Received: 11 February 2022 | Accepted: 17 August 2022

DOI: 10.1002/ajb2.16064

RESEARCH ARTICLE

Botany

Cryptic species diversity in a widespread neotropical tree genus: The case of *Cedrela odorata*

Kristen N. Finch[1] | **F. Andrew Jones**[1,2] | **Richard C. Cronn**[3]

[1]Department of Botany and Plant Pathology, Oregon State University, 2082 Cordley Hall, 2701 SW Campus Way, Corvallis, OR 97331, USA

[2]Smithsonian Tropical Research Institute, Balboa, Ancon, Republic of Panama

[3]USFS PNW Research Station, 3200 SW Jefferson Way, Corvallis, OR 97331, USA

Correspondence

Kristen N. Finch, Oregon State University, Department of Botany and Plant Pathology, 2082 Cordley Hall, 2701 SW Campus Way, Corvallis, OR 97331 USA.
Email: finchkn@uw.edu

Abstract

Premise: Reconciling the use of taxonomy to partition morphological variation and describe genetic divergence within and among closely related species is a persistent challenge in phylogenetics. We reconstructed phylogenetic relationships among *Cedrela odorata* (Meliaceae) and five closely allied species to test the genetic basis for the current model of species delimitation in this economically valuable and threatened genus.

Methods: We prepared a nuclear species tree with the program SNPhylo and 16,000 single-nucleotide polymorphisms from 168 *Cedrela* specimens. Based on clades present and ancestral patterns ADMIXTURE, we designed nine species delimitation models and compared each model to current taxonomy with Bayes factor delimitation. Timing of major lineage divergences was estimated with the program SNAPP.

Results: The resulting analysis revealed that modern *C. odorata* evolved from two genetically distinct ancestral sources. All species delimitation models tested better fit the data than the model representing current taxonomic delimitation. Models with the greatest marginal likelihoods separated Mesoamerican *C. odorata* and South American *C. odorata* into two species and lumped *C. angustifolia* and *C. montana* as a single species. We estimated that *Cedrela* diversified in South America within the last 19 million years following one or more dispersal events from Mesoamerican lineages.

Conclusions: Our analyses show that the present taxonomic understanding within the genus obscures divergent lineages in *C. odorata* due in part to morphological differentiation and taxonomic distinctions that are not predictably associated with genetic divergence. A more accurate application of taxonomy to *C. odorata* and related species may aid in its conservation, management, and restoration efforts.

KEYWORDS

CITES, conservation, illegal logging, Meliaceae, multispecies coalescence, neotropical forests, next generation sequencing, SNPs, target capture, taxonomy

Forests of Central and South America support high levels of terrestrial biodiversity including an estimated 16,000 tree species within Amazonian rainforests alone (Pennington et al., 2015; Pennington and Lavin, 2016; Dick and Pennington, 2019). Loss of biodiversity is a consequence of unsustainable practices and the removal of living trees from forests around the world (Elias, 2012; Nellemann, 2012; van Zonneveld et al., 2018), and accordingly, neotropical forests have the largest number of protected tree species. Exploitative logging and land-use conversion can lead to forest fragmentation, which can have severe consequences on natural regeneration and available genetic diversity for adapting to climate change and more-rapidly evolving enemies (Millar and Libby, 1991; O'Neill et al., 2001; Lowe et al., 2003; Kometter et al., 2004; Muellner et al., 2011; Inza et al., 2012). Although laws are in place to

protect economically valuable tree species from over-exploitation and promote sustainable practices (e.g., US Lacey Act, EU Timber Regulation, and Australia Illegal Logging Prohibition Act), 35–72% of the wood sourced in the Amazon is thought to be acquired from illegal logging (Saunders and Reeve, 2014).

The neotropical tree species *Cedrela odorata* (Meliaceae) has been logged in regional timber trade for over 250 years and is a target of illegal logging (Pennington and Muellner, 2010). The primary threat to *C. odorata* is a high level of exploitative logging (Urrunaga et al., 2012), combined with decreased population connectivity (Muellner et al., 2010, 2011; Pennington and Muellner, 2010; Cavers et al., 2013). Although *Cedrela odorata* is distributed from Mexico to northern Argentina (24°N to 27°S) and Caribbean islands (Pennington and Muellner, 2010), it primarily occurs in low-density stands with approximately one individual per hectare in Peru, Costa Rica, Colombia, and Guyana (Tajikistan, 2019). Deforestation has led to a nearly 30% decline in the global distribution of *C. odorata* in the last 100 years, and this decline is expected to continue (Tajikistan, 2019).

In 2001, *C. odorata* was listed under the protections of CITES Appendix III requiring validated documentation of species identity and source for both export and import documentation, protecting populations in Bolivia, Brazil, Colombia, Guatemala, and Peru (Ferriss, 2014). Consequently, two other species (*C. fissilis* Vell. and *C. lilloi* syn. *angustifolia*) have also been listed under CITES Appendix III because their survival is threatened due to strong morphological similarities with *C. odorata* (Ferriss, 2014; UNEP-WCMC, 2015). As a result, differentiating among protected *C. odorata* and its "look-alikes" that appear in trade is challenging. *Cedrela* species are differentiated based on combinations of six morphological characteristics (e.g., size of fruit, density of leaflet hairs, and number of leaflet pairs), some of which are regionally plastic and all of which show high levels of variation within species and overlap among species, making field identification difficult for nearly all members of the genus (Pennington and Muellner, 2010). A greater challenge is differentiating among processed wood specimens without leaves, flowers, and fruit (Gasson, 2011), which leads to most *Cedrela* wood specimens being declared as "*C. odorata*" irrespective of the actual identity and thus confounding harvest estimates across all *Cedrela* species (Tajikistan, 2019). Legal and illegal trade in *C. odorata* wood impacts its congeners that are traded as *C. odorata* intentionally or accidentally. In 2019, CITES accepted a proposal to list all *Cedrela* species to Appendix II (including *C. odorata*), leading to increased global scrutiny for all wood and wood products identified as a *Cedrela* species (Tajikistan, 2019).

The most recent revision of the genus (Pennington and Muellner, 2010) describes 17 species based on morphological, ecological, and genetic evidence. Morphological descriptions are associated with genetic distinctiveness and phylogenetic position in some *Cedrela* species (Muellner et al., 2010; Pennington and Muellner, 2010). However, multiple phylogenies of *Cedrela* have revealed complexity and uncertainty in taxonomic position of *Cedrela* species arising from two observations. (1) *Cedrela odorata* specimens showing coherent morphology are paraphyletic, suggesting the existence of diverse lineages or cryptic species within *C. odorata* (i.e., related species that are classified under one species name due to high morphological similarities, even though they are as genetically divergent as other separate species) (Pennington and Muellner, 2010; Cavers et al., 2013). Although these have been repeatedly acknowledged in the literature, they are hidden in taxonomy, consequently disregarded on trade records, and likely neglected by conservation status assessments and range estimations. (2) Some distinct morphological species are nested within *C. odorata* and have more restricted geographic distributions (e.g., *C. nebulosa*; Pennington and Muellner, 2010), but the genetic distinctiveness of these species is uncertain. The contradictory application of *Cedrela* taxonomy, which recognizes narrow morphological variation but disregards distinct genetic lineages, could hinder the prescription of protections and conservation initiatives aimed at restricting trade, conservation, and restoring *Cedrela odorata* and allied species in neotropical forests.

Cryptic species

Multiple studies have reported the presence of cryptic diversity within *C. odorata* with one or more species in South America (Cavers et al., 2013) and Mesoamerica. The Mesoamerican *C. odorata* taxon (referred here as *C. odorata* s.s.) was first described by Gillies et al. (1997), who identified strong genetic differences between North Pacific populations and nearby Atlantic populations in Costa Rica. This genetic division was further supported using organellar and nuclear genetic markers (Cavers et al., 2003a, 2003b, 2013; Navarro et al., 2004, 2005; Muellner et al., 2010; Pennington and Muellner, 2010). The line of demarcation between these two populations followed the Cordilleras de Guanacaste, Talamanca, and the Central Mountains in Costa Rica, a region known to harbor cryptic diversity in reptiles (Lotzkat et al., 2011; Cadle, 2012; Doan et al., 2016), amphibians (Crawford et al., 2007; Arias and Kubicki, 2018), and other tree species (Cavers and Lowe, 2002). Most compellingly, Navarro et al. (2002) demonstrated that *C. odorata* from drier climate regions (Mexico to Pacific Costa Rica) were significantly larger across multiple traits (i.e., seed mass, shoot height and diameter, and leaf size) than trees from mesic regions (Atlantic Costa Rica to Panama) when grown in common gardens in Costa Rica. Climatic groups (dry vs. mesic) explained 52% of the total variance and 80% of genetic variance. This sharp transition in quantitative trait variation and genetic distance point to the central mountain ranges of Costa Rica as a potentially strong migration or reproductive barrier and/or a zone of

secondary contact between cryptic species (Cavers et al., 2013). Additional cryptic species (Cavers et al., 2013) have been suggested to occur within for South American *C. odorata*; however, observed genetic differentiation between Mesoamerican and South American *C. odorata* has not been incorporated into taxonomic treatments of the species.

Recent species descriptions

Before the most recent treatment of the genus (Pennington and Muellner, 2010), *Cedrela odorata* had historically been treated as a widespread and morphologically diverse species, showing geographic, altitudinal, and ecological structuring of morphological variation (Smith, 1960; Styles, 1981; Pennington and Muellner, 2010). Phylogenetic analyses revealed some genetic basis for morphological variation, and a revised taxonomy was applied to recognize fine-scale morphological variation, leading to the description of four new species among South American populations that were previously ascribed to *C. odorata* (Zapater et al., 2004; Pennington and Muellner, 2010). These descriptions made *C. odorata* a more cohesive morphological species, but these recently named taxa are more closely related to South American *C. odorata* than the putative cryptic species—South American *C. odorata* and Mesoamerican *C. odorata*—are to each other (Pennington and Muellner, 2010). The current treatment of these populations as distinct nominal species (e.g., *C. nebulosa* T.D. Penn. & Daza in 2010 and *C. saltensis* M.A. Zapater & del Castillo in 2004) in combination with the reluctance to recognize cryptic species of *C. odorata* in taxonomy underscores the persistent conflict between the use of taxonomy to partition morphological variation and the use of taxonomy to describe genetic divergence (Hey et al., 2003; Bickford et al., 2007; Struck et al., 2018).

We created a nuclear phylogeny of six species in the genus *Cedrela*, focusing on *C. odorata* and closely allied species in Mesosmerica and western South America. Using the species delimitation scheme presented in Pennington and Muellner (2010) as a hypothesis and multispecies coalescence analysis with thousands of single nucleotide polymorphisms (SNPs), we seek to (1) determine whether dense sampling of dispersed nuclear SNPs supports a monophyletic or paraphyletic resolution of *C. odorata*, (2) determine whether SNPs support the presence of cryptic species within *C. odorata*, (3) determine whether narrow endemic species represent distinct lineages or instead appear to be conspecific with *C. odorata*, and (4) estimate the timing of major lineage divergences. Because *C. odorata* and its congeners are protected and vulnerable to over-exploitation for timber (Muellner et al., 2010, 2011; Pennington and Muellner, 2010; Cavers et al., 2013; Ferriss, 2014; UNEP-WCMC, 2015; Tajikistan, 2019), it is important to have a clearly defined taxonomy so that regulators understand species distribution and which species are represented in wood trade. Moreover, *Cedrela* taxonomy is dynamic, with new species being described including *C. ngobe* Köcke, T.D. Penn. & Muellner (Köcke et al., 2015), *C. domatifolia* (Palacios et al., 2019), *C. microanthus*, and *C. pandeirensis* (Huamán Mera, 2014). Improving our understanding of the partitioning of genetic lineages in *Cedrela* species and within *C. odorata* can help improve estimates of biodiversity in the neotropics, leading to improved protection and conservation. Our analysis focuses on *Cedrela odorata*, but it provides a framework for how genealogies and taxonomy can be reinvestigated in the future using a full sampling of *Cedrela* species and other tropical tree lineages that likely include cryptic or poorly defined species (Dick et al., 2003; Dexter et al., 2010; Garcia et al., 2011; Scotti-Saintagne et al., 2013; Fine et al., 2014; Winterton et al., 2014; Gill et al., 2016).

MATERIALS AND METHODS

Samples, target enrichment, and sequencing

We prepared a genomic data set for six *Cedrela* species: *C. angustifolia* D.C., *C. fissilis* Vell., *C. montana* Mortiz ex Turcz., *C. nebulosa* T.D. Penn. & Daza, *C. odorata* L., and *C. saltensis* M.A. Zapater & del Castillo (Appendices S1, S2). Leaf tissue samples from 192 *Cedrela* individuals were obtained from herbarium specimens at the Missouri Botanical Garden Herbarium (MO). We extracted total genomic DNA from leaf fragments with the FastDNA Kit (MP Biomedicals, Santa Ana, CA, USA), quantified genomic DNA by fluorometry (Qubit, Thermo Fisher Scientific, Waltham, MA, USA), and sheared DNA to a modal size of 130 bp. Genomic DNA libraries were prepared using the NEBNext Ultra II Kit (New England Biolabs, Ipswich, MA, USA) and an input of genomic DNA ranging from 20 to 250 ng.

