TROPICAL TIMBER FORENSICS:
A MULTI-METHODS APPROACH TO TRACING BOLIVIAN *CEDRELA*

KATHELYN PAREDES VILLANUEVA

Propositions

1. To get high identification accuracies, unique fingerprints require unique approaches.
   (this thesis)
2. Signals of species and origin are there, methods to identify them not.
   (this thesis)
3. To arrive at sound ecological conclusions, field experience is more important than statistical significance.
4. With the timber tracing techniques, one should prioritize scientific knowledge over money.
5. Finding the variables that explain patterns in your data is like walking through a tropical forest. You always get distracted by random things.
6. Supervisors are like lianas, they swing between criticism and encouragement.
7. It is easier to classify people than plant species.
8. Being vegan is good for health and environment but bad for social life.

Propositions belonging to the thesis, entitled

"Tropical timber forensics: A multi-methods approach to tracing Bolivian *Cedrela*"

Kathelyn Paredes Villanueva
Wageningen, 22 November 2018

Tropical timber forensics:

A multi-methods approach to tracing Bolivian *Cedrela*

Kathelyn Paredes Villanueva

Thesis committee

Promotors

Prof. Dr P.A. Zuidema
Personal chair at the Forest Ecology and Forest Management Group
Wageningen University & Research

Prof. Dr F.J.J.M. Bongers
Personal chair at the Forest Ecology and Forest Management Group
Wageningen University & Research

Other members

Dr H. Beeckman, Royal Museum for Central Africa, Tervuren, Belgium
Prof. Dr R.G.A. Boot, Utrecht University
Dr L.W. Chatrou, Wageningen University & Research
Prof. Dr W. Peters, Wageningen University & Research

This research was conducted under the auspices of the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC).

Tropical timber forensics:

A multi-methods approach to tracing Bolivian *Cedrela*

Kathelyn Paredes Villanueva

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Thursday 22 November 2018
at 11 a.m. in the Aula.

Kathelyn Paredes Villanueva
Tropical timber forensics: A multi-methods approach to tracing Bolivian *Cedrela*, 182 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2018)
With references, with summaries in English, Dutch & Spanish

ISBN: 978-94-6343-522-2
DOI: https://doi.org/10.18174/461827

Para mi Bolivia

Table of Contents

From a small seed a mighty trunk may grow.

-Aeschylus

Chapter 1

General Introduction

Illegal logging, counteracting measures and alternative methods for its identification

Deforestation and forest degradation through land use change, development of infrastructure and illegal logging account for almost 20% of global greenhouse gas emissions - more than the entire global transport sector and second only to the energy sector (FAO, 2014; IPCC, 2014). Illegal logging is one of the drivers of forests loss resulting in loss of species, economic problems and a massive contribution to global warming (IUFRO, 2016). Almost half of all the tropical timber traded internationally is thought to be illegal (Hoare, 2015). To fight these illegal practices countries are implementing mechanisms such as the climate mitigation policy of reducing emissions from deforestation and forest degradation (REDD+) through enhanced forest management. In addition, specific attempts to address illegal logging were established driven by wood consumer countries such as the European Union Timber Regulation (EUTR), the European Union's Forest Law Enforcement, Governance and Trade (FLEGT) action plan and its Voluntary Partnership Agreement (VPA), U.S. Lacey Act and Australian Illegal Logging Prohibition Act. These legislations prohibit the importation of timber products that have been harvested or traded in violation of applicable foreign laws including those regulating the trade of endangered species (Convention on International Trade in Endangered Species of Wild Fauna and Flora, CITES). These measures require due diligence and evidence that the timber has not been illegally sourced. However, control of illegal timber trade remains a challenge.

Most legislative measures focus on combating international illegal trade although a high proportion (70-90%) of illegal tropical timber is traded in domestic markets (Cerutti and Lescuyer, 2011; Kishor and Lescuyer, 2012; Lescuyer *et al.*, 2014). It has been estimated that between 10% and 80% of the total timber trade is illegal (Seneca Creek Associates, 2004) and in some countries, such as Papua New Guinea, Democratic Republic of the Congo, Liberia and the countries of the Amazon region (Stark and Pang Cheung, 2006; Lawson and MacFaul, 2010; Wit *et al.*, 2010; Hoare, 2015; IUFRO, 2016), it is as high as 80-90%. Worldwide, the local control of harvesting and trade of timber species is mainly

based on certificates with declared origin. However, these control systems have been weakened by the frequent use of false declarations of species and geographic origin as these documents are prone to be falsified. To effectively combat fraud in illegal logging and trade, there is a need for forensic techniques to independently verify the origin in both local and international markets (Degen, 2007; UNODC, 2016). Techniques based on the intrinsic characteristics of the wood would help confirm the origin of the timber more reliably and would thereby support the combat against illegal timber trade.

To date, there are various promising methods based on timber properties to trace its origin (Table 1.1) such as mass spectrometry (Fidelis *et al.*, 2012; Lancaster and Espinoza, 2012; Espinoza *et al.*, 2015; Musah *et al.*, 2015; Deklerck *et al.*, 2017; Finch *et al.*, 2017), near-infrared spectroscopy (Braga *et al.*, 2011; Pastore *et al.*, 2011; Sandak *et al.*, 2011; Bergo *et al.*, 2016), stable isotopes (Horacek *et al.*, 2009; Kagawa and Leavitt, 2010; Förstel *et al.*, 2011), genetics (Degen *et al.*, 2006; Degen and Fladung, 2007; Lowe, 2007; Jolivet and Degen, 2012; Degen *et al.*, 2013; Lowe *et al.*, 2016) and wood anatomy (Gasson, 2011; Gasson *et al.*, 2011; Wheeler, 2011; Moya *et al.*, 2013). These methods, individually or in combination, could assist in tracing the geographical origin of timber and tree species (McGough, 2010). In this study, I assess the potential for timber tracing based on mass spectrometry, stable isotopes composition and genetic techniques (Table 1.1). Direct Analysis in Real Time Time-of-Flight- Mass Spectrometry (DART-TOFMS) analyses the chemical composition of heartwood. The DART source consists of an ionization technique that occurs at atmospheric pressure and is discussed by Cody *et al.* (2005). Once the molecules from the sample are ionized, they are directed towards the time-of-flight mass spectrometer (TOFMS) (Cody *et al.*, 2005). The mass spectrometer will then characterize the molecules from the sample by determining the mass to charge (m/z) of the ions in their protonated forms. The differentiation potential is based on the analysis of the masses (m/z) detected and the relative intensities obtained. Biological stable isotopes of carbon ($^{13}C/^{12}C$) and oxygen ($^{18}O/^{16}O$) can be analyzed by looking at the ratios variation imprinted in the wood. This variation is determined by the photosynthetic pathway for $\delta^{13}C$ (Sternberg *et al.*, 1984) and climate conditions for $\delta^{18}O$ (Dansgaard, 1964; Helle and Schleser, 2004). Finally, for the genetic analysis,

microsatellites –simple/short tandem repeats of <10 bp (Ellegren, 2004; Krebs *et al.*, 2014) – have been applied to characterize populations and estimate the probability of one sample of belonging to any of them (UNODC, 2016). Given that these three methods could provide different accuracies of identification, I also assessed if a combination of these approaches could increase accuracy.

Table 1.1 Tracing techniques applied globally, in the tropical forests and this thesis

Technique	Studied globally		Examples in the tropics		This thesis
	Spp ID	Geo ID	Spp ID	Geo ID	
Wood anatomy	✓	X	Moya *et al.* (2013)	X	
Mass spectrometry	✓	✓	Espinoza *et al.* (2015)	X	**Chapter 2**
Near infrared spectroscopy	✓	✓	Pastore *et al.* (2011)	X	
Stable isotopes	?	✓	X	Vlam *et al.* (2018)	**Chapter 3**
Dendrochronology	X	✓	X	X	
DNA barcoding	✓	X	?	X	
DNA fingerprinting	?	✓	?	Degen *et al.* (2013)	**Chapter 4**

Cedrela species, a widely distributed and intensively traded timber

For assessing the potential for identification of the selected timber tracing methods, I chose *Cedrela* (Meliaceae), one of the most important tropical timbers (Toledo *et al.*, 2008) that has been extensively harvested and traded (Mostacedo and Fredericksen, 1999; Richter and Dallwitz, 2000; Mostacedo and Fredericksen, 2001) and strongly affected by illegal logging. The genus *Cedrela* is distributed in the Neotropics from the Pacific coast of Mexico, through Central America, to Argentina (Tosi, 1960; Chaplin, 1980; Francis and Lowe, 2000; Toledo *et al.*, 2008). It has been traded as roundwood, lumber, veneer and imported in small quantities in Europe and larger quantities in North America (Wagenführ, 2007). Illegal logging of *Cedrela* has resulted in CITES-listing of several species in this genus (Compt and Christy, 2008). As a result, timber from these species can be traded internationally only if the appropriate permits have been obtained and presented for clearance at the port of entry or exit (CITES, 2017). *Cedrela* species present particular problems for identification due to wood-anatomical similarities

(Gasson, 2011; Gasson *et al.*, 2011; Moya *et al.*, 2013). As a result of this many threatened species are declared and traded under false names (Moya *et al.*, 2013). For authorities enforcing CITES, methods to differentiate between *Cedrela* species are needed.

Bolivia harbours as many as six different *Cedrela* species: *Cedrela angustifolia* Sessé & Moc. Ex DC., *Cedrela balansae* C. DC., *Cedrela fissilis* Vell., *Cedrela montana* Moritz ex Turcz., *Cedrela odorata* L., and *Cedrela saltensis* M.A. Zapater & del Castillo; all of which are exploited. Selective harvesting of these valuable timber species – together with *Swietenia macrophylla* and *Amburana cearensis* - has resulted in reduction and degradation of their populations in natural forests in Bolivia (Killeen *et al.*, 1993) (Gullison *et al.*, 1996). *Cedrela* species are distributed along different climatic zones, from moist to dry tropical forests, and from low to high altitudes (Mostacedo *et al.*, 2003; Navarro, 2011; Navarro-Cerrillo *et al.*, 2013). They are highly valued locally (Mostacedo and Fredericksen, 1999) and used in carpentry, fine furniture, doors, windows, joinery, musical instruments, carvings, coatings and plywood (Toledo *et al.*, 2008). *Cedrela* populations have declined considerably in recent years due to overexploitation (Mostacedo and Fredericksen, 1999; Mostacedo and Fredericksen, 2001). As a result, out of the six species, three are currently listed in Appendix III of CITES: *Cedrela odorata*, *C. fissilis*, and *C. angustifolia* (listed as *C. lilloi*) (CITES, 2017). All trade in specimens of species included in Appendix III requires the prior grant and presentation of an export permit. An export permit will only be granted when the specimen was not obtained in violation of the laws for the protection of fauna and flora of the State which has included that species in Appendix III. When importing any specimen of a species included in Appendix III, it will additionally require the prior presentation of a certificate of origin (CITES, 2018). Currently, the verification that declared information is in compliance of local regulations and CITES wood identification is based on visual inspection and revision of documentation.

This thesis focuses on evaluating timber tracing techniques based on intrinsic properties of wood. For each of the studied methods (DART, stable isotopes and genetics), I assess the accuracy for species and geographical origin identification in relation to the spatial

resolution of *Cedrela* species in Bolivia. I will also evaluate the potential of combining these three methods to trace the provenance of these species. The geographical distribution of *Cedrela* species along South America makes the application of the tracing methods promising as they could also be applicable and extended to other countries. Furthermore, this represents an opportunity to create synergies that would be able to secure the complementary efforts to control illegal logging, certify well-managed forests and maintain and enhance carbons stocks that many countries are already implementing with CITES, REDD+, FLEGT and forest certification (Putz *et al.*, 2012). Hence the tracing techniques assessed in this research would be applicable not only in Bolivia but also in the countries implementing these initiatives.

Illegal timber trade, trade regulations and control systems in Bolivia

The major economic activities in developing countries like Bolivia are based on the use of natural resources. The total production of wood and wood products accounts for 3% of the national GDP and about 50 thousand people in the country are directly involved in operations related to timber forest activities such as extraction, transport and processing of wood (Malky, 2005). Many forest products, besides the great value and utility in the country, are exported to other countries. For example, export of timber products, e.g. floors, furniture boards, sawn wood and posts, had a value of USD 51.2 million during 2017 (CFB, 2018). A high percentage of the exported products comes from certified forest management. Currently, Bolivia is in the fifth position -981,862 hectares- of the tropical forests certified by the Forest Stewardship Council (CFV, 2017). Forest certification is an important promoter of the forest management practices in compliance with national and international laws.

The current Bolivian legislation stipulates that logging activities should be done with prior inventory and census of trees, and with an analysis of the population structures and ecosystem conditions that will be impacted (BOLFOR/FMT, 2003). Harvesting is based on

the obligatory presentation of a management plan that follows the inventory. The management plan proposes rules and parameters for operation in a given area in accordance with the regulations and it is duly approved by the competent authority (Ley Forestal 1700, 1996). The transportation of the harvested timber is accompanied by a corresponding certificate of origin declaration. The forest certificate of origin (CFO) is the only document proving the legal origin of forest products. It is a sworn statement which certifies the origin of forest products and supports their transport, storage, processing and marketing. It is non-transferable and only issued by the competent authority. If this certificate is absent or not officially signed the transportation will be under penalty of confiscation according to Forestry Law (Ley Forestal 1700, 1996), Article 74th, Chapter I: Of plans and programs for supply management and processing of raw material. Despite legal harvesting limitations, highly valuable timbers remain at high risk because of continued illegal logging and timber trade: for example in 2012 and 2016 the percentage of illegal logging was 92% and 64%, respectively (ABT, 2017). These high illegality rates are an indication that control systems for logging are still being violated. A common problem of illegal logging is the "informality" and the small scale in which it operates (Hoare, 2015). Declaration of timber origin accompanying timber from these operations is in some cases false: an endangered species could have been harvested or timber may have been logged outside the permitted area of harvesting. The sale of certificates of origin to cut and extract timber without any permission through bribery is common at the checkpoints. A truck, for example, can carry up to 15,000 board feet of lighter species and 11,000 board feet of heavy wood– species which are coming from natural stands to local markets in Santa Cruz, Oruro and La Paz (Jose Ledezma, pers. Comm.). Since all *Cedrela* species have light wood, this makes it easy to mix wood from different regions and thereby masking their true origin. In recent years, the local government implemented a digital system for the emission of the certificates of origin in which the harvested timber and forest products are registered by the user. Although this digital system processes the information faster, it is dependent on the person who enters the data and hence still vulnerable to manipulation.

Illegally sourced timber products can enter at any stage of the timber supply chain (Lowe *et al.*, 2016). The traditional control and monitoring systems of timber are intended as a

means to control timber use in forest areas and its transport. The implementation of the local and international regulations presents major challenges because such systems are prone to manipulation of falsified data along the chain of custody. Furthermore, the system to control the timber provenance based on the respective declaration of origin is weakened by the lack of methods that are able to verify the origin certificates in connection with regulations (Degen and Fladung, 2007). At this point it is therefore urgent to have an effective technique to identify illegality and control timber trade.

Objectives and hypotheses

The overall objective of the project is to support the implementation of sustainable forest management in Bolivia by developing methods to detect illegal logging. The specific research objectives of the project are as follows:

1. To evaluate the effectiveness of Direct Analysis in Real Time Time-of-Flight Mass Spectrometry (DART-TOFMS) to identify *Cedrela* species and trace the regional origin of timber from these species.

2. To assess the applicability and spatial resolution of DNA fingerprinting (using microsatellites) to trace the origin of *Cedrela odorata* within Bolivia.

3. To evaluate the applicability and spatial resolution of stable isotope analyses to trace the origin of *Cedrela fissilis* and *Cedrela odorata* within Bolivia.

These objectives are addressed in three core chapters:

Chapter 2: In this chapter I assess the potential of Direct Analysis in Real Time Time-of-Flight Mass Spectrometry (DART-TOFMS) to identify *Cedrela* species and sites of origin. To this end, I try to answer the following research questions: (1) To what extent can Bolivian *Cedrela* species be differentiated based on wood chemical composition?; (2) To what extent can chemical composition help to differentiate timber from different sites in Bolivia?; (3) What is the accuracy for identification of each *Cedrela* species and site of

origin within Bolivia based on their chemical profiles? I expect to find distribution patterns of the wood composition that mirror local environmental conditions (Zobel and van Buijtenen, 1989; Wilkins and Stamp, 1990; Mosedale and Ford, 1996; Moya and Calvo-Alvarado, 2012) as the geographical sites of the collected samples may have different environmental conditions. I also expect that each *Cedrela* species will present specific chemicals that distinguish it from others (Chatterjee *et al.*, 1971; Maia *et al.*, 2000; Lago *et al.*, 2004; Eason and Setzer, 2007; Cordeiro *et al.*, 2012).

Chapter 3: This chapter covers an assessment of stable isotopes ($\delta^{13}C$ and $\delta^{18}O$) to identify the geographical origin of two of the most intensively exploited tropical timbers: *Cedrela fissilis* and *C. odorata.* I also explore the potential of identification of *Cedrela* species in Bolivia based on their isotopic fingerprints. My research questions are: (1) At what scale do Bolivian *Cedrela odorata* and *C. fissilis* populations differ in their oxygen and carbon stable isotopes characteristics?; (2) To what extent will this differentiation level allow us to identify *Cedrela* species and track timber origin at different spatial resolution?; (3) How can the relationship between rainfall, altitude and the isotopic ratio be used to predict the likelihood of belonging to specific sites? I expect that isotopic ratios will vary throughout the sampling sites given that isotopic fractionation varies according to environmental conditions (McCarroll and Loader, 2004), altitude and regional climatic regimes (Dansgaard, 1964; Förstel and Hützen, 1983; Gonfiantini *et al.*, 2002). I also expect that the biggest differences of isotopes ratios will be found for the sites with higher altitudes in comparison to the lowlands because precipitation variation is more homogeneous in the latter.

Chapter 4: To investigate the spatial distribution and genetic variation in three sites of *C. odorata* in Bolivia, I address the following questions: (1) Do the study sites represent distinct genetic groups and follow a spatial pattern across the species distribution?; (2) To what extent does this genetic structure allow successful timber tracing at a regional and nation-wide level? I expect that site discrimination will increase with increasing distances due to the genetic structure and discrimination (Gillies *et al.*, 1999; Jolivet and Degen, 2012; Vlam *et al.*, 2018) and that protected areas still harbour a larger historic genetic diversity

given that logging may result in a decrease of genetic variation (Finkeldey and Ziehe, 2004) and allelic frequencies (Cornuet and Luikart, 1996; Rajora *et al.*, 2000).

Combining approaches

It has been suggested that combining tools will improve accuracy (von Scheliha and Zahnen, 2011; Dormontt *et al.*, 2015; Johann Heinrich von Thünen Institute, 2015). In addition, there has been a lot of discussion about comparing different tracing methods although they have only been tested individually and in relation to different statistical methods for data analysis (Degen *et al.*, 2017; Deklerck *et al.*, 2017; Finch *et al.*, 2017). If multiple tracing methods addressing the same question could be used alongside each other, then the accuracy of source identification could be improved. In *Chapter 5*, I address the questions (1) Is there any difference in the identification accuracy between the three studied techniques: Direct Analysis in Real Time Time-of-Flight Mass Spectrometry, microsatellites, and stable isotopes?; (2) Does combining the three methods improve identification accuracy of species and geographic origin? To answer these questions, I compare the identification accuracies between the methods applied in this study individually and combined. Considering the different types of data and that both environment and genetics contribute to variation in trees (Jara, 1995; Sultan, 2000; Johnson and Agrawal, 2005; Andrew *et al.*, 2010), I expect that the tracing techniques would be complementary and that the different sites signals imprinted in wood would be singularly captured by each method.

Study species

The genus *Cedrela* belongs to the group of softwood and valuable timber species in Bolivia. In general, they are deciduous species and partially light demanding (Brienen and Zuidema, 2006; Brienen *et al.*, 2010). They rapidly grow in forest clearings in variable soils and topography but require good drainage (Mostacedo *et al.*, 2003). They can be found from moist to dry tropical forests (Figure 1.1) and within a wide range of

positions in the storey (Mostacedo *et al.*, 2003; Navarro, 2011; Cavers *et al.*, 2013; Navarro-Cerrillo *et al.*, 2013). *C. odorata* is found in moist subtropical forests. *C. fissilis* is abundant in eastern Bolivia, mainly in warm, temperate, moist forests and subtropical forests. Both species are found at altitudes from 0 to 1500 m.a.s.l. (meters above sea level). *C. angustifolia*, and *C. montana*, are generally located at altitudes from 1900 to 2200 m.a.s.l. Individuals of *C. balansae* are found in both warm, temperate moist forests and subtropical moist forests at altitudes from 0 to 1700 m.as.l. Finally, *C. saltensis* is also found at higher altitudes from 700 to 2700 m.a.s.l.

Figure 1.1 *Cedrela* species collected from different forest types in Bolivia

C. angustifolia

C. balansae

C. fissilis

C. montana

C. odorata

C. saltensis

Cedrela trees can be as tall as 40 meters. They have shallow roots in the upper soil layers (Noldt *et al.*, 2001). The trunk and crown of all these species are similar (Toledo *et al.*, 2008). The *C. odorata* trunk has a wider tabular base. In contrast, *C. fissilis* has a cylindrical trunk from the base to the top. Alternate leaves are arranged spirally; they are paripinnate, sometimes imparipinnate. The petiole range from 4 to 12 cm and the rachis from 30 to 46 cm. Leaves have from 12 to 26 opposite or sub-alternate leaflets which can be whole and glabrous or pubescent. *Cedrela* trees are monoecious, with female (that open first) and male flowers in the same inflorescence (Toledo *et al.*, 2008). They are pollinated by insects (Bawa *et al.*, 1985; Howard *et al.*, 1995). The fruits are located in the top of the branches and when they mature their valves open from top to bottom releasing around 50 seeds (Toledo *et al.*, 2008). Their light wing-shaped seeds are dispersed by wind over long distances (Mostacedo *et al.*, 2003; Toledo *et al.*, 2008).

Study sites

This study was carried out on 6 *Cedrela* species: *C. angustifolia*, *C. balansae*, *C. fissilis*, *C. montana*, *C. odorata*, and *C. saltensis*; collected from 12 sites covering the *Cedrela* distribution in Bolivia (Figure 1.2). These sites are located in different forest types from moist to dry tropical forests: Amazon, Chaqueño, Chiquitano, Yungas, and Tucuman. Annual rainfall ranges from 579 to 2008 mm (SENAMHI, 2018b) and sample altitudes from 127 to 2175 m.a.s.l. (Table 1.2). Dry season covers the months June-August and rainy season the months November-February. A total of 288 samples were analysed in this thesis. Collection of samples was carried out during the years 2015 and 2016. Samples were collected from natural forest areas: community sites and roads with remnant *Cedrela* trees and protected areas. Dominant trees that naturally grew in these sites were kept as the main criteria for selection of samples. In addition, distant trees were preferred for genetic analysis (minimum distance 26 m) and different distance gradients for the chemical analyses (minimum distance 2.5 m) with stable isotopes and Direct Analysis in Real Time Time-of-Flight Mass Spectrometry (DART-TOFMS). The

phenology of the *Cedrela* trees varied even between trees in the same sampling site. Given such cases, when collection of leaves was not possible cambium was collected.

Table 1.2 *Cedrela* sampling sites and sample sizes throughout its distribution in Bolivia. Altitude was estimated as the mean from all the collected samples. Annual rainfall was provided by the Servicio Nacional de Meteorología e Hidrología (SENAMHI, 2018b) for the period 2005-2016.

Species	Sites	Sample size	Botanical samples	Mean altitude (m.a.s.l.)	Annual rainfall (mm)
C. angustifolia	Monteagudo	2	2	2146	816
	Postrervalle	13	12	2021	978*
C. balansae	Concepción	10	10	415	1031
	Villamontes	10	3	588	983
C. fissilis	Bajo Paraguá	21	2	292	579
	Concepción	34	13	462	1031
	Espejos	27	6	497	1503
	Guarayos	34	22	238	1402
	Roboré	21	13	634	1096
	Yapacaní	10	5	318	1888
C. montana	Postrervalle	2	2	2026	978*
C. odorata	Cobija	36	2	267	1952
	Riberalta	34	1	158	1797
	Rurrenabaque	26	5	443	2008
C. saltensis	Monteagudo	8	4	1595	816
Total		**288**	**102**		

* SENAMHI database limited to period 1982-1992.

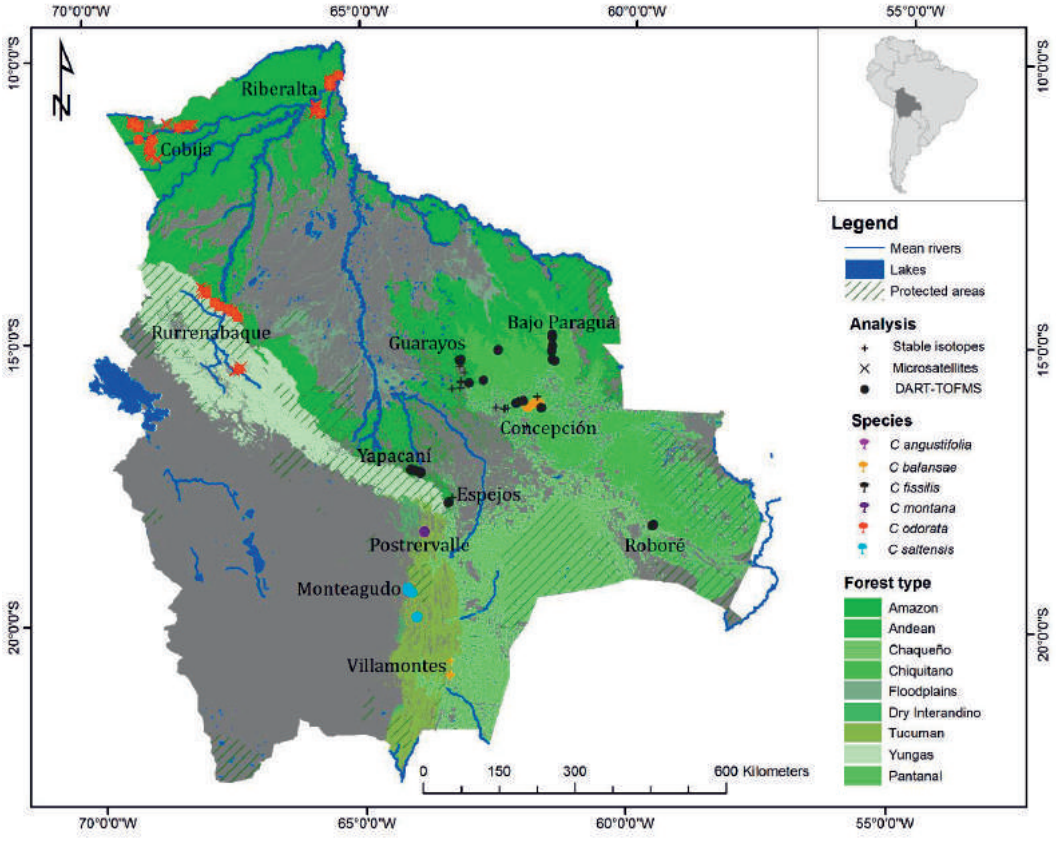


Figure 1.2 Study area: sampling sites throughout the *Cedrela* distribution in Bolivia. *Cedrela* species included in the thesis (colours) and type of analysis applied (symbols).

Forest cover: Autoridad de Bosques y Tierra, 2015.

Thesis outline

This thesis comprises a general introduction (**Chapter 1**), three research chapters (**Chapters 2** to **4**) and the general discussion (**Chapter 5**). Research chapters include analysis of 3 different tracing methods with samples collected at different distances to assess their scale and precision. In the first research chapter, **Chapter 2**, I evaluate the potential of chemical properties by Direct Analysis in Real Time Time-of-Flight Mass

Spectrometry (DART-TOFMS) to identify *Cedrela* species and their geographical origin. In **Chapter 3**, I assess the potential of $\delta^{13}C$ and $\delta^{18}O$ stable isotopes to identify species and geographical origin. In addition, I evaluate their relationship with rainfall and altitude. In **Chapter 4,** I analyse the genetic properties in *C. odorata* using microsatellites to identify its geographical origin at a country level. Finally, **Chapter 5** covers a comparison of the 3 methods used for species and sites identification in each of the research chapters. In addition, I test a combination approach in which I integrate different methods. This combination approach allowed me to assess whether accuracies are improved. The need for tracing at different spatial scales is also discussed and the implications for timber tracing are explored. I conclude with recommendations based on the findings of this thesis and discussions with stakeholders in both producing and consuming countries.

Chapter 2

Chemical differentiation of Bolivian *Cedrela* species as a tool to trace illegal timber trade

Kathelyn Paredes-Villanueva, Edgard Espinoza, Jente Ottenburghs, Mark G. Sterken, Frans Bongers, and Pieter A. Zuidema

Published in Forestry: An International Journal of Forest Research (2018) 91(4): cpy019

Abstract

Combating illegal timber trade requires the ability to identify species and verify geographic origin of timber. Forensic techniques that independently verify the declared species and geographic origin are needed, as current legality procedures are based on certificates and documents that can be falsified. Timber from the genus *Cedrela* is among the most economically valued tropical timbers worldwide. Three *Cedrela* species are included in the Appendix III of CITES: *C. fissilis*, *C. odorata*, and *C. angustifolia* (listed as *C. lilloi*). *Cedrela* timber is currently traded with false origin declarations and under a different species name, but tools to verify this are lacking. We used Direct Analysis in Real Time Time-of-Flight Mass Spectrometry (DART-TOFMS) to chemically identify *Cedrela* species and sites of origin. Heartwood samples from six *Cedrela* species (the three CITES-listed species plus *C. balansae*, *C. montana*, and *C. saltensis*) were collected at 11 sites throughout Bolivia. Mass spectra detected by DART-TOFMS comprised 1062 compounds; their relative intensities were analysed using Principal Component Analyses (PCA), Kernel Discriminant Analysis (KDA), and Random Forest analyses to check discrimination potential among species and sites. Species were identified with a mean discrimination error of 15-19%, with substantial variation in discrimination accuracy among species. The lowest error was observed in *C. fissilis* (Mean=4.4%). Site discrimination error was considerably higher: 43-54% for *C. fissilis* and 42-48% for *C. odorata*. These results provide good prospects to differentiate *C. fissilis* from other species, but at present there is no scope to do so for other tested species. Thus, discrimination is highly species specific. Our findings for tests of geographic origin suggest no potential to discriminate at the studied scale and for the studied species. Cross-checking results from different methods (KDA and Random Forest) reduced discrimination errors. In all, the DART-TOFMS technique allows independent verification of claimed identity of certain *Cedrela* species in timber trade.

Keywords: Illegal logging, *Cedrela*, mass spectrometry, discriminant analysis, Random Forest

Introduction

Illegal trade in timber is a worldwide environmental problem, resulting in damage of natural resources and economic loss. It has been estimated that 10% to 80% of the total timber trade is illegal (Seneca Creek Associates, 2004) and in some countries, such as Papua New Guinea, Liberia, and the Amazon countries (Stark and Pang Cheung, 2006; Lawson and MacFaul, 2010; Wit *et al.*, 2010), this percentage has been as high as 80-90% of all logging operations. The most common type of fraud concerns false declarations of species and geographic origin, as current legal procedures are generally based on certificates and documents which can be falsified. Most legislative measures focus at combating international illegal trade but a high proportion (70-90%) of illegal tropical timber is traded in domestic markets (Cerutti and Lescuyer, 2011; Kishor and Lescuyer, 2012; Lescuyer *et al.*, 2014). Clearly, there is a need for forensic techniques to independently verify the origin of traded timber in both domestic and international markets.

The genus *Cedrela* (Meliaceae) delivers one of the most important tropical timbers (tropical cedar), but illegal logging of *Cedrela* has resulted in CITES-listing of several species in this genus (Compt and Christy, 2008). As a result, timber from these species can be traded internationally only if the appropriate permits have been obtained and presented for clearance at the port of entry or exit (CITES, 2017). The problem is that CITES-listed and non-listed *Cedrela* species are harvested and traded under the same name (Moya *et al.*, 2013) and are often confused due to wood-anatomical similarities (Gasson, 2011; Gasson *et al.*, 2011; Moya *et al.*, 2013). For authorities enforcing CITES, methods to differentiate *Cedrela* species are needed.

Bolivia harbours as many as six *Cedrela* species, in different climatic zones, from moist to dry tropical forests, and from low to high altitudes (Mostacedo *et al.*, 2003; Navarro, 2011; Navarro-Cerrillo *et al.*, 2013): *Cedrela angustifolia* Sessé & Moc. Ex DC., *Cedrela balansae* C. DC., *Cedrela fissilis* Vell., *Cedrela montana* Moritz ex Turcz., *Cedrela odorata* L., and *Cedrela saltensis* M.A. Zapater & del Castillo. *Cedrela* species are highly valued locally

(Mostacedo and Fredericksen, 1999) and used in carpentry, fine furniture, doors, windows, joinery, musical instruments, carvings, coatings and plywood (Toledo *et al.*, 2008). However, *Cedrela* populations have declined considerably in recent years due to overexploitation (Mostacedo and Fredericksen, 1999; Mostacedo and Fredericksen, 2001). As a result, out of the six species, three are currently listed in Appendix III of CITES: *C. odorata*, *C. fissilis*, and *C. angustifolia* (listed as *C. lilloi* C. DC.) (CITES, 2017). Despite legal harvesting limitations, these species remain at high risk because of continued illegal logging and timber trade (ABT, 2017). The high incidence of illegal trade indicates that control systems have limited effectiveness and methods for independent verification of species and legal origin are needed.

Chemical analysis tools, such as mass spectrometry (Fidelis *et al.*, 2012), near-infrared spectroscopy (Braga *et al.*, 2011; Bergo *et al.*, 2016), and stable isotopes (Kagawa and Leavitt, 2010; Förstel *et al.*, 2011; Vlam *et al.*, 2018), can be used to discriminate species and verify the geographical origin of traded timber. For example, previous studies used a specific mass spectrometer to discriminate species that cannot be identified based on wood anatomy in the Americas (Espinoza *et al.*, 2015), Africa (Deklerck *et al.*, 2017), and Asia (McClure *et al.*, 2015). In this study, we focus on chemical characterization by Direct Analysis in Real Time (DART) coupled with Time-of-Flight Mass Spectrometry (TOFMS). This technique has the potential to assist in enforcing protection of *Cedrela* species as it cannot be falsified, in contrast to current certificates used for declaration of species origin. In DART analysis, the mass spectrometer quickly identifies the chemical components by the differing mass to charge (*m/z*) of ions/compounds from specimens, without the need for sample preparation. The resulting chemical spectra can be used as a reference database for species identifications. Because this methodology has a high potential to identify species and locations, our aim is to test its applicability to differentiate *Cedrela* timber obtained from different species and geographic provenances.

We answer the following research questions: (1) To what extent can Bolivian *Cedrela* species be differentiated based on wood chemical composition?; (2) To what extent can

chemical composition help to differentiate timber sourced from different sites in Bolivia?; (3) What is the accuracy for identification of each *Cedrela* species and site of origin within Bolivia based on their chemical profiles? As the geographical sites of the collected samples may have different environmental conditions, we expect to find distribution patterns of the wood composition that mirror these conditions (Zobel and van Buijtenen, 1989; Wilkins and Stamp, 1990; Mosedale and Ford, 1996; Moya and Calvo-Alvarado, 2012). We also expect that each *Cedrela* species will present specific chemicals that distinguish it from others (Chatterjee *et al.*, 1971; Maia *et al.*, 2000; Lago *et al.*, 2004; Eason and Setzer, 2007; Cordeiro *et al.*, 2012).

Methods

Study site and species

We studied heartwood samples from 6 *Cedrela* species in Bolivia, from 11 sites. In total we sampled 127 trees. Altitude of the sites ranged from 145 m.a.s.l. (meters above sea level) in Riberalta to 2022 m.a.s.l. in Postrervalle (Table 2.1). We selected sites taking into account the distribution of the study species and we maintained a minimum of 70 km distance between all site pairs to maximize the sampling coverage across the country (Table 2.1 and Figure 2.1). The maximum distance between pairs of sampled sites was 1300 km (Cobija-Roboré). We used these samples to perform two types of tests: differentiation of species and differentiation of geographic origin. In the species identification analyses, we included all *Cedrela* species in the sample collection to analyse cross-species discrimination. For the geographic origin analysis, we only included the two species with the largest sample sizes that we had sampled at multiple sites: *C. odorata* from 3 sites and *C. fissilis* from 6 sites. The maximum distance between pairs of sites was 80 km for *C. fissilis* (Espejos-Yapacaní) and 425 km for *C. odorata* (Riberalta-Rurrenabaque). Minimum distances between pairs of sites were 70 km (Concepción-Guarayos) for *C. fissilis* and 285 km (Riberalta-Cobija) for *C. odorata*. We performed a stratified random sampling: in each of the *Cedrela* populations found, trees

of diameter ≥10 cm were randomly selected with a minimum distance among trees of at least 50 m in order to obtain a homogeneous sampling in each site and to reduce genetic noise and confounding impact of sampling relatives on site (Gillies *et al.*, 1999). This random selection of samples covered different types of forest strata.

Table 2.1 *Cedrela* species and sites included in the study. Sample size refers to the number of trees sampled; botanical samples to the number of trees from which botanical samples were obtained for verification of identification by taxonomists.

Species	Sites	Sample size	Botanical samples	Altitude (m.a.s.l.)
C. angustifolia	Monteagudo	2	2	1705
	Postrervalle	13	12	2022
C. balansae	Concepción	10	10	432
C. fissilis	Bajo Paraguá	10	*	287
	Concepción	13	13	432
	Espejos	6	6	553
	Guarayos	13	9	260
	Roboré	10	*	632
	Yapacaní	10	5	318
C. montana	Postrervalle	2	2	2022
C. odorata	Cobija	10	*	274
	Riberalta	10	*	145
	Rurrenabaque	10	4	309
C. saltensis	Monteagudo	8	2	1705
Total		**127**	**65**	

*No botanical samples were collected, but identification was based on previous collections.

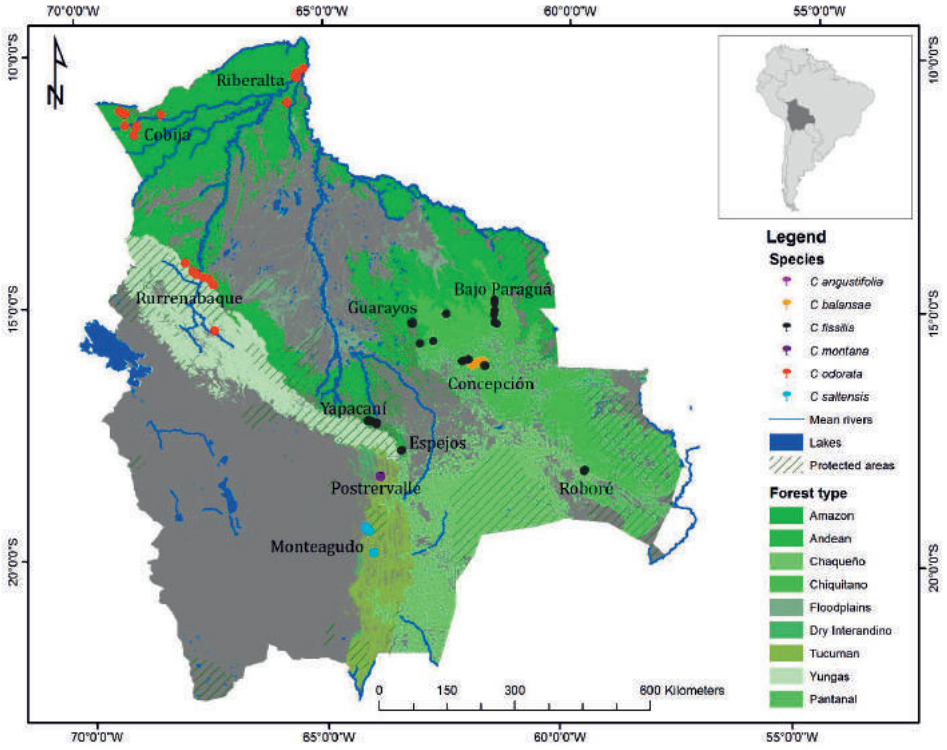


Figure 2.1 Locations of sampled trees belonging to six *Cedrela* species in Bolivia. Forest cover: Autoridad de Bosques y Tierra, 2015.

