



**HAL**  
open science

## Timber tracking of *Jacaranda copaia* from the Amazon Forest using DNA fingerprinting

Lorena F. M. Capo, Bernd Degen, Celine Blanc-Jolivet, Niklas Tysklind, Stephen Cavers, Malte Mader, Barbara R.V. Meyer-Sand, Kathelyn Paredes-Villanueva, Euridice N. Honorio Coronado, Carmen García-Dávila, et al.

### ► To cite this version:

Lorena F. M. Capo, Bernd Degen, Celine Blanc-Jolivet, Niklas Tysklind, Stephen Cavers, et al.. Timber tracking of *Jacaranda copaia* from the Amazon Forest using DNA fingerprinting. 2024. hal-04664901

**HAL Id: hal-04664901**

**<https://hal.inrae.fr/hal-04664901v1>**

Preprint submitted on 30 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

1 Article

## 2 Timber tracking of *Jacaranda copaia* from the Ama- 3 zon Forest using DNA fingerprinting

4 Lorena F.M. Capo <sup>1,\*</sup>, Bernd Degen <sup>2</sup>, Celine Blanc-Jolivet <sup>2</sup>, Niklas Tysklind <sup>3</sup>, Stephen Cavers <sup>4</sup>,  
5 Malte Mader <sup>2</sup>, Barbara R.V. Meyer-Sand <sup>1</sup>, Kathelyn Paredes-Villanueva <sup>5</sup>, Eurídice N. Honorio-  
6 Conrado <sup>6</sup>, Carmen R. García-Dávila <sup>7</sup>, Valerie Troispoux <sup>8</sup>, Adline Delcamp <sup>8</sup>, and Alexandre M.  
7 Sebbenn <sup>1,9,\*</sup>

8 <sup>1</sup> Departamento de Fitotecnia, Faculdade de Engenharia de Ilha Solteira, Universidade Estadual  
9 Paulista, Ilha Solteira, SP, Brazil

10 <sup>2</sup> Thünen Institute of Forest Genetics, Sieker Landstrasse 2, 22927 Grosshansdorf, Ger-  
11 many

12 <sup>3</sup> INRAE, UMR0745 EcoFoG, AgroParisTech, Cirad, CNRS, Université des Antilles,  
13 Université de Guyane, Campus Agronomique, Avenue de France, BP97387 Kourou  
14 Cedex, France

15 <sup>4</sup> Centre for Ecology and Hydrology Edinburgh, Penicuik, UK

16 <sup>5</sup> Carrera de Ingeniería Forestal, Laboratorio de Dendrocronología, Facultad de Cien-  
17 cias Agrícolas, Universidad Autónoma Gabriel René Moreno, Km 9 carretera al  
18 Norte, El Vallecito, Santa Cruz, Bolivia

19 <sup>6</sup> Royal Botanic Gardens, Kew, Richmond, London, TW9 3AE, UK

20 <sup>7</sup> Instituto de Investigaciones de la Amazonía Peruana (IIAP), Av. José A. Quiñones  
21 km 2.5, Iquitos, Peru

22 <sup>8</sup> Plateforme Génome Transcriptome de Bordeaux, INRA Site de Pierroton Bâtiment  
23 Artiga, 69 route d'Arcachon, 33610 Cestas, France

24 <sup>9</sup> Instituto Florestal de São Paulo, CP 1322, São Paulo, SP 01059-970, Brazil

25 \* Correspondence: Alexandre M. Sebbenn, Instituto Florestal de São Paulo, CP  
26 1322, São Paulo, SP 01059-970, Brazil. E-mail: alexandresebbenn@yahoo.com.br

27 **Citation:** To be added by editorial  
28 staff during production.

29 Academic Editor: Firstname Last-  
30 name

31 Received: date

32 Accepted: date

33 Published: date

34 **Publisher's Note:** MDPI stays neu-  
35 tral with regard to jurisdictional  
36 claims in published maps and institu-  
37 tional affiliations.



38 **Copyright:** © 2022 by the author  
39 Submitted for possible open access  
40 publication under the terms and con-  
41 ditions of the Creative Commons At-  
42 tribution (CC BY) license ([https://cre-  
43 ativecommons.org/licenses/by/4.0/](https://creativecommons.org/licenses/by/4.0/)).  
44  
45

**Abstract:** We investigated the utility of nuclear and cytoplasmic single nucleotide polymorphism (SNP) markers for timber tracking of the intensively logged and commercialized Amazonian tree *Jacaranda copaia*. Samples of 832 trees from 38 population samples in Bolivia, Brazil, French Guiana, and Peru were genotyped at 113 nuclear, 11 chloroplastic and 4 mitochondrial SNP markers. Bayesian cluster analyzes were carried out to group individuals into homogeneous genetic groups for self-assignment tests of groups of individuals or of individuals to their population of origin. Cluster analysis based on all SNP markers detected seven main genetic groups. Genetic differentiation was high among populations (0.484) and among genetic groups (0.415), and populations showed a strong isolation-by-distance pattern, favoring the group or individual tracking to the correct site of origin. Self-assignment test of groups of individual for all loci was able to determine the population origin of all samples with 100% accuracy. Self-assignment test of individuals was able to assign the origin of 94.5–100% of individuals (accuracy: 91.7–100%). Our results show that the use of 128 SNP makers is suitable to correctly determine the origin sampling site of *J. copaia* timber and they are a useful tool for customs and local and international police.

**Keywords:** illegal logging; forensics; SNP markers; timber tracking; tropical trees; *Jacaranda copaia*

---

## 1. Introduction

Timber logging, whether legal or illegal, contributes to biodiversity loss. Illegal logging is also an economic problem for the legal market, as extremely low-priced wood competes with legal logging, where costs are higher [1,2]. International laws have been established in countries around the world to avoid the importation of illegal timber, as for example, the timber regulation of the United States (Lacey Act), European Union (The EU Timber Regulation No. 995/2010), and Australia (Illegal Logging Prohibition Act), which declare illegal to import timber and its derived products (furniture, musical instruments, etc) originated from illegal timber extraction [2]. These regulations require that importing companies declare the country and specific place of origin, names of the exact species of all plants contained in their wood products, and guarantee that these wood products have a legal origin and have been extracted in a sustainable way in accordance with the laws of the country of origin. In addition to the legislation targeting illegal timber trade, the Convention on International Trade in Endangered Species- CITES [3,4] allows international trade of species listed in Appendix II only when sourced from plantations.

The main problems in the illegal timber trade are in the documentation of the origin of the wood, having falsified certificates and documentation, and with missing or false information [5]. According to the Brazilian Federal Police, in 2021, 90% of the wood extracted from the Amazon Forest had an illegal origin [5,6]. Timber tracking methods are required to fight against illegal timber trade, but methods based only on documentation are sensitive to manipulation and forgery. Therefore, timber export companies and institutions responsible for controlling the origin of imported timber, such as customs, federal police, and Interpol need reliable tools to prove and confirm the declared origin of wood and its derivative products, when traded internationally.

To be able to track the species, country, and specific site origin of imported timber, several methods exist, such as chemical analysis or mass spectrometry [7–9], stable isotopes [10], near-infrared spectroscopy [11], wood anatomy [12,13], and DNA fingerprint [14–21], all exploiting differences in the profiles among species and specific origins. These methods have been shown strong potential for species, country, and site origin determination, especially DNA fingerprinting. Some methods, such as wood anatomy, isotopic profiling, and spectrometric methods are limited in their capacity to identify the origin of timber due to variations among tissue types, individual sample age, individual and population genetic differences and/or

92 environmental influences on timber composition [20]. Thus, genetic  
93 methods that allow correct identification of tree species and tracking of  
94 timber origin are essential for controls on the legality of timber by  
95 public authorities, industry, and trading companies.

96 Here, we evaluate the use of a DNA fingerprinting method to track  
97 the intensively exploited and high value wood of the Neotropical pio-  
98 neer tree *Jacaranda copaia* (Aubl.) D. Don. (Bignoniaceae). The genus *Jac-*  
99 *aranda* sp. has ca. 53 species, all Neotropical, and most are found in the  
100 *Cerrado* around the Amazon Forest. *J. copaia* is the only species of *Jaca-*  
101 *randa* widely distributed in the Amazon Forest [22]. The species occurs  
102 from Central America to western South America, from Belize to Bolivia,  
103 including Brazil, French Guiana, and Peru [22]. In Brazil, the species is  
104 found in the states of Acre, Amapá, Amazonas, Maranhão, Mato  
105 Grosso, Pará, Rondônia, and Roraima, and populations usually have  
106 more than one tree individuals per hectare [22]. *J. copaia* is a fast-grow-  
107 ing species, with an average annual increase of 2.05 cm in diameter at  
108 breast height (DBH) and 1.98 m in height. Adult trees have straight  
109 stems, reaching 106 cm in DBH and 45 m in height [22,23]. Its wood is  
110 light and used for furniture [23,24]. The species is hermaphroditic and  
111 self-incompatible. About 40 species of bees, wasps, butterflies, and  
112 hummingbirds were detected as potential pollinators, although *Eu-*  
113 *glossa* spp. and *Centris* spp. bees was identified as the main pollinators  
114 in the Tapajós National Forest, Brazil [25]. **Reproduction occurs mainly**  
115 **by outcrossing (> 94%)** [26,27]. Each fruit can have up to 250 seeds and  
116 the winged seeds are dispersed by the wind [25]. Mating between re-  
117 lated trees has been reported and attributed to the fact that populations  
118 have intrapopulation spatial genetic structure, ranging from 50 to 500  
119 m, due to the short-distance seed dispersal (up to 100 m), but also due  
120 to short pollen dispersal distance, ranging from 34–90 m [27].

121 Here, samples of *J. copaia* collected in 43 sites in Bolivia, Brazil,  
122 French Guiana, and Peru were used to characterize genetic diversity  
123 and population structure, and to test the ability of chloroplast, mito-  
124 chondrial, and nuclear single nucleotide polymorphism (SNPs) mark-  
125 ers to determine the **geographic origin of wood** of the species harvested  
126 from the Amazon Forest.

## 127 2. Materials and methods

### 128 2.1. Sampling

129 **There is considerable confusion in the literature regarding the**  
130 **naming of samples and populations, therefore, we specify here our**  
131 **naming convention: “individual sample” corresponds to the biological**  
132 **material sampled from a single individual tree; “sampling sites” are lo-**  
133 **ocations where individuals are collected; “population samples” are the**  
134 **collection of individuals sampled at a sampling site; “reporting**  
135 **groups” are the biological populations or genetic groups (i.e. groups of**  
136 **interbreeding individuals with a common ancestry) as estimated from**  
137 **population genetic analyses (i.e. STRUCTURE) and used to assign indi-**

138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166

viduals; “country” are all population samples collected within the political borders of a country. Countries may include several reporting groups, and reporting groups may range across different countries.