Libraries were pooled into equimolar 24-plex pools for hybridization capture target enrichment ("target capture") with custom MYbaits (Arbor Biosciences, Ann Arbor, MI, USA). Methods for the development of the hybridization probes used in this study were described previously (Finch et al., 2019). Target capture was performed with a hybridization temperature of 65°C to ensure specificity. Enriched targets were amplified using the KAPA HiFi HotStart ReadyMix PCR Kit (v. 5.13; KAPA Biosystems, Boston, MA, USA) and NEXTFlex Primers (Bioo Scientific, Austin, TX, USA). We prepared two 96-plex pools of target capture libraries by combining four 24-plexes in equimolar ratios. Pools were sequenced with paired-end 100-bp reactions on the Illumina HiSeq 3000 (Center for Quantitative Life Sciences, Oregon State University, Corvallis, OR, USA).

Sequence yield was estimated by counting raw sequence reads per specimen (Appendix S1; Finch, 2018; Finch et al., 2019). We selected sequenced reads from a single individual, *C. odorata* 300, to serve as a reference nuclear gene sequence for further analysis. This individual was

selected because it originated in Peru, a region of general interest for this study due to high *Cedrela* species richness in Peru and specific interest driven by conservation and protection concerns, and sequence yield was high compared to other samples (Appendix S1). We assembled sequenced reads resulting from target capture with *C. odorata* 300 de novo described by Finch et al. (2020). Briefly, sequencing *C. odorata* 300 generated 1.4×10^7 paired reads (2.8×10^9bp) for the de novo assembly. The resulting *C. odorata* 300 reference contig sequences were filtered to remove sequences lacking the hybridization probe sequences or were homologous to organelle sequences. The remaining 9139 assembled contigs were used as our reference sequence for SNP variant calling. The mean contig length for the reference assembly was 982.5 bp (range: 156–4053 bp) for a total assembly length of 9.0×10^6bp (N50 = 983 bp; 36.1% GC; Finch et al., 2020).

Assessment of on-target yield

Sequence reads were considered "on target" if they shared 90% identity with the contigs from the *C. odorata* 300 reference. To make this assessment, we mapped single-end, captured DNA sequence reads from *Cedrela* herbarium specimens to the nuclear reference with bbmap.sh from BBTools (v. 36.14; DOE Joint Genome Institute; https://jgi.doe.gov/data-and-tools/bbtools) in local alignment mode. Mapping conflicts were resolved by retaining one mapping location selected at random, maintaining greater depth of coverage rather than discarding all multi-mapping reads. We used the bbmap.sh "coveragestats" parameter to obtain coverage estimates for all *C. odorata* 300 reference contigs. Depth was calculated by multiplying the sequence read length (101 bp) by the number of sequenced reads mapping to each target contig (plus and minus strands), divided by the covered target length (covered bases; Finch et al., 2019). We used a one-way ANOVA to compare on-target yield and depth across *Cedrela* species and to confirm that the baits showed no species-specific bias.

Assessment of target capture efficiency

Hybridization probe sequences were designed using a leaf transcriptome assembly derived from a Mexican *C. odorata* in the living collection at the New York Botanical Garden (NYBG; specimen CEOD-NYBG; Finch et al., 2019). The selection of this individual was driven by the logistical feasibility of obtaining flash-frozen tissue from a living specimen for RNA extraction. Using this assembly, we were able to assess the taxon-specific enrichment bias of our *C. odorata*-derived hybridization probes in other *Cedrela* species by mapping 10^6 organelle-depleted reads to the *C. odorata* 300 reference and estimating resulting read depth. This analysis was performed for specimens that yielded at least 10^6 organelle-depleted sequence reads ($n = 151$). Sequence reads with 90% chloroplast identity to our draft chloroplast genome for CEOD-NYBG (Finch et al., 2019) were removed using bbmap.sh, and 10^6 single-end sequence reads were randomly selected with reformat.sh (BBTools). We mapped subsampled reads by individual to the *C. odorata* 300 reference in local alignment mode with bbmap.sh at 90% identity; the coveragestats parameter allowed us to estimate mean depth for the six species groups. Depth was calculated as described above. We used a one-way ANOVA to test for differences in mean depth among species.

Alignment and variant detection

To assess variation via SNPs, we made a variant call format file (vcf; Danecek et al., 2011) using sequence reads from target capture and a protocol combining SAMtools (v. 1.9; Li et al., 2009) and the Genome Analysis Toolkit (v. 3.7; GATK; McKenna et al., 2010). Sequence reads from 192 *Cedrela* specimens were aligned to the *C. odorata* 300 target capture reference using BWA-MEM (v. 0.7.12-r1039; Li and Durbin, 2010; Li, 2013). SAMtools was used to convert and sort alignment files, GATK to define and realign insertions and deletions, and bcftools (v. 1.9 mpileup; Li, 2011) to prepare a vcf, excluding indel calls to speed vcf generation. To avoid dubious inferences due to missing information, we excluded *Cedrela* specimens with more than 10% missing information after assessing sequence yield per individual, leaving 167 trees for analysis (Appendix S1). We also included DNA sequencing reads from the *C. odorata* individual CEOD-NYBG used to design the target capture probes for this study (Finch et al., 2019), and we included sequence reads from two *Swietenia mahagoni* (L.) Jacq. (Meliaceae) individuals (Finch et al., 2019) to serve as outgroup sequences. This resulted in 168 *Cedrela* and two *S. mahagoni* specimens for the current analysis (Appendix S1).

Phylogenetic analyses of nuclear SNPs

We used our vcf containing 170 *Cedrela* and *Swietenia* specimens and the program SNPhylo (v. 20140701; Lee et al., 2014) to generate a species tree from nuclear SNPs. SNPhylo accepts a vcf input, filters low quality data, extracts SNPs in putative linkage equilibrium, aligns SNPs with MUSCLE (v. 3.8.31; Edgar, 2004), and generates a maximum likelihood species tree with PHYLIP (v. 3.697 DNAML; Felsenstein, 1993). Our SNPhylo species tree was generated using default parameters and by assigning *S. mahagoni* as the outgroup. To obtain bootstrap support values for the nuclear species tree generated with SNPhylo, we used the PHYLIP alignment file produced by SNPhylo and inferred a maximum likelihood species tree with 1000 bootstrap replicates via RAxML (v. 8.2.10; Miller et al., 2011; Stamatakis, 2014). For this analysis, we used the 16-state, general time reversible (GTR), secondary-structure-substitution model and

estimated invariable sites with a gamma distribution of rate variation among sites (GTR-I-Γ). All other parameters were default. A bootstrap majority rule consensus tree was generated with Mesquite (v. 3.31 [build 859]; Maddison and Maddison, 2017) with the required frequency of clades set to 0.5, and some bootstrap support values (converted to percentage) were superimposed over the SNPhylo maximum likelihood tree. Tree graphics were generated in R (v. 3.6.1; R Core Team, 2013) (Appendix S3: list of R packages).

Ancestry estimation from SNP data

We estimated the structure of ancestral lineages across our samples with ADMIXTURE (v. 1.3.0; Alexander et al., 2009). Like the program STRUCTURE (Pritchard et al., 2000), ADMIXTURE estimates the optimal number of ancestral populations (K) among individuals and membership coefficients to ancestral populations. Ancestry was estimated with the quality filtered vcf output from SNPhylo (default parameters: minimum depth of coverage = 5, linkage disequilibrium threshold = 0.1, minor allele frequency threshold = 0.1, and missing rate = 0.1) that was further filtered via VCFtools (v. 0.1.17; Danecek et al., 2011) to remove SNPs with more than two alleles, SNPs with a quality lower than 30 (defined as $-10\ \mathrm{Log}_{10}$ [Probability of incorrect SNP call]), SNPs with a minor allele represented in a single individual, and SNPs containing missing information. To further ensure linkage equilibrium among SNPs, we then thinned SNPs using a window of 5000 bases to retain approximately one SNP per contig from the target capture reference for *C. odorata* 300 ($n = 9139$ contigs and 9139 SNPs). Finally, we removed *S. mahagoni* individuals, converted the SNP data set to PLINK format (v. 1.90b5.2; Purcell et al., 2007) and performed ancestry prediction with ADMIXTURE for K values of 1 (one ancestral population) through to 10 (10 ancestral populations). The optimal K (number of ancestral populations) was determined via 5-fold cross validation and measured in prediction error (Alexander et al., 2009). Ancestry proportions for each individual were plotted as bar plots in R. We showed geographic distributions of ancestry proportions by converting bar plots into pie charts and superimposing them over the geographic source of each individual with R package scatterpie. Maps were drawn using the base map shapefiles from the World Borders Dataset (http://thematicmapping.org/downloads/world_borders.php).

Species delimitation

We used Bayes factor delimitation (BFD*; Battey and Klicka, 2017; Song et al., 2017; Leaché and Ogilvie, 2018) to compare species delimitation models for *C. odorata* and closely allied species. BFD*, implemented in the programs SNAPP (v. 1.3.0; Bouckaert and Bryant, 2012; Leaché and Bouckaert, 2018) and BEAST2 (v. 2.6.1; Bouckaert et al., 2014), generates multispecies coalescent models based on SNPs. In this way, we compared eight models with taxa showing alternative species definitions (Figure 3). Each delimitation model was ranked based on a marginal likelihood score and log Bayes factors (Leaché and Ogilvie, 2018) to determine the species relationship model that best fits the data. Log Bayes factors were estimated by calculating the difference in marginal likelihood estimate for each model and the marginal likelihood estimate for the model representing current taxonomy (model C), and multiplying that difference by 2 (Stange et al., 2017). We also calculated log Bayes factors to compare two models with near equal marginal likelihood estimates across replicates.

Based on ancestral patterns from the ADMIXTURE analysis and clades present in the nuclear species tree, we designed nine species delimitation schemes for *C. odorata* and other closely related species of *Cedrela* (Figure 3). Model C uses current taxonomy, while models 1, 2, and 3 each differ from current taxonomy by one change: (1) model 1 tests the definition of *C. odorata* by considering it as two taxa, one from Mesoamerica (abbreviated *C. odorata* M), and a second from southeastern Costa Rica, Panama, and the rest of South America (abbreviated *C. odorata* SA); (2) model 2 tests the definition of *C. nebulosa* and *C. saltensis* as a one taxon, instead of two (Figure 3); and (3) model 3 tests the definition of *C. angustifolia* and *C. montana* as a one taxon, instead of two. Models 1.2, 1.3, and 2.3 are each two steps away from model C and represent combinations of models 1, 2, and 3 (e.g., model 1.2 combines changes to current taxonomy that were presented in models 1 and 2; Figure 3). Model 4 treats *C. odorata* as two taxa, one composed of *C. odorata* M and a second composed of *C. odorata* SA combined with *C. nebulosa* and *C. saltensis*. Model 4.1 is similar to model 4, but it treats *C. angustifolia* and *C. montana* as one taxon.