Preliminary analyses of sapwood and heartwood showed a wider variation of compounds in heartwood (70.0629-1086.567 *m/z*) compared with sapwood with a dominance of sugars and starch (69.0285-958.4909 *m/z*) that were not species-specific. Based on these results, we decided to only include heartwood samples in our analyses. A single heartwood sample was collected from each tree using a 5 mm diameter increment borer (Haglöf) at 50-100 cm stem height. Species were morphologically identified *in situ* with the help of local guides. In addition, botanical samples were collected for species confirmation when identification in the field was not possible. This was done for 53% of the sampled trees. The voucher preparation and confirmation of the species based on herbarium collections were carried out by an experienced botanist, A. Araujo Murakami at the Museo de Historia Natural Noel Kempff Mercado (Bolivia).

Chemical analysis

We used Direct Analysis in Real Time Time-of-Flight- Mass Spectrometry (DART-TOFMS) to differentiate *Cedrela* species and to explore if geographical origin could be determined based on chemical composition of heartwood. The DART source consists of an ionization technique that occurs at atmospheric pressure and is discussed by Cody *et al.* (2005). Once the molecules from the sample are ionized, they are directed towards the time-of-flight mass spectrometer (TOFMS) (Cody *et al.*, 2005). The mass spectrometer will then characterize the molecules from the sample by determining the mass to charge (*m/z*) of the ions in their protonated forms.

The principal ionization mechanisms for DART-TOFMS have been thoroughly discussed and it has been used to identify timber species with an accuracy of 70% to 95% (Lancaster and Espinoza, 2012; Evans *et al.*, 2017). To describe the chemotaxonomic relationship of our *Cedrela* samples, mass spectra were acquired using a DART ion source (IonSense, Saugus, MA, USA) coupled to a JEOL AccuTOF time-of-flight mass spectrometer (JEOL USA, Peabody, MA, USA) in positive ion mode. To check if preparation of wood was needed, we tested the maximum number of compounds by soaking wood in methanol versus using wood with no previous treatment. We did not observe any enhancement with previous preparation of wood samples (data not shown). Hence, we decided to use untreated heartwood samples. We cut slivers of heartwood no wider than 4 mm from each sample with a scalpel. These slivers were held in the DART helium gas stream for 8 seconds. A mass calibration standard of polyethylene glycol 600 (Ultra, Kingstown, RI, USA) was run between each 5 samples. The DART source parameters were: needle voltage, 3.5 kV; electrode 1 voltage, 150 V; electrode 2 voltage, 250 V; and gas heater temperature, 350°C. The mass spectrometer settings included: rings lens voltage, 5 V; orifice 1 voltage, 20 V; orifice 2, 5 V; cone temperature, 120°C; peaks voltage, 600 V; ion guide bias, 28 V; focus lens voltage, -120 V; reflectron voltage, 870 V; pusher voltage, 778 V; pulling voltage, -778 V; suppression voltage, 0.00 V; flight tube voltage, -7000 V; and detector voltage, 2000 V. Spectra covered the mass range of 70 to 1100 mass-to-charge ratios (*m/z*) and were obtained at 1 scan per second. The

helium flow rate for the DART source was 2.0 mL s^{-1}. The resolving power of the mass spectrometer, as stated by the manufacturer, was ±2.0 millimass units (mmu). The diagnostic compounds for spectrum classification were selected with 250 mmu and 1% threshold (Deklerck *et al.*, 2017). TSS Unity, a mass-spec data-processing software (Shrader Software Solutions, Inc., Grosse Pointe Park, MI, USA), was used to export the data as text files for further analysis.

Statistical analysis

Our analysis of the masses (*m/z*) detected and relative intensities obtained from the DART TOFMS consisted of several steps. To evaluate if chemotypes can enable differentiation of *Cedrela* species (research question 1: species identification) we first evaluated the existence of species specific compounds, reduced the sample-compound data matrix using Principal Component Analysis (PCA), and finally performed a discriminant analysis to classify the species, determine the importance of each compound, and predict sample assignment. For these analyses we used Kernel Discriminant Analyses with package ks 1.10.5 (Duong, 2007, 2017), and Random Forest model with package randomForest 4.6.12 (Liaw and Wiener, 2002) and dplyr 0.7.4 (Wickham *et al.*, 2017) in R version 3.3.3 (R Development Core Team, 2017). We used PCA in order to reduce the number of variables (compounds) into principal components that can then be used as input for a first type of discriminant analysis (KDA). A second discriminant analysis (Random Forest) was used to identify the most important compounds that can differentiate between species. Both discriminant analyses were based on randomized samples and variables in every run. The classification results allowed us to assess the classification success by evaluating frequencies of correct and erroneous identifications. For the analysis of geographic origin (research question 2: geographic origin identification) for *C. fissilis* and *C. odorata* (Table 2.1), we followed the same steps. Based on the classifications, cross validation errors were estimated for species and site assignments (research question 3: identification accuracy).

In detail the method involved four main steps. First, we produced a heat-map graph to visualize the chemical profiles (or chemotypes) of the specimens and to verify whether

heartwood samples of a particular species contain diagnostic molecules (expressed as mass-to-charge ratio: *m/z*) that allow it to be distinguished from other species. The heatmap is a graphical representation of the raw mass spectra measured by DART-TOFMS and is created using the Mass Mountaineer software (RBC Software, Peabody, MA, USA). It illustrates the mass-to-charge ratio (*m/z*) of the detected compounds and their intensities in a spectrum.

Second, to reduce the large data matrix into a set of variables so that the variation within each set is maximized (Gotelli and Ellison, 2004), a PCA was necessary for the set which consisted of 125 samples and 1062 compounds. PCA aims to find the linear combinations of variables by using the covariance matrix of data. The first axis reflects the linear fit capturing most of the variation and the successive orthogonal axes reflect the linear capturing of remaining variation not captured in each of the previous components. We extracted six principal components from the sample-molecule matrix, reflecting the greatest variation in the data matrix. The loadings of all 125 samples on the first six axes were retained and this new matrix was used as input in the discriminant analysis. We excluded *C. montana* due to its small sample size (2 individuals).

Third, we performed Kernel Discriminant Analysis (KDA) to test species identification and geographic origin. As KDA cannot cope with more than 6 variables, we performed the PCA analysis described above, and used the first 6 PCA axes. KDA separates the samples based on an *a priori* classification assignment (to species and sites classes) and looks for the optimal non-linear combination of variables (here the 6 component loadings) for maximal separation of the samples in the six dimensional space (Baudat and Anouar, 2000). KDA's learning algorithm uses Bayes discriminant rule which allocates a point x in the sample space to one (and only one) of the sampled populations. Each population is associated to a kernel density which was estimated implementing a diagonal data-driven (constrained, symmetric and positive-definite) bandwidth matrix (Duong, 2007). This learning algorithm needs to be trained in order to assess the discrimination power of KDA. Therefore, our data were split in two sets: 80% for training and 20% for testing the model. The pre-smoothed data were then applied to

estimate a Smoothed Cross Validation (SCV) error (Duong, 2007) as a different procedure to test correctness of the assignment tests. This delivers the classification error (%) which is the probability that samples are incorrectly assigned to a provenance. A cross validation error of 0% indicates that all the samples were correctly assigned.

Finally, we used Random Forest analysis to generate a sample classification model in which splits are based on just one chemical compound. One Random Forest run created 500 'Random Forests' which are used to obtain a final model (Breiman, 2001; Liaw and Wiener, 2002). As with KDA, the algorithm uses 80% of the dataset for training and 20% for model validation. Every run of Random Forest uses a different training set and may lead to different results. Therefore, we ran Random Forest 100 times and averaged the results. In this way, a total of 50.000 Random Forests (100 runs x 500 Random Forests) were built. For each run, the model provided a list of compounds, with their value of importance. We selected the most important compounds that occurred in >40% of the runs and calculated their frequency. These tentative assignments were based on 351 molecules described either for the *Cedrela* genera or the Meliaceae family (Afendi *et al.*, 2012). Chemical composition in wood can vary not only among species but also for a given tree species or even a given tree (Pettersen, 1984), but heartwood extractive and exudates can also be species specific (Hillis, 1987). Therefore we used the list of the most important compounds to check if any species indicative compound was present.

Random Forest analysis allowed us to identify specific chemical compounds that separate one species or site from the other. The Out-of-Bag (OOB, take one out) error rate and species class error were estimated for each of the 100 runs and used to calculate the standard deviation (SD) of these estimates. The OOB estimate is equivalent to the SCV error of the KDA analysis.

Both KDA and Random Forest analyses generated confusion matrices showing the frequency at which each species/site was wrongly classified. In addition, the total of samples tested for each species after 100 randomization runs allowed us to check with what species a single sample could be confused. Finally, the mean errors per species for

site identification across the 100 runs were obtained together with their corresponding standard deviation.

Results

A total of 1062 ions were characterized and their respective intensities were described, in 6 *Cedrela* species from 11 sites across Bolivia, from the DART-TOMFS spectra. The results were analysed for species and sites identification separately. A first inspection of chemical data in the heatmap (Figure 2.2) suggests species-specific patterns in the chemical profiles. Further cross-checking with the actual mass spectra confirmed that the *C. odorata* samples had a higher intensity of compounds with molecular masses around *m/z* 212 and 480, *C. fissilis* had higher intensities for compounds in the *m/z* 484-502 range, *C. balansae* at *m/z* 478 and 680, and *C. angustifolia* showed high intensities for compounds at *m/z* 212 and 400. Although the samples of *C. montana* showed distinctive ions at *m/z* 275 and 398, this species was excluded from further analyses due to small sample size.

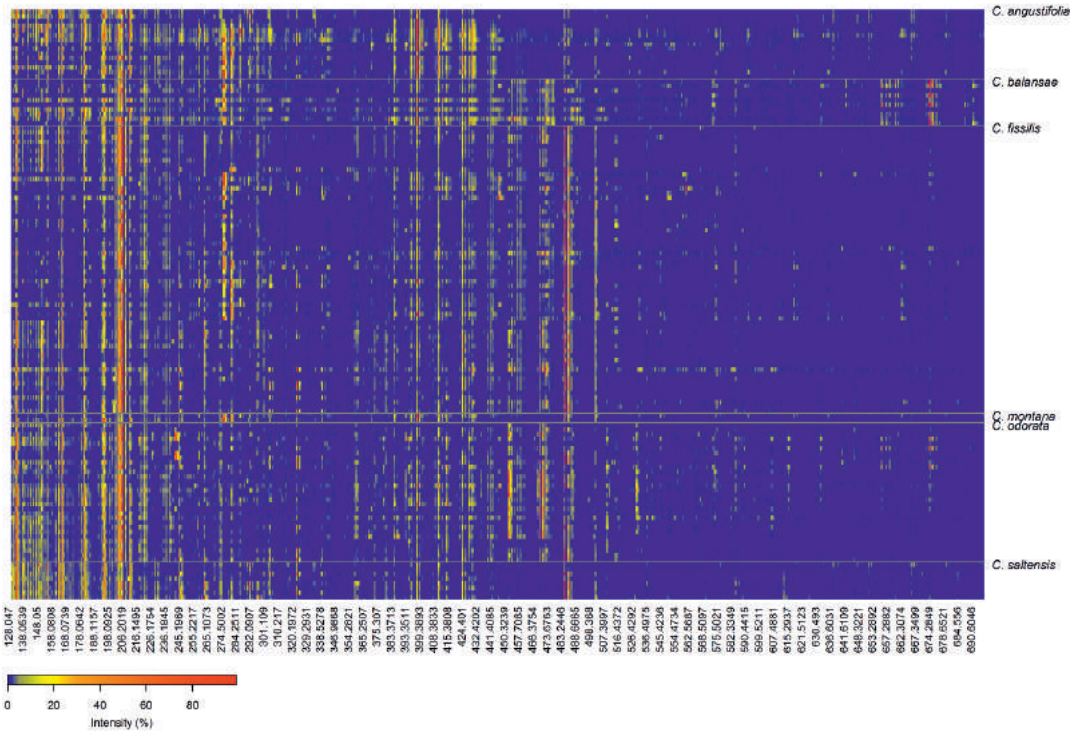


Figure 2.2 Heatmap of the output of the DART-TOMFS for 6 *Cedrela* species in Bolivia. Each row represents one sample (one tree). Each column represents a specific mass-to-charge ratio (*m/z*) of an ion. Colour gradient represents relative compound intensity (relative to the most abundant compound).

Identification of species

The analysis for species differentiation included five species*: C. angustifolia*, *C. balansae*, *C. fissilis*, *C. odorata*, and *C. saltensis* (Table 2.1). The PCA analysis showed that the six most important components together explained 72.1% of the variation across the samples and that the samples were reasonably well separated in the PCA space (Figure 2.3). The variances explained by the 6 principal components (PCs) were: 24%, 20%, 10%, 8%, 6%, and 4% for PC 1-6 correspondingly. These six components were used as input for the KDA. The KDA (of the 80% sample) resulted in a clear separation of the species (Table 2.2).

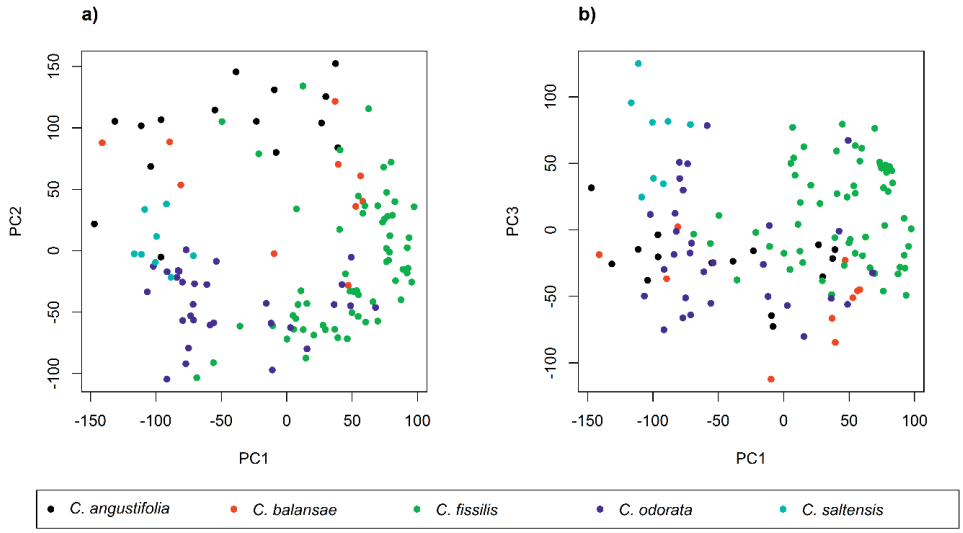


Figure 2.3 Results of Principal Component Analysis used for KDA analyses for species. Scatterplots combining a) PC1 and PC2, and b) PC1 and PC3.

Table 2.2 Error classification for species. Mean Smoothed Cross Validation (SCV) error, mean Out-of-bag (OOB) error for classification and their corresponding standard deviations (SD) were estimated after 100 runs for KDA and Random Forest respectively.

Species	KDA		Random Forest	
	Mean error (%)	SD (%)	Mean error (%)	SD (%)
C. angustifolia	26.5	28.0	33.9	7.9
C. balansae	46.1	36.8	42.4	17.7
C. fissilis	8.7	7.1	4.4	1.8
C. odorata	22.3	17.5	15.8	5.3
C. saltensis	20.2	32.4	29.6	17.2
Mean	**18.9**	**7.0**	**14.9**	**2.3**

The KDA had a total mean error of 19% for the SCV test (Table 2.2). Species-specific errors differed strongly, from 8.7% for *C. fissilis* to 46.1% for *C. balansae*. The mean error per species (OOB) from the 100 Random Forest analyses was 15%, representing a mean identification accuracy of 85% (Table 2.2; Supplementary Data Figure S.2.1). Again, these

errors differed substantially between species with the lowest value of 4.4% for *C. fissilis* and highest error of 42.4% for *C. balansae* (Supplementary Data Figure S.2.1).

In the KDA analysis, identification errors for *C. angustifolia* and *C. balansae* included wrong assignments to all the other species. *C. fissilis* was wrongly identified as all the species except as *C. saltensis*. *C. odorata* was mostly identified as *C. fissilis* (118 samples out of 562) and in some cases wrongly identified as *C. saltensis* (11 samples out of 562). It was rarely classified as *C. angustifolia* (1 samples out of 562) and never as *C. balansae*. *C. saltensis* was mostly confused with *C. odorata* (33 samples out of 188), in some cases with *C. angustifolia* (10 samples out of 188), rarely as *C. balansae* (2 samples out of 188) but never as *C. fissilis* (Supplementary Data Table S.2.1).

From the Random Forest analyses, the most important compounds for species discrimination were selected (Supplementary Data Table S.2.3). In total, 15 compounds were most frequent in over 58% of the runs (100 runs). For some compounds we were able to infer the molecular formula and make tentative assignments.

In most cases of the Random Forest analyses, each species was confused with three other species (Supplementary Data Table S.2.2): *C. angustifolia* was mostly classified as *C. fissilis* or *C. saltensis* and on one occasion as *C. balansae*. *C. balansae* was confused with all species except for *C. saltensis*. A similar pattern holds for *C. fissilis*, although this species was mostly confused with *C. odorata*. Vice versa, *C. odorata* was mostly confused with *C fissilis*, in addition to two samples that were mistakenly identified as *C. saltensis*. Finally, *C. saltensis* was confused with all species, except for *C. balansae*.

Identification of geographic origin

The analysis for geographic origin was done for *C. fissilis* and *C. odorata* separately. Classification performance was higher for Random Forest compared to Kernel Discriminant analysis. Furthermore, Random Forest showed similar error rates for both species while Kernel Discriminant showed a difference of 6.3% between *C. fissilis* and *C. odorata* (Table 2.3a and b). The error rate for site identification was highly variable for

both methods. The first six PCs were selected from the PCA analysis as (Figure 2.4) they explained the highest variance: 78.9% in the case of *C. fissilis* and 86.2% for *C. odorata*.

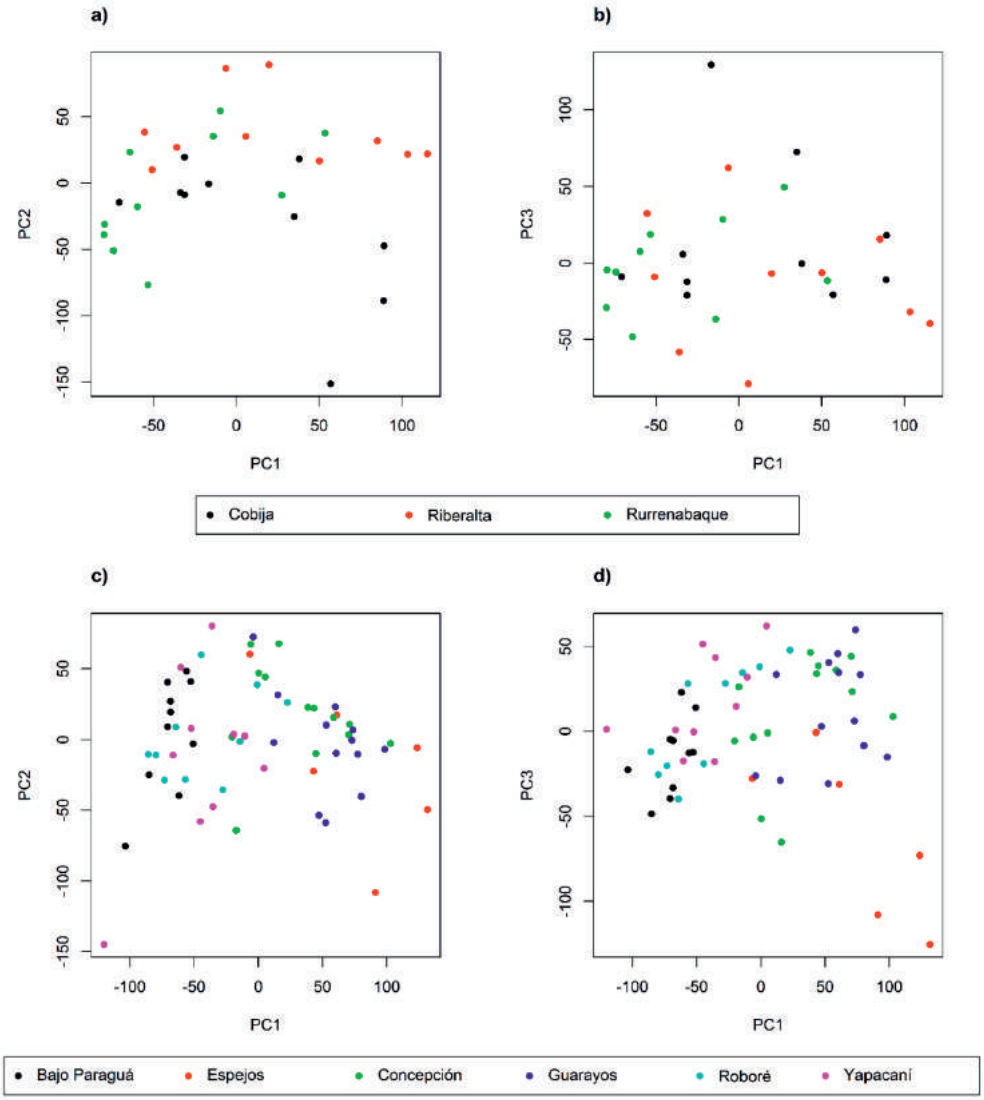


Figure 2.4 Results of Principal Component Analysis used for KDA analyses for geographic origin. Scatterplots combining (a) PC1 and PC2, (b) PC1 and PC3 for *C. odorata*, and (c) PC1 and PC2 and d) PC1 and PC3 for *C. fissilis*.

Table 2.3 Error classifications for sites of *C. fissilis* (a) and *C. odorata* (b) based on KDA with 6 PCs, and Random Forest analyses. Mean classification and standard deviation (SD) were estimated using the classification error per site after 100 runs with different training and testing sets.

	KDA		Random Forest	
	Mean error (%)	**SD (%)**	**Mean error (%)**	**SD (%)**
a) *C. fissilis* sites				
Bajo Paraguá	45.8	36.5	38.5	16.9
Espejos	43.3	44.4	86.7	11.7
Concepción	37.5	36.8	23.5	13.3
Guarayos	39.7	32.2	36.9	17.5
Roboré	80.9	33.3	57.3	17.9
Yapacaní	60.9	36.8	48.2	20.7
Mean	**53.9**	**12.5**	**42.7**	**4.8**
b) *C. odorata* sites				
Cobija	47.8	38.5	60.7	15.8
Riberalta	38.4	38.6	40.9	19
Rurrenabaque	48.5	38.5	30.4	21.6
Mean	**47.7**	**19.7**	**42.4**	**8.6**

KDA classification errors for *C. fissilis* samples were on average 53.9% (range 39.7% to 80.9%), while those for *C. odorata* averaged 47.7% (range 38.4% to 48.5%, Table 2.3). Roboré and Yapacaní showed the highest total mean error for sites discrimination of *C. fissilis* samples (Table 2.3a, Supplementary Data Figure S.2.1c), and Concepción and Guarayos the lowest. Rurrenabaque showed the highest mean error and Riberalta the lowest error for *C. odorata* sample classification (Table 2.3b, Supplementary Data Figure S.2.1d).

There was misclassification between 3-4 other sites of origin (Supplementary Data Table S.2.4) with the trained algorithm in KDA. However, some sites showed chemical characteristics clearly distinct from other sites. For example, samples from Roboré and Bajo Paraguá were distinct from Espejos but this site was often confused with Concepción and Guarayos. Samples from Bajo Paraguá and Espejos were distinct from

each other but wrongly assigned to Concepción and Yapacaní. Guarayos and Espejos were distinct from Bajo Paraguá but were wrongly assigned to Concepción and Yapacaní.

Similarly to KDA, there was misclassification between 2-3 other sites of origin in the Random Forest analyses (Supplementary Data Table S.2.5). For example, Bajo Paraguá was distinct from 3 sites: Espejos, Concepción and Guarayos but some samples were wrongly assigned to Roboré and Yapacaní. Roboré samples had a higher chance of being wrongly assigned to Bajo Paraguá compared with Yapacaní. Samples from Espejos were wrongly assigned to all sites except Bajo Paraguá. The highest error for the identification of *C. fissilis* sites using Random Forest was observed in Espejos followed by Roboré and Yapacaní while Concepción showed the best performance with the lowest error rate (Supplementary Data Figure S.2.1e and f, Table 2.3).

On the other hand, *C. odorata* showed the highest classification error for Cobija and lowest error for Rurrenabaque. Samples from Cobija were confused with samples from Rurrenabaque and Riberalta (Supplementary Data Table S.2.6). However, Rurrenabaque samples were mostly assigned to Riberalta followed by Cobija. Riberalta was confused by the other two sites but it had the highest number of correct assignments.

Although Random Forest included a higher number of samples from different sites compared with KDA (24 samples), it performed similarly in error rates and assignments. Samples from Cobija were mostly wrongly assigned to Rurrenabaque and to a lesser extent to Riberalta (Supplementary Data Table S.2.7). Samples from Rurrenabaque were wrongly assigned to Riberalta and Cobija, at roughly equal frequencies. With this method, Rurrenabaque showed the highest number of correct assignments. In each of the 100 Random Forest analyses, the most important compounds for site discrimination were elected (Supplementary Data Table S.2.8).

Discussion

To combat the illegal trade in timber, independent methods to identify species and verify geographical origin need to be developed. In this study, we assessed the effectiveness of DART-TOFMS spectra followed by multivariate statistical analysis to determine the potential for differentiating *Cedrela* species and geographic origin of *Cedrela* timber. Overall species differentiation error was 15-19% (range for two statistical methods), while that for geographic origin was significantly higher (42-54%). These discrimination errors are higher compared with previous studies that applied DART-TOFMS, which reached discrimination errors of less than 10% for species discrimination (Lancaster and Espinoza, 2012; Musah *et al.*, 2015; Evans *et al.*, 2017) and of ~30% in distinguishing between sites of origin (Finch *et al.*, 2017). We also found strong differences in discrimination error between species. Possible explanations for these differences include (1) low sample sizes for some species, (2) variation within species, (3) misidentification by the curator, or (4) variation across the sites where the species are found (e.g., some species are found together as *C. fissilis* and *C. balansae*). We will discuss these possible causes below.

Low sample size can lead to higher error rates. This is exemplified by *C. montana*, of which only two samples were collected. Including this species in the analyses increased the error of species identification from 15 to 30%. Yet other studies that applied DART-TOFMS in species with small sample sizes have successfully discriminated between species (Lancaster and Espinoza, 2012; McClure *et al.*, 2015; Wiemann and Espinoza, 2017). This discrepancy depends on the degree of chemical variation which is much smaller in some species than in others. This variability was evidenced by *C. fissilis* and *C. odorata* which showed the lowest error rates in the species discrimination analysis compared with *C. angustifolia* and *C. balansae* which showed the highest error rates. This indicates that the accuracy of discrimination is highly species specific which thwarts extrapolating these results to other species and sites. Nevertheless, a more accurate conclusion can be reached by identifying representative chemical compounds in a heatmap. This graphical overview facilitates the discovery of particular trends, such as

species-specific chemicals. Another possible source of error is misidentification by the curator. This possible observer bias could be solved by having multiple curators identify and compare herbarium samples before further analysis. In this study, the samples identified were based on a large herbarium collection and previous identifications of *Cedrela* samples throughout Bolivia.

The low accuracy of site discrimination may also be caused by local conditions such as climate, soil characteristics and nutrient availability which seem to affect tree performance and composition (Gentry, 1995; Medina *et al.*, 1995; Oliveira-Filho *et al.*, 1998; Toledo *et al.*, 2011). In the Meliaceae family, Noldt *et al.* (2001) found that some species were more sensitive to environmental conditions due to root systems in the upper soil layers. The *Cedrela* samples in our study also showed superficial tree roots and site-specific growth variation (Paredes-Villanueva *et al.*, 2016) which indicates that these trees display site-specific characteristics that may have played an important role in wood formation. Such site characteristics vary from large scale e.g. ecosystem under different climatic regimes to small scale *e.g.*, the micro site factors that contribute to tree development (Reifsnyder *et al.*, 1971). The scale variation of site identification may have played a role in our discrimination among sites: *C. odorata* sites were more distant than *C. fissilis* sites. This was confirmed when only Bajo Paraguá, Roboré and Yapacaní (the most distant sites of *C. fissilis*) were analysed: the accuracy remained similar with Random Forest (57%) and increased to 53% with KDA (data not shown). These results suggest that discriminating between more distant regions or locations may result in higher accuracies than discriminating among neighboring sites.

Apart from these external factors that influence discrimination error, the two statistical analyses we used (Random Forest and KDA) also resulted in different error rates. These errors can be reduced by comparing the probabilities of being assigned to another group. Therefore, KDA and Random Forest would best be used alongside each other as triangulation methods. Comparing and cross-checking results between groups and statistical methods will increase the certainty in identifying species and site of origin. In addition, results should also be complemented by other independent statistical tools.

Consistent results of these statistical methods could increase the confidence of correct identification when analysing the spectra generated by DART-TOFMS. Decision Trees (Kamiński *et al.*, 2018; Therneau *et al.*, 2018) or other machine learning algorithms that would also provide information of the less abundant chemicals could also be used as multiple approximation methods.

Finally, DART-TOFMS is a qualitative analysis; in order to investigate the role of distance, rainfall, altitude and the chemical composition of *Cedrela* trees in predicting the likelihood of belonging to the conditions of a specific site, it is necessary to apply a quantitative chemical approach. Such an analysis, in which the effects of sample size and time on the detection accuracy of the chemical signals are measured, will allow us to interpret the resulting molecular mass spectra across different spatial and temporal scales. The within-the-tree variation and among-site differentiation of the chemical compounds of the same species represents a great potential for more specific characterization.

All samples in this study were collected in Bolivia, a country that is severely impacted by illegal trade in timber, including *Cedrela* species. The methods used in this study showed the high potential of mass spectrometry for use in *Cedrela* species identification in Bolivia, with the highest confidence in identifying *C. fissilis*. DART TOFMS analysis can easily separate *Cedrela* genus trees from the other look-alike species, like *Swietenia macrophylla* King and *Carapa guianensis* Aubl. (Braga *et al.*, 2011; Bergo *et al.*, 2016), and this would help when false declarations and documents are being used. Previous studies also found that most of the difficulties of *Cedrela* identifications were at the species level rather than at the genus level (Gasson, 2011). Also, the accuracy of identification between samples from the genera *Dalbergia* and *Machaerium* was >95% (Lancaster and Espinoza, 2012; Espinoza *et al.*, 2015). This suggests that DART TOFMS analysis may perform better in distinguishing between *Cedrela* and other look-alike genera, but suffers in species specific assignment within the taxa.

Conclusion

Cedrela species belong to a timber genus that has been overexploited in the last couple of years. The regulation of their trading has presented many challenges, given that the identification of those species that belong to the CITES list is difficult because of similar wood anatomical characteristics. Our approach offers a strategy for improving identification certainty of *Cedrela* species by using a complementary approach contributing to their proper forest management and conservation. DART-TOFMS offers an alternative for identification and chemical discrimination among such species. There are several statistical methods to analyse the data generated by DART-TOFMS. Consistent results of two statistical methods (discriminant analyses: KDA and Random Forest) were found in this study, and applying both methods on the same dataset is recommended. Our results reveal potential for *Cedrela* species assignment (81-85% accuracy), particularly for *C. fissilis* (95.6%). Our results also show that discrimination of geographical origin is not possible due to low assignment (with accuracies of 46-57% for *C. fissilis* and 52-58% for *C. odorata*). Thus, the mass spectrometric approach used here can help to identify species provenance of certain Bolivian *Cedrela* timbers, but not geographic provenance within the country.

Acknowledgements

We would like to thank the Museo de Historia Natural Noel Kempff Mercado for their support on the species identification. This research was financed by the NFP/Nuffic fellowship (the Netherlands). Fieldwork was logistically supported by Universidad Autónoma Gabriel René Moreno and financed by Alberta Mennega Stichting and The Rufford Foundation. We are also grateful to Christoph Ruttkies and Tarn Duong for their recommendations and assistance with the statistical analysis. We are also thankful to Pieter Baas and Nicolien Sol for their help during the collection of samples and analysis of data.

Supplementary Data

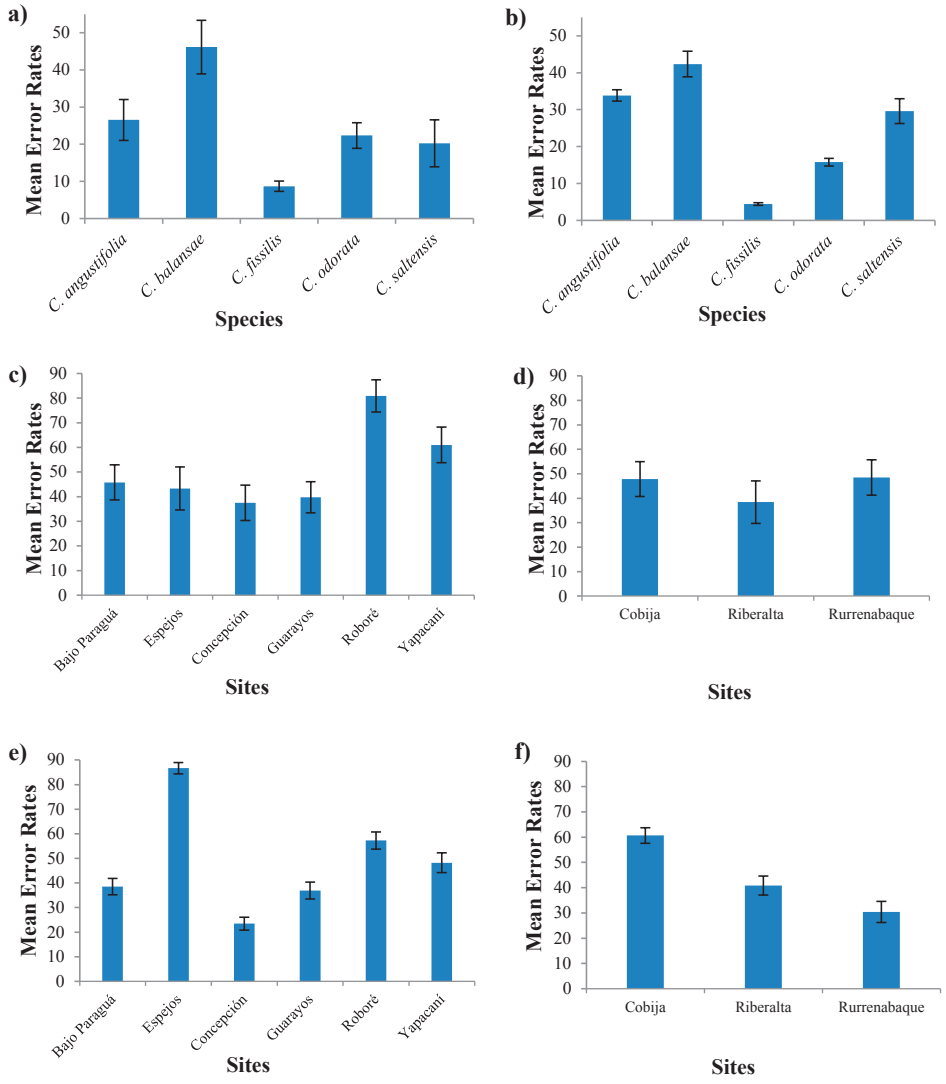


Figure S.2.1 Mean error rates of a) the Kernel Discriminant Analysis and b) Random Forest analysis for species analyses. Mean error rates for c) Kernel Discriminant Analysis per site for *C. fissilis* and d) Kernel Discriminant Analysis per site for *C. odorata*, e) Random Forest analyses per site for *C. fissilis* and f) Random Forest analyses per site for *C. odorata*. The whiskers show the standard error of the data.

Table S.2.1 Confusion matrix of species discrimination and frequency of species (%) in each randomized classification table using KDA

KDA		*C. angustifolia*	*C. balansae*	*C. fissilis*	*C. odorata*	*C. saltensis*	Total
C. angustifolia	Total samples	232	23	20	20	20	315
	%	73.7	7.3	6.3	6.3	6.3	100
C. balansae	Total samples	44	112	44	12	10	222
	%	19.8	50.5	19.8	5.4	4.5	100
C. fissilis	Total samples	32	18	1107	56	0	1213
	%	2.6	1.5	91.3	4.6	0	100
C. odorata	Total samples	1	0	118	432	11	562
	%	0.2	0	21.0	76.9	2.0	100
C. saltensis	Total samples	10	2	0	33	143	188
	%	5.3	1.1	0	17.6	76.1	100

Table S.2.2 Confusion matrix of species discrimination and frequency of species (%) in each randomized classification table using Random Forest

Random Forest		*C. angustifolia*	*C. balansae*	*C. fissilis*	*C. odorata*	*C. saltensis*	Total
C. angustifolia	Total samples	796	1	236	0	162	1195
	%	66.6	0.1	19.7	0	13.6	100
C. balansae	Total samples	92	470	177	54	0	793
	%	11.6	59.3	22.3	6.8	0	100
C. fissilis	Total samples	4	4	4752	211	0	4971
	%	0.1	0.1	95.6	4.2	0	100
C. odorata	Total samples	0	0	369	2010	2	2381
	%	0	0	15.5	84.4	0.1	100
C. saltensis	Total samples	91	0	13	81	475	660
	%	13.8	0	2.0	12.3	72.0	100

Table S.2.3 List of the 15 most important chemical compounds obtained from 50,000 runs of Random Forests. The numbers are the mass-to-charge ratios (*m/z*).

m/z	% of runs including the compound	Molecular formula	Tentative assignments
501.278	100	$C_{28}H_{38}O_8$ -H	3,7-Dideacetylkhivorin
500.265	100	$C_{28}H_{34}O_7$ +NH_4	Gedunin
484.245	100	$C_{27}H_{34}O_9$ – H_2O	Cedrodorin
483.244	100	$C_{28}H_{34}O_7$ +H	Gedunin
469.344	99	$C_{27}H_{32}O_7$ +H	Mexicanolide
528.412	92	-	-
451.337	92	$C_{27}H_{32}O_7$ - OH	Mexicanolide
229.200	84	-	-
227.095	83	$C_{15}H_{24}$ +Na	delta-Cadinene
357.136	79	$C_{21}H_{24}O_5$ +H	-
470.335	74	$C_{27}H_{36}O_8$ –H_2O	Swiemahogin A
452.307	74	$C_{27}H_{34}O7$ –H_2O	Methyl angolensate
527.418	71	$C_{29}H_{36}O_{10}$ -OH	6-Acetoxycedrodorin
471.347	67	$C_{27}H_{34}O_7$ +H	Methyl angolensate
507.399	58	$C_{30}H_{36}O_7$ -H	Mahonin

Table S.2.4 Confusion matrix of sites discrimination and frequency of sites (%) in each randomized classification table using KDA in *C. fissilis*

KDA		Bajo Paraguá	Espejos	Concepción	Guarayos	Roboré	Yapacaní	Total
Bajo Paraguá	**Total samples**	101	0	1	0	82	25	209
	%	48.3	0	0.5	0	39.2	12.0	100
Espejos	**Total samples**	0	83	39	30	0	3	155
	%	0	53.5	25.2	19.4	0	1.9	100
Concepción	**Total samples**	0	0	144	49	17	43	253
	%	0	0	56.9	19.3	6.7	17.0	100
Guarayos	**Total samples**	0	2	95	157	4	6	264
	%	0	0.8	36.0	59.5	1.5	2.3	100
Roboré	**Total samples**	120	0	40	0	36	17	213
	%	56.3	0	18.8	0	16.9	8.0	100
Yapacaní	**Total samples**	48	22	8	0	50	78	206
	%	23.3	10.7	3.9	0	24.3	37.9	100

Table S.2.5 Confusion matrix of sites discrimination and frequency of sites (%) in each randomized classification table using Random Forest in *C. fissilis*

Random Forest		Bajo Paraguá	Espejos	Concepción	Guarayos	Roboré	Yapacaní	Total
Bajo Paraguá	**Total samples**	507	0	0	0	235	65	807
	%	62.8	0	0	0	29.1	8.1	100
Espejos	**Total samples**	0	68	216	181	1	3	469
	%	0	14.5	46.1	38.6	0.2	0.6	100
Concepción	**Total samples**	0	12	805	221	0	0	1038
	%	0	1.2	77.6	21.3	0	0	100
Guarayos	**Total samples**	0	11	342	649	0	0	1002
	%	0	1.1	34.1	64.8	0	0	100
Roboré	**Total samples**	369	0	13	0	361	66	809
	%	45.6	0	1.6	0	44.6	8.2	100
Yapacaní	**Total samples**	186	0	17	0	152	420	775
	%	24.0	0	2.2	0	19.6	54.2	100

Table S.2.6 Confusion matrix of sites discrimination and frequency of sites (%) in each randomized classification table using KDA in *C. odorata*

KDA		Cobija	Riberalta	Rurrenabaque	Total
Cobija	**Total samples**	107	23	80	210
	%	51.0	11.0	38.1	100
Riberalta	**Total samples**	6	114	74	194
	%	3.1	58.8	38.1	100
Rurrenabaque	**Total samples**	34	69	93	196
	%	17.3	35.2	47.4	100

Table S.2.7 Confusion matrix of sites discrimination and frequency of sites (%) in each randomized classification table using Random Forest in *C. odorata*

Random Forest		Cobija	Riberalta	Rurrenabaque	Total
Cobija	**Total samples**	319	182	288	789
	%	40.4	23.1	36.5	100
Riberalta	**Total samples**	181	485	137	803
	%	22.5	60.4	17.1	100
Rurrenabaque	**Total samples**	112	117	579	808
	%	13.9	14.5	71.7	100

Table S.2.8 List of the most important chemical compounds obtained from 50,000 of Random Forests to identify site of origin for *C. fissilis* and C. *odorata*. The numbers are the mass-to-charge ratios (*m/z*).

