ITS and *trnH-psbA* as Efficient DNA Barcodes to Identify Threatened Commercial Woody Angiosperms from Southern Brazilian Atlantic Rainforests

Mônica Bolson[1☉], Eric de Camargo Smidt[1☉], Marcelo Leandro Brotto[2], Viviane Silva-Pereira[1☉]*

1 Departamento de Botânica, Setor de Ciências Biológicas, Universidade Federal do Paraná, Curitiba, Paraná, Brazil, 2 Museu Botânico Municipal de Curitiba, Curitiba, Paraná, Brazil

☉ These authors contributed equally to this work.
* visilvapereira@gmail.com

Abstract

The *Araucaria* Forests in southern Brazil are part of the Atlantic Rainforest, a key hotspot for global biodiversity. This habitat has experienced extensive losses of vegetation cover due to commercial logging and the intense use of wood resources for construction and furniture manufacturing. The absence of precise taxonomic tools for identifying *Araucaria* Forest tree species motivated us to test the ability of DNA barcoding to distinguish species exploited for wood resources and its suitability for use as an alternative testing technique for the inspection of illegal timber shipments. We tested three cpDNA regions (*matK*, *trnH-psbA*, and *rbcL*) and nrITS according to criteria determined by The Consortium for the Barcode of Life (CBOL). The efficiency of each marker and selected marker combinations were evaluated for 30 commercially valuable woody species in multiple populations, with a special focus on Lauraceae species. Inter- and intraspecific distances, species discrimination rates, and ability to recover species-specific clusters were evaluated. Among the regions and different combinations, ITS was the most efficient for identifying species based on the 'best close match' test; similarly, the *trnH-psbA* + ITS combination also demonstrated satisfactory results. When combining *trnH-psbA* + ITS, Maximum Likelihood analysis demonstrated a more resolved topology for internal branches, with 91% of species-specific clusters. DNA barcoding was found to be a practical and rapid method for identifying major threatened woody angiosperms from *Araucaria* Forests such as Lauraceae species, presenting a high confidence for recognizing members of *Ocotea*. These molecular tools can assist in screening those botanical families that are most targeted by the timber industry in southern Brazil and detecting certain species protected by Brazilian legislation and could be a useful tool for monitoring wood exploitation.

Introduction

The Brazilian Atlantic Rainforest is one of 34 global biodiversity hotspots and is among the most seriously threatened biomes on the planet. Of the original forest, only 7 to 8% currently remains as small and biologically impoverished remnants that are exposed to strong anthropogenic pressures along the Brazilian coast [1]. Some important compilations of inventories of the floristic diversity have recorded more than 16,000 species of vascular plants in Atlantic Rainforest, with more them a half being endemic to the biome [1], List of Species of Brazilian Flora (http://floradobrasil.jbrj.gov.br/). One of its forest formations, the Mixed Ombrophilous Forest (MOF), also known as the *Araucaria* Forest, is found in the southern region of Brazil; however, well-preserved fragments of MOF occupy only 0.8% of their original range [2, 3]. Although much of this drastic reduction is associated with the expansion of agriculture and livestock, a major contribution to the intensive exploitation is the fact that MOF has been the main source of wood for furniture making and civil construction in the southern states of Brazil for more than 150 years [3, 4].

Based on lists of species published in 162 floristic and phytosociological studies, the current diversity of these forests in Paraná State is estimated to be 244 shrub-arboreal species [5]. Among the arboreal angiosperms associated with *Araucaria angustifolia* (Bertol.) Kuntze, which characterizes this forest formation, the Myrtaceae and Lauraceae families are the richest in species number, with 18 and 14 species, respectively, families that include some species complexes with problematic taxonomic delimitations and great variation regarding habits and forms [5]. In Lauraceae in particular, the precise identification of different *Ocotea* species has remained a difficult task for non-specialists, as the vegetative morphologies of these plants are very similar, their flowers are very small, and species complex delimitations are poorly defined [6, 7, 8].

The Lauraceae species from MOF, including *Ocotea porosa* (Nees & Mart.) Barroso, popularly known as "imbuia", have historically been used in civil construction and for manufacturing furniture. Indeed, four species of *Ocotea* Aubl. and three species of *Nectandra* Rol. ex Rottb are of great interest to the timber industry and are intensely exploited and threatened [5]. The Publication of Resolution n° 278 in May 24, 2001, by the National Environmental Council (CONAMA) of Brazil, suspended the extraction or use of threatened species in the Atlantic Rainforest biome and prohibited any commercial use of a extensive list of approx. 420 native Brazilian forest species [9, 10]. Among the commercially used MOF wood species included on this list are *Ocotea catharinensis* Mez, *O. odorifera* (Vellozo) Rohwe, *O. porosa* and *Cedrella fissilis* Vel [1, 10]. Despite laws restricting the exploitation of these forest resources, clandestine logging continues almost unrestrained, and many other types of natural resources are still being exploited [9], with special focus on the internal market. Due to this global problem, since 2013, new European Union rules limit the import of illegally harvested timber and products derived from illegal timber on the EU market (http://www.euflegt.efi.int/eu-timber-regulation), including some native Brazilian species.

However, timber identification of tree species is a great challenge for monitoring and controlling the wood trade and for inspection by authorities to prevent their commercial use and illegal export. Although the most popular and routine method for identifying species is based on wood anatomy, taxonomically related species often have similar wood structure, which makes this strategy rather inefficient. Moreover, this difficulty is amplified when considering a genus or species complex with unsolved taxonomy, for which the species limits are not clearly defined. Because of their biological and economic relevance and the critical situation for the most important woody species and related taxa in MOF, these plants are ideal candidates for identification using DNA barcoding techniques. Because some taxa are protected by Brazilian Legislation [10] and listed in the Red List of Brazilian Flora [11], the feasibility of assessing

DNA from dry wood could allow a broader application of DNA barcoding for wood certification or forensics tracking [12].

The molecular technique of DNA barcoding was developed as a rapid and practical way of identifying organisms at the species level using specific DNA sequences [13]. This molecular tool can be used as a forensic resource for the recognition of an unidentified organism from small amounts of processed tissue via the sequencing of a standardized genomic region. These sequences can then be compared to a set of sequences produced from previously well-identified specimens that can be easily accessed for taxonomic confirmation [14]. The molecular identification of animal species through barcoding is already well established and utilizes a portion of the gene sequence of mitochondrial cytochrome oxidase subunit I (COI) as the universal barcode standard [13]. For plants, however, the establishment of a single DNA region as a universal barcode does not appear to be a simple or realistic goal [15, 16].

Different from the situation in animals, the COI marker cannot be used in plants because of the overall lower mutation rate of plant mtDNA [17]. According to previous studies, a combination of plastid markers should be used because of the general inefficiency shown by any single marker alone [18]. In terms of biological and evolutionary features, the greatest difficulties involve the differences in evolutionary rates between taxa [19], the high frequency of interspecific hybridization events that are widespread in angiosperms [20], the incipient diversification processes in species complexes, and the poorly established interspecific limits of endemic and highly diverse plant groups [21, 22]. Finally, another factor that limits the establishment of a universal DNA barcode in plants, particularly if it is intended for applications to tree communities, is the slow evolutionary rates of perennial plants [17, 23]. Initially, the *matK* and *rbcL* regions of the plastid genome were considered promising regions for barcoding [18, 24, 25, 26]. However, low molecular resolution at the species level and problems with *matK* amplification prompted CBOL to recommend the inclusion of *trnH-psbA* [27] and, later, nrITS for different groups of angiosperms [28, 29, 30, 31].

Given the evident difficulty in determining universal regions for all terrestrial plant groups, Kress et al. [16] suggested that this molecular tool would have a greater potential if it was standardized and applied to specific or regional floristic groups. In such cases, the success of DNA barcoding could be guaranteed after determining the complementary regions (or combinations of regions) adequate for each situation. Indeed, a number of studies have demonstrated that barcodes can be satisfactorily applied to study the diversity [32] and evolution of ecological communities [33, 34], to characterize species in floristic surveys [35] and to discriminate taxa in families with complex taxonomy [36, 37, 38, 39]. Barcoding has also been useful as a forensic tool to determine the origin of commercialized natural products, such as medicinal plants [40, 41], teas [42], and foods [43]. Furthermore, the recent reports on the barcoding of commercially important species of trees [44, 45], forests communities [46, 47, 48] and CITES-protected timber species [49] have emphasized the feasibility of employing this approach as a routine method for identifying species of economic interest.