Due to computational limitations, this analysis used 15 individuals from clades depicted on the SNPhylo nuclear species tree and 1000 randomly selected SNP loci. To evaluate the effect of individual selection on the BFD* analysis and results, we replicated the full analysis of nine species delimitation models using four sets of 15 individuals representing genetic information from four combinations of 30 specimens. Loci were held constant across models and replicates. The 30 individuals selected (identified in red, Figure 1; identified by stars, Appendix S1) showed low amounts of admixture and missing information, and we avoided specimens that we suspected are taxonomic misidentifications. An extensible markup language (xml) file to run SNAPP via BEAST2 was prepared with the SNPhylo filtered vcf and snapp_prep.rb (described by Stange et al., 2017; https://github.com/mmatschiner/snapp_prep). We did not adjust sampling parameters for mutation rate, θ, coalescence rate, or clock rate for this analysis. Instead, we assessed whether parameter adjustments were necessary after examining trace plots for model convergence; in all cases, we found that no further parameter adjustments were necessary (effective sample

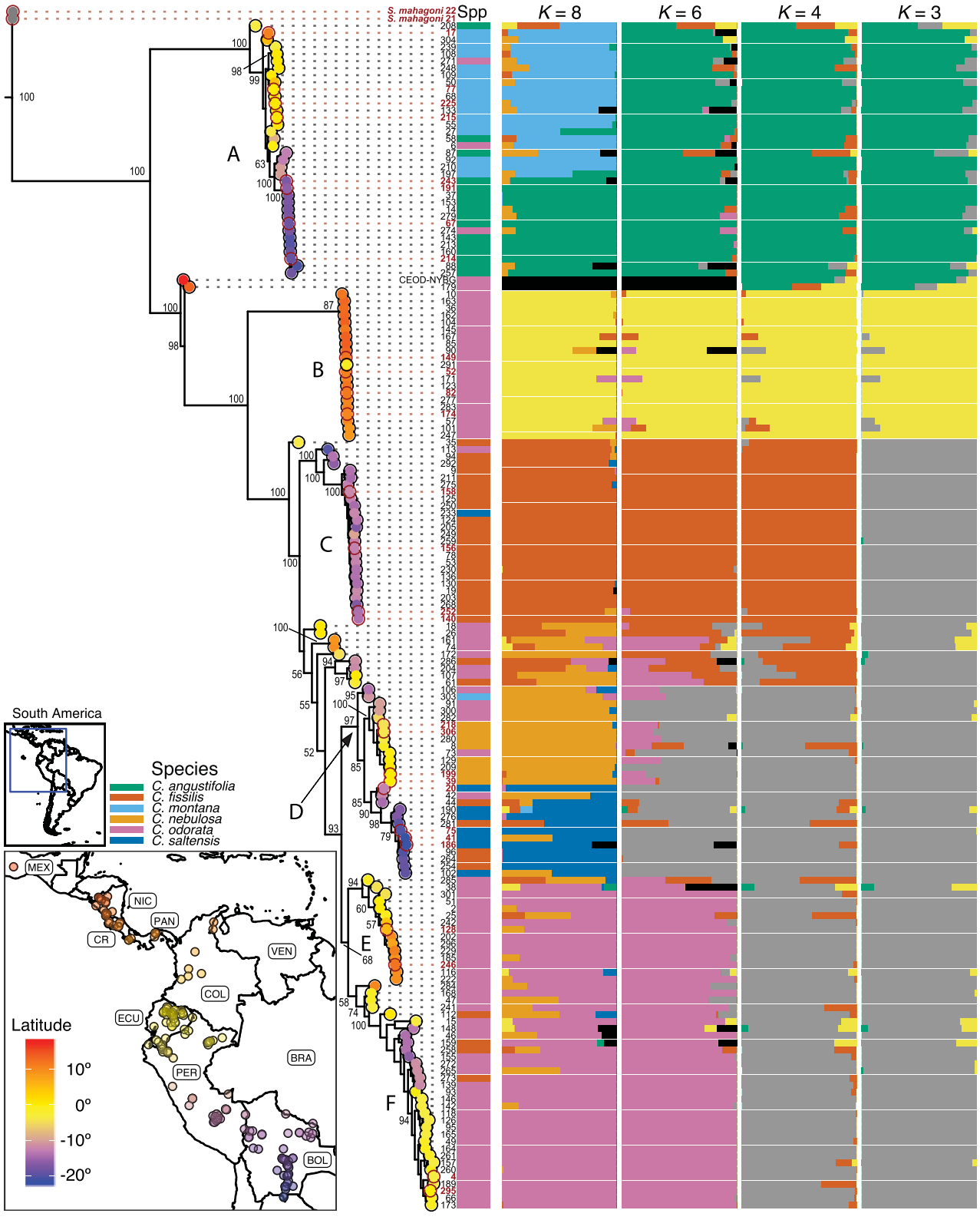

FIGURE 1 Maximum likelihood nuclear species tree, species identifications, and results of ancestry estimation for $K = 3$, 4, 6, and 8. Tips, or individual specimens, are color-coded by latitude listed on the herbarium specimen label, and correspond to the map, showing a portion of South America and the geographic positions of all specimens included in the species tree. The colors in the column labeled "Spp" indicate the species identification listed on the specimen label and correspond to the legend labeled "Spp." The bar plots show ancestry models indicated by $K = 3$, 4, 6, and 8. For ease of interpretation, we have used a similar color palette for ancestral populations such that proportions correlate with species identifications where possible, but the colors in these bar plots do not indicate species identity. Numbers at nodes indicate bootstrap support values for 1000 replicates with the same data and RAxML. Specimens in red were selected for species delimitation with BFD*.

size estimates > 200). In all analyses, *S. mahagoni* was used as the outgroup, and analyses were not supplied with a guide tree because incorrect guide trees might bias the analysis (Leaché et al., 2014; but also see Grummer et al., 2014; Leaché and Ogilvie, 2018). We ran SNAPP via BEAST2 to estimate species trees with a chain length of 100,000 Markov chain Monte Carlo (MCMC) iterations, storing every 50th tree after discarding the first 10,000 iterations as burn in (10%). BFD* speciation models were assessed after 50 stone sampling steps (Leaché and Ogilvie, 2018).

Absolute divergence time estimation

We estimated a time-calibrated phylogeny with SNAPP via BEAST2 for the 30 individuals used in BFD* replicates and the same 1000 SNP loci. To prepare the xml guide file, we provided the snapp_prep.rb script with a file containing time constraints to calibrate the species tree to absolute time using a strict clock model (Stange et al., 2017). For this analysis, we used three fossil calibration points as constraints. First, the stem node of the subfamily Cedreloideae was calibrated at a mean age of 51.9 million years ago (Ma) based on fossil fruit and seeds of *Toona sulcata* (Meliaceae) recovered from the London Clay fossil site in the United Kingdom (Chandler, 1964). We used this fossil to calibrate the stem node of the Cedreloideae rather than the crown node of tribe Cedreleae (Muellner et al., 2009) based on the interpretation of Koenen et al. (2015). For this parameter estimation, we instructed BEAST2 to sample from a lognormal distribution with a standard deviation of 0.04 to capture the maximum (56.1 Ma at the 97.5th percentile) and minimum (47.9 Ma at the 2.5th percentile) age range for this fossil. Second, we constrained the stem age of Cedreleae at a mean age of 31.05 based on the Eocene/Oligocene leaves and seeds of *Cedrela lancifolia* from the Florissant fossil beds of Colorado (MacGinitie, 1953; Manchester, 2001). This time calibration point was sampled from a lognormal distribution with a standard deviation of 0.05 to capture the putative age range for the fossil (34.2 Ma at the 97.5th percentile; 28.1 Ma at the 2.5th percentile). Lastly, we constrained the stem node leading to the outgroup, *S. mahagoni*, with a fossil flower of *Swietenia miocenica* that was found preserved in amber in Chiapas, Mexico (Castañeda-Posadas and Cevallos-Ferriz, 2007). BEAST2 sampled this parameter from lognormal distribution with a mean of 24.25 Ma and a standard deviation of 0.038 (26.1 Ma at the 97.5th percentile; 22.5 Ma at the 2.5th percentile). We estimated this reduced-species tree with chain length of 1,000,000 MCMC iterations storing every 5000th tree after discarding the first 100,000 iterations as burn in (10%); these steps were done to ensure model convergence and to prepare a maximum clade credibility tree with sampled trees via TreeAnnotator (v. 2.6.0; Helfrich et al., 2018).

RESULTS

Sequencing

Target capture and short read sequencing of 192 *Cedrela* specimens resulted in 1.6×10^9 paired sequence reads (3.2×10^{11}bp total) with a mean yield of 8.3×10^6 paired reads per specimen (range: 7.1×10^4–4.7×10^7; Appendix S1). On average, 6.4×10^4 reads mapped to each contig from the target capture reference *C. odorata* 300 (range: 450–9.8×10^5 reads; Appendix S1) for an average depth of 44.5× per target (range: 1.5 × –308.7×; Appendix S1). There were no "failed" targets, and at least 450 single-end reads mapped to targets. Using normalized inputs (10^6 organelle-depleted reads), the estimated individual mean depth was 13.5× (range: 8.5×–17.5×) per million reads, and we did not observe a significant difference in mean depth among species (one-way ANOVA; $F_{5,145} = 0.301$, $P = 0.912$). Moreover, the depth of coverage for each species converged on the same mean, indicating an absence of measurable enrichment bias across *Cedrela* species (Appendix S4).

Nuclear species tree and ancestry inferences

Initial filtering by SNPhylo resulted in a vcf with 9.9×10^5 SNPs (46.2% of detected non-indel variants). From these, 1.6×10^4 SNPs met the linkage disequilibrium threshold and were used to generate the nuclear species tree (Figure 1). These SNPs derived from 8941 contigs of the *C. odorata* 300 reference, and 1 to 10 SNPs per contig were selected (mean = 1.7 SNPs/contig).

Common patterns arose when comparing the maximum likelihood tree based on nuclear SNPs and the plots showing ADMIXTURE-based ancestry proportions. The ADMIXTURE analysis determined that a model with three ancestral populations ($K = 3$ groups) was the optimal number to represent these data (Figure 1). The $K = 3$ model showed the lowest cross-validation error among tested models ($K = 1$ through 10) and a log likelihood (lnL) of −115,949.1192. This model supported ancestral population groups that coincide with taxonomic divisions between (1) *C. angustifolia* plus *C. montana*, (2) Mesoamerican *C. odorata*, and (3) South American *C. odorata* plus allied taxa (*C. fissilis*, *C. nebulosa*, and *C. saltensis*; Figure 1).

Notably, ADMIXTURE at $K = 3$ (Figure 1) showed that modern *C. odorata* is best modeled as arising from two ancestral groups. One ancestral population is composed of *C. odorata* specimens that are primarily distributed across Nicaragua and Costa Rica on the northwestern side of the Cordilleras de Talamanca and the Central Mountains (i.e., Mesoamerica, with the exception of one individual from Ecuador; *C. odorata* 291; Figure 2A, shown by a yellow pie chart in Ecuador). This group forms a strongly supported clade (Figure 1, Clade B). *Cedrela odorata* samples from southeast Costa Rica, Panama, and South America are

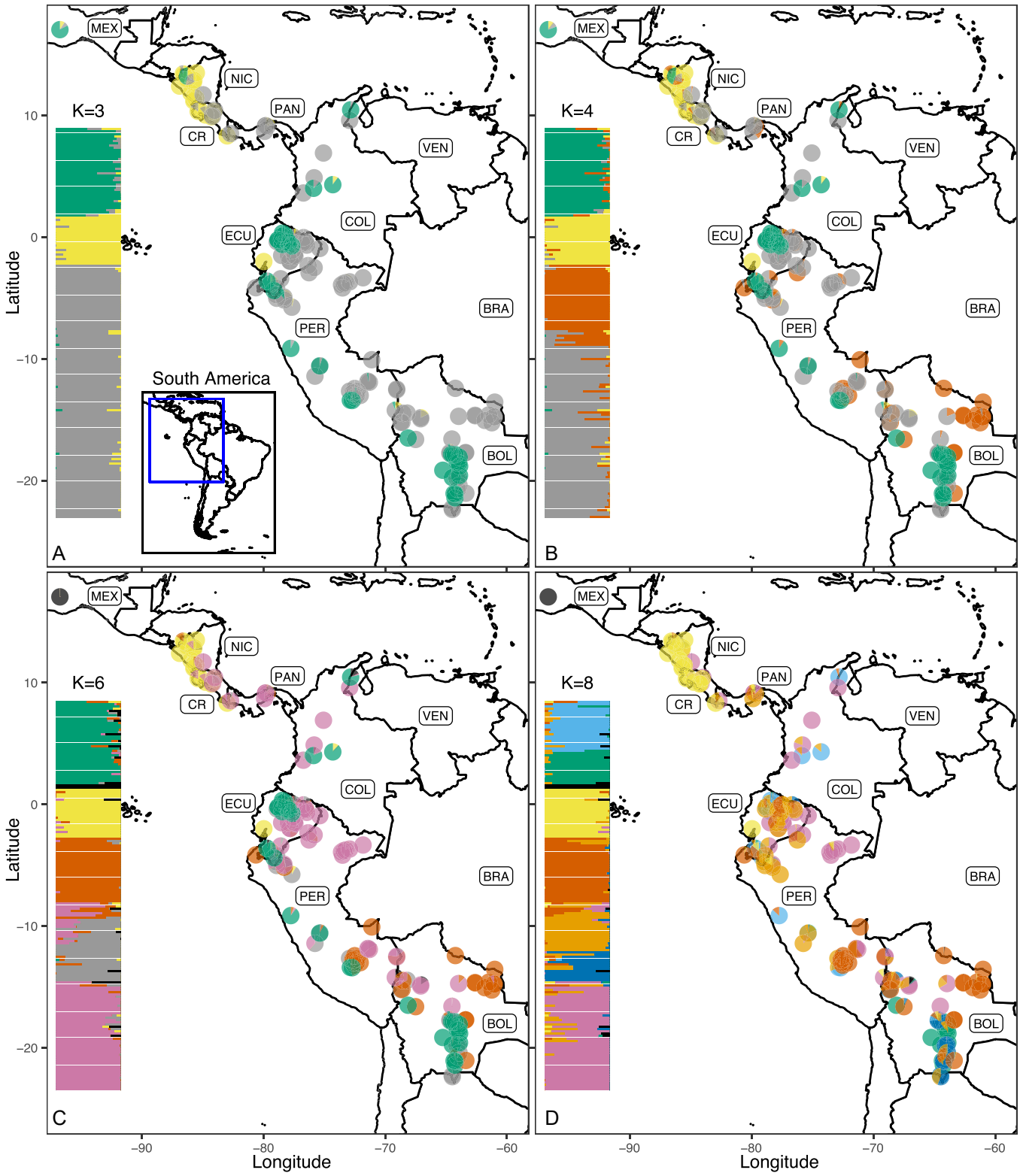

FIGURE 2 Maps show the geographic distribution of ancestral populations as pie charts. The legend corresponds to the ADMIXTURE models presented in Figure 1: (A) $K = 3$, (B) $K = 4$, (C) $K = 6$, (D) $K = 8$.

estimated to have arisen from a separate ancestral source that also included all samples from *C. fissilis*, *C. nebulosa*, and *C. saltensis* (Figure 1, shown by gray bars under $K = 3$). The majority of specimens identified as *C. angustifolia* and *C. montana* form a highly supported clade (Figure 1, Clade A, turquoise bars) that is sister to the rest of *Cedrela*, and these specimens were estimated to have arisen from a third ancestral source.