C. fissilis				***C. odorata***			
m/z	**% of runs including the compound**	**Molecular formula**	**Tentative assignment**	***m/z***	**% of runs including the compound**	**Molecular formula**	**Tentative assignment**
149.123	100	-	-	468.307	97	-	-
122.075	100	-	-	467.344	96	-	-
121.067	100	-	-	527.418	88	$C_{29}H_{36}O_9$ - H	Methyl 3beta-acetoxy-6-hydroxy-1-oxomeliac-14-enoate
109.098	100	-	-	673.281	77	$C_{35}H_{46}O_{14}$ - OH	Meliacarpinin D
279.165	100	-	-	583.22	77	-	-
123.044	99	-	-	81.035	63	-	-
280.164	97	-	-	486.349	58	-	-
274.112	96	-	-	99.044	58	-	-
274.5	95	-	-	192.142	58	-	-
150.072	94	-	-	470.335	50	$C_{27}H_{34}O_7$	Methyl angolensate
95.087	84	-	-	117.053	48	-	-
135.103	83	$C_{10}H_{14}$ +H	p-Cymene	303.449	44	-	-
206.201	79	-	-	528.412	43	-	-
104.069	53	-	-	469.344	41	$C_{27}H_{32}O_7$ +H	Mexicanolide
275.276	50	-	-				
81.035	47	-	-				
379.292	44	-	-				
204.186	40	$C_{15}H_{26}O$ – H_2O	T-Muurolol				

Chapter 3

Isotopic characterization of *Cedrela* species to verify regional provenance of Bolivian timber

Kathelyn Paredes-Villanueva, Arnoud Boom, Jente Ottenburghs, Frans Bongers, and Pieter A. Zuidema

Published with the doi: 10.3959/2021-17, and the corrections made.

Abstract

Illegal trade of tropical timber leads to economic and biodiversity losses worldwide. To date, the most common type of fraud concerns false declarations sites of origin based on falsifiable documents. With concerns about sustainable provenancing of tropical timber, there is a need for independent tools to check the source of timber and verify compliance with international and national regulations. We evaluated the potential for one of these tools – isotopic tracing – for identifying four *Cedrela* species: *C. balansae*, *C. fissilis*, *C. odorata*, and *C. saltensis*; and for identifying geographic origin for two of the most intensively exploited tropical timbers: *C. fissilis* and *C. odorata.* We studied differences in $\delta^{13}C$ and $\delta^{18}O$ stable isotope signatures between 11 forest sites (163 trees sampled) distributed to cover species distributions within Bolivia. For most samples (90% of the sample set) we quantified isotope fractions of 10-year bulk samples; but we also evaluated annual fluctuations in isotopic signature for the last 10 years. We assessed the relationship between the stable isotopes composition and precipitation and altitude, but did not find significant correlations between annual data and these variables suggesting that they may not always be the limiting factor of wood composition in the sampled trees. However, 10-year-bulk data did show higher positive correlations with altitude and higher negative correlations with rainfall suggesting that long-term and more stable data better represented site-specific isotopic composition. To explore the isotopic site differentiation a Kernel Discriminant Analyses was used. In spite of the strong influence of climate and elevation on isotopic values, KDA results showed a low discrimination success: 37.5% accuracy for *C. odorata* and 29.5% for *C. fissilis* sites. This suggests that discrimination of geographical origin is not possible due to low differentiation –and highly variable annual isotopic data– among sites because of species and site-specific isotopic imprints. Based on our findings we recommend to assess which variables – besides rainfall and altitude – might be controlling isotopic signatures before considering environmental variables as proxies of stable isotopes in trees. Discrimination potential could be increased by adding other isotopes proxies or combining them with rare earth and trace elements.

Keywords: Illegal logging, *Cedrela*, stable isotopes, discriminant analysis, tropical timber

Introduction

Illegal logging practices lead to loss of natural resources, community conflicts and economic problems. The economic problems arise by the distortion of local and global markets that illegal low-priced timber cause (ABT Associates Inc., 2006; Blaser, 2010). Illegal timber trade threatens sustainable management of tree species and forest areas. To date, the most common type of illegal timber trade is the fraud concerning false declarations of sites of origin based on falsifiable documents. With increasing sustainability concerns, there is a need for independent tools to verify the source of timber in connection with (inter)national regulations (Degen and Fladung, 2007). A technique based on the intrinsic characteristics of the wood would help confirming the origin of the timber more reliably and would help combatting illegal timber trade.

In recent years, stable isotopic composition of biological materials has proven to be useful to trace the geographic origin. Isotopic measurements were first introduced for wine provenance and adulteration tests (Versini *et al.*, 1997) and were further applied to milk, meat, honey, juices, spirits and other foods (Rossmann, 2001; Oulhote *et al.*, 2011). Recently, they have also been applied to determine drugs origin at a regional and country level (Ehleringer *et al.*, 2000; Booth *et al.*, 2010). Although their application in tracing timber has already been established (Förstel and Hützen, 1983; Boner *et al.*, 2007; Kagawa and Leavitt, 2010; Förstel *et al.*, 2011; Vlam *et al.*, 2018), the spatial resolution of isotopic tracing is still unknown for most tropical timbers.

For assessing the spatial resolution of timber origin identification, we selected *Cedrela* species, an intensively harvested and traded tropical timber (Richter and Dallwitz, 2000) that is strongly affected by illegal logging. It has been traded as roundwood, lumber, veneer and imported in small quantities in Europe and larger quantities in America (Wagenführ, 2007). Overexploitation of the species has caused the decline of *Cedrela* populations in recent years (Mostacedo and Fredericksen, 1999; Mostacedo and Fredericksen, 2001). To support protection and regulation the Convention on

International Trade in Endangered Species of Wild Fauna and Flora (CITES) listed *Cedrela* species in Appendix III (Compt and Christy, 2008).

Several members of this genus are distributed throughout most of tropical America and the Caribbean islands (Wagenführ, 2007) and Bolivia harbours six species. These species can be found along different climatic and environmental gradients, from moist to dry tropical forests, and along a wide altitudinal range (Mostacedo *et al.*, 2003; Navarro, 2011; Navarro-Cerrillo *et al.*, 2013). In this study we covered this wide range of altitudes and rainfall across Bolivia to assess whether the isotopic composition (based on $\delta^{13}C$ and $\delta^{18}O$) of *Cedrela* populations can be used to differentiate between sites of origin.

Isotopes are the different forms or mass numbers of a single chemical element due to different numbers of neutrons in the nucleus. Some isotopes are stable while others are unstable –extremely high numbers of neutrons are unable to be held together in their nuclei- and radioactive, releasing particles and energy to decay into a more stable form (Cook *et al.*, 2018). Among the most important biological stable isotopes are carbon ($^{13}C/^{12}C$) and oxygen ($^{18}O/^{16}O$). Lighter isotopes (with less neutrons) form weaker chemical bonds, are more chemically active and evaporate/diffuse quicker than heavy isotopes. When isotopes are partitioned differently between two phases or two substances, the enrichment or higher abundance of one isotope relative to another in a chemical or physical process is called fractionation (Cook *et al.*, 2018). Plants convert CO_2 in carbohydrates during photosynthesis. During this process in trees, the Rubisco enzyme preferentially binds $^{12}CO_2$ and thus discriminates against $^{13}CO_2$. While photosynthesis drives ^{13}C values in plant tissue process, evapotranspiration and composition of source water taken up by the trees determine $\delta^{18}O$ values in plant tissue. The fractionation of $\delta^{18}O$ can also take place during the water cycle, through partitioning of precipitation waters in the system and the return of the water to the atmosphere by evapotranspiration. Trees record the results of fractionation in their tissue and tree rings, depending on environmental variations (Helle and Schleser, 2004).

The isotopic fractionation varies according to environmental conditions (McCarroll and Loader, 2004) and topographic and regional climatic regimes (Förstel and Hützen,

1983). Therefore we expect that $\delta^{13}C$ in plants will be determined by the CO_2 fixation pathway (Sternberg *et al.*, 1984) and $\delta^{18}O$ isotopic variation will be related to rainfall regimes (Dansgaard, 1964) and altitude (Gonfiantini *et al.*, 2002). Moreover, the patterns of oxygen and carbon isotopic spatial variation might allow us to match these properties to *Cedrela* samples of unknown origin at a regional level. We also expect that the highest differences of isotopes ratios will be found on the sites with higher altitudes where *Cedrela* is distributed in comparison to the lowlands where precipitation variations are more homogeneous. Our research questions are: (1) At what scale Bolivian *Cedrela odorata* and *C. fissilis* populations differ in their oxygen and carbon stable isotopes characteristics?; (2) To what extent this differentiation level will allow to identify *Cedrela* species and track timber origin at different spatial resolution?; (3) How can the relationship between rainfall, altitude and the isotopic ratio be used to predict the likelihood of belonging to the conditions of specific sites? To address these questions we collected *Cedrela* samples across Bolivia and assessed the isotopic variation in relation to rainfall, altitude and site of origin.

Methodology

Sampling sites

We performed the species identification analysis based on four *Cedrela* species: *C. balansae*, *C. fissilis*, *C. odorata*, and *C. saltensis*; and the origin identification based on two *Cedrela* species: *C. fissilis* and *C. odorata*. Samples were randomly collected from eleven sites throughout their distribution in Bolivia (Figure 3.1). A total of 163 trees were randomly selected and 2 wood cores were sampled from each individual tree. All the core samples were collected at breast height (130 cm) using a 5 mm diameter increment borer (Haglöf). Individuals of large/dominant trees per site were selected for extracting the wood samples as there is a height gradient for leaf $\delta^{13}C$ values associated with light intensity and humidity (Ometto *et al.*, 2002). For the origin identification of *C. fissilis* and *C. odorata*, minimum sample distance, ranging from 3 to 676 m among trees within a population and from 78 to 421 km among sites, were considered to maximize the

sampling coverage across the country. The maximum distance between sample sites for *C. fissilis* was 501 km and for *C. odorata* 613 km. In addition, sampling was carried out between sites along roads or/and in villages to obtain a gradual variation between sampled populations. This approach allowed us to assess the discrimination within sites and to compare the isotopic composition at different distance of samples.

In addition, botanical samples were collected for species confirmation when identification in the field was not possible. This was done for 27% of the sampled trees. The voucher preparation and confirmation of the species based on herbarium collections were carried out by an experienced botanist at the Museo de Historia Natural Noel Kempff Mercado (Bolivia). Climate data were obtained from the reporting agency of Servicio Nacional de Meteorología e Hidrología (SENAMHI); Table 3.1).

Table 3.1 *Cedrela* species and number of samples collected from 11 study sites in Bolivia

Species	Sites	Sample size for 10-year bulk analysis	Sample size for annual analysis	Botanical samples	Average altitude of samples (m.a.s.l.)	Annual precipitation during last 10 years of tree growth (mm)
C. balansae	Villamontes	10		3	588	-
C. fissilis	Bajo Paraguá	21	3	2	292	-
	Concepción	21	3	*	468	1044.9
	Espejos	21	3	*	481	1552.0
	Guarayos	21	4	13	224	1420.8
	Roboré	21	3	13	635	1100.3
	Yapacaní	10		5	318	1860.2
C. odorata	Cobija	10		*	274	1998.8
	Riberalta	10		*	145	1795.8
	Rurrenabaque	10		4	309	2057.2
C. saltensis	Monteagudo	8		4	1595	-
Total		**163**	**16**	**44**		

*No botanical samples were collected, but identification is based on previous collections.

- No climate data available for the studied period during the analyses.

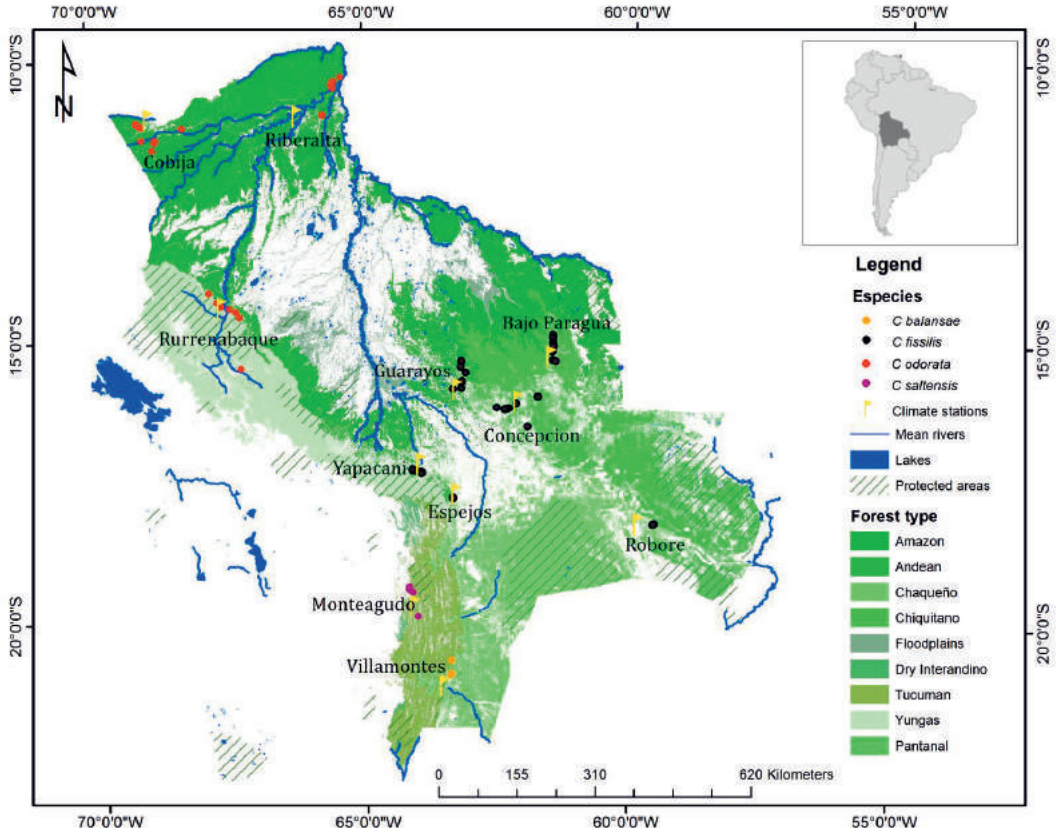


Figure 3.1. Sampling sites of *Cedrela* species and closest climate stations in Bolivia. Forest cover: Autoridad de Bosques y Tierra, 2015.

Samples were air dried and carefully polished using sandpaper with grit sizes ranging from 26.8 to 425 µm (Orvis and Grissino-Mayer, 2002). *Cedrela* species exhibited visible tree rings, and they were composed of ring-porous wood that had parenchyma bands in the boundaries as found by previous studies (Worbes, 1999; Paredes-Villanueva *et al.*, 2016). Tree-ring boundaries were identified on the increment cores and marked with a lead pencil. In each core, tree-rings were identified using a binocular microscope (Leica MZ 125) coupled to a cold light source; and assigned to the calendar year in which their growth started: from September of the current year to August of the following year according to Schulman (1956). *Cedrela* annual ring formation has been demonstrated in

previous studies (Worbes, 1999; Dünisch *et al.*, 2002; Brienen and Zuidema, 2005; Bräuning *et al.*, 2009; Paredes-Villanueva *et al.*, 2016). Annual ring widths within each increment core were accurately measured, compared, and matched using TSAP/LINTAB (Frank Rinn, Heidelberg, Germany) software/hardware combination to a resolution of 0.01 mm. Missing and false rings, suggested by cross-dated samples, were identified visually and revised by using the same software.

Isotope analyses

For the analysis of $\delta^{13}C$ and $\delta^{18}O$ from wood, flakes were extracted from a bulk of the last 10 rings formed in each sampled tree collected from Bajo Paraguá, Concepción, Espejos, Guarayos, Roboré and Yapacaní using a gauge and microtome. Flakes from annual tree rings were also sampled from 3-4 trees per site (10% of the sample set) for all the sites except Yapacaní. Samples were collected from widely distributed trees in each site. Cellulose was extracted from the organic wood samples using the adapted Jayme-Wise method (Wieloch *et al.*, 2011) as described in Vlam *et al.* (2018). Cellulose was then oven-dried at 50°C. Subsequently, for $\delta^{13}C$, cellulose samples were combusted in a continuous flow mode with an element analyzer; and for $\delta^{18}O$, cellulose was pyrolyzed in a glassy carbon, both coupled to a mass spectrometer (Sercon Hydra 20-20). The values of the isotopes ratios ($\delta^{13}C$ and $\delta^{18}O$) per site were represented in parts per thousand (‰) according to the international Vienna Standard Mean Ocean Water (VSMOW) for oxygen and Pee Dee Belemnite (PDB) standards for carbon.

Statistical analysis

The resulting isotopic values were evaluated in relation to local environmental variables, annual precipitation and altitude. For this purpose, Pearson correlation analysis was performed on the annual measurements in *C. fissilis* and on the 10-year-bulk measurements for *C. fissilis* and *C. odorata*. For the pooled data analysis, both annual precipitation and annual isotopic data were averaged by the year of measurement and included in the 10-year-bulk dataset. Data from Bajo Paraguá site were not included in this correlation analysis since there was no data available for the studied period.

We performed several data analysis steps with the $\delta^{13}C$ and $\delta^{18}O$ data to evaluate if they can differentiate and identify *Cedrela* sample sites (research question 1: site differentiation; research question 2: sample origin identification). The data analyses were divided over annual and 10-year-bulk sample sets. We first evaluated the $\delta^{13}C$ and $\delta^{18}O$ annual data by averaging the isotopic values and weighing them with their corresponding ring width. As the years corresponding to wood formation are generally unknown in confiscated timber, both approaches allowed us to evaluate if the annual variation has any influence on the general isotopic patterns across the study sites. Weighting correction would be preferred over simple ring averaging if there would be an effect of tree ring width on annual isotopic values. After selecting the type of annual data (weighted or averaged), these were used together with the 10-year-bulk data as input for the discriminant analysis (KDA). Scatterplots were performed using ggplot2 (Wickham, 2016) and ggpubr (Kassambara, 2018) packages in R version 3.4.3 (R Development Core Team, 2017).

Second, the site differentiation potential was assessed by performing a discriminant analysis on the $\delta^{13}C$ and $\delta^{18}O$ data to classify the sites and predict sample assignment. For these analyses we used Kernel Discriminant Analyses with package ks 1.10.5 (Duong, 2007, 2017) in R version 3.3.3 (R Development Core Team, 2017). The discriminant analysis was based on randomized samples and variables in every run. The classification results allowed us to assess the site differentiation potential measured by the frequency in which each species is assigned to either a correct or erroneous class (research question 1: site differentiation).

KDA separates the samples based on an *a priori* classification assignment (to specific site classes) and looks for the optimal non-linear combination of variables (component loadings) for maximal separation of the samples in a two dimensional space (Baudat and Anouar, 2000). KDA's learning algorithm uses Bayes discriminant rule (Duong, 2007) and needs to be trained in order to assess its discrimination power. Therefore, our data was split in two sets: 80% for training and 20% for testing the model. Smoothed Cross Validation (SCV) error (Duong, 2007) was applied to test correctness of the site

assignments (research question 2: sample origin identification). KDA analyses generated confusion matrices showing the frequency at which each site was wrongly classified. After 100 randomizations runs, we checked with what site a single sample origin could be confused most. The classification error was expressed in percentage (%) where a cross validation error of 0% indicates that all the samples were correctly assigned and that the method separated the sites correctly in the dataset. Finally, the average errors per site identification across the 100 runs were obtained together with their corresponding standard deviation. Similarly, we followed these steps considering only annual samples to evaluate the consistency of discrimination accuracy with the 10-bulk sample data.

Results

To test the potential of $\delta^{13}C$ and $\delta^{18}O$ for identification of timber origin, the *C. fissilis* and *C. odorata* samples were assessed in two groups (1) annual samples from each of the ten most recent tree rings per tree and (2) bulk of mixed samples from the last ten dated tree rings per tree. Kernel analysis (KDA) was applied for the assessment of isotopic discrimination. In addition, for validating the power of both discriminant analyses, 80% of the dataset was split for training and 20% for testing the KDA model. Results were assessed based on their classification error per species per site.

Relationship between $\delta^{13}C$ and $\delta^{18}O$ and environmental variables

The relationship between precipitation and latitude and stable isotopes ($\delta^{13}C$ and $\delta^{18}O$) varied depending on the level of resolution analyzed. Annual isotopic values of both $\delta^{13}C$ and $\delta^{18}O$ in C. *fissilis* showed no significant correlation with precipitation (Table 3.2, Supplementary Data Figure S.3.2) and altitude (Table 3.2). However, when the 10-most-recent years were pooled this species showed higher and more significant correlations. Correlations between 10-year-bulk samples of $\delta^{13}C$ and precipitation and altitude were significant. Only correlation between 10-year-bulk samples of $\delta^{18}O$ and precipitation showed slightly lower values. Similarly, *C. odorata* showed a high correlation with both

environmental variables with lower values between annual precipitation and $\delta^{18}O$ bulk samples.

Table 3.2 Exploratory correlation analysis between stable isotopes ($\delta^{18}O$ and $\delta^{13}C$) in *C. fissilis* and *C. odorata* and environmental variables (annual precipitation and altitude). Analyses were separated for annual and 10-year-bulk samples.

	Precipitation		**Altitude**	
	Tree $\delta^{18}O$	**Tree $\delta^{13}C$**	**Tree $\delta^{18}O$**	**Tree $\delta^{13}C$**
Annual samples				
Cedrela fissilis	r = 0.02 t_{128} = 0.23 p = 0.82	r = 0.02 t_{128} = 0.24 p = 0.81	r = 0.20 t_{158} = 2.61 p = 0.01	r = 0.10 t_{158} = 1.32 p = 0.19
10-year bulk samples				
Cedrela fissilis	r = -0.27 t_{92} = -2.66 p = 0.01	r = -0.48 t_{92} = -5.31 p = 7.50e-07	r = 0.44 t_{92} = 4.73 p = 8.11e-06	r = 0.41 t_{92} = 4.32 p = 3.95e-05
Cedrela odorata	r = -0.27 t_{92} = -2.66 p = 0.01	r = -0.48 t_{92} = -5.31 p = 7.50e-07	r = 0.44 t_{92} = 4.73 p = 8.11e-06	r = 0.41 t_{92} = 4.32 p = 3.95e-05

Isotopic annual variation on the 10 most recent rings per tree

Isotopic ratios could vary between years. In our data set, the standard deviations for the annual samples ranged from 0.26 to 0.80 (Supplementary Data Table S.3.1). To deal with this variation and as the years corresponding to wood formation are generally unknown in confiscated timber, we explored the transformation of annual data into 10-year-bulk data by averaging or weighting. We also assessed whether this transformation had an effect on the dominant signals across the study sites. There was a significant correlation between weighed and averaged annual isotopic data for $\delta^{18}O$ (Pearson correlation, r = 0.91, t_{14} = 8.44, p = <0.001 and $\delta^{13}C$ (Pearson correlation, r = 0.9, t_{14} = 7.76, p = <0.001; Figure 3.2). Moreover, both transformations followed the same isotopic pattern as the 10-year-bulk samples (Supplementary Data Figure S.3.1). Weighting correction would be preferred over averaging if there would be an effect of tree ring width on annual isotopic values. However, there was no significant correlation between tree ring width and

annual isotopic values for $\delta^{13}C$ (Pearson correlation, r = -0.09, t_{158} =-1.20, p = 0.23) nor $\delta^{18}O$ (Pearson correlation, r = -0.13, t_{158} = -1.63, p = 0.11). Therefore, we decided to use the averaged annual values combined with the 10-year-bulk data set for further analyses.

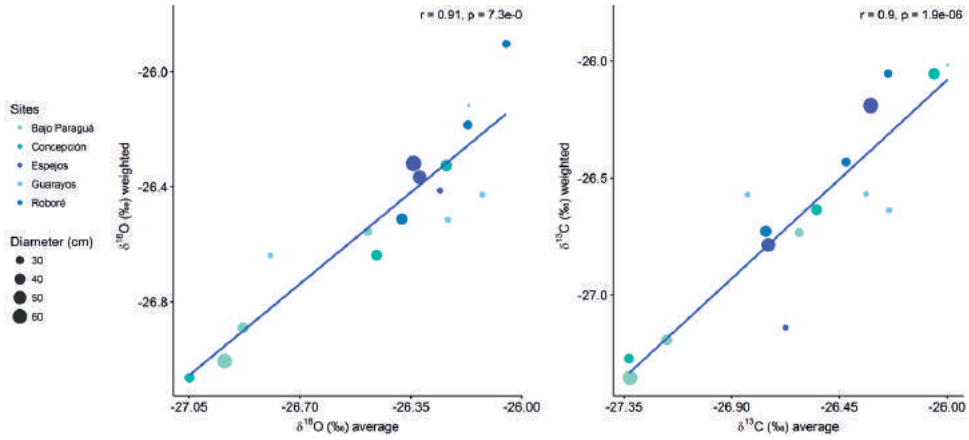


Figure 3.2 Scatterplots of weighted and averaged data of $\delta^{18}O$ and $\delta^{13}C$ stable isotopes for *C. fissilis* annual samples

In summary, the distribution of the isotopic values, represented by averaged plus 10-year-bulk samples, were clustered together and species specific (Figure 3.3). Compared to *C. odorata*, $\delta^{13}C$ and $\delta^{18}O$ values were higher for *C. fissilis* and mainly fluctuated around -28‰ and -26‰ (Figure 3.2). Extreme values in this species ranged from -28‰ to -25‰ for $\delta^{18}O$ and from -24‰ to -29‰ for $\delta^{13}C$ stable isotopes. *C. odorata* values mainly ranged from -29.5‰ to -27.5‰. Its extreme values ranged from -29‰ to -26‰ for $\delta^{18}O$ and from -26‰ to -31‰ $\delta^{13}C$ stable isotopes.

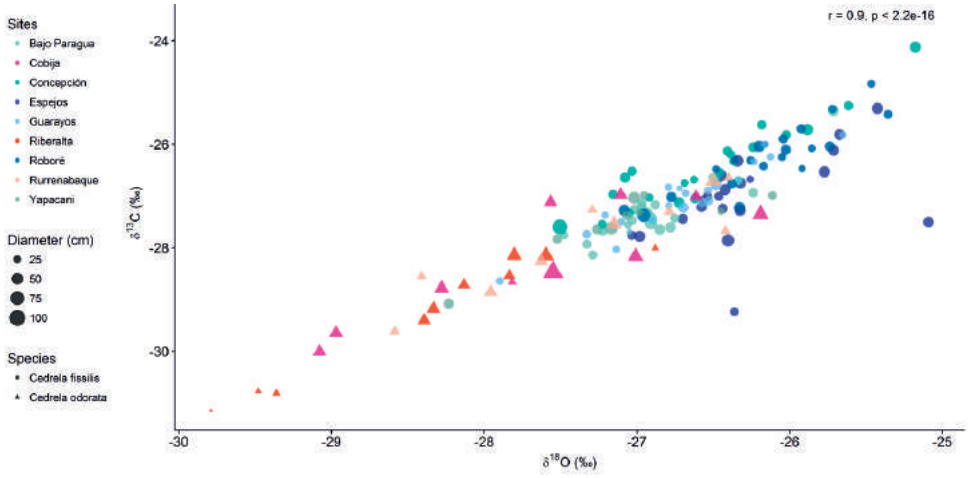


Figure 3.3 Scatterplot of $\delta^{13}C$ and $\delta^{18}O$ stable isotopes for 10-year-bulk samples of *C. fissilis* (triangles) and *C. odorata* (circles)

As size and age could have an effect on isotopic ratios (Brienen *et al.*, 2017), we assessed correlation between diameter at breast height (DBH) and stable isotopes (Supplementary Data Table S.3.2). This correlation was not significant in *C. odorata* ($\delta^{18}O$ Pearson correlation, r = 0.31, t_{28} = 1.73, p = 0.09; $\delta^{13}C$ Pearson correlation, r = 0.31, t_{28} = 1.71, p = 0.09) with only Riberalta reporting higher correlation ($\delta^{18}O$ Pearson correlation, r = 0.53, t_8 = 1.75, p = 0.12; $\delta^{13}C$ Pearson correlation, r = 0.67, t_8 = 2.57, p = 0.03). Correlation between DBH and stable isotopes was not significant in *C. fissilis* ($\delta^{18}O$ Pearson correlation, r = -0.03, t_{113} = -0.33, p = 0.74; $\delta^{13}C$ Pearson correlation, r = -0.06, t_{113} = -0.61, p = 0.54) with only Roboré reporting higher and significant correlation ($\delta^{18}O$ Pearson correlation, r = -0.55, t_{19} = -2.83, p = 0.01; $\delta^{13}C$ Pearson correlation, r = -0.56, t_{19} = -2.95, p = 0.01).

Kernel Discriminant Analysis (KDA) for 10-year-bulk samples per tree for site identification

The classification matrix based in 10-year-bulk samples per tree of *C. fissilis* compared true and estimated sites of origin showing the correct and wrong assignments of samples to each site (Table 3.3). Each of the 100 runs provided a mean SCV error across all sites. These wrong estimated sites averaged to a total classification error of 70.5%. Yapacaní

was the site with the least discriminative isotopic composition: only 0.3% was correctly assigned to their corresponding site of origin and most of them wrongly assigned to Bajo Paraguá (60.1%). On the other hand, samples originated on Bajo Paraguá were mostly correctly assigned (56.7%) to their site of origin. Part of these samples was wrongly assigned to Guarayos site (21.1%) and to the rest of the sample sites (4-7%).

Table 3.3 Classification matrix for each sample site for 10-year-bulk samples from *C. fissilis* species. Values in boxes indicate the assignment of sample origin averaged over 100 runs and their standard deviation (SD). Blue boxes indicate the correct assignments of samples.

Site classification (%)	Bajo Paraguá (est.)	Concepción (est.)	Espejos (est.)	Guarayos (est.)	Roboré (est.)	Yapacaní (est.)
Bajo Paraguá (true)	**56.7**	5.1	7.8	21.1	4.1	5.3
SD	27.7	12.0	12.6	25.1	9.7	10.9
Concepción (true)	19.7	**17.1**	4.0	21.2	35.2	2.9
SD	20.9	19.3	13.0	19.9	23.5	9.3
Espejos (true)	20.4	1.0	**18.8**	34.0	23.0	2.8
SD	22.6	4.8	23.6	29.0	25.7	8.5
Guarayos (true)	33.6	2.9	5.1	**33.2**	20.9	4.3
SD	26.2	8.0	12.5	24.6	23.9	10.2
Roboré (true)	22.6	10.8	13.9	9.5	**43.2**	0.0
SD	24.3	15.6	16.5	17.0	26.9	0.0
Yapacaní (true)	60.1	3.5	18.5	17.5	0.0	**0.3**
SD	36.4	12.1	26.9	29.2	0.0	3.3

For *C. odorata* sites, the classification matrix with true and estimated sites of origin averaged to a total classification error of 62.5%. Most of the samples originally from Riberalta were correctly assigned (72.3%) with some confusion with Cobija or Rurrenabaque (10-17%, Table 3.4). The latter two sites showed correct assignments below 36%. Samples from Rurrenabaque were almost equally classified to the three studied sites. The samples from Cobija were mostly wrongly assigned to Rurrenabaque (48.8%) and Riberalta (33.4%). Only 17.8% of the samples were correctly classified.

Table 3.4 Classification matrix for each sample site for 10-year-bulk samples from *C. odorata* species. Values in boxes indicate the assignment of sample origin averaged over 100 runs and their standard deviation (SD). Blue boxes indicate the correct assignments of samples.

Site classification (%)	Cobija (est.)	Riberalta (est.)	Rurrenabaque (est.)
Cobija (true)	**17.8**	33.4	48.8
SD	31.3	37.2	38.8
Riberalta (true)	10.3	**72.3**	17.4
SD	24.2	37.3	31.3
Rurrenabaque (true)	30.8	33.3	**35.9**
SD	37.3	35.3	35.1

Kernel Discriminant Analysis (KDA) for 10-year-bulk samples per tree for species identification

Isotopic data of $\delta^{13}C$ and $\delta^{18}O$ showed values grouped by species, especially those from *C. fissilis* and *C. odorata* (Figure 3.3 and Supplementary Data Figure S.3.3). KDA of isotopic data on four *Cedrela* species allowed discrimination of *C. fissilis* with high accuracy (93.8%, Table 3.5). Oppositely, lowest discrimination potential was showed in *C. odorata* with a mean error of 41%, *C. saltensis* with 88.3% and *C. balansae* with the highest error rate of 98.6%.

Table 3.5 Classification matrix from 10-year-bulk samples of *Cedrela* species. Values in boxes indicate the assignment of sample origin averaged over 100 runs and their standard deviation (SD). Blue boxes indicate the right assignments of samples.

Species classification (%)	*C. balansae* (est.)	*C. fissilis* (est.)	*C. odorata* (est.)	*C. saltensis* (est.)
***C. balansae* (true)**	**1.4**	57.5	23.3	17.8
SD	7.2	34.5	28.5	27.4
***C. fissilis* (true)**	2.8	**93.8**	2.8	0.6
SD	3.2	4.4	3.2	1.8
***C. odorata* (true)**	0.0	41.1	**58.9**	0.0
SD	0.0	21.3	21.3	0.0
***C. saltensis* (true)**	16.0	72.3	0.0	**11.8**
SD	29.1	35.3	0.0	28.0

Discussion

Stable isotopes prospective to trace timber origin

In this study, we assessed isotopic differentiation in *Cedrela* species and the potential to trace the location origin of *C. odorata* and *C. fissilis* wood in Bolivia. We also assessed the relationship between rainfall and altitude and $\delta^{13}C$ and $\delta^{18}O$ isotopes from *Cedrela*'s trees to predict the likelihood of belonging to the conditions of specific sites. Considering that isotopic fractionation varies according to environmental conditions (McCarroll and Loader, 2004) and topographic and regional climatic regimes (Förstel and Hützen, 1983), we hypothesized that the patterns $\delta^{13}C$ and $\delta^{18}O$ isotopes in *Cedrela* would mirror a synchronous spatial variation and allow us to trace the origin of the timber. Although our results showed high variation of $\delta^{18}O$ and $\delta^{13}C$ isotopes in annual samples, there were species specific isotopic signals in 10-year-bulk samples. Further analysis on the species distribution sites indicated a limited isotopic potential for tracing *Cedrela* timber origin: classification accuracy varied from 0.33 to 56.7% for *C. fissilis* and from 17.8 to 72.3% for *C. odorata*. This variation in classification accuracy can be explained by species specific signals or site specific conditions such as rainfall, altitude, light availability, soils and large scale climatic factors (e.g., El Nino-Southern Oscillation, ENSO) (Kurita *et al.*, 2009). Rather than relating a single variable to isotopic variation, we recommend to take the effects of all environmental conditions into account. Furthermore, the effects of these conditions on isotopes may differ between small and large scales.

One source of variation is related to small scale influences on $\delta^{18}O$ and $\delta^{13}C$ isotopes. Annual isotopic data can change every year given the different environmental conditions influencing tree development (McCarroll and Loader, 2004; West *et al.*, 2006). We expected that isotopic variation would be related to rainfall regimes (Dansgaard, 1964) and altitude (Gonfiantini *et al.*, 2002), given the heterogeneity of these variables in our sampling sites. However, we found no significant correlation between annual isotopic data and rainfall or altitude. Low correlations of rainfall with $\delta^{13}C$ and $\delta^{18}O$ in trees have also been found in previous studies (Poussart and Schrag, 2005; Cullen and Grierson,

2007; Schollaen *et al.*, 2013; Baker *et al.*, 2015). First, such low correlations can be explained by the annual switch of one environmental variable to another being more limiting during tree growth (Helle and Schleser, 2004; Paredes-Villanueva *et al.*, 2016). For example, given the fact that *Cedrela* is a relatively light sensitive tree (Brienen and Zuidema, 2006; Brienen *et al.*, 2010), light availability might show a greater effect on tree development compared to rainfall in some sites (Brienen *et al.*, 2010). The annual switching of environmental variables can be taken into account by analyzing annual isotopic data. However, in our study annual variation did not improve discrimination (data not shown), and even added noise. Second, low correlations between rainfall and isotopic variation might also depend on differences in scale (Kurita *et al.*, 2009). Studying isotopes on a small scale while the limiting factor works on a large scale can affect statistical analysis and consequent discrimination between sites. Therefore, a large-scale perspective will reflect variable characteristics of the sites of origin more reliably.

Large scale variations represented dominant and more constant patterns in our dataset when using the pooled annual samples. The 10-year-bulk samples showed a higher correlation across the study sites. In addition, our results showed different isotopic distributions for both species, with lower values for *C. odorata* (~28‰ $\delta^{18}O$, ~29‰ $\delta^{13}C$) compared to *C. fissilis* (~26‰ $\delta^{18}O$, ~27‰ $\delta^{13}C$). ^{13}C stable isotope composition in plants is affected by photosynthesis. In this process, isotope discrimination against ^{13}C from the atmosphere (≈-8‰) occurs during diffusion of CO_2 into de leaf (-4.4‰ fractionation) and its fixation into carbohydrates (-30‰ fractionation). The latter process occurs most commonly by the C_3 photosynthesis pathway where CO_2 binds to the enzyme Rubisco (biochemical cycle known as the Calvin Cycle) that preferentially binds to $^{12}CO_2$ (Helle and Schleser, 2004; Cook *et al.*, 2018). Under humid conditions trees have wider stomatal openings resulting in a reduction of $\delta^{13}C$ values while under dry conditions stomata might close resulting in increasing $\delta^{13}C$ values. This was in line with our results, $\delta^{13}C$ values from 10-year-bulk samples that ranged from -31‰ for *C. odorata* (which prefers moist conditions) to -24‰ for *C. fissilis* (which is more abundant in drier conditions).