Cambium or leaf samples were collected from 832 individual trees from 43 sampling sites (2–31 individuals) in the Amazon rainforest in four countries: Bolivia, Brazil, French Guiana, and Peru (Table 1, Figure 1). The minimum distance between individual trees was 50 m and the distance between sampling sites ranged from 23–2,648 km. All sampled trees were georeferenced using GPS and the plant material collected was stored in a plastic bag labeled with silica gel. In Bolivia, population samples were collected from 5 sampling sites on farms and forestry concessions. In Brazil, population samples were collected from 12 sampling sites in national forests, parks, extractive reserves, and ecological stations, with support from the Chico Mendes Institute of Biodiversity; In French Guiana, population samples were collected by the Institut national de recherche pour l’agriculture, l’alimentation et l’environnement (INRAE) together with the forest authorities (Office National des Forêts, ONF) in 4 sampling sites. Six additional individuals were collected in three sites; In Peru, population samples were collected from 19 sampling sites in national forests, parks, extractive reserves, ecological stations, farms, and forestry concession. However, due to the low sample size at 5 sampling sites (2 trees), for genetic analysis, these individuals were grouped with the closest sampling sites. We also divided the Brazilian samples into West (six population samples) and East (six population samples) origins and Peru into North (nine population samples) and South (eight population samples) origins. All individual samples were registered in a database at the Thünen Institute (SampleDataBase, Grosshansdorf, Germany).



167  
168  
169  
170  
171  
172  
173  
174

**Figure 1.** Spatial distribution of sampling locations for *Jacaranda copaia* in South America.

**Table 1.** Information on the population sample size (*n*); sampling site latitude and longitude; population sample abbreviation (Abbrev); and sample size after combining population samples with only two individuals with the closest population sample (*n*1).

Country	Sampling Site / population sample	<i>n</i>	Latitude	Longitude	Abbrev	<i>n</i> 1
1-F. Guiana	Counami	30	5,41543	-53,175	1FG-Co	32
2-F. Guiana	Sinamary	2	5,2884	-52,916		
3-F. Guiana	Piste de Paul Isnard	27	5,33216	-53,957	2FG-Is	29
4-F. Guiana	Acapou	2	5,27343	-54,218		
5-F. Guiana	Route de Cacao	30	4,56779	-52,406	3FG-Ro	32
6-F. Guiana	Regina	2	4,13118	-52,088		
7-F. Guiana	Saut Maripa	28	3,87833	-51,857	4FG-Ma	28
8-Brazil	ESEC de Maraca-RR	31	3,37032	-61,444	5BW-Ma	31
9-Brazil	Flona de Anauá e arredores-Rorainópolis-RR	28	-0,9339	-60,451	6BW-An	28
10-Brazil	AMATA Flona do Jamari-RO	8	-9,4014	-62,911	7BW-Ja	8



11-Brazil	ESEC do Jarí	15	-0,4955	-52,829	8BW-Jr	15
12-Brazil	Resex Chico Mendes-Xapuri-AC (AMATA-Flona do Jamari)	16	-10,504	-68,595	9BW-Xa	16
13-Brazil	Resex Chico Mendes-Comunidade Cumaru-Assis-AC	15	-10,772	-69,647	10BW-Co	15
14-Brazil	FLONA Amapá-AP	20	0,52785	-51,128	11BE-Am	20
15-Brazil	PARNA da Ana Avilhanas-AM	11	-2,5345	-60,837	12BE-Av	11
16-Brazil	Flona de Tapajós-PA	27	-2,8687	-54,92	13BE-Ta	27
17-Brazil	Resex Tapajós-Arapins-PA	11	-3,0792	-55,278	14BE-Ar	11
18-Brazil	FLONA Tefé-AM	4	-3,5248	-64,972	15BE-Te	4
19-Brazil	FLONA do Carajás	23	-6,0628	-50,059	16BE-Ca	23
20-Peru	Dpto Loreto, Maynas, El Napo, Huiririma Native Community	26	-2,4761	-73,744	17PN-Hu	26
21-Peru	Huaman Urco	27	-3,3128	-73,198	18PN-Ur	27
22-Peru	Dpto Loreto, Maynas, Las Amazonas, Est. Biológica Madre-selva	28	-3,6312	-72,233	19PN-Ma	28
23-Peru	Dpto Loreto, Mayna, Iquitos, Comunidad Campesina Yarina	28	-3,827	-73,567	20PN-Ya	28
24-Peru	Allpahuayo	2	-3,9544	-73,422		
25-Peru	Dpto Loreto, Mar. Ramón Castilla, C. Poblado Unión Progre-sista	27	-3,9727	-70,841	21PN-Pr	29
26-Peru	Dpto Loreto, Requena, Jenaro Herrera Research Centre	11	-4,8966	-73,646	22PN-Re	11
27-Peru	Jenaro Herrera	25	-4,9158	-73,649	23PN-He	25
28-Peru	Dpto Loreto, Alto Amazonas, Jeberos, Centro Poblado Jeberos	26	-5,2598	-76,317	24PN-Je	26
29-Peru	Shucushuyacu	27	-6,0199	-75,854	25PN-Sh	27
30-Peru	Dpto Ucayali, Cor. Portillo, Con. Forestal-Oxigeno para el Mundo	29	-8,8869	-74,034	26PS-Po	29
31-Peru	Dpto Ucayali, Padre Abad, Macuya Forestry Research Station	30	-8,8766	-75,014	27PS-Pa	30
32-Peru	Dpto Ucayali, Atalaya, Tahuania, Concesión Forestal-Javier Díaz	29	-9,9803	-73,817	28PS-Di	29
33-Peru	Dpto Ucayali, Atalaya, Raymondi, Comunidad San Juan de Inuya	12	-10,582	-73,071	29PS-In	12
34-Peru	Dpto Madre de Dios, Tahuamanu, Concesión Forestal Madera-cre	31	-11,145	-69,758	30PS-Md	33
35-Peru	Ibéria	2	-11,299	-69,524		
36-Peru	Dpto Madre de Dios, P.N. Manu, Est. Biológica Cocha Cashu	15	-11,903	-71,403	31PS-Ca	15
37-Peru	Dpto Madre de Dios, Manu, Estación Biológica Los Amigos	30	-12,565	-70,088	32PS-Am	30
38-Peru	Dpto Madre de Dios, R. Nac. Tambopata, La Torre-Sandoval	24	-12,832	-69,284	33PS-Ta	24
39-Bolivia	Riberalta, MABET	15	-10,442	-65,55	34Bo-Ri	15
40-Bolivia	Riberalta, El Desvelo	11	-11,093	-65,746	35Bo-De	11
41-Bolivia	Cobija, Road - Bella Vista	13	-11,198	-68,287	36Bo-Vi	13
42-Bolivia	Riberalta, El Chorro	5	-11,514	-66,327	37Bo-Ch	5
43-Bolivia	Rurrenabaque, Área Protegida Madidi	29	-14,162	-67,905	38Bo-Ma	29

175

176

177

178

179

180

181

## 2.2. DNA extraction and SNPs analysis

Cambium and leaf samples collected in Brazil were sent the Laboratory of Population Genetics and Forestry of São Paulo State University in Ilha Solteira, Brazil (UNESP-FEIS) for DNA isolation. Samples collected in French Guiana, Peru, and Bolivia were sent to Thünen Institute facilities in Großhansdorf, Germany, for DNA isolation. DNA

isolation from leaf and cambium was carried out according to Dumolin et al. [28]. The individual samples were screened for 128 SNP and insertion-deletion INDEL markers (nCpMtSNPs) according to Sebbenn et al. [29], using the MassARRAY® iPLEX™ (Assay Design Suite v2.0 [Agena Bioscience™, San Diego, EUA]) genotyping, including 113 nuclear SNPs (nSNPs), 11 chloroplast SNPs (CpSNPs), and 4 mitochondrial SNPs (MtSNPs). The combined nuclear, chloroplast, and mitochondrial SNP markers will hereafter be referred to as nCpMtSNP markers.

### 2.3. Genetic diversity

The genetic diversity for nSNPs was determined for each population sample and country by the total number of alleles ( $K$ ), percent of polymorphic loci ( $P\%$ ), observed ( $H_o$ ), and expected ( $H_e$ ) heterozygosity. The mean fixation index ( $F$ ) was estimated to quantify the inbreeding coefficient within each population sample or country and the statistical significance of the  $F$  values was determined permuting alleles among individuals. For CpMtSNPs, the genetic diversity was determined for each site by  $K$  and percent of polymorphic loci ( $P\%$ ). We also estimated the percent of polymorphic loci ( $P\%_{128}$ ), observed for all nCpMtSNP markers. These analyses were carried out using the GDANT 1b software [30].

### 2.4. Bayesian clustering analysis

The Bayesian clustering method implemented in the software STRUCTURE v.2.3.4 [31] was used for the 128 nCpMtSNP markers to check for the number of reporting groups (*i.e.* genetic populations). We set the length of burn-in and Markov chain Monte Carlo to 10,000 and tested  $K$  values from 2–10 for 30 times. We used the admixture model with correlated allele frequencies. The optimal number of reporting groups was estimated with the  $\Delta K$  method [32]. For each optimal  $K$ , data from the 30 STRUCTURE runs was permuted with CLUMPP v.1.1.2 [33] to obtain the final  $Q$  values for each individual. The optimal number of reporting groups was determined with the  $\Delta K$  method described by Evanno et al. [31] using the software CLUMPAK [33].

### 2.5. Genetic differentiation among populations, countries and clusters

The overall and pairwise genetic differentiation ( $F_{st}$ ) was estimated among all population sample (locals), clusters determined by STRUCTURE analysis, among countries, and population sample within country, and for all nCpMtSNP, nSNP, and CpMtSNPs markers. These analyses were carried out using the GDANT 1b software [30]. The pairwise  $F_{st}$  and spatial distance among population samples with sample size higher than 16 individuals and among clusters determined by STRUCTURE analysis, was used to investigate the isolation by distance (IBD) gene dispersal, using the Spearman correlation coefficient



( $\rho$ ). We also assessed the IBD by the estimate the spatial genetic structure (SGS) based on calculation of the coancestry coefficient ( $\theta_{ij}$ ), described in Loiselle et al. [34], between mean pairs of individuals within 10 distance-classes (0–198, 199–441, 442–655, 656–860, 861–1068, 1069–1409, 1410–1838, 1839–2770, 2771–2478, and 2479–2646 km) determined using the same number of pairs per classes, and using the SPAGED1.2 software [35]. The statistical significance of the average  $\theta_{ij}$  of each distance class was derived by comparing the limits of the confidence interval at 95% probability for the average  $\theta_{ij}$  for each distance class, estimated permuting (1000 times) genotypes between distances classes using the SPAGED1 software.

