Molecular evidence of cryptic speciation, historical range expansion, and recent intraspecific hybridization in the Neotropical seasonal forest tree *Cedrela fissilis* (Meliaceae)

Magali Gonçalves Garcia, Roberta Santos Silva, Maria Antonia Carniello, Joseph William Veldman, Ana Aparecida Bandini Rossi, Luiz Orlando de Oliveira

**Abstract**

Molecular phylogeography can lead to a better understanding of the interaction between past climate events, large-scale vegetation shifts, and the evolutionary history of Neotropical seasonal forests. The endangered timber tree species *Cedrela fissilis* is associated with seasonal forests and occurs throughout South America. We sampled *C. fissilis* from 56 sites across the species' range in Brazil and Bolivia and obtained sequence data for nuclear and chloroplast DNA. Most specimens (149 out of 169) exhibited intraindividual polymorphism for the nuclear internal transcribed spacer (ITS). Cloning and an array of complementary sequence analyses indicated that the multiple copies of ITS were functional paralogs – concerted evolution in *C. fissilis* appeared to be incomplete. Independent Bayesian analyses using either ITS or cpDNA data revealed two separate phylogenetic lineages within *C. fissilis* that corresponded to populations located in separate geographic regions. The divergence occurred in the Early Pliocene and Late Miocene. We argue that climate-mediated events triggered dispersal events and split ancestral populations into at least two large refugial areas of seasonal forest that were located to the east and west of the present day Cerrado. Upon recent climate amelioration, formerly isolated lineages reconnected and intraspecific hybridization gave rise to intraindividual polymorphism and incomplete concerted evolution in *C. fissilis*.

**1. Introduction**

Interest in Neotropical seasonal forests is growing because they are important areas of endemism (Prado, 2000; Pennington et al., 2006) and are threatened by high rates of deforestation due to urban expansion, conversion to agricultural use, and forest fires (Steininger et al., 2001; Fundação SOS Mata Atlântica and INPE, 2009). Moreover, little is known about their evolutionary history. Seasonal forests occur in areas that experience a severe dry season (up to five consecutive months), or in areas where precipitation shows little fluctuation, but where winter temperatures (<15 °C) cause physiological drought (Veloso et al., 1991; Oliveira-Filho and Fontes, 2000).

Although seasonal forests are disjunct in their distribution across the Neotropics (Fig. 1), distant blocks of such forest are related floristically (Bigarella et al., 1975; Ab’Saber, 1977). Where they do occur, seasonal forests are often intermingled with patches of other types of vegetation, which include evergreen forests, savannas, and shrublands (e.g., Pennington et al., 2000; Killeen et al., 2006; Oliveira-Filho et al., 2006). Despite their geographical proximity to other vegetation types, Neotropical seasonal forests contain many phylogenetically unrelated species that are absent from evergreen forests and savannas in neighbouring areas (Prado, 2000; Pennington et al., 2006). The 'Pleistocenic Arc' hypothesis (Prado and Gibbs, 1993) explains these patterns by suggesting that present-day seasonal forests are relicts formed by vicariance of a much larger, single formation that covered South America during the Pleistocene. An alternative hypothesis suggests that rare long-distance dispersal of seasonal forest species might explain the disjunct distribution of floristically related blocks (Gentry, 1982; Mayle, 2004).

The disjunct distribution of seasonal forests at the continental scale, as well as their often patchy distribution at regional scales, suggests that these forests are spatially and temporally dynamic. Indeed, seasonal forests have both contracted (Carnaval and Moritz, 2008; Carnaval et al., 2009) and expanded (Auler et al., 2004; Mayle, 2004) in response to climatic changes during the Quaternary period (Ledru et al., 2007). These shifts in vegetation were triggered by cy-
Cedrela (Meliaceae) is a tree genus that probably evolved in seasonal forests. **Cedrela** is a Neotropical, monophyletic genus whose diversification started in the Oligocene and Early Miocene and intensified in the Late Miocene and Early Pliocene (Muellner et al., 2010). At present, most of the 17 species of **Cedrela** have restricted distributions in seasonal forests of Central and South America; two species, **Cedrela odorata** and **C. fissilis**, are widespread and occur in both moist evergreen forests and seasonal forests (Muellner et al., 2009). The IUCN Red List (IUCN, 2010) categorizes **C. odorata** as ‘Vulnerable A1cd+2cd’ and **C. fissilis** as ‘Endangered A1acd+2cd’ species. Recent molecular studies have identified three genetically distinct entities in **C. odorata** that are morphologically indistinguishable (Muellner et al., 2009).

**Cedrela fissilis** Vell. (Meliaceae), which is known commonly in Brazil as ‘cedro branco,’ is a valuable timber species that has long suffered from overharvesting. **Cedrela fissilis** occurs throughout South America (Pennington et al., 1981), but its presence in Central America (Costa Rica and Panama) is uncertain (Pennington and Muellner, 2010). In Brazil, **C. fissilis** is associated with seasonal forests, but it also occurs in forest–savanna ecotones, gallery forests, and moist evergreen forests that are adjacent to seasonal forests (Carvalho, 1994). Large canopy trees occur at low densities (1–3 trees/ha) in old-growth forests of Southern Brazil (Carvalho, 1994). In secondary forests, densities of **C. fissilis** are often much higher (Smith, 1960). **C. fissilis** exhibits a number of traits that favour cross-pollination and long-distance gene flow and are important to its biogeography. These traits include: protogynous dichogamy; asynchronous floral anthesis among both inflorescences of the same tree and individuals; pollination by moths and bees; and wind-dispersed seeds that are released from high in the forest canopy (Carvalho, 1994).