An ancestral cluster representative of *C. fissilis* becomes evident at $K = 4$ groups (lnL = −108,324.45), and 28 of the specimens with an ancestry proportion greater than 50% for this group (34 total) were present on Clade C (Figure 1) and were mostly *C. fissilis* distributed across southern Peru and northern Bolivia (Figure 2B). This admixture class (Figure 1, orange bars under $K = 4$) was also observed in some South American specimens identified as "*C. odorata*" present in a grade between Clades C and D (e.g., *C. odorata* 18, 26, 161, 74, 172, 204, 107; Figure 1).

The model for $K = 6$ groups (lnL = −96,295.81) highlights two additional ancestral groups with strong support. The first is formed by two samples: CEOD-NYBG (*C. odorata* from Mexico; Figure 2C) and *C. odorata* (179) from Nicaragua (Figure 1, black bars under $K = 6$). These individuals form a grade between *C. angustifolia*/*C. montana* (Clade A) and the rest of *Cedrela* (Figure 1; Clades B, C, D, E, and F), and their resolution was strongly supported across bootstrap replicates. We explored the identity of these individuals with additional sequencing and analysis; preliminary results suggest that they are likely to be misidentified specimens of one of the Mesoamerican endemic *Cedrela* species, rather than *C. odorata* (Finch, 2019b). The $K = 6$ model also shows the genetic distinctiveness of ancestral groups corresponding to South American *C. odorata* (pink bars) and *C. nebulosa*/*C. saltensis* (gray bars), and some specimens identified as *C. fissilis*, *C. odorata*, and one *C. montana*, that may be misidentifications (Figure 1, also gray). The gray "*C. nebulosa*/*C. saltensis*" group forms a strongly supported clade that includes all the herbarium-identified *C. nebulosa* and nearly all *C. saltensis* (Figure 1, Clade D).

Increasing K to 8 groups (lnL = −91,055.11) maintained similar patterns in ancestral population arrangement along the species tree but showed that Clades A and D were represented by two ancestral populations each (Figure 1). $K = 8$ revealed the distinctiveness of *C. nebulosa* and *C. saltensis* specimens appearing in Clade D (Figure 1, gold and navy bars, respectively, under $K = 8$), dividing them into separate ancestral populations. The majority of *C. nebulosa* specimens originated from higher latitudes than *C. saltensis* specimens, which originated in Bolivia (Figure 2D). $K = 8$ also reflected to the taxonomic distinction represented by *C. angustifolia* and *C. montana* (Figure 1, turquoise and light blue bars, respectively). As observed with *C. nebulosa* and *C. saltensis*, the genetic divergence in *C. angustifolia* and *C. montana* is associated with geographic separation (Figures 1 and 2D).

Species delimitation with BFD*

Model C in Figure 3 depicts current taxonomy with six *Cedrela* species, and it shows the lowest mean marginal likelihood estimate compared to all other species delimitation schemes tested with BFD*, indicating that current taxonomy is the poorest representation of these genomic data relative to all alternative classification models tested (Table 1, Figure 3). The simplest alternative model 4.1 with five taxa (Figure 3) provided the best fit of the data in three of four replicates (i.e., replicated for individuals), yielding the highest estimated mean marginal likelihood (−6984.09; Table 1) and the highest support when compared to current taxonomy (log Bayes factors ranged from −242 to −164; Table 1, Figure 3). A difference of greater than 10 log Bayes factors indicates very strong evidence that the model with a greater marginal likelihood estimate (MLE) is a better representation of the data (Leaché and Ogilvie, 2018). In Model 4.1, *C. odorata* from Mesoamerica and South America are separate taxa, *C. nebulosa* and *C. saltensis* are part of the South American *C. odorata* group, and *C. angustifolia* and *C. montana* are combined as a one taxon. Appendix S5 shows a revised geographic distribution of our specimens under the taxonomic splits of model 4.1.

Model 1.3 provided the second-best representation of the data, and this model favored the recognition of *C. odorata* from Mesoamerica and South America as separate taxa, *C. nebulosa* and *C. saltensis* as separate taxa from South American *C. odorata*, and *C. angustifolia* and *C. montana* as one taxon (See Appendix S6 for geographic distribution). Interestingly, model 1.3 had a lower average marginal likelihood across replicates but showed higher support for one set of individuals (Table 1), suggesting that intraspecific variation may affect the results of BFD* for taxa showing the smallest divergences. For example, direct comparison of models 4.1 and 1.3 showed that model 4.1 best represented the data for individual sets 1 and 2; model 4.1 was a slightly better representation of the data for individual set 3 (Table 2); conversely, model 1.3 was the best representation of the data for individual set 4 (Table 2).

Overall, the modest difference in relative support across the lowest-ranking models (model 2, 3, and 2.3) compared to model C indicates the limited effect of dividing or combining South American *Cedrela* lineages (*C. angustifolia*, *C. montana*; *C. nebulosa*, *C. saltensis*) relative to the larger effect of splitting *C. odorata* into Mesoamerican and South American taxa (Table 1, Figure 3). All models that classify Mesoamerican *C. odorata* and South American *C. odorata* as distinct taxa provide convincing evidence for the existence of cryptic species.

Cedrela species divergence

We used SNAPP and BEAST2 to sample time-calibrated species trees with 30 *Cedrela* and *S. mahagoni* individuals and 1000 SNP loci. Species tree topologies were nearly congruent across sampling iterations, and nearly all nodes on the maximum clade credibility tree (Figure 4) showed posterior probabilities of 1.0, excepting the node separating *C. nebulosa* and *C. saltensis*. These two species are closely related to each other and to South American *C. odorata*, showing approximately 55% posterior probability (Figure 4). The clade containing *C. odorata*, *C. nebulosa*,

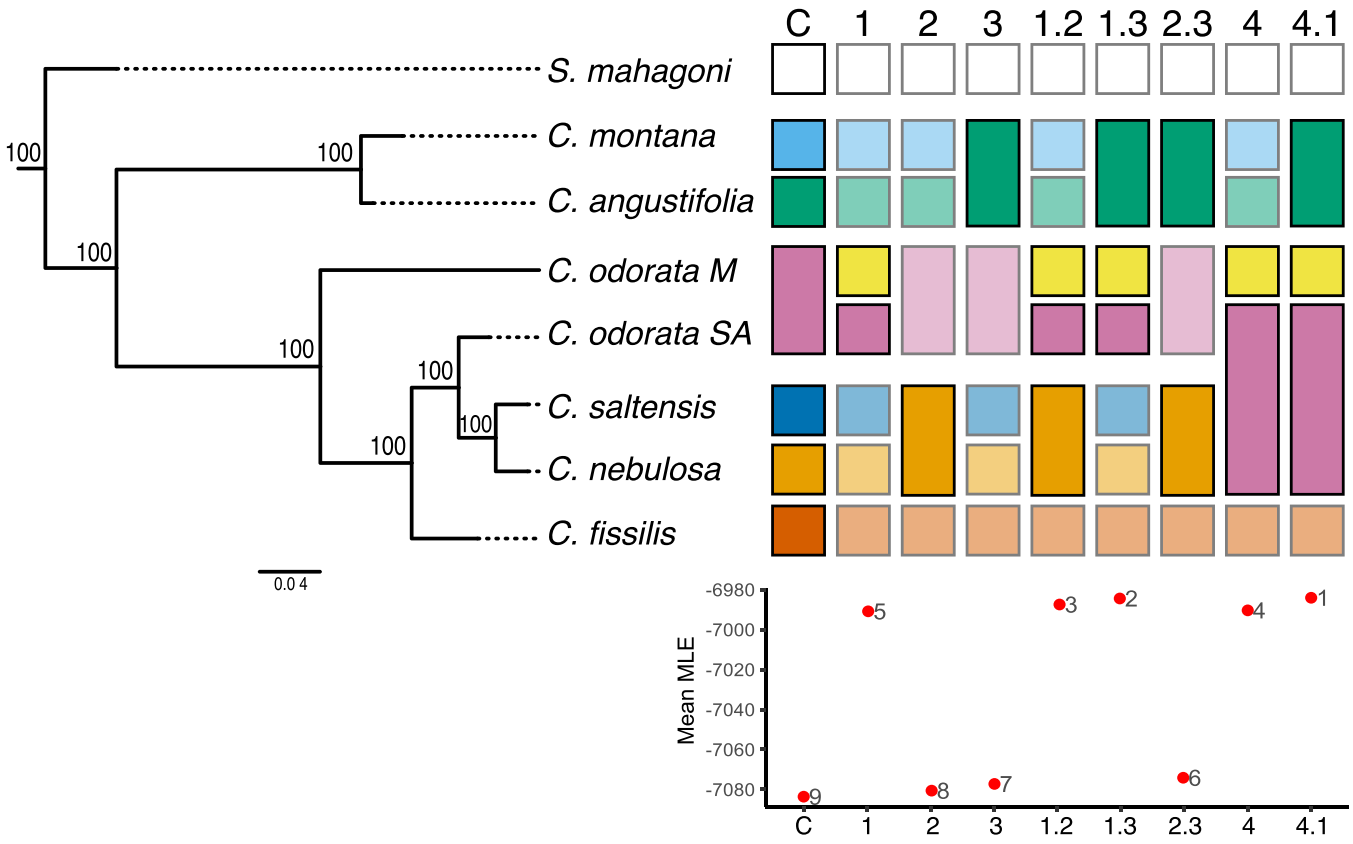

FIGURE 3 Species delimitation models tested with BFD* and their marginal likelihood scores shown in the scatter plot. The phylogeny is a reduced version of the maximum likelihood species tree generated with one specimen from each taxon and RAxML via the CIPRES server with 1000 bootstrap replicates and *Swietenia mahagoni* as the outgroup. *Cedrela odorata* M represents *C. odorata* s.s., and *C. odorata* SA represents *C. odorata* from southeastern Costa Rica, Panama, and South America. Colored blocks show different taxonomic concepts for *C. odorata* and allied species. The scatterplot shows estimated mean marginal likelihood estimates (MLE) across BFD* replicates for each species delimitation model (Table 1). Numbers near mean MLE estimates are model rankings with number 1 indicating the model that we determined to be the best representation of the data (but also see Table 2).

TABLE 1 BFD* results summary. The mean marginal likelihood estimates (MLE) for each species delimitation model were generated by estimating the mean MLE across BFD* replicates with four sets of individuals (set). Log Bayes factors estimates reported for each set.

Model	No. of taxa	Mean MLE	MLE range	2 ln (Bayes factors) by set			
				1	2	3	4
4.1	5	−6984.09	−7173.81 to −6840.07	**−179**	**−242**	**−211**	−164
1.3	7	−6984.44	−7177.98 to −6841.90	−171	−238	−210	**−174**
1.2	7	−6987.46	−7179.63 to −6845.24	−168	−231	−209	−161
4	6	−6990.39	−7180.38 to −6846.48	−166	−229	−201	−149
1	8	−6990.88	−7185.13 to −6846.07	157	230	197	159
2.3	5	−7074.15	−7252.61 to −6934.32	−22	−16	−24	−14
3	6	−7077.24	−7257.16 to −6934.86	−13	−11	−15	−13
2	6	−7080.60	−7259.44 to −6940.27	−8	−1	−12	−2
Current taxonomy	7	−7083.62	−7263.44 to −6941.46	–	–	–	–

and *C. saltensis* resolved as sister to the South American taxon *C. fissilis*. Together, these species represent the majority of specimens originating from South America. The sister lineage to this South American clade is "*C. odorata* M," (*C. odorata* specimens from Mesoamerica [Nicaragua, northwestern Costa Rica]) identified at $K = 3$ in the ADMIXURE analysis (yellow ancestral group; Figure 1, Clade B). In agreement with the SNPhylo nuclear species

tree, the species complex represented by *C. angustifolia* and *C. montana* resolved as sister to the rest of the *Cedrela* in this data set (Figure 4).