## 2.6. Genetic assignment

No wood DNA samples were used here. The purpose of our study is to determine if the population structure and genetic differentiation among *J. copaia* across the species range is enough to allow correct determination of the origin of an individual sample. Provided that DNA can be extracted from the target material, the same analytical procedures can be used for DNA collected from wood stored by timber companies, sawn wood, instruments musicals, furniture, etc. Similar procedures have been used in other studies for the same purpose [16–19]. The Bayesian method [36] implemented in GeneClass 2.0 [37] was used to assign population samples (group of individuals collected in a sampling site) and individuals to its reporting group and country. The genetic grouping determined by the STRUCTURE analysis was used to group individuals into reporting groups. Both population sample and individual sample assignment were carried out for all nCpMtSNP markers, and the most likely reporting group determined by the highest score by the Bayesian criteria was used as an indicator of the power of the markers to compute the proportion of correctly assigned population sample or individuals in self-assignments tests [38]. Here, the individual samples from the population samples (reference data) were self-classified to the reporting groups and countries, using the leave-one-out approach [39]. We also estimated the overall,  $\geq 80\%$  and  $\geq 95\%$  score rate of population samples and individual samples assigned to correct origin reporting group.

## 3. Results

### 3.1. Genetic diversity

The rate of missing data ranged among loci from 0.1 to 47.1%, with average of 4.4% (Table S1) and among population samples ranged from 0.8 to 32.3%, average of 4.6% (Table 2). Across population samples, the total number of alleles ( $K$ ) for 113 nSNPs ranged from 122–226 alleles (mean of 183) and for 15 chloroplast and mitochondrial loci (CpMtSNPs) from 14–19 alleles, with a mean of 15.9 alleles (Table 2).

The percent of polymorphic loci ( $P\%$ ) for 113 nSNPs was higher (ranged among populations from 23–88.5%) than for 15 CpMtSNPs (0–26.7%) and 128 nCpMtSNP markers (20.3–78.1%, mean of 55.9%). The observed heterozygosity ( $H_o$ ) ranged from 0.023–0.343 (mean of 0.178), the expected heterozygosity ( $H_e$ ) for nSNPs ranged among populations from 0.028–0.348 (mean of 0.204). The values of  $H_e$  for nSNPs were high ( $\geq 50\%$  of the maximum value for biallelic loci, 0.5) for 21 populations, where the highest values were observed for Brazil and Bolivia populations. The mean intrapopulation fixation index ( $F$ ) ranged from -0.086–0.295 and was significantly ( $P < 0.05$ ) higher than zero in six of the 38 populations, suggesting inbreeding. At country level, for nSNPs,  $K$  and  $P\%$  were highest in Brazil for nSNPs (226, 100%, respectively) and CpMtSNPs (30, 100%, respectively), and lowest in French Guiana (nSNPs= 200, 76.1%, respectively; CpMtSNPs= 18; 20%, respectively). Bolivia presented the highest  $H_o$  (0.319) and  $H_e$  (0.359) values and lowest  $F_{IS}$  (0.113), where French Guiana presented the lowest  $H_o$  (0.095) and  $H_e$  (0.192) values and the highest  $F_{IS}$  (0.506).

**Table 2.** Genetic diversity of *J. Copaia* at different population samples and countries for 113 nuclear (nSNPs), 15 chloroplast and mitochondrial (CpMtSNPs), and 128 nuclear, chloroplast and mitochondrial (nCpMtSNPs) markers. \* denotes  $P < 0.05$ ;  $n$  is the sample size;  $M\%$  is the percentage of missing data;  $P\%$  is the percentage of polymorphic loci;  $P\%_{128}$  is the percentage of polymorphic loci for all 128 nCpMtSNPs;  $K$  is the total number of alleles;  $H_o$  and  $H_e$  are the observed and expected heterozygosity, respectively;  $F$  is the fixation index.

Sample	$n$	nSNPs						CpMtSNPs		nCpMtSNPs
		$M(\%)$	$K$	$P\%$	$H_o$	$H_e$	$F$	$K$	$P\%$	$P\%_{128}$
1FG-Co	32	6.6	122	50.4	0.023	0.028	0.005	16	6.7	45.3
2FG-Is	29	4.9	190	69	0.109	0.197	0.25*	17	13.3	62.5
3FG-Ro	32	5.7	196	71.7	0.094	0.202	0.295*	16	6.7	64.1
4FG-Ma	28	3.5	195	72.6	0.148	0.254	0.286*	16	13.3	65.7
5BW-Ma	31	2.4	193	82.3	0.299	0.309	0.005	16	6.7	73.4
6BW-An	28	1.5	195	70.8	0.291	0.292	-0.022	15	0	72.7
7BW-Ja	8	32.3	207	80.5	0.194	0.298	0.198*	17	13.3	72.6
8BW-Jr	15	4.3	207	83.2	0.205	0.287	0.057	16	13.3	75
9BW-Xa	16	4.5	204	72.6	0.278	0.29	-0.021	16	6.7	75
10BW-Cu	15	1.8	147	83.2	0.261	0.275	0.016	15	0	73.5
11BE-Am	20	1.1	204	75.2	0.215	0.231	0.052	15	0	66.4
12BE-Av	11	7.5	207	61.9	0.278	0.29	-0.021	15	0	64.8
13BE-Ta	27	2.8	197	83.2	0.261	0.275	0.016	15	0	74.2
14BE-Ar	11	3.1	183	69	0.238	0.276	0.078	15	0	68
15BE-Te	4	1.4	207	83.2	0.141	0.257	0.057	15	0	73.5
16BE-Ca	23	5.3	190	81.4	0.275	0.282	-0.019	15	0	75
17PN-Hu	26	2.4	149	30.1	0.053	0.059	0.017	17	13.3	28.1
18PN-Ur	27	5.0	147	31.9	0.06	0.066	0.018	17	13.3	29.7
19PN-Ma	28	4.4	147	30.1	0.054	0.066	0.031	19	20	28.9
20PN-Ya	28	2.5	147	30.1	0.059	0.064	0.024	16	6.7	27.4
21PN-Pr	29	4.7	142	30.1	0.065	0.069	0.016	19	26.7	29.7

22PN-Re	11	1.8	146	25.7	0.043	0.065	0.072	15	0	22.7
23PN-He	25	3.8	145	29.2	0.05	0.056	0.018	16	20	28.1
24PN-Je	26	7.9	142	28.3	0.041	0.061	0.087*	19	26.7	28.1
25PN-Sh	27	3.2	147	26.5	0.047	0.053	0.028	14	0	23.4
26PS-Po	29	13.9	146	30.1	0.046	0.075	0.087*	16	6.7	27.4
27PS-Pa	30	3.1	140	30.1	0.063	0.064	0.02	14	0	26.6
28PS-Di	29	4.9	139	23.9	0.054	0.061	0.024	15	0	21.1
29PS-In	12	4.8	213	23	0.056	0.061	0.017	15	0	20.3
30PS-Ma	33	4.5	206	88.5	0.297	0.315	0.045	15	0	78.1
31PS-Ca	15	2.4	213	82.3	0.271	0.31	0.073	15	0	72.7
32PS-Am	30	4.5	209	88.5	0.312	0.317	-0.01	15	0	78.1
33PS-Ta	24	3.2	209	85	0.284	0.314	0.08	15	0	75
34Bo-Mb	15	2.8	207	84.1	0.321	0.324	-0.026	16	6.7	75
35Bo-De	11	1.3	206	84.1	0.297	0.312	-0.012	16	6.7	75
36Bo-Vi	13	0.8	205	82.3	0.313	0.313	-0.023	15	0	72.7
37Bo-Ch	5	1.4	212	82.3	0.34	0.348	-0.086	19	26.7	75.8
38Bo-Ma	29	1.3	226	87.6	0.322	0.313	-0.032	15	0	77.3
Overall	832	4.4	183	100	0.178	0.204	0.086*	15.9	6.7	55.9
F. Guiana	121	5.2	200	76.1	0.095	0.192	0.506*	18	20	69.5
Brazil	209	4.2	226	100	0.261	0.354	0.264*	30	100	100
Peru	429	4.7	217	92.9	0.111	0.222	0.498*	23	53.3	88.3
Bolivia	73	1.5	218	92.9	0.319	0.359	0.113*	20	33.3	85.9

295

### 296 *3.2. Bayesian cluster analysis*

297 The results of the STRUCTURE analysis showed that based on delta K, the best representation of the  
 298 data is obtained with a K of 4 or 9 different reporting groups (i.e. genetic populations) (Figure 2).  
 299 With 5 groups, there is less mixture of gene pools within individuals. With K= 5, 85% of all individ-  
 300 uals had a gene pool membership coefficient above 0.9, while for K= 4, only 76% of all individuals  
 301 had gene pool membership coefficients above 0.9. The STRUCTURE analysis based on CpMtSNPs  
 302 showed a clear distinction between French Guiana and North Brazil sampling sites (red) from North  
 303 Peru, South Brazil (6BW-An; 7B-Ja), East Bolivia (34Bo-Ri, 35Bo-De) (green and white), and from  
 304 South Peru (29P-In; 30P-Md, 31P-Ca, 32P-Am, 33P-Ta), Southwest Brazil (9BW-Xa, 10B-Co) and  
 305 Bolivia (36Bo-Vi, 37Bo-Ch; 38Bo-Ma) (blue) (Figure 2A). The STRUCTURE analysis based on nSNP  
 306 markers, revealed that there are a few population samples with mixtures of individuals from differ-  
 307 ent gene pools (e.g. Flona do Jamari (7BW-Ja), Resex Tapajos-Arapins (14BE-Ar), and ESEC  
 308 Maraca (5BW-Ma) in Brazil and Piste de Paul Isnard (2FG-Is), Route de Cacao (3FG-Ro) and  
 309 Saut Maripa (4FG-Ma) in French Guiana in Figure 3B). We found one genetic group (K3) which  
 310 we only found in two Brazilian sampling sites (8BW-Jr, 11BE-Am) and another genetic group (K1)  
 311 distributed more in the middle of the Amazon basin (Figure 2B).