It is likely that an understanding of the evolutionary history of seasonal forests in the Neotropics will benefit from investigation of the molecular phylogeography of plant species (e.g., Caetano et al., 2008; Oliveira et al., 2010). Molecular phylogeographic studies on *Carapichea ipecacuanha* (Rubiacaeae), an understory shrub with a disjunct distribution in seasonal forests, have suggested the long-term persistence of the species in the Atlantic range. In contrast, the colonization of the Amazonian range by *C. ipecacuanha* seems recent, and probably occurred from a single parental source after a strong genetic bottleneck (Oliveira et al., 2010). In *C. ipecacuanha*, the internal transcribed spacer (ITS) was found to exhibit intraspecific polymorphism in a range-dependent manner. Intraspecific hybridization in the Atlantic range led to specimens with multiple functional copies of ITS, as well as pseudogenes, whereas specimens in the Amazonian range, with a single lineage, lacked polymorphisms in ITS (Queiroz et al., 2011). **Cedrela fissilis** and *C. ipecacuanha* are codistributed throughout much of the range of *C. ipecacuanha* and it is plausible that the two species, although differing with respect to several life history traits and ecological requirements, could have experienced a common history of climatic changes. We anticipated that molecular data from *C. fissilis* would be even more informative than those of *C. ipecacuanha* because of the widespread distribution of the former relative to the latter species: **C. fissilis** occurs in many areas (Carvalho, 1994) where *C. ipecacuanha* is absent (Oliveira et al., 2010).

In this study, we explored the evolutionary history of seasonal forests through molecular analysis of nuclear and chloroplast
genes of specimens of C. fissilis collected from across the species range in Brazil and Bolivia. We also carried out phylogenetic analyses that included the other 16 species of Cedrela. The questions addressed by this research were as follows: (1) Is there any evidence for range-dependent intraspecific polymorphism and pseudogenes of ITS in C. fissilis? (2) Is C. fissilis a monophyletic species relative to the other members of the genus? (3) Does C. fissilis consist of a single evolutionary lineage, or is there sufficient lineage differentiation to indicate cryptic speciation within C. fissilis? (4) To what extent is the distribution of genetic diversity in C. fissilis associated with geography and vegetation type? (5) Is there evidence for the existence of recently colonized areas or refugia? Addressing these questions should shed light on the evolutionary history of seasonal forests and contribute to the elaboration and testing of biogeographic hypotheses to understand better a threatened ecosystem that strongly requires conservation.

2. Materials and methods

2.1. Sampling strategy and DNA extraction

In this study, 169 specimens of Cedrela fissilis, collected from 56 sites in Brazil and Bolivia, were analysed (Fig. 1). The sampling sites were chosen to cover most of the geographical range of the species. Most sites were located within seasonal forests (39 out of 56 sites), whereas others were located in adjacent vegetation (four in savannas; 13 in moist evergreen forest). To define the vegetation formations in which the species occurred, we consulted the Terrestrial Ecoregions of the World database from the World Wildlife Fund (Olson et al., 2001). The habitat classifications for Brazil from the Brazilian Institute of Geography and Statistics (IBGE, 2004) provided details about vegetation at a local scale.

Most sampling sites were located in seasonal forests on either the east or the west of a savanna biome of South America, which corresponds to the open canopy vegetation known as Cerrado (Ecoregion NT0704; Fig. 1). Additional sampling sites located within the Cerrado or in transitional zones provided supplementary specimens. At the western side of the Cerrado, most sampling sites were within two ecoregions: Madeira-Tapajós moist forests (NT0135) and Chiquitano dry forests (NT0212). Hereafter, these sites are referred to collectively as the Chiquitano range. At the eastern side of the Cerrado, the sampling sites were located in five ecoregions: Araucaria moist forests (NT0101), Bahia coastal forests (NT0103), Bahia interior forests (NT0104), Paraná-Paralba interior forests (NT0150), and Atlantic dry forests (NT0202). Hereafter, these sites are referred to collectively as the Atlantic range. The Atlantic range had an outlier that was located within the Pernambuco coastal forests (NT0151), which are a block of seasonal forests of northeastern Brazil in an area dominated by Caatinga (NT1304), namely, scrubland vegetation.

The sizes of population samples varied from 1 to 8, and depended on the number of specimens available for sampling within a given population (Appendix A, online supplemental material). Leaf samples were transported to the laboratory while still fresh and then kept at −80 °C. Alternatively, they were dried immediately using silica gel and kept at room temperature until the DNA was extracted. Total genomic DNA was extracted following the procedure described in Rossi et al. (2009). Genomic DNA from all specimens was archived in our laboratory at the Federal University of Viçosa, Brazil.

2.2. Assembly of ITS datasets

The entire ITS region of the nuclear 18S-26S ribosomal RNA genes (which included the 5.8S gene) was amplified in accordance with standard PCR protocols (Oliveira et al., 2010; Muellner et al., 2009), using primer IT54 described by White et al. (1990) and primer ITS.1E by Baum et al. (1998). PCR products were cleaned using ExoSAP IT (USB; 3 μl of enzyme per 9 μl of reaction). Sequencing was performed by Macrogen Inc., South Korea (www.macrogen.com), using the same primers as in the PCR amplifications. Sequences were imported into the program Sequencer 4.8 (Gene Codes) for editing. Complete sequence alignments were performed with the introduction of gaps to compensate for the presence of insertions/deletions (indels).

