



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

Kathelyn Paredes-Villanueva^{1,2} · G. Arjen de Groot^{1,3} · Ivo Laros^{1,3} · Jan Bovenschen^{1,3} · Frans Bongers¹ · Pieter A. Zuidema¹

Received: 15 June 2018 / Revised: 18 January 2019 / Accepted: 20 March 2019 / Published online: 23 April 2019
© The Author(s) 2019

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 seven microsatellites (SSRs), we studied the spatial distribution and genetic diversity 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 three 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

Communicated by C. Kulheim

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11295-019-1339-4>) contains supplementary material, which is available to authorized users.

✉ Kathelyn Paredes-Villanueva
kathypavi@gmail.com

- ¹ Forest Ecology and Forest Management Group, Wageningen University & Research, Droevendaalsesteeg 3, PO Box 47, 6700 AA Wageningen, The Netherlands
- ² Carrera de Ingeniería Forestal, Universidad Autónoma Gabriel René Moreno, Facultad de Ciencias Agrícolas, Km 9 al Norte, El Vallecito, Santa Cruz, Bolivia
- ³ Wageningen Environmental Research, Wageningen University & Research, Droevendaalsesteeg 3a, 6708 PB Wageningen, the Netherlands

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 DNA fingerprint 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; Paredes-Villanueva et al. 2018), 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 fingerprint methods, a large genetic diversity throughout the distribution of the timber species concerned is required. Previous studies suggest that the population genetic diversity 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 because the current system is vulnerable to manipulation.

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 one of 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). Our main objective was to test the applicability of microsatellites to differentiate between *Cedrela* timber from different geographical origin. Therefore, we investigated the spatial distribution and genetic diversity in three sites, each composed by three 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 nationwide scale? We expected to find a pattern of isolation by distance in which genetic differentiation increases with geographical distance, allowing us to differentiate between sites of origin (Wright 1943).

Methodology

Sampling

Samples of cambium and leaves of *C. odorata* were collected from 81 randomly selected trees which were distributed around three northern towns in Bolivia: 30 trees around Cobija, 29 around Riberalta, and 22 around Rurrenabaque (Fig. 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 analyze if there was genetic difference at smaller scales, samples were collected within each site from three subgroups each composed of up to ten trees. The subgroups 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 1).

