplast, and mitochondrial SNPs to achieve a theoretical correct self-assignment rate for Azobé between West and Central Africa of 86%, while Honorio Coronado et al. (2020) reported theoretical success rates for Cumaru (*Dipteryx* spp.) of 91–100% to the correct genetic cluster and 69–92% to the country of origin. Our study, however, revealed moderate success (36-43%) at site-level (<10km) assignments. Yet, at a finer spatial scale than genetic cluster or country level assignments, as conducted by Honorio Coronado et al. (2020) and Blanc-Jolivet et al. (2021), our plastid-wide SNP set achieved significant success, with 83% of Tali trees accurately traced within <100km (primarily within Cameroon) and 90% of Azobé trees within 200-300 km (across Cameroon, Gabon, and Congo) (Chapter 4). The patterns observed for the haplotype distribution and frequencies of Tali (Chapter 2) and Azobé (Chapters 3) influenced the blind samples assignment results (Chapter 4). Hence, haplotypes distribution and frequency are good indicators of the spatial resolution and predictors of the success rate of haplotypes for timber tracing.

In conclusion, the ability of using haplotypes for timber tracing depends critically on the genetic spatial structure present in the species, the spatial coverage of reference samples, the granularity present in the samples, and the availability of markers that detect (fine) genetic structure. Part of the limitation of haplotypes for geographic tracing of timber may be overcome when combining them with chemical tracing methods. In Chapter, 5 we compared and combined the plastome-wide SNPs set with multi-element and isotopic profiling tracing methods. While multi-element analysis outperformed the other two methods at the site levels self-assignment (73%), the haplotypes were more region-specific than the other two methods. Each of the three tracing methods revealed clear spatial patterns, with each method highlighting unique sites or clusters, with different underlaying drivers, resulting in a mosaic of spatial variation. Interestingly, when combining all three methods this remarkably increased accuracy, up to 88% correct at the site level and 94% within 100 km. This result demonstrates the effectiveness of multi-method approaches in precise timber origin assignment. Notably, the genetic clusters significantly enhanced the spatial resolution of

chemical tracing, with misassignment distances reduced to less than 500 km (Chapter 5).

The assignment accuracy achieved in Chapter 5 is among the highest values reported for genetic tracing studies but reaches these levels at considerably finer spatial scales. For example, Honorio Coronado et al. (2020) reported 91–100% accuracy for Cumaru (*Dipteryx* spp.) at the correct genetic cluster level, while Vlam et al. (2018) achieved 92% precision for Tali blind samples from three concessions, of which two were only 14 km apart. In the legal prosecution of timber-related cases, two primary objectives arise: to either support or dispute a trade claim and/or to determine the timber's origin. The critical considerations include defining sufficient evidence for trade claims, determining when it is crucial to pinpoint origin, and specifying the desired level of resolution (concession, country region, country, or broader region). In both cases, different methods, or a combination of them, may provide evidence. The use of plastome-wide SNP sets facilitates fine- to medium-scale tracing precision (0-500 km). However, further steps are crucial before they can be applied in forensic settings. Future work should aim to develop statistical tools for analysing dependent haplotype data and its combination with other data types (e.g. chemical), and pursue forensic developmental validation of the pSNP reference sets, adhering to the guidelines set by the Scientific Working Group on DNA Analysis Methods (SWGDAM, 2012).

6.6 THE ETHICS OF TIMBER TRACING

The development of timber tracing methods, particularly for tropical timber, requires comprehensive reference data, based on extensive fieldwork across countries and international collaboration. This collaboration is vital as it taps into local knowledge of culture, environment, flora, and logistics, ensuring an equitable and effective approach. While the global North has access to more significant funding and resources, the global South faces the brunt of illegal timber harvesting and thus has a keen interest in developing systems to prevent forest resource depletion. However, the centralized leadership of projects in the global North may conflict with the logical placement of research activities in affected countries, raising issues of fairness and practicality.

The Timtrace project illustrated these challenges, especially when the COVID-19 pandemic

forced us to adapt sampling strategies to ensure progress. As a result of the pandemic, the initial plan for direct sample collection by PhD candidates evolved into a collaborative model with industry and academia, leading to successful sampling campaigns in Cameroon and Indonesia. This model highlighted the importance of cooperation and the need for decentralized sampling coordination, permit acquisition, engagement with timber companies, and logistical organization tailored to each country's context. The Wageningen team and local Timtrace partners fostered our collaboration through shared knowledge and training sessions. I personally conducted training in Cameroon and Indonesia, as well as online for Congo and Gabon. The partnership's strength was further amplified as we leveraged our network, with in-person training led by local teams. Specifically, a collaborator from Dschang University in Cameroon shared their expertise with the Institute for Research in Tropical Ecology team in Gabon. Subsequently, this knowledge was extended to train the Congolese Timtrace team at Marien Ngouabi University, illustrating a dynamic exchange and strengthening of skills and collaboration across our network. The pandemic thus changed the activities within this PhD thesis, and underscored the necessity of collaborative approaches and the acknowledgment of local parties in the co-creation of projects, communication, and scientific publications.

During the past decades, several initiatives have produced guidelines for timber tracing studies. For instance, this was done by the Global Timber Tracking Network (GTTN) and the World Forest ID (WFID). These initiatives aim to compile extensive databases of DNA profiles and chemical data crucial for timber tracing. However, maximizing the utility of these databases requires prioritizing data sharing, adhering to the Findable, Accessible, Interoperable, and Reusable (FAIR) principles (Wilkinson et al., 2016) and navigating regulatory landscapes surrounding data ownership, such as the Nagoya Protocol and Access and Benefit Sharing (ABS). Establishing clear data sharing agreements and logistical frameworks is essential for promoting transparency, accountability, and fostering trust among stakeholders. This comprehensive approach not only advances scientific understanding but also enhances the ethics and effectiveness of timber tracing efforts globally.

6.7 TIMBER TRACING IN THE FUTURE

The regulation of the timber sector to ensure transparency and legality has been a longstanding challenge due to the complexity of global supply chains and the widespread illegal practices. International regulations such as the European Union Deforestation Regulation (EUDR) and the US Lacey Act aim to address these issues. Yet, enforcement remains difficult due to the lack of transparency and the occurrence of forged documentation within the supply chain. The combination of population genetics, stable isotope ratios, and multi-element analysis has yielded highly reliable results for tracing tropical timber origin in Central Africa, as a means of generating independent evidence for legality of timber. The combination of methods has demonstrated unprecedented accuracy levels and can address timber claim questions down to the site level, when the area is covered in the reference data. Particularly in regions with limited geographical barriers and environmental variation, such as Central Africa, the combination can achieve highly accurate tracing results. Integration of these and other methods is thus essential, and one approach could involve stacking probability maps with genetic and chemical data.

Genetic tracing studies counts with various methods for detecting Single Nucleotide Polymorphisms (SNPs), which can be automated using technologies like SNP arrays. In this thesis, genotyping for a broad sample set to obtain whole plastomes was achieved through Whole Genome Sequencing (WGS) showcasing an approach to minimize costs without sacrificing depth. Future studies could explore more targeted strategies, such as target enrichment before WGS or designing specific primers for each SNP for amplicon-based sequencing or SNP array development. This tailored approach offers a balance between comprehensive genomic analysis and cost-effectiveness, enhancing the efficiency of genetic tracing in larger sample sets.

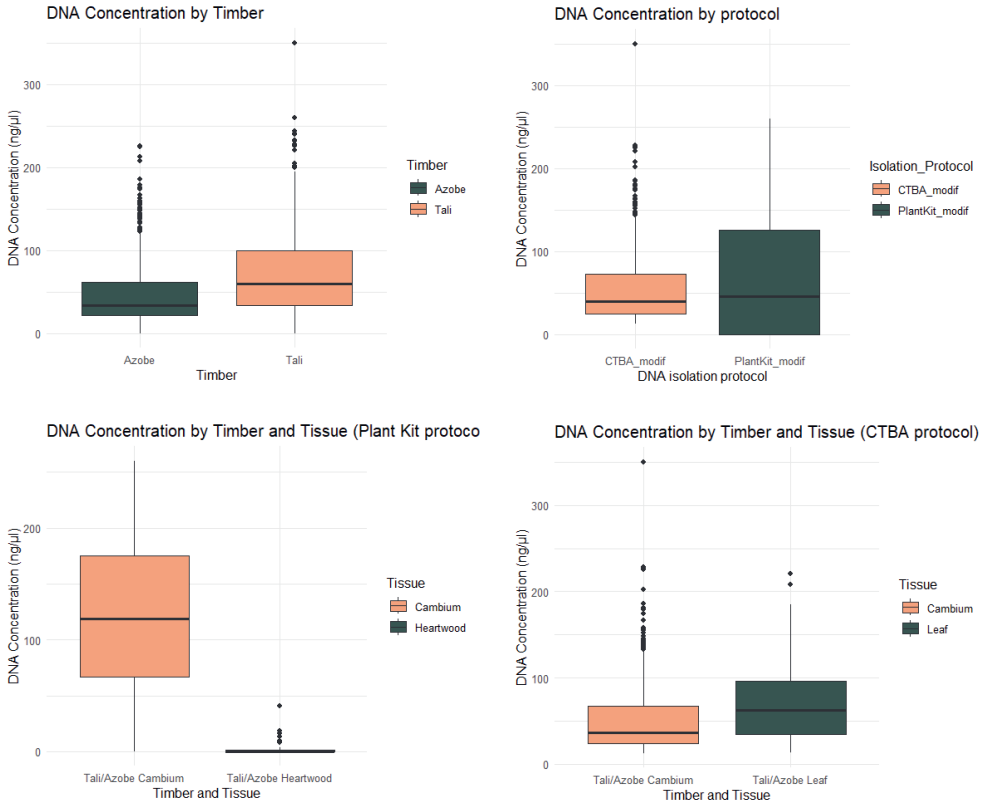
Moving forward also integrating visual, blockchain, and remote sensing data may contribute to achieving comprehensive traceability and verification. For example, He et al. (2019) employed BLOG and WEKA for wood species identification . BLOG, a supervised machine learning algorithm, uses logic formulae for species identification, and used a training and testing sets (He et al., 2019). WEKA workbench assists in classification and clustering for