312

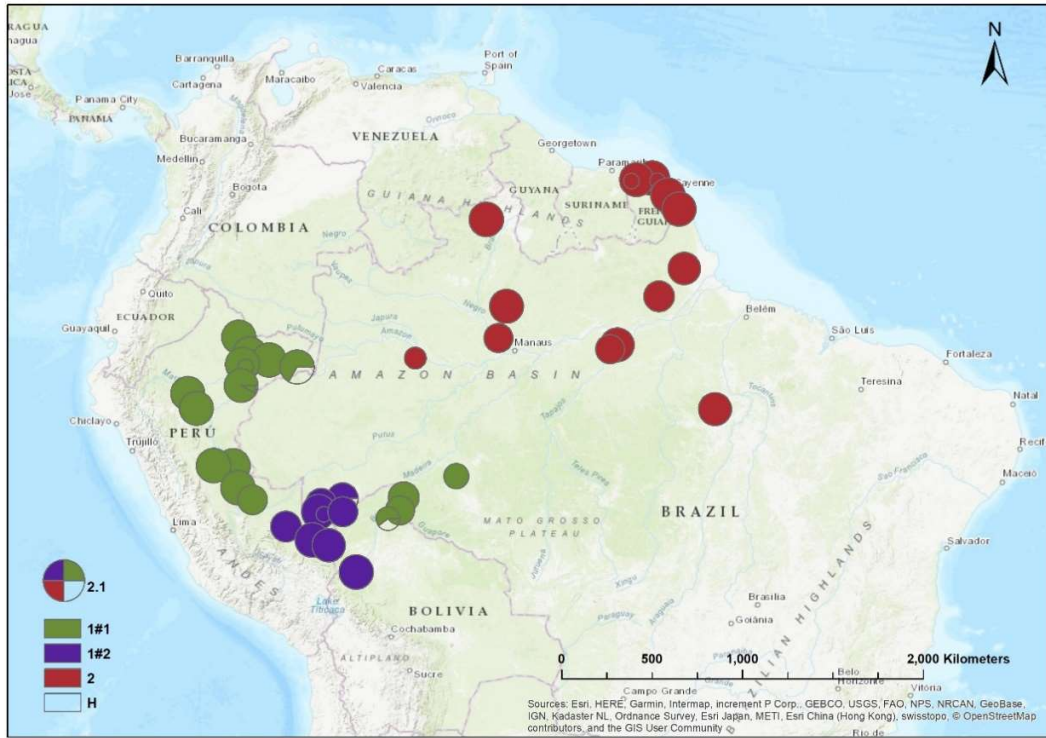
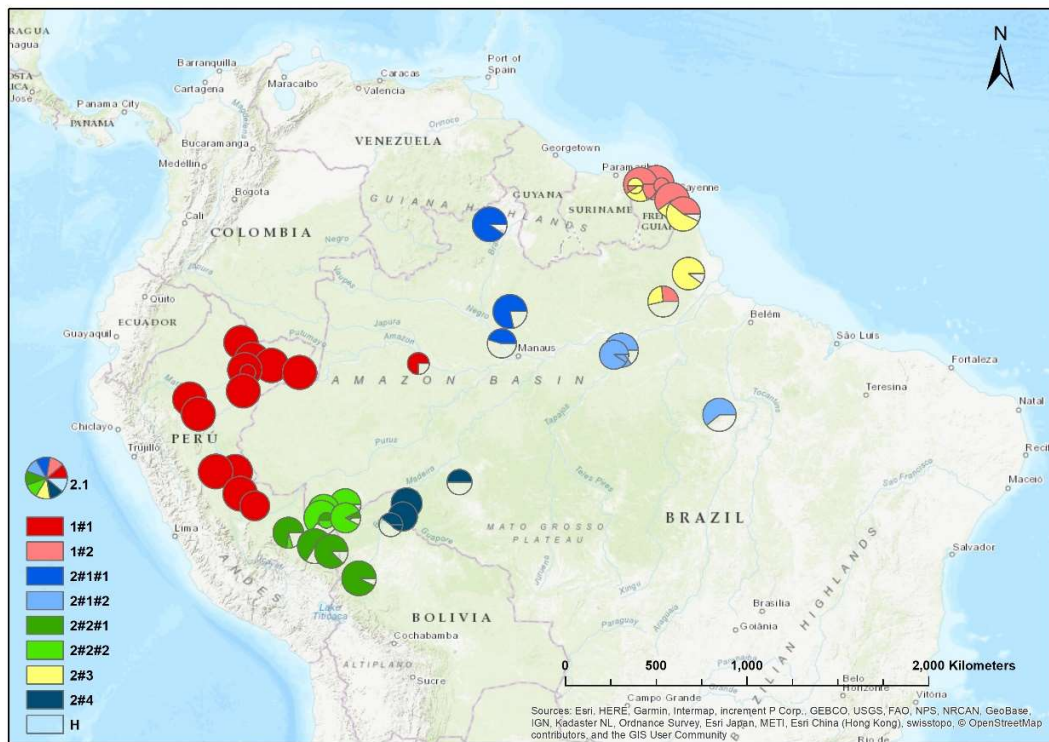


Figure 3A

313

314



315

316

Figure 3B



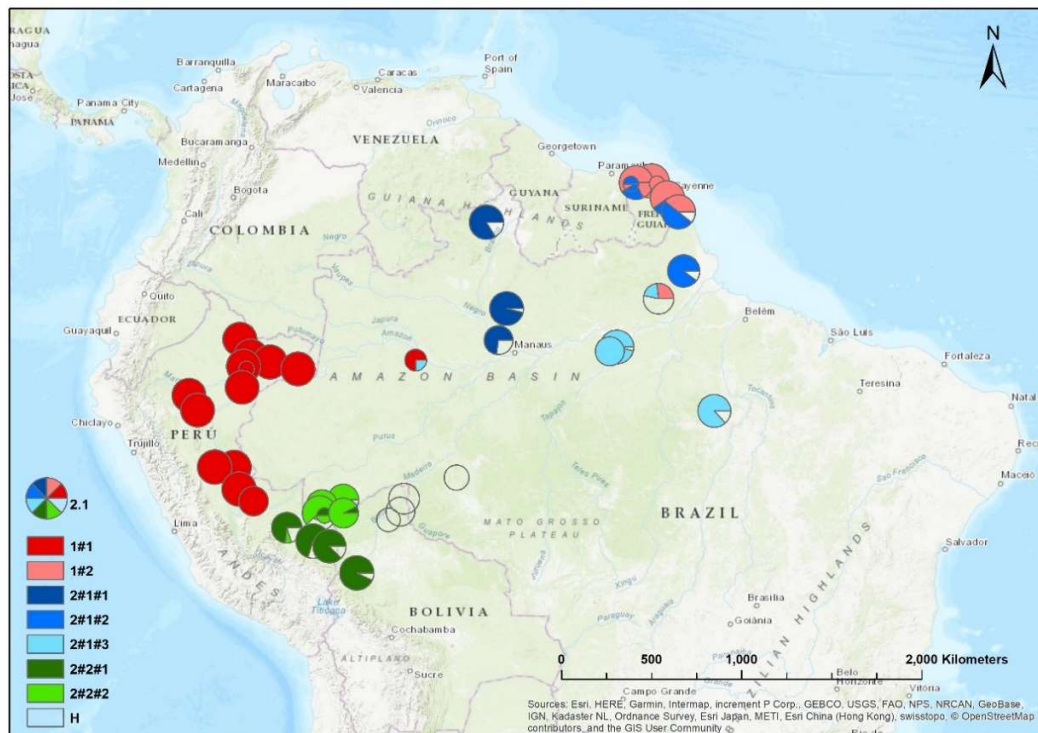


Figure 3C

**Figure 2.** Population structure of *Jacaranda copaia* in South America showing the Spatial distribution of reporting groups (i.e. genetic clusters) based on CpMtSNP (A, K=4), nSNP (B; K=9), and nCpMtSNP (C; K=8) markers estimated by STRUCTURE (K= 4 to 9).

### 3.3. Population differentiation

To determine differences between nuclear SNPs (nSNPs) and chloroplast and mitochondrial markers (CpMtSNPs) in estimating genetic differentiation ( $F_{st}$ ) among population samples, we compared the results among all 128 nCpMtSNPs, 113 nSNPs, and 15 CpMtSNPs (Table 3). The results showed that the  $F_{st}$  among all population samples and countries was higher for CpMtSNPs (0.942 and 0.695, respectively) than for all nCpMtSNPs (0.484 and 0.295, respectively) and nSNPs (0.415 and 0.233, respectively). The  $F_{st}$  was also higher among clusters determined by STRUCTURE analysis for CpMtSNPs (0.896) than CpMtSNPs (0.401) and nSNPs (0.315). Among population samples within Brazil, Peru, and Bolivia, the  $F_{st}$  was also higher for CpMtSNPs (0.974, 0.741, and 0.735, respectively) than for all nCpMtSNPs (0.3, 0.466, and 0.142, respectively) and nSNPs (0.224, 0.456, and 0.107, respectively). In contrast, among population samples within French Guiana, the  $F_{st}$  was higher for nCpMtSNPs (0.12) and nSNPs (0.117) than CpMtSNPs (0.004). The results also showed that the estimates of  $F_{st}$  for nCpMtSNPs, nSNPs, and CpMtSNPs was generally highest within Peru and Brazil, where the number of sampling sites was highest (Peru= 17, Brazil= 12), indicating that the number of sampling sites is important to detect high population differentiation within countries. The highest  $F_{st}$  for CpMtSNPs and nCpMtSNPs than for nSNPs indicated that the combination of 113 nuclear and 15 cytoplasmatic SNP markers increases the capacity to detect genetic differences among populations.



341 **Table 3.** Genetic differentiation ( $F_{st}$ ) among all population samples, all countries, all population samples within  
 342 countries for all nCpMtSNPs (128), nSNPs (113), and CpMtSNPs (15).

Sample	$np$	nCpMtSNPs (128)	nSNPs (113)	CpMtSNPs (15)
All populations	38	0.484 ± 0.043*	0.415 ± 0.032*	0.942 ± 0.042*
<u>Genetic groups</u>	8; 9; 4€	0.401 ± 0.044*	0.315 ± 0.03*	0.896 ± 0.094*
Countries	4	0.295 ± 0.036*	0.233 ± 0.022*	0.695 ± 0.144*
French Guiana	4	0.120 ± 0.017*	0.117 ± 0.017*	0.011 ± 0.002
Brazil	12	0.299 ± 0.049*	0.224 ± 0.03*	0.925 ± 0.103*
Peru	17	0.466 ± 0.056*	0.456 ± 0.056*	0.741 ± 0.267*
Bolivia	5	0.142 ± 0.034*	0.107 ± 0.024*	0.735 ± 0.383*

343 \*  $P < 0.05$ ;  $np$  is number of populations;  $\pm$  is the 95% standard error, 1.96SE; € is the number the genetic groups  
 344 determined for nCpMtSNPs, nSNPs, and CpMtSNPs, using Bayesian cluster analyzes.