Discrimination analysis in this study covered a spatial resolution of 3 m to 501 km for *C. fissilis* and 85 m to 613 km for *C. odorata* samples. Discrimination of sites presented wide ranges of classification accuracy: from 0.3% to 56.7% in *C. fissilis* and from 17.8% to 72.3% in *C. odorata* samples. The average accuracy was 29.5% for *C. fissilis* and 37.5% for *C. odorata* site classification. Similar results have been found for site identification of *Erythrophleum* species in Cameroon and Congo Republic with an average accuracy of 35% (Vlam *et al.*, 2018). The same study also found high variability in the classification accuracy ranging from 46% to 99% at lower spatial scales (14-216 km) (Vlam *et al.*, 2018). This suggests that some sites and species have greater potential to be traced than others.

While the potential of stable isotopes to trace origin has previously been demonstrated in timber species (Förstel and Hützen, 1983; Boner *et al.*, 2007; Kagawa and Leavitt, 2010; Förstel *et al.*, 2011; Vlam *et al.*, 2018), high resolution tracing of timber origin remains a challenge. In this study, a wide range of altitudes and rainfall regimes were included in the analysis: 224-635 m.a.s.l. of altitude and 1045-1860 mm of precipitation in *C. fissilis* sampled sites and 145-309 m.a.s.l. of altitude and 1796-2057 mm of precipitation in *C. odorata* sampled sites. In addition, we used averaged annual precipitation and elevation to assess if more dominant site characteristic will be mirrored in isotopic properties ($\delta^{13}C$ and $\delta^{18}O$) of trees. Despite significant correlations between these variables, low discrimination accuracies among sites suggest that other governing factors at larger scales, like soils properties and El Nino-Southern Oscillation (ENSO), influence site specific variations. Research on $\delta^{13}C$ and $\delta^{15}N$ of coca (*Erythroxylum coca*) from countries in South America (Ehleringer *et al.*, 2000) found significant ratio differences among the sample sites and attributed this to the differences of soils (for $\delta^{15}N$) and length of wet season (for $\delta^{13}C$). Soil characteristic showed effect on tree growth (Toledo *et al.*, 2011) in dry forests where soil water availability and nutrients seem to impact growth (Medina *et al.*, 1995; Oliveira-Filho *et al.*, 1998). This impact shows a spatial gradient (Murphy and Lugo, 1986; Ceccon *et al.*, 2006) that progressively shifts in wetter forests where light variation is more important for growth (Engelbrecht *et al.*, 2007; Brienen *et al.*, 2010). ENSO variability was also found to

influence isotopic composition in *Cedrela* species from Bolivia (Vuille and Werner, 2005; Brienen *et al.*, 2012; Baker *et al.*, 2015). Vuille and Werner (2005) found that $\delta^{18}O$ was significantly more enriched during the El Niño (reduction in precipitation) and more depleted during La Niña events. How stable isotopes composition varies will depend on the site and species-specific response to these different changing environmental factors (Helle and Schleser, 2004; Poussart and Schrag, 2005; van der Sleen *et al.*, 2017). Up to now, several stable isotopes analyses have been done in Bolivia (Brienen *et al.*, 2012; Nijmeijer, 2012; van der Sleen, 2014; Baker *et al.*, 2015; van der Sleen *et al.*, 2015) focusing on different species and sites. However, it is required to assess the species and site specific variation before performing discrimination analysis for timber tracking purposes.

Considerations for stable isotope applications in timber tracking

Some considerations have to be taken to analyze stable isotopes data regardless of whether samples are bulked: light effect, strength of precipitation-isotopes correlation and ontogenetic variation. First, if a species is light demanding, it may present different isotopic composition when the samples are collected at disturbed areas or when the timber sample analyzed was formed during clear opening events (Brienen *et al.*, 2010; van der Sleen *et al.*, 2014). In our case, samples were collected in both places, where there was previous intervention or use, and dense forests with closed canopy. However, variation/sensitivity of isotopes values depends on their response to what variable (s) control and limit growth (Helle and Schleser, 2004).

Second, one could ideally use isotope values from climate data as proxies, e.g. precipitation. *Cedrela* has shallow roots and we expected to find a correlation between $\delta^{18}O$ in trees and precipitation (Brienen *et al.*, 2012). However, in our study, precipitation did not show high variation among sites nor significant correlation with the isotopic ratios. This suggests that precipitation amount may not always be the primary source explaining $\delta^{18}O$ values in our samples. Other possible oxygen sources – that can have effect on isotopic variation in plants – include humidity, soil moisture, groundwater (Aggarwal *et al.*, 2004). Therefore it is recommended to assess which of these variables

are controlling the isotopic signature under study before considering them as proxies of stable isotopes in trees. However, these analyses were outside the scope of our study.

Third, isotopic variation also depends on the age or size (ontogenetic development phase) underlying the long term trends (Broadmeadow *et al.*, 1992; McCarroll and Loader, 2004; van der Sleen *et al.*, 2015; Brienen *et al.*, 2016). We tested if this effect could have influenced the discrimination potential of our isotopes samples. However, the correlation between diameter at breast height (DBH) and stable isotopes was not significant in *C. odorata,* nor in *C. fissilis*. Removing extreme trees with small or large DBH (2 *C. odorata* samples with <15 and >70) did not help much as it improved mean discrimination accuracy by only 2.2%. Similarly, removing 4 *C. fissilis* samples with <20 and >70 improved mean discrimination accuracy by only 1.4%.

In addition, during the analysis of timber with unknown origin, estimating the possible period of tree harvesting will be a minimum requirement to correct for possible effects of the increasing CO_2 concentrations during post-industrial years which increases the amount of $\delta^{13}C$ that trees incorporated into their biomass (van der Sleen *et al.*, 2017). Finally, the lack of variation in $\delta^{13}C$ and $\delta^{18}O$ could be overcome by assessing multiple tree-isotopes proxies or combining them with rare earth and trace elements (English *et al.*, 2001; Kelly *et al.*, 2005; Joebstl *et al.*, 2010) to increase discrimination among sites.

Conclusion

Cedrela species have been overexploited the last couple of years. *Cedrela* species have been listed among the most important traded timbers and are widely distributed in Central and South America. In Bolivia, these species are found at locations with different climatic and environmental gradients, and our sampling covered a wide range of altitudes and rainfall regimes across Bolivia to assess whether the isotopic composition (based on $\delta^{13}C$ and $\delta^{18}O$) of *Cedrela* populations can be used to differentiate between sites of origin. Stable isotopes composition of wood has the potential to hold non-manipulative information on the geographic and taxonomic origin. This potentially

improves the verification of timber origin certificates in connection with regulations. We used two types of data to assess $\delta^{13}C$ and $\delta^{18}O$ variation in *C. fissilis* and *C. odorata* isotopic composition: annual and 10-year-bulk samples. Discrimination analysis (KDA) reveals no potential for discrimination of geographical origin. Sites assignment was low: 37.5% accuracy for *C. odorata* and 29.5% for *C. fissilis* sites. This suggests that discrimination of geographical origin in the region sampled is not possible due to low differentiation among sites dependent on the species and site-specific isotopic imprint. However, our results suggest that it is possible to distinguish species based on isotopes, especially *C. fissilis* with high accuracy (93.8%). We conclude that stable isotopes cannot help to identify sites provenance of *Cedrela* timber within Bolivia but show a great potential for discrimination of *C. fissilis* from the rest of the species.

Acknowledgements

We would like to thank the Museo de Historia Natural Noel Kempff Mercado for their support on the species identification. This research was financed by the NFP/Nuffic fellowship (the Netherlands). Fieldwork was logistically supported by Universidad Autónoma Gabriel René Moreno and financed by Alberta Mennega Stichting and The Rufford Foundation. This research was supported by the International Foundation for Science (IFS), Stockholm, Sweden, through a grant to Kathelyn Paredes Villanueva for a second campaign of fieldwork and laboratory analyses.

Supplementary Data

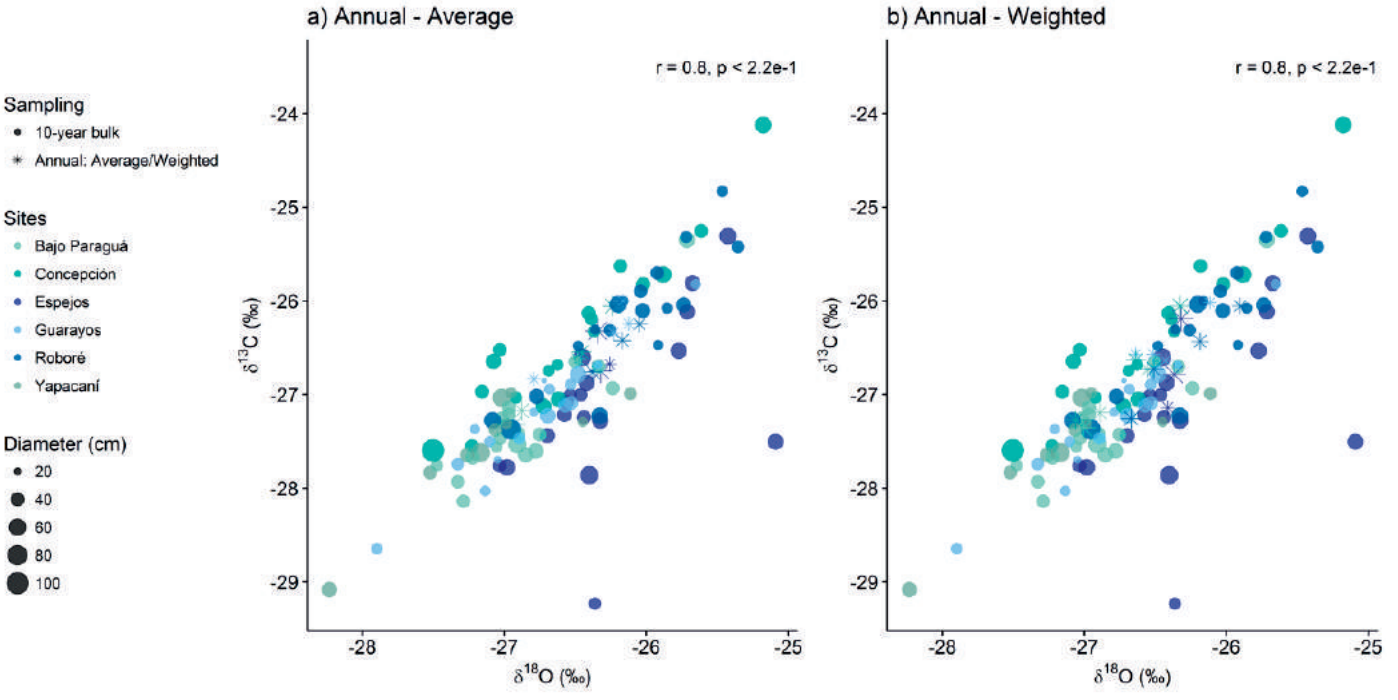


Figure S.3.1 Scatterplot of average and weighed data for $\delta^{13}C$ and $\delta^{18}O$ stable isotopes in *C. fissilis*

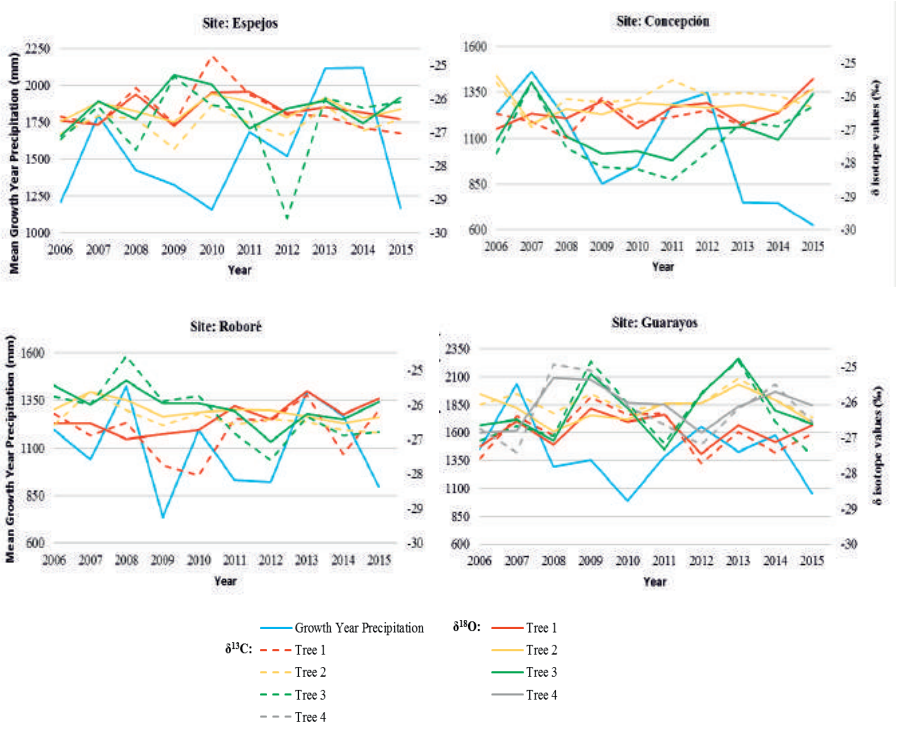


Figure S.3.2 Relationship between annual precipitation and annual samples of $\delta^{13}C$ and $\delta^{18}O$ stable isotopes in *C. fissilis*

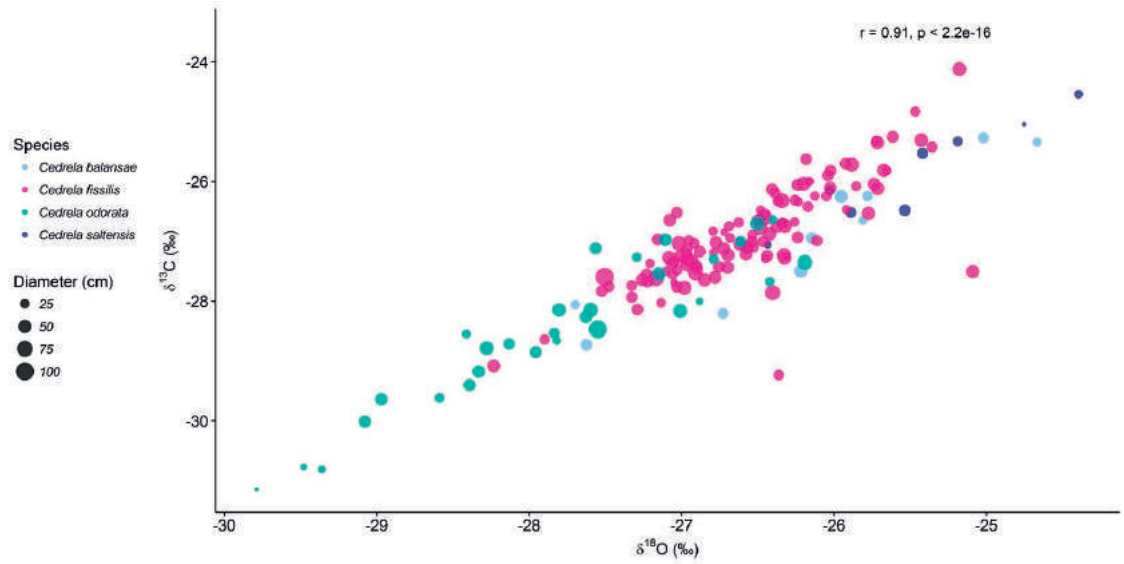


Figure S.3.3 Scatterplot of 10-year bulk samples of $\delta^{13}C$ and $\delta^{18}O$ stable isotopes in *Cedrela* species

Table S.3.1 Standard deviations (SD) of *C. fissilis* samples per tree and per site

Site	Sample	SD of $\delta^{18}O$ per tree	SD of $\delta^{18}O$ per site	SD of $\delta^{13}C$ per tree	SD of $\delta^{13}C$ per site
Bajo Paraguá	1	0,40	0,43	0,71	0,65
	2	0,30		0,41	
	3	0,58		1,19	
Concepción	4	0,36	0,64	0,61	0,81
	5	0,41		0,62	
	6	0,40		0,53	
Espejos	7	0,45	0,43	0,42	0,83
	8	0,40		0,39	
	9	0,75		0,93	
Guarayos	10	0,45	0,61	0,75	0,76
	11	0,26		0,35	
	12	0,50		0,84	
	13	0,43		0,61	
Roboré	14	0,40	0,42	0,32	0,69
	15	0,80		0,95	
	16	0,57		0,84	

Table S.3.2 Pearson correlation between diameter at breast height (DBH) and stable isotopes in *C. fissilis* and *C. odorata*

	$\delta^{18}O$				$\delta^{13}C$			
Pearson Correlation	**r-squared**	**t-value**	**degrees of freedom**	**p-value**	**r-squared**	**t-value**	**degrees of freedom**	**p-value**
Cobija	0.19	0.55	8	0.59	-3.56e-5	-0.0001	8	0.99
Riberalta	0.53	1.75	8	0.12	0.67	2.57	8	0.03
Rurrenabaque	0.18	0.51	8	0.62	0.20	0.57	8	0.58
All sites of *C. odorata*	0.31	1.73	28	0.09	0.31	1.71	28	0.09
Espejos	0.21	0.92	19	0.37	0.32	1.48	19	0.16
Bajo Paraguá	0.20	0.88	19	0.39	0.09	0.39	19	0.70
Concepción	-0.14	-0.61	19	0.55	-0.08	-0.35	19	0.73
Roboré	-0.55	-2.83	19	0.01	-0.56	-2.95	19	0.01
Yapacaní	-0.33	-0.98	8	0.36	-0.18	-0.52	8	0.62
Guarayos	-0.01	-0.03	19	0.97	-0.12	-0.52	19	0.61
All sites of *C. fissilis*	-0.03	-0.33	113	0.74	-0.06	-0.61	113	0.54

Chapter 4

Genetic differences among *Cedrela odorata* sites in Bolivia provide limited potential for fine-scale timber tracing

Kathelyn Paredes-Villanueva, G. Arjen de Groot, Ivo Laros, Jan Bovenschen, Frans Bongers, and Pieter A. Zuidema

Submitted for publication

Abstract

Illegal trade of tropical timber leads to biodiversity and economic losses worldwide. There is a need for forensic tools that allow tracing the origin of timber and verifying compliance with international and national regulations. We evaluated the potential for genetic tracing of *Cedrela odorata*, one of the most traded Neotropical timbers, within Bolivia. Using a set of eight microsatellite (SSR) we studied the spatial distribution and genetic variation and tested whether populations show sufficient genetic discrimination for timber tracing at a national level. Cambium and leaves were sampled from 81 *C. odorata* trees from 3 sites, at 268-501 km distance. To explore genetic differentiation, Bayesian clustering and Principal Component Analysis (PCA) were employed. To infer the origin of samples we conducted Kernel Discriminant Analysis (KDA) based on a PCA that included all alleles and a manual assessment of site-unique alleles. The PCA showed three distinct genetic clusters but only one of them corresponded with one of the sampled sites. The KDA based on allele frequency had a 33.7% mean classification error, with a considerably lower error (8.2%) for the site which matched with one genetic cluster. The blind test on unique alleles led to a similar classification error (30%). The occurrence of multiple genetic clusters within sites suggests that Bolivian *C. odorata* populations contain several parental lines, resulting in limited potential for forensic tracing at a national level. Based on our findings we recommend for additional sampling across the spatial range of *C. odorata* within the country to support the development of forensic techniques for this species.

Keywords: *Cedrela odorata*, DNA, timber tracing, unique alleles, provenancing

Introduction

Illegal logging and illegal timber trade is a worldwide environmental problem, resulting in biodiversity and economic loss. It has been estimated that 10% up to 80% of the total timber trade is illegal (Seneca Creek Associates, 2004) and in some countries, such as Papua New Guinea, Liberia, and those comprising the Amazon (Stark and Pang Cheung, 2006; Lawson and MacFaul, 2010; Wit *et al.*, 2010), this is as high as 80-90%. Up to now, the control of harvesting and trade of timber species has been carried out based on certificates with declared origin. However, these systems have been weakened by the frequent use of false declarations of species and geographic origin as these documents are prone to be falsified. The system to control the timber provenance based on the respective declaration of origin is dependent on the person who enters the data and hence vulnerable to manipulation. As a result of this scenario, timber is being harvested from unauthorized areas by falsely declaring their origin. To effectively combat fraud in illegal logging and trade, there is a need for forensic techniques that use timber properties as fingerprints to independently verify the origin in both local and international markets (Degen, 2007; UNODC, 2016).

There are various potential methods based on timber properties to trace its origin such as mass spectrometry (Fidelis *et al.*, 2012), near-infrared spectroscopy (Braga *et al.*, 2011; Bergo *et al.*, 2016), stable isotopes (Kagawa and Leavitt, 2010), and DNA (Degen *et al.*, 2006; Degen and Fladung, 2007; Lowe, 2007; Jolivet and Degen, 2012; Degen *et al.*, 2013; Lowe *et al.*, 2016). DNA fingerprints hold a potential to trace a sample to a stump, population or region of origin. Its relation with spatial patterns could assist in validating the harvesting origin throughout the chain of custody and endorse legal declarations (Degen *et al.*, 2006; Degen and Fladung, 2007; Lowe, 2007; Jolivet and Degen, 2012; Degen *et al.*, 2013; Lowe *et al.*, 2016).

For implementing precise DNA fingerprints methods, a large genetic variability throughout the distribution of the timber species concerned is required. Previous studies suggest that the population genetic variability could influence the identification

resolution from site specific to country level (Lemes *et al.*, 2003; Novick *et al.*, 2003; Jolivet and Degen, 2012; Degen *et al.*, 2013; Lowe *et al.*, 2016). Genetic geographical variability at a regional scale has been found on previous studies of *Cedrela* species (Cavers *et al.*, 2003; Cavers *et al.*, 2004). A spatial analysis of *Cedrela odorata*'s cpDNA in Central and South America found geographical genetic structuring within species suggesting influence of geographic and climatic drivers (Cavers *et al.*, 2013).

Cedrela odorata is one of the most important tropical timbers (Spanish cedar) and has been affected by illegal logging. It is distributed throughout tropical America and the Caribbean islands. Due to its continuous overexploitation throughout its distribution range it has recently been protected by the Convention on International Trade in Endangered Species of Wild Fauna and Flora in Appendix III (Compt and Christy, 2008). However, a high proportion (70-90%) of illegal tropical timber is traded nationally (Cerutti and Lescuyer, 2011; Kishor and Lescuyer, 2012; Lescuyer *et al.*, 2014) and much of the illegal logging occurs in non-permitted areas or close to official logging concessions (ABT, 2017). Effective conservation and control strategies are required for the protection of these endangered species. Methods to assist in tracking and verifying the origin of timber within countries are needed.

Bolivia is one of the most important habitats for *C. odorata*. These species can be found along different altitudinal, climatic and environmental gradients, from moist to dry tropical forests (Mostacedo *et al.*, 2003; Navarro, 2011; Navarro-Cerrillo *et al.*, 2013; Paredes-Villanueva *et al.*, 2016). *C. odorata* is the most commercially high valuable species in the country (Mostacedo and Fredericksen, 1999) as its softwood is used in carpentry, joinery, musical instruments, carvings and plywood (Toledo *et al.*, 2008). However, it has been heavily affected by illegal logging (Mostacedo and Fredericksen, 1999; Toledo *et al.*, 2008; Navarro-Cerrillo *et al.*, 2013). We investigated the spatial distribution and genetic variation in three sites, each composed by 3 subgroups, of *C. odorata* in Bolivia. We addressed the following questions: (1) Do the study sites represent distinct genetic groups and follow a spatial pattern across the species

distribution?; (2) To what extent does this genetic structure allow successful timber tracing at regional and nation-wide scale?

Methodology

Sampling

Samples of cambium and leaves of *C. odorata* were collected from 81 randomly selected trees which were distributed around 3 Northern towns in Bolivia: 30 trees around Cobija, 29 around Riberalta, and 22 around Rurrenabaque (Figure 4.1). For the purpose of this paper, the sampling sites will be referred by the name of the town. Sampled trees were homogeneously distributed with a minimum distance of 26-98 m between individuals to minimize the probability of sampling relatives (Gillies *et al.*, 1999). Distance ranges between sample sites were 304-419 km between Riberalta and Cobija, 501-621 km between Riberalta and Rurrenabaque, and 268-534 km between Cobija and Rurrenabaque. To analyse if there was genetic difference at smaller scales, samples were collected from 3 sub-groups within each site composed of 10 trees each. The sub-groups were at opposite sides and at different distances determined by the presence of the species in the area. The collected samples were then dried in silica gel (Chase and Hills, 1991) for transporting and storing. When species identity was uncertain, additional botanical samples were collected and transported with a botanical press for taxonomic identification. The voucher preparation and confirmation of the species based on herbarium collections were carried out by an experienced botanist, A. Araujo Murakami at the Museo de Historia Natural Noel Kempff Mercado (Bolivia). Once verified, correctly identified samples were included in the dataset.

Table 4.1 Sampling characteristics: sample size, tree diameter range, minimum and maximum distance between trees per site, and number of samples collected for species identification

Site	Sampled size	Tree diameter range (cm)	Min. distance (m)	Max. distance (km)	Samples for confirmation in herbarium
Cobija	30	1-130	98	122	1
Riberalta	29	4-146	26	88	2
Rurrenabaque	22	0.5-64	42	178	4

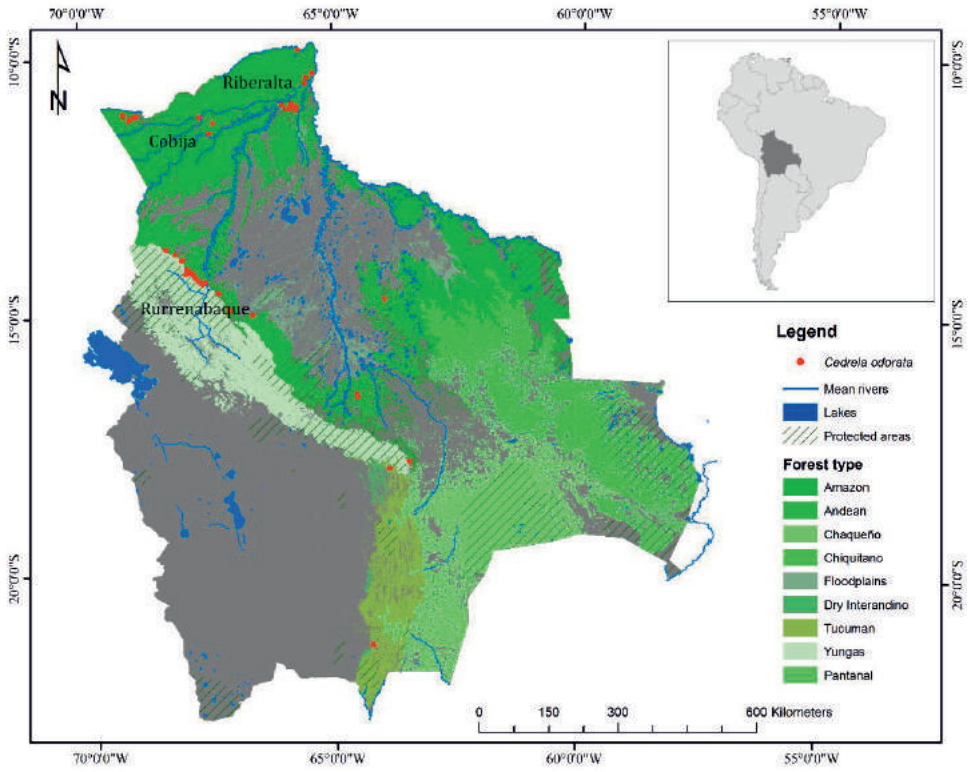


Figure 4.1 Sampling sites (Cobija, Riberalta and Rurrenabaque) based on *Cedrela odorata* distribution in Bolivia based on herbarium collections. Source: Museo de Historia Natural Noel Kempff Mercado, 2017. Forest cover: Autoridad de Bosques y Tierra, 2015.

DNA extraction

Working with free of contamination equipment and very conserved primer pairs improved DNA quality and optimized the PCR conditions (Deguilloux *et al.*, 2002). Cambium and leaf samples were cut into small pieces using a sterile scalpel and under sterile conditions to avoid contamination with other plant DNA. Then, 10-50 mg of cambium/leaf pieces were incubated in liquid nitrogen with one stainless steel bead (5 mm; Qiagen, the Netherlands) for 5 min to freeze and then grinded using a mixer-mill apparatus Type MM 300 (Westburg) for 2 min at 30 units shaking-speed (1/s frequency). DNA extraction was performed with the DNeasy Plant Mini Kit (DNeasy Plant Handbook 10/2012) following Rachmayanti *et al.* (2006), Hernández *et al.* (2008) and Degen *et al.* (2013) adding 3.1% (w/v) polyvinylpyrrolidone (PVP) into AP1 lysis buffer of DNeasy Plant Mini Kit (Qiagen, the Netherlands). DNA quantity and quality was determined using NanoDrop™ 2000/2000c Spectrophotometer.

Nuclear microsatellite analysis

Nuclear microsatellites are variable markers that permit individual tree genotyping and suggested for forest crime applications (Deguilloux *et al.*, 2002). For our analysis there were primers available (Hernández *et al.*, 2008; Hernández Sánchez, 2008; Cárdenas *et al.*, 2015). Up to now, microsatellites or Simple Sequence Repeats (SSR, Tautz, 1989) for *Swietenia humilis* (White and Powell, 1997a, b), *Swietenia macrophylla* (White and Powell, 1997a; Lemes *et al.*, 2002), *Cedrela odorata* (White and Powell, 1997a; Hernández *et al.*, 2008; Hernández Sánchez, 2008), and *Cedrela fissilis* (Gandara, 2009) have been developed and also tested on *Cedrela balansae* and *Cedrela saltensis* (Soldati *et al.*, 2013; Soldati *et al.*, 2014). Therefore, we assessed 8 primer pairs of SSRs developed for *Cedrela odorata* and *Cedrela fissilis* (Table 4.2). We considered patterns of high polymorphism loci for the selection of these SSRs. Up to now microsatellite amplification was done in separate PCR's for each marker, we designed multiplex PCR's for this purpose using Multiplex Manager v1.2 software (Holleley and Geerts, 2009). Seven selected SSR loci (locus Ced61a failed to show consistent results in the amplification

tests) were used for the detection and separation of fragments in a sequencer. Genotyping of the leaves and cambium samples were carried out in two PCR reactions.

Table 4.2 Oligonucleotides and primers used for all *Cedrela* samples

Multiplex PCR	Locus	5' label of forward primer	Primer Concentration in PCR (nM)
-	Ced61a	NED	200
I	Ced95	6FAM	300
I	Ced131	PET	200
I	Ced41	VIC	200
II	Ced2	PET	200
II	Ced18	6FAM	200
II	Ced44	VIC	200
II	Ced54	NED	200

Extracted DNA concentrations were 50 ng/µl in average (NanoDrop™ 2000/2000c Spectrophotometer) and were diluted 5 times with elution buffer before use in PCR. Selected loci were amplified using 10.0 µl PCR reactions, containing 2.0 µl template DNA, 1.9 µl H2O, 1 µl primermix (see details in Table 4.2), 0.1 µl BSA (20 mg/ml), and 5 µl QIAGEN Multiplex PCR Mastermix. In this PCR reaction mixture, we combined the primers into two multiplexes for each sample. Multiplex I targeted 3 loci (Ced95, Ced131 and Ced41) and Multiplex II targeted 4 loci (Ced2, Ced18, Ced44 and Ced54). Reactions for Multiplex I and II were run according to the following protocol: 15 min at 95°C, then 10 cycles of 30 s at 94°C, 180 s at 62°C minus 1°C per cycle, 60 s at 72°C; followed by 35 cycles of 30 s at 94°C, 180 s at 52°C, 60 s at 72°C, and finally 30 min at 60°C. PCR products were directly analyzed using an automated ABI Prism® Genetic Analyzer (Applied Biosystems) coupled to 3730 Series Data Collection Software 4. Fragment sizes were scored using GeneMarker® v.2.6.7 (SoftGenetics) software.

Data analysis

After the validation and calibration of the pre-selected SSR markers, the number of alleles (N_A), allele frequencies (p_{ij}), the observed heterozygosity (H_o), expected heterozygosity (H_e) and fixation index ($F_{is}=1-(H_o/H_e)$) were calculated using FSTAT

v.2.9.3 (Goudet, 1995). From a total number of 91 samples, 81 samples were selected after excluding samples without allelic information in more than one locus.

Genetic differentiation among sampling sites was explored through two methods: STRUCTURE and Principal Component Analysis (PCA). One method employed a Bayesian algorithm in a model-based clustering with STRUCTURE v.2.3.4 (Pritchard *et al.*, 2000) to analyse the genetic structure of *C. odorata* from 3 sites in Bolivia using 7 microsatellites (Table 4.2). Four independent runs were performed for the complete sample set and each of the sites with different number of groups (*K*), each with 50,000 iterations in the burning period and 50,000 in Monte Carlo Markov Chain (MCMC) iterations assuming both admixture and no-admixture models with both correlated and independent allele frequencies models. *Delta K* statistics were used to define the optimal number of groups (*K*) by STRUCTURE HARVESTER Web v0.6.94 (Earl and vonHoldt, 2012). The CLUMPP output files from the previous analysis were exported to Excel files to visualize the number of groups and genetic clusters. Individual samples were then classified as belonging to 1 cluster based on the highest assignment percentage, e.g. an individual assigned for 50% to cluster 1 and 25% to cluster 2 and 3, was classified as belonging to cluster 1. The second method employed a PCA based on all (143) the alleles present in the data set to explore genetic differentiation among sampling sites. The alleles were scored on a presence/absence binary dataset which was then analysed in a logisticPCA model by using logisticPCA v.0.2 package in R version 3.3.3. Once applying this model, plots were prepared with ggplot v.2.2.1 package. To check the markers performance, tests of Hardy-Weinberg equilibrium (HWE) and linkage equilibrium (LE) were performed using the program GENEPOP v.4.2 (Raymond and Rousset, 1995). Deviance from HWE in one marker might indicate null alleles. Deviance from LE might mean that two markers are located closely together on the genome and their inheritance is not independent. For measuring the fixation and genetic differentiation among sites, Wright's F_{ST} (Wright, 1951) was calculated.

Characterization of sites and genotype assignment

To check the consistency of the results, we used two independent approaches to infer the origin of samples. The first approach included a discriminant analysis based on all (143) the alleles as a whole present in the data set: the Principal Components –previously used for genetic differentiation in PCA– were applied as variables in a Kernel Discriminant Analysis (KDA) by using the package ks version 1.10.6 (Duong, 2007, 2017) in R version 3.4.3 (R Development Core Team, 2017). KDA learning algorithm uses Bayes discriminant rule which allocates a point *x* in the sample space to one (and only one) of the sampled sites (Duong, 2007). This learning algorithm needed to be trained in order to assess the discrimination power of KDA. Therefore, our data was split in two sets: 97% of the sample set used for training and 3% as testing set with 10,000 randomizations of both data sets. Smoothed Cross Validation method (Duong, 2007) was also applied for assessing the discrimination analysis. This showed the classification error, expressed in percentage (%), for the probability of each sample to belong to the sampling sites.

In the second approach, we made use of the presence of characteristic or unique alleles, which were defined as alleles that occurred within the reference dataset only in a single, or in two out of three study sites. We performed a blind test, in which the genetic profiles of all samples were re-coded by the second author (GAG) and 10 samples (i.e. profiles) were randomly drawn from the reference dataset. The remaining samples were then used by the first author (KPV) as a reference to infer the origin of the 10 selected samples. Inference of origin was manually analysed based on a fixed set of rules: 1) In case the test sample contained an allele that occurred in two sites, the sample was considered to originate from one of these two sites. 2) In case the profile of a test sample contained an allele (at any of the loci) that was found only in the reference data for one of the sites, this site was considered to be its origin. When combining this information for all observed alleles in the test sample, in most cases only one possible origin remained. In case information from two alleles in a profile resulted in contradictory conclusions, or in case two possible sites of origin remained, the result was scored as inconclusive. For

the samples for which a conclusion was reached, KPV then checked the outcome with GAG to check whether the conclusion was correct.

Results

Genetic diversity and characterization of sites

The final dataset contained microsatellite profiles of 81 *C. odorata* samples from 3 spatially separated sites in Bolivia, based on 7 microsatellite loci per profile, together harbouring a total of 143 different alleles. The observed genetic structure within sites did not match with the sub-groups distinguished during sampling. The mean number of alleles (A) per nSSR locus was 14.2 for all the 3 sites analysed. The expected heterozygosity (H_e) varied from 0.83 to 0.89 for all the analysed loci (Table 4.3). The mean observed heterozygosity (H_o) was lower than the mean H_e in all three sampling sites, resulting in a positive fixation index (F_{is}) for especially Cobija and Rurrenabaque (0.14 and 0.11, respectively), indicating spatial substructure and non-random mating within these populations. This is in line with the results of the STRUCTURE analysis (Figure 4.2 and 4.3; blue, red and yellow dots), showing the presence of multiple genetic clusters in these populations. F_{is} nearly equalled zero (0.02) for Riberalta, indicating little substructure, again in line with the STRUCTURE results that classified all individuals of this population to the same genetic cluster (Figure 4.2 and 4.3, blue dots).

Genetic characterization and geographic structure of sites

The average pairwise breed genetic differentiation test (F_{st}) allowed us to assess the magnitude of the genetic differentiation among sites. Riberalta and Rurrenabaque sampling sites showed the highest differentiation (F_{st}= 0.10), followed by the Riberalta and Cobija (F_{st}= 0.06) sites. Cobija and Rurrenabaque showed the lowest genetic differentiation (F_{st}= 0.04). Consistently, the allelic richness (A) for Cobija was the highest among all sites (Table 4.3).

Table 4.3 Average expected heterozygosity (H_e), average observed heterozygosity (H_o), mean number of alleles (A), number of alleles with a minimal sample size of 21 diploid individuals (A_r), and the average inbreeding coefficient or fixation index (F_{is}). Sample size were: 30 trees in Cobija, 29 in Riberalta, and 22 in Rurrenabaque.

Loci/sites		Ced18	Ced44	Ced54	Ced02	Ced95	Ced41	Ced131	Average per site
H_e	Riberalta	0.42	0.93	0.90	0.84	0.93	0.93	0.86	**0.83**
	Cobija	0.82	0.96	0.85	0.86	0.96	0.92	0.90	**0.89**
	Rurrenabaque	0.83	0.93	0.92	0.84	0.94	0.85	0.86	**0.88**
H_o	Riberalta	0.32	0.86	0.90	0.83	1.00	0.97	0.78	**0.81**
	Cobija	0.55	0.93	0.60	0.54	0.93	0.86	0.86	**0.75**
	Rurrenabaque	0.60	0.86	0.75	0.71	0.86	0.81	0.81	**0.77**
A	Riberalta	3	16	15	9	16	18	9	**12.29**
	Cobija	15	23	14	15	21	18	12	**16.86**
	Rurrenabaque	12	16	14	11	17	12	11	**13.29**
A_r	Riberalta	2.93	14.75	13.81	8.43	14.85	16.35	8.44	**11.37**
	Cobija	12.55	20.13	11.49	13.31	18.92	15.87	11.35	**14.80**
	Rurrenabaque	11.77	15.81	14.00	10.91	16.68	11.77	10.82	**13.11**
	Total	*10.29*	*20.05*	*14.10*	*13.24*	*19.35*	*17.02*	*12.15*	***15.17***
F_{is}	Riberalta	0.11	0.08	0.00	0.02	-0.07	-0.03	0.08	**0.02**
	Cobija	0.27	0.03	0.26	0.32	0.02	0.06	0.03	**0.14**
	Rurrenabaque	0.23	0.07	0.17	0.14	0.08	0.03	0.05	**0.11**

Structure analysis using 7 microsatellites showed clear groupings of the 81 samples through the analysis of the ΔK (Evanno *et al.*, 2005). The ΔK calculated assuming an admixture model with both independent and correlated allele frequencies showed a consistent decrease in the ΔK curve after K=3. We selected this optimal K value as the number of genetic clusters (Supplementary Data Figure S.4.1) for our study species. Results considering no admixture among sites also showed similar results (data not shown).