We selected six specimens whose sequences were associated with electropherograms that showed multiple overlapping double peaks and cloned the PCR products into the vector pGEM T-Easy (Promega). For each specimen, we resubmitted six clones for sequencing and received usable data from 34 clones. Products from the remaining 120 specimens that gave rise to multiple overlapping double peaks for both strands were not subjected to cloning and, thus, were excluded from subsequent DNA analyses.

We set the limits of the two ITS regions (ITS1 and ITS2) and the 5.8S gene in C. fissilis in accordance with the following three databases: RFam (Gardner et al., 2008), The European Ribosomal Database (Wuyts et al., 2004), and The ITS Database (Selig et al., 2008). Subsequently, the flanking regions of the 18S and 28S DNAs were trimmed to produce an ITS dataset with a size that ranged from 591 to 639 bases. This variation in size was due to the presence of three indels within the ITS1 region: a 36 bp indel (in two clones of specimen PLT75), a 48 bp indel (in specimen VNI38), and a 51 bp indel (in specimen AJU28). The aligned ITS1 subset contained 251 bases, of which 27 were polymorphic (24 substitutions and three indels); the aligned 5.8S subset spanned 156 bases and lacked polymorphic sites; the ITS2 subset was comprised of 232 bases, of which 14 were polymorphic (14 substitutions, no indels). Henceforth, we will refer to this dataset as ITS dataset A. ITS dataset A contained 77 sequences: 20 sequences obtained through direct sequencing of specimens whose electropherograms showed no ambiguities, 23 sequences obtained through direct sequencing of specimens whose electropherograms contained a single site showing a double peak (coded with IUPAC ambiguity codes: R, Y, K, S, or M), and 34 sequences obtained from cloning (GenBank accesses JPR22184 to JPR22260).

We also assembled a second ITS dataset, ITS dataset B, in which the single ambiguities were resolved; that is, each of the 23 sequences that harboured single double peaks was separated into a pair of new sequences. These sequences were identical to the respective parental sequence except that each sequence of the pair included one of the two possible nucleotide codes instead of the formerly ambiguous code. ITS dataset B contained 100 sequences.

2.3. Analyses of secondary structures in the ITS region

We used ITS dataset B to investigate further the intraspecific polymorphism in ITS in Cedrela fissilis and to determine whether pseudogenes were present. We examined the ITS1 and 5.8S gene regions of each sequence for the occurrence of unexpected substitutions within evolutionarily conserved motifs. We then performed analyses of secondary structure for both the 5.8S gene and the ITS2 regions and looked for the presence of aberrant structures that may suggest a nonfunctional copy.

The program RDP3 (Heath et al., 2006) was used to scan ITS data- set B for recombination using seven different recombination detection methods. Next, we split ITS dataset B into three smaller subsets, which corresponded to the ITS1, 5.8S, and ITS2 regions. The DnaSP v5 software (Librado and Rozas, 2005) was used to define haplotypes within each of the three subsets. We used sequence alignments in Sequencer to identify unexpected nucleotide substitutions within known conserved angiosperm motifs in both the 5.8S gene and the
ITS1 region. In the 5.8S gene, we inspected the following three 5.8S motifs that are conserved in angiosperms: M1 (5'-CGATGAAAGCCT AGC-3') and M3 (5'-TTTGGCAGC-3') from Harpke and Peterson (2008), and M2 (5'-GATTGCGAAGCCT-3') from Jobes and Thien (1997). In the ITS1 region, we inspected the conserved angiosperm motif 5'-GGCRY(4–7n)GYGYCAAGGAA-3' from Liu and Schardl (1997). In the ITS1 region, we inspected the conserved angiosperm motif 5'-GGCRY(4–7n)GYGYCAAGGAA-3' from Liu and Schardl (1994).

Both the secondary structure and the minimum free energy (ΔG) of haplotypes within the 5.8S and ITS2 subsets were estimated with the online version of the program MFOLD 3.2 – RNA-folding (Mathews et al., 1999; Zuker, 2003), with the default conditions for temperature (37 °C) and ionic conditions. The secondary structures and ΔG values of the ITS2 region were predicted by homology modelling using the secondary structure of an accession of C. fissilis (GIZ1981686) that is available in The ITS2 Database (Selig et al., 2008). Subsequently, the secondary structures for the 5.8S gene and ITS2 region were obtained with the help of the program 4SALE (Seibel et al., 2008) and edited for printing with the program 4SALE (Seibel et al., 2008) and edited for printing with the program 4SALE (Seibel et al., 2008).