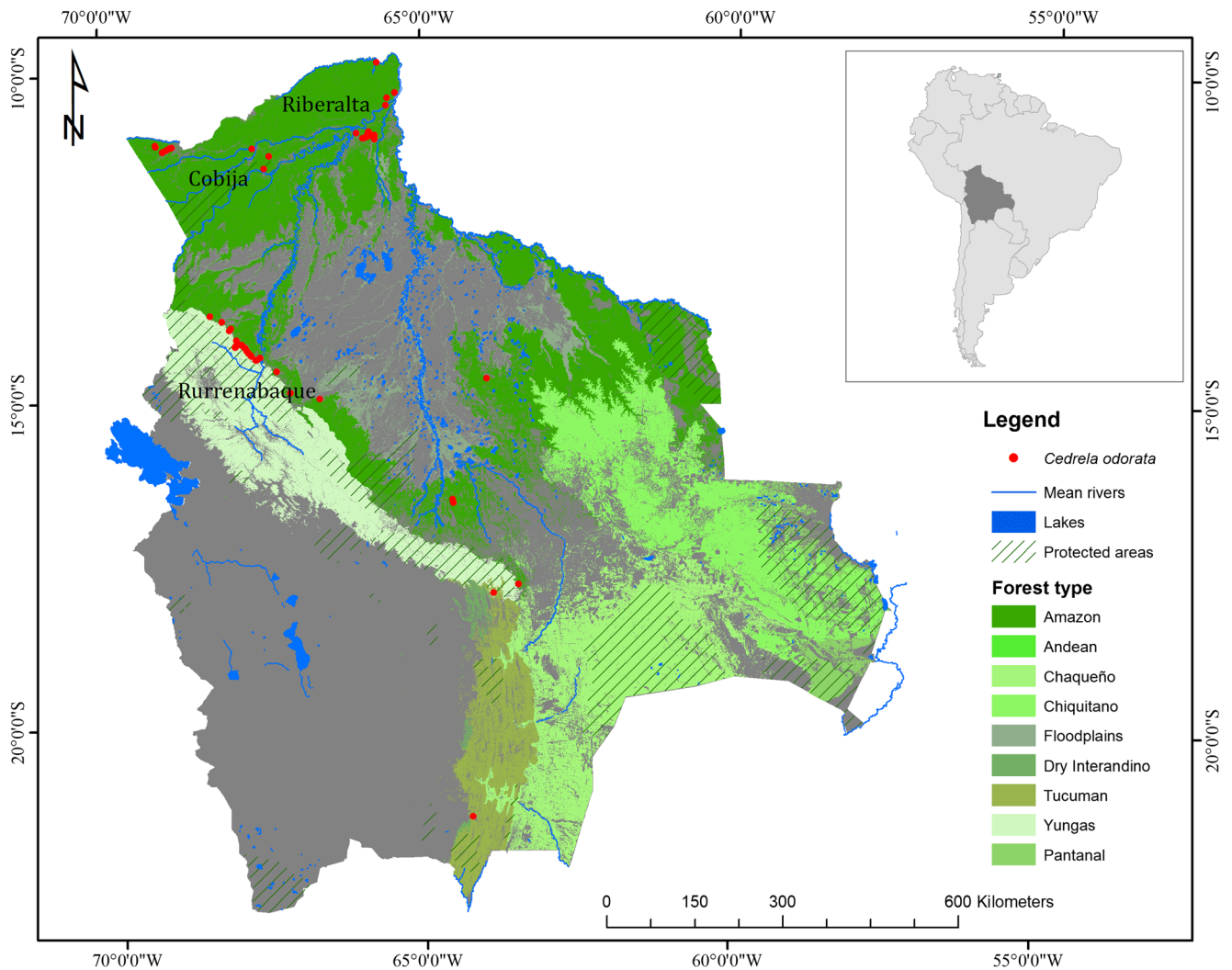


Fig. 1 Sampling sites (Cobija, Riberalta, and Rurrenabaque) based on *Cedrela odorata* distribution in Bolivia based on herbarium collections. Source: Museo de Historia Natural Noel Kempff Mercado, 2017. Forest cover: Autoridad de Bosques y Tierra, 2015

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 DNAs. 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 unit 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

Table 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

(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). Microsatellites or simple sequence repeats (SSRs, Tautz, 1989) have already been developed for *Swietenia humilis* (White and Powell 1997a, 1997b), *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) and have also been tested on *Cedrela balansae* and *Cedrela saltensis* (Soldati et al. 2013; Soldati et al. 2014). Therefore, we assessed eight primer pairs of SSRs developed for *Cedrela odorata* and *Cedrela fissilis* (Table 2). We considered patterns of high polymorphism loci for the selection of these SSRs. Up to now, microsatellite amplification was done in separate PCRs for each marker; we designed multiplex PCRs 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 was carried out in two PCR reactions.

Extracted DNA concentrations were 50 ng/μl in average (NanoDrop™ 2000/2000c spectrophotometer) and were diluted five 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 primer mix (see details in Table 2), 0.1 μl BSA (20 mg/ml), and 5 μl QIAGEN Multiplex PCR Master Mix. In this PCR reaction mixture, we combined the primers into two multiplexes for each

sample. Multiplex I targeted three loci (Ced95, Ced131, and Ced41), and Multiplex II targeted four 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), mean number of alleles per locus (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 analyze the genetic structure of *C. odorata* from three sites in Bolivia using seven microsatellites (Table 2). Four independent runs were performed for the complete sample set and each of the sites with different number of groups ($K = 1$ through 18), 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 frequency models. ΔK 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 one cluster based