For final classification of each sample to the three genetic clusters, the admixture model was selected as we suspected mixing of the genetic provenances (Falush *et al.*, 2003). Using these models settings, similarity in the classification results between 5 iterations was 99.7%. Genetic cluster 1 (blue, Figure 4.2 and 4.3) showed an almost homogenous genetic composition for all the samples in Riberalta site. Genetic cluster 2 and 3 covered both Rurrenabaque and Cobija (yellow and red respectively, Figure 4.2 and 4.3).

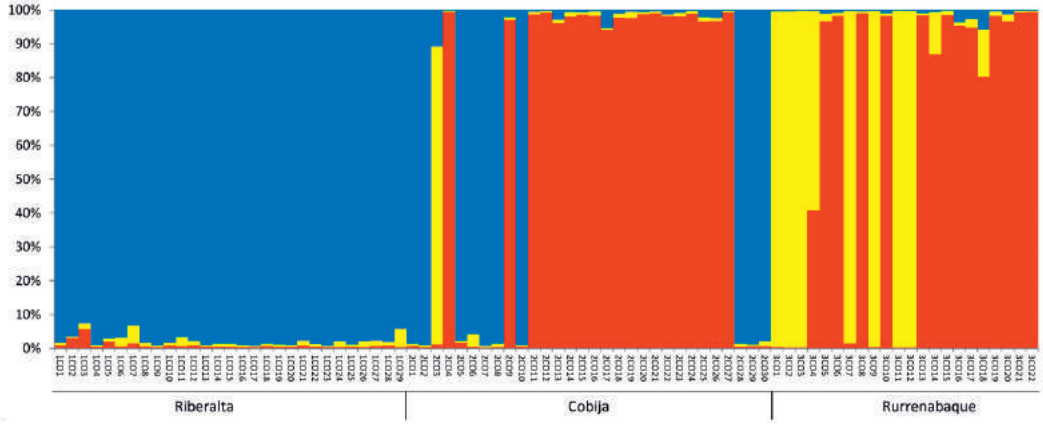


Figure 4.2 Site genetic structure of 81 *Cedrela odorata* samples with admixture and at *K*=3 based on the analysis of allele frequency in STRUCTURE software. Names refer to the study sites and colours to genetic clusters: Cluster 1= blue, Cluster 2= yellow, and Cluster 3= red.

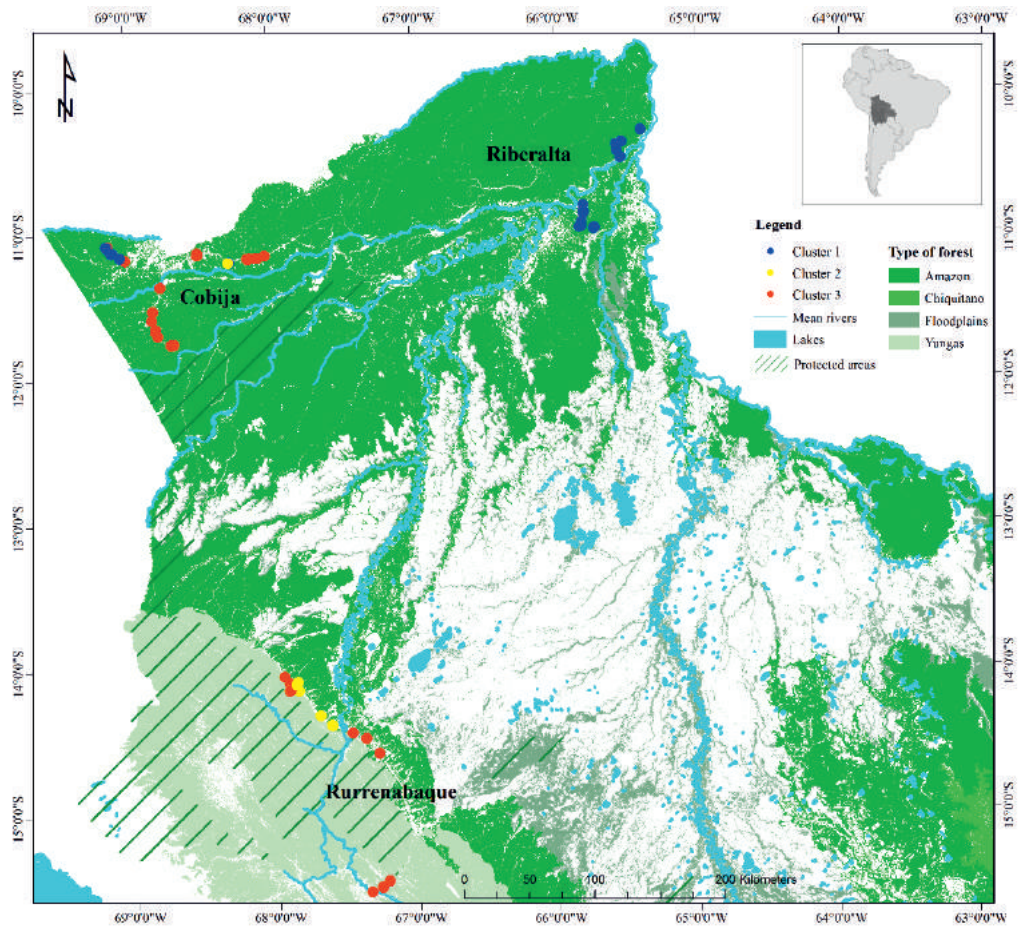


Figure 4.3 Genetic characterization of 81 sampled *Cedrela odorata* trees at three sites in Bolivia. Coloured filled dots refer to the dominant genetic clusters: Cluster 1= blue, Cluster 2= yellow, and Cluster 3= red. Samples covered two forest types: Amazon and Yungas. Forest cover: Autoridad de Bosques y Tierra, 2015.

Comparable to the STRUCTURE results, the Principal Component Analysis (PCA) showed 3 genetic clusters based on the allelic composition. It showed a clear separation of the genetic cluster 2 (yellow circle, Figure 4.4) which mainly grouped samples from Rurrenabaque. The composition of the genetic cluster 3 (red circle, Figure 4.4) was shared between Cobija and Rurrenabaque. Genetic cluster 1 (blue circle, Figure 4.4) was shared between Riberalta and Cobija.

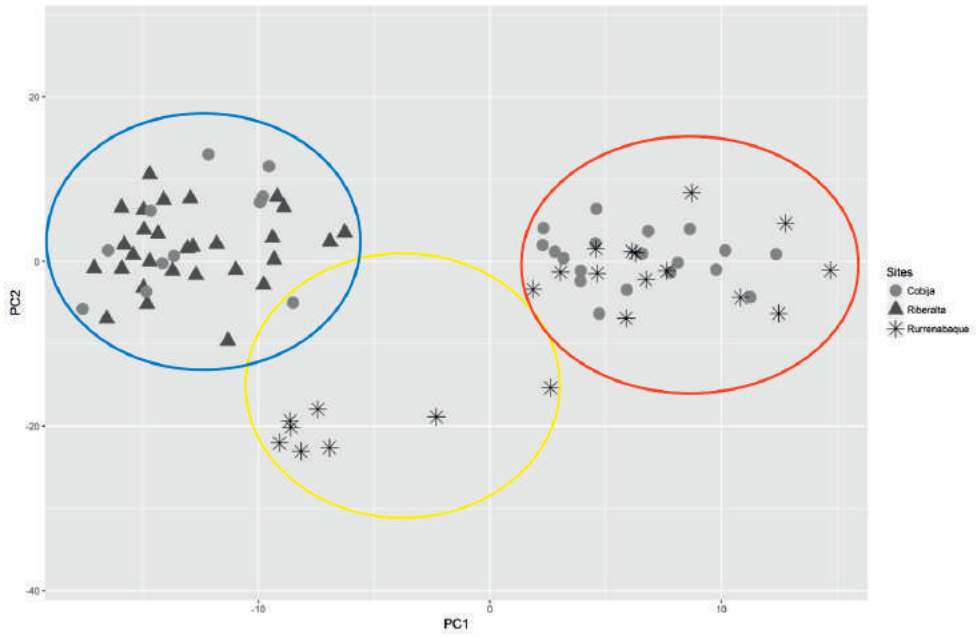


Figure 4.4 Principal Component Analysis (PCA) showing three genetic clusters (coloured circles) based on presence/absence of alleles in 81 *Cedrela odorata* trees at three geographic sites (symbols)

Genotype assignment

From the PCA analysis using all the alleles (first approach), two Principal Components were used as variable input for the Kernel Discriminant Analysis (KDA), as the PCA was composed on binary data based on the presence/absence of alleles in each of the sites. By using 97% of the sample set as training we maximized the allelic information to train the model. After 10,000 runs of randomizing both training and testing set, we found stable results for the assignment of all the samples to each sampling site. The Smoothed Cross Validation presented Cobija as the site with highest classification error of 49.2%, followed by Rurrenabaque with 43.7% and Riberalta with the lowest classification error of 8.2% (Figure 4.5). The mean classification error of assigning one sample to its true origin was 33.7%. These results confirmed the admixture structure presented previously on the cluster analysis in STRUCTURE.

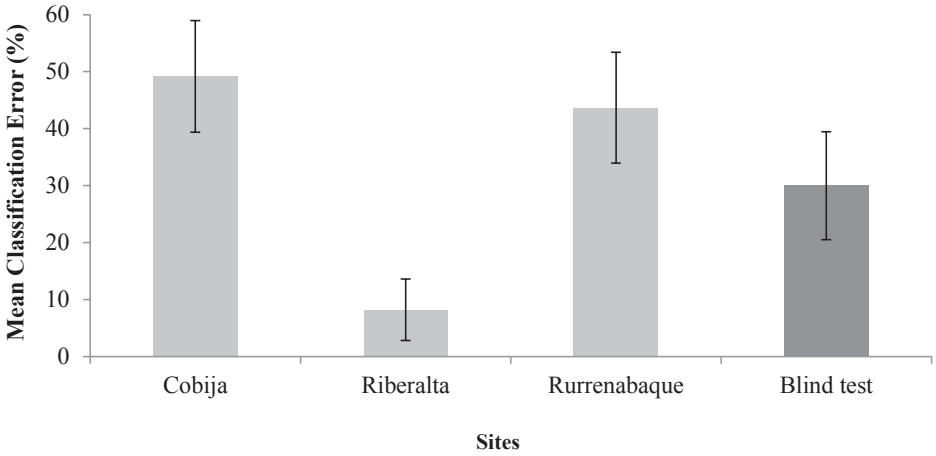


Figure 4.5 Classification errors of samples to 3 sites in Bolivia. *Light grey bars* indicate mean classification error of the sampling sites in Bolivia after 10,000 randomization of the training and test set with Kernel Discrimination Analysis (KDA). *Dark grey bar* indicates mean classification error of manual blind test, and *whiskers* the standard deviation (SD).

The alleles' reference dataset of *C. odorata* showed the presence of unique alleles for each site (Supplementary Data Table S.4.1). These unique alleles were used for a manual classification of a blind set of samples. The blind test (Supplementary Data Table S.4.2) was performed using the unique alleles and 10 samples were assigned to possible sites of origin or suspected provenance. The high amount of unique alleles in the sampling sites allowed us to test whether these could be used as characteristic of specific site (Supplementary Data Table S.4.1). From the total, 70% of the samples were correctly assigned to the site of origin (Supplementary Data Table S.4.3). For 30% of the samples, the allele composition gave conflicting suggestion of the site of origin. For example, BLIND08 was correctly assigned, however it presented unique alleles characteristic of both Cobija and Rurrenabaque. On the contrary, BLIND03 sample was wrongly assigned to Riberalta, as it presented a unique allele (locus Ced95-allele 230) characteristic of Riberalta and Cobija but also other alleles (locus Ced02-allele 161) characteristic of Riberalta and Rurrenabaque, and (Ced131-allele 100) only for sites of Cobija and Rurrenabaque (Supplementary Data Table S.4.3).

Discussion

There is a clear need for within-country provenancing of tropical timbers, at scales <100 km. We assessed whether *C. odorata* trees from three regions in Bolivia – Cobija, Riberalta and Rurrenabaque – could be genetically distinguished (question 1). We also tested the level of discrimination and resolution (question 2) for tracing timber origin. We found classification errors ranging from 10% to 50%. The mean classification error of assigning one sample to its true origin was 33.7% based on allele frequency and 30% based on unique alleles' blind test. These results are in a similar range to a previous study on *Swietenia macrophylla* species, which showed an error assignment of 29.3% to site of origin (Degen *et al.*, 2013). Other study on *Hymenaea courbaril* found mean classification error of 11-12% (Chaves *et al.*, 2018). The samples of Degen *et al.* (2013) and Chaves *et al.* (2018) were analysed at large geographical scales: within and among countries: 5-5660 km and 156-1384 km, respectively, from the closest to the farthest sample. In contrast, we sampled on a finer geographical scale (within Bolivia 26 m-619 km) which revealed several limitations. Assignment accuracy showed to be dependent not only on the number of sites sampled and the genetic differentiation (Cavers *et al.*, 2005) but also on the alignment of the latter with the spatial organization. Our results suggest that species-specific traits and the analysis approach may affect the origin-identification success. Below we discuss possible explanations for our classification success.

First, results from different species have shown varying results, independent of spatial scale. They suggest that species-specific traits affecting population genetics, such as pollination/seed dispersal syndrome, mating strategy, and mutation rates, might have affected the genetic variation of the sampled sites (Hedrick, 2011) and potential for tracing. Our study species is pollinated by insects (Bawa *et al.*, 1985; Howard *et al.*, 1995) and its light wing-shaped seeds are dispersed by wind over long distances (Mostacedo *et al.*, 2003; Toledo *et al.*, 2008). The seed dispersal followed by a limited pollen dispersal trait best explained the multiple clusters in some of our populations (substructure in Rurrenabaque and Cobija sites), as very local mating patterns (inbreeding) could have

kept the introduced genetic cluster intact. In addition, the lack of spatial substructure can also be due to regular gene flow from neighbouring domestic trees or conservation areas. Although samples were collected in natural forests to reduce the possibility that trees were recruited from non-natural seed sources, the Cobija samples were close to communities where trees may have been planted. Similarly, Rurrenabaque samples were located in disturbed sites along the border of the Madidi National Park. Gene flow from the national park to the disturbed areas may have contributed to genetic variation (Soliani *et al.*, 2016). It has been found that reduction of tree populations due to logging may result in decrease of genetic variation (Finkeldey and Ziehe, 2004) and allelic frequencies (Cornuet and Luikart, 1996; Rajora *et al.*, 2000) hence it is expected that protected areas still harbour a larger historic genetic diversity. This structure should be taken into account during future analyses that include conservation areas as a reference since populations further from them may differ in genetic composition. It is also recommended to get information about enrichment planting activities in the past before sample collection or data analysis.

Second, the resolution and approach used in the study, such as frequency of alleles and the presence of unique alleles, may have influenced the level of precision for identifying the origin of the wood. The Bayesian clustering approach to estimate allele frequency provides the likelihood of one allele of belonging to one genetic cluster iteratively (Pritchard *et al.*, 2000). If this likelihood is low, this approach tries to assign the allele to another genetic cluster. When genetic clusters have shared alleles the estimations may result in low accuracy for site origin assignment during the KDA analyses. Alternatively, unique alleles in each sampling site could be used for site diagnosis. However, care should be taken because the inference of sample origin may lose power when the unique alleles are used together with allele composition and frequency. Site characterization based on alleles depends on the type of data making up the training sets. Vlam *et al.* (2018) suggested that increasing assignment accuracy may be not only due to the presence of large numbers of unique (private) alleles but also on the way that these are analysed (PCA vs. Bayesian clustering analysis).

To objectively and confidently pinpoint the site of origin, a blind sample should only be assigned to its origin if unique alleles occur in only one genetic cluster. The classification error based on the presence of unique/characteristic alleles per site could be higher if a blind sample originates from an unknown site. Therefore, the currently available reference dataset has a limited contribution to identify the origin of a sample. This indicates that, although our sites were widely dispersed, a bigger and more widespread dataset is needed, namely a reference database with samples across the distribution of the species including neighbouring countries. Such an extensive dataset would show to what extent Bolivian samples come from local areas or are illegally imported.

Both the spatial resolution and the assignment accuracy depend on geographical scale and sampling scheme. Previous studies found that identifying the country of origin was more accurate (82.2%) than identifying the site of origin individually (70.7%; Degen *et al.*, 2013) and that only after merging common genetic clusters did mean error decrease from 67-72% to 48-49% (Jolivet and Degen, 2012) and from 95.5 to 4.2% error (Vlam *et al.*, 2018). The 81 samples used in this study showed differences on spatial discrimination due to the high degree of genetic mixing within sites. To improve insights into the degree of genetic mixing, it is necessary to adjust sampling strategies according to the level of differentiation and local variation which in turn consider global standardized methods. Standardizing sampling and data collection will make possible to use genetic data from other countries when necessary. By expanding the database to other countries and by sampling in between sites (e.g. sampling across a gradient) will be possible to identify site of origin at finer scales and with a higher precision.

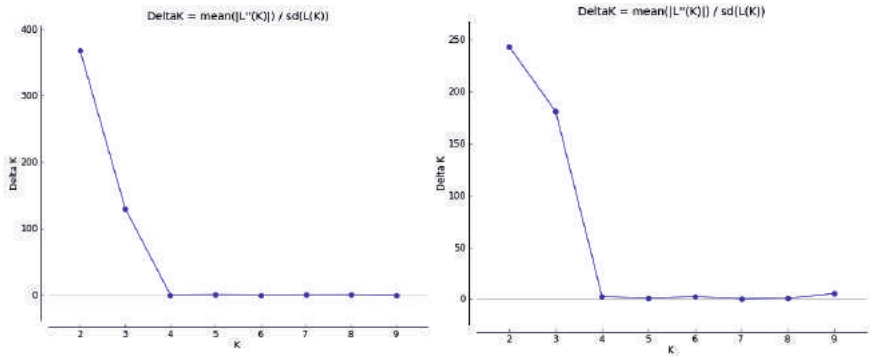
In conclusion, this study shows that (a) microsatellites can be used to define genetic clusters of *C. odorata* and study provenances; (b) the studied *C. odorata* populations within Bolivia have multiple parental lines; (c) the current reference dataset has limited use for tracing. Based on our findings we recommend for additional sampling across the spatial range of *C. odorata* in this part of South America to support the development of forensic techniques for this species.

Acknowledgements

We would like to thank the Museo de Historia Natural Noel Kempff Mercado for their support on the species identification and scientific guidance. This research was implemented under the MMAYA/VMABCCGDF/DGBAP/MEG N° 0280/2016 authorization and financed by the NFP/Nuffic fellowship (the Netherlands). Fieldwork was logistically supported by Universidad Autónoma Gabriel René Moreno and financed by Alberta Mennega Stichting and The Rufford Foundation. We thank Jente Ottenburghs for his assistance in the statistical analysis and reviewing previous versions of this manuscript.

Supplementary Data

Figure S.4.1 Number of groups (*K*) determined by *ΔK* assuming admixture model with both independent and correlated allele frequencies



a) Admixture model, independent allele frequencies; b) Admixture model, correlated allele frequencies

Table S.4.1 Reference set of alleles' occurrence in three sampling sites in Bolivia. All the present alleles (N) are listed for each of the 7 loci.

Locus: Ced18				Locus: Ced44			
	Riberalta	Cobija	Rurrenabaque		Riberalta	Cobija	Rurrenabaque
N	**29**	**30**	**22**	**N**	**29**	**30**	**22**
p: 124	0.241	0.083	0.318	p: 178	0	0.017	0.045
p: 126	0.724	0.3	0.023	p: 180	0	0.017	0
p: 128	0.034	0	0	p: 182	0.086	0	0.182
p: 130	0	0.3	0.273	p: 184	0.017	0.017	0
p: 132	0	0.033	0.023	p: 186	0.052	0	0.136
p: 134	0	0	0.023	p: 187	0.052	0.05	0.023
p: 136	0	0.017	0	p: 188	0	0.017	0.023
p: 138	0	0.033	0.068	p: 190	0	0.017	0.045
p: 140	0	0.033	0.045	p: 191	0	0.017	0.045
p: 142	0	0.017	0.023	p: 192	0.017	0	0
p: 144	0	0.033	0.045	p: 193	0.138	0	0
p: 146	0	0.017	0.068	p: 195	0	0.05	0.023
p: 148	0	0.067	0	p: 197	0.069	0.033	0.045

p: 150	0	0.017	0.068
p: 152	0	0.017	0
p: 154	0	0.017	0.023
p: 156	0	0.017	0

Locus: Ced54

	Riberalta	Cobija	Rurrenabaque
N	**29**	**30**	**22**
p: 182	0	0.017	0.024
p: 184	0.069	0	0
p: 186	0.034	0	0
p: 187	0.052	0.25	0.19
p: 188	0	0.017	0.048
p: 189	0	0.183	0.048
p: 190	0.034	0.05	0.143
p: 192	0.052	0.033	0.119
p: 194	0.034	0.017	0.024
p: 196	0.19	0.15	0.071
p: 198	0.017	0.017	0.095
p: 200	0.052	0	0.024
p: 202	0.052	0.2	0.024
p: 204	0.207	0.017	0.119
p: 206	0.138	0.017	0
p: 208	0.034	0.017	0.048
p: 210	0.017	0	0
p: 212	0	0	0.024
p: 220	0.017	0.017	0

Locus: Ced95

	Riberalta	Cobija	Rurrenabaque
N	**29**	**30**	**22**
p: 192	0	0	0.023
p: 194	0	0.033	0.023
p: 196	0	0	0.023
p: 198	0.086	0	0.159
p: 200	0.017	0.033	0.114
p: 202	0.069	0.067	0.068
p: 204	0	0.017	0.023
p: 206	0.034	0.017	0.068
p: 208	0.121	0	0
p: 210	0.034	0.067	0.023
p: 199	0.052	0.033	0.023
p: 201	0	0.067	0.091
p: 203	0	0.083	0.114
p: 205	0	0.067	0.068
p: 207	0.017	0.05	0.045
p: 209	0	0.033	0.045
p: 211	0.017	0.017	0
p: 213	0.069	0.067	0.045
p: 215	0.069	0.067	0
p: 217	0.086	0.083	0
p: 219	0.103	0.033	0
p: 221	0.121	0.1	0
p: 223	0.034	0.017	0
p: 225	0	0.05	0

Locus: Ced02

	Riberalta	Cobija	Rurrenabaque
N	**29**	**30**	**22**
p: 147	0	0.017	0.045
p: 149	0	0	0.318
p: 153	0	0.345	0.227
p: 155	0	0.017	0
p: 157	0.017	0.017	0.068
p: 159	0.017	0.034	0.068
p: 161	0.155	0.017	0.023
p: 163	0.31	0	0.045
p: 165	0.121	0.103	0.068
p: 167	0.121	0.086	0.068
p: 169	0	0.052	0.045
p: 171	0.103	0.103	0
p: 173	0.103	0.034	0
p: 175	0.052	0.086	0.023
p: 177	0	0.017	0
p: 179	0	0.034	0
p: 191	0	0.034	0

p: 212	0.052	0.017	0.045
p: 214	0.034	0.033	0.023
p: 216	0	0.067	0.091
p: 218	0	0.083	0.114
p: 220	0	0.067	0.091
p: 222	0.017	0.067	0.023
p: 224	0	0.033	0.045
p: 226	0	0	0
p: 228	0.017	0.017	0.045
p: 230	0.069	0.083	0
p: 232	0.121	0.1	0
p: 234	0.086	0.033	0
p: 236	0.086	0.067	0
p: 238	0.121	0.05	0
p: 240	0.034	0.033	0
p: 242	0	0.017	0

Locus: Ced131

	Riberalta	Cobija	Rurrenabaque
N	**29**	**30**	**22**
p: 88	0	0.033	0
p: 89	0	0	0.205
p: 90	0	0.033	0.023
p: 92	0	0	0.159
p: 94	0.103	0.083	0
p: 96	0.069	0.067	0.25
p: 98	0	0.117	0.091
p: 100	0.017	0.183	0.114
p: 102	0.259	0.15	0.045
p: 104	0.155	0.15	0.045
p: 106	0.138	0.1	0.023
p: 108	0.121	0.033	0.023
p: 110	0.121	0.033	0.023
p: 112	0.017	0.017	0

Locus: Ced41

	Riberalta	Cobija	Rurrenabaque
N	**29**	**30**	**22**
p: 122	0.069	0.067	0
p: 124	0.172	0.1	0
p: 126	0.086	0.067	0
p: 128	0.017	0.067	0.295
p: 130	0.034	0	0.045
p: 132	0.086	0.217	0.227
p: 134	0.034	0.017	0.045
p: 136	0	0.083	0.091
p: 138	0.017	0.1	0
p: 140	0.069	0.017	0.136
p: 142	0.069	0.033	0.023
p: 144	0.052	0.05	0
p: 146	0.017	0.017	0
p: 148	0.034	0.05	0
p: 150	0.086	0.017	0
p: 152	0.052	0.017	0.023
p: 156	0.069	0	0
p: 158	0	0.033	0.023
p: 160	0	0	0.045
p: 162	0	0.033	0.023
p: 164	0.017	0	0
p: 166	0	0.017	0
p: 168	0.017	0	0.023

Table S.4.2 Genetic profile of ten blind samples to identify sample site of origin. This information was compared to the reference set of alleles' occurrence of the three sampling sites in Bolivia for site of origin inference. Two alleles for each of the 7 loci were provided.

	BLIND01	BLIND02	BLIND03	BLIND04	BLIND05	BLIND06	BLIND07	BLIND08	BLIND09	BLIND10
Ced18	126	126	126	124	126	124	130	146	130	124
	130	126	126	124	126	124	134	146	138	126
Ced44	205	193	213	182	197	182	199	178	197	213
	209	219	221	190	211	182	203	191	201	219
Ced54	202	204	196	192	206	196	187	204	187	190
	206	206	202	196	206	198	212	204	189	200
Ced02	153	163	161	149	173	149	165	149	153	171
	153	171	169	149	173	149	167	153	153	173
Ced95	220	208	230	198	212	198	214	192	212	230
	224	234	238	204	228	198	218	206	216	236
Ced41	124	130	124	128	124	128	132	132	132	124
	132	142	152	128	148	128	140	136	160	142
Ced131	100	94	100	89	104	89	96	90	100	108
	100	100	104	89	104	92	96	100	102	110

Table S.4.3 Blind test for assignment of unique alleles to sites of origin. Inference of origin was based on the following conditions: 1) In case the test sample contained an allele that occurred in two sites, the sample was considered to originate from one of these two sites. 2) In case the profile of a test sample contained an allele (at any of the loci) that was found only in the reference data for one of the sites, this site was considered to be its origin. In case the resulting information from two alleles in a profile resulted in contradictory conclusions, or in case two possible sites of origin remained, the result was scored as inconclusive (question mark symbol).

	BLIND01	BLIND02	BLIND03	BLIND04	BLIND05	BLIND06	BLIND07	BLIND08	BLIND09	BLIND10
Ced18	Riberalta, Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Cobija, Rurrenabaque	Cobija, Rurrenabaque	Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque
	Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque	?	Cobija, Rurrenabaque	Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque
Ced44	Cobija, Rurrenabaque	Riberalta	Riberalta, Cobija, Rurrenabaque	Ribertalta, Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Ribertalta, Rurrenabaque	Riberalta, Cobija	Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque
	Cobija, Rurrenabaque	Riberalta, Cobija	Riberalta, Cobija	Cobija, Rurrenabaque	Cobija	Ribertalta, Rurrenabaque	Cobija, Rurrenabaque	Cobija, Rurrenabaque	Cobija, Rurrenabaque	Riberalta, Cobija
Ced54	Riberalta, Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Riberalta	Riberalta, Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque
	Riberalta	Riberalta	Riberalta, Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Riberalta	Riberalta, Cobija, Rurrenabaque	?	Riberalta, Cobija, Rurrenabaque	Cobija, Rurrenabaque	Riberalta, Rurrenabaque
Ced02	Cobija, Rurrenabaque	Riberalta, Rurrenabaque	Riberalta, Rurrenabaque	Rurrenabaque	Riberalta, Cobija	Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Rurrenabaque	Cobija, Rurrenabaque	Riberalta, Cobija
	Cobija, Rurrenabaque	Riberalta, Cobija	Cobija, Rurrenabaque	Rurrenabaque	Riberalta, Cobija	Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Cobija, Rurrenabaque	Cobija, Rurrenabaque	Riberalta, Cobija
Ced95	Riberalta, Cobija	Riberalta	Riberalta, Cobija	Riberalta, Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Riberalta, Rurrenabaque	Riberalta, Cobija	?	Riberalta, Cobija, Rurrenabaque	Riberalta, Cobija
	Riberalta, Cobija	Riberalta, Cobija	Riberalta, Cobija	Cobija	Cobija, Rurrenabaque	Riberalta, Rurrenabaque	Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Cobija, Rurrenabaque	Riberalta, Cobija
Ced41	Riberalta, Cobija	Riberalta,	Riberalta, Cobija	Riberalta, Cobija,	Riberalta, Cobija	Riberalta, Cobija,	Riberalta, Cobija,	Riberalta, Cobija,	Riberalta, Cobija,	Riberalta, Cobija

		Rurrenabaque		Rurrenabaque		Rurrenabaque	Rurrenabaque	Rurrenabaque	Rurrenabaque	
	Riberalta, Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Riberalta, Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Riberalta, Cobija	Riberalta, Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Cobija, Rurrenabaque	Rurrenabaque	Riberalta, Cobija, Rurrenabaque
Ced131	Cobija, Rurrenabaque	Riberalta, Cobija	Cobija, Rurrenabaque	Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Cobija	Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque
	Cobija, Rurrenabaque	Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque	Riberalta, Cobija, Rurrenabaque
Assignment	**Riberalta**	**Riberalta**	**Riberalta**	**Rurrenabaque**	**Riberalta**	**Rurrenabaque**	**Cobija**	**Rurrenabaque**	**Rurrenabaque**	**Riberalta**
True origin	**Cobija**	**Riberalta**	**Cobija**	**Rurrenabaque**	**Riberalta**	**Rurrenabaque**	**Rurrenabaque**	**Rurrenabaque**	**Rurrenabaque**	**Riberalta**
Blind sample	2CO17	1CO19	2CO1	3CO12	1CO13	3CO11	3CO10	3CO14	3CO21	1CO25

Chapter 5

General discussion

Illegal logging contributes to forest degradation, loss of biodiversity and economic problems worldwide (Hoare, 2015; IUFRO, 2016). In most timber exporting countries, methods to certify and check timber species and their origin are currently based on paper certificates or digital control systems. In most importing countries, checks include more sophisticated methods such as physical barcoding and radio frequency identification (RFID) tagging (Seidel *et al.*, 2012; Dormontt *et al.*, 2015). However, relying on non-manipulable methods to identify and track timber remains a challenge. There is a great need for methods that use inherent wood characteristics to verify timber identity and source, thereby providing better evidence and opportunities for effectively counteracting the overharvesting and illegal trade of tropical timber.

Alternative methods to verify timber origin based on wood properties have shown promising results, including wood anatomy (Gasson, 2011; Gasson *et al.*, 2011; Wheeler, 2011; Moya *et al.*, 2013), mass spectrometry (Fidelis *et al.*, 2012; Lancaster and Espinoza, 2012; Espinoza *et al.*, 2015; Musah *et al.*, 2015; Deklerck *et al.*, 2017; Finch *et al.*, 2017), near-infrared spectroscopy (Braga *et al.*, 2011; Pastore *et al.*, 2011; Sandak *et al.*, 2011; Bergo *et al.*, 2016), stable isotopes (Horacek *et al.*, 2009; Kagawa and Leavitt, 2010; Förstel *et al.*, 2011), genetics (Degen *et al.*, 2006; Degen and Fladung, 2007; Lowe, 2007; Jolivet and Degen, 2012; Degen *et al.*, 2013; Lowe *et al.*, 2016) and dendrochronology (Domínguez-Delmás *et al.*, 2013). In this thesis, I assessed 3 methods to identify the species and geographic origin of tropical timber. Two methods were based on the chemical properties – Direct Analysis in Real Time Time-of-Flight-Mass Spectrometry (DART-TOFMS, **Chapter 2**) and stable isotopes ($\delta^{13}C$ and $\delta^{18}O$, **Chapter 3**) – while the third method was based on genetic composition using microsatellites (**Chapter 4**). I addressed the question of species identification with DART-TOFMS and stable isotopes, and geographic origin identification with each of the aforementioned methods. Spatial resolution for tracing timber origin was also assessed. Finally, I performed a comparison and a combination of the results of the various methods applied in this thesis (**this Chapter**).

In this chapter, I synthesize and compare results from the previous chapters. Comparison was done separately for species and site identification. Then, I combine the different methods. To do this, I re-run each method separately with a subset of the data to keep the sample size equal to the combined analyses. I discuss the importance of scale for timber tracing as well as similar tracing methods that can be applied to other biological materials or goods. Finally, based on the results, I provide recommendations for improving timber tracing methods.

Species identification: techniques and results

Species identification of traded tropical timber is challenging due to the lack of fertile plant material, which is usually necessary for plant identification (Dormontt *et al.*, 2015; Lowe *et al.*, 2016). Wood anatomic characteristics have been used as the most common method for verification (Gasson, 2011; Gasson *et al.*, 2011; Wheeler, 2011; Moya *et al.*, 2013). However, this requires high-level expertise and identification is usually made to the genus rather than the species level. Reliable species identification remains an important barrier to tracing as the wood of many tropical species looks similar and there is a lack of trained personnel.

I assessed the effectiveness of DART-TOFMS spectra and stable isotopes ($\delta^{13}C$ and $\delta^{18}O$) followed by multivariate statistical analysis to differentiate six similar *Cedrela* species, one of the most important and extensively traded tropical timbers (**Chapter 2**). Overall, the range for average species differentiation error with DART and stable isotopes was 19-24% (Table 5.1). Because stable isotopes have not been previously tested for tropical species identification, I cannot compare my results with published data. In contrast, the DART discrimination errors are higher compared to previous studies which reported an error of less than 10% for species discrimination (Lancaster and Espinoza, 2012; Musah *et al.*, 2015; Evans *et al.*, 2017).

Table 5.1 Kernel Discrimination Analysis (KDA) of *Cedrela* species based on DART-TOFMS (Direct Analysis in Real Time Time-of-Flight- Mass Spectrometry) and stable isotopes ($\delta^{13}C$ and $\delta^{18}O$). Same results as Chapter 2 and 3. Mean classification and standard deviation (SD) were estimated using the classification error per site after 100 runs with 80% of the data set for training in DART-TOFMS and stable isotopes (10-year-bulk samples of $\delta^{13}C$ and $\delta^{18}O$ data). Principal components were used as inputs for DART-TOFMS as reducing variables was needed.

Method	**DART-TOFMS**		**ISOTOPES**	
Species	**Mean Error (%)**	**SD (%)**	**Mean Error (%)**	**SD (%)**
C. angustifolia	26.5	28.0	*	*
C. balansae	46.1	36.8	98.6	7.2
C. fissilis	8.7	7.1	6.2	4.4
C. odorata	22.3	17.5	41.1	21.3
C. saltensis	20.2	32.4	88.3	28.0
Mean	**18.9**	**7.0**	**23.5**	**6.4**

*Small sample size

Large differences found in discrimination error between species can be explained by (1) low sample sizes for some species, (2) high variation within species, and/or (3) misidentification by the curator. I will discuss these causes below.

Low sample size can lead to high error rates. This was exemplified by including *C. montana*, with only two samples, in the discrimination analysis resulting in an increase in error rate from 15 to 30%. However, other studies applying DART-TOFMS to small sample sizes (3-7) have successfully discriminated between species (Lancaster and Espinoza, 2012; McClure *et al.*, 2015; Wiemann and Espinoza, 2017). This discrepancy depends on the degree of chemical variation, which is much smaller in some species than others, and indicates that the accuracy of discrimination is highly species-specific. In this thesis, DART-TOFMS (**Chapter 2**) and stable isotopes (**Chapter 3**) demonstrated limited potential to differentiate among *Cedrela* species. Only *C. fissilis* showed good prospects for discrimination from other species using both methods. Previous studies have found

that identification accuracy between samples from different genera such as *Dalbergia* and *Machaerium* was over 95% (Lancaster and Espinoza, 2012; Espinoza *et al.*, 2015). This suggests that DART-TOFMS analysis may perform better in distinguishing between *Cedrela* and similar-looking genera (e.g. *Carapa guianensis* Aubl. and *Swietenia macrophylla* King.), but is inadequate for discerning between species within genera.

DART-TOFMS and stable isotopes techniques did not show much improvement for differentiating between the *Cedrela* species compared to wood anatomy. However, care should be taken not to generalize these results, because these techniques can work well at discriminating between other species (Lancaster and Espinoza, 2012; Espinoza *et al.*, 2015). Consideration should also be taken to select data analyses because small but important diagnostic chemical compounds can be inadvertently excluded when running the statistical analysis. To avoid this problem I recommend the use of statistical machine learning tools together with a visual inspection of the compounds (for example, with the use of heatmap) before reaching a conclusive species classification.

Site identification: techniques and results

Site identification was performed using three methods: DART-TOFMS and stable isotopes, which are based on chemical composition and described in **Chapters 2** and **3,** respectively, and microsatellites, a genetics-based method described in **Chapter 4**. I applied these techniques to two of the most intensively exploited tropical timbers: *C. fissilis* and *C. odorata.* These species also had the largest sample sizes at multiple sites. Initially, the analyses on each tracing method were done separately for each species to enable comparison of results.

For *C. fissilis*, site classification error was higher using stable isotopes (70.48%) than mass spectrometry (DART-TOFMS) data (53.9%). Similar results were obtained for *C. odorata* using stable isotopes and mass spectrometry (62.5% and 47.7%, respectively). For the latter species, genetic (microsatellites) data resulted in a discrimination error of 33.7% (Table 5.2). Other studies have used multiple techniques to identify sample origin

(Degen, 2007; von Scheliha and Zahnen, 2011). However, because these studies applied different formats to report their statistical analyses, I cannot compare them with my results and assess the effectiveness of each method. In general, genetic techniques demonstrate higher accuracy than other methods. For example, site identification in *Erythrophleum* in Cameroon and the Republic of the Congo had an average accuracy of 35% using stable isotopes ($\delta^{13}C$, $\delta^{18}O$ and $\delta^{15}N$) and 92% using microsatellites (Vlam *et al.*, 2018). Another way to achieve higher accuracies is to use a different combination of stable isotopes. For example, Förstel *et al.* (2011) used $\delta^{13}C$, $\delta^{18}O$ and $\delta^{2}H$ to discriminate the origin of *Tectona grandis* to 100% among sites in Brazil, Costa Rica, Honduras and Panama.

Table 5.2 Error classification for sites of *C. fissilis* and *C. odorata* in Bolivia based on Kernel Discrimination Analysis (KDA). Same results as Chapter 2, 3, and 4. Mean classification and standard deviation (SD) were estimated using the classification error per site after 100 runs with 80% of the data set for training in DART-TOFMS and stable isotopes (10-year-bulk samples of $\delta^{13}C$ and $\delta^{18}O$ data). Genetic (microsatellite) analysis was based on 97% of the data for training set using 10,000 randomizations. Principal components were used as inputs for the DART-TOFMS and genetics methods.