345

#### 346 3.4. Isolation by distance

347 The Spearman correlation coefficient ( $\rho$ ) between pairwise  $F_{st}$  and spatial distance between popu-  
 348 lation samples ( $n > 16$  individuals) based on 128 nCpMtSNP markers was significantly ( $P < 0.001$ )  
 349 higher than zero ( $\rho = 0.626$ ), indicating an isolation by distance (IBD) genetic pattern (Figure 3A). The  
 350 Spearman correlation coefficient ( $\rho$ ) between pairwise  $F_{st}$  and spatial distance between clustered lo-  
 351 cals, as determined by STRUCTURE analysis was also significantly ( $P < 0.001$ ) greater than zero for  
 352 nCpMtSNPs ( $\rho = 0.521$ , Figure 3B) and nSNPs ( $\rho = 0.552$ , Figures 3C), reinforcing the occurrence of an  
 353 isolation by distance genetic pattern. The estimate of pairwise coancestry for nSNP and nCpMtSNP  
 354 markers between individuals within ten distance classes was significantly higher than zero up to the  
 355 distance of 655 km, where for CpMtSNP markers was significant up to 1068 km, and non-significant  
 356 or significantly lower than zero in the other distance classes (Figure S1).

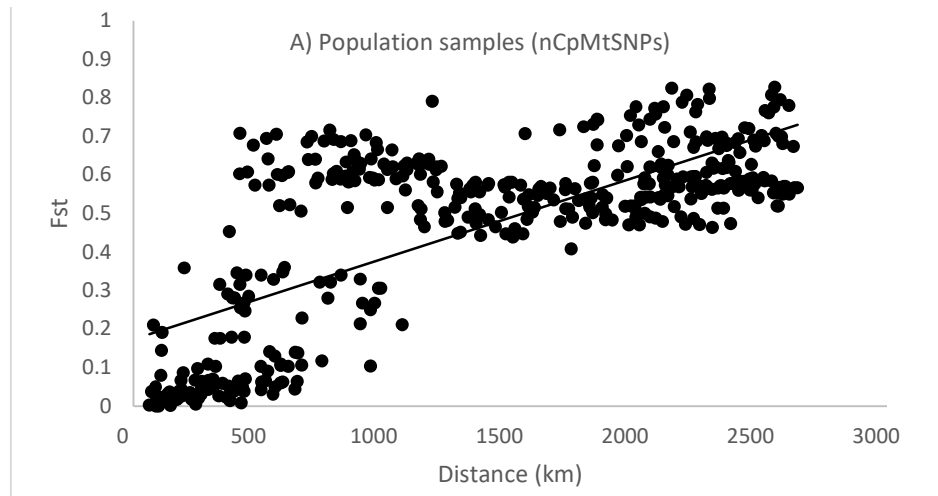
357

358

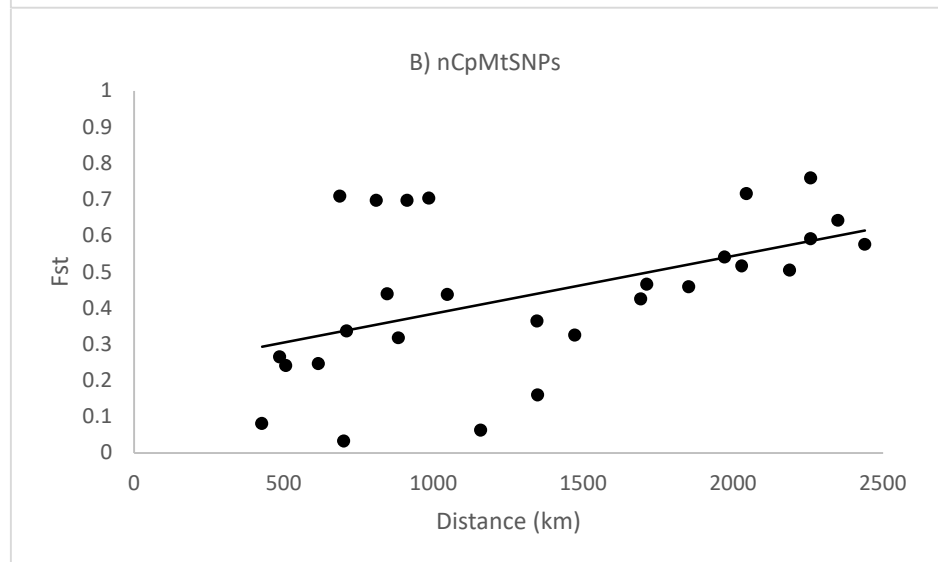
359

360

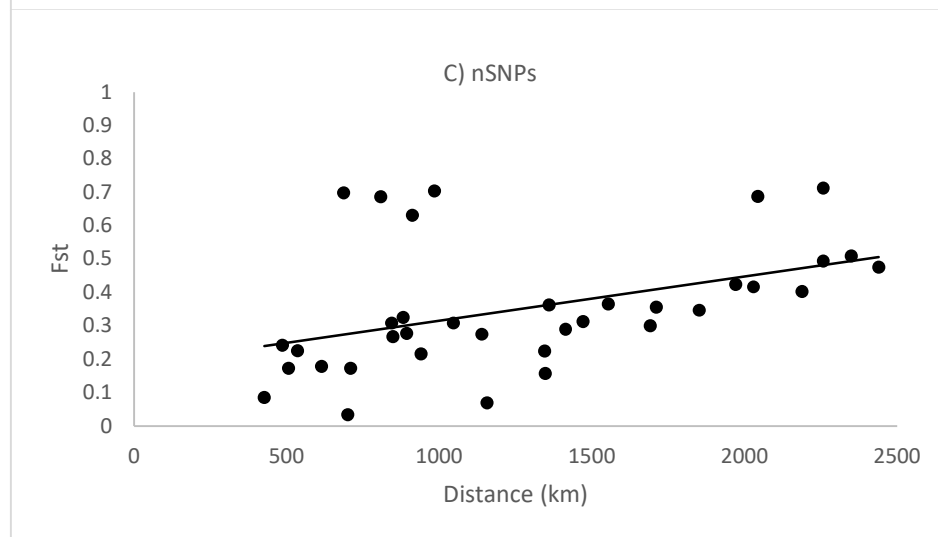
361



362

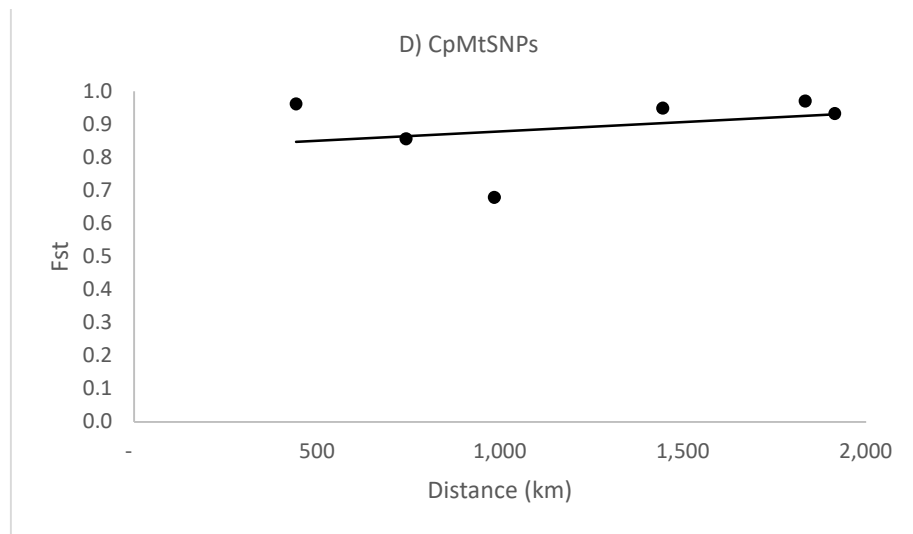


363



364

365



366  
367

368 **Figure 3.** Pattern of isolation by distance in *Jacaranda copaia* population samples (A) and genetic populations  
 369 determined by cluster analysis using STRUCTURE analysis (B, C, D).  $F_{ST}$  is the pairwise genetic differentiation  
 370 between population samples (sample size higher than 16 individuals) for all 128 nCpMtSNP markers (A) and the  
 371 pairwise genetic differentiations between clustered locals as determined by STRUCTURE analysis for all 128  
 372 nCpMtSNP (B), 113 nSNP (C), and 15 CpMtSNP (D), markers. The Spearman correlation coefficient ( $\rho$ ) was  
 373 significantly higher than zero ( $P < 0.01$ ) between pairwise  $F_{st}$  and spatial distance among populations (A) for all  
 374 128 nCpMtSNP markers (0.645), as well as between  $F_{st}$  and spatial distance between clustered locals as deter-  
 375 mined by STRUCTURE analysis (B, C, and D) for nCpMtSNPs (0.521) and nSNPs (0.552), where non-significantly  
 376 correlation was observed for CpMtSNsP (0.142).

377

### 378 3.5. Genetic assignment

379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396

We performed self-assignment tests based on seven main genetic clusters determined in STRUCTURE analysis (Figure 2C) using all 128 nCpMtSNP markers. Due to the genetic similarity and small sample size of the 15BE Te site (Figure 2A, B), it was grouped with the 12 population samples from Northern Peru for statistical analyses. Self-assignment tests were carried out to assign or exclude the seven clusters as the origin of individuals or groups of individuals (locations grouped by cluster) from our sample. The reporting groups assignment test was able to self-assign 100% of population samples (score of 100%) to the correct origin for seven reporting groups and four countries (Table 4). Self-assignment test of individuals was able to self-assign 94.5–100% (mean of 98.4%) of individuals to the correct origin reporting group (score ranging from 91.7–100%, mean of 97.3%). The rate of correct individual assignment for scores  $\geq 80\%$  and  $\geq 95\%$  were high (95.5 and 89.9%, respectively). The individual sample assignment test was able to self-assign from 87.6% of individual from French Guiana to 100% of individuals from Bolivia to the correct origin country (score ranging from 93.4–99.9%). The rate of correct individual assignment to the correct country source with scores  $\geq 80\%$  and  $\geq 95\%$  ranged from 87.8–100% and 72.8–100%, respectively.

397  
398

**Table 4.** Corrected population sample and individual self-assignment rate and scores for reporting groups and countries (mean score) for all 128 nCpMtSNP markers. Rate means the percent of individuals assigned to the

399 correct source reporting group or country; Rates > 80 and > 95% are the rates of individuals assigned to the  
 400 correct source reporting group or country, with a score values  $\geq$  than 80 and 95%, respectively.