2.4. Assembly of the cpDNA dataset

Three regions of the chloroplast genome were investigated. The trnT–trnl spacer was amplified using standard PCR protocols (Mueller et al., 2009; Oliveira et al., 2010). The initial amplification used primers A and D, which were described by Taberlet et al. (1991). Samples that yielded no reaction product or ambiguous results were re-examined by two independent amplification reactions. The first reaction was performed using the primers A2, described by Cronn et al. (2002), and A5 (5'-GATGCCTTACCGATTTCG-3'), which is complementary to the primer C described by Taberlet et al. (1991). This primer pair amplifies the intergenic spacer between trnT (UGU) and the 5’ exorn of trnl (UAA) (Taberlet et al., 1991). The second reaction was performed with primers C and D and amplified the intron of trnl (UAA) (Taberlet et al., 1991). The trnS–trnG intergenic spacer was amplified using primer pair trnS/trnG. The psbB, psbT, and psbN genes were amplified using primer pair psbB/psbF. Primer pairs trnS/trnG and psbB/psbF were described by Hamilton (1999). The PCR products were cleaned and sequenced, and the sequences edited, following the same protocols that were used for the ITS products. We refrained from using indels that were bordered by mononucleotide repeats (e.g., polyA) as a source of information in subsequent statistical analyses. The removal of these sites from subsequent analyses was intended to eliminate effects that can arise from experimental error or evolutionary lability associated with this type of indel (see Mast et al., 2001).

2.5. Phylogenetic analyses

Bayesian analyses were carried out independently for ITS and cpDNA datasets because of the presence of more than one ITS copy per specimen; moreover, we also had an uneven number of samples in each dataset. We conducted two independent phylogenetic analyses for the ITS sequences, one based on ITS dataset A and a second based on an extended version of ITS dataset A. The extended version contained additional sequences from other members of the genus Cedrela, with Toona ciliata as an outgroup. These sequences were obtained from GenBank (Appendix B, online supplementary material). Before each analysis was run, identical sequences that belonged to the same taxon were combined into a single terminal sequence. However, the original specimen codes were kept for future reference and presented in the tree. MrModeltest v2.3 (Nylander, 2004) was used with the two versions of ITS dataset A separately. The Akaike Information Criteria (Akaike, 1973) indicated that GTR + I + G was the best-fit model among the 24 models of molecular evolution for both analyses. Using the GTR + I + G model, MrBayes v3.1.2 (Ronquist and Huelsenbeck, 2003) was used to estimate Bayesian phylogenies. Each Bayesian analysis was performed using two simultaneous runs of 10 million generations each, with one cold and three heated chains in each run. Sampling was carried out once every 1000 trees; the first 250 trees were discarded as burn-in samples. For each analysis, a 50% majority-rule consensus tree of the two independent runs was obtained with posterior probabilities that were equal to bipartition frequencies.

We also conducted a phylogenetic analysis that was based on the cpDNA dataset. In this case, we did not include additional sequence data from GenBank because the sequences that were available provided data for only two regions (namely, the trnS–trnG intergenic spacer and the psbB, psbT, and psbN genes). Moreover, alignment analyses had indicated that most of the parsimony-informative characters in our cpDNA dataset were located in the trnT–trnl region. Again, identical sequences that belonged to the same taxon were combined into a single terminal sequence, but the original specimen codes were kept for future reference and presented in the tree. MrModeltest v2.3 (Nylander, 2004) was used to analyse the cpDNA dataset. The Akaike Information Criteria (Akaike, 1973) indicated that HKY was the best-fit model among the 24 models of molecular evolution. Using the HKY model, MrBayes v3.1.2 (Ronquist and Huelsenbeck, 2003) was used to estimate an unrooted Bayesian phylogeny. Bayesian analysis was performed using two simultaneous runs of 2 million generations each, with one cold and three heated chains in each run. Sampling was carried out once every 1000 trees; the first 250 trees were discarded as burn-in samples. A 50% majority-rule consensus tree of the two independent runs was obtained with posterior probabilities that were equal to bipartition frequencies.

2.6. Divergence dating

To estimate the time of divergence between lineages of C. fissilis, we used the relaxed clock method as implemented in BEAST 1.6.1 (Drummond and Rambaut, 2007). The BEAST analysis explicitly assumes distinct evolutionary models for population-level data (such as a coalescent-based model) and species-level data (such as the Yule process speciation) (Drummond and Rambaut, 2007). As a consequence, we refrained from merging our intraspecific data (full ITS dataset A) with interspecific data (Appendix B, online supplementary material) and, instead, we chose two sequences (COL268 and SMA73) from our ITS dataset A to combine with the interspecific ITS dataset. The resulting ‘time ITS dataset’ contained a total of 25 sequences: 23 for the tribe Cedreleae (20 sequences of Cedrela and three of Toona) and one sequence of each of the two outgroups, Khaya anthotheca and Swietenia macrophylla. The Akaike Information Criteria (Akaike, 1973) in MrModeltest 2.3 (Nylander, 2004) indicated HKY + G to be the best-fit model of molecular evolution. For the molecular clock model, we chose the uncorrelated log-normal relaxed clock option. Yule process speciation was used as the tree prior. We followed Mueller et al. (2010) and calibrated the BEAST analysis with the following normally distributed priors: the date of the most recent common ancestor (MRCA) of Cedreleae was set to 48.6 million years ago (Mya) (Reid and Chandler, 1933; Chandler, 1964), with an SD of 1; the date of the MRCA of Cedrela set to 33.62 Mya (Meyer and Manchester, 1997), with an SD of 1;
and the date of the MRCA of *Khaya* and *Swietenia* was set to 22.5 Mya (Castañeda-Posadas and Cevallos-Ferriz, 2007), with an SD of 1. The analysis was run for 10 million generations, with samples taken every 2000 generations. These settings ensured that both model parameters and time estimates were sampled adequately (Effective Sample Size, ESS, values were well above 500 for all statistics in Tracer 1.5).