Method	**DART-TOFMS**		**GENETICS**		**ISOTOPES**	
Sites/species	**Mean Error (%)**	**SD (%)**	**Mean Error (%)**	**SD (%)**	**Mean Error (%)**	**SD (%)**
***Cedrela fissilis* sites**						
Bajo Paraguá	45.8	36.5			43.3	27.7
Espejos	43.3	44.4			81.2	23.6
Concepción	37.5	36.8			82.9	19.3
Guarayos	39.7	32.2			66.8	24.5
Roboré	80.9	33.3			56.8	26.9
Yapacaní	60.9	36.8			99.7	3.3
Mean	**53.9**	12.5			**70.5**	9.4
***Cedrela odorata* sites**						
Cobija	47.8	38.5	49.2	50.0	82.3	31.3
Riberalta	38.4	38.6	8.2	27.5	27.8	37.3
Rurrenabaque	48.5	38.5	43.7	49.6	64.1	35.1
Mean	**47.7**	19.7	**33.7**	25.1	**62.5**	18.1

Possible explanations for differences in accuracy among the studied methods are local or large-scale environmental conditions and species-inherent characteristics. Local environmental conditions such as light, soil properties, rainfall/temperature regimes, and altitude may play a role in defining tree performance and wood composition (Förstel and Hützen, 1983; Gentry, 1995; Medina *et al.*, 1995; Oliveira-Filho *et al.*, 1998; McCarroll and Loader, 2004; Toledo *et al.*, 2011). However, the interaction of these factors with species-specific traits will define to what extent genetics or chemical characteristics will distinguish one site from another. One example is the root systems in the Meliaceae family. Noldt *et al.* (2001) found that some species were more sensitive to environmental conditions due to **root systems** concentrated in the upper soil layers. The *Cedrela* samples in our study also showed superficial tree roots and site-specific growth variation (Paredes-Villanueva *et al.*, 2016), indicating that these trees display site-specific characteristics that may have played an important role in wood formation. Another example is the **light requirement** of trees. Brienen *et al.* (2010) found a strong effect of light on *Cedrela*'s tree development, which supports *Cedrela*'s role as a relatively light-demanding canopy tree (Brienen and Zuidema, 2006; Brienen *et al.*, 2010). Finally, large-scale climatic drivers may also play a role in defining tree composition associated with site conditions. Low discrimination accuracies among sites suggest that factors operating at larger scales, like the El Niño-Southern Oscillation (**ENSO**), influence site-specific variation. ENSO variability was found to influence isotopic composition in *Cedrela* species from Bolivia (Vuille and Werner, 2005; Brienen *et al.*, 2012; Baker *et al.*, 2015). Each environmental signal on tree wood composition is determined by their effect on limiting growth (Helle and Schleser, 2004; Paredes-Villanueva *et al.*, 2016) and their scale (Kurita *et al.*, 2009).

Species reproductive strategies and ontogeny can also play an important role in the differentiation of timber origin. For example, results from different species have shown varying results, independent of spatial scale. This suggests that species-specific traits such as **pollination syndrome, seed dispersal syndrome, and mating strategy** can affect the genetic variation within sampled sites (Hedrick, 2011) and hence the potential for tracing. The *Cedrela* species used in this study are **pollinated by insects** (Bawa *et al.*,

1985; Howard *et al.*, 1995) and their **light, wing-shaped seeds** are dispersed by wind over long distances (Mostacedo *et al.*, 2003; Toledo *et al.*, 2008). The seed dispersal followed by a restricted pollination best explained the genetic clustering in some of our populations (substructure in Rurrenabaque and Cobija sites), as very local mating patterns (inbreeding) could have kept the established genetic cluster intact. In addition, the lack of spatial genetic substructure can also be due to regular gene **flow from neighbouring domestic trees** or **conservation** areas. Although samples were collected in natural forests to reduce the possibility that trees were recruited from non-natural seed sources, the Cobija samples were close to communities where trees may have been planted. Similarly, Rurrenabaque samples were located in human-disturbed sites along the border of Madidi National Park. Gene flow from the national park to disturbed areas may have contributed to genetic variation (Degen *et al.*, 2006; André *et al.*, 2008; Soliani *et al.*, 2016). Finally, ontogenetic characters such as individual **age or size** could affect tree growth variation (van der Sleen, 2014; Brienen *et al.*, 2017). I tested whether this effect influenced the discrimination potential of the isotope samples and found that diameter at breast height (DBH) and stable isotopes were not significantly correlated in *C. odorata*. When extreme DBH values were removed (2 samples with <15 and >70), mean discrimination accuracy only improved by 2.2%. DBH and stable isotopes were also not significantly correlated in *C. fissilis*. When extreme DBH values were removed (4 samples with <20 and >70) mean discrimination accuracy improved by a mere 1.4%. These results suggest that age is not an important factor affecting *Cedrela* species discrimination.

In this thesis, three methods were used to also assess site discrimination of *Cedrela* timber. Discrimination analysis (KDA) was ineffective for discriminating the geographical origin of *Cedrela* using DART-TOFMS and stable isotopes within Bolivia or at smaller scales. Genetic methods, in contrast, showed a higher discrimination of *Cedrela odorata* origin with 66.7% accuracy. *Cedrela* samples were distributed from highlands to lowlands and from dry to moist tropical forests. I expected that the highly variable environmental conditions (e.g. altitude and rainfall) of sample origins would be mirrored in *Cedrela* wood composition. However, this was not the case and my results suggest

complex interactions between species-specific traits and the environment. If future timber tracing efforts aim to use environmental variables as proxies for site origin, it is recommended that these complex interactions are taken into account. For example, one environmental variable that captures these interactions is evapotranspiration rate, which can consequently be used as a proxy for wood properties.

Comparing and combining timber tracing methods: challenges and opportunities

Determination of species and origin by each method can be accomplished by species-specific wood composition (this thesis). Previous studies have suggested that DART can be used either for species (Lancaster and Espinoza, 2012; Espinoza *et al.*, 2015) or site identification (Finch *et al.*, 2017). In this thesis, only *C. fissilis* could be discriminated with high accuracy among the 6 *Cedrela* species studied (**Chapter 2**) and the discrimination method was ineffective at discerning sites of origin. In contrast, stable isotopes were found to be suitable to trace timber origin (Kagawa and Leavitt, 2010; Förstel *et al.*, 2011). However, our assessment of $\delta^{13}C$ and $\delta^{18}O$ stable isotopes in both *C. fissilis* and *C. odorata* (**Chapter 3**) showed error rates of 60 to 70% for site identification. Overall, our results from evaluating three different methods suggest that site origin and wood composition are highly species-specific. When genetic methods (microsatellites) were applied to discriminate *C. odorata* sites of origin (**Chapter 4**), only one site showed high accuracy (Riberalta). Despite the formation of distinct genetic clusters, the low accuracy resulted from the mismatch of these clusters with their corresponding sample sites.

Criteria for selecting one method over others also includes the time of analysis, sample requirements needed, and equipment price and availability (Kagawa and Leavitt, 2010; Dormontt *et al.*, 2015). However, there is no scientific methodology capable of addressing all forensic questions about timber species identification and origin (Dormontt *et al.*, 2015) with high accuracies. To answer such questions, different methods have been applied individually but rarely have they been tested in combination

(von Scheliha and Zahnen, 2011; Johann Heinrich von Thünen Institute, 2015). Combining techniques to answer questions of species identification and site origin has been suggested as a viable approach to improve accuracy (von Scheliha and Zahnen, 2011; Dormontt *et al.*, 2015; Johann Heinrich von Thünen Institute, 2015). For example, samples of *Milicia excelsa* and *Entandrophragma cylindricum* species from Cameroon were characterized simultaneously using both stable isotope and genetic approaches. Combined results correctly identified ~94% of blind samples, a higher discrimination rate than that achieved by either method independently (von Scheliha and Zahnen, 2011; Dormontt *et al.*, 2015).

One of the challenges in using several methods simultaneously is the availability of a data set with all the measurements from these methods, especially given that sample size may decrease after DNA extraction, cellulose extraction, and other processes during the sample analyses. In this thesis, sample size limited the number of combined analyses to two tracing methods given that some samples were not assessed with all methods. This was due to the method-specific problems faced during either DNA or cellulose extraction or because the dominant trees selected for the isotope analysis were not the same as the trees with high quality DNA selected for the genetic analysis.

Despite these challenges, I managed to build a dataset that was suitable for the combined analysis of stable isotopes and DART-TOFMS. After Kernel Discriminant Analysis was applied to those species and sites for which stable isotopes ($\delta^{13}C$ and $\delta^{18}O$) and DART-TOFMS were measured both individually and in combination, accuracy improved only in some cases (Table 5.3). For species identification, combination of DART-TOFMS and stable isotopes improved accuracy relative to the isotopic method but not relative to DART-TOFMS (Supplementary Data Table S.5.1). For site identification, applying stable isotopes alone showed a slightly better performance than DART-TOFMS to discriminate the Bajo Paraguá site but not the Roboré and Yapacaní sites. However, combining both methods improved accuracy in geographic origin identification by 1-19% for Bajo Paraguá and Roboré but showed higher error for Yapacaní in comparison to using each method individually (Supplementary Data Table S.5.2). These results demonstrate little

or no improvement in identification accuracy when methods are combined and that it could even be detrimental in some cases. If using a method individually does not provide adequate identification accuracy, deciding whether or not to combine methods can be challenging. It might be unknown or there might not be preliminary data that indicate whether combining methods will lead to a greater precision.

Table 5.3 Classification errors of *Cedrela* species and sites in Bolivia using both stable isotopes and DART-TOFMS individually and in combination in Kernel Discrimination Analysis (KDA). Mean classification and standard deviation (SD) were estimated using the classification error per site after 100 runs with 80% of the data set for training in DART-TOFMS and stable isotopes (10-year-bulk samples of $\delta^{13}C$ and $\delta^{18}O$ data). A total of 60 samples were used for species identification and 30 samples for site identification. Each dataset was composed of the same sample size: 30 and 10 samples of species and sites group, respectively.

Species	Isotopes		DART		Isotopes + DART	
	Mean Error (%)	**SD (%)**	**Mean Error (%)**	**SD (%)**	**Mean Error (%)**	**SD (%)**
C. fissilis	**11.8**	12.8	**4.8**	7.9	**6.1**	8.4
C. odorata	**30.6**	18.6	**10.3**	13.9	**12.8**	13.8
Mean SCV Error	**21.4**	9.4	**7.3**	7.1	**9.7**	7.5
Sites						
Bajo Paraguá	**48.3**	36.6	**50.4**	39.1	**47.5**	38.1
Roboré	**59.5**	37.9	**49.6**	36.8	**40.8**	36.9
Yapacaní	**64.1**	39.1	**50.8**	39.1	**61.9**	37.4
Mean SCV Error	**58.5**	17.7	**52.3**	17.1	**53.7**	19.6

Another challenge regarding the combination of methods is standardizing analytical methods between different datasets. Studies on the individual application of the different methods have shown that identification accuracy can vary depending on the statistical analysis used. For example, use of Random Forest models to discriminate *Pericopsis* species with DART-TOFMS showed a higher discrimination power than assessing LOOCV (leave-one-out-cross-validation) in a KDA with a difference of 0.3-5% (Deklerck *et al.*, 2017). Moreover, the use of nearest neighbour analysis outperformed the commonly used frequency and Bayesian approach by 5-7% for site identification in *Sapelli* (Degen *et al.*, 2017). These results together with our comparison of KDA with Random Forests in

the DART-TOFMS dataset (**Chapter 2**) and the Bayesian approach in the microsatellites *Cedrela* dataset (**Chapter 4**) reinforce the idea that models should be selected based on their applicability and performance with particular data. Every method uses a different approach to optimize identification accuracy. Because the performance of each statistical method is context-specific (Montgomery *et al.*, 2007; Oulhote *et al.*, 2011), the choice of different models to analyse pooled data or the application of a standardized method might result in lower accuracies.

Finally, variation of wood properties depends on genetic (**Chapter 4**) and environmental factors (**Chapter 2** and **3**). In the case of isotopes, there is more uncertainty because variation is dependent on many variables and the interplay of complex physiological interactions. Moreover, little is known about the effect of the atmospheric values of pre- and post-industrial $\delta^{13}C$ on isotopic discrimination (Helle and Schleser, 2004). Understanding the factors causing this variation will greatly improve the selection of the best methods to use individually or in combination based on the specific case and context.

How and when to combine the different timber tracing methods remains a challenge and is currently under debate among experts. Previous studies suggest that combining methods could improve identification accuracy (von Scheliha and Zahnen, 2011). For example, preliminary analyses suggested that NIRS (near-infrared spectroscopy) could ideally complement stable isotopes of *Tectona grandis* and *Swietenia* species from Latin America and Asia (Johann Heinrich von Thünen Institute, 2015) to improve accuracy. However, our study showed that identification accuracy might not only depend on the species and sites under analysis but also on the type of statistical analysis. These differences in accuracy should be taken into account if pooling methods are going to be performed.

Need for tracing at different spatial scales

Illegal trade in timber is a worldwide environmental problem that causes negative impacts at both large and small scales. It has been estimated that 10% to 80% of the total timber trade is illegal (Seneca Creek Associates, 2004) and in some countries such as Papua New Guinea, Liberia, and those comprising the Amazon (Stark and Pang Cheung, 2006; Lawson and MacFaul, 2010; Wit *et al.*, 2010), this percentage has been as much as 80-90% of all logging operations. Although most legislative measures focus on combating international illegal trade, a high proportion (70-90%) of illegal tropical timber is traded in domestic markets (Cerutti and Lescuyer, 2011; Kishor and Lescuyer, 2012; Lescuyer *et al.*, 2014). Furthermore, much illegal logging occurs in non-permitted areas or close to official logging concessions (ABT, 2017). Both the large and small scale dimensions of illegal logging could be addressed with forensic techniques that independently verify the origin of traded timber in both domestic and international markets.

In Bolivia, as in other timber producing countries, illegal logging and trade continue to be high (ABT, 2017) despite a decrease in illegality from 92% in 2012 to 47% in 2017 (ABT, 2017). The mode of smuggling at smaller scales is related to a different mechanism where most timber is commercialized locally or transported in private vehicles to cities. Timber commercialized in this way does not represent high quantities or economic value and is generally not accompanied by proper legal documentation. Such small-scale illegal trade is considered to be informal harvesting and is usually difficult to identify and control. The high percentage of illegality that still exists in the industry indicates a critical need for better controls at both local and national scales.

Clearly, independent tools to verify tropical timber species and origin are also needed at the national level. Are the forensics methods proposed in this thesis able to work within the country? I applied chemical (DART-TOFMS and stable isotopes) and genetic (microsatellites) analyses to properties inherent in wood to assess their potential for species and origin identification. I also assessed different scales at which differentiation

could be achieved. Such site characteristics vary from large scale (e.g., ecosystems under different climate regimes) to small scale (e.g., microsite factors that contribute to tree development; Reifsnyder *et al.*, 1971). For example, topography, soil chemistry, hydrology and microclimate drive variation within forests at small spatial scales (Tiessen *et al.*, 1994; Gentry, 1995; Xia *et al.*, 2016; Jucker *et al.*, 2018) and variation in tree performance and composition (Medina *et al.*, 1995; Oliveira-Filho *et al.*, 1998; Toledo *et al.*, 2011; Paredes-Villanueva *et al.*, 2016). In addition to the effects of these factors, I found the variation in identification errors to be dependent on several factors particular to each of the methods studied.

First, differences in site identification accuracies could be explained, in part, by the variation in distances between sample sites. Site discrimination error is expected to be lower for samples from distant sites, which was exemplified by my result showing *C. odorata* sites to be more distant from one another than *C. fissilis* sites. Spatial resolution covered by DART-TOFMS measurement was from 7 m to 503 km among *C. fissilis* samples and 85 m to 613 km for *C. odorata* samples within Bolivia. Site discrimination error was 43-54% for *C. fissilis* and 42-48% for *C. odorata*. When only Bajo Paraguá, Roboré and Yapacaní (the most distant sites of *C. fissilis*) were analyzed with DART-TOFMS, the accuracy remained similar with Random Forest (57%) and increased to 53% with KDA (data not shown). Previous studies found discrimination errors of 24-30% at a distance of ~35–65 km of *Pseudotsuga menziesii* sites (Cascade and Coast) in Western Oregon (Finch *et al.*, 2017). Overall, these results suggest that discriminating between more distant regions or locations may result in higher accuracies than discriminating among neighboring sites.

Second, spatial resolution covered by *C. odorata* genetic data (microsatellites) ranged from 26 m to 619 km and classification errors ranged from 10% to 50%. These results show a similar error range to a previous study on *Swietenia macrophylla*, which showed an error assignment of 29.3% to site of origin (Degen *et al.*, 2013). Another study on *Hymenaea courbaril* found a mean classification error of 11-12% (Chaves *et al.*, 2018). The samples of Degen *et al.* (2013) and Chaves *et al.* (2018) were analysed at large

geographical scales: with sample distance of 5-5660 km and 156-1384 km, respectively. Our analyses revealed several limitations and the assignment accuracy was dependent not only on the number of sample sites and genetic differentiation (Cavers *et al.*, 2005) but also on the spatial alignment of the different genetic clusters with the sites distribution. Both the spatial resolution and assignment accuracy depend on geographical scale and sampling scheme. In addition, previous studies found that identifying the country of origin was more accurate (82.2%) than identifying the specific site of origin (70.7%; Degen *et al.*, 2013) and that only after merging common genetic clusters did mean error decrease from 67-72% to 48-49% (Jolivet and Degen, 2012) and from 95.5 to 4.2% error (Vlam *et al.*, 2018). The 81 samples used in this study showed differences in spatial discrimination due to the high degree of genetic mixing within sites. To improve insights into the degree of genetic mixing, it is necessary to adjust sampling strategies according to the level of differentiation and local variation. It is also important to study the history of the sampling sites. For example, how long have sites been isolated from each other and how much gene flow has occurred? This can be analyzed with isolation-with-migration models or demographic analyses using Approximate Bayesian Computation (ABC) models (Beaumont, 2010; Hey, 2010).

Third, isotopic discrimination analysis covered a spatial resolution of 3 m to 501 km for *C. fissilis* and 85 m to 613 km for *C. odorata* samples resulting in classification errors that range from 43% to 99.7% for *C. fissilis* and from 28% to 82% for *C. odorata* samples. The average accuracy was 30% for *C. fissilis* and 38% for *C. odorata* site classification. Similar results have been found for site identification of *Erythrophleum* species in Cameroon and the Republic of the Congo with an average accuracy of 35% (Vlam *et al.*, 2018). This same study also found high variability in the classification accuracy ranging from 46% to 99% at lower spatial scales (14-216 km; Vlam *et al.*, 2018). Despite the fact that our sites covered a wide range of altitudes and rainfall regimes (224-635 m.a.s.l. of altitude and 1045-1860 mm of precipitation in *C. fissilis* sites and 145-309 m.a.s.l. of altitude and 1796-2057 mm of precipitation in *C. odorata* sites), results indicated that some sites and species have greater potential to be traced than others. In addition, low discrimination accuracies among sites suggested that factors operating at larger scales, like soils and El

Niño-Southern Oscillation (ENSO), influence site-specific variations. Research on $\delta^{13}C$ and $\delta^{15}N$ in coca (*Erythroxylum coca*) from countries in South America found significant ratio differences among the sample sites and attributed this to differences of soils (for $\delta^{15}N$) and the length of wet season (for $\delta^{13}C$; Ehleringer *et al.*, 2000). Soil characteristics such as soil water availability and nutrients can affect tree growth (Toledo *et al.*, 2011) particularly in dry forests (Medina *et al.*, 1995; Oliveira-Filho *et al.*, 1998). This impact shows a spatial gradient (Murphy and Lugo, 1986; Ceccon *et al.*, 2006) along which soil characteristics become less important as one shifts from dry to wetter forests. Eventually light variation becomes a more important factor for growth in wet forests (Engelbrecht *et al.*, 2007; Brienen *et al.*, 2010). ENSO variability was also found to influence isotopic composition in *Cedrela* species from Bolivia (Vuille and Werner, 2005; Brienen *et al.*, 2012; Baker *et al.*, 2015). Vuille and Werner (2005) found that $\delta^{18}O$ was significantly more enriched during the El Niño when precipitation decreases and depleted during La Niña events. How stable isotopes composition varies will depend on the site and species-specific response to these changing environmental factors (Helle and Schleser, 2004; Poussart and Schrag, 2005; van der Sleen *et al.*, 2017). The species and site specific variation should be assessed before performing isotopic discrimination analysis for timber tracing purposes.

In conclusion, comparison of the three methods applied in this study – DART-TOFMS, microsatellites and stable isotopes – suggests that the identification accuracy is highly dependent not only on the chemical and genetic variation but also on the alignment of the different clusters with the spatial organization (i.e. study sites distribution). The spatial variation in *Cedrela* wood composition in Bolivia does not necessarily follow demographic or political boundaries. The classic recommendation has been to increase the sample size of datasets. However, increasing sample size does not ensure improvement of resolution or discrimination among sites/species. Based on the results of this thesis, accuracy is highly species and site specific. Increasing sample size will probably not change the outcome of this analysis. A more effective approach would be to conduct a pilot study and power analysis to determine how large the sample size should be and check if it makes biological sense. In addition, a more practical approach would be

to treat each chemically or genetically distinct cluster as a different unit in which initiatives to stop illegal timber harvesting are coordinated by the related governments. To this end, standardized data collection and sample measurement are needed at larger scales together with sampling strategies that are adjusted to site variation at smaller scales. If the resolution will be tested or a combination of methods applied, sampling schemes should not be restricted to method-specific sampling criteria, such as sampling from distant trees to minimize the probability of sampling relatives (Gillies *et al.*, 1999). A larger and more widespread dataset is needed, namely a reference database following a standardized sampling scheme across the distribution of the species at different distances including neighbouring countries. Standardized sampling and data collection will enable the use of data from other locations and countries when necessary. By expanding the database for a species to other countries and sampling between sites (e.g. sampling across a gradient) it will also be possible to identify the site of origin at finer scales and with higher precision.

Applications of tracing methods to other biological studies and implications of forensics use for timber tracing methods

The methods applied in this thesis – DART-TOFMS, stable isotopes and microsatellites – showed great potential for tracing different goods and wildlife products. For example, DART was used to discriminate among biofuel (soy and canola) feedstocks (98% accuracy in LOOCV), insect species (100% accuracy in LOOCV), psychoactive plant products (*Datura*, *Brugmansia* and *Hyocyamus* seeds; 96% accuracy in LOOCV), and *Eucalyptus* (leaves) species (83% accuracy in LOOCV; Musah *et al.*, 2015). Isotopic analyses, which have also been applied to multiple goods and forensic investigations (Rossmann, 2001; West *et al.*, 2010; Oulhote *et al.*, 2011), were first used to determine the origin of wine and perform adulteration tests (75% accuracy in LDA; Versini *et al.*, 1997). Further applications were performed on honeysuckle plants (96% cross-validation of multi-elements in LDA; Fan *et al.*, 2018), teas (100% accuracy with PCA, and hierarchical cluster analysis; Cengiz *et al.*, 2017), and cocaine (90-96% with bivariate

mean and standard deviation parameters; Ehleringer *et al.*, 2000). Microsatellites have also been applied to individual forensic identification of parrots (0.0001-0.000003% in exclusion probability) and their parentage (99.5-100% accuracy exclusion probability; Jan and Fumagalli, 2016). In general, identification using the different methods varied with a very low accuracy of 0.0001% for identification of parrots using microsatellites to 100% accuracy for insect species identification using DART and tea origin identification using stable isotopes. These results suggest a great potential for identification of timber samples or products. This was confirmed by previous analyses that tested the applicability of chemical and genetic properties in wood traceability (Boner *et al.*, 2007; Degen, 2007; Kagawa and Leavitt, 2010; Förstel *et al.*, 2011; Lancaster and Espinoza, 2012; Degen *et al.*, 2013; Musah *et al.*, 2015).

The results of the methods used in this thesis showed a great potential for tropical timber traceability and for forensic purposes. Timber traceability applications could be used to either confirm the origin of legal timber or identify illegal logging. The latter should be linked to forensic procedures. The procedural rules and evidentiary limitations and requirements may differ in each jurisdiction. Determining if laboratory analytical methods could be admissible will be determined by the reliability of their results but to date such requirements are not included in regulations concerning wood identification (UNODC, 2016). In general, it has been suggested that an identification accuracy higher than 90% is necessary to provide reliable results and to use these techniques as forensic methods. Although this requirement is difficult to be achieved up to date, results could still provide useful timber origin/species information when measurements are done in real-world conditions and errors and risk of inaccuracies are reported (Koehler, 2017). Interpretation and communication of results should be restricted to the evidence from the data and following an approach accepted by legal systems. A guide for communication of scientific results of timber tracing analyses in a clear and concise manner was presented by UNODC (2016).

The translation of the scientific methods assessed in this thesis will require us to take practical decisions. The first decision involves strengthening of the law enforcement in

timber producing countries. Tropical countries rely on regulations related to forest management, conservation (e.g. national red lists of plant species), and export of forest products. However, even relying on these regulations, enforcement is not effective due to the lack of readily available tools to check paper-based declarations. The techniques assessed in this thesis could be used along the different stages of the chain of custody (Lowe *et al.*, 2016). The second decision requires transfer of knowledge by capacity building and training of local forest stakeholders. This will allow identification of illegal logging at earlier stages of the timber chain of custody in a more effective way. Lastly, it is necessary to include guidelines that identify steps to follow when a case of suspected origin is present. These guidelines should not only be implemented at an international level but should also be connected to local regulations. As Seymour and Busch (2018) said "...*the biggest part of the problem, needs to be the biggest part of the solution*".

In the last years, forest certifications were a good approach to promote sustainable forest management and ensure that harvesting is not counterproductive. However, the evaluations take time and effort and they are still paper-based, a system prone to substitution, inclusion or mixing with prohibited timber (Johnson and Laestadius, 2011). The proposed alternative techniques could help to protect the integrity of forest companies and all certification schemes (bodies) (Lowe *et al.*, 2016) in this regard. What can we do as researchers? Many scientists tend to prefer working from their ivory towers. However, the academic efforts should be connected to the practical concerns of the daily forest management tasks ensuring that the timber tracing tools are ready when the law enforcers need them.

Recommendations for timber tracing initiatives

Given the conflictive high demand of timber, "the solution to reducing illegal logging may ultimately rest on future means of distinguishing between legally and illegally harvested timber" (ABT Associates Inc., 2006). With this research I obtained data from *Cedrela*'s chemical, isotopic and genetic distribution and description of timber tracing methods in Bolivia, to be used for legal procedures by local authorities. Future initiatives should include strengthening participation of consuming countries, collaboration among laboratories and collaboration with accredited forensics laboratories for specific cases, all of which will require standardization of methods.

Firstly, involving local stakeholders in producing countries is crucial as they are part of the initial process of timber harvesting. The more effective identifications of origin and species are those implemented at early stages of the timber chain of custody (Lowe *et al.*, 2016). Currently, the timber tracing methods assessed in this thesis are already being used by some consuming countries but their applicability in producing countries is limited. While the provision of equipment could potentially give a boost to their use in these countries, it would not ensure their long-term application and the survival of local laboratories. Instead, capacity building could be focused on identifying strengths or techniques that have already started or have been developed and on people that could commit to not only apply the acquired knowledge but also transfer it to other local stakeholders. One first step to involve local stakeholders could be strengthening already existing capacities, e.g. identification based on wood anatomy. Wood anatomy can reliably determine genus and occasionally species (UNODC, 2016). However, it requires expertise and needs taxonomically validated reference collections (Dormontt *et al.*, 2015). This technique is mainly used on highly traded species and could potentially be applied to other commercial timber species. When identification is not possible with local efforts, a second step would involve collaboration with other laboratories to use techniques that are not readily available *in situ*.

Finally, it will be needed, to collaborate with other laboratories and join task forces, a standardization of sampling and collection of data. Additional samples from the field are not necessary when the density is appropriate (von Scheliha and Zahnen, 2011) and samples are more widespread. One aspect in which sampling does present challenges is when the sampling aims are different. Whether the sampling is designed to solely understand the technique or to resemble real-world conditions may limit their application. Establishing clear sampling objectives on a step-by-step basis will require time and resources but will ultimately result in more applicable and efficient tools. While these methods are still under development or even when low accuracies are the outcome, information could be cross-referenced with declared documentation and information collected in the field and submitted to a true/false analysis of the claimed origin. This thesis represents a first step for techniques exploration and comparison in the context of a producing country like Bolivia and my results will provide more understanding of the techniques and tools for decision making on how and when they could assist with identification of timber species and geographical origin effectively.

Supplementary Data

Table S.5.1 Classification matrix of *Cedrela* species based on stable isotopes, DART-TOFMS, and a combination of both methods

Classification (%)	Isotopes		DART		Isotopes + DART	
	C. fissilis	*C. odorata*	*C. fissilis*	*C. odorata*	*C. fissilis*	*C. odorata*
C. fissilis	**88.18**	11.82	**95.21**	4.79	**93.88**	6.12
SD	12.81	12.81	7.91	7.91	8.41	8.41
C. odorata	30.64	**69.36**	10.32	**89.68**	12.82	**87.18**
SD	18.62	18.62	13.95	13.95	13.75	13.75

Table S.5.2 Classification matrix of sites of origin based on stable isotopes, DART-TOFMS, and a combination of both methods

Classification (%)	Isotopes			DART			Isotopes + DART		
	Bajo Paraguá	Roboré	Yapacaní	Bajo Paraguá	Roboré	Yapacaní	Bajo Paraguá	Roboré	Yapacaní
Bajo Paraguá	**51.67**	31.58	16.75	**49.58**	34.00	16.42	**52.50**	23.17	24.33
SD	36.55	32.59	30.37	39.09	36.56	29.10	38.08	30.47	31.42
Roboré	43.50	**40.50**	16.00	41.25	**50.42**	8.33	35.75	**59.17**	5.08
SD	38.56	37.98	26.98	33.21	36.76	19.14	35.79	36.88	15.35
Yapacaní	47.58	16.50	**35.92**	36.83	14.00	**49.17**	43.08	18.83	**38.08**
SD	42.31	31.83	39.14	39.16	26.90	39.11	39.64	29.46	37.35

What we are doing to the forests of the world is but a mirror reflection of what we are doing to ourselves and to one another.

— Chris Maser, Forest Primeval: The Natural History of an Ancient Forest

References

ABT, 2017. Audiencia pública inicial y parcial de rendición de cuentas (de enero hasta agosto, gestión 2017) [PowerPoint slides]. In. Autoridad de Fiscalización y Control Social de Bosques y Tierra (ABT), Santa Cruz, Bolivia.

ABT Associates Inc., 2006. Illegal logging: a market-based analysis of trafficking in illegal timber. Trends in Organized Crime 10, 61-64.

Afendi, F.M., Okada, T., Yamazaki, M., Hirai-Morita, A., Nakamura, Y., Nakamura, K., Ikeda, S., Takahashi, H., Altaf-Ul-Amin, M., Darusman, L.K., Saito, K., Kanaya, S., 2012. KNApSAcK family databases: integrated metabolite–plant species databases for multifaceted plant research. Plant and Cell Physiology 53, e1(1-12).

Aggarwal, P.K., Fröhlich, K., Kulkarni, K.M., Gourcy, L.L., 2004. Stable isotope evidence for moisture sources in the asian summer monsoon under present and past climate regimes. Geophysical Research Letters 31.

André, T., Lemes, M.R., Grogan, J., Gribel, R., 2008. Post-logging loss of genetic diversity in a mahogany (*Swietenia macrophylla* King, Meliaceae) population in Brazilian Amazonia. Forest Ecology and Management 255, 340-345.

Andrew, R.L., Wallis, I.R., Harwood, C.E., Foley, W.J., 2010. Genetic and environmental contributions to variation and population divergence in a broad-spectrum foliar defence of *Eucalyptus tricarpa*. Annals of Botany 105, 707-717.

Baker, J.C.A., Hunt, S.F.P., Clerici, S.J., Newton, R.J., Bottrell, S.H., Leng, M.J., Heaton, T.H.E., Helle, G., Argollo, J., Gloor, M., Brienen, R.J.W., 2015. Oxygen isotopes in tree rings show good coherence between species and sites in Bolivia. Global and Planetary Change 133, 298-308.

Baudat, G., Anouar, F., 2000. Generalized discriminant analysis using a kernel approach. Neural Computation 12, 2385-2404.

Bawa, K.S., Bullock, S.H., Perry, D.R., Coville, R.E., Grayum, M.H., 1985. Reproductive biology of tropical lowland rain forest trees. II. Pollination systems. American Journal of Botany 72, 346-356.

Beaumont, M.A., 2010. Approximate Bayesian computation in evolution and ecology. Annual Review of Ecology, Evolution, and Systematics 41, 379-406.

Bergo, M.C.J., Pastore, T.C.M., Coradin, V.T.R., Wiedenhoeft, A.C., Braga, J.W.B., 2016. NIRS identification of *Swietenia macrophylla* is robust across specimens from 27 countries. IAWA Journal 37, 420-430.

Blaser, J., 2010. Forest law compliance and governance in tropical countries: a region-by-region assessment of the status of forest law compliance and governance in the tropics, and recommendations for improvement. In: Sarre, A. (Ed.). FAO/ITTO, p. 27.

BOLFOR/FMT, 2003. Diámetros mínimos de corta en bosques tropicales de Bolivia: recomendaciones basadas en la investigación forestal. In. Proyecto BOLFOR Forest Management Trust, Santa Cruz, Bolivia, p. 8.

Boner, M., Sommer, T., Erven, C., Förstel, H., 2007. Stable isotopes as a tool to trace back the origin of wood. In: Degen, B. (Ed.), Proceedings of the international workshop "Fingerprinting methods for the identification of timber origins. Johann Heinrich von Thünen-Institut, pp. 3-5.

Booth, A.L., Wooller, M.J., Howe, T., Haubenstock, N., 2010. Tracing geographic and temporal trafficking patterns for marijuana in Alaska using stable isotopes (C, N, O and H). Forensic Science International 202, 45-53.

Braga, J.W.B., Pastore, T.C.M., Coradin, V.T.R., Camargos, J.A.A., da Silva, A.R., 2011. The use of near Infrared Spectroscopy to identify solid wood specimens of (Cites Appendix II). IAWA Journal 32, 285-296.

Bräuning, A., Volland-Voigt, F., Burchardt, I., Ganzhi, O., Nauss, T., Peters, T., 2009. Climatic control of radial growth of *Cedrela montana* in a humid mountain rainforest in southern Ecuador. Erdkunde 63, 337-345.

Breiman, L., 2001. Random forests. Machine Learning 45, 5-32.

Brienen, R.J., Helle, G., Pons, T.L., Guyot, J.-L., Gloor, M., 2012. Oxygen isotopes in tree rings are a good proxy for Amazon precipitation and El Niño-Southern Oscillation variability. Proceedings of the National Academy of Sciences 109, 16957-16962.

Brienen, R.J., Zuidema, P.A., 2005. Relating tree growth to rainfall in Bolivian rain forests: a test for six species using tree ring analysis. Oecologia 146, 1-12.

Brienen, R.J.W., Gloor, E., Clerici, S., Newton, R., Arppe, L., Boom, A., Bottrell, S., Callaghan, M., Heaton, T., Helama, S., Helle, G., Leng, M.J., Mielikäinen, K., Oinonen, M., Timonen, M., 2017. Tree height strongly affects estimates of water-use efficiency responses to climate and CO_2 using isotopes. Nature Communications 8, 288.

Brienen, R.J.W., Schöngart, J., Zuidema, P.A., 2016. Tree rings in the tropics: insights into the ecology and climate sensitivity of tropical trees. In, Tropical Tree Physiology: Adaptations and Responses in a Changing Environment. Springer International Publishing, Cham, pp. 439-461.

Brienen, R.J.W., Zuidema, P.A., 2006. Lifetime growth patterns and ages of Bolivian rain forest trees obtained by tree ring analysis. Journal of Ecology 94, 481-493.

Brienen, R.W., Zuidema, P., Martínez-Ramos, M., 2010. Attaining the canopy in dry and moist tropical forests: strong differences in tree growth trajectories reflect variation in growing conditions. Oecologia 163, 485-496.

Broadmeadow, M.S.J., Griffiths, H., Maxwell, C., Borland, A.M., 1992. The carbon isotope ratio of plant organic material reflects temporal and spatial variations in CO_2 within tropical forest formations in Trinidad. Oecologia 89, 435-441.

Cárdenas, D., Arboleda, N.C., Tunjano, S.S., Barrera, L.Q., Rodríguez, M.B., Rodríguez, S.G., Rincón, L.M., Martín, L.E.R., Castañeda, M.R., Álvarez, H.A., Peinado, Á.J.V., Matajira, J.C.C., Sánchez, A.G., Echeverry, J.C.G., Arias, Á.M., Gutierrez, C.A., Rivera, L.L., Velásquez, M.M., Pedraza, L.M., Villate, G.C.M., 2015. Planes de manejo para la

conservación de abarco, caoba, cedro, palorosa y canelo de los andaquíes. Instituto Amazónico de Investigaciones Científicas, SINCHI, Colombia, p. 201.

Cavers, S., Degen, B., Caron, H., Lemes, M.R., Margis, R., Salgueiro, F., Lowe, A.J., 2005. Optimal sampling strategy for estimation of spatial genetic structure in tree populations. Heredity 95, 281-289.

Cavers, S., Navarro, C., Lowe, A., 2003. Chloroplast DNA phylogeography reveals colonization history of a Neotropical tree, *Cedrela odorata* L., in Mesoamerica. Molecular Ecology 12, 1451-1460.

Cavers, S., Navarro, C., Lowe, A., 2004. Targeting genetic resource conservation in widespread species: a case study of *Cedrela odorata* L. Forest Ecology and Management 197, 285-294.

Cavers, S., Telford, A., Arenal Cruz, F., Pérez Castañeda, A., Valencia, R., Navarro, C., Buonamici, A., Lowe, A., Vendramin, G., 2013. Cryptic species and phylogeographical structure in the tree *Cedrela odorata* L. throughout the Neotropics. Journal of Biogeography 40, 732-746.

Ceccon, E., Huante, P., Rincón, E., 2006. Abiotic factors influencing tropical dry forests regeneration. Brazilian Archives of Biology and Technology 49, 305-312.

Cengiz, M.F., Turan, O., Ozdemir, D., Albayrak, Y., Perincek, F., Kocabas, H., 2017. Geographical origin of imported and domestic teas (*Camellia sinensis*) from Turkey as determined by stable isotope signatures. International Journal of Food Properties 20, 3234-3243.

Cerutti, P.O., Lescuyer, G., 2011. The domestic market for small-scale chainsaw milling in Cameroon: present situation, opportunities and challenges. Center for International Forestry Research (CIFOR), Bogor, Indonesia, p. 38.

CFB, 2018. Comercio exterior de productos forestales de Bolivia 2017. In. Cámara Forestal de Bolivia, Santa Cruz de la Sierra, Bolivia.

CFV, 2017. Empresas forestales certificadas en Bolivia. In, Boletín informativo. Diciembre 2017. Consejo Boliviano para la Certificación Forestal Voluntaria (CFV), Santa Cruz, Bolivia.

Chaplin, G.E., 1980. Progress with provenance exploration and seed collection of *Cedrela* spp.

Chase, M.W., Hills, H.H., 1991. Silica gel: an ideal material for field preservation of leaf samples for DNA studies. Taxon 40, 215-220.

Chatterjee, A., Chakrabortty, T., Chandrasekharan, S., 1971. Chemical investigation of *Cedrela toona*. Phytochemistry 10, 2533-2535.

Chaves, C.L., Degen, B., Pakull, B., Mader, M., Honorio, E., Ruas, P., Tysklind, N., Sebbenn, A.M., 2018. Assessing the ability of chloroplast and nuclear DNA gene markers to verify the geographic origin of Jatoba (*Hymenaea courbaril* L.) timber. Journal of Heredity 109, 543–552.

CITES, 2017. Convention on International Trade in Endangered Species of Wild Fauna and Flora. Appendices I, II and III. In. Convention on International Trade in Endangered Species of Wild Fauna and Flora.

CITES, 2018. Regulation of trade in specimens of species included in Appendix III. In. Convention on International Trade in Endangered Species of Wild Fauna and Flora.

Cody, R.B., Laramée, J.A., Durst, H.D., 2005. Versatile new ion source for the analysis of materials in open air under ambient conditions. Analytical chemistry 77, 2297-2302.