DNA sequence analysis, utilizing cross-validation model (He et al., 2019). Such machine learning techniques could be tested for origin identification, and adapting haplotype networks to project new individual haplotypes as assignments. Data interpolation methods have been proposed to bridge the gaps between sites, suggesting a network of reference populations across species distributions, although these methods will struggle with distinct or rare haplotypes. Machine learning techniques, combined with haploscapes (maps of the haplotype distribution and their probabilities as provide by GeoAssign) methods could improve scalability and broaden their application across research fields (Truszkowski et al., 2023) and pave the way towards innovative phylogenetics and timber tracing. Isotopic, haplotypic, and chemical profile distribution maps would enable taking tracing techniques to a higher level, from site verification to identifying the region of origin with greater certainty (Truszkowski et al., 2023). Development of uncertainty maps alongside these prediction maps is essential, explicitly incorporating uncertainty into the assignment process. To realize these stacked maps, the inclusion of more sites across all methods is crucial (Boeschoten et al., 2023). Using a multitude of methods would make it much easier to fill in gaps in coverage, e.g., regions where it is impossible or too dangerous to sample. This flexibility would also allow for the integration of maps produced by different projects or organizations, enhancing collaboration and expanding the application of combined forensic tracing methods to other species and geographic regions. Testing the effectiveness of combined tracing methods in other species and regions is imperative. While our study demonstrates promising results in Central Africa, it is essential to evaluate the complementarity of tracing methods in environments with different environmental gradients and geographic barriers. Regions with limited environmental variation, such as Central Africa, may benefit most from combined methods due to the limited efficacy of individual techniques.

Determining the specific question to answer is crucial in cases where the species or origin of timber is disputed. Do we need to verify the likelihood of timber originating from a particular claimed area or country, or identify its source without such claims of origin? Engaging in solution-based research and collaborating with policymakers, the timber industry and practitioners, such as customs officials, emphasize the importance of providing decision-makers with the necessary knowledge to support conservation efforts effectively (Di Marco et al., 2017). Moreover, by focusing on solution-oriented research and adapting aspects like

graduate education, research funding, and media strategies to more effectively address the biodiversity loss (Fonseca et al., 2021), such as the trade of deforestation related products like timber.

Supporting information



Supporting figure S6.1. ANOVA tests for DNA concentrations. Box plots comparing A) Species (*Lophira* spp vs *Erythrophleum* spp)($P < 0.001$); B) DNA Isolation Protocol (CTBA Vs Plant Kit)($P < 0.05$); Plant Tissues C) Cambium Vs Heartwood ($P < 0.001$) and D) Cambium vs Leaf ($P < 0.001$).

References

Aboul-Maaty, N.A.-F., Oraby, H.A.-S., 2019. Extraction of high-quality genomic DNA from different plant orders applying a modified CTAB-based method. Bulletin of the National Research Centre 43, 25. https://doi.org/10.1186/s42269-019-0066-1

Akhmetzyanov, L., Copini, P., Sass-Klaassen, U., Schroeder, H., Groot, G.A. de, Laros, I., Daly, A., 2020a. DNA of centuries-old timber can reveal its origin. Scientific Reports 10. https://doi.org/10.1038/s41598-020-77387-2

Akhmetzyanov, L., Sánchez-Salguero, R., García-González, I., Buras, A., Dominguez-Delmás, M., Mohren, F., Ouden, J. den, Sass-Klaassen, U., 2020b. Towards a new approach for dendroprovenancing pines in the Mediterranean Iberian Peninsula. Dendrochronologia 60. https://doi.org/10.1016/j.dendro.2020.125688

Allen, S. T., Sprenger, M., Bowen, G. J., & Brooks, J. R. (2022). Spatial and Temporal Variations in Plant Source Water: O and H Isotope Ratios from Precipitation to Xylem Water. (pp. 501–535). doi:doi: 10.1007/978-3-030-92698-4 18.

Anderson, E.C., 2010. Assessing the power of informative subsets of loci for population assignment: Standard methods are upwardly biased. Molecular Ecology Resources 10, 701–710. https://doi.org/10.1111/j.1755-0998.2010.02846.x

Bandelt, H.J., Forster, P., Rohl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution 16, 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036

Blanc-Jolivet, C., Kersten, B., Bourland, N., Guichoux, E., Delcamp, A., Doucet, J.L., Degen, B., 2018. Development of nuclear SNP markers for the timber tracking of the African tree species Sapelli, Entandrophragma cylindricum. Conservation Genetics Resources 10, 539–541. https://doi.org/10.1007/s12686-017-0872-4

Blanc-Jolivet, C., Liesebach, M., 2015. Tracing the origin and species identity of Quercus robur and Quercus petraea in Europe: a review. Silvae Genet. 64, 182–193. https://doi.org/10.1515/sg-2015-0017

Blanc-Jolivet, C., Mader, M., Bouda, H.-N., Massot, M., Daïnou, K., Yene, G., Opuni-Frimpong, E., Degen, B., 2021. Development of new SNP and INDEL loci for the valuable African timber species Lophira alata. Conservation Genet Resour 13, 85–87. https://doi.org/10.1007/s12686-020-01173-5

Blanc-Jolivet, C., Yanbaev, Y., Kersten, B., Degen, B., 2018. A set of SNP markers for timber tracking of Larix spp. in Europe and Russia. Forestry 91, 614–628. https://doi.org/10.1093/forestry/cpy020

Boeschoten, L.E., 2023. Chemical methods for timber tracing. Wageningen University. https://doi.org/10.18174/633639

Boeschoten, L.E., Sass-Klaassen, U., Vlam, M., Comans, R.N.J., Koopmans, G.F., Meyer-Sand, B.R.V., Tassiamba, S.N., Tchamba, M.T., Zanguim, H.T., Zemtsa, P.T., Zuidema, P.A., 2022. Clay and soil organic matter drive wood multi-elemental composition of a tropical tree species: Implications for timber tracing. Science of The Total Environment 849, 157877. https://doi.org/10.1016/j.scitotenv.2022.157877

Boeschoten, L.E., Vlam, M., Sass-Klaassen, U., Meyer-Sand, B.R.V., Adzkia, U., Bouka, G.D.U., Ciliane-Madikou, J.C.U., Obiang, N.L.E., Guieshon-Engongoro, M., Loumeto, J.J., Mbika, D.M.F., Moundounga, C.G., Ndangani, R.M.D., Bourobou, D.N., Rahman, M.M., Siregar, I.Z., Tassiamba, S.N., Tchamba, M.T., Toumba-Paka, B.B.L., Zanguim, H.T., Zemtsa, P.T., Zuidema, P.A., 2023. A new method for the timber tracing toolbox: applying multi-element analysis to determine wood origin. Environ. Res. Lett. 18, 054001. https://doi.org/10.1088/1748-9326/acc81b

Boom, A.F., Migliore, J., Kaymak, E., Meerts, P., Hardy, O.J., 2021. Plastid introgression and evolution of African miombo woodlands: New insights from the plastome-based phylogeny of Brachystegia trees. Journal of Biogeography 48, 933–946. https://doi.org/10.1111/jbi.14051

Bouka, G.U.D., Doumenge, C., Ekué, M.R.M., Daïnou, K., Florence, J., Degen, B., Loumeto, J.J., McKey, D., Hardy, O.J., 2022. Khaya revisited: Genetic markers and morphological analysis reveal six species in the widespread taxon K. anthotheca. TAXON 71, 814–832. https://doi.org/10.1002/tax.12720

Bucci, G., Vendramin, G. G., & Ellstrand, N. C. (2000). Delineation of genetic zones in the European Norway spruce natural range: Preliminary evidence. Molecular Ecology, 9 , 923– 934. doi:doi: 10.1046/j.1365-294X.2000.00946.x.

But, G.W.-C., Wu, H.-Y., Siu, T.-Y., Chan, K.-T., Wong, K.-H., Lau, D.T.-W., Shaw, P.-C., 2023. Comparison of DNA extraction methods on CITES-listed timber species and application in species authentication of commercial products using DNA barcoding. Sci Rep 13, 151. https://doi.org/10.1038/s41598-022-27195-7

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 421. https://doi.org/10.1186/1471-2105-10-421

Cernusak, L. A. et al. (2022). Do 2H and 18O in leaf water reflect environmental drivers differently? New Phytologist , (pp. 41–51). doi:doi: 10.1111/nph.18113.

Chabot, A. A., Hobson, K. A., van Wilgenburg, S. L., McQuat, G. J., & Lougheed, S. C. (2012). Advances in linking wintering migrant birds to their breeding-ground origins using combined analyses of genetic and stable isotope markers. PLoS ONE, 7 . doi:doi: 10.1371/journal.pone.0043627.

Cuperus, J.T., 2022. Single-cell genomics in plants: current state, future directions, and hurdles to overcome. Plant Physiology 188, 749–755. https://doi.org/10.1093/plphys/kiab478

D'Andrea, R., Corona, C., Poszwa, A., Belingard, C., Dom´ınguez-Delm´as, M., Stoffel, M., Crivellaro, A., Crouzevialle, R., Cerbelaud, F., Costa, G., & Paradis-Grenouillet, S. (2023). Combining conventional tree-ring measurements with wood anatomy and strontium isotope analyses enables dendroprovenancing at the local scale. Science of the Total Environment, 858 . doi:doi: 10.1016/j.scitotenv.2022.159887.