TABLE 2 Head-to-head comparison of models 1.3 and 4.1. Evidence that a species delimitation model is a better representation of the data can be assessed with log Bayes factors. If 2 ln (Bayes factors) equals: (1) <2 weak evidence, (2) >2 positive evidence, (3) 5–10 strong evidence, (4) >10 very strong evidence that the model with a greater MLE is a better representation of the data (Leaché and Ogilvie, 2018).

Set	MLE model 1.3	MLE model 4.1	2 ln (Bayes factors) or $2(MLE_{1.3} - MLE_{4.1})$
1	−7177.98	−7173.81	−8
2	−6841.90	−6840.07	−4
3	−7063.67	−7062.93	−1
4	−6854.22	−6859.53	11

BEAST2 estimated a median age of 40.13 Ma for the divergence of *S. mahagoni* and *Cedrela* (95% highest posterior density [HPD] interval 42.44–38.23 Ma; Figure 4) for the time of divergence of *S. mahagoni* and *Cedrela*. The median age of the divergence of *C. angustifolia* and *C. montana* from the rest of *Cedrela* was estimated at 18.95 Ma (95% HPD interval 20.13–17.59 Ma; Figure 4). We estimate that our *C. angustifolia* and *C. montana* samples likely diverged from each other comparatively recently (median age 750,000 yr; 95% HPD interval 1.22–0.18 Ma; Figure 4). *Cedrela odorata* from Mesoamerica likely diverged from the South American clades around 9.67 Ma (95% HPD interval 11.05–7.95 Ma; Figure 4). *Cedrela fissilis* diverged from *C. odorata* SA, *C. nebulosa*, and *C. saltensis* around 4.85 Ma (95% HPD interval 5.74–3.94 Ma; Figure 4). The divergence of *C. odorata* SA from *C. nebulosa* and *C. saltensis* likely occurred in the last 2.70 Myr (95% HPD interval 3.25–2.00 Ma; Figure 4), and in the same interval

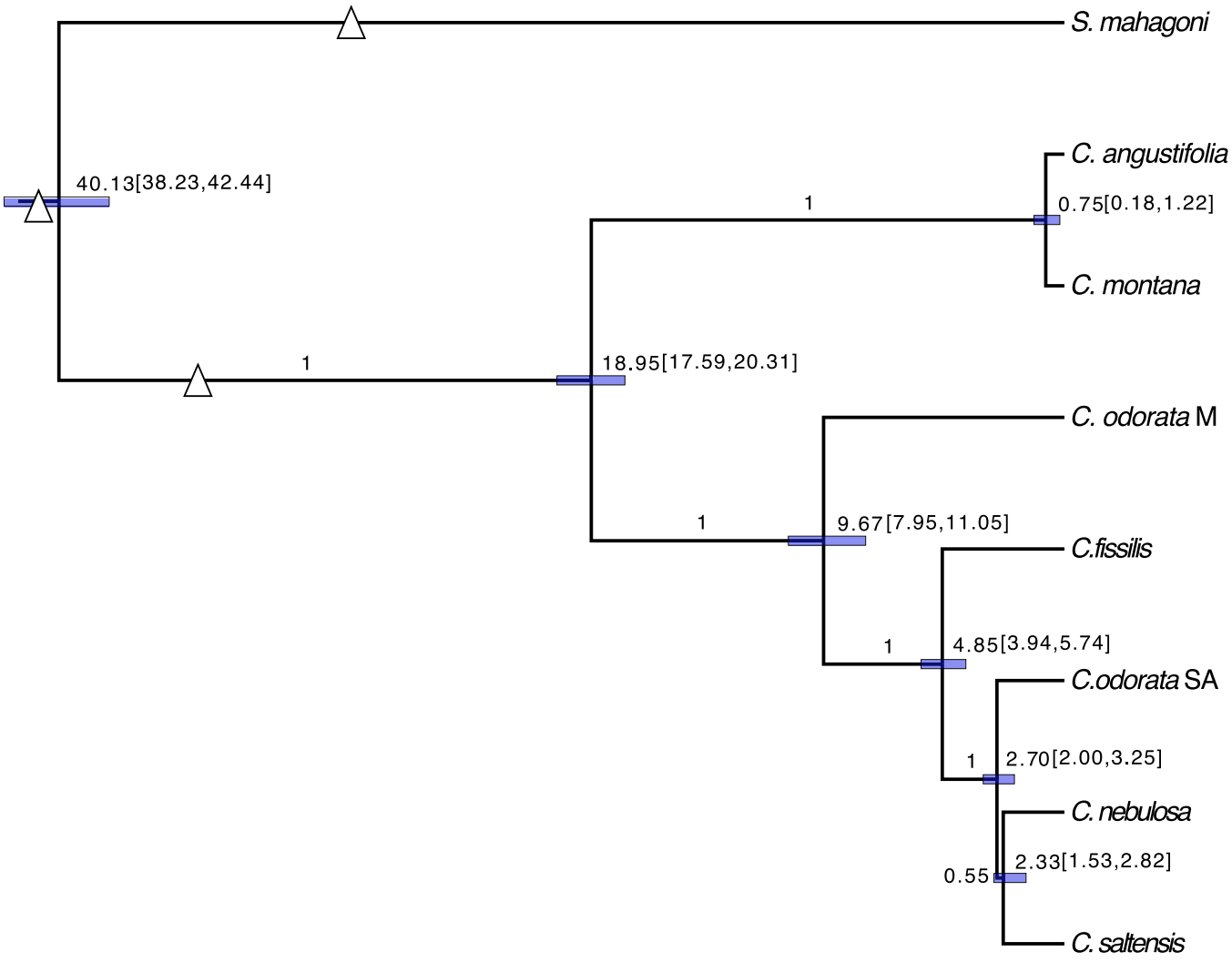

FIGURE 4 Maximum clade credibility tree for 30 *Cedrela* and *Swietenia mahagoni* specimens via SNAPP and BEAST2. Numbers on the branches indicate posterior probabilities for species tree topology. Numbers at the right of nodes indicate median node ages. The 95% HPD interval around node ages are bracketed and indicated by node bars. Triangles on branches indicated time calibration positions 51.9 Ma corresponding to the age of the stem node of the Cedreloideae, and 31.05 Ma corresponding to the stem age of Cedreleae, and 24.25 Ma constraining constrained the stem node leading to the outgroup, *S. mahagoni*.

our *C. nebulosa* and *C. saltensis* samples also diverged (median age 2.33 Ma; 95% HPD interval 2.82–1.53 Ma; Figure 4).

DISCUSSION

Cryptic species within *C. odorata*

Using a dense sample of unlinked genomic SNPs from multiple individuals, we found evidence supporting the existence of cryptic species within *C. odorata*. Herbarium specimens identified as *C. odorata* clearly derived from two divergent ancestral populations (Figure 1), with one group of descendants restricted to Mesoamerica and a second group of descendants more widely distributed across Panama through South America. Species delimitation modelling strongly favors all species delimitations that treat these lineages as separate taxa (Table 1, Figure 3). These results add compelling genomic evidence to previous reports of a quantitative genetic discontinuity within *C. odorata* that is localized near the central mountain ranges of Costa Rica (Figures 1, 2; Gillies et al., 1997; Navarro et al., 2002, 2004; Muellner et al., 2009, 2010; Cavers et al., 2013). Our *C. odorata* specimens from Nicaragua and northwestern Costa Rica formed a genetic cluster that was separate from, and sister to, all *C. odorata*, *C. fissilis*, *C. nebulosa*, and *C. saltensis* specimens collected in southeastern Costa Rica, Panama, and South America (Figures 1 and 2). Given the totality of evidence, we conclude that trees from Nicaragua and northwestern Costa Rica (Figure 1, Clade B) are likely representative of *C. odorata* s.s. (Muellner et al., 2010).

Navarro et al. (2002) described quantitative and physiological traits that differ in response to aridity and drought (e.g., seed size and mass, leaf size, growth rate); these traits appear to differentiate Mesoamerican (*C. odorata* M) and South American *C. odorata* (*C. odorata* SA) into distinct groups when plants are grown in common gardens in Costa Rica. The strong quantitative and adaptive differentiation between these groups led the authors to suggest that *C. odorata* may be in the process of "forming new species in Costa Rica." We find consistently high genomic divergence among these groups (Figure 1), a pattern that contradicts a hypothesis of "nascent divergence". Instead, our data support the initial suggestion by Cavers et al. (2003b) and subsequent examination of the fossil record (Koecke et al., 2013; Muellner-Riehl and Rojas-Andrés, 2022), that considers Costa Rica a region of secondary contact between two morphologically similar *Cedrela* taxa that diverged more distantly in the past (Figure 4).

If the Mesoamerican lineage represents the distinctive *C. odorata* s.s., then what is the southern lineage? In recent phylogenetic analyses of *Cedrela*, Muellner et al. (2009) and Cavers et al. (2013) included one French Guianan *C. odorata* (collection *Mori* et al., *19208* Kew; GenBank FJ462463, FJ462495, FJ462527) that provided preliminary evidence of cryptic species in *C. odorata*. This specimen phylogenetically resolved as closely related to *C. odorata* specimens from Costa Rica, Panama, and Ecuador and was a close relative to South American *C. saltensis* and *C. nebulosa*. Our *C. odorata* specimens from southeastern Costa Rica, Panama, and across South America show similar phylogenetic affinities as this French Guianan specimen. It is possible that this lineage is related to the oldest *Cedrela odorata* specimen on record originating in South America—specifically, *Cedrela guianensis* A.Juss.—which was collected from French Guiana in 1830 (Smith, 1960; Pennington and Muellner, 2010) and later determined to be synonymous with *C. odorata* based on morphology (Smith, 1960). Given the strong genetic distinction observed within the two lineages of *C. odorata*, modest population sampling from French Guiana or surrounding regions could reveal whether our specimens from the cryptic South American "*C. odorata*" and specimens from near the site of Adrien De Jussieu's *C. guianensis* are the same lineage.

Recognizing cryptic species in taxonomy is a contentious issue that has no widespread consensus. A recent review (Struck et al., 2018) estimated that 116 of 600 research articles ascribed new taxonomic concepts to cryptic species; of this total, seven considered plant species. For plants, some have suggested annotating identification keys for the species to indicate that a set of morphological characters describes two distinct genetic lineages, but that cryptic species should not be identified by name unless distinguishing morphological or anatomical characters can be identified (Ross, 1974; Paris et al., 1989). Taxonomic reassignment of South American *C. odorata* (to *C. guianensis* or another taxon) would remedy the paraphyletic resolution of *C. odorata* and better reflect the monophyletic ancestry of *C. odorata* and allied species in the neotropics. Given the genetic distinctiveness of these lineages, re-evaluation of morphological characters from taxa that have been genotyped and identified to lineage may help identify diagnostic traits that were overlooked in the past. We recently developed a SNP assay that can be used to classify *C. odorata* to lineage (e.g., *C. odorata* s.s. versus *C. odorata* SA) with a high degree of accuracy (Finch et al., 2020), and this assay should enable such a comparison.

In contrast to findings presented in prior studies, the presence of additional cryptic taxa within South American *C. odorata* were not detected in this study. Muellner et al. (2009) described "cryptic" groups represented by (1) the two specimens from French Guiana and Ecuador (see above) and (2) two specimens from Brazil and Venezuela. Later, Cavers et al. (2013) hypothesized that the Andes range formed a barrier to gene flow and that two cryptic species were isolated on each side of the Andes, with one corresponding to the cryptic taxon proposed by Muellner et al. (2009). Our genome-wide sampling shows that *C. odorata* specimens from Colombia and Venezuela resolve similarly and in a strongly supported clade that includes individuals from Ecuador and Peru on the Amazonian side

of this hypothetical vicariance barrier (Figure 1, Clade E). Unfortunately, our sampling did not include many *C. odorata* specimens from the Pacific slope of the Andes in Ecuador and Colombia. Future studies should include these areas before ruling out the existence of additional cryptic taxa within *C. odorata*.

BFD* supports a simplified *Cedrela* taxonomy

In addition to estimating higher marginal likelihood scores for models showing the separation of *C. odorata* into two species (*C. odorata* s.s. and *C. odorata* SA), our results indicate that *C. angustifolia* and *C. montana* show weak genetic differentiation and could potentially be recognized as a single taxon (Table 1, Figure 3). These species have been regarded as separate species since the description of *C. montana* in 1858 (Pennington and Muellner, 2010). Our genome-wide data reveal these taxa derived from a single ancestral lineage at large values of K (e.g., $K = 6$; Figure 1), and show a best fit with species delimitation models that treat them as one species (model 4.1; Table 1, Figure 3). Pennington and Muellner (2010) reported that the vast majority of *C. angustifolia* and *C. montana* specimens can be placed without difficulty based on leaflet width, shape, bases, and fruit size except for those in southern Ecuador and northern Peru where they coexist (Pennington and Muellner, 2010). We identified three specimens originating in Peru—a zone of overlap between *C. angustifolia* and *C. montana* (Pennington and Muellner, 2010)—that appear to be *C. montana* forms misidentified as *C. angustifolia*. *Cedrela angustifolia* and *C. montana* may represent a cline of morphological and genetic variation. To date, genetic evidence supporting their taxonomic distinctiveness relies on evidence from linked plastid genes (Pennington and Muellner, 2010; Cavers et al., 2013). Our previous phylogenetic analysis based on chloroplast genomes also showed distinct haplotypes and pronounced geographic structuring (Finch, 2019b). However, our results from 8941 unlinked nuclear genomic SNPs fail to support this distinction and instead suggest that a reasonable interpretation is to consider these taxa as infrataxa of a single species. If taxonomy were simplified to reflect this finding, *C. angustifolia* has historical priority (Pennington and Muellner, 2010), and intraspecific taxon names could be applied to recognize the distinctive morphologies at the northern (*C. angustifolia* subsp. *angustifolia*) and southern (*C. angustifolia* subsp. *montana*) geographic extremes of this lineage.