Compt, J., Christy, T., 2008. The 14th meeting of the Conference of the Parties to CITES. Traffic Bulletin 21, 101.

Cook, C.S., Erkkila, B.R., Chakraborty, S., Tipple, B.J., Cerling, T.E., Ehleringer, J.R., 2018. Stable isotope biogeochemistry and ecology. Laboratory Manual. University of Utah, Salt Lake City, USA, p. 173.

Cordeiro, J.R., Martinez, M.I.V., Li, R.W.C., Cardoso, A.P., Nunes, L.C., Krug, F.J., Paixão, T.R.L.C., Nomura, C.S., Gruber, J., 2012. Identification of four wood species by an electronic nose and by LIBS. International Journal of Electrochemistry 2012, 5.

Cornuet, J.M., Luikart, G., 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144, 2001-2014.

Cullen, L.E., Grierson, P.F., 2007. A stable oxygen, but not carbon, isotope chronology of *Callitris columellaris* reflects recent climate change in north-western Australia. Climatic Change 85, 213-229.

Dansgaard, W., 1964. Stable isotopes in precipitation. Tellus 16, 436-468.

Degen, B., 2007. Proceedings of the international workshop “Fingerprinting methods for the identificarion of timber origins". Johann Heinrich von Thünen-Institut, p. 66.

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.

Degen, B., Blanc, L., Caron, H., Maggia, L., Kremer, A., Gourlet-Fleury, S., 2006. Impact of selective logging on genetic composition and demographic structure of four tropical tree species. Biological Conservation 131, 386-401.

Degen, B., Fladung, M., 2007. Use of DNA-markers for tracing illegal logging. In: Degen, B. (Ed.), Proceedings of the international workshop “Fingerprinting methods for the identificarion of timber origins. Johann Heinrich von Thünen-Institut, pp. 6-14.

Degen, B., Ward, S., Lemes, M., Navarro, C., Cavers, S., Sebbenn, A., 2013. Verifying the geographic origin of mahogany (*Swietenia macrophylla* King) with DNA-fingerprints. Forensic Science International: Genetics 7, 55-62.

Deguilloux, M., Pemonge, M., Petit, R., 2002. Novel perspectives in wood certification and forensics: dry wood as a source of DNA. Proceedings of the Royal Society of London B: Biological Sciences 269, 1039-1046.

Deklerck, V., Finch, K., Gasson, P., Van den Bulcke, J., Van Acker, J., Beeckman, H., Espinoza, E., 2017. Comparison of species classification models of mass spectrometry data: Kernel Discriminant Analysis vs Random Forest; a case study of Afrormosia (*Pericopsis elata* (Harms) Meeuwen). Rapid Communications in Mass Epectrometry 31, 1582-1588.

Domínguez-Delmás, M., Nayling, N., Ważny, T., Loureiro, V., Lavier, C., 2013. Dendrochronological dating and provenancing of timbers from the Arade 1 shipwreck, Portugal. International Journal of Nautical Archaeology 42, 118-136.

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.

Dünisch, O., Bauch, J., Gasparotto, L., 2002. Formation of increment zones and intraannual growth dynamics in the xylem of *Swietenia macrophylla*, *Carapa guianensis*, and *Cedrela odorata* (Meliaceae). IAWA Journal 23, 101-119.

Duong, T., 2007. ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R. Journal of Statistical Software 21, 1-16.

Duong, T., 2017. ks: kernel smoothing. In. https://cran.r-project.org/web/packages/ks/.

Earl, D.A., vonHoldt, B.M., 2012. Structure harvester: a website and program for visualizing structure output and implementing the Evanno method. Conservation Genetics Resources 4, 359-361.

Eason, H.M., Setzer, W.N., 2007. Bark essential oil composition of *Cedrela tonduzii* C. DC. (Meliaceae) from Monteverde, Costa Rica Records of Natural Products 1, 24-27.

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.

Ellegren, H., 2004. Microsatellites: simple sequences with complex evolution. Nature Reviews Genetics 5, 435-445.

Engelbrecht, B.M., Comita, L.S., Condit, R., Kursar, T.A., Tyree, M.T., Turner, B.L., Hubbell, S.P., 2007. Drought sensitivity shapes species distribution patterns in tropical forests. Nature 447, 80-82.

English, N.B., Betancourt, J.L., Dean, J.S., Quade, J., 2001. Strontium isotopes reveal distant sources of architectural timber in Chaco Canyon, New Mexico. Proceedings of the National Academy of Sciences 98, 11891-11896.

Espinoza, E.O., Wiemann, M.C., Barajas-Morales, J., Chavarria, G.D., McClure, P.J., 2015. Forensic analysis of CITES-protected *Dalbergia* timber from the Americas. IAWA Journal 36, 311-325.

Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14, 2611-2620.

Evans, P.D., Mundo, I.A., Wiemann, M.C., Chavarria, G.D., McClure, P.J., Voin, D., Espinoza, E.O., 2017. Identification of selected CITES-protected Araucariaceae using DART TOFMS. IAWA Journal 38, 266-S263.

Falush, D., Stephens, M., Pritchard, J.K., 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164, 1567-1587.

Fan, L., Chen, L., Ding, R., Wang, L., Zhang, B., 2018. Geographical discrimination of honeysuckle (*Lonicera japonica* Thunb.) from China by characterization of the stable isotope ratio and multielemental analysis. Analytical Letters 51, 2507-2516.

FAO, 2014. Agriculture, forestry and other land use emissions by sources and removals by sinks. In. Climate, Energy and Tenure Division, FAO, p. 75.

Fidelis, C.H.V., Augusto, F., Sampaio, P.T.B., Krainovic, P.M., Barata, L.E.S., 2012. Chemical characterization of rosewood (*Aniba rosaeodora* Ducke) leaf essential oil by comprehensive two-dimensional gas chromatography coupled with quadrupole mass spectrometry. Journal of Essential Oil Research 24, 245-251.

Finch, K., Espinoza, E., Jones, F.A., Cronn, R., 2017. Source identification of western Oregon Douglas-fir wood cores using mass spectrometry and random forest classification. Applications in Plant Sciences 5, 1600158.

Finkeldey, R., Ziehe, M., 2004. Genetic implications of silvicultural regimes. Forest Ecology and Management 197, 231-244.

Förstel, H., Boner, M., Höltken, A., Fladung, M., Degen, B., Zahnen, J., 2011. Fighting illegal logging through the introduction of a combination of the isotope method for identifying the origins of timber and DNA analysis for differentiation of tree species. WWF Germany, Berlin, Germany, p. 99.

Förstel, H., Hützen, H., 1983. Oxygen isotope ratios in German groundwater. Nature 304, 614–616.

Francis, J., Lowe, C., 2000. Bioecología de árboles nativos y exóticos de Puerto Rico y las Indias Occidentales. Reporte Técnico General IITF-15. In. USDA, Servicio Forestal, Departamento de Agricultura de los Estados Unidos, Instituto Internacional de Dasonomía Tropical, Río Piedras, Puerto Rico, p. 583.

Gandara, F., 2009. Diversidade genética de populações de cedro (*Cedrela fissilis* Vell. (Meliaceae)) no Centro-Sul do Brasil. In. Escuela Superior de Agricultura, Universidad de San Pablo, Brasil, Piracicaba, p. 87.

Gasson, P., 2011. How precise can wood identification be? Wood anatomy's role in support of the legal timber trade, especially cites. IAWA Journal 32, 137-154.

Gasson, P., Baas, P., Wheeler, E., 2011. Wood anatomy of CITES-listed tree species. IAWA Journal 32, 155-198.

Gentry, A.H., 1995. Diversity and floristic composition of neotropical dry forests. In: Bullock, S.H., Mooney, H.A. (Eds.), Seasonally dry tropical forests. Cambridge University Press, pp. 146-194.

Gillies, A., Navarro, C., Lowe, A., Newton, A.C., Hernandez, M., Wilson, J., Cornelius, J.P., 1999. Genetic diversity in Mesoamerican populations of mahogany (*Swietenia macrophylla*), assessed using RAPDs. Heredity 83, 722-732.

Gonfiantini, R., Roche, M.-A., Olivry, J.-C., Fontes, J.-C., Zuppi, G.M., 2002. The altitude effect on the isotopic composition of tropical rains. Chemical Geology 181, 147-167.

Gotelli, N., Ellison, A., 2004. A primer of ecological statistics. Sinauer Associates Inc. Sunderland, p. 614.

Goudet, J., 1995. FSTAT (version 1.2): a computer program to calculate F-statistics. Journal of Heredity 86, 485-486.

Gullison, R., Panfil, S., Strouse, J., Hubbell, S., 1996. Ecology and management of mahogany (*Swietenia macrophylla* King) in the Chimanes Forest, Beni, Bolivia. Botanical Journal of the Linnean society 122, 9-34.

Hedrick, P., 2011. Genetics of populations. Jones & Bartlett Learning, p. 675.

Helle, G., Schleser, G.H., 2004. Interpreting climate proxies from tree-rings. In: Fischer, H., Kumke, T., Lohmann, G., Flöser, G., Miller, H., von Storch, H., Negendank, J.F.W. (Eds.), The climate in historical times: towards a synthesis of Holocene proxy data and climate models. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 129-148.

Hernández, G., Buonamici, A., Walker, K., Vendramin, G.G., Navarro, C., Cavers, S., 2008. Isolation and characterization of microsatellite markers for *Cedrela odorata* L. (Meliaceae), a high value neotropical tree. Conservation Genetics 9, 457-459.

Hernández Sánchez, L.G., 2008. Genetic diversity and mating system analysis of *Cedrela odorata* L. (Meliaceae) populations under different human dominated landscapes and primary forests. In, Manejo y Conservación de Bosques Naturales y Biodiversidad, Escuela de Posgrado. Programa de Educación para el Desarrollo y la Conservación del Centro Agronómico Tropical de Investigación y Enseñanza, Costa Rica.

Hey, J., 2010. Isolation with migration models for more than two populations. Molecular Biology and Evolution 27, 905-920.

Hillis, W.E., 1987. Heartwood and tree exudates. Springer-Verlag, p. 268.

Hoare, A., 2015. Tackling illegal logging and the related trade. What progress and where next? In. Chatham House, London.

Holleley, C.E., Geerts, P.G., 2009. Multiplex Manager 1.0: a cross-platform computer program that plans and optimizes multiplex PCR. BioTechniques 46, 511-517.

Horacek, M., Jakusch, M., Krehan, H., 2009. Control of origin of larch wood: discrimination between European (Austrian) and Siberian origin by stable isotope analysis. Rapid Communications in Mass Spectrometry: RCM 23, 3688-3692.

Howard, F., Nakahara, S., Williams, D., 1995. Thysanoptera as apparent pollinators of West Indies mahogany, *Swietenia mahagoni* (Meliaceae). In, Annales des Sciences Forestières. EDP Sciences, pp. 283-286.

IPCC, 2014. Climate change 2014: synthesis report. Contribution of working groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. In. IPCC, Geneva, Switzerland, p. 112.

IUFRO, 2016. Illegal logging and related timber trade – dimensions, drivers, impacts and responses. A global scientific rapid response assessment report. In: Kleinschmit, D., Mansourian, S., Wildburger, C., Purret, A. (Eds.), IUFRO World Series, Vienna, p. 145.

Jan, C., Fumagalli, L., 2016. Polymorphic DNA microsatellite markers for forensic individual identification and parentage analyses of seven threatened species of parrots (family Psittacidae). PeerJ 4, e2416.

Jara, L.F., 1995. Mejoramiento forestal y conservación de recursos genéticos forestales (Tomo I). In. CATIE, Turrialba, Costa Rica, p. 174.

Joebstl, D., Bandoniene, D., Meisel, T., Chatzistathis, S., 2010. Identification of the geographical origin of pumpkin seed oil by the use of rare earth elements and discriminant analysis. Food Chemistry 123, 1303-1309.

Johann Heinrich von Thünen Institute, v., 2015. Development & implementation of a species identification and timber tracking system with DNA fingerprints and isotopes in Africa. In. International Tropical Timber Organization (ITTO), p. 232.

Johnson, A., Laestadius, L., 2011. New laws, new needs: the role of wood science in global policy efforts to reduce illegal logging and associated trade. IAWA Journal 32, 125-136.

Johnson, M.T., Agrawal, A.A., 2005. Plant genotype and environment interact to shape a diverse arthropod community on evening primrose (*Oenothera biennis*). Ecology 86, 874-885.

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.

Jucker, T., Bongalov, B., Burslem, D., Nilus, R., Dalponte, M., Lewis, S.L., Phillips, O.L., Qie, L., Coomes, D.A., 2018. Topography shapes the structure, composition and function of tropical forest landscapes. Ecology Letters 21, 989-1000.

Kagawa, A., Leavitt, S.W., 2010. Stable carbon isotopes of tree rings as a tool to pinpoint the geographic origin of timber. Journal of Wood Science 56, 175-183.

Kamiński, B., Jakubczyk, M., Szufel, P., 2018. A framework for sensitivity analysis of decision trees. Central European Journal of Operations Research 26, 135-159.

Kassambara, A., 2018. ggpubr: 'ggplot2' based publication ready plots. In. http://www.sthda.com/english/rpkgs/ggpubr.

Kelly, S., Heaton, K., Hoogewerff, J., 2005. Tracing the geographical origin of food: the application of multi-element and multi-isotope analysis. Trends in Food Science and Technology 16, 555-567.

Killeen, J., Garcia, E., Beck, S., 1993. Guía de arboles de Bolivia. Herbario Nacional de Bolivia, Missouri Botanical Garden. Quipus, SRL, La Paz, p. 958.

Kishor, N., Lescuyer, G., 2012. Controlling illegal logging in domestic and international markets by harnessing multi-level governance opportunities. International Journal of the Commons 6, 255-270.

Koehler, J.J., 2017. Forensics or fauxrensics: ascertaining accuracy in the forensic sciences. Arizona State Law Journal 49, 1369-1416.

Krebs, J.E., Goldstein, E.S., Kilpatrick, S.T., 2014. Lewin's Genes XI. Jones & Bartlett Learning, United States of America, p. 940.

Kurita, N., Ichiyanagi, K., Matsumoto, J., Yamanaka, M.D., Ohata, T., 2009. The relationship between the isotopic content of precipitation and the precipitation amount in tropical regions. Journal of Geochemical Exploration 102, 113-122.

Lago, J.H.G., de Ávila, P., de Aquino, E.M., Moreno, P.R.H., Ohara, M.T., Limberger, R.P., Apel, M.A., Henriques, A.T., 2004. Volatile oils from leaves and stem barks of *Cedrela fissilis* (Meliaceae): chemical composition and antibacterial activities. Flavour and Fragrance Journal 19, 448-451.

Lancaster, C., Espinoza, E., 2012. Analysis of select *Dalbergia* and trade timber using direct analysis in real time and time-of-flight mass spectrometry for CITES enforcement. Rapid Communications in Mass Spectrometry 26, 1147-1156.

Lawson, S., MacFaul, L., 2010. Illegal logging and related trade: indicators of the global response. Chatham House, London, UK, p. 132.

Lemes, M.R., Brondani Rp Fau - Grattapaglia, D., Grattapaglia, D., 2002. Multiplexed systems of microsatellite markers for genetic analysis of mahogany, *Swietenia macrophylla* King (Meliaceae), a threatened neotropical timber species. The Journal of Heredity 93, 287–291.

Lemes, M.R., Gribel, R., Proctor, J., Grattapaglia, D., 2003. Population genetic structure of mahogany (*Swietenia macrophylla* King, Meliaceae) across the Brazilian Amazon, based on variation at microsatellite loci: implications for conservation. Molecular Ecology 12, 2875-2883.

Lescuyer, G., Ndotit, S., Ndong, L.B.B., Tsanga, R., Cerutti, P.O., 2014. Policy options for improved integration of domestic timber markets under the voluntary partnership agreement (VPA) regime in Gabon. In. Center for International Forestry Research (CIFOR), Bogor, Indonesia, p. 4.

Ley Forestal 1700, 1996. Ley de 12 de julio de 1996 Honorable Congreso Nacional, Bolivia.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R news 2, 18-22.

Lowe, A., 2007. Can we use DNA to identify the geographic origin of tropical timber? In: Degen, B. (Ed.), Proceedings of the International Workshop "Fingerprinting methods for the identification of timber origins". Bonn Press. Sonderheft 321, Bonn, Germany, pp. 15-19.

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.

Maia, B.H.L.N.S., Paula, J.R.d., Sant'Ana, J., Silva, M.F.d.G.F.d., Fernandes, J.B., Vieira, P.C., Costa, M.d.S.S., Ohashi, O.S., Silva, J.N.M., 2000. Essential oils of *Toona* and *Cedrela* species (Meliaceae): taxonomic and ecological implications. Journal of the Brazilian Chemical Society 11, 629-639.

Malky, A., 2005. Diagnósticos sectoriales. Sector forestal en Bolivia. In. Unidad de Análisis de Políticas Sociales y Económicas (UDAPE), La Paz, Bolivia, p. 49.

McCarroll, D., Loader, N.J., 2004. Stable isotopes in tree rings. Quaternary Science Reviews 23, 771-801.

McClure, P.J., Chavarria, G.D., Espinoza, E., 2015. Metabolic chemotypes of CITES protected *Dalbergia* timbers from Africa, Madagascar, and Asia. Rapid Communications in Mass Spectrometry 29, 783-788.

McGough, N., 2010. Reflections of fingerprinting techniques. Genetic and isotopic fingerprinting methods – practical tools to verify the declared origin of wood. In, Documentation of the international conference Eschborn.

Medina, E., Mooney, H.A., Bullock, S.H. (Eds.), 1995. Seasonally dry tropical forests. Cambridge University Press, Cambridge.

Montgomery, J., Evans, J.A., Cooper, R.E., 2007. Resolving archaeological populations with Sr-isotope mixing models. Applied Geochemistry 22, 1502-1514.

Mosedale, J.R., Ford, A., 1996. Variation of the flavour and extractives of European oak wood from two French forests. Journal of the Science of Food and Agriculture 70, 273-287.

Mostacedo, B., Fredericksen, T., 2001. Regeneración y silvicultura de bosques tropicales en Bolivia. Editora El País, Santa Cruz, Bolivia, p. 221.

Mostacedo, B., Fredericksen, T.S., 1999. Regeneration status of important tropical forest tree species in Bolivia: assessment and recommendations. Forest Ecology and Management 124, 263-273.

Mostacedo, B., Justiniano, J., Toledo, M., Fredericksen, T., 2003. Guía dendrológica de especies forestales de Bolivia. Proyecto de Manejo Forestal Sostenible, Santa Cruz, Bolivia, p. 245.

Moya, R., Calvo-Alvarado, J., 2012. Variation of wood color parameters of *Tectona grandis* and its relationship with physical environmental factors. Annals of Forest Science 69, 947-959.

Moya, R., Wiemann, M.C., Olivares, C., 2013. Identification of endangered or threatened Costa Rican tree species by wood anatomy and fluorescence activity. Revista de Biología Tropical 61, 1113-1156.

Murphy, P.G., Lugo, A.E., 1986. Ecology of tropical dry forest. Annual Review of Ecology and Systematics 17, 67-88.

Musah, R.A., Espinoza, E.O., Cody, R.B., Lesiak, A.D., Christensen, E.D., Moore, H.E., Maleknia, S., Drijfhout, F.P., 2015. A high throughput ambient mass spectrometric approach to species identification and classification from chemical fingerprint signatures. Scientific Reports 5, 11520.

Navarro-Cerrillo, R.M., Agote, N., Pizarro, F., Ceacero, C.J., Palacios, G., 2013. Elements for a non-detriment finding of *Cedrela* spp. in Bolivia — A CITES implementation case study. Journal for Nature Conservation 21, 241-252.

Navarro, G., 2011. Clasificación de la vegetación de Bolivia. Centro de Ecología Difusión Simón I. Patiño, Santa Cruz, p. 713.

Nijmeijer, A., 2012. Physiological responses of a tropical tree to elevated CO_2: a century long evaluation of *Pseudolmedia laevis* trees using stable carbon isotope values from tree rings. In. Wageningen University, The Netherlands.

Noldt, G., Bauch, J., Koch, G., Schmitt, U., 2001. Fine roots of *Carapa guianensis* Aubl. and *Swietenia macrophylla* King: cell structure and adaptation to the dry season in Central Amazonia. Journal of Applied Botany 75, 152-158.

Novick, R.R., Dick, C.W., Lemes, M.R., Navarro, C., Caccone, A., Bermingham, E., 2003. Genetic structure of Mesoamerican populations of big-leaf mahogany (*Swietenia macrophylla*) inferred from microsatellite analysis. Molecular Ecology 12, 2885-2893.

Oliveira-Filho, A.T., Curi, N., Vilela, E.A., Carvalho, D.A., 1998. Effects of canopy gaps, topography, and soils on the distribution of woody species in a central Brazilian deciduous dry forest. Biotropica 30, 362-375.

Ometto, J.P., Flanagan, L.B., Martinelli, L.A., Moreira, M.Z., Higuchi, N., Ehleringer, J.R., 2002. Carbon isotope discrimination in forest and pasture ecosystems of the Amazon Basin, Brazil. Global Biogeochemical Cycles 16, 56-51-56-10.

Orvis, K.H., Grissino-Mayer, H.D., 2002. Standardizing the reporting of abrasive papers used to surface tree-ring samples. Tree-Ring Research 58, 47-50.

Oulhote, Y., Le Bot, B., Deguen, S., Glorennec, P., 2011. Using and interpreting isotope data for source identification. Trends in Analytical Chemistry 30, 302-312.

Paredes-Villanueva, K., López, L., Navarro-Cerrillo, R.M., 2016. Regional chronologies of *Cedrela fissilis* and *Cedrela angustifolia* in three forest types and their relation to climate. Trees - Structure and Function 30, 1581-1593.

Pastore, T.C.M., Braga, J.W.B., Coradin, V.T.R., Magalhães, W.L.E., Okino, E.Y.A., Camargos, J.A.A., de Muñiz, G.I.B., Bressan, O.A., Davrieux, F., 2011. Near infrared spectroscopy

(NIRS) as a potential tool for monitoring trade of similar woods: discrimination of true mahogany, cedar, andiroba, and curupixá. Holzforschung 65, 73-80.

Pettersen, R.C., 1984. The chemical composition of wood. American Chemical Society, Washington, D.C., pp. 57-126.

Poussart, P.F., Schrag, D.P., 2005. Seasonally resolved stable isotope chronologies from northern Thailand deciduous trees. Earth and Planetary Science Letters 235, 752-765.

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945-959.

Putz, F.E., Zuidema, P.A., Synnott, T., Peña-Claros, M., Pinard, M.A., Sheil, D., Vanclay, J.K., Sist, P., Gourlet-Fleury, S., Griscom, B., Palmer, J., Zagt, R., 2012. Sustaining conservation values in selectively logged tropical forests: the attained and the attainable. Conservation Letters 5, 296-303.

R Development Core Team, 2017. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. In, http://www.R-project.org/.

Rachmayanti, Y., Leinemann, L., Gailing, O., Finkeldey, R., 2006. Extraction, amplification and characterization of wood DNA from Dipterocarpaceae. Plant Molecular Biology Reporter 24, 45-55.

Rajora, O.P., Rahman, M.H., Buchert, G.P., Dancik, B.P., 2000. Microsatellite DNA analysis of genetic effects of harvesting in old-growth eastern white pine (*Pinus strobus*) in Ontario, Canada. Molecular Ecology 9, 339-348.

Raymond, M., Rousset, F., 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. Journal of Heredity 86, 248-249.

Reifsnyder, W.E., Furnival, G., Horowitz, J., 1971. Spatial and temporal distribution of solar radiation beneath forest canopies. Agricultural Meteorology 9, 21-37.

Richter, H., Dallwitz, M., 2000. Commercial timbers: descriptions, illustrations, identification, and information retrieval. In English, French, German, Portuguese, and Spanish. Version: 25th June 2009. In, http://delta-intkey.com.

Rossmann, A., 2001. Determination of stable isotope ratios in food analysis. Food Reviews International 17, 347-381.

Sandak, A., Sandak, J., Negri, M., 2011. Relationship between near-infrared (NIR) spectra and the geographical provenance of timber. Wood Science and Technology 45, 35-48.

Schollaen, K., Heinrich, I., Neuwirth, B., Krusic, P.J., D'Arrigo, R.D., Karyanto, O., Helle, G., 2013. Multiple tree-ring chronologies (ring width, $\delta^{13}C$ and $\delta^{18}O$) reveal dry and rainy season signals of rainfall in Indonesia. Quaternary Science Reviews 73, 170-181.

Schulman, E., 1956. Dendroclimatic changes in semiarid America. University of Arizona Press, Tucson, Arizona, p. 142.

Seidel, F., Fripp, E., Adams, A., Denty, I., 2012. Tracking sustainability: review of electronic and semi-electronic timber tracking technologies. In, Technical Series. International Tropical Timber Organization (ITTO), p. 60.

SENAMHI, 2018a. Servicio Nacional de Meteorología e Hidrología, Bolivia. In. http://www.senamhi.gob.bo/.

SENAMHI, 2018b. Sistema de Procesamiento de Datos Meteorológicos (SISMET). In. http://senamhi.gob.bo/index.php/sismet, Servicio Nacional de Meteorología e Hidrología, Bolivia.

Seneca Creek Associates, L., 2004. "Illegal" logging and global wood markets: the competitive impacts on the U.S. wood products industry. American Forest & Paper Association, p. 163.

Seymour, F., Busch, J., 2018. Why forests? Why now? New developments for the New Year. In. https://www.cgdev.org/blog/why-forests-why-now-new-developments-new-year.

Soldati, M.C., Fornes, L., Van Zonneveld, M., Thomas, E., Zelener, N., 2013. An assessment of the genetic diversity of *Cedrela balansae* C. DC. (Meliaceae) in Northwestern Argentina by means of combined use of SSR and AFLP molecular markers. Biochemical Systematics and Ecology 47, 45-55.

Soldati, M.C., Inza, M.V., Fornes, L., Zelener, N., 2014. Cross transferability of SSR markers to endangered *Cedrela* species that grow in Argentinean subtropical forests, as a valuable tool for population genetic studies. Biochemical Systematics and Ecology 53, 8-16.

Soliani, C., Vendramin, G.G., Gallo, L.A., Marchelli, P., 2016. Logging by selective extraction of best trees: does it change patterns of genetic diversity? The case of *Nothofagus pumilio*. Forest Ecology and Management 373, 81-92.

Stark, T., Pang Cheung, S., 2006. Sharing the blame: global consumption and China's role in ancient forest destruction. In. Greenpeace International and Greenpeace China, Beijing, China, p. 63.

Sternberg, L.O., Deniro, M.J., Johnson, H.B., 1984. Isotope ratios of cellulose from plants having different photosynthetic pathways. Plant Physiology 74, 557-561.

Sultan, S.E., 2000. Phenotypic plasticity for plant development, function and life history. Trends in Plant Science 5, 537-542.

Therneau, T., Atkinson, B., Ripley, B., 2018. rpart: recursive partitioning and regression trees. In. https://CRAN.R-project.org/package=rpart.

Tiessen, H., Chacon, P., Cuevas, E., 1994. Phosphorus and nitrogen status in soils and vegetation along a toposequence of dystrophic rainforests on the upper Rio Negro. Oecologia 99, 145-150.

Toledo, M., Chevallier, B., Villaroel, D., Mostacedo, B., 2008. Ecología y silvicultura de especies menos conocidas: cedro, *Cedrela* spp. Proyecto BOLFOR II/Instituto Boliviano de Investigación Forestal, Santa Cruz, Bolivia, p. 30.

Toledo, M., Poorter, L., Peña-Claros, M., Alarcón, A., Balcázar, J., Leaño, C., Licona, J.C., Llanque, O., Vroomans, V., Zuidema, P., 2011. Climate is a stronger driver of tree and forest growth rates than soil and disturbance. Journal of Ecology 99, 254-264.

Tosi, J.A., Jr, 1960. Zonas de vida natural en el Perú. Memoria explicativa sobre el mapa ecológico del Perú. In, Bol. Téc. 5. Instituto Interamericano de las Ciencias Agrícolas de la E.E.A., Lima, Perú, p. 271.

UNODC, 2016. United Nations Office on Drugs and Crime. Best practice guide for forensic timber identification. In. International Consortium on Combating Wildlife Crime, New York, USA, p. 214.

van der Sleen, J.P., 2014. Environmental and physiological drivers of tree growth : a pan-tropical study of stable isotopes in tree rings. In. Wageningen University, Wageningen, p. 174.

van der Sleen, P., Groenendijk, P., Vlam, M., Anten, N.P., Boom, A., Bongers, F., Pons, T.L., Terburg, G., Zuidema, P.A., 2015. No growth stimulation of tropical trees by 150 years of CO_2 fertilization but water-use efficiency increased. Nature Geoscience 8, 24-28.

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.

van der Sleen, P., Zuidema, P.A., Soliz-Gamboa, C.C., Helle, G., Pons, T.L., Anten, N.P.R., 2014. Understanding causes of tree growth response to gap formation: $\delta^{13}C$-values in tree rings reveal a predominant effect of light. Trees - Structure and Function 28, 439-448.

Versini, G., Monetti, A., Reniero, F., 1997. Wine–nutritional and therapeutic benefits. In: Watkins, T. (Ed.). American Chemical Society Washington, DC.

Vlam, M., de Groot, G.A., 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.

von Scheliha, S., Zahnen, J., 2011. Genetic and isotopic fingerprinting methods – practical tools to verify the declared origin of wood. In. Documentation of the international conference Eschborn, p. 28.

Vuille, M., Werner, M., 2005. Stable isotopes in precipitation recording South American summer monsoon and ENSO variability: observations and model results. Climate Dynamics 25, 401-413.

Wagenführ, R., 2007. Holzatlas. Fachbuchverlag Leipzig im Carl Hanser Verlag, Germany, p. 816.

West, J.B., Bowen, G.J., Cerling, T.E., Ehleringer, J.R., 2006. Stable isotopes as one of nature's ecological recorders. Trends in Ecology & Evolution 21, 408-414.

West, J.B., Bowen, G.J., Dawson, T.E., Tu, K.P., 2010. Isoscapes: understanding movement, pattern, and process on Earth through isotope mapping. Springer Netherlands, p. 487.

Wheeler, E.A., 2011. Inside Wood – a web resource for hardwood anatomy. IAWA Journal 32, 199-211.

White, G., Powell, W., 1997a. Cross-species amplification of SSR loci in the Meliaceae family. Molecular Ecology 6, 1195-1197.

White, G., Powell, W., 1997b. Isolation and characterization of microsatellite loci in *Swietenia humilis* (Meliaceae): an endangered tropical hardwood species. Molecular Ecology 6, 851-860.

Wickham, H., 2016. ggplot2: elegant graphics for data analysis. Springer-Verlag New York, p. 213.

Wickham, H., Francois, R., Henry, L., Müller, K., RStudio, 2017. dplyr: a grammar of data manipulation. A fast, consistent tool for working with data frame like objects, both in memory and out of memory. In, https://CRAN.R-project.org/package=dplyr.

Wieloch, T., Helle, G., Heinrich, I., Voigt, M., Schyma, P., 2011. A novel device for batch-wise isolation of α-cellulose from small-amount wholewood samples. Dendrochronologia 29, 115-117.

Wiemann, M.C., Espinoza, E.O., 2017. Species verification of *Dalbergia nigra* and *Dalbergia spruceana* samples in the wood collection at the Forest Products Laboratory. Research Paper FPL–RP–690. In. Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, p. 10.

Wilkins, A.P., Stamp, C.M., 1990. Relationship between wood colour, silvicultural treatment and rate of growth in *Eucalyptus grandis* Hill (Maiden). Wood Science and Technology 24, 297-304.

Wit, M., Van Dam, J., Cerutti, P.O., Lescuyer, G., Kerrett, R., McKeown, J.P., 2010. Chainsaw milling: supplier to local markets - a synthesis. In: (ETFRN), E.T.F.R.N. (Ed.), ETFRN News. ETFRN, Wageningen, The Netherlands, p. 16.

Worbes, M., 1999. Annual growth rings, rainfall-dependent growth and long-term growth patterns of tropical trees from the Caparo Forest Reserve in Venezuela. Journal of Ecology 87, 391-403.

Wright, S., 1951. The genetical structure of populations. Annals of Eugenics 15, 323-354.

Xia, S.-W., Chen, J., Schaefer, D., Goodale, U.M., 2016. Effect of topography and litterfall input on fine-scale patch consistency of soil chemical properties in a tropical rainforest. Plant and Soil 404, 385-398.

Zobel, B.J., van Buijtenen, J.P., 1989. Wood variation: its causes and control. Springer Berlin Heidelberg, Berlin, Heidelberg, p. 363.

To dwellers in a wood, almost every species of tree has its voice as well as its feature.

— Thomas Hardy, Under the Greenwood Tree

Summary

Summary

Illegal trade of tropical timber leads to economic and biodiversity losses worldwide. It has been estimated that 10% to 80% of the total timber trade is illegal and, in some countries, such as Papua New Guinea, Liberia, and the Amazon countries, this percentage could be as high as 80-90% of all logging operations. To date, the most common type of fraud concerns false declarations of species and geographic origin, as current legal procedures are generally based on certificates and documents which can be falsified. Most legislative measures focus on combating international illegal trade but a high proportion (70-90%) of illegal tropical timber is traded in domestic markets. With concerns about sustainable procurement of tropical timber, clearly there is a need for effective control strategies and forensics tools to independently verify the source of traded timber in compliance with international and national regulations and counteract the illegal harvesting and trade of tropical timber. To this end, chemical analysis (Direct Analysis in Real Time Time-of-Flight Mass Spectrometry), genetics (microsatellites), and stable isotopes ($\delta^{13}C$ and $\delta^{18}O$) were applied in this thesis as they allow tracking the origin of timber. In this study, I selected one of the major South American timber genera, *Cedrela*, that includes species which are listed in the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) to assess the potential for timber tracking based on chemical analysis, DNA techniques, and stable isotopes composition of wood.

The genus *Cedrela* (Meliaceae) delivers one of the most important Neotropical timbers (tropical cedar), but illegal logging of *Cedrela* has resulted in CITES-listing of several species in this genus. As a result, timber from these species can be traded internationally only if the appropriate permits have been obtained and presented for clearance at the port of entry or exit. The problem is that CITES-listed and non-listed *Cedrela* species are harvested and traded under the same name and are often confused due to wood-anatomical similarities. For authorities enforcing CITES, methods to differentiate *Cedrela* species are needed.

Bolivia harbours as many as six *Cedrela* species, in different climatic zones, from moist to dry tropical forests, and from low to high altitudes: *Cedrela angustifolia* Sessé & Moc. Ex

DC., *Cedrela balansae* C. DC., *Cedrela fissilis* Vell., *Cedrela montana* Moritz ex Turcz., *Cedrela odorata* L., and *Cedrela saltensis* M.A. Zapater & del Castillo. *Cedrela* species are highly valued locally and used in carpentry, fine furniture, doors, windows, joinery, musical instruments, carvings, coatings and plywood. However, *Cedrela* populations have declined considerably in recent years due to overexploitation and exportation of large quantities of timber obtained from these species. As a result, out of the six species, three are currently listed in Appendix III of CITES: *C. odorata*, *C. fissilis* and *C. angustifolia* (listed as *C. lilloi* C. DC.). Despite legal harvesting limitations, these species remain at high risk because of continued illegal logging and timber trade. The high incidence of illegal trade indicates that control systems have limited effectiveness and methods for independent verification of species and legal origin are needed.

To assess the potential of chemical properties in wood for species and geographical origin identification, I used two approaches: Direct Analysis in Real Time (DART) coupled with Time-of-Flight Mass Spectrometry (TOFMS) and stable isotopes ($\delta^{13}C$ and $\delta^{18}O$). First, in DART-TOFMS analysis (**Chapter 2**), the mass spectrometer identifies the chemical components by the differing mass to charge (*m/z*) of ions/compounds from specimens. Heartwood samples from six *Cedrela* species (the three CITES-listed species plus *C. balansae*, *C. montana*, and *C. saltensis*) were collected at 11 sites throughout Bolivia. The resulting chemical spectra detected by DART-TOFMS comprised 1062 compounds; their relative intensities were analysed using Principal Component Analyses (PCA), Kernel Discriminant Analysis (KDA), and Random Forest analyses to check discrimination potential among species and sites. Species were identified with a mean discrimination error of 15-19%, with substantial variation in discrimination accuracy among species. The lowest error was observed in *C. fissilis* (Mean = 4.4%). Site discrimination error was considerably higher: 43-54% for *C. fissilis* and 42-48% for *C. odorata*. These results provide good prospects to differentiate *C. fissilis* from other species, but not so for the other tested species. Our findings suggest that discrimination is highly species specific and has no potential to geographically discriminate at the studied scale and for the studied species. DART-TOFMS technique allowed the independent verification of claimed identity of certain *Cedrela* species in timber trade.

Second, stable isotopes ($\delta^{13}C$ and $\delta^{18}O$) were also analysed throughout the range of two intensively exploited tropical timbers: *C. fissilis* and *C. odorata* to evaluate potential as a provenancing tool (**Chapter 3)**. We studied differences in $\delta^{13}C$ and $\delta^{18}O$ stable isotope signatures between 9 forest sites (145 trees sampled) to cover species distributions within Bolivia. Our results suggest that it is possible to distinguish species based on isotopes, especially *C. fissilis* with high accuracy (93.8%). We also assessed the relationship between the stable isotopes composition and precipitation and altitude, but did not find significant correlations between annual data and these variables suggesting that they may not always be the limiting factor of wood composition in the sampled trees. However, 10-year-bulk data did show higher correlations suggesting that long-term and more stable data better represented site-specific isotopic composition. To explore the isotopic site differentiation, a KDA was used and resulted in a low discrimination success: 37.5% accuracy for *C. odorata* and 29.5% for *C. fissilis* sites. This suggests that discrimination of geographical origin was not possible due to low differentiation among sites because of species and site-specific isotopic imprints. Our findings suggest that other variables – besides rainfall and altitude – might be controlling isotopic signatures. This strengthened the importance of identifying the variables with most influence on isotopic variation before their use as proxies.

To assess the potential of genetic tools for identification of geographical origin, we used microsatellites (SSR). A set of eight SSR's was applied to study if the spatial distribution and genetic variation were sufficient for discrimination of timber at a national level (**Chapter 4**). Cambium and leaves were sampled from 81 *C. odorata* trees from 3 sites, at 268-501 km distance. Genetic differentiation was assessed by Bayesian clustering and PCA. The PCA showed three distinct genetic clusters but only one of them corresponded with one of the sampled sites. The KDA based on allele frequency including all alleles had a 33.7% mean classification error, with a lower error (8.2%) for the site which matched with one genetic cluster. A blind test on site-unique alleles led to a similar accuracy (30%). The occurrence of multiple genetic clusters within sites suggested that Bolivian *C. odorata* populations contain several parental lines, resulting in limited potential for forensic genetic tracing at a national level.

Finally, our results using three different methods suggest that site and wood composition are highly species-specific. To assess if combining tools that answer the same question will improve accuracy, I combined methods from samples with more data available: stable isotopes ($\delta^{13}C$ and $\delta^{18}O$) and DART-TOFMS (**Chapter 5**). After applying KDA to the species and sites that have measurements of stable isotopes and DART-TOFMS, accuracy improved only in some cases. For species identification, combination of DART-TOFMS and stable isotopes improved the accuracy in comparison to the isotopic method but not in comparison to DART-TOFMS. For site identification, applying stable isotopes alone showed a slightly better performance than DART-TOFMS to discriminate Bajo Paraguá site but not for the rest of the sites (Roboré and Yapacaní). However, by combining both methods, accuracy for geographic origin identification was improved in 1-19% for two sites (Bajo Paraguá and Roboré) but showed higher error for one site (Yapacaní) in comparison to using each method individually. Our findings suggest that a combination of methods to identify *Cedrela* species and sites could be favourable in some cases but detrimental in others.