### 3. Results

Sequencing of the target trnT–trnL region produced two overlapping segments, which yielded a total aligned sequence of 1498 bases (GenBank accessions JF922130–JF922156), of which 21 were polymorphic (16 substitutions; five indels of 1 bp, 5 bp, 7 bp, 21 bp, and 22 bp). Sequencing of the trnS–trnG intergenic spacer was carried out and its alignment was 716 bases long (GenBank accessions JF922261 to JF922286), of which five were polymorphic (four substitutions; one indel of 1 bp). Sequencing of the trnS–trnG intergenic spacer was carried out and its alignment was 689 bases long (GenBank accessions JF922157–JF922183), of which one was a substitution. The final cpDNA dataset resulted from the concatenation of these three chloroplast regions, contained 27 sequences, and was 2903 bases long.

#### 3.1. Inter- and intra-individual polymorphism for ITS

DNA alignments revealed that the ITS region contained 41 polymorphic sites, all of which were base substitutions. The 41 polymorphic sites were spread along ITS1 and ITS2; no variability was observed within the 5.8S gene. Across sequence alignments, each of the 41 polymorphic sites showed at least one ambiguity in at least one specimen; that is, multiple sequences showed overlapping double peaks in both strands at the same site of the marker. Therefore, the elimination of ambiguous sites from subsequent analyses would reduce the variability within the dataset to zero. Each site that gave rise to double peaks was flanked on both sides by stretches of clean sequences (Appendix C, online supplementary material). The pattern of double peaks in the electropherograms suggested that intratandem polymorphism was present in most of the specimens of *C. fissilis* that we had sampled.

#### 3.2. Structural analyses of the ITS region

In *C. fissilis*, the conserved ITS1 angiosperm motif 5’-GGCRY(4–7)nGYGYCAAGGAA-3’ of Liu and Schardl (1994) appeared as 5’-GGGCC(GAGCY)GCCCAAGGAA-3’ and contained no further substitutions. The three 5.8S motifs that are conserved in angiosperms were free of substitutions, and motif M3 (5’-TTTGAAYGCA-3’), described by Harpke and Peterson (2008), appeared as 5’-TTTGAAACGCA-3’.

No nucleotide substitutions or indels were observed within the 5.8S gene. This single 5.8S haplotype exhibited the predicted secondary structure for a functional 5.8S gene, with the presence of the five conserved helices, a CG content of 55.76%, and a ΔG of −10.20 kcal/mol (Appendix D, online supplementary material). Further alignments with sequences obtained from GenBank indicated that the 5.8S gene of 17 other members of the genus exhibited a sequence that was identical to the 5.8S haplotype of *C. fissilis*.

Structural analyses of the ITS2 region identified seven secondary structures that displayed highly similar topologies (Appendix E, online supplementary material). The secondary structures each exhibited four helices, which we numbered I to IV following Coleman (2007). Nucleotide substitutions were identified in all four helices, but they were found more commonly in Helix I and Helix III. Helix II exhibited the most conserved secondary structure and harboured a characteristic pyrimidine–pyrimidine bulge (C–U) at the base. The ΔG required for the secondary structures varied from −90.0 to −97.4 (mean = −93.99 kcal/mol), and the CG content of these sequences ranged from 72.4% to 74.1% (mean = 73.3%).

Two of the seven secondary structures (A1 and B1) were recovered at high frequencies, whereas the remaining five were rare (Table 1). Extension of the loop at the end of Helix II distinguished A1 from B1. There was a clear correspondence between the geographic range where the specimens were sampled and the frequency of a given secondary structure for the ITS2: A1 was prevalent among sequences from the Atlantic range, whereas B1 was more frequent among sequences sampled in the Chiquitano range. Only five sequences from the Chiquitano range displayed the A1 secondary structure, and they were recovered from clones (SRA113_1, SRA113_3, SRA113_6, PLA75_2, and PLA75_6). Similarly, the six sequences from the Atlantic range that displayed the B1 secondary structure were also recovered from clones (DOU117_1, DOU117_2, ITA124_1, ITA124_3, ITA124_4, and ITA124_6). The individuals SRA113, PLA75, DOU117, and ITA124 exhibit intratandem polymorphism for ITS; in addition of the sequences with the more typical secondary structure (A1 for the Atlantic range or B1 for the Chiquitano range), each of those four individuals also contained the rare paralogs we recovered only through cloning.

#### 3.3. Bayesian analyses based on the ITS data

The Bayesian consensus tree for the full extended ITS dataset (Fig. 2) showed the locations of the 77 sequences from *C. fissilis* in relation to 26 other sequences from *Cedrela*. *Toona ciliata* (FJ462488) was used as an outgroup. The tree displayed a topology that suggested an association between member composition and geography, with strong support from posterior probabilities (PP). Overall, the most basal positions contained sequences from Mexican and Central American species, intermediary positions were filled by sequences from South American species other than *C. fissilis*, and terminal clades contained sequences from *C. fissilis*. All 77 sequences of ITS dataset A were grouped together in a strongly supported clade (PP = 98%), with one sequence of *C. odorata* (FJ462471) and two sequences of *Cedrela balansae* (FJ462473) and (FJ462474). Henceforth, we will refer to this clade as the *C. fissilis* clade.