The efficiency of the barcoding approach was previously tested for Amazonian trees [48] and for Sapindaceae species from the Atlantic Rainforest [45], with both studies including species typically exploited for their wood. In these two cases, the technique appeared to contribute to the recognition of woody species, mainly those one with taxonomic confusion. Considering the economic importance of the arboreal flora of MOF and the continued illegal exploitation, barcoding the main woody tree species of these forests could be employed for controlling legal wood harvesting, detecting protected species in seized cargo, and certifying legally extracted species. Within this context, we expect this technique to be capable of correctly identifying at the species level approximately 30 important wood species distributed among 17 families. In particular, recognition of Lauraceae members at least at the genus level would be considered a

great achievement because it is the most diverse family in this area and is known to be taxonomically problematic. Such an endeavor would require the following: (*i*) the establishment of a regional and multipopulational genomic DNA bank of the most important woody species in MOF; (*ii*) the determination of which DNA fragments (or combination of fragments) are useful as barcodes for this group of plants; and (*iii*) the development of a high-quality reference database of DNA barcode sequences that is easily accessible to non-academic users.

This study, which is the first to test commercially viable woody species in the Atlantic Rainforest of southern Brazil, evaluated the utility of combinations of plastidial fragments of the *matK* and *rbcL* coding regions, the intergenic spacer *trnH-psbA*, and the nuclear internal transcribed spacer (ITS) to determine (*i*) the universality of the primers and to quantify their amplification and sequencing success rates, (*ii*) their patterns of intraspecific and interspecific variation, and (*iii*) their ability to correctly identify taxa of interest.

Materials and Methods

Plant selection and sampling

A list of species prepared by Isernhagen et al. [5] was used to identify and select the principal timber species from MOF utilized in the commercial wood industry [50]. Based on the results, thirty arboreal angiosperms species were sampled in the field. At least three individuals from different populations were collected in Paraná State, Brazil, resulting in 112 specimens (S1 Table). This research received permission for field collection in protected areas in Paraná State, issued by Chico Mendes Institute for Biodiversity Conservation (ICMBio, Ministério do Meio Ambiente), the agency responsible for regulating activities involving natural resources in Brazil.

Samples of all specimens were deposited in the herbarium of the Botany Department of the Federal University of Paraná (UPCB), a depository of genetic patrimony, with permanent license for the storage of biological and genetic samples from Brazilian flora (59/2004—CGEN Council for the Genetic Heritage Management, Ministério do Meio Ambiente). Specialists performed identifications by comparison with herbarium samples deposited in the UPCB and Municipal Botanical Museum of Curitiba (MBM) herbaria, which contains the most important reference collection of southern vegetation in Brazil. Duplicates were sent to national (HUEFS, RB, SP) and international (NY) herbaria. Leaf tissue samples from all specimens were stored either at -80°C or in silica gel.

DNA extraction, amplification, and sequencing

Total genomic DNA was extracted from fragments of fresh leaf tissue according to the 2X CTAB protocols [51]. Samples of total DNA were deposited in the Total DNA Bank (Laboratory of Systematic and Molecular Ecology of Plants–UFPR) and stored in ultrafreezers (-80°C).

The following fragments were used for barcodes: *matK*, with the primers *matK 3F_Kim f* and *matK 1R_Kim r* or *matK 5R* and *matK 2.1F* [15]; *rbcL*, with the primers *rbcLa_f* [52] and *rbcLa_r* [53]; *trnH-psbA*, with the primers *trnHf_05* [54] and *psbA3 f* [55]; and ITS, with the primers ITS92 and ITS75 [56] or ITS18_F [57] and ITS26_R [58].

Polymerase chain reactions (PCRs) were performed in 20 μL reaction mixtures containing 1 X buffer, 1.5 mM MgCl$_2$, 0.2 mM each dNTP, 0.8 μM each primer, 0.4 mg/ml BSA, 0.65 unit Taq DNA polymerase, and 20–50 ng genomic DNA; 1 M betaine and 2% DMSO were added to the ITS reaction. The amplification program for each region is provided in S2 Table. The PCR products were purified with PEG 20%, and the sequencing reactions were performed by Macrogen Inc. (Seoul, South Korea—http://dna.macrogen.com) using the same primers as used for PCR (S2 Table).

Barcoding analyses

The forward and reverse sequences were edited, and consensus sequences were obtained using the Staden Package software [59]. Multiple sequence alignments were performed using Clustal W program [60] and MEGA6 software [61] with the standard parameters, and the matrices were confirmed manually. To facilitate the alignment of the matrices for the two spacers (*trnH-psbA* and ITS), the sequences were organized according to the taxa classification by APG III [62]. All of the sequences generated in this work were submitted to International Nucleotide Sequence Database Collaboration GenBank (Accessions, *rbcL*: KF561901—KF561971; *matK*: KF555384—KF555448; ITS: KF420934—KF421013; *trnH- psbA*: KF421014—KF421108 (S2 Table)).

Distance-based methods. The genetic variability of each marker was described by the median length (bp) and total alignment length (bp); the number of polymorphic sites and number of parsimony informative sites (PIC), both discounting gaps; the number of sites with gaps; nucleotide diversity (Pi); and the number of haplotypes in each region using DnaSP 5.10 [63].

The barcoding analyses were conducted separately for each region and combined in the following arrangement: (*i*) coding regions (*matK* + *rbcL*); (*ii*) non-coding regions (*trnH-psbA* + ITS); (*iii*) plastid genome regions (*matK* + *rbcL* + *trnH-psbA*); and (*iv*) the four regions combined (*matK* + *rbcL* + *trnH-psbA* + ITS). In the two latter combinations, only species that had sequences from at least two and three regions, respectively, were considered.

The presence of 'barcoding gaps' between the interspecific and intraspecific distances were evaluated using frequency histograms based on the uncorrected paired p-distances obtained using the TAXONDNA program and the 'best close match' test [64].

The 'best close match' test was used to determine the closest correspondence of a target sequence with all of the others in a set of aligned data, there by establishing a similarity limit based on the distribution of the frequencies of the intraspecific and interspecific distances. A limiting value was calculated separately for each matrix (all makers separately and in combination), and the relative frequencies of all of the intraspecific distances were calculated to determine a limiting value below 95% [64]. In such tests, sequences can be classified as follows: "correct", when the genetic distances between the target sequence and other sequences of the same species fit within 95% of the calculated limit; "ambiguous", when a sequence lies within this percentage but is similar to that of another species; "incorrect", when a target sequence does not encounter a sequence of the same species within the calculated limit; and "incongruent", when the sequence does not lie within the 95% range of the calculated limit [65].

Monophyly-based methods. The following different phylogenetic approaches were used to evaluate whether the samples from different individuals form a species-specific clusters: distance methods applying a paired p-distance with the Neighbor-joining algorithm (NJ) [66, 67]; Maximum Likelihood (ML) methods using an evolutionary model selected with Bayesian Information Criterion [61]; and Maximum Parsimony (MP), which uses heuristic searches with 2000 replicates and random additions retaining 20 trees per replication with TBR algorithm, followed by an analysis of TBR to explore all of the topologies obtained in the initial analysis, establishing a limit of 10,000 trees. A strict consensus tree was constructed from multiple equally parsimonious trees. Support analyses utilized 1000 bootstrap replicates (BP) [68] in the NJ and MP analyses. For computational reasons, only 100 bootstrap replicates were used for ML. The NJ and ML analyses were performed using MEGA6 software [61], and the MP analyses were performed using PAUP v.4.0b10a software [69] with Fitch parsimony [70]. The phylogenetic trees were rooted in Lauraceae, basal angiosperm family, in all of the analyses.

Table 1. Sample sizes, success rates of amplification and sequencing, and the molecular characteristics of the four markers evaluated for woody species from MOF.

Barcode	N of samples tested	N of samples amplified (%)	N of samples sequenced (%)	Median length (bp)	Alignment length (bp)	Number of polymorphic sites*	Number of parsimony informative sites (PIC)*	Number of sites with gaps	Nucleotide diversity (Pi)	N accessions (N species)	Number of haplotypes
ITS	112	103(91)	80(78)	656.4	850	179	156	544	0,207	80(29)	44
trnH-psbA	108	106(98)	89(84)	482.2	707	101	96	584	0,31	89(30)	44
matK	102	75(74)	65(86)	817.5	939	438	376	255	0,168	65(28)	52
rbcL	70	68(97)	68(100)	571.8	580	156	144	17	0,069	68(30)	33

*excluding those with gaps/missing data.

doi:10.1371/journal.pone.0143049.t001

Results

Amplification, sequencing success and makers features

A genomic DNA collection was assembled using 112 individuals from 30 different timber angiosperms species of high commercial value in the MOF. PCR was performed for all 112 accessions using ITS, 108 using *trnH-psbA*, 102 using *matK* and a lower sampling of 70 using *rbcL*. The amplification reactions were performed with high success (>90%) for *trnH-psbA*, *rbcL*, and ITS, whereas *matK* demonstrated lower reaction efficiency, with successful amplification of only 74% of the samples ([Table 1](#)). Multiple attempts per sample were required for *matK* for adjustment of the DNA template concentration and identifying ideal primer pairs, which demonstrated significant variability in amplification success among the different botanical families.