Dadzie, A.M., Livingstone, D.S., Opoku, S.Y., Takrama, J., Padi, F., Offei, S.K., Danquah, E.Y., Motamayor, J.C., Schnell, R.J., Kuhn, D.N., 2013. Conversion of Microsatellite Markers to Single Nucleotide Polymorphism (SNP) Markers for Genetic Fingerprinting of Theobroma cacao L. Journal of Crop Improvement 27, 215–241. https://doi.org/10.1080/15427528.2012.752773

Daïnou, K., Blanc-Jolivet, C., Degen, B., Kimani, P., Ndiade-Bourobou, D., Donkpegan, A.S.L., Tosso, F., Kaymak, E., Bourland, N., Doucet, J.L., Hardy, O.J., 2016. Revealing hidden species diversity in closely related species using nuclear SNPs, SSRs and DNA sequences - a case study in the tree genus Milicia. BMC Evol. Biol. 16, 1–15. https://doi.org/10.1186/s12862-016-0831-9

Daniell, H., Lin, C.-S., Yu, M., Chang, W.-J., 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol 17, 134. https://doi.org/10.1186/s13059-016-1004-2

Degen, B., Blanc-Jolivet, C., Stierand, K., Gillet, E., 2017. A nearest neighbour approach by genetic distance to the assignment of individual trees to geographic origin. Forensic Science International: Genetics 27, 132–141. https://doi.org/10.1016/j.fsigen.2016.12.011

Degen, B., Fladung, M., 2007. Use of DNA-markers for tracing illegal logging, in: Proceedings of the International Workshop "Fingerprinting Methods for the Identification of Timber Origins" Held in Bonn - Germany, October 8-9 2007. pp. 6–14.

Degen, B., Ward, S.E., Lemes, M.R., Navarro, C., Cavers, S., Sebbenn, A.M., 2013. Verifying the geographic origin of mahogany (Swietenia macrophylla King) with DNA-

fingerprints. Forensic Science International: Genetics 7, 55–62. https://doi.org/10.1016/j.fsigen.2012.06.003

Deng, Y., Liu, T., Xie, Y., Wei, Y., Xie, Z., Shi, Y., Deng, X., 2020. High Genetic Diversity and Low Differentiation in Michelia shiluensis, an Endangered Magnolia Species in South China. Forests 11, 469. https://doi.org/10.3390/f11040469

DeRaad, D. A. (2022). *Snpfiltr* : An R package for interactive and reproducible SNP filtering. Molecular Ecology Resources, 22 , 2443–2453. doi:doi: 10.1111/1755-0998.13618.

Di Marco, M., Chapman, S., Althor, G., Kearney, S., Besancon, C., Butt, N., Maina, J.M., Possingham, H.P., Rogalla von Bieberstein, K., Venter, O., Watson, J.E.M., 2017. Changing trends and persisting biases in three decades of conservation science. Global Ecology and Conservation 10, 32–42. https://doi.org/10.1016/j.gecco.2017.01.008

Dieki, R., Eyang-Assengone, E.-R., Makouloutou-Nzassi, P., Bangueboussa, F., Emvo, E.N., Akue, J.P., 2022. Comparison of six methods for Loa loa genomic DNA extraction. PLOS ONE 17, e0265582. https://doi.org/10.1371/journal.pone.0265582

Dormontt, E.E., Boner, M., Braun, B., Breulmann, G., Degen, B., Espinoza, E., Gardner, S., Guillery, P., Hermanson, J.C., Koch, G., Lee, S.L., Kanashiro, M., Rimbawanto, A., Thomas, D., Wiedenhoeft, A.C., Yin, Y., Zahnen, J., Lowe, A.J., 2015. Forensic timber identification: It's time to integrate disciplines to combat illegal logging. Biological Conservation 191, 790–798. https://doi.org/10.1016/j.biocon.2015.06.038

Dormontt, E.E., Jardine, D.I., van Dijk, K.J., Dunker, B.F., Dixon, R.R.M., Hipkins, V.D., Tobe, S., Linacre, A., Lowe, A.J., 2020. Forensic validation of a SNP and INDEL panel for individualisation of timber from bigleaf maple (Acer macrophyllum Pursch). Forensic Science International: Genetics 46. https://doi.org/10.1016/j.fsigen.2020.102252

Drivelos, S. A., & Georgiou, C. A. (2012). Multi-element and multi-isotope-ratio analysis to determine the geographical origin of foods in the European Union. TrAC Trends in Analytical Chemistry, 40 , 38–51. doi:doi: 10.1016/j.trac.2012.08.003.

Duminil, J., Brown, R.P., Ewédjè, E.E.B., Mardulyn, P., Doucet, J.L., Hardy, O.J., 2013. 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 Evolutionary Biology 13. https://doi.org/10.1186/1471-2148-13-195https://doi.org/10.1016/j.ympev.2011.09.020

Duminil, J., Heuertz, M., Doucet, J.L., Bourland, N., Cruaud, C., Gavory, F., Doumenge, C., Navascués, M., Hardy, O.J., 2010. CpDNA-based species identification and

phylogeography: Application to African tropical tree species. Mol. Ecol. 19, 5469–5483. https://doi.org/10.1111/j.1365-294X.2010.04917.x

Duminil, J., Kenfack, D., Viscosi, V., Grumiau, L., Hardy, O.J., 2012. Testing species delimitation in sympatric species complexes: The case of an African tropical tree, Carapa spp. (Meliaceae). Molecular Phylogenetics and Evolution 62, 275–285. https://doi.org/10.1016/j.ympev.2011.09.020

Duminil, J., Mona, S., Mardulyn, P., Doumenge, C., Walmacq, F., Doucet, J.-L., Hardy, O.J., 2015. Late Pleistocene molecular dating of past population fragmentation and demographic changes in African rain forest tree species supports the forest refuge hypothesis. J. Biogeogr. 42, 1443–1454. https://doi.org/10.1111/jbi.12510

Dumolin, S., Demesure, B., Petit, R.J., 1995. Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. Theor. Appl. Genet. 91, 1253–1256. https://doi.org/10.1007/BF00220937

Eba'a Atyi, H.H., Lescuyer, 2022. The Forests of the Congo Basin: State of the Forests 2021. Center for International Forestry Research (CIFOR). https://doi.org/10.17528/cifor/008700

Ehleringer, J. R., Casale, J. F., Lott, M. J., & Ford, V. L. (2000). Tracing the geographical origin of cocaine: Cocaine carries a chemical fingerprint from the region where the coca was grown. Nature, 408 , 311–312. doi:doi: 10.1038/35042680.

Erythrophleum africanum (Welw. ex Benth.) Harms [WWW Document], n.d. . PROTA4U. URL https://prota.prota4u.org/protav8.asp?g=pe&p=Erythrophleum+africanum (accessed 2.8.24).

Ewédjè, E.E.B.K., Jansen, S., Koffi, G.K., Staquet, A., Piñeiro, R., Essaba, R.A., Obiang, N.L.E., Daïnou, K., Biwolé, A.B., Doucet, J.L., Hardy, O.J., 2020. Species delimitation in the African tree genus Lophira (Ochnaceae) reveals cryptic genetic variation. Conservation Genetics 21, 501–514. https://doi.org/10.1007/s10592-020-01265-7

Fahey, P.S., Fowler, R.M., Udovicic, F., Cantrill, D.J., Bayly, M.J., 2021. Use of plastid genome sequences in phylogeographic studies of tree species can be misleading without comprehensive sampling of co-occurring, related species. Tree Genetics and Genomes 17. https://doi.org/10.1007/s11295-021-01524-9

Fonseca, C.R., Paterno, G.B., Guadagnin, D.L., Venticinque, E.M., Overbeck, G.E., Ganade, G., Metzger, J.P., Kollmann, J., Sauer, J., Cardoso, M.Z., Lopes, P.F.M., Oliveira, R.S., Pillar, V.D., Weisser, W.W., 2021. Conservation biology: four decades of problem- and

solution-based research. Perspectives in Ecology and Conservation 19, 121–130. https://doi.org/10.1016/j.pecon.2021.03.003

CITES [WWW Document], 2024. URL https://cites.org/eng (accessed 2.29.24).

GBIF.org (2021). GBIF Occurrence Download https://doi.org/10.15468/dl.wpwm8v. Date accessed: 2021-11-25.

Global Timber Tracking Network, 2016. URL https://globaltimbertrackingnetwork.org/ (accessed 2.1.24).

Glover, K.A., Hansen, M.M., Lien, S., Als, T.D., Høyheim, B., Skaala, Ø., 2010. A comparison of SNP and STR loci for delineating population structure and performing individual genetic assignment. BMC Genet. 11. https://doi.org/10.1186/1471-2156-11-2

Gonzálvez, A., Llorens, A., Cervera, M.L., Armenta, S., de la Guardia, M., 2009. Elemental fingerprint of wines from the protected designation of origin Valencia. Food Chem. 112, 26–34. https://doi.org/10.1016/j.foodchem.2008.05.043

Gorel, A.-P., 2019. Climatic niche evolution within tropical tree lineages: the case of the Erythrophleum genus (Fabaceae - Caesalpinioideae) in Africa. Liège Univertité.

Gorel, A.P., Duminil, J., Doucet, J.L., Fayolle, A., 2019. Ecological niche divergence associated with species and populations differentiation in Erythrophleum (Fabaceae, Caesalpinioideae). Plant Ecol. Evol. 152, 41–52. https://doi.org/10.5091/plecevo.2019.1543Gruber, B., Adamack, A.T., 2015. landgenreport: a new r function to simplify landscape genetic analysis using resistance surface layers. Molecular Ecology Resources 15, 1172–1178. https://doi.org/10.1111/1755-0998.12381

Gori, Y., Stradiotti, A., & Camin, F. (2018). Timber isoscapes. A case study in a mountain area in the Italian Alps. PLOS ONE, 13 , e0192970. doi:doi: 10.1371/journal.pone.0192970.

Gruber, B., & Adamack, A. (2022). PopGenReport. Harris, I., Osborn, T., Jones, P., & Lister, D. (2020). Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. Scientific Data, 7 .