on the highest assignment percentage, e.g., an individual assigned for 50% to cluster 1 and 25% to clusters 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 dataset to explore genetic differentiation among sampling sites. The alleles were scored on a presence/absence binary dataset which was then analyzed in a logistic PCA 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 performance of the markers, tests of Hardy-Weinberg equilibrium (HWE) and linkage equilibrium (LE) were performed using the program GENEPOP v.4.2 (Raymond and Rousset 1995).

Table 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

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, both Wright's F_{ST} (Wright 1951) and Hedrick's G'_{ST} (Hedrick 2005) were calculated using GenAEx v.6.5 (Peakall and Smouse 2012).

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 dataset: 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 datasets. 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 only in a single or in two out of three study sites within the dataset composed of all the alleles. We performed a blind test, in which the genetic profiles of all samples were re-coded by the second author (GAG) and ten samples (i.e., profiles) were randomly drawn from the dataset. The remaining samples were then used by the first author (KPV) as a reference to infer the origin of the ten selected samples. Inference of origin was manually analyzed 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 dataset 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 three spatially separated sites in Bolivia, based on seven microsatellite loci per profile, together harboring a total of 143 different alleles. The mean number of alleles (A) per nSSR locus was 14.2 for all the three sites analyzed. The expected heterozygosity (H_e) varied from 0.83 to 0.89 for all the analyzed loci (Table 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), suggesting inbreeding or Wahlund effects due to spatial genetic structure within populations. This is in line with the results of the STRUCTURE analysis (Figs. 2 and 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 (Figs. 2 and 3, blue dots). In general, the observed genetic structure within sites did not match with the subgroups distinguished during sampling.