High-quality sequences were obtained for most of the amplified DNA samples. *rbcL* exhibited the highest efficiency, with 100% sequencing success and *matK* (86%) and *trnH-psbA* (84%) intermediate efficiency; in contrast, ITS showed the least efficiency, with 78% success ([Table 1](#)). The four DNA regions chosen were effectively amplified and sequenced for most of the species, generating a database of 302 sequences; however, it was not possible to obtain sequences for all of the DNA regions for all accessions ([S1 Table](#)). When we discarded the accessions that were not amplified and/or sequenced for two or more fragments, analyses were performed with a final matrix composed of 80 sequences of ITS, 89 of *trnH-psbA*, 65 of *matK* and 68 of *rbcL*, from 101 accessions ([Table 1](#)).

Based on multiple alignments of all sequences for each region, *matK* presented the largest length (939 bp) and *rbcL* the shortest length (580 bp) ([Table 1](#)). The number of polymorphic sites, discounting gaps, varied for individual regions: from 101 in *trnH-psbA* to 438 in *matK*. The largest number of parsimony informative characters was provided by *matK* (376), followed by ITS (156) and *rbcL* (144). *trnH-psbA* showed the highest number of gaps (584) and the highest nucleotide diversity (0.31), followed by ITS with 544 gaps and nucleotide diversity of 0.207. The greatest number of haplotypes was generated by *matK* (52), followed by ITS and *trnH-psbA* (44 each) ([Table 1](#)). The haplotypes provided by these three markers were species-specific, except for some Lauraceae species. *rbcL* and *trnH-psbA* provided only one haplotype for all *Nectandra* species and *matK* provided two, but one shared between the three species (*N. grandiflora*, *N. lanceolata*, *N. megapotamica*). For *Ocotea rbcL* provided four, *trnH-psbA* seven and *matK* eight haplotypes, being one or more haplotypes shared between four from five species analyzed, including those most important taxa (*O. porosa*, *O. puberula*, *O. odorifera* and *O. catharinensis*). ITS provided the most specific haplotypes, with only one haplotype shared between all accessions

of *Nectandra megapotamica* and *N. lanceolata*. All other species were characterized by one (*O. catharinensis*, *O. odorifera*, *O. puchella*) to three exclusive ITS haplotypes (*O. porosa*).

## Comparative performance of barcodes

Testing discrimination of species based on distance methods. As estimated from p-distances, variations in intra- and interspecific distances were observed among the four DNA regions. The presence or absence of a barcoding gap for individual or combined regions was represented by the frequency distribution of the genetic distance between intraspecific sequences and between interspecific sequences (Fig 1). When loci were individually analyzed, ITS presented the best barcode gap performance, with 71% of pairwise interspecific p-distances greater than 0.05 and 85.7% of pairwise intraspecific p-distances lower than 0.05. Although *trnH-psbA* provided the highest values of interspecific genetic distances, with a range varying from zero to a maximum of 0.19, this was combined with a high frequency of low distances at the interspecific level. Nonetheless, complete discontinuity was not found between the intra- and interspecific ranges for these two individual markers. The best barcode gap performance appears to have resulted from the combination of ITS + *trnH-psbA*. Conversely, *rbcL* showed the worst performance, with a total overlap of intra- and interspecific variation; the same unsatisfactory result was observed for *matK* and for the combination *matK* + *rbcL* (Fig 1).

The results of the 'best close match' test for individual regions revealed ITS as the best at species identification (83%), followed by *rbcL* (68%), *trnH-psbA* (65%) and *matK* (52%). The combined regions showed correct identification of nearly 76% for *trnH-psbA* + ITS, 71% for *matK* + *trnH-psbA* + *rbcL*, and 69% for *matK* + *trnH-psbA* + *rbcL* + ITS (Fig 2). In contrast, the *matK* + *rbcL* combination exhibited the worst performance for correct identification and the largest proportion of incorrect identifications. Ambiguous identification rates were relatively low (< 9%) for most of the individual regions and combinations.

Testing discrimination of species based on phylogenetic trees. Determination of the percentages of species-specific clusters was based on the reconstruction of phylogenetic trees according to three analyses (NJ, ML, and MP) for the four regions and four combinations tested. The phylogenetic trees generated from the three analyses differed in terms of their topologies and degrees of resolution, though there was little difference between the percentages of species-specific clusters obtained with ML and MP. The trees generated using NJ did not have good resolution and only demonstrated well-resolved topology for the ITS region and the combination *trnH-psbA* + ITS (77% support). The low resolution and low support values for the NJ trees prompted us to concentrate our efforts on ML and MP analyses (Table 2; S3 Table).

The topology generated by MP demonstrated the greatest support values for the ITS region, which was capable of recovering 96% of species-specific clusters with ≥70% of bootstrap support, followed by *trnH-psbA* (72%) and *rbcL* (68%). Conversely, ML revealed a greater resolution for *rbcL* (72%). As the worst performance, *matK* demonstrated the same percentages with both ML and NJ (50%) (Table 2). When compared with MP analysis, ML showed a more resolved topology for internal branches in the combined regions, *trnH-psbA* + ITS, with 91% of species-specific clusters and ≥70% of bootstrap support (Fig 3). Therefore, we choose to explore in deeper detail the *trnH-psbA* + ITS tree generated by ML; we present the ITS tree obtained by MP in the supporting information section (S1 Fig).

The *trnH-psbA* + ITS tree inferred by ML discriminated 100% of 16 sampled families (Fig 3). Among them, there was only one specimen analyzed for Rutaceae (*Zanthoxylum*), Lythraceae (*Lafoensia*), Aquifoliaceae (*Ilex*) and Proteaceae (*Roupala*). With more than one sample per species, families Cunoniaceae (*Lamanonia*), Rosaceae (*Prunus*), Malvaceae (*Luehea*), Erythroxylaceae (*Erythroxylum*), Lamiaceae (*Vitex*), Bignoniaceae (*Jacaranda*), Myrsynaceae
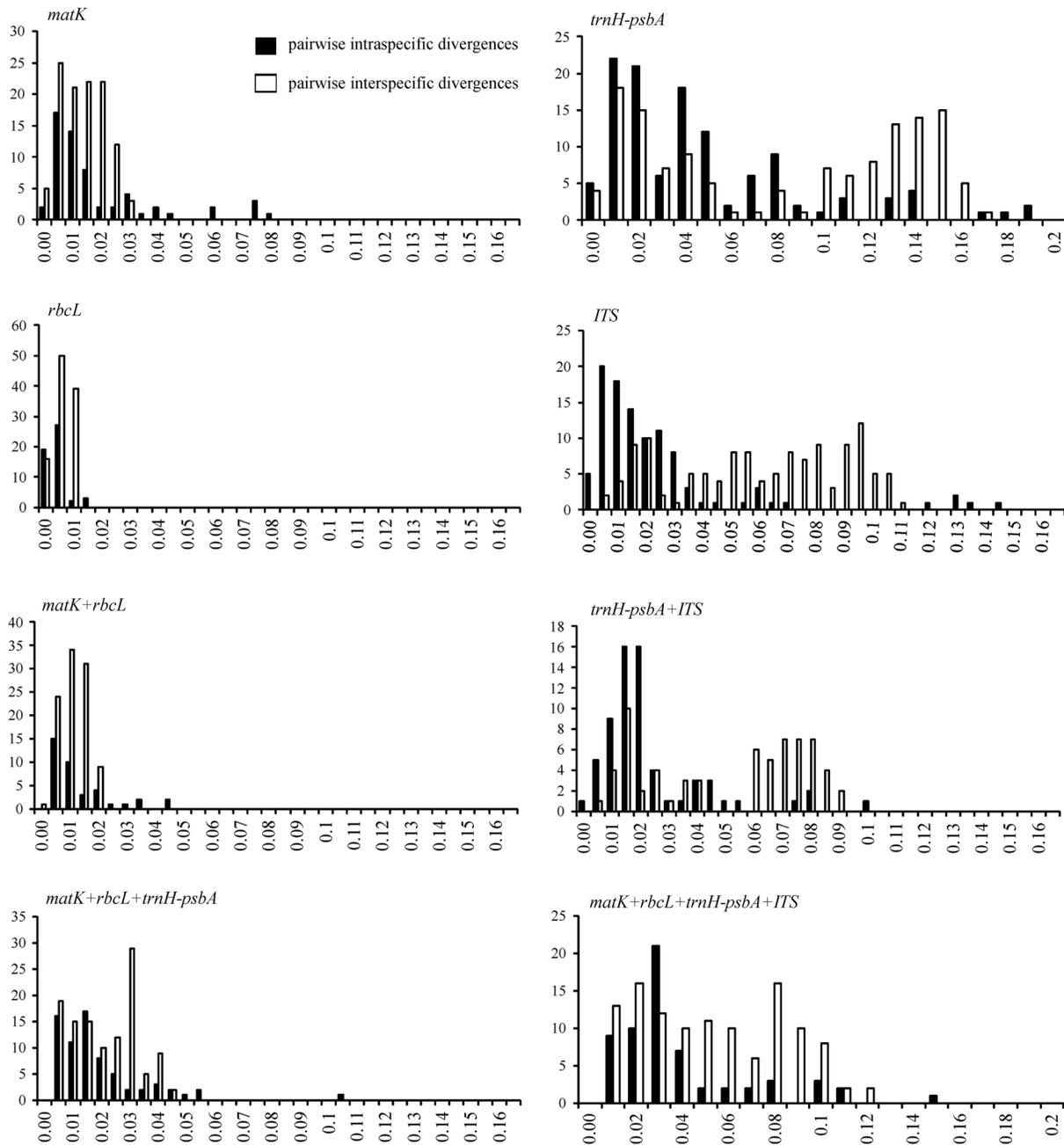
Fig 1. Histograms of the frequencies (*y*-axes) of pairwise intraspecific (black bars) and interspecific (white bars) divergences based on the p-distance (*x*-axes) for individual and combined ITS, *trnH-psbA*, *matK* and *rbcL* markers.