401

Reporting groups	Rate	Score	Rate	Score	Rate: Score> 80%	Rate: Score>95%
1	100	100	100	99.9	100	99.4
2	100	100	100	100	100	100
3	100	100	98.6	99.8	100	98.6
4	100	100	96.0	93.2	89.7	69.2
5	100	100	100	100	100	100
6	100	100	94.5	91.7	82.6	67.4
7	100	100	100	98.2	96.5	94.7
Overall	100	100	98.4	97.5	95.5	89.9
Country						
French Guiana	100	100	87.6	98.1	99.1	99.1
Brazil	100	100	98.1	99.9	100	98.5
Peru	100	100	98.1	93.4	87.8	72.8
Bolivia	100	100	100	99.9	100	100

402 Reporting groups:

403 1: 1FG-Co, 2FG-Is, 3FG-Ro, 4FG-Ma, 11BE-Am, 8BW-Jr;

404 2: 13BE-Ta, 14BEAr, 16BE-Ca;

405 3: 5BW-Ma, 12BE-Av, 6BE-An, 15BE-Te;

406 4: 17PN-Hu, 18PN-Ur, 19PN-Ma, 20PN-Ya, 21PN-Pr, 22PN-Re, 23PN-He, 24PN-Je, 25PN-Sh;

407 5: 7BW-Ja, 34Bo-Ri, 35Bo-De, 37Bo-Ch;

408 6: 26PS-Pa, 27PS-Pr, 28PS-Di, 29PS-I;

409 7: 9BW-Xa, 10BW-Co, 30PS-Md, 31PS-Ca, 32PS-Am, 33PS-Ta, 36Bo-Vi, 38Bo-Ma.

410

## 4. Discussion

411

### 4.1. Genetic diversity

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

The illegal timber trade, both species-specific and illegally sourced, has become a major problem for tropical forests. Control of the chain of custody for timber originating from the Amazonian Forest, as well as timber from other biomes and regions of the planet, is urgent to verify the species, countries, and place of origin declared in the transport and export documentation and restrict or even eradicate illegal logging. Our study shows the potential of DNA fingerprinting to track the country and origin of timber, as well as a method to follow and verify the chain of custody for *J. copaia* timber products. Our study shows only moderate levels of genetic diversity for nSNPs and low levels of genetic diversity for CpMtSNPs for *J. copaia*. However, the observed ( $H_o$ ) and expected ( $H_e$ ) heterozygosity were generally moderate to high, respectively, and are within the reported pattern identified for other Neotropical trees using nSNP markers. For example, in *Carapa* sp.  $H_o$  has been observed ranging among populations from 0.017–0.264 and  $H_e$  ranging from 0.02–0.251 [41], in *Dipteryx* sp.  $H_o$  has been observed ranging among populations from 0.03–0.39 and  $H_e$  ranging from 0.03–0.29 [19], and in *Handroanthus* sp.  $H_o$  has been observed

430 ranging among populations from 0.15–0.276 and  $H_e$  ranging from  
431 0.071–0.371 [42]. The genetic diversity of *J. copaia* was especially low in  
432 some Peru population samples, due to the low percent of polymorphic  
433 loci ( $P\%_{128}$ ) for all nCpMtSNP markers. The estimate of intrapopulation  
434 fixation index ( $F$ ) indicated inbreeding in some populations.  
435 Inbreeding has also been observed for other Neotropical trees using  
436 SNP markers (Table S2). However, because *J. copaia* is self-incompatible  
437 [25], and our population samples were taken from geographically  
438 distant trees, the observed inbreeding is likely an artefact of the  
439 Wahlund effect [43] due to mixtures of samples from different  
440 subpopulations, as previously shown for *J. copaia* [44].

441

#### 442 4.2. Population genetic differentiation

443 The presence of population genetic differentiation and intrapopulation spatial genetic structure  
444 (SGS), or the occurrence of isolation by distance patterns (IBD), is key for the assignment of timber to  
445 different origins [1,45], including country, population, and regions within countries. High genetic  
446 differentiation among the different reporting groups increases the success of genetic assignment [1,  
447 45]. Our results showed strong genetic differentiation among all reporting groups and between  
448 countries, as well as SGS and a pattern of IBD. The  $F_{st}$  for all nCpMtSNP, nSNPs, and CpMtSNPs  
449 was generally high, especially for CpMtSNPs. The  $F_{st}$  was higher among reporting groups than  
450 among countries, as well as the mean assignment test of individuals to correct origin reporting  
451 groups was higher than among correct country.

452 Within countries for all nCpMtSNP and CpMtSNP markers,  $F_{st}$  was lowest among French  
453 Guiana population samples. In agreement, the lowest self-assignment of individuals was found in  
454 French Guiana (87.6%). Brazil, Peru and Bolivia showed a high rate of self-assignment of individuals  
455 to the correct country of origin (> 98%). The results showed that the  $F_{st}$  among all population sam-  
456 ples, reporting groups, and among countries was higher for CpMtSNPs than for nSNPs. This can be  
457 attributed to the fact that CpMtSNPs have lower variability due to uniparental inheritance, while  
458 nuclear SNPs have higher diversity due to biparental inheritance and recombination. Thus,  
459 CpMtSNPs can show low variation within or among closed reporting groups, where nSNPs can pre-  
460 sent high variation and share many alleles among reporting groups, as can be observed in Figure 2.  
461 Furthermore, the percentage of polymorphic loci was lower for CpMtSNPs than nSNPs, where the  
462 highest polymorphism was observed, and many population samples shared the same alleles, which  
463 reduces genetic differences between population samples. These results indicate the importance of

464 using both nSNP and CpMtSNP markers to detect greater population genetic structure and thus in-  
465 crease the potential for assigning wood origins to population samples, reporting groups, and coun-  
466 tries.

#### 468 4.3. Genetic assignment and practical application

469 The results indicate a high power to correctly assign reporting  
470 groups at both levels: among countries and reporting groups (100%).  
471 Our results confirm that this approach offers the possibility of high lev-  
472 els of success when grouping timber samples based on their origin sam-  
473 pling site. This success can be attributed to the general wide genetic  
474 differentiation among countries and reporting groups. Even when the  
475 differentiation was only moderate ( $F_{st} < 0.2$ ) (French Guiana, Bolivia),  
476 differences in allele frequencies between analyzed groups were enough  
477 to produce high scores for the determination of the correct  
478 origin. However, it is important to emphasize that this is true so long  
479 as the reference data (collection of samples of individuals from  
480 sampling sites encompassing the species range) set contains genetic  
481 data of the timber's original population. Therefore, the success in  
482 determining the origin of *J. copaia* timber, as well as other species,  
483 greatly depends on the reference data set developed by the authorities  
484 involved in controlling the origin of timber (reference population  
485 samples). The reference data set should include genetic data from all  
486 sites where harvesting is legally practiced.

487 The results for assignment test of individuals for all nCpMtSNPs  
488 was lower than for assignment test of groups of individuals, but also  
489 showed a high power of correct assignment among countries and re-  
490 porting groups. In general, the results for the assignment test of  
491 individuals indicate that this approach has the power to determine the  
492 reporting group of origin, and in some cases estimate the specific site  
493 of timber origin. When using CpMtSNP markers, it is necessary to de-  
494 velop a large number of loci to improve both population sample and  
495 individual assignment tests. Similar results of better self-assignment  
496 tests at the population sample level rather than at the individual was  
497 reported for *Hymenaea* sp. [18]. For *Swietenia macrophylla*, based on  
498 nuclear microsatellite loci (SSRs), the assignment test was higher at the  
499 country (82%) than population sample (71%) level [1]. For the  
500 Malaysian *Gonystylus bancanus* tree, the self-assignment rate using a set  
501 of 16 nSSRs was lower (55%) than that observed here at the population  
502 sample level [46]. For SNP markers of *Entandrophragma cylindricum*, the  
503 assignment at the country level ranged from 66–74%, depending on the  
504 assignment method [15]. Many other studies on tropical and temperate  
505 tree species have been developed using microsatellite and SNP markers  
506 to track timber showing that, thanks to spatial genetic structure and  
507 genetic differentiation, DNA fingerprinting is the most effective tool to  
508 track the country and population of origin of timber [45–51].



510 Finally, we recommend that the timber sector add such genetic  
511 controls as an independent audit beyond the paper-based proof of the  
512 chain of custody [1]. It is important to note that the power of the genetic  
513 reference data to detect false declarations reaches 100% if more than  
514 one sample with the same declaration is tested.

#### 515 4.4. Conclusion

516 The genetic differentiation ( $F_{st}$ ) among all population samples,  
517 reporting groups, and countries was generally high, especially based  
518 on CpMtSNP markers. Furthermore, there is a strong isolation by  
519 distance pattern among population samples, favoring the tracking of  
520 population samples or individual samples to the correct site. For self-  
521 assignment tests, we were able to obtain 100% accuracy to the country  
522 and reporting group for all samples using all 128 nCpMtSNPs and 113  
523 nSNPs. Our results show that the use of all nCpMtSNP or nSNP  
524 markers are suitable to correctly determine the country and reporting  
525 group of origin for *J. copaia* timber, thus offering a very useful tool for  
526 customs and local and international police. The *J. copaia* reference data-  
527 base of our study represents a robust assignment tool available to tim-  
528 ber companies or government agencies to test and validate origin dec-  
529 larations. We recommend the use of the method described herein for  
530 other native tropical species, since it is highly effective in identifying  
531 the origin of wood, thus helping the police and other relevant agencies  
532 in the definition of illegally deforested areas, as well as unsustainable  
533 extraction.

534 **Supplementary Materials:** The following supporting information can be  
535 downloaded at: [www.mdpi.com/xxx/s1](http://www.mdpi.com/xxx/s1), Figure S1. Spatial genetic structure  
536 (SGS) of *Jacaranda copaia* for all sample individuals in all populations for all 128  
537 nCpMtSNP, 113 nSNP, and 15 CpMtSNP markers. Unfilled circles indicated  
538 mean  $\theta_{ij}$  values significant different than zero ( $P < 0.05$ ) of the hypothesis of  
539 absence of SGS ( $H_0: \theta_{ij} = 0$ ); vertical bars show the standard error at 95% of  
540 probability; **Table S1.** Population sample size (n), missing data (%), genetic di-  
541 versity, and differentiation among population samples ( $F_{st}$ ) for each and mean  
542 nCpMtSNP markers: k= number of alleles;  $H_o$ = observed heterozygosity;  $H_e$ =  
543 expected heterozygosity; **Table S2.** Population genetic diversity and population  
544 structure ( $F_{st}$ ) for Neotropical tree species based on SNP markers.