Genetic characterization and geographic structure of sites

The average pairwise genetic differentiation tests (F_{st} and G'_{st}) allowed us to assess the magnitude of the genetic differentiation among sites. Riberalta and Rurrenabaque sampling sites showed the highest differentiation ($F_{st} = 0.10$; $G'_{st} = 0.68$), followed by the Riberalta and Cobija ($F_{st} = 0.06$; $G'_{st} = 0.40$) sites. Cobija and Rurrenabaque showed the lowest genetic differentiation ($F_{st} = 0.04$; $G'_{st} = 0.34$). Differentiation values were significant ($P < 0.05$) for all population pairs. Consistently, the allelic richness (A) for Cobija was the highest among all sites (Table 3).

Structure analysis using seven 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 (Online Resource 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 model settings, similarity in the classification results between five iterations was 99.7%. Genetic cluster 1 (blue,

Table 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	0.83	
	Cobija	0.82	0.96	0.85	0.86	0.96	0.92	0.89	
	Rurrenabaque	0.83	0.93	0.92	0.84	0.94	0.85	0.88	
H_o	Riberalta	0.32	0.86	0.90	0.83	1.00	0.97	0.81	
	Cobija	0.55	0.93	0.60	0.54	0.93	0.86	0.75	
	Rurrenabaque	0.60	0.86	0.75	0.71	0.86	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
F_{is}	Total	10.29	20.05	14.10	13.24	19.35	17.02	12.15	15.17
	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	

Figs. 2 and 3) showed an almost homogenous genetic composition for all the samples in Riberalta site. Genetic clusters 2 and 3 covered both Rurrenabaque and Cobija (yellow and red, respectively, Figs. 2 and 3).

Comparable with the STRUCTURE results, the PCA showed three genetic clusters based on the allelic composition. It showed a clear separation of the genetic cluster 2 (yellow circle, Fig. 4) which mainly grouped samples from Rurrenabaque. The composition of the genetic cluster 3 (red circle, Fig. 4) was shared between Cobija and Rurrenabaque. Genetic cluster 1 (blue circle, Fig. 4) was shared between Riberalta and Cobija.

Genotype assignment

From the PCA analysis using all the alleles (first approach), two principal components were used as variable input for the 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 sets, 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 the highest classification error of 49.2%, followed by

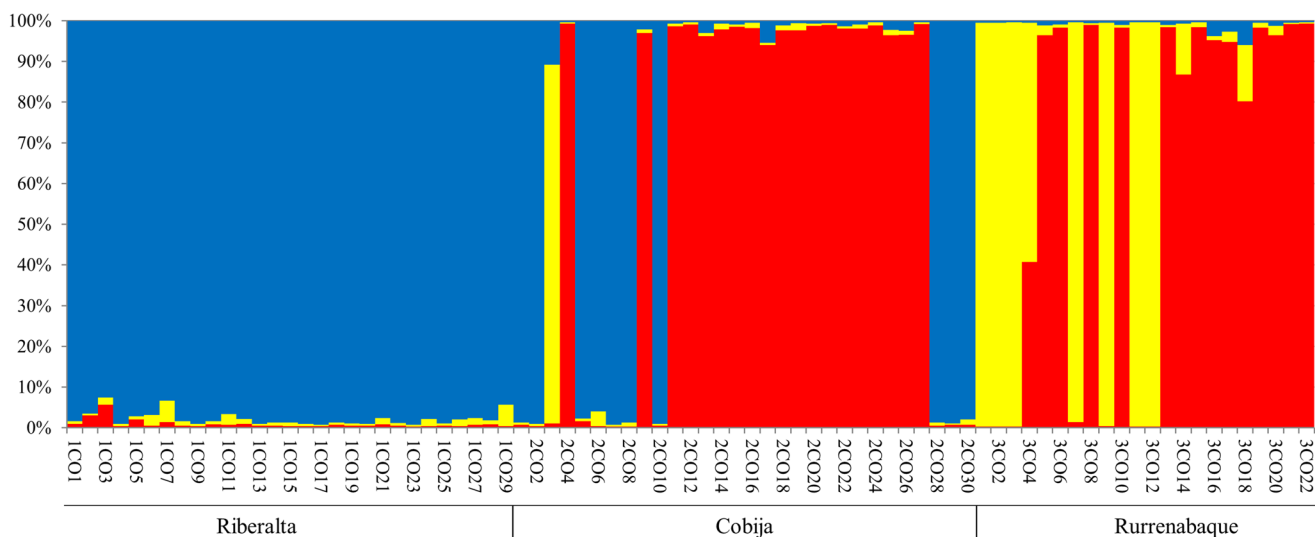


Fig. 2 Site genetic structure of 81 *Cedrela odorata* samples with admixture and at $K=3$ based on the analysis of allele frequency in STRUCTURE software. Names refer to the study sites and colours to genetic clusters: cluster 1 = blue, cluster 2 = yellow, and cluster 3 = red