(*Myrsine*) and Fabaceae (*Mimosa*) were discriminated with high support (100% BP). Families with more than one species in different genera, such as Meliaceae (*Cedrela* and *Cabralea*), Anacardiaceae (*Lithraea* and *Schinus*), Sapindaceae (*Cupania* and *Matayba*), Myrtaceae (*Myrcia* and *Campomanesia*), were also discriminated with high support (99–100% BP). Only Lauraceae, with three genera (*Cryptocarya*, *Ocotea* and *Nectandra*) and nine species sampled, was not completely discriminated. The genus *Cryptocarya* was distinguished as a sister group of *Ocotea* and *Nectandra*, with 100% bootstrap, yet species of *Nectandra*, which were also
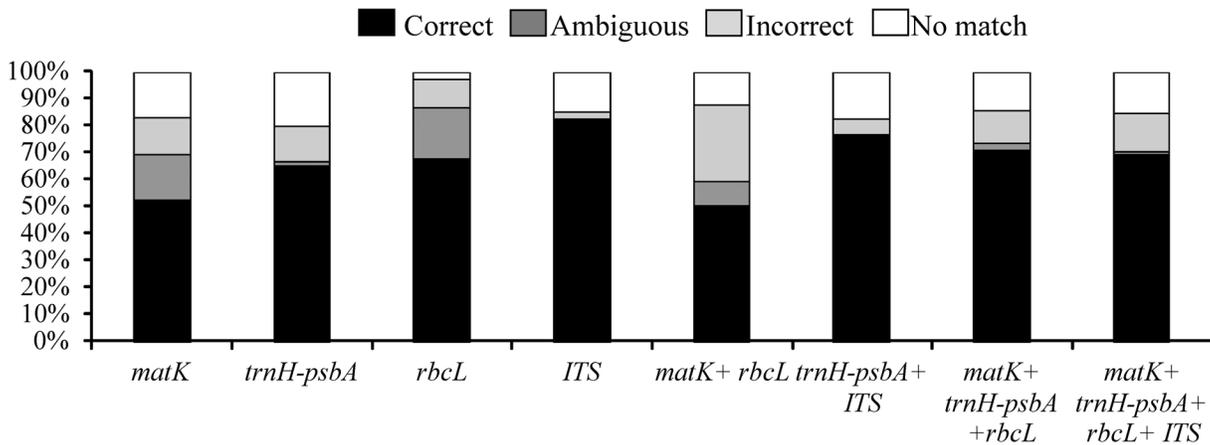
Fig 2. Success of species identification based on individual and combined analyses of ITS, *trnH-psbA*, *matK* and *rbcL* markers using the taxondna program and the 'best close match' method.

identified with 100% support, were included within the genus *Ocotea*. *Nectandra* presented an intermediate degree of discrimination (76–85% BP) between *N. megapotamica* and *N. grandiflora*, though samples of *N. lanceolata* were not grouped together and appeared as successive sister groups of the other species in the genus. Of the five sampled *Ocotea* species, *O. porosa* was discriminated with good support (87% BP), followed by *O. catharinensis* and *O. odorifera* with moderate support (73–87% BP). The multiple individuals of *Ocotea pulchella* (93% BP) and *O. puberula* (100% BP) were grouped as a highly supported species-specific clade (97% BP) together with *Nectandra* species, making *Ocotea* paraphyletic in our analyses.

Discussion

This is the first report of the efficacy of four barcode loci in identifying commercial wood tree species from MOF in the southern Brazilian Atlantic Rainforest. Combining the characteristics of molecular variation, intra- and interspecific divergence patterns, the 'best close match' test and the ability to recover species-specific clusters in phylogenetic trees, our results support the designation of ITS as the principal barcode locus for the forensic identification of the main explored timber species in MOF. The inclusion of *trnH-psbA* as an additional locus is recommended because, together with ITS, the barcode combines nuclear and plastidial genomic information and confers more reliability to the data set though not increasing the power of taxonomic identification of species.

Table 2. Percentage of species-specific clusters using different tree-based methods (NJ, MP, and ML) with ≥70% bootstrap support for single DNA regions and different combinations.

DNA region	*N* species	NJ	ML	MP
matK	20	50	50	45
trnH-psbA	25	56	64	72
rbcL	25	68	72	68
ITS	22	77	91	96
matK + *rbcL*	20	65	65	65
trnH-psbA + ITS	22	77	91	86
matK + *trnH-psbA* + *rbcL*	25	64	76	76
matK + *trnH-psbA* + *rbcL* + ITS	24	62.5	75	75
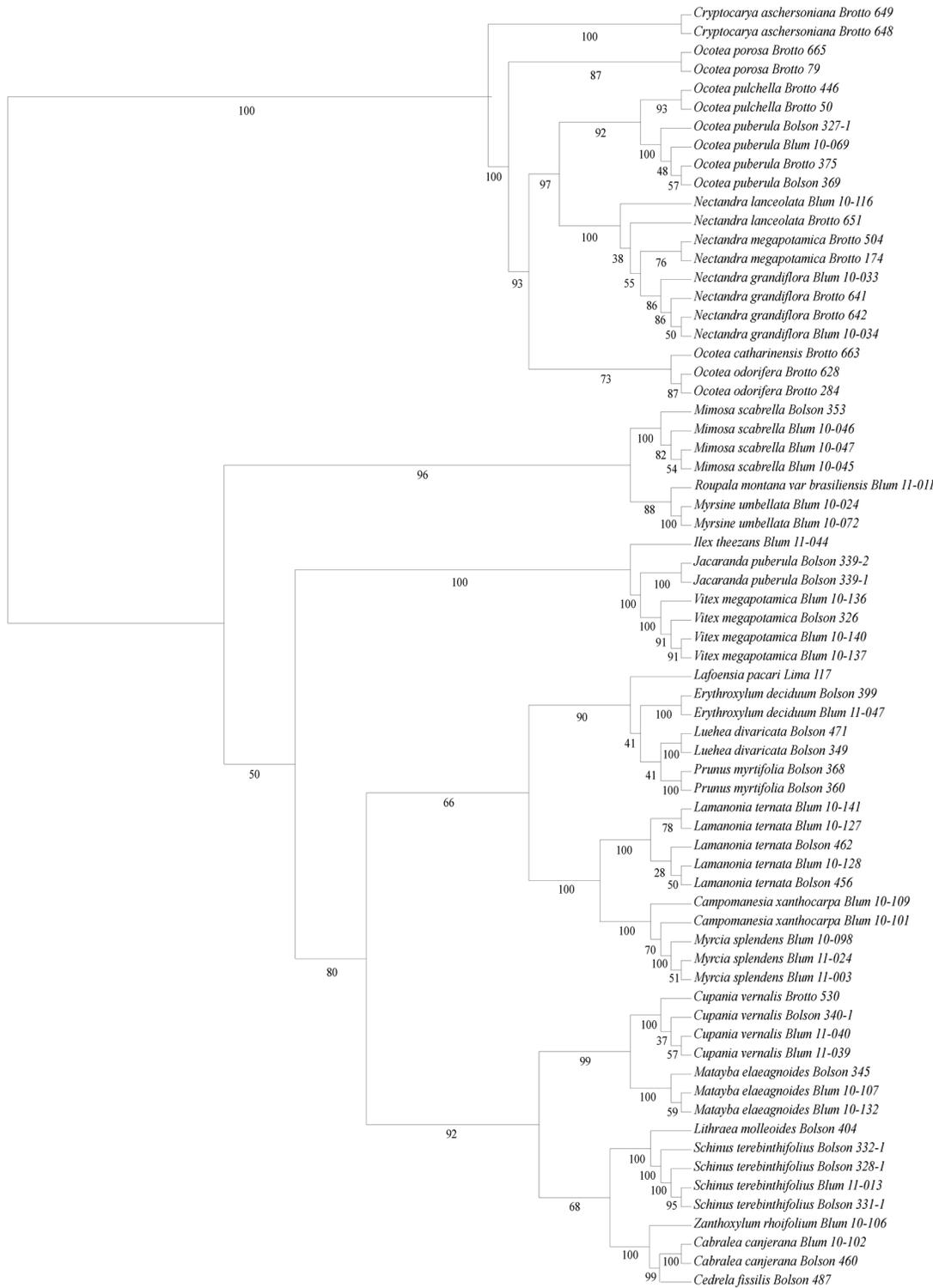
Fig 3. Cladogram for *trnH-psbA* + ITS generated by Maximum Likelihood. The bootstrap values ≥70% are shown under the branches. The species name is followed by the accession number of the specimen.