Our interpretation of the specific status of *C. nebulosa* and *C. saltensis* is also complicated by variable results across replicate analyses. These species are reported to be morphologically and ecologically distinct from each other and from *C. odorata*, differing primarily in density of leaf indumentum and petiole length (Pennington and Muellner, 2010). However, 8941 unlinked nuclear SNPs showed that these species resolve as well-supported lineages nested within *C. odorata* SA (Figure 1: Clade D) and that they derive from the same ancestral population as *C. odorata* SA (Figure 1). Species delimitation models that treat these taxa conspecific with *C. odorata* SA (Figure 3) showed the highest marginal likelihoods in three of four BFD* replicates (model 4.1; Table 1), while one of four replicates showed higher support for treating *C. nebulosa* and *C. saltensis* as distinct species (model 1.3; Table 2). It is possible that we selected specimens that show stronger genetic structure in one of four replicate sets (Table 2; set 4). It is important to note that multispecies coalescence analysis based on dense genomic data typically use a small number of individuals (often one). Our use of four combinations of individuals in multispecies coalescence analysis was made possible by the availability of multiple high-quality specimens, a large number of unlinked SNPs, and substantial computing resources. This analysis highlights the genotypic variability present within these taxa and the potential pitfalls of using small numbers of individuals or genomic targets to infer the phylogenetic history of this complex group.

A criticism of multispecies coalescence methods such as BFD* is that they can "over-delimit" species by identifying population structure that results in taxon splitting and taxonomic inflation (Sukumaran and Knowles, 2017; Stanton et al., 2019; but also see Leaché et al., 2019). In contrast, our application of this method favors lumping these nominal taxa into one species (*C. odorata* SA; i.e., *C. guianensis*), or alternatively adopting the use of intraspecific taxon names for *C. nebulosa* and *C. saltensis* to represent genetic lineages within South American *C. odorata* (i.e., *C. guianensis* subsp. *nebulosa* and *C. guianensis* subsp. *saltensis*; model 4.1). Given these findings and recent estimated divergence times of *C. nebulosa* and *C. saltensis* and their divergence from *C. odorata* SA within the divergence interval of *C. odorata* SA from *C. fissilis* (Figure 4), we think *C. nebulosa* and *C. saltensis* may be morphotypes of *C. odorata* SA that could be regarded as subspecies. In our view, model 4.1 is a more conservative conclusion on the delimitation of these species, especially given that BFD* is sensitive to population structure.

It is possible that ecological isolation led to the morphological differentiation observed among our specimens labeled *C. nebulosa* and *C. saltensis*. The morphological distinctiveness of these two taxa appears supported by our genomic data because we found only one clearly misclassified *C. saltensis* specimen (*C. saltensis* 233, which we predict is *C. fissilis*; Figure 1, clade C). However, genotypes with admixture proportions nearly identical to *C. nebulosa* and *C. saltensis* are found in specimens classified as *C. fissilis*, *C. odorata*, and even *C. montana* (Figure 1, clade D), pointing to the practical difficulty of applying taxonomic keys to these taxa. One or more factors—recency of divergence, long generation time, or gene flow in the ancestral lineages—may have led to the observed pattern where *C. nebulosa* and *C. saltensis* resolve as a clade within *C. odorata* SA (Rieseberg and Brouillet, 1994; De Queiroz, 2007; Knowles and Carstens, 2007; Frankham et al., 2012;

Naciri and Linder, 2015; Pennington and Lavin, 2016). This pattern of well-supported lineages nested within geographically widespread and paraphyletic taxa is common across neotropical rainforest tree species (Dexter et al., 2010; Fine et al., 2014; Winterton et al., 2014). Future genetic and morphological studies in *Cedrela* should include more individuals and populations conforming to the morphological descriptions of *C. nebulosa* and *C. saltensis* to evaluate the distinctiveness of these lineages and their relationship to *C. odorata* SA.

Cedrela in South America

Muellner-Riehl and Rojas-Andrés (2022) provided a detailed review of the *Cedrela* fossil record that illustrates the boreotropical origins of the genus supported by Eocene fossils in North America (e.g., Figure 1 of Muellner-Riehl and Rojas-Andrés, 2022). They suggest that *Cedrela* became restricted to Central and South America by the early Oligocene following its extinction in North America driven by climate cooling. Multiple molecular reconstructions (Muellner et al., 2009, 2010; Koecke et al., 2013; Koenen et al., 2015; Muellner-Riehl and Rojas-Andrés, 2022) and fossil evidence (specifically *Cedrela* from Salto de Tequendama, Colombia) (Hooghiemstra et al., 2006), support ecological diversification of *Cedrela* in the more recent Miocene, leading to extant lineages and a contemporary distribution in Central and South America. This timeframe coincides with hypotheses favoring an early (i.e., 23 Ma) and intermittent closure of the Central American Seaway and with evidence of early Miocene emergence of the Isthmus of Panama (Cody et al., 2010; Bacon et al., 2015; Montes et al., 2015; McGirr et al., 2020). Our divergence estimates and interpretations add to a growing list of studies finding that show neotropical forest are composed of plant lineages that colonized South America from elsewhere during the Cenozoic (Pennington and Dick, 2004; Dick and Pennington, 2019). A recent model and illustration of the Miocene Central American Seaway (fig. 6 of McGirr et al., 2020) accompany multiple reviews of the fossil record and molecular clock analyses (Cody et al., 2010; Bacon et al., 2015) that support Miocene colonization of South America by plant families from Central America via early land emergence or long-distance dispersal.

Cedrela angustifolia and *C. montana* represent one of the earliest extant lineages in *Cedrela*, forming a monophyletic lineage that is sister to the remainder of the genus. We estimated that the common ancestor of these two *Cedrela* lineages diverged approximately 18.95 Ma (Figure 4). This dispersal event in South America would require a migration path on land (e.g., via an intermittent "proto-Isthmus"; McGirr et al., 2020) or long-distance dispersal, possibly aided by the winged, wind-dispersed seeds characteristic of the genus. Our combination of robust genomic and population sampling highlights the large genetic distance that separates this distinctive lineage from the remainder of South American *Cedrela*. The magnitude of genetic divergence in this lineage could be indicative of long-distance dispersal across high-resistance migration routes, such as the mid-Miocene Central American Seaway, followed by prolonged isolation.

The next major radiation in *Cedrela* occurred 9.67 Ma (Figure 4) when Mesoamerican *C. odorata* s.s. (*C. odorata* M) diverged from the lineage that gave rise to South American lineages of *C. odorata* SA, *C. fissilis*, *C. nebulosa*, and *C. saltensis*. Given recent reconstructions (McGirr et al., 2020), the Isthmus of Panama may have been readily traversed during this period, allowing a lineage related to *C. odorata* M to expand into South America, which then ultimately radiated and gave rise to all South American taxa included in this analysis. Given the centrality of Mesoamerican and Panamanian populations to the *Cedrela* migration story, future research should focus on this region to clarify the relationships on and around this continental junction.

Our divergence estimates for the remaining lineages largely coincide with Koecke et al. (2013), who suggested that South American *Cedrela* diversified within the last 6 Ma. We estimated that the lineages leading to *C. fissilis* and *C. odorata* SA/*C. nebulosa*/*C. saltensis* diverged 4.85 Ma (Figure 4). Additional *Cedrela* species not included here that are more closely related to *C. fissilis* (i.e., *C. balasae* and *C. weberbaurii*), but not included in our analysis, likely diverged in the same interval (Koecke et al., 2013). The divergence event separating *C. odorata* SA from the lineage leading to sister taxa *C. nebulosa* and *C. saltensis* occurred 2.70 Ma (Figure 4) and the divergence of our *C. nebulosa* and *C. saltensis* samples occurred at approximately the same time (2.33 Ma; Figure 4). In our view, the recent divergence of *C. odorata* SA/*C. nebulosa*/*C. saltensis* specimens, combined with frequent contradictions between taxonomic classification and genetic affinities (e.g., Figure 2), make it likely that these three taxa can be less ambiguously defined as one species with two subspecies (Figure 3, Model 4.1) versus the current taxonomy of three separate species (e.g., Figure 3, Models 1 and 1.3). Due to our sparse, individual-based sampling for *C. odorata* and the closely allied species, our power to infer the biogeography of *Cedrela* is limited. Still our analysis shows that *C. odorata* SA, *C. fissilis*, *C. nebulosa*, *C. saltensis*, and unsampled taxa represent one more example of a recent radiation contributing to the high biodiversity of neotropical forests (Richardson et al., 2001; Koenen et al., 2015; Males, 2017; Dick and Pennington, 2019).

Cedrela timber

Recently, the entire genus *Cedrela* was added to the list of CITES Appendix II species at the 18th Conference of the Parties (Tajikistan, 2019). This change means that CITES authorities in exporting countries must confirm the legality and obtain an export permit from the Management Authority of the Contracting State for all *Cedrela*, irrespective of the species-level taxonomy (Contracting

States, 1973; https://cites.org/eng/cop/18/prop/index.php). Before the elevation of *Cedrela* to Appendix II, recognizing the South American lineage of *C. odorata* as a distinct species with a different name (e.g., *C. guianensis*) could have led to the creation of a new taxon not currently protected by CITES, but the new policy change makes it possible to re-examine *Cedrela* taxonomy without concern for the impact new definitions will have on policies aiding its protection.

While elevation to CITES II is intended to deter the harvest and trade of *Cedrela,* it may lead to an increase in illegal *Cedrela* trade since elevated listings signal scarcity or rarity and can increase demand (Challender et al., 2022). Additionally, stricter requirements may encourage concealment of the identity of imports. For example, CITES-II export permits must be based on a scientific "non-detriment" finding or descriptive evidence that the removal of the specimens was not at the detriment of the species (Contracting States, 1973). This finding requires knowledge of the species distribution, ecological role, and practices to promote its regeneration; it also requires increased law enforcement capacity to validate and track legal trade (van der Hout, 2015). In our view, a taxonomic model that more accurately represents genetic lineages, their interactions, and their distributions would make the species names more prescriptive and increase the ability of management authorities to assess non-detriment findings in *Cedrela*. Our preferred species circumscription for our samples (Figure 3, model 4.1) includes five species rather than the current seven, and this simplified set of names could benefit the assessments for non-detriment findings.

While species of *Cedrela* now have heightened protection under CITES II, the discordance between expert identifications on herbarium labels and inferences of genetic affinities based on genome-scale phylogenetic resolutions (Figure 1) highlight the immense challenge of accurately identifying widespread timber species in trade. Of the species tested here, *C. odorata* had the highest rate of misidentification, ranging between 32% (if *C. nebulosa* and *C. saltensis* are treated as *C. odorata*) to 39% (if *C. nebulosa* and *C. saltensis* are considered distinct from *C. odorata*). *Cedrela fissilis* showed a comparable misidentification rate of 31%, with specimens resolving as either *C. odorata* (6) or *C. saltensis* (5). The *C. angustifolia/C. montana* lineage shows the lowest misidentification rate if these species are treated as a single taxon (3%); if treated as separate taxa, *C. angustifolia* has a higher misidentification rate (19%) than *C. montana* (5%). The sources of these herbarium specimens primarily include leaves, but leaves alone usually lack the discriminating features needed to provide accurate classification in *Cedrela*. This highlights the challenge facing managing authorities issuing export permits and importers who are required to provide scientific names for imported *Cedrela* wood products, as wood anatomy has even fewer distinguishing features than leaves. Given that expert identification of *Cedrela* can result in misidentification rates approaching 30% for the most-commonly traded timber species (*C. odorata* and *C. fissilis*), clear challenges exist for governments that seek to enforce regulations that protect *Cedrela* species that may be comingled in a shipment and for companies that are trying to exercise due-diligence and meet requirements for wood legality. Considering the inherent difficulty identifying the visible and hidden complexity demonstrated by *Cedrela*, current estimates of trade in these species are likely to be inaccurate. This problem is not unique to *Cedrela*; many valuable tropical timber genera present even greater taxonomic challenges (e.g., *Shorea* [Dipterocarpaceae], with ~200 species; *Eucalyptus* [Myrtaceae] with ~700 species).