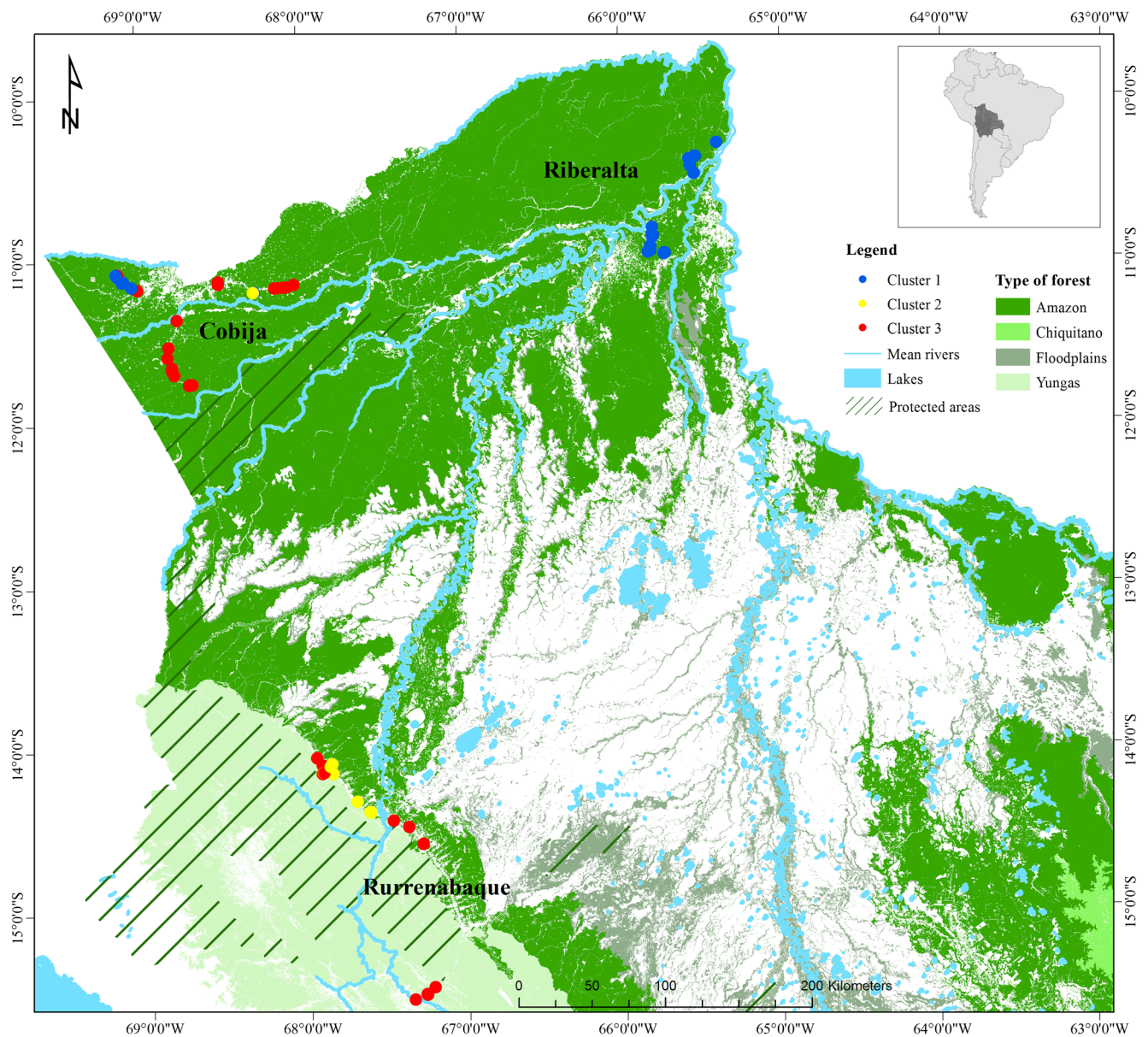


Fig. 3 Genetic characterization of 81 sampled *Cedrela odorata* trees at three sites in Bolivia. Colored filled dots refer to the dominant genetic clusters: cluster 1 = blue, cluster 2 = yellow, and cluster 3 = red. Samples

covered two forest types: Amazon and Yungas. Forest cover: Autoridad de Bosques y Tierra, 2015

Rurrenabaque with 43.7% and Riberalta with the lowest classification error of 8.2% (Fig. 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.