To explore the geographic distributions and phylogenetic relationship of members of the *C. fissilis* clade further, we developed an unrooted Bayesian consensus tree using sequences found.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Geographic distribution of the seven secondary structures of ITS2 in <em>Cedrela fissilis</em>.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secondary structure</td>
<td>Number of haplotypes</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>A1</td>
<td>7</td>
</tr>
<tr>
<td>A2</td>
<td>1</td>
</tr>
<tr>
<td>A3</td>
<td>1</td>
</tr>
<tr>
<td>A4</td>
<td>1</td>
</tr>
<tr>
<td>B1</td>
<td>5</td>
</tr>
<tr>
<td>B2</td>
<td>1</td>
</tr>
<tr>
<td>B3</td>
<td>1</td>
</tr>
</tbody>
</table>
within that clade and presented it in conjunction with a map of Brazil and Bolivia (Fig. 3). The unrooted tree revealed the presence of two major groups (PP = 97%); because these two major groups may or may not be monophyletic, we avoid referring to them as clades. The West group (Fig. 3, in green) consisted mostly of sequences from specimens collected in the Chiquitano range: nine sequences from Madeira-Tapajós moist forests (NT0135) and 20 sequences from Chiquitano dry forests (NT0212). In addition, the West group contained two cloned sequences (ITA124_3 and ITA124_5, see blue arrows in Fig. 3) from a specimen found at the westernmost edge of an ecoregion of Atlantic origin (Paraná-Paraiá interior forests – NT0150), and two sequences from specimens of C. fissilis from Paraguay (FJ518893) and Peru (FJ462475), respectively. The East group (Fig. 3, in blue) contained the remaining 46 sequences from C. fissilis, in addition to two sequences from C. balansae from Argentina (FJ462473) and Paraguay (FJ462474) and a sequence of C. odorata from Brazil (FJ462471). Most of the sequences in the East group were derived from specimens collected in ecoregions located along the Atlantic coast. However, four sequences obtained through cloning were from two specimens (PLA75_2, PLA75_6, SRA113_1, and SRA113_3; see green arrows in Fig. 3) from the Chiquitano range.

3.4. Bayesian analysis based on concatenated cpDNA data

The unrooted Bayesian consensus tree for cpDNA and associated PP are presented together with a map of Brazil and Bolivia that shows the geographic distribution of 27 specimens of C. fissilis (Fig. 4). The tree revealed the presence of two major groups, with high support (PP = 100%). The sequences from the Chiquitano range clustered almost exclusively in the West group (Fig. 4, in green), whereas the sequences from the Atlantic range clustered almost exclusively in the East group (Fig. 4, in blue). The few exceptions were as follows: (i) specimen SRA113 was placed into the East group and (ii) specimens PAL164 and MAN284 were placed into the West group. Specimen SRA113 was collected within the core of the Chiquitano dry forests (NT0212) in Eastern Bolivia; it showed intraspecific polymorphism for ITS and displayed paralogs of either Chiquitano or Atlantic origin (Fig. 3). PAL164 was collected in an isolated block of seasonally dry forest located in Central Brazil and surrounded by savanna vegetation (Cerrado – NT0704). This disjunct block is part of the Paraná-Paraiá interior forests (NT0150), an ecoregion of Atlantic origin. Meanwhile, specimen MAN284 was collected in Northern Minas Gerais, a transition zone that comprises an enclave of seasonally dry forests (Atlantic...
dry forests – NT0202), savanna vegetation (Cerrado – NT0704), and scrubland (Caatinga – NT1304).

3.5. Estimated date of divergence

The BEAST analysis yielded a maximum clade credibility tree (Fig. 5) that showed that the sequences of the *C. fissilis* clade formed a strongly supported clade (PP = 98%). In contrast, there was only weak support (PP < 60%) for the basal placement of the sequence of *C. odorata* (FJ462471) from Brazil in this clade. Within the *C. fissilis* clade, there were two sub-clades (PP = 100%), which corresponded to the West and East groups that we had identified earlier (Figs. 2 and 3). The *C. fissilis* sequences COL268 (from the West group) and SMA73 (from the East group) that were used in the analysis were obtained through direct sequencing of specimens whose electropherograms presented no ambiguities. In the unrooted Bayesian consensus tree, these two sequences occupied most derived placements (Fig. 3). The BEAST analysis suggested that the divergence between the West and East groups of *C. fissilis* took place about 10 Mya (Fig. 5).

4. Discussion

4.1. Multiple, functional copies of ITS

In our analyses, *C. fissilis* did not display the expected within-species homogeneity for copies of ITS. The nuclear ribosomal DNA (nrDNA) consists of tandem units of the genes 18S, 5.8S, and 28S, each of which encodes a ribosomal RNA molecule. Two regions (ITS1 and ITS2) flank the two sides of the 5.8S gene and thus separate the three ribosomal genes (i.e., 18S – ITS1 – 5.8S – ITS2 – 28S). In plant genomes, it has been proposed that the hundreds to thousands of tandemly repeated copies of nrDNA that are present in the genome (Rogers and Bendich, 1987) are homogenized to a single sequence through unequal crossing over and a high frequency of gene conversion, a phenomenon called concerted evolution (Elder and Turner, 1995). When carried to completion, concerted evolution removes paralogous sequences from the genome, and thereby eliminates intranidividual polymorphism. As a consequence, nrDNA tandem repeats presumably become homogeneous within species and different between species (Dover, 1986). However, in 149 out of 169 specimens of *C. fissilis*, the ITS sequences showed at least one ambiguity; that is, at polymorphic sites, the electropherograms displayed overlapping double peaks for both strands. Only 20 specimens were free of ambiguities. It appears that multiple, paralogous copies of ITS coexist within the genome of most specimens of *C. fissilis* from Brazil and Bolivia.