CONCLUSIONS

We re-examined the phylogenetic relationships among six species in the historically overexploited neotropical tree genus *Cedrela*. Our data are best represented by a species delimitation model that (1) recognizes Mesoamerican *C. odorata* s.s. and the previously proposed cryptic taxon containing *C. odorata* from South America as separate entities, (2) regards *C. angustifolia* and *C. montana* as a single species, and (3) combines *C. nebulosa* and *C. saltensis* into South American *C. odorata*. We suggest that *Cedrela* taxonomy be modified to reflect this simplified scheme, perhaps by the application of infrataxa to distinguish morphological and/or geographic distinctness of some nominal species. Our findings suggest that South American *Cedrela* is a young group and that it diversified in the last 19 Myr, potentially following long-distance dispersal from Mesoamerica into South America. The divergence of *C. odorata* s.s. from the other South American species occurred within the last 10 Myr. We found that South American *C. odorata* and *C. fissilis* diverged within the last 4–6 Myr, suggesting that these lineages derived from a common ancestor with *C. odorata* s.s. and represent an immigrant group of species that successfully radiated across much of South America in the recent past. Our findings provide a framework for the assessment of newly described *Cedrela* species that were not included here. Our study joins others in a growing area of research that utilizes natural history collections and genomics to re-examine phylogenetic relationships among closely related neotropical tree species.

AUTHOR CONTRIBUTIONS

K.N.F., R.C.C., and F.A.J. jointly designed the study. K.N.F. and R.C.C. collected data with support from F.A.J. K.N.F. led manuscript preparation with guidance and contributions from R.C.C. and F.A.J.

ACKNOWLEDGMENTS

Funding for this study was provided by U.S. Agency for International Development award 19318814Y0010-140001 to the U.S. Forest Service International Programs and R.C.C.; U.S.D.A. Forest Service Pacific Northwest Research Station, National Science Foundation award DEB 1257976

to F.A.J.; and the Moldenke Endowment (Botany and Plant Pathology Department, Oregon State University) to K.N.F. The authors thank MO, NYBG, and Fairchild Tropical Botanic Garden (FTBG) for aiding in specimen collection, especially James Solomon of MO; Dennis Stevenson, Samantha Frangos, and Lisa DeGironimo of NYBG; and Whitney Meier, Bob Brennan, Michael Hass, and Brett Jestrow of FTBG. We extend thanks to Aaron Liston, Scott Kelchner, and George Kral for providing friendly reviews. We appreciate the support and sequencing services provided by the Center for Quantitative Life Sciences at Oregon State University and the Genomics & Cell Characterization Core Facility at University of Oregon. Finally, we thank the two reviewers and the editor who generously volunteered their time to provide helpful comments and judicious suggestions that improved the quality of our manuscript.

DATA AVAILABILITY STATEMENT

Raw sequence data associated with this study is available from the NCBI GenBank BioProject Archive under accession number PRJNA369105. Complementary data sets with custom scripts used to generate sequencing metrics are available through the Oregon State University Scholars Archive. Access information is available in the Literature Cited section of this article.

ORCID

Kristen N. Finch http://orcid.org/0000-0003-2098-7546

REFERENCES

Alexander, D. H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19: 1655–1664.

Arias, E., and B. Kubicki. 2018. A new moss salamander, genus *Nototriton* (Caudata: Plethodontidae), from the Cordillera de Talamanca, in the Costa Rica–Panama border region. *Zootaxa* 4369: 487–500.

Bacon, C. D., D. Silvestro, C. Jaramillo, B. T. Smith, P. Chakrabarty, and A. Antonelli. 2015. Biological evidence supports an early and complex emergence of the Isthmus of Panama. *Proceedings of the National Academy of Sciences* 112: 6110.

Battey, C. J., and J. Klicka. 2017. Cryptic speciation and gene flow in a migratory songbird Species Complex: insights from the Red-Eyed Vireo (*Vireo olivaceus*). *Molecular Phylogenetics and Evolution* 113: 67–75.

Bickford, D., D. J. Lohman, N. S. Sodhi, P. K. L. Ng, R. Meier, K. Winker, K. K. Ingram, and I. Das. 2007. Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution* 22: 148–155.

Bouckaert, R., and D. Bryant. 2012. A rough guide to SNAPP. Website: https://github.com/BEAST2-Dev/SNAPP/releases/download/v1.2.0/SNAPP.pdf [accessed 10 February 2022].

Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, et al. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLOS Computational Biology* 10: e1003537.

Cadle, J. E. 2012. Cryptic species within the *Dendrophidion vinitor* complex in Middle America (Serpentes: Colubridae). *Bulletin of the Museum of Comparative Zoology* 160: 183–240.

Castañeda-Posadas, C., and S. R. S. Cevallos-Ferriz. 2007. *Swietenia* (Meliaceae) flower in Late Oligocene–Early Miocene amber from Simojovel de Allende, Chiapas, Mexico. *American Journal of Botany* 94: 1821–1827.

Cavers, S., and A. J. Lowe. 2002. Regional scale genetic structure within two Central American tree species: the influence of geography, biology and geological history. *In* B. Degen, M. D. Loveless, and A. Kremer [eds.], Modelling and experimental research on genetic processes in tropical and temperate forests; conference proceedings, Kourou, French Guiana, 2000, 179–190. Embrapa Amazonia Oriental, Belem, Brazil.

Cavers, S., C. Navarro, and A. J. Lowe. 2003a. A combination of molecular markers identifies evolutionarily significant units in *Cedrela odorata* L. (Meliaceae) in Costa Rica. *Conservation Genetics* 4: 571–580.

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

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

Challender, D. W. S., D. Brockington, A. Hinsley, M. Hoffmann, J. E. Kolby, F. Massé, D. J. D. Natusch, et al. 2022. Mischaracterizing wildlife trade and its impacts may mislead policy processes. *Conservation Letters* 15: e12832.

Chandler, M. E. J. 1964. The lower tertiary floras of southern England. *Bulletin of the British Museum (Natural History), Geological Series* 12: 1–151.

Cody, S., J. E. Richardson, V. Rull, C. Ellis, and R. T. Pennington. 2010. The Great American Biotic Interchange revisited. *Ecography* 33: 326–332.

Contracting States. 1973. Text of the Convention on International Trade in Endangered Species of Wild Fauna and Flora. Website: https://cites.org/eng/disc/text.php

Crawford, A. J., E. Bermingham, and P. S. Carolina. 2007. The role of tropical dry forest as a long-term barrier to dispersal: a comparative phylogeographical analysis of dry forest tolerant and intolerant frogs. *Molecular Ecology* 16: 4789–4807.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27: 2156–2158.

De Queiroz, K. 2007. Species concepts and species delimitation. *Systematic Biology* 56: 879–886.

Dexter, K. G., T. D. Pennington, and C. W. Cunningham. 2010. Using DNA to assess errors in tropical tree identifications: How often are ecologists wrong and when does it matter? *Ecological Monographs* 80: 267–286.

Dick, C. W., K. Abdul-Salim, and E. Bermingham. 2003. Molecular systematic analysis reveals cryptic tertiary diversification of a widespread tropical rain forest tree. *American Naturalist* 162: 691–703.

Dick, C. W., and R. T. Pennington. 2019. History and geography of neotropical tree diversity. *Annual Review of Ecology, Evolution, and Systematics* 50: 279–301.

Doan, T. M., A. J. Mason, T. A. Castoe, M. Sasa, and C. L. Parkinson. 2016. A cryptic palm-pitviper species (Squamata: Viperidae: *Bothriechis*) from the Costa Rican highlands, with notes on the variation within *B. nigroviridis*. *Zootaxa* 4138: 271–290.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–1797.

Elias, P. 2012. Logging and the law: How the U.S. Lacey Act helps reduce illegal logging in the Tropics. Union of Concerned Citizens, Cambridge, MA, USA.

Felsenstein, J. 1993. PHYLIP (phylogeny inference package), version 3.5 c. University of Washington, Seattle, WA, USA.

Ferriss, S. 2014. An analysis of trade in five CITES-listed taxa. The Royal Institute of International Affairs, Chatham House, London, UK.

Finch, K. N. 2018. Dataset for genomic resources for the neotropical tree genus Cedrela (Meliaceae) and its relatives [Data set]. Oregon State University, Corvallis, OR, USA. Website: https://doi.org/10.7267/NV935820Q

Finch, K. N. 2019a. Dataset for predicting the geographic origin of Spanish cedar (Cedrela odorata L.) based on DNA variation (version 1) [Data

set]. Oregon State University Corvallis, OR, USA. Website: https://doi.org/10.7267/TQ57NX45Z

Finch, K. N. 2019b. Genomic resources for phylogenetics, species delimitation, and geographic localization of neotropical tree species Cedrela odorata L. (Meliaceae). Ph.D. dissertation, Oregon State University, Corvallis, OR, USA.

Finch, K. N., R. C. Cronn, M. C. Ayala Richter, C. Blanc-Jolivet, M. C. Correa Guerrero, L. De Stefano Beltrán, C. R. García-Dávila, et al. 2020. Predicting the geographic origin of Spanish cedar (*Cedrela odorata* L.) based on DNA variation. *Conservation Genetics* 21: 625–639.

Finch, K. N., F. A. Jones, and R. C. Cronn. 2019. Genomic resources for the Neotropical tree genus *Cedrela* (Meliaceae) and its relatives. *BMC Genomics* 20: 58.

Fine, P. V. A., F. Zapata, and D. C. Daly. 2014. Investigating processes of neotropical rain forest tree diversification by examining the evolution and historical biogeography of the Protieae (Burseraceae). *Evolution* 68: 1988–2004.

Frankham, R., J. D. Ballou, M. R. Dudash, M. D. B. Eldridge, C. B. Fenster, R. C. Lacy, J. R. Mendelson, et al. 2012. Implications of different species concepts for conserving biodiversity. *Biological Conservation* 153: 25–31.

Garcia, M. G., R. S. Silva, M. A. Carniello, J. W. Veldman, A. A. B. Rossi, and L. O. de Oliveira. 2011. Molecular evidence of cryptic speciation, historical range expansion, and recent intraspecific hybridization in the Neotropical seasonal forest tree *Cedrela fissilis* (Meliaceae). *Molecular Phylogenetics and Evolution* 61: 639–649.

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

Gill, B. A., B. C. Kondratieff, K. L. Casner, A. C. Encalada, A. S. Flecker, D. G. Gannon, C. K. Ghalambor, et al. 2016. Cryptic species diversity reveals biogeographic support for the 'mountain passes are higher in the tropics' hypothesis. *Proceedings of the Royal Society, B, Biological Sciences* 283: 20160553.

Gillies, A. C. M., J. P. Cornelius, A. C. Newton, C. Navarro, M. Hernández, and J. Wilson. 1997. Genetic variation in Costa Rican populations of the tropical timber species *Cedrela odorata* L., assessed using RAPDs. *Molecular Ecology* 6: 1133–1145.

Grummer, J. A., R. W. Bryson, and T. W. Reeder. 2014. Species delimitation using Bayes factors: simulations and application to the *Sceloporus scalaris* species group (Squamata: Phrynosomatidae). *Systematic Biology* 63: 119–133.

Helfrich, P., E. Rieb, G. Abrami, A. Lücking, and A. Mehler. 2018. TreeAnnotator: versatile visual annotation of hierarchical text relations. Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018), Miyazaki, Japan.

Hey, J., R. S. Waples, M. L. Arnold, R. K. Butlin, and R. G. Harrison. 2003. Understanding and confronting species uncertainty in biology and conservation. *Trends in Ecology & Evolution* 18: 597–603.

Hooghiemstra, H., V. M. Wijninga, and A. M. Cleef. 2006. The paleobotanical record of Colombia: implications for biogeography and biodiversity. *Annals of the Missouri Botanical Garden* 93: 297–325.

Huamán Mera, A. 2014. Structure and genetic diversity of Cedrela (Meliaceae) on the upper Parana river. Ph.D. dissertation, Universidade Federal de Viçosa, Minas Gerais, Brazil.

Inza, M. V., N. Zelener, L. Fornes, and L. A. Gallo. 2012. Effect of latitudinal gradient and impact of logging on genetic diversity of *Cedrela lilloi* along the Argentine Yungas Rainforest. *Ecology and Evolution* 2: 2722–2736.

Knowles, L. L., and B. C. Carstens. 2007. Delimiting species without monophyletic gene trees. *Systematic Biology* 56: 887–895.

Köcke, A. V., A. N. Muellner-Riehl, O. Cáceres, and T. D. Pennington. 2015. *Cedrela ngobe* (Meliaceae), a new species from Panama and Costa Rica. *Edinburgh Journal of Botany* 72: 225–233.

Koecke, A. V., A. N. Muellner-Riehl, T. D. Pennington, G. Schorr, and J. Schnitzler. 2013. Niche evolution through time and across continents: The story of Neotropical *Cedrela* (Meliaceae). *American Journal of Botany* 100: 1800–1810.