The *C. odorata* dataset showed the presence of unique alleles for each site (Online Resource 2). These unique alleles were used for a manual classification of a blind set of samples. The blind test (Online Resource 3) was performed using the unique alleles, and ten 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 (Online Resource 2). From the total, 70% of the samples were correctly assigned to the site of origin (Online Resource 4). 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 (Online Resource 4).

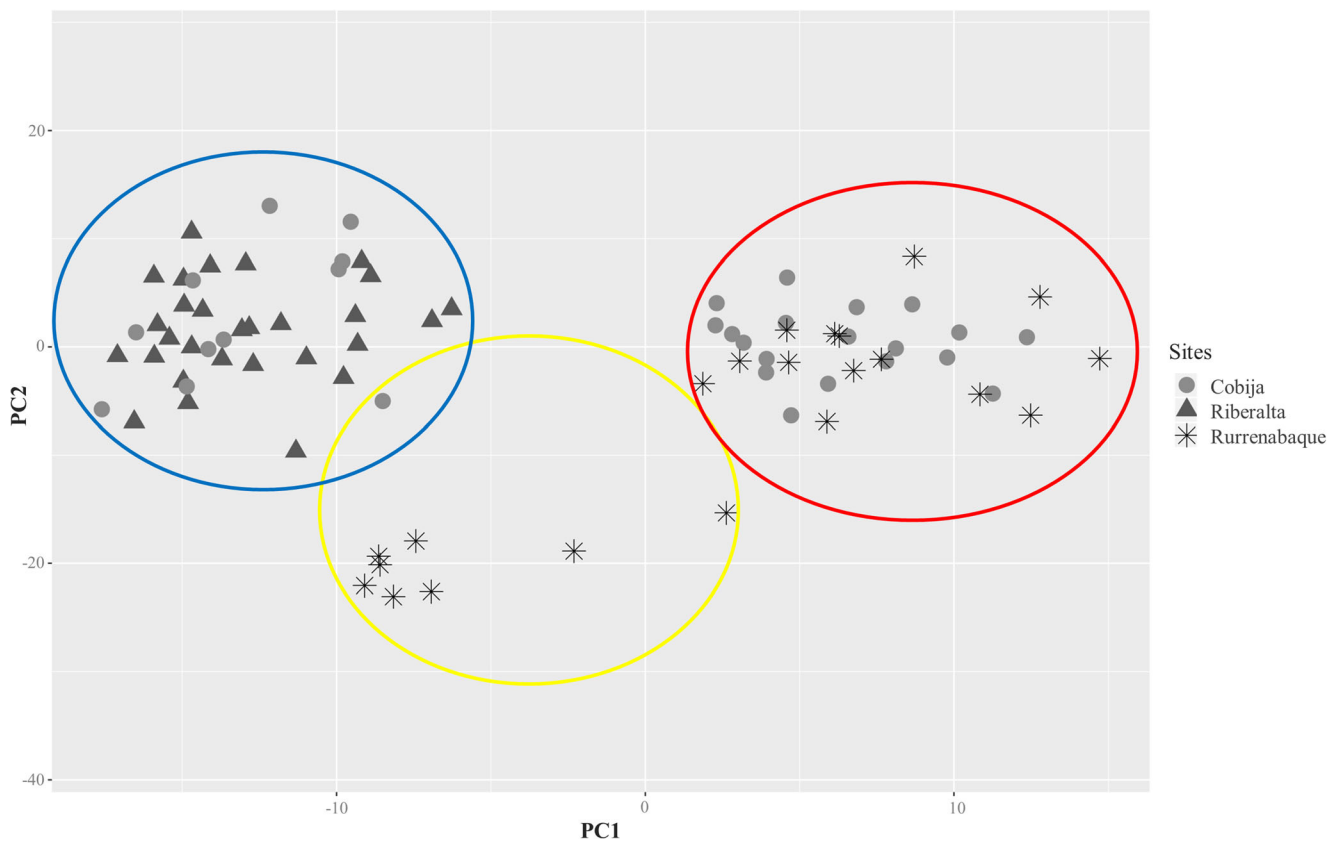


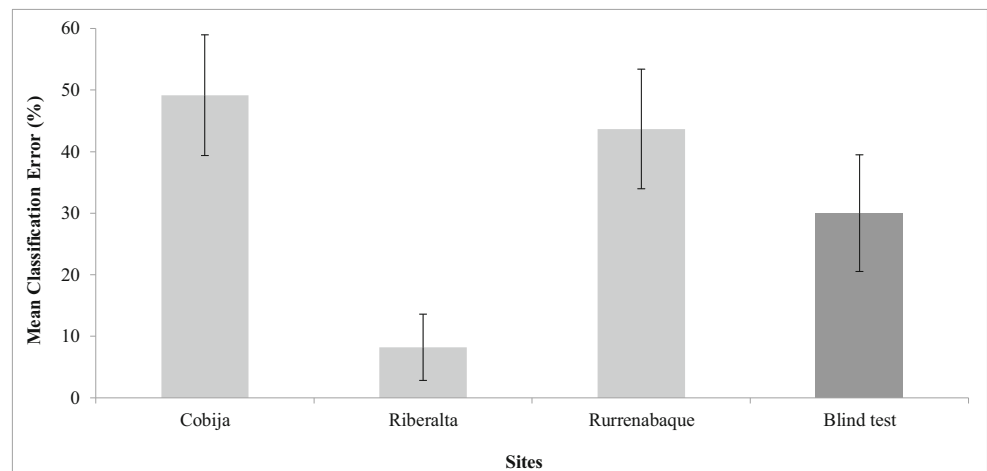
Fig. 4 Principal component analysis (PCA) showing three genetic clusters (colored circles) based on presence/absence of alleles in 81 *Cedrela odorata* trees at three geographic sites (symbols)

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 analyzed at large geographical scales: within and

Fig. 5 Classification errors of samples to three 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)



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 diversity 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 complete genetic differentiation among populations can also be due to regular gene flow from neighboring domesticated 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 diversity (Soliani et al. 2016). It has been found that reduction of tree population size due to logging may result in decrease of genetic diversity (Finkeldey and Ziehe 2004; Ratnam et al. 2014) and allelic frequencies (Cornuet and Luikart 1996; Rajora et al. 2000); hence, it is expected that protected areas still harbor 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 analyzed (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 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 neighboring 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, and (c) the dataset developed in this study 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 No. 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.