doi:10.1371/journal.pone.0143049.g003

Amplification, sequencing success and makers features

According to the CBOL plant working group, an ideal DNA barcode must combine conserved regions for universal primer design, show high rates of PCR amplification and sequencing, and present genetic variability that is high enough to distinguish sequences at the species level yet sufficiently conserved among individuals of the same species [13, 14, 18]. The first step in determining whether a DNA fragment can serve as a barcode is to evaluate its universal applicability by quantifying PCR and sequencing success. Indeed, the ease of obtaining a high-quality sequence determines whether DNA barcoding can become a viable technique for monitoring woody cargo and product inspection by regulatory agencies. In this regard, the *rbcL* plastid DNA region produced more successful reactions and specific amplification and resulted in simple and high-quality sequencing for MOF species, showing the best performance, with 100% of the samples successfully sequenced. Although previous studies have described the use of one or two primer pairs to sequence ≥90% of samples [28, 35], only one was necessary to amplify all of our accessions.

In our study, *trnH-psbA* was easily amplified from 98% of the samples and showed intermediate success in sequencing (84%). A similar pattern has been reported for other land plant groups, with *trnH-psbA* PCR amplification and sequencing rates high enough to be considered as barcode loci [16, 47]. Although this marker has been rejected as barcodes in some cases due to the length (> 1000 bp) and difficulty in bidirectional sequencing [18, 21], we were able to obtain satisfactory success rates of PCR amplification and sequencing as well as high-quality bidirectional sequences.

Among the four fragments, ITS presented a high success rate for amplification (91%) but the worst performance for sequencing (78%). Similar results were found in other studies in which the authors did not report difficulty in amplifying ITS but did encounter problems in sequencing this locus in Cycadales species and other gymnosperms [36], in tree species of India [47], and in tropical Amazonian trees [48]. This common pattern of ITS has been linked to secondary structure formation, multiple copies from paralogs, endophytic fungal contamination or wide variations in sequence lengths, leading to low-quality sequences [30]. In fact, this drawback has been used as an argument for considering this nuclear spacer as less acceptable as a standard DNA barcode in some research (e.g., [38]). Despite showing sequencing problems in some studies, this region has been suggested to be a good complement to core barcode fragments, particularly in cases in which the laboratory expense for producing sequences was not so costly or the very high species-discriminating rate compensated for the investment [31, 45, 47, 48]. Although the major problem related to ITS is the occurrence of divergent paralogs, resulting in nonspecific amplifications that are often identified by the presence of multiple bands on an agarose gel [30], this issue was not observed for the woody species from MOF. Considering that we did not encounter indications of this and the aforementioned problems associated with ITS, the low rate of sequencing is the only inconvenience that needs to be addressed.

Although *matK* exhibited the lowest amplification performance, this locus generated an intermediate sequencing percentage (86%). Neither of the two most frequently used *matK* primer pairs was efficient for *Erythroxylum deciduum* A. St.-Hil. (Erythroxylaceae) and *Vitex megapotamica* (Spreng.) Moldenke (Lamiaceae) DNA amplification. Many studies indicate that *matK* is a key marker for discriminating specific groups [71,72]; nevertheless, many authors have questioned its utility as a barcode due to poor amplification and sequencing performance and problems related to the universality of the primers [22, 36, 37, 39, 73, 74]. Based on the accumulated knowledge in the literature as well as our own experience, to obtain good results with *matK*, it is necessary to utilize different published primers according to the plant

family or to propose new primers for the samples of interest [75, 76]. For MOF woody species, multiple attempts per sample were required to adjust the concentrations of DNA template and to identify the ideal primers pairs for *matK*, a marker that demonstrated significant variability in amplification success among different families. Because of these characteristics, *matK* appears to be an impractical choice as a forensic barcode for plants that are phylogenetically unrelated, as is the case for the woody species from MOF.

In terms of molecular variation, *rbcL* was the most conserved sequence among the four regions, as indicated by the slight difference between the median sequence length and alignment length and the very low number of gaps in the final matrix. Although the *trnH-psbA* intergenic spacer demonstrated the lowest number of polymorphic sites, this locus presented the highest number of gaps due to the great variation in sequence length. As reported in many studies [18, 21, 22, 24, 77], this molecular pattern makes it somewhat difficult to align a matrix containing a large number of distinct taxonomic families, and this region has been not recommended for barcoding in those cases. Accordingly, we opted to carry out the alignment by grouping sequences according to orders and families following the APG III classification [62], resulting in a more reliable matrix.

When we compare the amplification and sequencing success of different DNA regions is observed that it is relatively harder to obtain sequences of ITS and *matK*; however, as discussed below, ITS was the best marker in identifying species from MOF. In this way, the low sequencing rate of ITS is a difficulty that we must deal with.

Comparative performance of markers

Testing discrimination of species based on distance methods. The combination of the *rbcL* and *matK* regions has been recommended as a universal barcode for land plants [18, 53] and when compared to individual regions, has been shown by a number of studies to result in better resolution for discriminating between species [22, 38, 73]. However, this combination was able to correctly identify only 50% of the commercially important woody species from the MOF, demonstrating less accuracy than the individual regions (*matK*, 52%; *rbcL*, 68%) and indicating it as an inappropriate barcode gap. This combination also demonstrated similar low efficiency (50%) in discriminating species of *Primula* L. [73] and Labiatae [39] within a strict taxonomic context, making it difficult to justify the use of core barcodes to distinguish between these closely related species. In a floristic approach, however, these markers were able to discriminate approximately 70% of a large list of different families from the flora of Wales [35]. Despite the low variation, *rbcL* was more efficient in distinguishing among woody MOF species using TAXONDNA when compared to *matK*, which showed a high number of variable sites and the worst performance in recognizing taxa. This result was unexpected because *matK* is known to be a rapidly evolving coding region, and this marker has been used as a barcode with an impressive ability to discriminate among angiosperm species [18, 24, 39, 71, 74]. Regardless, our results could be explained by a technical artifact due to the low amplification/sequencing success of *matK*, resulting in a smaller number of analyzed taxa, combined with the effect of sampling bias when studying a non-related set of families, which could be better recognized by a slower evolving marker such as *rbcL*.

ITS exhibited the best species discrimination performance according to the 'best close match' test using TAXONDNA; this single fragment was more efficient in identifying species than any other region or combination for the woody species studied. As has been observed in other studies, ITS provided a good alternative for the molecular identification of phylogenetically unrelated species and, at the same time, provided accurate identification of closely related species [31, 40, 41, 45, 78]. Although *trnH-psbA* did not show great performance alone, in

combination with ITS, it was capable of providing accurate identification for more than 75% of woody MOF species.