Intranidividual polymorphism for ITS may be associated with nonfunctional nrDNA copies, or pseudogenes, in addition to functional copies (Bailey et al., 2003; Oliveira et al., 2010). Features such as low CG content, high rates of substitutions and deletions within highly conserved motifs, and aberrant secondary structures are suggestive of paralogs that have escaped from functional constraints and become pseudogenes (Bailey et al., 2003; Rosselló et al., 2007; Oliveira et al., 2010). In *C. fissilis*, the ITS paralogs exhibited none of these features; we could not identify any pseudogenes among the sequences that we analysed. Thus, the hypothesis that multiple, functional copies of ITS exist within the genome of some specimens of *C. fissilis* could not be rejected.

4.2. Incomplete concerted evolution in *C. fissilis*

Under certain circumstances, the cohesiveness of the tandem repeats of nrDNA is disturbed and concerted evolution remains
incomplete. This gives rise to intraindividual polymorphism, such as that observed in \textit{C. fissilis}, which persists over time. In plants, hybridization, polyploidy, and long generation time are factors that contribute to incomplete concerted evolution and probably contributed to intraindividual polymorphism in \textit{C. fissilis}. Both interspecific and intraspecific hybridization bring together different alleles from the parental sources into a single genome and therefore can give rise to intraindividual polymorphism (Koch et al., 2003; Muir et al., 2001). Long generation time, a common feature of tropical tree species such as \textit{C. fissilis}, is thought to reduce the rates of concerted evolution (Sang et al., 1995). Polyploidization could favour the maintenance of intraindividual polymorphism because tandemly repeated copies of nrDNA on different chromosomes are less susceptible to the homogenizing effects of concerted evolution (Campbell et al., 1997). Given that the number of loci and chromosomal locations of nrDNA in \textit{C. fissilis} are unknown, the extent to which polyploidization contributed to the intraindividual polymorphism that we identified remains unclear. Nonetheless, the high and variable range of chromosome numbers within \textit{Cedrela} (Styles and Vosa, 1971) suggests that ancient polyploidy probably contributed to the incomplete concerted evolution in \textit{C. fissilis}.

4.3. Cryptic speciation within \textit{C. fissilis}

Monophyly of \textit{C. fissilis} relative to other species of \textit{Cedrela} was not supported by our Bayesian analysis because sequences from two other species (\textit{C. odorata} and \textit{C. balansae}) were grouped together with sequences of \textit{C. fissilis} and formed a well-supported clade (Fig. 2). It is interesting to note that the non-fissilis components clustered in the East group only, and clustering seemed to have a geographically component (Fig. 3). The sequence of \textit{C. odorata} (FJ462471, from Brazil) was identical to the sequence of \textit{C. fissilis} from north-eastern Brazil (GUA204). Two sequences of \textit{C. balansae} from Argentina and Paraguay (FJ462473 and FJ462474, respectively) also clustered together with the \textit{C. fissilis} sequences of the East group. The presence of multiple evolutionary lineages within \textit{C. fissilis} suggests cryptic speciation.

Bayesian analyses using either ITS (Fig. 3) or cpDNA (Fig. 4) datasets yielded trees of highly similar topologies. Each tree had major, well-supported groups, which suggested that \textit{C. fissilis} underwent one major diversification event. With a few exceptions (see Section 3.3), the placement of sequences produced almost exclusive groups that corresponded to geography: the West group contained sequences from the Chiquitano range, whereas the East group contained sequences from the Atlantic range. The split of \textit{C. fissilis} into West and East groups occurred during the Early Pliocene and Late Miocene, which overlaps with the period in which the diversification of \textit{Cedrela} intensified in Central America and northern parts of South America (Muellner et al., 2010).

The phylogenetic relationship between \textit{C. odorata} and \textit{C. fissilis} remains unclear, mostly because species delimitation for both \textit{C. odorata} and \textit{C. fissilis} seems to be unresolved. Earlier studies pointed to the existence of genetically distinct but morphologically indistinguishable entities of \textit{C. odorata} throughout the species range (Muellner et al., 2010; Muellner et al., 2009). In the present study, we provide evidence that the clade that contained \textit{C. fissilis} and \textit{C. odorata} (FJ462471, from Brazil) shared an MRCA with the clade of \textit{C. odorata} from Central America (FJ462468, from El Salvador; FJ462467, from Belize) at about 20 Mya. Larger samplings within \textit{C. odorata} and \textit{C. balansae} to conduct additional DNA and morphometric analyses are required to determine the relationships and delimitations.
among these three species. *Cedrela angustifolia* Sessé & Moc. ex DC. (syn = *Cedrela lilloi* C. DC.) is another species of *Cedrela* that occurs in Southern Brazil; it is categorized as an ‘Endangered A1a+2cd’ species (IUCN, 2010). The *C. fissilis* clade shared an MRCA with the clade of *Cedrela montana* (FJ462480, from Peru) and *C. angustifolia* (FJ462478, from Bolivia) at about 28 Mya (Fig. 5), which suggested that, over time, multiple dispersal events brought *Cedrela* to the forests of what is currently known as Brazil.