545 **Author Contributions:** Conceptualization, Lorena F.M. Capo and Alexandre  
546 M. Sebbenn, Bernd Degen, Celine Blanc Jolivet, and Stephen Cavers; Method-  
547 ology, Alexandre M. Sebbenn, Celine Blanc-Jolivet; Data validation, Malte  
548 Mader, Valerie Troispoux, and Celine Blanc-Jolivet; Formal Analysis, Alexan-  
549 dre M. Sebbenn and Celine Blanc-Jolivet; Investigation, Alexandre M. Sebbenn  
550 and Celine Blanc-Jolivet; Resources, Bernd Degen; Writing – Original Draft  
551 Preparation, Lorena F.M. Capo and Alexandre M. Sebbenn; Writing – Review  
552 & Editing, Lorena F.M. Capo, Alexandre M. Sebbenn and Niklas Tysklind; Su-  
553 pervision, Bernd Degen; Project Administration, Bernd Degen, Celine Blanc-  
554 Jolivet; Sampling, **Barbara R.V. Meyer-Sand**, Kathelyn Paredes-Villanueva,  
555 Euridice N. Honorio-Conorado, Carmen R. García-Dávila, Niklas Tysklind and  
556 Valérie Troispoux.

**Funding:** This research was supported by the German Federal Ministry of Food and Agriculture in the frame of the “Large scale project on genetic timber verification”. Genotyping was done at the Genomic and Sequencing Facility of Bordeaux (grants from the Conseil Regional d’Aquitaine n° 20030304002FA and 20040305003FA, the European Union, FEDER n° 2003227 and Investissements d’avenir, n° ANR-10-EQPX-16-01. [Support for the](#)

[UMR EcoFoG was provided by CEBA: ANR-10-LABX-25-01.](#) Please add: “This research received no external funding” or “This research was funded by NAME OF FUNDER, grant number XXX” and “The APC was funded by XXX”. Check carefully that the details given are accurate and use the standard spelling of funding agency names at <https://search.crossref.org/funding>. Any errors may affect your future funding.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author due to [prevent data from being used by private companies to make a profit](#).

**Acknowledgments:** Permits ICMBIO n° 47960, 49369, registration CGen n° A16E4DA (Brazil), MAYA/VMABCCGDF/DGBAP/MEG n° 0280/2016 (Bolivia), R.D. n° 001-2016-SERNANP-DGANP, R.D. n° 001A-2015-SERFOR-DGG-SPFFS-DGSPF, and Contract n° 001-2016-SERFOR-DGGSPFFS-DGSPF (Peru) were granted for the collection, transport, and genetic analyses of samples. We are grateful to colleagues from IIAP, INRAE French Guiana, CIRAD Paracou Station, Office National des Forêts, and Thünen Institute for their assistance during field work and laboratory work. In Bolivia, samples were identified by Museo de Historia Natural Noel Kempff Mercado.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Degen, D.; Ward, S.E.; Lemes, M.R.; Navarro, C.; Cavers, S.; Sebbenn, A.M. Verifying the geographic origin of mahogany (*Swietenia macrophylla* King) with DNA-fingerprints. *Forensic Sci. Int-Gen.* 2013, 7, 55–62. <https://doi.org/10.1016/j.fsigen.2012.06.003>
2. Lowe, A.J.; Dormontt, E.E.; Bowie, M.J.; Degen, B.; Gardner, S.; Thomas, D.; Clarke, C.; Rimbawanto, A.; Wiedenhoef, A.; Yin, Y. Opportunities for improved transparency in the timber trade through scientific verification. *BioScience* 2016, 66, 990–998. <https://doi.org/10.1093/biosci/biw129>
3. CITES, Comércio Internacional das Espécies da Flora e Fauna Selvagens em Perigo de Extinção, 2021. Disponível em: < <https://cites.org/esp/app/appendices.php>>. Acesso em 1 de junho de 2023.
4. Lescuyer, G.; Ndotit, S.; Ndong, L.B.B.; Tsanga, R.; Cerutti, P.O. Policy options for improved integration of domestic timber markets under the voluntary partnership agreement (VPA) regime in Gabon. Center for International Forestry Research (CIFOR), 2014.
5. Federal Police Brazil, GOV, 2022. Disponível em:< <https://www.gov.br/pf/pt-br/search?SearchableText=madeira%20ilegal%20extra%C3%ADda%20da%20Amaz%C3%B4nia>>. Acesso em: 17 de maio de 2022.
6. Farias, Elaíze. 2019. “Amazônia em Chamas: 90% da madeira exportada são ilegais, diz Polícia Federal.” *Amazônia Real*, September 13. Accessed July 31, 2023. <https://amazoniareal.com.br/amazonia-em-chamas-90-da-madeira-exportada-sao-ilegais-diz-policia-federal/67>
7. Perez Coelho, M.; Sanz, J.; Cabezudo, M. Analysis of volatile components of oak wood by solvent extraction and direct thermal desorption-gas chromatography-mass spectrometry. *J. Chromat.* 1997, 778, 427–434. [https://doi.org/10.1016/S0021-9673\(97\)00286-0](https://doi.org/10.1016/S0021-9673(97)00286-0)

- 605 8. Paredes-Villanueva, K.; Espinoza, E.; Ottenburghs, J.; Sterken, M.G.; Bongers, F.; Zuidema, P.A. Chemical  
606 differentiation of Bolivian *Cedrela* species as a tool to trace illegal timber trade. *Forestry: An Internat. J.*  
607 *For. Resear.* 2018, 91, 603–613. <https://doi.org/10.1093/forestry/cpy019>
- 608 9. Deklerck, V.; Finch, K.; Gasson, P.; Van den Bulcke, J.; Van Acker, J.; Beeckman, H.; Espinoza, E. Compar-  
609 ison of species classification models of mass spectrometry data: Kernel discriminant analysis vs random  
610 forest; A case study of Afrormosia (*Pericopsis elata* (Harms) Meeuwen). *Rapid Commun. Mass Spectrom.*  
611 2017, 1582–1588. <https://doi.org/10.1002/rcm.7939>
- 612 10. Kagawa, A.; Leavitt, S.W. Stable carbon isotopes of tree rings as a tool to pinpoint the geographic origin  
613 of timber. *J. Wood Sci.* 2010, 56, 3, 175–183. <https://doi.org/10.1007/s10086-009-1085-6>
- 614 11. Bergo, M.C.J.; Pastore, T.C.M.; Coradin, V.T.R.; Wiedenhoeft, A.C.; Braga, J.W.B. NIRS identification of  
615 *Swietenia macrophylla* is robust across specimens from 27 countries. *IAWA J.* 2016, 37, 3, 420–430.  
616 <http://dx.doi.org/10.1163/22941932-20160144>
- 617 12. Gasson, P.; Baas, P.; Wheeler, E. Wood anatomy of Cites-listed tree species. *IAWA J.* 2011, 32, 2, 155–198.  
618 <http://dx.doi.org/10.1163/22941932-90000050>
- 619 13. Moya, R.; Wiemann, M.C.; Olivares, C. Identification of endangered or threatened Costa Rican tree species  
620 by wood anatomy and fluorescence activity. *Rev. Biol. Trop.* 2013, 61, 1113–1156.
- 621 14. Tnah, L.H.; Lee, S.L.; Ng, K.K.S.; Tani, N.; Bhassu, S.; Othman, R.Y. Geographical traceability of an im-  
622 portant tropical timber (*Neobalanocarpus heimii*) inferred from chloroplast DNA. *For. Ecol. Manage.* 2009,  
623 258, 1918–1923. <https://doi.org/10.1016/j.foreco.2009.07.029>
- 624 15. Jolivet, C.; Degen, D. Use of DNA fingerprints to control the origin of sapelli timber (*Entandrophragma*  
625 *cylindricum*) at the forest concession level in Cameroon. *Forensic Sci. Int-Gen.* 2012, 6, 487–493.  
626 <https://doi.org/10.1016/j.fsigen.2011.11.002>
- 627 16. Dormontt, E.E.; Boner, M.; Braun, B.; Breulmann, B.; Degen, B.; Espinoza, E.; Gardner, S.; Guillery, P.;  
628 Hermanson, J.C.; Koch, G.; Lee, S.L.; Kanashiro, M.; Rimbawanto, A.; Thomas, D.; Wiedenhoeft, A.C.; Yin,  
629 Y.; Zahnen, J.; Lowe, A.J. Forensic timber identification: It's time to integrate disciplines to combat illegal  
630 logging. *Biol. Conserv.* 2015, 191, 790–798. <https://doi.org/10.1016/j.biocon.2015.06.038>
- 631 17. Blanc-Jolivet, C.; Yanbaev, Y.; Kersten, B.; Degen, B. A set of SNP markers for timber tracking of *Larix* spp.  
632 in Europe and Russia. *Forestry: An. Int. J. For. Resear.* 2018, 91, 614–628. <https://doi.org/10.1093/forestry/cpy020>
- 633 18. Chaves, C.L.; Degen, B.; Pakull, B.; Mader, M.; Honorio, E.; Ruas, P.; Tysklind, N.; Sebbenn, A.M. As-  
634 ssuming the ability of chloroplast and nuclear DNA gene markers to verify the geographic origin of Jatoba  
635 (*Hymenaea courbaril* L.) timber. *J. Hered.* 2018, 109, 543–552. <https://doi.org/10.1093/jhered/esy017>
- 636 19. Coronado, E.N.H.; Blanc-Jolivet, C.; Mader, M.; García-Dávila, C.R.; Gomero, D.A.; del Castillo, D.T.;  
637 Llampazo, G.F.; Pizango, G.H.; Sebbenn, A.M.; Meyer-Sand, B.R.V.; Paredes-Villanueva, K.; Tysklind, N.;  
638 Troispoux, V.; Massot, M.; Carvalho, C.; Lima, H.C.; Cardoso, D.; Degen, D. SNP markers as a successful  
639 molecular tool for assessing species identity and geographic origin of trees in the economically important  
640 South American legume Genus *Dipteryx*. *J. Hered.* 2020, 111, 4, 346–356.  
641 <https://doi.org/10.1093/jhered/esaa011>
- 642 20. Vlam, M.; de Groot, G.A.; Boom, A.; Copini, P.; Laros, I.; Veldhuijzen, K.; Zakamdi, D.; Zuidema, P.A.  
643 Developing forensic tools for an African timber: Regional origin is revealed by genetic characteristics, but  
644 not by isotopic signature. *Biol. Conserv.* 2018, 220, 262–271. <https://doi.org/10.1016/j.biocon.2018.01.031>
- 645 21. Abeele, S.V.; Hardy, O.J.; Beeckman, H.; Ilondea, B.A.; Janssens, S.B. Genetic markers for species conser-  
646 vation and timber tracking: Development of microsatellite primers for the Tropical African tree species  
647 *Prioria balsamifera* and *Prioria oxyphylla*. *Forests* 2019, 10, 1037. <https://doi.org/10.3390/f10111037>
- 648 22. Loureiro, A.A.; Silva, M.F.; Alencar, J.C. Essências madeireiras da Amazonia. INPA, 1979, v.1.
- 649 23. Sampaio, P.T.B.; Barbosa, A.P.; Fernandes, N.P. Ensaio de espaçamento com caroba - *Jacaranda copaia*  
650 (AUBL.) D. Don. Bignoniaceae. *Acta Amaz.* 1989, 9, 383–383. <https://doi.org/10.1590/1809-43921989191389>
- 651 24. Gentry, A.H. Bignoniaceae – Part II (Tribe Tecomeae). *Flora Neotropica* 1992, 25, 2, 1–370.  
652 <https://www.jstor.org/stable/i400045>
- 653