Gruber, B., Adamack, A.T., 2015. landgenreport: a new r function to simplify landscape genetic analysis using resistance surface layers. Molecular Ecology Resources 15, 1172–1178. https://doi.org/10.1111/1755-0998.12381

Hardy, O.J., Born, C., Budde, K., Daïnou, K., Dauby, G., Duminil, J., Ewédjé, E.-E.B.K., Gomez, C., Heuertz, M., Koffi, G.K., Lowe, A.J., Micheneau, C., Ndiade-Bourobou,

D., Piñeiro, R., Poncet, V., 2013. Comparative phylogeography of African rain forest trees: A review of genetic signatures of vegetation history in the Guineo-Congolian region. Comptes Rendus Geoscience, The impact of a major environmental crisis on species, populations and communities: the fragmentation of African forests at the end of the Holocene 345, 284–296. https://doi.org/10.1016/j.crte.2013.05.001

Hardy, O.J., Delaide, B., Hainaut, H., Gillet, J.-F., Gillet, P., Kaymak, E., Vankerckhove, N., Duminil, J., Doucet, J.-L., 2019. Seed and pollen dispersal distances in two African legume timber trees and their reproductive potential under selective logging. Mol. Ecol. 28, 3119–3134. https://doi.org/10.1111/mec.15138

He, T., Jiao, L., Yu, M., Guo, J., Jiang, X., Yin, Y., 2019. DNA barcoding authentication for the wood of eight endangered Dalbergia timber species using machine learning approaches. Holzforschung 73, 277–285. https://doi.org/10.1515/hf-2018-0076

Henry, R.J. (Ed.), 2005. Plant diversity and evolution: genotypic and phenotypic variation in higher plants, 1st ed. CABI Publishing, UK. https://doi.org/10.1079/9780851999043.0000

Hoare, A., & Uehara, T. H. K. (2022). Establishing fair and sustainable forest economies: Lessons learned from tackling illegal logging. Technical Report September, Royal Institute of International Affairs London. URL: https://chathamhouse.soutron.net/Portal/Public/en-GB/RecordView/Index/191315. doi:doi: 10.55317/9781784135386.

Hoare, A., 2015. Tackling Illegal Logging and the Related Trade - What Progress and Where Next?

Hoare, A., Young, D., Uehara, T., Seidu, K., Birikorang, G., Soh, L. W., & Kamga, J. K. (2020). Forest sector accountability in Cameroon and Ghana Exploring the impact of transparency and participation Energy, Environment and Resources Programme. Technical Report, .

Höltken, A.M., Schröder, H., Wischnewski, N., Degen, B., Magel, E., Fladung, M., 2012. Development of DNA-based methods to identify CITESprotected timber species: A case study in the Meliaceae family. Holzforschung 66, 97–104. https://doi.org/10.1515/HF.2011.142

Honorio Coronado, E.N., Blanc-Jolivet, C., Mader, M., García-Dávila, C.R., Aldana Gomero, D., del Castillo Torres, D., Flores Llampazo, G., Hidalgo Pizango, G., Sebbenn, A.M., Meyer-Sand, B.R.V., Paredes-Villanueva, K., Tysklind, N., Troispoux, V., Massot, M., Carvalho, C., de Lima, H.C., Cardoso, D., Degen, B., 2020. SNP Markers as a Successful

Molecular Tool for Assessing Species Identity and Geographic Origin of Trees in the Economically Important South American Legume Genus Dipteryx. J. Hered. 111, 346–356. https://doi.org/10.1093/jhered/esaa011

Hu, H., Al-Shehbaz, I.A., Sun, Y., Hao, G., Wang, Q., Liu, J., 2015. Species delimitation in Orychophragmus (Brassicaceae) based on chloroplast and nuclear DNA barcodes. TAXON 64, 714–726. https://doi.org/10.12705/644.4

Hu, H., Hu, Q., Al-Shehbaz, I.A., Luo, X., Zeng, T., Guo, X., Liu, J., 2016. Species Delimitation and Interspecific Relationships of the Genus Orychophragmus (Brassicaceae) Inferred from Whole Chloroplast Genomes. Frontiers in Plant Science 7.

Hung, K.H., Lin, C.H., Ju, L.P., 2017. Tracking the geographical origin of timber by DNA fingerprinting: A study of the endangered species Cinnamomum kanehirae in Taiwan. Holzforschung 71, 853–862. https://doi.org/10.1515/hf-2017-0026

Illegal Logging and Deforestation | Basic Page [WWW Document], 2022. . U.S. Agency for International Development. URL https://www.usaid.gov/biodiversity/illegal-logging-and-deforestation (accessed 1.8.24).

International Barcode of Life – Illuminate Biodiversity, n.d. URL https://ibol.org/ (accessed 4.14.23).

INTERPOL, 2020. Forestry crime: targeting the most lucrative of environmental crimes [WWW Document]. URL https://www.interpol.int/en/News-and-Events/News/2020/Forestry-crime-targeting-the-most-lucrative-of-environmental-crimes (accessed 2.22.24).

INTERPOL, 2023. Illegal wildlife trade has become one of the 'world's largest criminal activities' [WWW Document]. URL https://www.interpol.int/en/News-and-Events/News/2023/Illegal-wildlife-trade-has-become-one-of-the-world-s-largest-criminal-activities (accessed 2.23.24).

Iwanaga, H., Teshima, K.M., Khatab, I.A., Inomata, N., Finkeldey, R., Siregar, I.Z., Siregar, U.J., Szmidt, A.E., 2012. Population structure and demographic history of a tropical lowland rainforest tree species Shorea parvifolia (Dipterocarpaceae) from Southeastern Asia. Ecology and Evolution 2, 1663–1675. https://doi.org/10.1002/ece3.284

Jiao, L., Lu, Y., He, T., Li, J., Yin, Y., 2019. A strategy for developing high-resolution DNA barcodes for species discrimination of wood specimens using the complete chloroplast genome of three Pterocarpus species. Planta 250, 95–104. https://doi.org/10.1007/s00425-019-03150-1

Jolivet, C., & Degen, B. (2012). Use of DNA fingerprints to control the origin of sapelli timber (Entandrophragma cylindricum) at the forest concession level in Cameroon. Forensic Science International: Genetics, 6 , 487–493. doi:doi: 10.1016/j.fsigen.2011.11.002.

Jombart, T., 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24, 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Kelchner, S.A., 2000. The Evolution of Non-Coding Chloroplast DNA and Its Application in Plant Systematics, Source: Annals of the Missouri Botanical Garden.

Kjær, K.H., Winther Pedersen, M., De Sanctis, B., De Cahsan, B., Korneliussen, T.S., Michelsen, C.S., Sand, K.K., Jelavić, S., Ruter, A.H., Schmidt, A.M.A., Kjeldsen, K.K., Tesakov, A.S., Snowball, I., Gosse, J.C., Alsos, I.G., Wang, Y., Dockter, C., Rasmussen, M., Jørgensen, M.E., Skadhauge, B., Prohaska, A., Kristensen, J.Å., Bjerager, M., Allentoft, M.E., Coissac, E., Rouillard, A., Simakova, A., Fernandez-Guerra, A., Bowler, C., Macias-Fauria, M., Vinner, L., Welch, J.J., Hidy, A.J., Sikora, M., Collins, M.J., Durbin, R., Larsen, N.K., Willerslev, E., 2022. A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA. Nature 612, 283–291. https://doi.org/10.1038/s41586-022-05453-y

Kleinschmit, D., Mansourian, S., Wildburger, C., Purret, A. (Eds.), 2016. llegal Logging and Related Timber Trade – Dimensions, Drivers, Impacts and Responses. A Global Scientific Rapid Response Assessment Report., IUFRO World Series. International Union of Forest Research Organizations (IUFRO), Vienna.

Koné, I., Lambert, J.E., Refisch, J., Bakayoko, A., 2008. Primate Seed Dispersal and its Potential Role in Maintaining Useful Tree Species in the Taï Region, Côte-d'Ivoire: Implications for the Conservation of Forest Fragments. Trop. Conserv. Sci. 1, 293–305. https://doi.org/10.1177/194008290800100309

Koopman, W.J.M., Kuiper, I., Klein-Geltink, D.J.A., Sabatino, G.J.H., Smulders, M.J.M., 2012. Botanical DNA evidence in criminal cases: Knotgrass (Polygonum aviculare L.) as a model species. Forensic Science International: Genetics 6, 366–374. https://doi.org/10.1016/j.fsigen.2011.07.013

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923

Lehmann, M. M., Schuler, P., Cormier, M.-A., Allen, S. T., Leuenberger, M., & Voelker, S. (2022). The Stable Hydrogen Isotopic Signature: From Source Water to Tree Rings. (pp. 331–359). doi:doi: 10.1007/978-3-030-92698-4 11.

Li, X., Yang, Y., Henry, R.J., Rossetto, M., Wang, Y., Chen, S., 2015. Plant DNA

barcoding: from gene to genome. Biological Reviews 90, 157–166. https://doi.org/10.1111/brv.12104

Lompo, D., Vinceti, B., Konrad, H., Gaisberger, H., Geburek, T., 2018. Phylogeography of African locust bean (Parkia biglobosa) reveals genetic divergence and spatially structured populations in West and Central Africa. Journal of Heredity 109, 811–824. https://doi.org/10.1093/jhered/esy047

Low, M.C., Schmitz, N., Boeschoten, L.E., Cabezas, J.A., Cramm, M., Haag, V., Koch, G., Meyer-Sand, B.R.V., Paredes-Villanueva, K., Price, E., Thornhill, A.H., Brusselen, J.V., Zuidema, P.A., Deklerck, V., Dormontt, E.E., Shapcott, A., Lowe, A.J., 2022. Tracing the world's timber: the status of scientific verification technologies for species and origin identification. IAWA Journal 44, 63–84. https://doi.org/10.1163/22941932-bja10097

Lowe, A.J., Cross, H.B., 2011. The ApplicATion of DnA meThoDs To Timber TrAcking AnD origin VerificATion, IAWA Journal.