When all possible combinations were evaluated for all fragments explored in the woody species from MOF, we determined that the best combinations were *trnH-psbA* + ITS and *matK* + *trnH-psbA* + *rbcL*. These combinations did not generate ambiguous identifications, with only low levels of incorrect identification, recognizing >76% and >71% of the species, respectively. Roy et al. [22] similarly reported good results with *trnH-psbA* + ITS for identifying species of Indian *Berberis* L and in a large-scale biodiversity inventory of tropical tree species of India [47]. In spite of the significant sequencing investment required for the four regions, we conclude that none of the combinations was more efficient than ITS alone for the MOF woody species examined in this study.

Testing discrimination of species based on phylogenetic trees. One of the most efficient ways to determine the utility of a DNA region as barcode is to verify whether it can detect species-specific clusters of species from the same genus and family using a phylogenetic analysis. In general, commercially important woody species of different genera are grouped within their respective families; this also occurred with species for which only a single individual was analyzed. Most species with multiple individuals form species-specific clusters, including those of Lauraceae, with a high confidence to identify the most threatened *Ocotea* species.

In this regard, MP analysis produced cladograms with better resolution for ITS and with the greatest percentage of species-specific clusters, considering all clusters with more than 70% of bootstrap. Using the same criterion of support, the four different combinations tested presented better results by ML analysis. *trnH-psbA* + ITS and *matK* + *trnH-psbA* + *rbcL* demonstrated the best topology, with 91% and 76% of the species being grouped into species-specific clusters, respectively. Moreover, all of the cladograms from the three analyses of the combination *trnH-psbA* + ITS produced only species-specific clusters.

Recent reports of barcoding commercially important species of trees [44], forest communities [46, 47] and CITES-protected timber species [49] have been illuminated the feasibility of using this approach as a routine method for identifying species of economic interest. As an example, in studies carried out with Meliaceae, ITS provided the best resolution for interpreting species-specific clades in *Cedrela*, a broad genus known for its importance as a wood supply [49]. In addition to Lauraceae and Meliaceae, Sapindaceae from the Atlantic Rainforest [45] and Amazonian trees [48] are species of great economic importance, and many taxonomic challenges need to be solved. In these cases, barcoding techniques appear to contribute to the recognition of key woody species, mainly those with some taxonomic confusion. For species threatened by illegal exploitation, the use of ITS as a barcode affords the possibility of adding an auxiliary tool for monitoring the use of their wood worldwide [45, 48, 49].

In our study, the ITS region and the combination *trnH-psbA* + ITS efficiently distinguished with high support three genera of Lauraceae (*Cryptocarya*, *Ocotea* and *Nectandra*) from woody species from MOF. Even for closely related species of *Ocotea*, the ITS region and the combination *trnH-psbA* + ITS presented similar potential for distinguishing among species by detecting species-specific clusters for all taxa. In this way, our results also indicated barcoding as a promising tool to molecularly recognize the main woody *Ocotea* species that are actually explored and threatened using the fragments studied to date. The same DNA regions were recently tested in 42 species from 11 genera of Lauraceae, and the results showed that *trnH-psbA* could efficiently identify approximately 82% of the species [79], except for some *Ocotea* such as *O. odorifera* and *O. puberula* (Rich.) Nees, the only two commercially viable woody species from MOF examined.

Some studies have stressed the non-monophyly origin of some Lauraceae genera, a fact that is responsible for the lack of resolution with regard to the systematics of this family [6, 7]. A proper phylogenetic study on Brazilian representatives of this family would require increased

sampling and the inclusion of other genera, such as *Caryodaphnopsis*, *Clinostemon*, *Dicypellium* and *Mezilaurus*, which are not native to MOF or are not explored as sources of wood. For *Ocotea* in particular, which is composed of some species complexes with poorly defined limits among taxa, more works focusing on morphological and genetic variation at population level are required to address the relationship between closely related species. In this way, DNA barcoding could play an important role in the molecular recognition of important species in a way that is complementary to traditional methods of identifying timber, with no claim of resolving taxonomic and evolutionary questions. The feasibility of using dry tissue fragments for DNA extraction and genetic analysis has a major impact on barcoding as a forensic tool in the regulation of endangered tree species trade [12], including some Lauraceae species listed in the Red List of Brazilian Flora [11], or wood Eco-certification for product manufacturing.

Given the difficulty of classic taxonomic identification of very close *Ocotea* and *Nectandra* species found in MOF, it is expected that other species also suffer from collection pressure, even though they are not as useful to industry. Although only *Ocotea porosa*, *O. catharinensis* and *O. odorifera* are on the list of taxa banned for exploration, other *Ocotea* species could also be partially impacted, which may indicate targeted extraction. In this case, we cannot assert that the observed excellent resolution of certain barcodes for detecting species-specific clusters will be maintained if the species list is extended to all *Ocotea* or *Nectandra* species found in southern Brazil. Nonetheless, although this tool may not accurately identify all species, recognizing a genus is a major breakthrough for establishing rapid methods for screening seized cargo and the supervision of manufactured products. In cases in which a specific identification of another species of Lauraceae not represented here is essential, we recommend the use of other highly variable regions that have undergone rapid evolution [27] in association with established barcode regions, tiered barcoding system [80] and/or studies at population levels. According to Feng et al. [80] a tiered barcoding system was an efficient tool for *Populus*, which is an ecologically and economically important genus of trees, but distinguishing between wild species is relatively difficult due to extensive interspecific hybridization and introgression, and the high level of intraspecific morphological variation.

Conclusion

One of the main objectives for establishing these molecular barcodes is to provide a practical forensic tool to determine the legality of the raw materials extracted from protected forests [34, 46, 47, 48] or from specific botanical families of interest to the timber industry [44, 45, 49]. The use of the ITS region could increase the laboratory time required for editing sequences or repeating sequencing reactions. Regardless, based on the practicality and viability of barcoding as a routine technique, we recommend the ITS region for barcode identification of these species, as it demonstrated a high degree of identification success using TAXONDNA and a better resolution of species-specific clusters in the phylogenetic tree generated by Maximum Parsimony, a computational method currently used in practice. Although exercising caution with regard to affirming this technique as the ultimate solution for identifying wood resources from MOF, our results show a great potential for recognizing those families most frequently explored, with high support at the genus level and relative confidence at the species level. This could be the first step in providing instruments for Brazilian authorities to follow the example of the new European Union rules for the timber trade, which since 2013, prohibits operators in Europe from placing illegally harvested timber and products derived from illegal timber on the EU market. Considering the great diversity of tree species in tropical forests, these molecular tools can assist in screening those botanical families most targeted by the timber industry and in detecting those species protected by Brazilian legislation.

Supporting Information

S1 Fig. Cladograms for ITS generated by Maximum Parsimony. The bootstrap values ≥70% are shown under the branches. The species name is followed by the accession number of the specimen.
(TIF)

S1 Table. Collection details for the commercial woody specimens used from MOF and accession numbers to International Nucleotide Sequence Database Collaboration Gen-Bank. Plant classification follows APG III (2009);—indicates no sequence obtained.
(DOCX)

S2 Table. Primers and PCR conditions for plastid and nuclear DNA sequence amplifications in this study. The complete references can be found in the text.
(DOCX)

S3 Table. Percentage of species-specific clusters (SSC) and non-species-specific clusters (NSSC) using different tree-based methods (NJ, MP, and ML) with different bootstrap support for single DNA regions and different combinations.
(DOCX)

Acknowledgments

Author Contributions

Conceived and designed the experiments: VSP ECS. Performed the experiments: MB ECS. Analyzed the data: MB ECS VSP. Contributed reagents/materials/analysis tools: MLB ECS VSP. Wrote the paper: MB VSP. Field collections and taxonomic identification: MB MLB.

References

1. Stehmann JR, Forzza RC, Salino A, Sobral M, Costa DP, Kamino LHY. Plantas da Floresta Atlântica. 1st ed. Rio de Janeiro: Jardim Botânico do Rio de Janeiro; 2009.

2. Veloso HP, Rangel-Filho ALR, Lima ICA. Classificação da vegetação brasileira, adaptada a um sistema universal. 1st ed. Rio de Janeiro: IBGE; 1991.