- 654 25. Maues, M.M.; Oliveira, P.E.A.M.; Kanashiro, M. Pollination biology in *Jacaranda copaia* (Aubl.) D. Don.  
655 (Bignoniaceae) at the “Floresta Nacional do Tapajós”, Central Amazon. Rev. Brasil. Bot. 2008, 31, 517–527.  
656 <https://doi.org/10.1590/S0100-84042008000300015>
- 657 26. James, T.; Veges, S.; Aldrich, P.; Hamrick, J.L. Mating systems of three tropical dry Forest tree species.  
658 Biotropica 1998, 30, 587–594. <https://doi.org/10.1111/j.1744-7429.1998.tb00098.x>
- 659 27. Vinson, C.C.; Kanashiro, M.; Harris, S.A.; Boshier, D.H. Impacts of selective logging on inbreeding and  
660 gene flow in two Amazonian timber species with contrasting ecological and reproductive characteristics.  
661 Mol. Ecol. 2015, 24, 38–53. <https://doi.org/10.1111/mec.13002>
- 662 28. Dumolin, S.; Demesure, B.; Petit, R.J. Inheritance of chloroplast and mitochondrial genomes in pediculate  
663 oak investigated with an efficient PCR method. Theor. Appl. Genet. 1995, 91, 1253–1256.  
664 <https://doi.org/10.1007/BF00220937>
- 665 29. Sebbenn, A.M.; Blanc-Jolivet, C.; Mader, M. Meyer-Sand, B.R.V.; Paredes-Villanueva, K.; Coronado,  
666 E.N.H.; Garcia-Davila, C.; Tysklind, N.; Troispoux, V.; Delcamp, A.; Degen, D. Nuclear and plastidial SNP  
667 and INDEL markers for genetic tracking studies of *Jacaranda copaia*. Conserv. Genet. Resour. 2019, 11, 117,  
668 341–343. <https://doi.org/10.1007/s12686-019-01097-9>
- 669 30. Degen, B.; Blanc-Jolivet, C.; Stierand, K.; Gillet, E. A nearest neighbour approach by genetic distance to  
670 the assignment of individual trees to geographic origin. Forensic Sci. Int. Genet. 2017, 27, 132–141.  
671 <https://doi.org/10.1016/j.fsigen.2016.12.011>
- 672 31. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software  
673 STRUCTURE: a simulation study. Mol. Ecol. 2005, 14, 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- 674 32. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype  
675 data. Genetics 2000, 155, 945–959. <https://doi.org/10.1093/genetics/155.2.945>
- 676 33. Kopelman, N.M.; Mayzel, J.; Jakobsson, M.; Rosenberg, N.A.; Mayrose, I. Clumpak: a program for identi-  
677 fying clustering modes and packaging population structure inferences across K. Mol. Ecol. Resour. 2015,  
678 15, 1179–1191. <https://doi.org/10.1111/1755-0998.12387>
- 679 34. Loiselle, B.A.; Sork, V.L.; Nason, J.; Graham, C. Spatial genetic structure of a tropical understory shrub,  
680 *Psychotria officinalis* (Rubiaceae). Am. J. Bot. 1995, 82, 1420–1425. <https://doi.org/10.1002/j.1537-2197.1995.tb12679.x>
- 681 35. Hardy, O.J.; Vekemans, X. SPAGeDI: A versatile computer program to analyses spatial genetic structure  
682 at the individual or population levels. Mol. Ecol. Notes 2002, 2, 618–620. <https://doi.org/10.1046/j.1471-8278>
- 683 36. Rannala, B.; Mountain, J.L. Detecting immigration by using multilocus genotypes. Proc. Natl. Acad. Sci.  
684 U.S.A. 1997, 94, 9197–9201. <https://doi.org/10.1073/pnas.94.17.9197>
- 685 37. Piry, S.; Alapetite, A.; Cornuet, J.M.; Paetkau, D.; Baudouin, L.; Estoup, A. GENECLASS2: a software for  
686 genetic assignment and first-generation migrant detection. J. Hered. 2004, 95, 536–539.  
687 <https://doi.org/10.1093/jhered/esh074>
- 688 38. Efron, B. Estimating the error rate of a prediction rule – improvement on cross validation. J. Am. Stat.  
689 Assoc. 1983, 78, 316–331. <https://doi.org/10.2307/2288636>
- 690 39. Cornuet, J.M.; Piry, S.; Luikart, G.; Estoup, A.; Solignac, M. New methods employing multilocus geno-  
691 types to select or exclude populations as origins of individuals. Genetics 1999, 153, 1989–2000.  
692 <https://doi.org/10.1093/genetics/153.4.1989>
- 693 40. Tysklind, N.; Blanc-Jolivet, C.; Mader, M.; Meyer-Sand, B.R.V.; Paredes-Villanueva, K.; Coronado, E.N.H.;  
694 Garcia-Davila, C.; Sebbenn, A.M.; Caron, K.; Troispoux, V.; Guichoux, E.; Degen, B. Development of nu-  
695 clear and plastid SNP and INDEL markers for population genetic studies and timber traceability of *Carapa*  
696 species. Conserv. Genet. Resour. 2019, 11, 337–339. <https://doi.org/10.1007/s12686-019-01090-2>
- 697 41. Meyer-Sand, B.R.V.; Blanc-Jolivet, C.; Mader, M.; Paredes-Villanueva, K.; Tysklind, N.; Sebbenn, A.M.;  
698 Guichoux, E.; Degen, D. Development of a set of SNP markers for population genetics studies of Ipe  
699 (*Handroanthus* sp.), a valuable tree genus from Latin America. Conserv. Genet. Resour. 2018, 10, 779–781.  
700 <https://doi.org/10.1007/s12686-017-0928-5>
- 701  
702  
703  
704

- 705 42. Wahlund, S. Composition of populations and correlation appearances viewed in relation to the studies of  
706 inheritance. *Hereditas* 1928, 11, 65–106.
- 707 43. Ogden, R.; Linacre, A. Wildlife forensic science: A review of genetic geographic origin assignment. *Forensic*  
708 *Sci. Int. Genet.* 2015, 18, 152–159. <https://doi.org/10.1016/j.fsigen.2015.02.008>
- 709 44. Scotti-Saintagne, C.; Dick, C.W.; Caron, H.; Vendramin, G.G.; Troispoux, V.; Sire, P.; Casalis, M.; Buonam-  
710 ici, A.; Valencia, R.; Lemes, M.R.; Gribel, R.; Scotti, I. Amazon diversification and cross-Andean dispersal  
711 of the widespread Neotropical tree species *Jacaranda copaia* (Bignoniaceae). *J. Biogeogr.* 2013b, 40, 707–719.  
712 <https://doi.org/10.1111/j.1365-2699.2012.02797.x>
- 713 45. Cornuet, J.M.; Piry, S.; Luikart, G.; Estoup, A.; Solignac, M. New methods employing multilocus geno-  
714 types to select or exclude populations as origins of individuals. *Genetics* 1999, 153, 1989–2000.  
715 <http://dx.doi.org/10.1093/genetics/153.4.1989>
- 716 46. Ng, K.K.; Lee, S.L.; Tnah, L.H.; Nurul-Farhanah, Z.; Ng, C.H.; Lee, C.T.; Tani, N.; Diway, B.; Lai, P.S.;  
717 Khoo, E. Forensic timber identification: a case study of a cites listed species, *Gonyostylus bancanus* (Thyme-  
718 laeaceae). *Forensic Sci. Int. Genet.* 2016, 23, 197–209. <https://doi.org/10.1016/j.fsigen.2016.05.002>
- 719 47. Jardine, D.I.; Blanc-Jolivet, C.; Dixon, R.R.M.; Dormontt, E.E.; Dunker, B.; Gerlach, J.; Kersten, B.; Van Dijk,  
720 K.J.; Degen, B.; Lowe, A.J. Development of SNP markers for Ayous (*Triplochiton scleroxylon* K. Schum) an  
721 economically important tree species from tropical West and Central Africa. *Conserv. Genet. Resour.* 2016,  
722 8, 129–139. <https://doi.org/10.1007/s12686-016-0529-8>
- 723 48. Pakull, B.; Mader, M.; Kersten, B.; Ekue, M.R.M.; Dipelet, U.G.B.; Paulini, M.; Bouda, Z.H.N.; Degen, B.  
724 Development of nuclear, chloroplast and mitochondrial SNP markers for *Khaya* sp. *Conserv. Genet. Re-*  
725 *sour.* 2016, 8, 283–297. <https://doi.org/10.1007/s12686-016-0557-4>
- 726 49. Schroeder, H.; Cronn, R.; Yanbaev, Y.; Jennings, T.; Mader, M.; Degen, B.; Kersten, B. Development of  
727 molecular markers for determining continental origin of wood from white oaks (*Quercus* L. sect. *Quercus*).  
728 *PloSOone* 2016, 11, e0158221. <http://dx.doi.org/10.1101/038562>
- 729 50. Paredes-Villanueva, K.; de Groot, G.A.; Laros, I.; Bovenschen, J.; Bongers, F.; Zuidema, P.A. Genetic dif-  
730 ferences among *Cedrela odorata* sites in Bolivia provide limited potential for fine-scale timber tracing. *Tree*  
731 *Genet. Genome.* 2019, 15, 33. <https://doi.org/10.1007/s11295-019-1339-4>
- 732 51. Pakull, B.; Schindler, L.; Mader, M.; Kerten, B.; Blanc-Jolivet, C.; Paulini, M.; Lemes, M.R.; Ward, S.E.;  
733 Navarro, C.M.; Cavers, S.; Sebbenn, A.M.; di Dio, O.; Guichoux, E.; Degen, D. Development of nuclear  
734 SNP markers for Mahogany (*Swietenia* spp.). *Conserv. Genet. Resour.* 2020, 12, 585–587.  
735 <https://doi.org/10.1007/s12686-020-01162-8v>
- 736