Lowe, A.J., Dormontt, E.E., Bowie, M.J., Degen, B., Gardner, S., Thomas, D., Clarke, C., Rimbawanto, A., Wiedenhoeft, A., Yin, Y., Sasaki, N., 2016. Opportunities for Improved Transparency in the Timber Trade through Scientific Verification. BioScience 66, 990–998. https://doi.org/10.1093/biosci/biw129

Lowe, A.J., Wong, K.N., Tiong, Y.S., Iyerh, S., Chew, F.T., 2010. A DNA method to verify the integrity of timber supply chains; Confirming the legal sourcing of merbau timber from logging concession to sawmill. Silvae Genetica 59, 263–268. https://doi.org/10.1515/sg-2010-0037

Mader, M., Pakull, B., Blanc-Jolivet, C., Paulini-Drewes, M., Bouda, Z.H.N., Degen, B., Small, I., Kersten, B., 2018. Complete chloroplast genome sequences of four meliaceae species and comparative analyses. International Journal of Molecular Sciences 19. https://doi.org/10.3390/ijms19030701

Maley, J., 1996. The African rain forest - Main characteristics of changes in vegetation and climate from the Upper Cretaceous to the Quaternary. Proc. R. Soc. Edinb. Sect. B Biol. Sci. 104, 31–73. https://doi.org/10.1017/S0269727000006114

Manel, S., Schwartz, M. K., Luikart, G., & Taberlet, P. (2003). Landscape genetics: Combining landscape ecology and population genetics. doi:doi: 10.1016/S0169-5347(03)00008-9.

Mascarello, M., Amalfi, M., Asselman, P., Smets, E., Hardy, O.J., Beeckman, H., Janssens, S.B., 2021. Genome skimming reveals novel plastid markers for the molecular

identification of illegally logged African timber species. PLoS ONE 16. https://doi.org/10.1371/journal.pone.0251655

Mascarello, M., Lachenaud, O., Amalfi, M., Smets, E., Hardy, O.J., Beeckman, H., Janssens, S.B., 2023. Genetic characterization of a group of commercial African timber species: From genomics to barcoding. PLoS ONE 18, e0284732. https://doi.org/10.1371/journal.pone.0284732

Matiz-Ceron, L., Reyes, A., Anzola, J., 2022. Taxonomical Evaluation of Plant Chloroplastic Markers by Bayesian Classifier. Front. Plant Sci. 12, 782663. https://doi.org/10.3389/fpls.2021.782663

May, C. (2017). Transnational crime and the developing world.

Meyer-Sand, B.R.V., Blanc-Jolivet, C., Mader, M., Paredes-Villanueva, K., Tysklind, N., Sebbenn, A.M., Guichoux, E., Degen, B., 2018. Development of a set of SNP markers for population genetics studies of Ipe (Handroanthus sp.), a valuable tree genus from Latin America. Conservation Genetics Resources 10, 779–781. https://doi.org/10.1007/s12686-017-0928-5

Migliore, J., Kaymak, E., Mariac, C., Couvreur, T.L.P., Lissambou, B.J., Piñeiro, R., Hardy, O.J., 2019. Pre-Pleistocene origin of phylogeographical breaks in African rain forest trees: New insights from Greenwayodendron (Annonaceae) phylogenomics. Journal of Biogeography 46, 212–223. https://doi.org/10.1111/jbi.13476

Morin, P.A., Mccarthy, M., 2007. Highly accurate SNP genotyping from historical and low-quality samples. Molecular Ecology Notes 7, 937–946. https://doi.org/10.1111/j.1471-8286.2007.01804.x

Nature, P. by, n.d. Illegal logging [WWW Document]. NEPCon - Preferred by Nature. URL https://www.preferredbynature.org/sourcinghub/info/illegal-logging-0 (accessed 1.8.24).

NCBI, 2024. National Center for Biotechnology Information- NCBI [WWW Document]. URL https://www.ncbi.nlm.nih.gov/ (accessed 2.29.24).

Ndiade-Bourobou, D., Daïnou, K., Hardy, O.J., Doumenge, C., Tosso, F., Bouvet, J.-M., 2020. Revisiting the North-South genetic discontinuity in Central African tree populations: the case of the low-density tree species Baillonella toxisperma. Tree Genet. Genomes 16, 15. https://doi.org/10.1007/s11295-019-1408-8

Nei, M., 1972. Genetic Distance between Populations. Am. Nat. 106, 283–292.

Nellemann, C., INTERPOL Environmental Crime Programme, 2012. Green carbon, black trade : Illegal logging, tax fraud and laundering in the world's tropicAbdel-Latif, A., Osman, G., 2017. Comparison of three genomic DNA extraction methods to obtain high DNA quality from maize. Plant Methods 13, 1. https://doi.org/10.1186/s13007-016-0152-4

Ng, C.H., Lee, S.L., Tnah, L.H., Siong, K.K., Lee, C.T., Diway, B., Khoo, E., 2017. Geographic origin and individual assignment of Shorea platyclados (Dipterocarpaceae) for forensic identification. PLoS ONE 12. https://doi.org/10.1371/journal.pone.0176158

Ng, C.H., Ng, K.K.S., Lee, S.L., Zakaria, N.-F., Lee, C.T., Tnah, L.H., 2022. DNA databases of an important tropical timber tree species Shorea leprosula (Dipterocarpaceae) for forensic timber identification. Sci Rep 12, 9546. https://doi.org/10.1038/s41598-022-13697-x

Ng, K.K.S., Lee, S.L., Tnah, L.H., Nurul-Farhanah, Z., Ng, C.H., Lee, C.T., Tani, N., Diway, B., Lai, P.S., Khoo, E., 2016. Forensic timber identification: A case study of a CITES listed species, Gonystylus bancanus (Thymelaeaceae). Forensic Science International: Genetics 23, 197–209. https://doi.org/10.1016/j.fsigen.2016.05.002

Nguyen, N.-L., Devendra, D., Szymańska, N., Greco, M., Angeles, I.B., Weiner, A.K.M., Ray, J.L., Cordier, T., De Schepper, S., Pawłowski, J., Pawłowska, J., 2023. Sedimentary ancient DNA: a new paleogenomic tool for reconstructing the history of marine ecosystems. Frontiers in Marine Science 10.

Nithaniyal, S., Newmaster, S.G., Ragupathy, S., Krishnamoorthy, D., Vassou, S.L., Parani, M., 2014. DNA Barcode Authentication of Wood Samples of Threatened and Commercial Timber Trees within the Tropical Dry Evergreen Forest of India. PLOS ONE 9, e107669. https://doi.org/10.1371/journal.pone.0107669

Nock, C.J., Waters, D.L.E., Edwards, M.A., Bowen, S.G., Rice, N., Cordeiro, G.M., Henry, R.J., 2011. Chloroplast genome sequences from total DNA for plant identification. Plant Biotechnology Journal 9, 328–333. https://doi.org/10.1111/j.1467-7652.2010.00558.x

Novaes, R.M.L., Rodrigues, J.G., Lovato, M.B., 2009. An efficient protocol for tissue sampling and DNA isolation from the stem bark of Leguminosae trees, Genetics and Molecular Research.

Ogden, R., Linacre, A., 2015. Wildlife forensic science: A review of genetic geographic origin assignment. Forensic Sci. Int. Genet. 18, 152–159. https://doi.org/10.1016/j.fsigen.2015.02.008

Ogden, R., McGough, H.N., Cowan, R.S., Chua, L., Groves, M., McEwing, R., 2009.

SNP-based method for the genetic identification of ramin Gonystylus spp. Timber and products: Applied research meeting CITES enforcement needs. Endangered Species Research 9, 255–261. https://doi.org/10.3354/esr00141

Oksanen, J. et al. (2020). vegan: Community Ecology Package. R package version 2.5-7.

Ouédraogo, D.-Y., Hardy, O.J., Doucet, J.-L., Janssens, S.B., Wieringa, J.J., Stoffelen, P., Angoboy Ilondea, B., Baya, F., Beeckman, H., Daïnou, K., Dubiez, E., Gourlet-Fleury, S., Fayolle, A., 2020. Latitudinal shift in the timing of flowering of tree species across tropical Africa: insights from field observations and herbarium collections. J. Trop. Ecol. 36, 159–173. https://doi.org/10.1017/S0266467420000103

Paradis, E., 2010. pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics 26, 419–420. https://doi.org/10.1093/bioinformatics/btp696

Paradis, E., Schliep, K., 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35, 526–528. https://doi.org/10.1093/bioinformatics/bty633

Paredes-Villanueva, K., Boom, A., Ottenburghs, J., Van Der Sleen, P., Manzanedo, R. D., Bongers, F., & Zuidema, P. A. (2022). Isotopic Characterization of Cedrela to Verify Species and Regional Provenance of Bolivian Timber. Tree-Ring Research, 78 , 73–89. doi:doi:10.3959/2021-17.

Paredes-Villanueva, K., de Groot, G. A., Laros, I., Bovenschen, J., Bongers, F., & Zuidema, P. A. (2019). Genetic differences among Cedrela odorata sites in Bolivia provide limited potential for fine-scale timber tracing. Tree Genetics and Genomes, 15 . doi:doi:10.1007/s11295-019-1339-4.

Piñeiro, R., Staquet, A., Hardy, O.J., 2015. Isolation of nuclear microsatellites in the African timber tree Lophira alata (Ochnaceae) and cross-amplification in L. lanceolata. Applications in Plant Sciences 3, 1500056. https://doi.org/10.3732/apps.1500056

Plouvier, D., 1997. Short Overview of the Situation of Tropical Moist Forests and Forest Management in Central Africa and Markets for African Timber. Journal on Tropical Forests and Forestry for Sustainable Development, BOS Nieuwletter 16, 42–49.

Porebski, S., Bailey, L.G., Baum, B.R., n.d. Modification of a CTAB DNA Extraction Protocol for Plants Containing High Polysaccharide and Polyphenol Components, Plant Molecular Biology Reporter.

Poulsen, J.R., Clark, C.J., Smith, T.B., 2001. Seed Dispersal by a Diurnal Primate Community in the Dja Reserve, Cameroon. Journal of Tropical Ecology 17, 787–808.

Rachmayanti, Y., Leinemann, L., Gailing, O., Finkeldey, R., 2006. Extraction, Amplification and Characterization of Wood DNA from Dipterocarpaceae.