Data archiving statement The binary dataset of alleles is available at <https://doi.org/10.5061/dryad.hh3f7k3>.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- ABT (2017) Audiencia pública inicial y parcial de rendición de cuentas (de enero hasta agosto, gestión 2017) [PowerPoint slides]. Autoridad de Fiscalización y Control Social de Bosques y Tierra (ABT), Santa Cruz, Bolivia
- Bawa KS, Bullock SH, Perry DR, Coville RE, Grayum MH (1985) Reproductive biology of tropical lowland rain forest trees II Pollination systems. *Am J Bot* 72:346–356
- Bergo MCJ, Pastore TCM, Coradin VTR, Wiedenhoef AC, Braga JWB (2016) NIRS identification of *Swietenia macrophylla* is robust across specimens from 27 countries. *IAWA J* 37:420–430
- Braga JWB, Pastore TCM, Coradin VTR, Camargos JAA, da Silva AR (2011) The use of near infrared spectroscopy to identify solid wood specimens of (Cites Appendix II). *IAWA J* 32:285–296
- Cárdenas D, Arboleda NC, Tunjano SS, Barrera LQ, Rodríguez MB, Rodríguez SG, Rincón LM, Martín LER, Castañeda MR, Álvarez HA, Peinado ÁJV, Matajira JCC, Sánchez AG, Echeverry JCG, Arias AM, Gutierrez CA, Rivera LL, Velásquez MM, Pedraza LM, Villate GCM (2015) Planes de manejo para la conservación de abarco, caoba, cedro, palorosa y canelo de los andaquíes. Instituto Amazónico de Investigaciones Científicas. SINCHI, Colombia
- Cavers S, Degen B, Caron H, Lemes MR, Margis R, Salgueiro F, Lowe AJ (2005) Optimal sampling strategy for estimation of spatial genetic structure in tree populations. *Heredity* (Edinb) 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. *Mol Ecol* 12:1451–1460
- Cavers S, Navarro C, Lowe A (2004) Targeting genetic resource conservation in widespread species: a case study of *Cedrela odorata* L. *For Ecol Manag* 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. *J Biogeogr* 40:732–746
- Cerutti PO, 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
- Chase MW, Hills HH (1991) Silica gel: an ideal material for field preservation of leaf samples for DNA studies. *Taxon* 40:215–220
- Chaves CL, Degen B, Pakull B, Mader M, Honorio E, Ruas P, Tysklind N, Sebbenn AM (2018) Assessing the ability of chloroplast and nuclear DNA gene markers to verify the geographic origin of Jatoba (*Hymenaea courbaril* L.) timber. *J Hered* 109:543–552
- Compt J, Christy T (2008) The 14th meeting of the Conference of the Parties to CITES. *Traffic Bull* 21:101
- Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 144:2001–2014
- Degen B (2007) Proceedings of the international workshop “Fingerprinting methods for the identification of timber origins”. Johann Heinrich von Thünen-Institut
- 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. *Biol Conserv* 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 identification of timber origins”. Johann Heinrich von Thünen-Institut. p. 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 Sci Int: 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. *Proc R Soc Lond B Biol Sci* 269:1039–1046
- Duong T (2007) ks: kernel density estimation and kernel discriminant analysis for multivariate data in R. *J Stat Softw* 21:1–16
- Duong T (2017) ks: kernel smoothing. <https://cran.r-project.org/web/packages/ks/>. <https://cran.r-project.org/web/packages/ks/>
- Earl DA, vonHoldt BM (2012) Structure harvester: a website and program for visualizing structure output and implementing the Evanno method. *Conserv Genet Resour* 4:359–361
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164:1567–1587
- Fidelis CHV, Augusto F, Sampaio PTB, Krainovic PM, Barata LES (2012) Chemical characterization of rosewood (*Aniba rosaeodora* Ducke) leaf essential oil by comprehensive two-dimensional gas chromatography coupled with quadrupole mass spectrometry. *J Essent Oil Res* 24:245–251
- Finkeldey R, Ziehe M (2004) Genetic implications of silvicultural regimes. *For Ecol Manag* 197:231–244
- Gandara F (2009) Diversidade genética de populações de cedro (*Cedrela fissilis* Vell. (Meliaceae)) no Centro-Sul do Brasil. Escola Superior de Agricultura, Universidad de San Pablo, Brasil
- Gillies A, Navarro C, Lowe A, Newton AC, Hernandez M, Wilson J, Cornelius JP (1999) Genetic diversity in Mesoamerican populations of mahogany (*Swietenia macrophylla*), assessed using RAPDs. *Heredity* (Edinb) 83:722–732
- Goudet J (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. *J Hered* 86:485–486
- Hedrick P (2011) Genetics of populations. Jones & Bartlett Learning
- Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution* 59:1633–1638
- Hernández G, Buonamici A, Walker K, Vendramin GG, Navarro C, Cavers S (2008) Isolation and characterization of microsatellite markers for *Cedrela odorata* L. (Meliaceae), a high value neotropical tree. *Conserv Genet* 9:457–459
- Hernández Sánchez LG (2008) Genetic diversity and mating system analysis of *Cedrela odorata* L. (Meliaceae) populations under different human dominated landscapes and primary forests. Programa de Educación para el Desarrollo y la Conservación del Centro Agronómico Tropical de Investigación y Enseñanza