3. Koch Z, Corrêa MC. Araucária: a floresta do Brasil Meridional. 1st ed. Curitiba: Olhar Brasileiro; 2002.

4. Reitz R, Klein RM, Reis A. Madeiras do Brasil. 1st ed. Florianópolis: Lunardelli; 1979.

5. Isernhagen I, Silva SM, Rodrigues W, Galvão F. Listagem de espécies arbustivo-arbóreas citadas nos trabalhos de fitossociologia florestal no Paraná, Brasil: uma contribuição aos programas de recuperação de áreas degradadas (RAD). In: Isernhagen I. A fitossociologia florestal no Paraná e os programas de recuperação de áreas degradadas: uma avaliação. Piracicaba: Instituto de Pesquisa e Estudo Florestais. M.Sc. Thesis. Universidade Federal do Paraná. 2001. pp. 511–134. Available: http://www.ipef.br/servicos/teses/arquivos/isernhagen,i.pdf

6. van der Werff H. A key to the genera of Lauraceae in the new world. Ann Mo Bot Gard. 1991; 78: 377–387.

7. Chanderbali AS, van der Werff H, Renner SS. Phylogeny and historical biogeography of Lauraceae: evidence from the chloroplast and nuclear genomes. Ann Mo Bot Gard. 2001; 88: 104–134.

8. Brotto ML, Cervi AC, Santos EP. O gênero Ocotea (Lauraceae) no estado do Paraná, Brasil. Rodriguésia. 2013; 64: 495–525.

9. Pires PTL. Aspectos legais. In: Sanquetta CR, Mattei E, editors. Perspectivas de recuperação e manejo sustentável das florestas de araucária. Curitiba: Multi-Graphic; 2006. pp. 53–104.

10. Ministério do Meio Ambiente–MMA. Instrução Normativa MMA n° 06. 23 Sep 2008. Available: http://www.mma.gov.br/sitio/index.php?ido = conteudo.monta&idEstrutura=179&idConteudo=8121&idMenu=8618. Accessed 11 November 2011.

11. Martinelli G, Moraes MA. Livro vermelho da flora do Brasil. Rio de Janeiro: Instituto de Pesquisa do Jardim Botânico do Rio de Janeiro; 2013. pp 1120.

12. Deguilloux MF, Pemonge MH, Petit RJ. Novel perspectives in wood certification and forensics: dry wood as a source of DNA. Proc R Soc Lond B Biol Sci. 2002; 269: 1039–1046.

13. Hebert PDN, Cywinska A, Ball SL, de Waard JR. Biological identifications through DNA barcodes. Proc R Soc Lond B Biol Sci. 2003; 270: 313–322.

14. Cowan RS, Chase MW, Kress WJ, Savolainen V. 300,000 species to identify: problems, progress and prospects in DNA barcoding of land plants. Taxon. 2006; 55: 611–616.

15. Ford CS, Ayres KL, Toomey N, Haider N, Stahl JVA, Kelly LJ, et al. Selection of candidate coding DNA barcoding regions for use on land plants. Bot J Linnean Soc. 2009; 159: 1–11.

16. Kress WJ, Erickson DL, Andrew Jones F, Swenson NG, Perez R, Sanjur O, et al. Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. Proc Natl Acad Sci USA. 2009; 106: 18621–18626. doi: 10.1073/pnas.0909820106 PMID: 19841276

17. Petit RJ, Hampe A. Some evolutionary consequences of being a tree. Annu. Rev. Ecol. Evol. Syst. 2006; 7: 187–214.

18. CBOL Plant Working Group. A DNA barcode for land plants. Proc Natl Acad Sci USA. 2009; 106: 12794–12797. doi: 10.1073/pnas.0905845106 PMID: 19666622

19. Smith SA, Donoghue MJ. Rates of molecular evolution are linked to life history in flowering plants. Science, 2008; 322: 86–89. doi: 10.1126/science.1163197 PMID: 18832643

20. Rieseberg LH, Wood TE, Baack E. The nature of plant species. Nature. 2006; 440: 524–527. PMID: 16554818

21. Hollingsworth ML, Clark A, Forrest LL, Richardson J, Pennington RT, Long DG, et al. Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. Mol Ecol Resour. 2009; 9: 439–457. doi: 10.1111/j.1755-0998.2008.02439.x PMID: 21564673

22. Roy S, Tyagi A, Shukla V, Kumar A, Singh UM, Chaudhary LB, et al. Universal plant DNA barcode loci may not work in complex groups: A case study with Indian Berberis species. PLOS ONE. 2010; 5: e13674. doi: 10.1371/journal.pone.0013674 PMID: 21060687

23. Yue J- X, Li J, Wang D, Araki H, Tian D, Yang S. Genome-wide investigation reveals high evolutionary rates in annual model plants. BMC Plant Biol. 2010; 10: 242–254. doi: 10.1186/1471-2229-10-242 PMID: 21062446

24. Lahaye R, van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, et al. DNA Barcoding the floras of biodiversity hotspots. Proc Natl Acad Sci USA. 2008; 105: 2923–2928. doi: 10.1073/pnas.0709936105 PMID: 18258745

25. Chase MW, Salamin N, Wilkinson M, Dunwell JM, Kesanakurthi RP, Haidar N, et al. Land plants and DNA barcodes: short-term and long-term goals. Philos Trans R Soc B-Biol Sci. 2005; 360: 1889–1895.

26. Newmaster SG, Fazekas AJ, Ragupathy S. DNA Barcoding in land plants: evaluation of rbcL in a multi-gene tiered approach. Can J Bot-Rev Can Bot. 2006; 84: 335–341.

27. Shaw J, Lickey EB, Schilling EE, Small RL. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms. The tortoise and the hare III. Am J Bot. 2007; 94: 275–288. doi: 10.3732/ajb.94.3.275 PMID: 21636401

28. Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci USA. 2005; 102: 8369–8374. PMID: 15928076

29. Yao H, Song J, Liu C, Luo K, Han J, Li Y, et al. Use of ITS2 region as the universal DNA barcode for plants and animals. PLOS ONE. 2010; 5: e13102. doi: 10.1371/journal.pone.0013102 PMID: 20957043

30. Hollingsworth PM. Refining the DNA barcode for land plants. Proc Natl Acad Sci USA. 2011; 108: 19451–19452. doi: 10.1073/pnas.1116812108 PMID: 22109553

31. Li DZ, Gao LM, Li HT, Wang H, Ge XJ, Liu JQ, et al. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. PNAS. 2011; 108: 19641–19646. doi: 10.1073/pnas.1104551108 PMID: 22100737

32. Kesanakurti PR, Fazekas AJ, Burguess KS, Percy DM, Newmaster SG, Graham SW, et al. Spatial patterns of plant diversity below-ground as revealed by DNA barcoding. Mol Ecol. 2011; 20: 1289–1302. doi: 10.1111/j.1365-294X.2010.04989.x PMID: 21255172

33. Bleeker W, Klausmeyer S, Peitinger M, Dienst M. DNA sequences identify invasive alien *Cardamine* at Lake Constance. Biol Conser. 2008; 141: 692–698.

34. Kress WJ, Erickson DL, Swenson NG, Thompson J, Uriarte M, Zimmerman JK. Advances in the use of DNA barcodes to build a community phylogeny for Tropical trees in a Puerto Rican forest dynamics plot. PLOS ONE. 2010; 5: e15409. doi: 10.1371/journal.pone.0015409 PMID: 21085700

35. de Vere N, Rich TCG, Ford CL, Trinder SA, Long C, Moore CW, et al. DNA barcoding the native flowering plants and conifers of Wales. PLOS ONE. 2012; 7: 1–12.

36. Sass C, Little DP, Stevenson DW, Specht CD. DNA Barcoding in the Cycadales: Testing the potential of proposed barcoding markers for species identification of cycads. PLOS ONE. 2007; 2: e1154. PMID: 17987130

37. Du ZY, Qimike A, Yang CF, Chen JM, Wang QF. Testing four barcoding markers for species identification of Potamogetonaceae. J Syst Evol. 2011; 49: 246–251.

38. Fu YM, Jiang WM, Fu CX. Identification of species within Tetrastigma (Miq.) Planch. (Vitaceae) based on DNA Barcoding techniques. J Syst Evol. 2011; 49: 237–245.