Rachmayanti, Y., Leinemann, L., Gailing, O., Finkeldey, R., 2009. DNA from processed and unprocessed wood: Factors influencing the isolation success. Forensic Science International: Genetics 3, 185–192. https://doi.org/10.1016/j.fsigen.2009.01.002

Rossi, F., Crnjar, A., Comitani, F., Feliciano, R., Jahn, L., Malim, G., Southgate, L., Kay, E., Oakey, R., Buggs, R., Moir, A., Kistler, L., Mateos, A.R., Molteni, C., Schulz, R., 2021. Extraction and high-throughput sequencing of oak heartwood DNA: Assessing the feasibility of genome-wide DNA methylation profiling. PLOS ONE 16, e0254971. https://doi.org/10.1371/journal.pone.0254971

Schneider, J.V., Jungcurt, T., Cardoso, D., Amorim, A.M., Töpel, M., Andermann, T., Poncy, O., Berberich, T., Zizka, G., 2021. Phylogenomics of the tropical plant family Ochnaceae using targeted enrichment of nuclear genes and 250+ taxa. TAXON 70, 48–71. https://doi.org/10.1002/tax.12421

Schroeder, H., Cronn, R., Yanbaev, Y., Jennings, T., Mader, M., Degen, B., Kersten, B., 2016. Development of molecular markers for determining continental origin of wood from White Oaks (Quercus L. sect. Quercus). PLoS ONE 11. https://doi.org/10.1371/journal.pone.0158221

Segers, A., 2018. Détermination des visiteurs floraux de trois essences ligneuses à haute valeur commerciale : Baillonella Toxisperma Pierre, Afzelia Bipindensis Harms et Erythrophleum Suaveolens (Guill. & Perr.) Brenan. Université de Liège.

Shi, C., Hu, N., Huang, H., Gao, J., Zhao, Y.-J., Gao, L.-Z., 2012. An Improved Chloroplast DNA Extraction Procedure for Whole Plastid Genome Sequencing. PLoS One 7, e31468. https://doi.org/10.1371/journal.pone.0031468

Shi, M.-M., Chen, X.-Y., 2012. Leading-edge populations do not show low genetic diversity or high differentiation in a wind-pollinated tree. Popul. Ecol. 54, 591–600. https://doi.org/10.1007/s10144-012-0332-7

Singh, N., Choudhury, D.R., Singh, A.K., Kumar, S., Srinivasan, K., Tyagi, R.K., Singh, N.K., Singh, R., 2013. Comparison of SSR and SNP Markers in Estimation of Genetic Diversity and Population Structure of Indian Rice Varieties. PLoS ONE 8, e84136. https://doi.org/10.1371/journal.pone.0084136

Singhal, D., Gupta, P., Sharma, P., Kashyap, N., Anand, S., Sharma, H., 2011. In-silico single nucleotide polymorphisms (SNP) mining of Sorghum bicolour genome. Afr. J. Biotechnol. 10, 580–583.

Smulders, M.J.M., Van 'T Westende, W.P.C., Diway, B., Esselink, G.D., Van Der Meer, P.J., Koopman, W.J.M., 2008. Development of microsatellite markers in Gonystylus bancanus (Ramin) useful for tracing and tracking of wood of this protected species. Molecular Ecology Resources 8, 168–171. https://doi.org/10.1111/j.1471-8286.2007.01914.x

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690. https://doi.org/10.1093/bioinformatics/btl446

SWGDAM, 2012. Validation guidelines for DNA analysis methods [WWW Document]. URL https://media.wix.com/ugd/4344b0_cbc27d16dcb64fd88cb36ab2a2a25e4c.pdf (accessed 2.27.24).

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., Vermat, T., Corthier, G., Brochmann, C., Willerslev, E., 2007. Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acids Res. 35, e14–e14. https://doi.org/10.1093/nar/gkl938

Takoutsing, B., & Heuvelink, G. B. (2022). Comparing the prediction performance, uncertainty quantification and extrapolation potential of regression kriging and random forest while accounting for soil measurement errors. Geoderma, . doi:doi: 10.1016/J.GEODERMA.2022.116192.

Tello, D., Gil, J., Loaiza, C.D., Riascos, J.J., Cardozo, N., Duitama, J., 2019. NGSEP3: Accurate variant calling across species and sequencing protocols. Bioinformatics 35, 4716–4723. https://doi.org/10.1093/bioinformatics/btz275

The European Parliament And The Council Of The European Union, 2023. 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.

Tnah, L.H., Lee, S.L., Ng, K.K.S., Bhassu, S., Othman, R.Y., 2012a. DNA extraction from dry wood of Neobalanocarpus heimii (Dipterocarpaceae) for forensic DNA profiling and timber tracking. Wood Science and Technology 46, 813–825. https://doi.org/10.1007/s00226-011-0447-6

Tnah, L.H., Lee, S.L., Ng, K.K.S., Bhassu, S., Othman, R.Y., 2012b. DNA extraction from dry wood of Neobalanocarpus heimii (Dipterocarpaceae) for forensic DNA profiling and timber tracking. Wood Sci Technol 46, 813–825. https://doi.org/10.1007/s00226-011-0447-6

Tnah, L.H., Lee, S.L., Ng, K.K.S., Tani, N., Bhassu, S., Othman, R.Y., 2009. Geographical traceability of an important tropical timber (Neobalanocarpus heimii) inferred from chloroplast DNA. Forest Ecology and Management 258, 1918–1923. https://doi.org/10.1016/j.foreco.2009.07.029

Truszkowski, J., Perrigo, A., Broman, D., Ronquist, F., Antonelli, A., 2023. Online tree expansion could help solve the problem of scalability in Bayesian phylogenetics. Systematic Biology 72, 1199–1206. https://doi.org/10.1093/sysbio/syad045

Turubanova, S., Potapov, P. V., Tyukavina, A., & Hansen, M. C. (2018). Ongoing primary forest loss in Brazil, Democratic Republic of the Congo, and Indonesia. Environmental Research Letters, 13 , 074028. doi:doi: 10.1088/1748-9326/aacd1c.

van der Sleen, P., Zuidema, P. A., & Pons, T. L. (2017). Stable isotopes in tropical tree rings: theory, methods and applications. Functional Ecology, 31 , 1674–1689. doi:doi: 10.1111/1365-2435.12889.

Veronesi, F., & Schillaci, C. (2019). Comparison between geostatistical and machine learning models as predictors of topsoil organic carbon with a focus on local uncertainty estimation. Ecological Indicators, 101 , 1032–1044. https://doi.org/10.1016/j.ecolind.2019.02.026

Vlam, M., Groot, G.A. de, Boom, A., Copini, P., Laros, I., Veldhuijzen, K., Zakamdi, D., Zuidema, P.A., 2018. Developing forensic tools for an African timber: Regional origin is revealed by genetic characteristics, but not by isotopic signature. Biological Conservation 220, 262–271. https://doi.org/10.1016/j.biocon.2018.01.031

Wang, Z., Erasmus, S. W., Dekker, P., Guo, B., Stoorvogel, J. J., & van Ruth, S. M. (2020). Linking growing conditions to stable isotope ratios and elemental compositions of Costa Rican bananas (Musa spp.). Food Research International, 129 . doi:doi: 10.1016/j.foodres.2019.108882.

Watkinson, C. J., Gasson, P., Rees, G. O., & Boner, M. (2020). The Development and Use of Isoscapes to Determine the Geographical Origin of Quercus spp. in the United States. Forests, 11 , 862. doi:doi: 10.3390/f11080862.

Watkinson, C. J., Rees, G. O., Gwenael, M. C., Gasson, P., Hofem, S., Michely, L., & Boner, M. (2022a). Stable Isotope Ratio Analysis for the Comparison of Timber From Two

Forest Concessions in Gabon. Frontiers in Forests and Global Change, 4 , 1–16. doi:doi: 10.3389/ffgc.2021.650257.

Webster, R., & Oliver, M. A. (2007). Geostatistics for environmental scientists.. John Wiley & Sons.

West, J. B., Bowen, G. J., Dawson, T. E., & Tu, K. P. (2010). Isoscapes. Dordrecht: Springer Netherlands. doi:doi: 10.1007/978-90-481-3354-3.

White, F. (Frank), 1983. The vegetation of Africa: a descriptive memoir to accompany the Unesco/AETFAT/UNSO vegetation map of Africa, Natural resources research ; 20. Unesco, Paris.

Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. https://doi.org/10.1038/sdata.2016.18

World Forest ID, 2021. World Forest ID - WFID [WWW Document]. World Forest ID. URL https://worldforestid.org (accessed 2.28.24).

Wright, M.N., Ziegler, A., 2017. ranger : A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. J. Stat. Soft. 77. https://doi.org/10.18637/jss.v077.i01

Wright, M.N., Ziegler, A., 2017. ranger : A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. J. Stat. Soft. 77. https://doi.org/10.18637/jss.v077.i01

Wynn, P. M., Loader, N. J., & Fairchild, I. J. (2014). Interrogating trees for isotopic archives of atmospheric sulphur deposition and comparison to speleothem records. Environmental Pollution, 187 , 98–105. doi:doi: 10.1016/j.envpol.2013.12.017

Xu, B., Wu, N., Gao, X.-F., Zhang, L.-B., 2012. Analysis of DNA sequences of six chloroplast and nuclear genes suggests incongruence, introgression, and incomplete lineage sorting in the evolution of Lespedeza (Fabaceae). Molecular Phylogenetics and Evolution 62,

346–358. https://doi.org/10.1016/j.ympev.2011.10.007

Yang, J., Vázquez, L., Chen, X., Li, H., Zhang, H., Liu, Z., Zhao, G., 2017. Development of Chloroplast and Nuclear DNA Markers for Chinese Oaks (Quercus Subgenus Quercus) and Assessment of Their Utility as DNA Barcodes. Frontiers in Plant Science 8.

Zhang, W., Fan, X., Zhu, S., Zhao, H., Fu, L., 2013. Species-Specific Identification from Incomplete Sampling: Applying DNA Barcodes to Monitoring Invasive Solanum Plants. PLOS ONE 8, e55927. https://doi.org/10.1371/journal.pone.0055927

Summary

The illegal logging and trade of tropical timber significantly hampers sustainable forest management and conservation and undermines legal timber exploitation. The complexity of the timber trade routes and the frequent mixture of timber from different species or geographic origins into wooden products offer opportunities for illegal timber trade. Various international regulations aim to curb this illegal trade. Yet, these legislative initiatives have not achieved the intended goals due to a significant gap between their intentions and enforcement. In recent years, researchers across the globe are addressing this challenge, by developing and testing methods that can verify the provenance of timber based on independent wood characteristics.