In conclusion, some of the factors that define timber tracing accuracies are not only the selection of the tracing techniques but also the statistical analysis approach, sampling design and scale of analysis. By expanding the database to other countries and by sampling in between sites it will be possible to identify site of origin at finer scales and with a higher precision. For this purpose a standardization of sampling procedures will be necessary. I highlight the importance of involving local stakeholders: joint forces fight against illegal logging more effectively. In this thesis I explored and compared different methods in the context of a tropical timber producing country like Bolivia. The results will provide more understanding of the techniques and tools for decision making on how and when they could assist with identification of timber species and geographical origin effectively.

Resumen

El comercio ilegal de maderas tropicales conduce a pérdidas económicas y de biodiversidad en todo el mundo. Se ha estimado que entre 10% y el 80% del comercio total de la madera es ilegal y en algunos países como Papúa Nueva Guinea, Liberia y los países amazónicos, este porcentaje ha sido tan alto como 80-90% de todas las operaciones de aprovechamiento. Hasta la fecha, el tipo de fraude más común es el de las declaraciones falsas sobre las especies y el origen geográfico, ya que los procedimientos legales actuales se basan generalmente en certificados y documentos vulnerables a la falsificación. Gran parte de las medidas legislativas se enfocan en combatir el comercio ilegal a nivel internacional, sin embargo un alto porcentaje (70-90%) de la madera tropical ilegal es comercializada dentro de mercados nacionales. Para velar por una procedencia sostenible de las maderas tropicales es necesario implementar estrategias de control efectivas así como herramientas de análisis forense para verificar independientemente la fuente de la madera comercializada (conforme a las normas nacionales e internacionales) y así contrarrestar el aprovechamiento y el comercio ilegal de madera tropical. Para ello, en la presente tesis, se aplicaron análisis químicos (Análisis Directo en Tiempo Real, Espectrometría de Masas en Tiempo de Vuelo; DART por sus siglas en inglés), genéticos (microsatélites) e isótopos estables ($\delta^{13}C$ y $\delta^{18}O$) que permiten la identificación del origen de la madera. En el presente estudio, se eligió uno de los mayores géneros en Sud América, *Cedrela*, el cual abarca especies que fueron incluidas en la Convención sobre el Comercio Internacional de Especies Amenazadas de Fauna y Flora Silvestres (CITES), para evaluar el potencial de seguimiento basado en análisis químicos, técnicas de ADN e isótopos estables en la madera.

El género *Cedrela* (Meliaceae) presenta una de las maderas neotropicales más importantes (cedro), pero su aprovechamiento ilegal ha resultado en que se incluya en la CITES a varias especies de este género. Como resultado, la madera de estas especies puede ser comercializada internacionalmente solo si los permisos apropiados han sido obtenidos y presentados para el despacho en el puerto de entrada o salida. El problema radica en que, tanto especies listadas como no listadas en la CITES, son aprovechadas y comercializadas bajo el mismo nombre y a menudo se confunden debido a las similitudes

anatómicas de la madera. Para que las autoridades implementen CITES se necesitan métodos que permitan diferenciar las distintas especies de *Cedrela*.

Bolivia alberga hasta seis especies de *Cedrela* en diferentes zonas climáticas, desde bosques tropicales secos hasta bosques tropicales húmedos, y de altitudes bajas a elevadas: *Cedrela angustifolia* Sessé & Moc. Ex DC., *Cedrela balansae* C. DC., *Cedrela fissilis* Vell., *Cedrela montana* Moritz ex Turcz., *Cedrela odorata* L., y *Cedrela saltensis* M.A. Zapater & del Castillo. Las especies de *Cedrela* son muy valoradas localmente y son utilizadas en carpintería, mueblería fina, puertas, ventanas, ebanistería, instrumentos musicales, tallados, revestimientos y contrachapados. Sin embargo, las poblaciones de *Cedrela* han disminuido considerablemente en los últimos años debido a su aprovechamiento intensivo y a la exportación en grandes cantidades de madera obtenida de estas especies. Como resultado, de las seis especies, actualmente tres figuran en el Apéndice III de la CITES: *C. odorata*, *C. fissilis* y *C. angustifolia* (listada como *C. lilloi* C. DC.). A pesar de las limitaciones legales para su corta, esta especie permanece en alto riesgo debido a su continua tala y comercio ilegal. La alta incidencia de su comercio ilegal es un indicador de la limitada eficacia de los sistemas de control actuales y se necesitan métodos independientes para verificar la legalidad de la especie y origen en cuestión.

Para evaluar el potencial de las propiedades químicas de la madera, que ayudan a la identificación de la especie y origen geográfico, se emplearon dos enfoques: el análisis directo en tiempo real (DART) acoplado a espectrómetro en tiempo de vuelo (TOFMS) y el análisis de isótopos estables ($\delta^{13}C$ y $\delta^{18}O$). Primeramente, en los análisis DART-TOFMS (**Capítulo 2**), el espectrómetro de masas identifica los componentes químicos por las diferentes relaciones masa a carga (*m/z*) de iones/compuestos en especímenes. Muestras de duramen de seis especies de *Cedrela* (las tres especies CITES además de *C. balansae*, *C. montana*, y *C. saltensis*) fueron recolectadas de 11 sitios de Bolivia. Los espectros químicos resultantes detectados por DART-TOFMS comprendieron 1062 compuestos; sus intensidades relativas fueron estudiadas a través del análisis de componentes principales (PCA), análisis discriminante de Kernel (KDA) y Random Forests para evaluar el potencial de discriminación entre especies y sitios. Las especies

fueron identificadas con un error de discriminación promedio de 15-19%, con una variación sustancial de precisión entre especies. El error de discriminación más bajo se observó en *C. fissilis* (Media = 4.4%). Los errores de identificación de sitio fueron considerablemente más altos: 43-54% para *C. fissilis* y 42-48% para *C. odorata*. Estos resultados ofrecen buenas perspectivas para diferenciar *C. fissilis* de otras especies, pero no así para las otras especies estudiadas. Los resultados sugieren que la discriminación es altamente específica para cada especie y que no se cuenta con potencial para discriminar geográficamente a la escala y para las especies estudiadas. La técnica de DART-TOFMS permitió verificar de manera independiente la identidad declarada de ciertas especies de *Cedrela* en el comercio de la madera.

Además, se analizaron isótopos estables ($\delta^{13}C$ y $\delta^{18}O$) en dos especies que han sido aprovechadas intensivamente: *C. fissilis* y *C. odorata* para evaluar el potencial de identificación de procedencias (**Capítulo 3**). Se estudiaron las diferencias en las huellas isotópicas de $\delta^{13}C$ y $\delta^{18}O$ entre 9 sitios boscosos (145 árboles muestreados) de manera que cubrieran la distribución de las especies de estudio en Bolivia. Los resultados obtenidos sugieren que es posible distinguir especies en base a la composición isotópica, especialmente *C. fissilis* con alta precisión (93.8%). También se evaluó la relación entre la composición de isótopos estables, la precipitación y la altitud, pero no se encontraron correlaciones significativas entre los datos anuales y dichas variables, lo cual sugiere que estos no siempre pueden ser los factores limitantes de la composición de la madera en los árboles muestreados. Sin embargo, los datos de 10 años combinados mostraron mayor correlación, lo cual sugiere que los datos a largo plazo y más estables representan mejor la composición isotópica de sitio. Para explorar la diferenciación isotópica entre sitios se utilizó un KDA que dio como resultado una discriminación baja: 37.5% de precisión para los sitios de *C. odorata* y 29.5% para los sitios de *C. fissilis*. Esto sugiere que la discriminación de origen geográfico no fue posible debido a la baja diferenciación isotópica por ser ésta específica del sitio y especie. Nuestros hallazgos sugieren que otras variables – además de la lluvia y la altitud – podrían estar controlando las huellas isotópicas. Estos resultados destacan la importancia de identificar las variables con mayor influencia en la variación isotópica antes de su aplicación como *proxies*.

Para evaluar el potencial de las herramientas genéticas para identificación de origen geográfico utilizamos microsatélites (SSR). Se aplicó un set de ocho SSR para evaluar si la distribución espacial y la variación genética pueden ser suficientes para la discriminación de la madera a nivel nacional (**Capítulo 4**). Se muestrearon *cambium* y hojas de 81 árboles de *C. odorata* provenientes de 3 sitios a 268-501 km de distancia. La diferenciación genética fue evaluada por agrupamiento (*clustering*) Bayesiano y PCA. El PCA mostró tres grupos genéticos distintos pero sólo uno de ellos correspondió con uno de los sitios muestreados. El análisis KDA basado en la frecuencia de todos los alelos tuvo un error de clasificación media de 33.7%, con un error menor (8.2%) para el sitio que coincidió con un *cluster* genético. Una prueba ciega (*blind test*) basada en los alelos únicos de sitio resultó en una precisión similar (30%). La ocurrencia de múltiples *clusters* genéticos dentro de los sitios sugiere que las poblaciones de *C. odorata* de Bolivia contienen varias líneas parentales, dando como resultado un potencial limitado para el seguimiento genético forense a nivel nacional.

Finalmente, usando tres métodos diferentes, los resultados sugieren que el sitio y composición de la madera son altamente específicos de cada especie. Para evaluar si la combinación de herramientas que responden a la misma pregunta podría mejorar la precisión se combinaron los métodos en las muestras que tenían más datos disponibles: isótopos estables ($\delta^{13}C$ y $\delta^{18}O$) y DART-TOFMS (**Capítulo 5**). Después de aplicar análisis KDA a las especies y sitios que tienen datos de isótopos estables y DART-TOFMS, la precisión mejoró solo en algunos casos. Para la identificación de especies, la combinación de isótopos y DART-TOFMS mejoró la precisión en comparación al método isotópico aplicado individualmente pero no en comparación DART-TOFMS. Para la identificación de sitio, la aplicación de isótopos estables de manera separada mostró un desempeño ligeramente mejor que DART-TOFMS para discriminar Bajo Paraguá, pero no para el resto de los sitios (Roboré y Yapacaní). Sin embargo, mediante la combinación de ambos métodos, la precisión para identificar el origen geográfico mejoró en 1-19% para dos sitios (Bajo Paraguá y Roboré), pero mostró mayor error para un sitio (Yapacaní) en comparación con el uso de cada método individualmente. Los resultados sugieren que la

combinación de métodos para identificar sitios y especies de *Cedrela* podría ser favorable en algunos casos, pero perjudicial en otros.

En conclusión, los factores que definen la precisión para el seguimiento de madera no comprenden solamente la selección de técnica de seguimiento en sí, sino también el enfoque de análisis estadístico, el diseño de muestreo y la escala del análisis. Al ampliar la base de datos hacia otros países y al incluir muestras entre los distintos sitios de estudio será posible identificar el lugar de origen a escalas más finas y con mayor precisión. Para este propósito será necesario contar con la estandarización de los procedimientos de muestreo. Es necesario destacar la importancia de involucrar a los actores locales: las fuerzas unidas luchan contra la tala ilegal de árboles de manera más eficaz. En la presente tesis, se exploraron y compararon diferentes métodos en el contexto de un país productor de maderas tropicales como Bolivia. Durante la toma de decisiones, los resultados proporcionarán mayor comprensión de las técnicas y herramientas sobre cómo y cuándo podrían ayudar en la identificación de especies maderables y origen geográfico de forma efectiva.

Samenvatting

Illegale handel in tropisch hardhout leidt wereldwijd tot economisch verlies en tot een afname in biodiversiteit. Naar schatting 10 tot 80 procent van alle houthandel is illegaal en in sommige landen, zoals Papua Nieuw-Guinea, Liberia en diverse Zuid-Amerikaanse landen, stijgt dit percentage zelfs tot 80-90% van alle houtkap. De meest voorkomende vorm van fraude betreft vervalste aangiften voor soorten en de geografische oorsprong van het hout. Deze fraude is mogelijk doordat de huidige legale procedures vooral gebaseerd zijn op certificaten en documenten die vervalst kunnen worden. De meeste wetgevende maatregelen zijn gericht op het bestrijden van illegale houthandel op internationale markten, maar een groot deel van de illegale handel (70-90%) vindt plaats op binnenlandse markten. De bezorgdheid rond duurzame herkomstbepaling van tropisch hardhout vraagt om effectieve controlestrategieën en forensische technieken om onafhankelijk de oorsprong van verhandeld hout te verifiëren in overeenstemming met internationale en nationale regelgeving en zo de illegale houtkap en houthandel tegen te werken. Daarom werden in deze dissertatie de volgende methoden getest: chemische analyse (Direct Analysis in Real Time Time-of-Flight Mass Spectometry), genetica (microsatellieten) en stabiele isotopen ($\delta^{13}C$ en $\delta^{18}O$). In deze studie selecteerde ik één van de belangrijkste Zuid-Amerikaanse plantengenera, *Cedrela* - waarvan enkele soorten door de Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) erkend zijn – om te bepalen in welke mate chemische analyses, DNA technieken en de samenstelling van stabiele isotopen de oorsprong van het hout kunnen achterhalen.

Het genus *Cedrela* levert een van de belangrijkste neotropische houtsoorten (Spaanse ceder), maar illegale houtkap zorgde ervoor dat enkele soorten uit dit genus op de CITES-lijst staan. Hierdoor kan hout van deze soorten alleen internationaal verhandeld worden als de juiste vergunningen verkregen worden om vervolgens te laten controleren aan de haven van vertrek of aankomst. Het probleem is dat *Cedrela*-soorten die wel en niet op de CITES-lijst staan onder dezelfde naam geoogst en verhandeld worden. Vaak worden deze soorten ook verward doordat de anatomische kenmerken van het hout gelijkaardig zijn. Voor autoriteiten die de CITES-regels volgen, zijn dus methoden nodig om onderscheid te maken tussen de verschillende *Cedrela*-soorten.

In Bolivia groeien zes *Cedrela*-soorten in verschillende klimaatzones, van natte naar droge tropische bossen, en van laagland naar hoogland: *Cedrela angustifolia* Sessé & Moc. Ex DC., *Cedrela balansae* C. DC., *Cedrela fissilis* Vell., *Cedrela montana* Moritz ex Turcz., *Cedrela odorata* L., en *Cedrela saltensis* M.A. Zapater & del Castillo. Deze soorten worden lokaal zeer gewaardeerd en worden gebruikt in timmerwerk, meubels, ramen, schrijnwerk, muziekinstrumenten, snijwerk, bekledingen en meubelplaten. De laatste jaren zijn *Cedrela* population sterk afgenomen door overexploitatie en grootschalige export. Hierdoor staan drie van de zes soorten momenteel op de Appendix III van CITES: *C. odorata*, *C. fissilis* en *C. angustifolia* (onder de naam *C. liloi* C. DC.). Ondanks legale restricties staan deze soorten nog steeds onder druk door aanhoudende illegale houtkap en houthandel. De hoge frequentie aan illegale houthandel geeft aan dat de huidige controlesystemen een beperkte effectiviteit hebben en dat methoden voor onafhankelijke verificatie van soorten en oorsprong nodig zijn.

Om te bepalen hoe groot het potentieel van chemische eigenschappen in het hout is om de soort en oorsprong te achterhalen, testte ik twee methoden: Direct Analysis in Real Time (DART) met Time-of-Flight Mass Spectrometry (TOFMS) en stabiele isotopen ($\delta^{13}C$ en $\delta^{18}O$). Ten eerste, in de DART-TOFMS analyse (**Hoofdstuk 2**) identificeert de massa spectrometer chemische componenten op basis van de verschillende massa's ten opzichte van de lading (*m/z*) van de ionen en chemische verbindingen in het hout. Het kernhout van zes *Cedrela*-soorten (drie CITES-soorten plus *C. balansae*, *C. montana* en *C. saltensis*) werd verzameld in 11 locaties doorheen Bolivia. De resulterende chemische profielen van de DART-TOFMS analyses bevatten 1062 chemische verbindigen. De relatieve intensiteit van deze verbindingen werd geanalyseerd met een Principal Component Analysis (PCA), Kernel Discriminant Analysis (KDA) en een Random Forest analysis om het in te schatten of er onderscheid gemaakt kan worden tussen soorten en plaatsen van herkomst. De identificatie van soorten had een gemiddelde foutenmarge van 15-19% met aanzienlijke variatie in nauwkeurigheid afhankelijk van de soort. *C. fissilis* had de laagste foutenmarge (gemiddeld 4.4%). De foutenmarge voor de plaats van herkomst was opmerkelijk groter: 43-54% voor *C. fissilis* en 42-48% voor *C. odorata*. Deze resultaten bieden goede vooruitzichten om onderscheid te maken tussen *C. fissilis*

en andere soorten, maar niet tussen de overige soorten onderling. Dit suggereert dat het onderscheidsvermogen zeer soort-specifiek is en dat er weinig potentieel is om onderscheid te maken tussen verschillenden plaatsen van herkomst voor de geografische schaal en de soorten in deze studie.

Ten tweede, analyses van stabiele isotopen ($\delta^{13}C$ en $\delta^{18}O$) doorheen de verspreiding van twee intensief gebruikte houtsoorten – *Cedrela fissilis* en *C. odorata* – werden uitgevoerd om het potentieel van isotopen voor herkomstbepaling in te schatten (**Hoofdstuk 3**). Ik bestudeerde de verschillen in $\delta^{13}C$ en $\delta^{18}O$ tussen 9 locaties (145 bomen) die de soortenverspreiding binnen Bolivia dekken. De resultaten suggereren dat het mogelijk is om onderscheid te maken tussen soorten op basis van isotopen, vooral voor *C. fissilis* had een hoge nauwkeurigheid (93.8%). Daarnaast bepaalde ik de relatie tussen samenstelling van stabiele isotopen, neerslag en hoogteligging. Er waren geen significantie correlaties tussen jaarlijkse gegevens en deze variabelen wat suggereert dat ze niet altijd een de limiterende factor zijn in de houtsamenstelling van de bemonsterde bomen. De gegevens van monsters die 10 jaar beslaan vertoonden daarentegen hogere correlaties wat aangeeft dat stabielere gegevens die op lange termijn verzameld zijn een beter beeld geven van de locatie-specifieke isotopensamenstelling. Om het onderscheid tussen isotopen van verschillende locaties te vergelijken, werd een KDA uitgevoerd. Deze analyse resulteerde in een lage succespercentages: 37.5% nauwkeurigheid voor *C. odorata* en 29.5% voor *C. fissilis* locaties. Dit suggereert dat het onderscheid tussen geografische locaties niet mogelijk was door de weinige verschillen in soort- en locatie-specifieke isotopensamenstelling. Deze bevindingen geven aan dat andere variabelen – naast regenval en hoogteligging – de samenstelling van isotopen bepalen. Daarom is het belangrijk om variabelen te identificeren die het meeste invloed hebben op de variatie in isotopen vooraleer men deze variabelen als proxies gebruikt.

Om het potentieel van genetische methoden voor herkomstbepaling in te schatten gebruikte ik microsatellieten (SSR). Een set van acht microsatellieten werd ontwikkeld om te bestuderen of de ruimtelijke verdeling en genetische variatie voldoende waren om onderscheid te maken tussen houtsoorten op een nationaal niveau (**Hoofdstuk 4**). Ik

verzamelde cambium en bladeren van 81 *C. odorata* bomen van drie locaties, op afstanden van 268 tot 501 kilometer. Genetische differentiatie werd bepaald met Bayesian clustering en PCA. De PCA toonde drie apart genetische clusters waarvan slechts een overeen kwam met een van de verzamellocaties. De KDA, gebaseerd op allelfrequenties van alle allelen, resulteerde in een gemiddelde foutenmarge van 33.7%, met een kleinere foutenmarge (8.2%) voor de locatie die overeen kwam met de genetische cluster. Een blinde test voor locatie-specifieke allelen leidde tot een gelijkaardige nauwkeurigheid (30%). Het voorkomen van meerdere genetische clusters binnen locaties suggereert dat de Boliviaanse *C. odorata* populaties verschillende parentale afstammingslijnen bevatten, waardoor het potentieel voor herkomstbepaling op nationaal niveau beperkt is.

De resultaten van de drie verschillende methodes geven aan dat de samenstelling van de locatie en het hout zeer soort-specifiek zijn. Om in te schatten of het combineren de verschillende methoden leidt tot een betere nauwkeurigheid, voegde ik monsters met voldoende gegevens samen: stabiele isotopen ($\delta^{13}C$ en $\delta^{18}O$) en DART-TOFMS (**Hoofdstuk 5**). Het toepassen van een KDA op deze gegevens verbeterde de nauwkeurigheid maar in sommige gevallen. Voor de identificatie van soorten verbeterde de combinatie van methoden de nauwkeurigheid ten opzichte van isotopen, maar niet DART-TOFMS. Voor de identificatie van locaties was de nauwkeurigheid van isotopen lichtjes beter in vergelijking met DART-TOFMS om onderscheid te maken tussen Bajo Paraguá en de andere locaties, maar niet tussen de overige locaties onderling (Roboré en Yapacaní). De combinatie van beide methoden resulteerde in een betere herkomstbepaling – een toename van 1-19% - voor twee locaties (Bajo Paraguá en Roboré), maar vertoonde een grotere foutenmarge voor de andere locatie (Yapacaní) in vergelijking met de individuele methoden. Deze bevindingen suggereren dat voor de identificatie van *Cedrela* soort en locaties een combinatie van methoden voordelig is in sommige gevallen, maar ongunstig in andere gevallen.

In conclusie, de nauwkeurigheid voor herkomstbepaling van hout wordt niet alleen bepaald door de gebruikte technieken, maar ook door de statistische analyses, de manier

waarop de monsters verzameld worden (sampling design) en de geografische schaal. Door de database uit te breiden naar andere landen en door de verzamelen tussen diverse locaties zal het mogelijk zijn om de locatie van herkomst op fijnere schaal en met grotere precisie te bepalen. Hiervoor is een standardisatie van verzamelingsmethodes nodig. Ik belicht dan ook hoe belangrijk het is om lokale belanghebbenden te betrekken: gezamenlijke strijdkrachten kunnen illegale houtkap doeltreffender bestrijden. In deze dissertatie onderzocht en vergeleek ik verschillende methoden in de context van een land dat tropische hardhout produceert, namelijk Bolivia. Deze resultaten van dit onderzoek zullen meer inzicht bieden waar en wanneer men doeltreffend kan ingrijpen in de identificatie van houtsoorten en hun herkomst.

Acknowledgements

Acknowledgements

Tracing timber origin is a topic I had wanted to study since long ago. Many people have contributed to this research before and during my PhD. Ole Kim Hansen, Erik Dahl Kjær, Rafael M. Navarro Cerrillo and Guillermo Palacios were the first people to whom I shared my ideas and who provided very useful comments, suggestions and opportunities to improve my proposal. Thank you all for your support and encouragement.

Gracias a mi familia por todo su apoyo. De alguna manera cada uno de ustedes me ha ayudado en mi investigación ya sea en el campo o durante mis trabajos en casa. Papá, gracias por todo el apoyo y motivación. Parte de lo que soy se lo debo a usted. Gracias a mis hermanas: Ruth, Adiel y Wendy por siempre estar ahí para mí. A Efraín y Brayan por el excelente trabajo de campo al que me acompañaron. A mis tíos y primos que siempre estuvieron animándome y haciéndome pasar un buen momento. Gracias a todos por aguantar mis momentos de estrés, aburrimientos, ausencias y locuras. Ustedes fueron mi gran motivación y soporte durante todo este tiempo.

Thanks to my promotors Pieter Zuidema and Frans Bongers. Your interest and trust to start with my PhD project was a great encouragement to continue with science. Your guidance, suggestions and quick response were always useful to me. I also appreciate the moments where you took me to my limits to express and defend my ideas and opinions! Thank you so much for listening and being there.

During my visits to Wageningen I made very valuable connections and friends. Thanks to my colleagues at the Forest Ecology and Forest Management group: FEM group, thank you so much for all the nice coffee breaks, discussions, and caring for me especially during my last months in Wageningen. Thank you, Joke, for all your help in every stage and moment of my PhD. Lu, Madelon, Estela, Merel, Peter, Alejandra, Juan Ignacio, Meike, José, Federico, Linar and Marlene: how nice it was to chat and spend time with you! I always left with a taste for more. Thanks a lot for listening, for your advice and tips. Thanks for always taking care of me by knocking on my office door, sending me a message or calling to check if everything was okay with me. Thanks to you I remained sound and lucid!

Time spent and talks with the REG group were always encouraging and enjoyable. Helen, Joost, Yingying, Yanjie and Linda, thanks a lot! Yin, thanks a lot for your care, the talks and for always checking if everything was okay with me.

Thanks, Patricia Meijer and Petra Kloppenburg for all your support and assistance through all this time whether I was in Wageningen or abroad!

Thanks, Jan, Ivo and Arjen for all your patience and for sharing your knowledge during my work in the genetics lab. Without you all the learning process would not have been as enjoyable as it was.

Lennart and Claudius from PE&RC, knowing that I could always count on your help and advice was always encouraging and relieving through all this period. Thanks for all the talks, trainings and events we had together. The fact of knowing that you honestly cared for all PhDs made me feel like family.

Jente, you accompanied me through all the stages of my PhD: fieldwork, lab work, statistical analyses and writing. Your patience and support meant a lot to me. You made a more enjoyable PhD journey out of each moment we shared and I will be forever grateful for this. Looking forward to more experiences and projects together! I also thank the Ottenburghs family for their care and support during my PhD.

Daniel Griffith, Lyndsay Collins and Ariana Arispe, thanks so much for checking my writing. *Su apoyo y revisiones me tranquilizó y animó mucho.*

A mis asistentes de campo y compañeras de lucha: Alejandra y Raquel, muchísimas gracias por todo el aguante durante las largas y pesadas horas en el monte; por las sugerencias cuando dudé qué decisiones tomar, por la ayuda incondicional, palabras y canciones que hacían del trabajo más divertido. Con ustedes supe, de verdad, lo que es disfrutar de la naturaleza a pesar de las circunstancias. También gracias a Estephanie, Suri, Jorge y Gary por los momentos compartidos, el apoyo, las palabras de aliento, consuelo, chistes…

Acknowledgements

Gracias inmensas a todas las personas que estuvieron de alguna manera involucradas durante el trabajo de campo. Mis guías y materos: don Florentino y Pablo Sergio Alvez en Cobija; Alberto Guzmán y don Danny en la Reserva El Tigre; Saúl Terrazas, Marcelo Yacuirí y Marco Antonio Teo Yarita en Cururú; Juan Pablo Pereira en Guarayos, don José Soqueré y don Oscar Prado en Concepción; Francisco Somoza en San Ignacio; Don Casimiro en Macharetí; Miguel Cuadiay en Riberalta; Lidio Daniel Parabá y David Gutierrez en Roboré; Juan Pérez en Espejos. A quienes nos transportaron en, hacia y entre los sitios de muestreo: Don Osvaldo Ribera, Antonio Cuéllar, don Gregorio, Miguel Angel Choque, José Poma, Efraín Bascopé, Francisco Paredes, Pedro Paredes, Abel Cuéllar, Rafael Pereira, don Medardo, Roberto Jurado, Jaime Paucara, Mario Tapia, a quienes nos pusieron buena música, los que se ofrecieron ayudar a colectar muestras, a quienes corrieron el riesgo de cruzar ese "super charco" y quienes se quedaron hasta el anochecer en el monte con nosotros sin queja alguna. Gracias por esas inolvidables aventuras. Siempre los recordaré con mucho cariño y emoción.

A don Marquito, Marcos Mendieta, gracias infinitas por el aguante en los días de trabajo duro y la paciencia para conmigo. Su siempre constante buen humor y tacto hicieron mis horas en el campo más llevaderas y siempre las llevo y las llevaré en mi corazón.

A las empresas madereras La Chonta, CIMAL, INPA Parket, MABET, Reserva el Tigre; a las propiedades de Modesto Galarza en Macharetí, Don Sabino en San Martín de Agua Rica, don Tati y Oscar Kerdi en Cobija, Marioly Acuña, Moisés Acuña y Fredy Cabrera en La Guardia; don Walter Paz de Concepción, don Sandro en Bajo Paraguá, Jesús Ortíz en Vilamontes; Propiedad Fortaleza en Guarayos; a las comunidades y hogares que abrieron sus puertas y nos permitieron colectar las valiosas muestras usadas para este estudio. Gracias por su confianza y apoyo.

A la Universidad Autónoma Gabriel René Moreno (UAGRM), Centro Tecnológico Avanzado Forestal (CTAF) y el Laboratorio de Dendrocronología, Instituto Boliviano de Investigación Forestal (IBIF), Servicio Nacional de Meteorología e Hidrología (SENAMHI), Servicio Nacional de Áreas Protegidas (SERNAP), Dirección General de Biodiversidad y Áreas Protegidas, Ministerio de Medio Ambiente y Agua (MMAyA). Gracias por el apoyo logístico y

de laboratorio, provisión de datos y asesoramiento durante la colecta y procesamiento de mis muestras en Bolivia. Al Museo de Historia Natural Noel Kempff Mercado por el apoyo para la identificación y exportación de mis muestras. De manera especial a Alejandro Araujo Murakami por todo el asesoramiento con la identificación del cedro y por todos los tips compartidos. Also, thanks to Pieter Baas and Nicolien Sol for helping with the exportation of some of my samples.

A Roberto Quevedo Sopepi, por su apoyo y confianza constante desde que estudié en la carrera. Aprecio mucho el que usted se las haya jugado en darme oportunidades en el Dendrolab y que me haya apoyado siempre en cada uno de mis nuevos estudios y afanes. Muchísimas gracias!

A Soledad Montaño y Lennart Meyer-Sand por haberme apoyado en parte de mi trabajo de campo. Gracias por todo el interés en aprender y tan grata compañía.

Thanks to the Global Timber Tracking Network (GTTN) for providing me the opportunity to participate and exchange ideas during their meetings. Jo van Brusselen, Gesche Schifferdecker, Nele Schmitz: your trust, words and support meant and mean a lot to me. Thanks so much!

The people I met during the courses at Wageningen University were so meaningful and valuable to me. Especially Claudine Diedericks and Daan van Vliet, who shared tips, advice and listened patiently. Thank you guys for always being there ready to help!

Jagoda, Aya, Eva: my soul sisters, thanks for your visits and spending time with me. You made Wageningen a cozy home for me.

A todos los que forman parte de Kairos, por sus palabras de aliento, oraciones y sinceros abrazos. De manera especial a Yandi, Maribel, Noelia, Gary, Pablo, Lyndsay, Jorge y Vivi: muchas gracias por animarme, escucharme, acompañarme y darme ese abrazo que tanto necesité durante todo este tiempo.

A las personas que estuvieron ahí. Gracias por todo el aguante! Significó mucho para mí!

Short biography

Kathelyn Paredes Villanueva was born on February 4, 1985 in Santa Cruz, Bolivia. As a child, she was always passionate about nature and in 2002, she decided to pursue a BSc. Degree in Forestry Engineering at Universidad Autónoma Gabriel René Moreno in Santa Cruz, Bolivia. She combined her studies with volunteering work at the Museo de Historia Natural Noel Kempff Mercado and El Vallecito/BIOFAN biotechnology laboratory. In addition, she worked with plant anatomy and morphology as a teaching assistant at the Forest Engineering education program. After graduating, she worked as extensionist for the Forest Fire Prevention, Control and Rehabilitation Program in the Santa Cruz Prefecture to support training courses about fire prevention with local indigenous communities, teachers and students. For some years, she has also operated as a consultant assisting in the implementation of environmental impact assessments and the creation of thematic maps using Geographic Information Systems (GIS).

In 2011, she obtained her Msc degree in Water Resources Management at San Simón University in Cochabamba, Bolivia. Her thesis research explored on the relationship between gender and water resource management in the Rio Grande basin within the Comarapa Municipality in Santa Cruz, Bolivia. In 2012, she finished a second Msc degree in Sustainable Management of Natural Forests at Universidad Autónoma Gabriel René Moreno in Santa Cruz (Bolivia) in which she applied dendrochronology to tropical forest species. Her thesis focused on the dendrochronological potential of *Machaerium scleroxylon* Tul. (morado) and tree rings as indicators of age and climatic fluctuations in Santa Cruz, Bolivia. During this research project, she supported the implementation of a Laboratory of Dendrochronology as part of a cooperation project between Universidad Autónoma Gabriel René Moreno (Bolivia) and Universidad de Córdoba (Spain). This initiative was financed by the Spanish Agency for International Development Cooperation (AECID). Later, she decided to continue with dendrochronological analysis of other tropical timber species as part of her PhD program at Universidad de Córdoba

(Spain). Alongside all her academic activities she has always been involved in extra activities to learn state-of-the-art methods for tracing timber origin. The last couple of years, she could apply some of these methods during her PhD at the Forest Ecology and Forest Management group at Wageningen University. She will continue to search for the best and most applicable timber tracing methods within a Bolivian framework.

List of publications

Paredes-Villanueva, K., Espinoza, E., Ottenburghs, J., Sterken, M.G., Bongers, F., Zuidema, P.A. 2018. Chemical differentiation of Bolivian *Cedrela* species as a tool to trace illegal timber trade. Forestry: An International Journal of Forest Research 91(5): 603-613.

Meyer-Sand, B.R.V., Blanc-Jolivet, C., Mader, M., **Paredes-Villanueva, K.**, Tysklind, N., Sebbenn, A., Guichoux, E., Degen, B. 2017. 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(4): 779-781.

Paredes-Villanueva, K., López, L., Navarro-Cerrillo, R.M. 2016. Regional chronologies of *Cedrela fissilis* and *Cedrela angustifolia* in three forest types and their relation to climate. Trees, Structure and Function 30(5): 1581-1593.

Paredes-Villanueva, K., López, L., Brookhouse, M., Navarro-Cerrillo, R.M. 2015. Rainfall and temperature variability in Bolivia derived from the tree-ring width of *Amburana cearensis* (Fr. Allem.) A.C. Smith. Dendrochronologia 35: 80-86.

Paredes-Villanueva, K., Sánchez-Salguero, R., Manzanedo, R.D., Quevedo Sopepi, R., Palacios, G., Navarro-Cerrillo, R.M. 2013. Growth rate and climatic response of *Machaerium scleroxylon* in a dry tropical forest in southeastern Santa Cruz, Bolivia. Tree-Ring Research 69(2):63-79.

Paredes, K. 2011. *Gender and Water Resource Management: Analysis of gender participation in the management of water resources in the Rio Grande basin within the Comarapa Municipality, Santa Cruz*. Forests and Gender. Gland, Switzerland: IUCN and New York, NY: WEDO. 122.pp

Paredes, K., Morales, I., Magariños, E. 2007. Evaluación de medios de cultivo *In Vitro* para *Amburana cearensis* (roble) y *Centrolobium tomentosum* (tejeyeque) en la fase de establecimiento. Documento Científico 2-2008. Proyecto de Manejo de Bosques en Bolivia (FOMABO), Programa de Investigaciones Forestales (PROINFOR). Santa Cruz, Bolivia.

PE&RC training and Education Statement

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 of literature (5 ECTS)

- Isotopes ecology, forest genetics, provenance, timber DNA extraction, wood chemical composition, forensic timber analysis, Bolivian endangered forest species (2015)

Writing of project proposal (4.5 ECTS)

- Alternative methods to determine the origin of wood in tropical forests: a case of Bolivian forests (2015)

Post-graduate courses (3.6 ECTS)

- Introduction to R for statistical analysis; PE&RC, WUR (2015)
- Bayesian statistics; PE&RC, WUR (2015)
- Multivariate analysis; PE&RC, WUR (2015)
- Statistics in DendRochRonology 2.0; Universität Erlangen-Nürnberg, Germany (2016)

Laboratory training and working visits (2.7 ECTS)

- Laboratory training and chemical analysis on timber samples from 10 different sites in Bolivia; U.S. Fish and Wildlife Service Forensics Laboratory (2015)
- Laboratory training and isotopes analysis for the collected samples; University of Leicester, UK (2016)

Invited review of (unpublished) journal manuscript (1 ECTS)

- Trees, Structure and function: dendrochronology (2017)

Deficiency, refresh, brush-up courses (6 ECTS)

- Genetic analysis, trends and concepts; Department of Plant Sciences, WUR (2017)

Competence strengthening / skills courses (7.2 ECTS)

- Information literacy including EndNote introduction; WUR (2014)
- PhD Competence assessment; WUR (2015)
- Data management planning; WUR (2015)
- Techniques for writing and presenting a scientific paper; WUR (2015)
- Scientific publishing; WUR (2016)
- Brain training; WUR (2016)
- Project and time management; WUR (2016)
- Writing grant proposals; WUR (2017)
- Brain training follow up; WUR (2017)
- Reviewing a scientific paper; WUR (2018)

PE&RC Annual meetings, seminars and the PE&RC weekend (3.75 ECTS)

- PE&RC Weekend for PhD candidates in their first years (2014)
- Vegetation-soil interactions symposium (2014)
- PE&RC Day: optimization of science; pressure & pleasure (2014)
- PE&RC Day: one's waste…another's treasure? (2015)
- National PhD day: enhance your career!; Delft University of Technology (2015)
- PE&RC Midterm weekend (2016)
- PE&RC Last year's weekend (2017)
- National PhD day (2016)
- PhD Workshop carousel (2018)

Discussion groups / local seminars / other scientific meetings (5.6 ECTS)

- Journal Club, FEM;
- Visit to U.S. Fish and Wildlife Service Forensics Laboratory and presentation of results of DART analysis to the group (2015)
- Presentation to FEM group, thesis proposal, discussion group (2015)
- Presentation to FEM group, DART analysis results, discussion group (2015)
- Workshop metabolomics in chemical ecology; Institute of Ecology (NIOO-KNAW), Wageningen University, the Netherlands (2016)
- R User discussion group; PE&RC (2017)

- GTTN Workshop on combining timber tracking tools & on securing reliable species and origin identification, Wageningen, the Netherlands (2018)
- GTTN Africa regional workshop; Yaoundé, Cameroon (2018)

International symposia, workshops and conferences (4.4 ECTS)

- International workshop on tree rings, isotopes and recent climate change in the Amazon basin; Manaus, Amazonas, Brazil (2016)
- Association for tropical biology and conservation conference: linking natural history and the conservation of tomorrow's tropical ecosystems; Kuching, Malaysia (2018)

Lecturing / Supervision of practicals / tutorials (9 ECTS)

- Supervision of internship/research, Wessel Rijkens (2015-2016)
- Supervision of thesis, Soledad Montaño Andia (2016-2018)
- Laboratory and analysis training to forestry students at the dendrochronology laboratory; UAGRM (2016)

Cover illustrations:

Tree rings:
https://pixabay.com/en/texture-tree-annual-rings-2119287/

Magnifier:
https://www.cestovatelskyobchod.cz/legami-lupa-magnifying-glass-vintage-memories

DNA:
https://www.scenarii.fr/fr/actuality/index/adn-de-marque

^{18}O isotope:
https://es.123rf.com/photo_68350126_los-isótopos-de-ox%C3%ADgeno-estructura-de-los-átomos.html

^{13}C isotope:
http://www.breakdradio.com/key-flaw-found-in-radioisotope-isochron-dating/

Oxygen molecule:
https://es.123rf.com/photo_31529545_ilustración-de-la-molécula-de-ox%C3%ADgeno-en-el-fondo-blanco-aislado.html

Carbon molecule:
https://sp.depositphotos.com/115063772/stock-illustration-molecule-of-carbon.html

Truck:
https://www.alamy.es/foto-camiones-de-madera-de-la-amazonia-a-traves-de-la-cordillera-de-los-andes-en-ecuador-55906018.html

The research described in this thesis was financially supported by a NUFFIC grant, NFP-PhD.14/61, to Kathelyn Paredes Villanueva

Financial support from International Foundation for Science (IFS), The Rufford Foundation and Alberta Mennega Stichting for fieldwork and laboratory analyses is gratefully acknowledged.

Cover design:
Nicole Escalante

Photography:
Kathelyn Paredes Villanueva
On page 8, 54, 80, 108, 150 and 176 by Raquel Cárdenas Medrano
On page 175 by Anthony Regis

Printed by:
Digiforce || ProefschriftMaken
Ede, The Netherlands