39. Theodoridis S, Stefanaki A, Tezcan M, Aki C, Kokkini S, Vlachonasios KE. DNAbarcoding in native plants of the Labiatae (Lamiaceae) family from Chios Island (Greece) and the adjacent Çesme-Karaburun Peninsula (Turkey). Mol Ecol Resour. 2012; 12: 620–633. doi: 10.1111/j.1755-0998.2012.03129.x PMID: 22394710

40. Chen SL, Yao H, Han JP, Liu C, Song JY, Shi L, et al. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLOS ONE. 2010; 5: e8613. doi: 10.1371/journal.pone.0008613 PMID: 20062805

41. Gao T, Yao H, Song JY, Liu C, Zhu YJ, Zhua Y, et al. Identification of medicinal plants in the family Fabaceae using a potential DNA barcode ITS2. J Ethnopharmacol. 2010; 130: 116–121. doi: 10.1016/j.jep.2010.04.026 PMID: 20435122

42. Stoeckle MY, Gamble CC, Kirpekar R, Young G, Ahmed S, Little FP. Commercial teas highlight plan DNA barcode identification successes and obstacles. Sci Rep. 2011; 1: 1–7.

43. Kumar S, Kahlon T, Chaudhary S. A rapid screening for adulterants olive oil using DNA barcodes. Food Chem. 2011; 127: 1335–1341. doi: 10.1016/j.foodchem.2011.01.094 PMID: 25214135

44. Simeone MC, Piredda R, Papini A, Vessella F, Schirone B. Application of plastid and nuclear markers to DNA barcoding of Euro-Mediterranean oaks (Quercus, Fagaceae): problems, prospects and phylogenetic implications. Bot J Linnean Soc. 2013; 172: 478–499.

45. Vivas CV, Moraes RCS, Alves-Araújo A, Alves M, Mariano-Neto E, van den Berg C, et al. DNA barcoding in Atlantic Forest plants: What is the best marker for Sapotaceae species identification? Genet Mol Biol. 2014; 37: 662–670. doi: 10.1590/S1415-47572014005000019 PMID: 25505841

46. Parmentier I, Duminil J, Kuzmina M, Philippe M, Thomas DW, Kenfack D, et al. How effective are DNA barcodes in the identification of African rainforest trees? PLOS ONE. 2013; 8: e54921. doi: 10.1371/journal.pone.0054921 PMID: 23565134

47. Tripathi AM, Tyagi A, Kumar A, Singh A, Singh S, Chaudhary LB, et al. The Internal Transcribed Spacer (ITS) region and trnhH-psbA are suitable candidate loci for DNA barcoding of tropical tree species of India. PLOS ONE. 2013; 8: e57934. doi: 10.1371/journal.pone.0057934 PMID: 23460915

48. Gonzalez MA, Baraloto C, Engel J, Mori SA, Pétronelli P, Riéra B, et al. Identification of Amazonian trees with DNA barcodes. PLOS ONE. 2013; 4: e7483.

49. Muellner AN, Schaefer H, Lahaye R. Evaluation of candidate DNA barcoding loci for economically important timber species of the mahogany family (Meliaceae). Mol Ecol Resour. 2011; 11: 450–460. doi: 10.1111/j.1755-0998.2011.02984.x PMID: 21481203

50. Inoue MT, Roderjan CV, Kuniyoshi SY. Projeto madeira do Paraná. Curitiba: FUPEF; 1984.

51. Doyle J, Doyle J. A rapid DNA isolation procedure for small amounts of leaf tissue. Phytochem Bull. 1987; 19: 810–815.

52. Levin RA, Wagner WL, Hoch PH, Nepokroeff M, Pires JC, Zimmer EA, et al. Family-level relationships of Onagraceae based on chloroplast *rbcL* and *ndhF* data. Am J Bot. 2003; 90: 107–115. doi: 10.3732/ajb.90.1.107 PMID: 21659085

53. Kress J, Erickson DL. A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. PLOS ONE. 2007; 6: 1–10.

54. Tate JA, Simpson BB. Paraphyly of Tarasa (Malvaceae) and diverse origins of the polyploid species. Syst Bot. 2003; 28: 723–737.

55. Sang T, Crawford DJ, Stuessy TF. Chloroplast DNA phylogeny, reticulate evolution and biogeography of Paeonia (Paeoniaceae). Am J Bot. 1997; 84: 1120–1136. PMID: 21708667

56. Desfeaux C, Maurice S, Henry JP, Lejeune B, Gouyon PH. The evolution of reproductive system in the genus *Silene*. Philos Trans R Soc B-Biol Sci. 1996; 263: 409–414.

57. Beyra-Matos A, Lavin M. Monograph of Pictetia (Leguminosae: Papilionoideae) and review of the Aeschynomeneae. Syst Bot. 1999; 56: 1–3.

58. Käss E, Wink M. Molecular phylogeny and phylogeography of Lupinus (Leguminosae) inferred from nucleotide sequences of the *rbcL* gene and ITS 1+2 regions of rDNA. Plant Syst Evol. 1997; 208: 139–167.

59. Staden R, Judge DP, Bonfield JK. Analysing Sequences Using the Staden Package and EMBOSS. In: Krawetz SA, Womble DD, editors. Introduction to Bioinformatics: A theoretical and practical approach. Totawa: Human Press; 2003.

60. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994; 22: 4673–4680. PMID: 7984417

61. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol. 2013; 30: 2725–2729. doi: 10.1093/molbev/mst197 PMID: 24132122

62. Angiosperm Phylogeny Group III (APG III). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Bot J Linnean Soc. 2009; 161: 105–121.

63. Librado P, Rozas J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009; 25: 1451–1452. doi: 10.1093/bioinformatics/btp187 PMID: 19346325

64. Meier R, Shiyang K, Vaidya G, Ng PKL. DNA Barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. Syst Biol. 2006; 55: 715–728. PMID: 17060194

65. Pettengill JB, Neel MC. An evaluation of candidate plant DNA barcodes and assignment methods in diagnosing 29 species in the genus Agalinis (Orobanchaceae). Am J Bot. 2010; 97: 1391–1406. doi: 10.3732/ajb.0900176 PMID: 21616891

66. Saitou N, Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987; 4: 406–425. PMID: 3447015

67. Srivathsan A, Meier R. On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature. Cladistics. 2011; 27: 1–5.

68. Felsenstein J. Confidence limits on phylogenies: an approach to using bootstrap. Evolution. 1985; 39: 783–791.

69. Swofford DL. PAUP: Phylogenetic Analysis using Parsimony. Version 4.0b. Champaign: Illinois Natural Survey; 2002.

70. Fitch WM. Towards defining the course of evolution: minimum change for a specific tree topology. Syst Zool. 1971; 20: 406–416.

71. Newmaster SG, Ragupathy S. Testing plant barcoding in a sister species complex of pantropical Acacia (Mimosoideae, Fabaceae). Mol Ecol Resour. 2009; 9: 172–180. doi: 10.1111/j.1755-0998.2009.02642.x PMID: 21564976

72. De Mattia F, Bruni I, Galimberti A, Cattaneo F, Casiraghi M, Labra M. A comparative study of different DNA barcoding markers for the identification of some members of Lamiaceae. Food Res Int. 2011; 44: 693–702.

73. Yan HF, Hao G, Hu CM, Ge XJ. DNA Barcoding in closely related species: A case study of Primula L. sect. Proliferae Pax (Primulaceae) in China. J Syst Evol. 2011; 49: 225–236.

74. Kelly LJ, Ameka GK, Chase MW. DNA Barcoding of African Podostemaceae (river-weeds): A test of proposed barcode regions. Taxon. 2010; 59: 251–260.

75. Li Y, Gao LM, Poudel RC, Li DZ, Forrest A. High universality of matK primers for barcoding gymnosperms. J Syst Evol. 2011; 49: 169–175.

76. Yu J, Xue JH, Zhou SL. New universal *matK* primers for DNA barcoding angiosperms. J Syst Evol. 2011; 49: 176–181.

77. Starr JR, Naczi RF, Chouinard BN. Plant DNA barcodes and species resolution in sedges (Carex, Cyperaceae). Mol Ecol Resour. 2009; 9: 151–63. doi: 10.1111/j.1755-0998.2009.02640.x PMID: 21564974

78. Xue CY, Li DZ. Use of DNA barcode sensu lato to identify traditional Tibetan medicinal plant Gentianopsis paludosa (Gentianaceae). J Syst Evol. 2011; 49: 267–270.

79. Liu Z, Chen SL, Song JY, Zhang SJ, Chen KL. Application of deoxyribonucleic acid barcoding in Lauraceae plants. Pharmacogn Mag. 2012; 8: 4–11. doi: 10.4103/0973-1296.93301 PMID: 22438656

80. Feng J, Jiang D, Shang H, Dong M, Wang G, He X, et al. Barcoding Poplars (Populus L.) from Western China. PLOS ONE 2013; 8: e71710. doi: 10.1371/journal.pone.0071710 PMID: 23977122