The research presented in this thesis contributed to addressing several critical barriers in the development of genetic tracing techniques for tropical timbers: the need for comprehensive reference databases, effective genetic markers, and enhanced spatial resolution for tracing.

Through extensive fieldwork in Central Africa, over 700 tree samples were collected and used to build robust reference databases for two major African timbers: Azobé (*Lophira alata*) and Tali (*Erythrophleum suaveolens* and *E. ivorense*). Next, whole plastomes were assembled for SNP detection. The SNP sets were then used to assess the possible use for differentiation between species (Tali) and for providing insights into the spatial genetic structure of the two studied timbers (Chapters 2 and 3). The thesis delivers key insights into timber traceability, notably through the differentiation of Tali species using plastid genome polymorphisms (Chapter 2). Furthermore, it provides evidence for a high potential to verify timber taxonomic claims for blind samples, i.e., samples of unknown taxonomic or geographic origin. In addition, using plastome-wide SNPs, I was able to accurately assign 72% of Azobé and 60% of Tali blind samples within 100 km of their true origin (Chapter 4).

In Chapter 5, we assessed to what extent the combination of genetic and chemical methods can improve assignment accuracy for Azobé. The results showed a substantial increase in tracing accuracy by integrating plastome-wide SNP analysis with multi-element data and stable isotopic data. In self assignments, 94% of the samples were accurately assigned to a site within 100 km when combining methods. These findings not only advance the precision of timber traceability but also underscore These findings advance the precision of timber traceability and underscore how different tracing methods may be complementary.

This PhD thesis has contributed to the advancement of timber traceability, by developing and testing genetic methods, and comparing and combining those with chemical methods. Despite the difficulties in extracting DNA from heartwood, the study demonstrates that obtaining genomic DNA of adequate quality enables highly precise tracing at fine scales, whether utilizing plastid Single Nucleotide Polymorphisms (pSNPs) alone or coupled with other tracing tools.

The research presented in this thesis would not have been possible without extensive collaborative across researchers, institutions, countries and disciplines. Finally, in this study I bridged academic research with practical application, focused on solution-oriented research and collaborated with policymakers and stakeholders from the industry. Such a comprehensive, multi-disciplinary and collaborative strategy is key to develop effective and fair tools to trace tropical timber and halt the trade in illegal timber.

Acknowledgements

This PhD journey has been a tapestry woven with the threads of collaboration, love, and perseverance, each strand vibrant with the contributions of cherished colleagues, friends, and family. I have grown personally and professionally, caring in me all I have experienced from hours of sample collection in the forests to understanding what true equitable long-lasting collaboration should look like. I am a better researcher and person for it!

First and foremost, I extend my heartfelt gratitude to my amazing family, who have been my pillars of strength and sources of endless inspiration. Mom, Rosa Amelia Uguccioni da Rocha, thank you for pouring in me the virtues of love, kindness, and care, and for your unwavering support. Dad, Vagner Venancio, you taught me resilience and courage in the face of life's challenges. To my beloved grandparents, especially my grandmothers Benedita Nunes Pereira and Neusa Aparecida Uguccioni da Rocha, thank you for your sacrifices and for providing a haven of love and prayers. Bruno, Junior, and Mayara, my blood and life siblings, having you in my life enriches me beyond words - our bond reaches to the moon and back. Your support and love were/are crucial for me, my love. To my family - Mirtes, Nilce, Deise, Lilian, Rosani, Paula, Rosana, Giovani, Rogerio, and all my cousins and partners - your unwavering support and presence have been essential to me. Brian, (g)uncle Valdir, and tia Sueli, your love and backing have been invaluable. I also thank my family-in-law, Nannette, Norbert and Fianna for your support and believing in me. Lennart, thank you for simply being you - always curious, energetic, empathetic, and caring. I cherish how we've grown together over these 13 years, learning from each other, supporting one another, and deepening both our friendship and our love. Thank you for believing in me when I didn't, and hold me strong through the hard times of this PhD journey and beyond! This PhD is a family achievement! Minha amada família e queridos amigos, eu amo vocês imensamente e sou grata por ter o privilégio de tê-los comigo nesta jornada! ESTE DOUTORADO É NOSSO!

In my academic journey, I am profoundly grateful for the guidance and support of my supervisors: Pieter, Rene, and Arjen. Your humanistic approach and willingness to embrace my uniqueness have profoundly shaped my professional path. To the Timtrace teams in the Netherlands, Cameroon, Gabon, Congo, Indonesia, and beyond—Ute, Mart, Prof. Tchamba, Herman, Pascaline, Steve, Dyana, Cynel, Nestor, Nathalie, Gael, Jannici, Mesly, Dieu-merci, Rita, Bijoux, Pak Iskandar, Fifi, Ulfa, Andrea, Majiidu, Miftah, Olivier, and Simon—I am

profoundly grateful for the incredible team I've had the fortune to work alongside. Your collaboration, ease of interaction, and support have greatly enriched both the Timtrace project and my personal and professional growth. Our discussions and exchanges have not only inspired and taught me so much, but have also laid the foundation for enduring relationships and collaborations. Thank you for your invaluable contributions and for making this journey truly amazing. I am deeply thankful to the forest companies and their collaborators who supported and facilitated our extensive sample collection efforts across several countries. Pallisco-CIFM, SFID, and Rougier Gabon, CUF, SEFAC Group, COBACAM community forest, RECODEMM Community Forest, Jua-Ikie, Ipendja, Tala-Tala, Loundoungou, Kabo, Missa, Mokabi Dzanga, Forêt de Mambili, Zone militaire (FCM), CEB, GSEZ, CBG, Wijma, amongst others. A special thanks to Pamphile, David Zakandi, Markus Pfannkuch, and Jules Pettang for their indispensable contributions. Your commitment to transparency, sustainable development, and forest research was pivotal this past 5 years. My gratitude extends to our dedicated field guides, botanists, drivers, and the local community chiefs, and communities who welcomed and guided us in their areas. I couldn't have asked for a better group of people to work with and develop friendships with. Thank you for embracing the project as your own, for the shared laughter, meals and the best (somewhat warm) beers after the long days in the forest. Your ongoing support through kind messages has been heartwarming. Thank you all for making this project (my PhD and Timtrace) possible!

Now, to my chosen family and cherished friends in the Netherlands! Ricardo, my dear friend, I miss you dearly and always will. Tony, where do I begin? Thank you for all that you are and for sharing yourself with me, for providing shelter while encouraging me to be the most authentic version of Bárbara possible - I love you. Laurinha, what a journey we've had, my friend! Thank you for embracing my noisy and intense self. Starting as colleagues, we navigated countless hardships and adventures, both in the jungle and beyond. Through long talks, abundant love, immense patience, and deep respect, I've gained a friend for life. I cherish and respect you immensely, and I'm so proud of both you and our friendship. Rozinho, dear friend, thank you for the deep random conversations, the personal growth, the unwavering support, and those late-night cartoon marathons - I love you! To my dear Ayşegül, Luis, Maira, and Nathalia, thank you for your loyal support, especially towards the end of my PhD. Your presence ensured that the finish line was within reach, closer than it

might have been without you. To my dear friends Daisy, Gabriela, Amanda, Adriana, Bruna, Stephan, Nellie, Vitor, Mireille, Sophie, Linda, Chenglong, Kassia, Ron, and Selcuk - I am incredibly blessed to have you all in my life. Thanks for your love, sharing your lives with me, and being such an important safety net, I cherish you all deeply. Thanks to the "MYBCT" band (Mithyzi, Yasmaine, -myself-, Chiara, and Thais) and associates Maira, Lis, and others for the fun-filled musical and foodie nights. You taught me one of the most vital lessons during this PhD journey: how to be "taken care of." Your friendships have been my sanctuary, helping me navigate through intense periods with love and patience. To my Droevendaal 97 friends—Tina, Bilal, Teo, Cata, Laura, and many others—thank you for making Droevendaal my second home during COVID times and lessening the loneliness of isolation. Finally, immense gratitude to my dear therapists, Luma and Mona, who have guided me through the highs and lows in my personal and professional life. Thanks to your help, I've come to understand myself better, to recognize and confidently claim my space in this world. My FEMily, including some GIS folks who became adopted members, has grown into more than just an academic family; you are a support network I can always rely on—Arjen, Louis, Tomo, Iris, Jorad, Mathieu, Susanna, Meike, Xiaohan, Ambra, Francisca, Johan, Heitor, Nicola, Marco, Jente, Jazz, Bo, Alejandra, Ursula, Shanshan, Sylvana, Carlinhos, Aldicir, Federico, Jose, Alan, Yanjun, Paulinha, Sietse, Rens, Marina, Louise, Joke, Eva, Masha, Cata, Caro, amongst many others. Thank you to all FEM professors and colleagues for the enriching academic discussions and chats during coffee breaks. Your patience with my weekly critiques of Dutch cuisine has been amusing and restoring. I am FEMily for life!

This PhD journey began long before I set foot in the Netherlands. Along my path, I've had the privilege of meeting incredible individuals from across the globe. To Luana, David, Guilherme, Guilherme Bloss, David, Butz, Michael, Maude, Audrey, Joyce, Jemisson, Stefanie, Iris, Julia Bastin, Gereon, Valerie, Valery, Eduardo, Charlotte, Wibke, Remi, Liria, Malin, Viki, Patricia, Hans Hoenika, and many others - thank you for being part of my life and making this journey truly remarkable. Your friendships and insights have been invaluable contributions to both my personal and professional growth.

To all, my deepest thanks for weaving this rich tapestry with me. This PhD is not just my achievement; it is ours.