Koenen, E. J. M., J. J. Clarkson, T. D. Pennington, and L. W. Chatrou. 2015. Recently evolved diversity and convergent radiations of rainforest mahoganies (Meliaceae) shed new light on the origins of rainforest hyperdiversity. *New Phytologist* 207: 327–339.

Kometter, R., M. Martinez, A. G. Blundell, R. Gullison, M. Steininger, and R. Rice. 2004. Impacts of unsustainable mahogany logging in Bolivia and Peru. *Ecology and Society* 9: 12.

Leaché, A. D., and R. R. Bouckaert. 2018. Species trees and species delimitation with SNAPP: a tutorial and worked example. Workshop on Population and Speciation Genomics, Český Krumlov, Czech Republic. Website: http://evomicsorg.wpengine.netdna-cdn.com/wp-content/uploads/2018/01/BFD-tutorial-1.pdf

Leaché, A. D., M. K. Fujita, V. N. Minin, and R. R. Bouckaert. 2014. Species delimitation using genome-wide SNP data. *Systematic Biology* 63: 534–542.

Leaché, A., and H. A. Ogilvie. 2018. Bayes factor delimitation of species (* with genomic data; BFD*): a tutorial and worked example. Workshop on Population and Speciation Genomics, Český Krumlov, Czech Republic. Website: https://evomics.org/wp-content/uploads/2016/01/BFDstar-tutorial1.pdf

Leaché, A. D., T. Zhu, B. Rannala, and Z. Yang. 2019. The spectre of too many species. *Systematic Biology* 68: 168–181.

Lee, T.-H., H. Guo, X. Wang, C. Kim, and A. H. Paterson. 2014. SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics* 15: 162.

Li, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27: 2987–2993.

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* https://arxiv.org/abs/1303.3997 [Preprint].

Li, H., and R. Durbin. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26: 589–595.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078–2079.

Lotzkat, S., J.-F. Bienentreu, A. Hertz, and G. Köhler. 2011. A new species of *Anolis* (Squamata: Iguania: Dactyloidae) formerly referred to as *A. pachypus* from the Cordillera de Talamanca of western Panama and adjacent Costa Rica. *Zootaxa* 3125: 1–21.

Lowe, A., B. Jourde, N. Colpaert, C. Navarro, and S. Cavers. 2003. Genetic diversity in Costa Rican populations of mahogany (*Swietenia macrophylla*) II: impact of logging on level, structure and dynamics. *Heredity* 90: 268–275.

MacGinitie, H. D. 1953. Fossil plants of the Florissant beds, Colorado. Carnegie Institution of Washington, Washington, D.C., USA.

Maddison, W. P., and D. R. Maddison. 2017. Mesquite: a modular system for evolutionary analysis, version. 3.31. Website: http://www.mesquiteproject.org

Males, J. 2017. Adaptive variation in vein placement underpins diversity in a major Neotropical plant radiation. *Oecologia* 185: 375–386.

Manchester, S. R. 2001. Update on the megafossil flora of Florissant, Colorado. *In* E. Evanof, K. M. Gregory-Wodzicki, and K. R. Johnson [eds.], Fossil flora and stratigraphy of the Florissant Formation, Colorado, Proceedings of the Denver Museum of Nature & Science, 137–161. Denver Museum of Nature & Science, Denver, CO, USA.

McGirr, R., M. Seton, and S. Williams. 2020. Kinematic and geodynamic evolution of the Isthmus of Panama region: implications for Central American Seaway closure. *GSA Bulletin* 133: 867–884.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, et al. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20: 1297–1303.

Millar, C. I., and W. J. Libby. 1991. Strategies for conserving clinal, ecotypic, and disjunct population diversity in widespread species. *Genetics and Conservation of Rare Plants* 149: 170.

Miller, M. A., W. Pfeiffer, and T. Schwartz. 2011. The CIPRES science gateway: a community resource for phylogenetic analyses. Proceedings of the 2011 TeraGrid Conference: extreme digital discovery, article 41. ACM Digital Library. Website: https://doi.org/10.1145/2016741.2016785

Montes, C., A. Cardona, C. Jaramillo, A. Pardo, J. C. Silva, V. Valencia, C. Ayala, et al. 2015. Middle Miocene closure of the Central American Seaway. *Science* 348: 226.

Muellner, A. N., T. D. Pennington, and M. W. Chase. 2009. Molecular phylogenetics of Neotropical Cedreleae (mahogany family, Meliaceae) based on nuclear and plastid DNA sequences reveal multiple origins of "*Cedrela odorata*". *Molecular Phylogenetics and Evolution* 52: 461–469.

Muellner, A. N., T. D. Pennington, A. V. Koecke, and S. S. Renner. 2010. Biogeography of *Cedrela* (Meliaceae, Sapindales) in Central and South America. *American Journal of Botany* 97: 511–518.

Muellner, A. N., H. Schaefer, and R. Lahaye. 2011. Evaluation of candidate DNA barcoding loci for economically important timber species of the mahogany family (Meliaceae). *Molecular Ecology Resources* 11: 450–460.

Muellner-Riehl, A. N., and B. M. Rojas-Andrés. 2022. Biogeography of Neotropical Meliaceae: geological connections, fossil and molecular evidence revisited. *Brazilian Journal of Botany* 45: 527–543.

Naciri, Y., and H. P. Linder. 2015. Species delimitation and relationships: the dance of the seven veils. *Taxon* 64: 3–16.

Navarro, C., S. Cavers, A. Pappinen, P. Tigerstedt, A. J. Lowe, and J. Merila. 2005. Contrasting quantitative traits and neutral genetic markers for genetic resource assessment of Mesoamerican *Cedrela odorata*. *Silvae Genetica* 54: 281–292.

Navarro, C., F. Montagnini, and G. Hernández. 2004. Genetic variability of *Cedrela odorata* Linnaeus: results of early performance of provenances and families from Mesoamerica grown in association with coffee. *Forest Ecology and Management* 192: 217–227.

Navarro, C., S. Ward, and M. Hernández. 2002. The tree *Cedrela odorata* (Meliaceae): a morphologically subdivided species in Costa Rica. *Revista de Biologia Tropical* 50: 21–29.

Nellemann, C. 2012. Green carbon, black trade: illegal logging, tax fraud and laundering in the world's tropical forests. United Nations Environment Programme, GRID-Arendal, Arendal, Norway.

O'Neill, G. A., I. Dawson, C. Sotelo-Montes, L. Guarino, M. Guariguata, D. Current, and J. C. Weber. 2001. Strategies for genetic conservation of trees in the Peruvian Amazon. *Biodiversity & Conservation* 10: 837–850.

Palacios, W. A., J. Santiana, and J. Iglesias. 2019. A new species of *Cedrela* (Meliaceae) from the eastern flanks of Ecuador. *Phytotaxa* 393: 84–88.

Paris, C. A., F. S. Wagner, and W. H. Wagner. 1989. Cryptic species, species delimitation, and taxonomic practice in the homosporous ferns. *American Fern Journal* 79: 46–54.

Pennington, R. T., and C. W. Dick. 2004. The role of immigrants in the assembly of the South American rainforest tree flora. *Philosophical Transactions of the Royal Society, B, Biological Sciences* 359: 1611–1622.

Pennington, R. T., M. Hughes, and P. W. Moonlight. 2015. The origins of tropical rainforest hyperdiversity. *Trends in Plant Science* 20: 693–695.

Pennington, R. T., and M. Lavin. 2016. The contrasting nature of woody plant species in different neotropical forest biomes reflects differences in ecological stability. *New Phytologist* 210: 25–37.

Pennington, T. D., and A. N. Muellner. 2010. A monograph of *Cedrela* (Meliaceae). DH Books, Milborne Port, UK.

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

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* 81: 559–575.

R Core Team. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Website: https://r-project.org

Richardson, J. E., R. T. Pennington, T. D. Pennington, and P. M. Hollingsworth. 2001. Rapid diversification of a species-rich genus of neotropical rain forest trees. *Science* 293: 2242–2245.

Rieseberg, L. H., and L. Brouillet. 1994. Are many plant species paraphyletic? *Taxon* 43: 21–32.

Ross, H. H. 1974. Biological systematics. Addison-Wesley Publishing, Boston, MA, USA.

Saunders, J., and R. Reeve. 2014. The EU timber regulation and CITES. Chatham House, London, UK.

Scotti-Saintagne, C., C. W. Dick, H. Caron, G. G. Vendramin, E. Guichoux, A. Buonamici, C. Duret, et al. 2013. Phylogeography of a species complex of lowland Neotropical rain forest trees (*Carapa*, Meliaceae). *Journal of Biogeography* 40: 676–692.

Smith, C. E. 1960. A revision of *Cedrela* (Meliaceae). *Fieldiana: Botany* 29: 295–341.

Song, S., J. Zhao, and C. Li. 2017. Species delimitation and phylogenetic reconstruction of the sinipercids (Perciformes: Sinipercidae) based on target enrichment of thousands of nuclear coding sequences. *Molecular Phylogenetics and Evolution* 111: 44–55.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.

Stange, M., M. R. Sánchez-Villagra, W. Salzburger, and M. Matschiner. 2017. Bayesian divergence-time estimation with genome-wide SNP data of sea catfishes (Ariidae) supports Miocene closure of the Panamanian Isthmus. *Systematic Biology* 67: 681–699.

Stanton, D. W. G., P. Frandsen, R. K. Waples, R. Heller, I.-R. M. Russo, P. A. Orozco-terWengel, C.-E. T. Pedersen, et al. 2019. More grist for the mill? Species delimitation in the genomic era and its implications for conservation. *Conservation Genetics* 20: 101–113.

Struck, T. H., J. L. Feder, M. Bendiksby, S. Birkeland, J. Cerca, V. I. Gusarov, S. Kistenich, et al. 2018. Finding evolutionary processes hidden in cryptic species. *Trends in Ecology & Evolution* 33: 153–163.

Styles, B. T. 1981. Swientenioideae. Meliaceae, with accounts of Swietenioideae and Chemotaxonomy, Flora neotropica, 360. New York Botanical Garden, Bronx, NY, USA.

Sukumaran, J., and L. L. Knowles. 2017. Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences, USA* 114: 1607–1612.

Tajikistan, M. 2019. Proposal for amendment of Appendix I or II for CITES CoP18 Prop. 57. Consideration of proposals for amendment of appendices I and II, 1–26. CITES, Colombo, Sri Lanka.

UNEP-WCMC [UNEP-World Conservation Monitoring Centre]. 2015. Overview of CITES Appendix III listings. UNEP-WCMC, Cambridge, UK.

Urrunaga, J. M., A. Johnson, I. D. Orbegozo, and F. Mulligan. 2012. The laundering machine. Environmental Investigation Agency, Washington D.C., USA.

van der Hout, P. 2015. Resource assessment and forest management plan for the CITES-listed species *Cedrela odorata* (red cedar) in Guyana. The International Tropical Timber Organization, Yokohama, Japan. Website: https://www.itto.int/files/itto_project_db_input/3084/Technical/TMT-SPD-014-13-R1-M-Resource%20Assessment.pdf

Winterton, C., J.-E. Richardson, M. Hollingsworth, A. Clark, N. Zamora, and P. RT. 2014. Historical biogeography of the neotropical legume genus *Dussia*: the Andes, the Panama Isthmus and the Chocó. *In* W. Stevens, O. Montriel, and P. Raven [eds.], Paleobotany and biogeography: a festschrift for Alan Graham in his 80th year, 128. Missouri Botanical Garden Press, St. Louis, MO, USA.

Zapater, M. A., E. M. Del Castillo, and T. D. Pennington. 2004. El genero *Cedrela* (Meliaceae) en la Argentina [The genus *Cedrela* (Meliaceae) in Argentina]. *Darwiniana* 42: 347–356.

van Zonneveld, M., E. Thomas, N. P. Castañeda-Álvarez, V. Van Damme, C. Alcazar, J. Loo, and X. Scheldeman. 2018. Tree genetic resources at risk in South America: a spatial threat assessment to prioritize populations for conservation. *Diversity and Distributions* 24: 718–729.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1. Specimen information and sequencing metrics for *Cedrela* data set.

Appendix S2. Map of *Cedrela* data set showing origins of the *Cedrela* specimens originated.

Appendix S3. Table with R packages used and package citations. Citations provided via R function citation().

Appendix S4. Density plot showing sequencing depth of coverage (log2 scale) distribution across enriched targets (contigs from the *C. odorata* 300 nuclear reference) for each species.

Appendix S5. Map shows a revised geographic distribution of our *Cedrela* specimens under the taxonomic splits of Model 4.1.

Appendix S6. Map shows a revised geographic distribution of our *Cedrela* specimens under the taxonomic splits of Model 1.3.

How to cite this article: Finch, K. N., F. A. Jones, and R. C. Cronn. 2022. Cryptic species diversity in a widespread neotropical tree genus: The case of *Cedrela odorata*. *American Journal of Botany* 109(10): 1622–1640. https://doi.org/10.1002/ajb2.16064