- Holleley CE, Geerts PG (2009) Multiplex Manager 1.0: a cross-platform computer program that plans and optimizes multiplex PCR. *BioTechniques* 46:511–517
- Howard F, Nakahara S, Williams D (1995) Thysanoptera as apparent pollinators of West Indies mahogany, *Swietenia mahagoni* (Meliaceae). In: *Annales des Sciences Forestières*, pp. 283–286. EDP Sciences
- 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 Sci Int: Genetics* 6:487–493
- Kagawa A, Leavitt SW (2010) Stable carbon isotopes of tree rings as a tool to pinpoint the geographic origin of timber. *J Wood Sci* 56:175–183
- Kishor N, Lescuyer G (2012) Controlling illegal logging in domestic and international markets by harnessing multi-level governance opportunities. *Int J Commons* 6:255–270
- Lawson S, MacFaul L (2010) *Illegal logging and related trade: indicators of the global response*. Chatham House, London
- Lemes MR, Brondani RP, Grattapaglia D (2002) Multiplexed systems of microsatellite markers for genetic analysis of mahogany, *Swietenia macrophylla* King (Meliaceae), a threatened neotropical timber species. *J Heredity* 93:287–291
- Lemes MR, 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. *Mol Ecol* 12:2875–2883
- Lescuyer G, Ndotit S, Ndong LBB, Tsanga R, Cerutti PO (2014) Policy options for improved integration of domestic timber markets under the voluntary partnership agreement (VPA) regime in Gabon. Center for International Forestry Research (CIFOR), Bogor, Indonesia, p 4
- Lowe A (2007) Can we use DNA to identify the geographic origin of tropical timber? In: *Proceedings of the International Workshop “Fingerprinting methods for the identification of timber origins” Bonn Press* (ed. Degen B), pp. 15–19. Sonderheft 321, Bonn, Germany
- Lowe AJ, Dormontt EE, Bowie MJ, 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
- Mostacedo B, Fredericksen TS (1999) Regeneration status of important tropical forest tree species in Bolivia: assessment and recommendations. *For Ecol Manag* 124:263–273
- Mostacedo B, Justiniano J, Toledo M, Fredericksen T (2003) *Guía dendrológica de especies forestales de Bolivia*. Segunda Edición edn, Proyecto de Manejo Forestal Sostenible, Santa Cruz
- Navarro-Cerrillo RM, Agote N, Pizarro F, Ceacero CJ, Palacios G (2013) Elements for a non-detriment finding of *Cedrela* spp. in Bolivia—a CITES implementation case study. *J Nat Conserv* 21:241–252
- Navarro G (2011) *Clasificación de la vegetación de Bolivia*. Centro de Ecología Difusión Simón I. Patiño, Santa Cruz
- Novick RR, Dick CW, Lemes MR, Navarro C, Caccone A, Birmingham E (2003) Genetic structure of Mesoamerican populations of big-leaf mahogany (*Swietenia macrophylla*) inferred from microsatellite analysis. *Mol Ecol* 12:2885–2893
- Paredes-Villanueva K, López L, Navarro-Cerrillo RM (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
- Paredes-Villanueva K, Espinoza E, Ottenburghs J, Sterken MG, Bongers F, Zuidema PA (2018) Chemical differentiation of Bolivian *Cedrela* species as a tool to trace illegal timber trade. *Forestry: An International Journal of Forest Research* 91: 603–613. <https://doi.org/10.1093/forestry/cpy019>
- Peakall R, Smouse P (2012) GenAIEx 6.5: genetic analysis in excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28:2537–2539
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- R Development Core Team (2017) R: a language and environment for statistical computing. R foundation for statistical computing, Vienna
- Rachmayanti Y, Leinemann L, Gailing O, Finkeldey R (2006) Extraction, amplification and characterization of wood DNA from Dipterocarpaceae. *Plant Mol Biol Report* 24:45–55
- Rajora OP, Rahman MH, Buchert GP, Dancik BP (2000) Microsatellite DNA analysis of genetic effects of harvesting in old-growth eastern white pine (*Pinus strobus*) in Ontario, Canada. *Mol Ecol* 9:339–348
- Ratnam W, Rajora OP, Finkeldey R, Aravanopoulos F, Bouvet J-M, Vaillancourt RE, Kanashiro M, Fady B, Tomita M, Vinson C (2014) Genetic effects of forest management practices: global synthesis and perspectives. *For Ecol Manag* 333:52–65
- Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J Hered* 86: 248–249
- 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
- Soldati MC, 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. *Biochem Syst Ecol* 47:45–55
- Soldati MC, Inza MV, 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. *Biochem Syst Ecol* 53:8–16
- Soliani C, Vendramin GG, Gallo LA, Marchelli P (2016) Logging by selective extraction of best trees: does it change patterns of genetic diversity? The case of *Nothofagus pumilio*. *For Ecol Manag* 373: 81–92
- Stark T, Pang Cheung S (2006) Sharing the blame: global consumption and China’s role in ancient forest destruction. Greenpeace International and Greenpeace China, Beijing, China, p 63
- Tautz D. (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.* 17(16):6463–6471.
- 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
- UNODC (2016) United Nations Office on Drugs and Crime. Best practice guide for forensic timber identification. p. 214. International Consortium on Combating Wildlife Crime, New York, USA
- Vlam M, de Groot GA, Boom A, Copini P, Laros I, Veldhuijzen K, Zakamdi D, Zuidema PA (2018) Developing forensic tools for an African timber: regional origin is revealed by genetic characteristics, but not by isotopic signature. *Biol Conserv* 220:262–271
- White G, Powell W (1997a) Cross-species amplification of SSR loci in the Meliaceae family. *Mol Ecol* 6:1195–1197
- White G, Powell W (1997b) Isolation and characterization of microsatellite loci in *Swietenia humilis* (Meliaceae): an endangered tropical hardwood species. *Mol Ecol* 6:851–860
- Wit M, Van Dam J, Cerutti PO, Lescuyer G, Kerrett R, McKeown JP (2010) Chainsaw milling: supplier to local markets—a synthesis. In: *ETFRN News* (ed. (ETFRN) ETFRN), p. 16. ETFRN, Wageningen, the Netherlands
- Wright S (1943) Isolation by distance. *Genetics* 28:114–138
- Wright S (1951) The genetical structure of populations. *Ann Eugenics* 15: 323–354

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