About the author

Bárbara Rocha Venancio Meyer-Sand was born in Ourinhos, Brazil, on 26 April 1991. She moved to Curitiba at the age of 6 where she grew up. She developed a love for nature and its beauties from a young age, growing up in a small farm and often going to her grandparents' farm in Guaratuba, in the Southern Atlantic Forest of Brazil. Bárbara was often found in the sea, forest or mangrove and playing with her cousins, which deepened her relationship with nature and people. After a BSc diploma in Industrial Engineering at the Federal University of Parana, with focus on the timber industry, she started her career as an industrial engineer in the timber industry.

Bárbara holds two MSc degrees that show her passion for forests, timber (she loves it!) and technology. The first is in Timber Engineering, specializing in International Timber Trade, from the École Supérieure du Bois in France. The second MSc is in Agronomy from São Paulo State University in Brazil. During her studies in France, she shifted her focus from trading timber to enhancing the sustainability of its trade. This change sparked her interest in conservation genetics. Since 2013, Bárbara has worked in the development of forensic methods aimed at curbing illegal timber trade and promoting trade in legally and sustainably sourced timber.

Before starting as a PhD candidate at Wageningen University in 2019, Bárbara helped developing timber tracing tools in the Brazilian Amazon and supporting timber legality on the Cameroonian domestic market. At Wageningen University, she continued developing timber tracing methods within the Timtrace project. She did so using the genetic information from chloroplast and also combined genetic and chemical methods. Her work aims to bridge the gap between research and practitioners, emphasizing the development of sustainable forest management and conservation strategies. Looking ahead, Bárbara is eager to further

connect society, science, policymakers, and practitioners in forest research. She also aims to facilitate and stimulate a more integrated and impactful approach of conducting science.

Bárbara's dedication extends beyond professional achievements; she is driven by a profound belief in the transformative power of knowledge to alter people's lives and the course of nature. She is active and vocal about issues of equity, inclusion, and diversity, and is committed to advancing a decolonized and discrimination-free science. Her passion for nature and people moves her to continuously efforts towards an equitable and sustainable future.

List of publications

Boeschoten LE, VlamM, Sass-Klaassen U, **Meyer-Sand BRV**, Adzkia U, Bouka GDU, Ciliane-Madikou JCU, Engone Obiang NL, Guieshon-Engongoro M, Loumeto JJ, Mbika DMMF, Moundounga CG, Ndangani RMD, Bourobou DN, Rahman MM, Siregar IZ, Tassiamba SN, Tchamba MT, Toumba-Paka BBL, Zanguim HT, Zemtsa PT and Zuidema PA. 2023. A new method for the timber tracing toolbox: applying multi-element analysis to determine wood origin. Environment Research Letters.

Boeschoten LE, Sass-Klaassen U, Vlam M,. Comans RNJ, Koopmans GF, **Meyer-Sand BRV**, Tassiamba SN, Tchamba M T, Zanguim HT, Zemtsa PT, Zuidema PA. 2022. Clay and soil organic matter drive wood multi-elemental composition of a tropical tree species: Implications for timber tracing. Science of the Total Environment 849:157877.

Low, M, Boeschoten LE, Cabezas JA, Cramm M, Deklerck V, Dormontt EE, Haag V, Koch G, **Meyer-Sand BRV**, Paredes-Villanueva K, Price E, Schmitz N, Shapcott A, Thornhill AH, Van Brusselen J, Zuidema PA, Lowe AJ. 2022. Tracing the world's timber: The status of scientific verification technologies for species and origin identification. IAWA Journal.

Honorio Coronado EN, Blanc-Jolivet C, Mader M, Garcia-Davila CR, Gomero DA, Castillo Torres D, Llampazo GF, Pizango GH, Sebbenn AM, **Meyer-Sand BRV**, Paredes-Villanueva K, Tysklind N, Troispoux V, Massot M, Carvalho C, de Lima HC, Cardoso D, Degen B (2020) Species Identity and Geographic Origin of Trees in the Economically Important South American Legume Genus *Dipteryx*. Journal of Heredity, esaa011.

Tysklind N, Blanc-Jolivet C, Mader M, **Meyer-Sand BRV**, Paredes-Villanueva K, Honorio Coronado EN, Garcia-Davila CR, Sebbenn AM, Caron H, Troispoux V, Guichoux E, Degen B (2019) Development of nuclear and plastid SNP and INDEL markers for population genetic studies and timber traceability of *Carapa* species. Conserv Genet Resources 11(3):337-339.

Honorio Coronado EN, Blanc-Jolivet C, Mader M, Garcia-Davila CR, Sebbenn AM, **Meyer-Sand BRV**, Paredes-Villanueva K, Tysklind N, Troispoux V, Massot M, Degen B (2019)

Development of nuclear and plastid SNP markers for genetic studies of *Dipteryx* tree species in Amazonia. Conserv Genet Resources 11(3):333-336.

Chaves CL, Blanc-Jolivet C, Sebbenn AM, Mader M, **Meyer-Sand BRV**, Paredes-Villanueva K, Honorio Coronado EN, Garcia-Davila CR, Tysklind N, Troispoux V, Massot M, Degen B (2019) Nuclear and chloroplastic SNP markers for genetic studies of timber origin for *Hymenaea* trees. Conserv Genet Resources 11(3):329-331.

Sebbenn AM, Blanc-Jolivet C, Mader M, **Meyer-Sand BRV**, Paredes-Villanueva K, Honorio Coronado EN, Garcia-Davila CR, Tysklind N, Troispoux V, Delcamp A, Degen B (2019) Nuclear and plastidial SNP and INDEL markers for genetic tracking studies of *Jacaranda copaia*. Conserv Genet Resources 11(3):341-343.

Meyer-Sand BRV, Blanc-Jolivet C, Mader M, Paredes-Villanueva K, Tysklind N, Sebbenn AM, Guichoux E, Degen B (2018) Development of a set of SNP markers for population genetics studies of Ipe (*Handroanthus* sp.), a valuable tree genus from Latin America. Conserv Genet Resources 10(4):779-781.

dos Reis AR, Lisboa LAM, Reis HPG, Barcelos JPD, Santos EF, Santini JMK, **Meyer-Sand BRV**, …, Lavres J (2018) Depicting the physiological and ultrastructural responses of soybean plants to Al stress conditions. Plant Physiology and Biochemistry, 130, 377–390. https://doi.org/10.1016/j.plaphy.2018.07.028

PE&RC Training and Education Statement

With the training and education activities listed below the PhD candidate has complied with the requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 32 ECTS (= 22 weeks of activities)

Review/project proposal (4.5 ECTS)

- Genetic methods applied to tracing of tropical timber species

Post-graduate courses (5.7 ECTS)

- Introduction to R for statistical analysis; PE&RC (2019)
- Bayesian statistics; PE&RC and WIMEK (2020)
- Multivariate analysis; PE&RC and WIMEK (2021)
- Workshop on population and speciation genomics, Cesky Krumlov; WPSG (2021)

Laboratory training and working visits (3.7 ECTS)

- MinIon training; IPB, Bogor, Indonesia (2020)
- Better safe than sorry awareness training on biological safety; Plant Sciences Group, WUR (2021)
- Stay labwork training; IPB, Bogor, Indonesia (2022)

Invited review of journal manuscripts (2 ECTS)

- Forests: forensic timber tracing (2020)
- Conservation genetics resources: SNPs marker development for timber tracing (2020)

Competence, skills and career-oriented activities (4.15 ECTS)

- Scientific writing; Wageningen in'to Languages (2022)
- Working on your PhD research in times of crisis; PE&RC (2020)
- Brain training; PE&RC (2019)

Scientific integrity/ethics in science activities (0.3 ECTS)

- Ethics in plant and environmental sciences; Wageningen Graduate Schools (2020)

PE&RC Annual meetings, seminars and PE&RC weekend/retreat (2.1 ECTS)

- PE&RC Weekend for first years (2019)
- PE&RC Symposium (2022)
- PE&RC Last year retreat (2023)

Discussion groups/local seminars or scientific meetings (6.9 ECTS)

- Final conference of the large scale project on timber verification (2019)
- PE&RC/FEM-PEN-WEC R discussion group (2019-2022)
- FEM Journal club (2019-2022)
- NWO Users committee (2019-2023)
- Global timber tracing network meeting (2020)
- Research seminars at IPB and CIFOR Indonesia (2020)
- Presentation on 2emes Journnees d'Etudes en Sciences et Technologies du Bois au Gabon (2021)
- Presentation at Universidade Federal de Vicosa (UFV) Brazil (2022)
- Atelier sur la tracabilité des bois tropicaux Gabon (2023)

International symposia, workshops and conferences (4.7 ECTS)

- Annual association for tropical biology and conservation meeting; Coimbatore, India (2023)
- British ecological society annual meeting; Belfast, Ireland (2023)

Societally relevant exposure (1.8 ECTS)

- Timtrace user committee meeting NOW (2019-2023)
- Knowledge clips and articles for Timtrace website (2020)
- NWO Photo story (2020)
- Presentation on tracing methods at the Dutch customs laboratory (2023)

Lecturing/supervision of practicals/tutorials (1.5 ECTS)

- Lecture on forensic techniques applied to timber tracing in Cameroon at Université de Dschang (2019)
- Lecture on forensic techniques applied to timber tracing at IPB (2019)
- Online training on sampling techniques at Université Marien NGOUAB (2020)
- Lecture on timber tracing forensic techniques, with focus on genetic methods at Utrecht University (2023)

BSc/MSc thesis supervision (6 ECTS)

- Characterization of the domestic timber market and the legal timber supply chain: case of the Mfoundi department Central Cameroon
- Study of the structural characteristics of tree species in Cameroon: case of Tali (*Erythrophleum spp*.) and Azobé (*Lophira alata Banks ex P.Gaertn*.) in three UFA of the Cameroon United Forest
- Extracting DNA from cambium samples of Azobé and Tali trees from Cameroon
- Genetic cluster analysis of *Erythrophleum spp.* from West and Central-Africa

The research described in this thesis was carried out at the Forest Ecology and Forest Management Group of Wageningen University and was financially supported by the Dutch Research Council (NWO-TTW-OTP-16427). Additional fieldwork support was received from the Alberta Mennega Foundation, FSC International and WorldForestID. the Wageningen University Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover design by Tarcio Vasconcelos

Cover art made by Tarcio Vasconcelos