4.4. Association of *C. fissilis* with geographic ranges

The low levels of structuring within the West group of *C. fissilis* suggested that the Chiquitano range harboured a genetically depauperate lineage. Moreover, the star-like topology that we identified for the West group can be taken to be an indicator of a large population expansion in the past (Marjoram and Donnelly, 1994). The hypothesis that the West group experienced a recent, rapid expansion is plausible given the existing evidence that stability of climate and vegetation in south-eastern Brazil provided refugial areas during climate cycles that affected other regions more dramatically (Carnaval and Moritz, 2008; Carnaval et al., 2009). The existence of substructuring within the East group was an additional indication that closely related lineages of *C. fissilis* evolved within the Atlantic range. It is likely that each of these lineages experienced a certain degree of genetic isolation within distinct refugial areas. To differentiate the closely related Atlantic lineages fully, we would need to analyse a much larger dataset. Further morphometric studies may reveal whether the genetic differentiation that we uncovered in *C. fissilis* has taxonomic relevance.

4.5. Genetic connectivity between formerly isolated lineages

During past climate fluctuations, Central Brazil was drier than the present-day Cerrado (Whitmore and Prance, 1987; Mayle, 2004; Pennington et al., 2004), and probably inhospitable to *C. fissilis*. Our data are consistent with a scenario in which climate-mediated events contributed to split the population of the parental source of *C. fissilis* into refugial areas where distinct lineages evolved during periods of genetic isolation. More recently (within the last few thousand years), a warmer wetter trend permitted novel vegetation to assemble in Central Brazil, when forests spread at the expense of savannas (Mayle, 2004). Currently, seasonal forests permeate the Cerrado in the form of gallery forests and provide...
otherwise isolated populations of plant (Oliveira-Filho and Ratter, 1995) and animal (Costa, 2003) species with genetic connectivity.

Data from this study provide evidence of relatively recent events in which C. fissilis dispersed from adjacent ecoregions into the Cerrado. Specimens collected in the Cerrado ecoregion did not form an exclusive group but, instead, clustered with specimens from nearby seasonal forests. Despite this general trend, exceptions to this pattern were found. Notably, the clustering of some Cerrado specimens illustrated the capability of C. fissilis to expand deeper into regions of Cerrado vegetation, as evidenced by the following: (i) specimens from distinct edges of the Atlantic range displayed both ITS paralogs and cpDNA that belonged to the Chiquitano lineage; (ii) specimens from within the core of the Chiquitano range exhibited both ITS paralogs and cpDNA that belonged to Atlantic lineages. Expansion eastwards brought the Chiquitano lineage towards Central Brazil, to the disjunct blocks of Paraná–Paráiba interior forests (NT0150) and into the Atlantic region (Atlantic dry forests – NT0202). Dispersal in both directions allowed both the Chiquitano lineage to reach Paraná–Paráiba interior forests (NT0150) in Southern Brazil and the Atlantic lineage to reach Chiquitano dry forests (NT0212) in Bolivia and Mato Grosso. We speculate that bidirectional migrations were possible owing to the Paraná River basin and Chaco, which at present straddle the Brazilian–Paraguayan–Argentinean border. Due to these long distance dispersal routes, the Chiquitano and Atlantic ranges are not completely genetically isolated ranges, even though they are separated by a 500–1500 km expanse of Cerrado vegetation.

The pattern of intraindividual polymorphism and incomplete concerted evolution that we uncovered for the ITS region suggest that intraspecific hybridization occurred when allopatric lineages of C. fissilis reconnected and established contact zones after the climate ameliorated in recent times. Support for the widespread occurrence of intraspecific hybridization within and among ranges came from the fact that the genomes of most specimens contained multiple, functional paralogs for ITS from distinct lineages. We attribute the lack of intraindividual polymorphism for ITS in some specimens as a relict from ancestral populations of seasonal forest refugia that have escaped intraspecific hybridization over time.

The phylogeography that we have described for C. fissilis is highly congruent with geographic patterns of genetic diversity for the understory shrub C. ipecacuanha (Oliveira et al., 2010). Expanded molecular studies that include even more species should help determine whether these patterns can be generalized to the evolutionary history of Neotropical seasonal forests through a framework of climatic cycles and shifts in vegetation that resulted in periods of geographic separation and genetic isolation that alternated with periods of range expansion and intraspecific hybridization.

Acknowledgments

We would like to express our sincere gratitude to the following people for providing assistance during field trips and sampling: Anderson Luis Moy (ECOPEF), Alan Sciamaroli (UFJD), Andressa Vasconcelos Flores, Buzi Cosenza (UEMG), Cristina Soares de Souza, Ernane R. Martins (UFMG), Gisele Benites Flor (FECLPP), Lourivaldo Vasconcelos Flores, Braz Cosenza (UEMG), Cristina Soares de Souza, Giseli Benites Flor (FECLPP), Helio de Brito (Fundação SOS Mata Atlântica, INPE), Rui C. Augusto (FUNDAÇÃO SOS Mata Atlântica/INPE), and Solon Longhi (UFSC). This study was part of MG Ph.D. Thesis with the support of a CNPq fellowship. LOO received a fellowship from CNPq (PQ 305710/2009-5) and grants from FAP-EMG (PPM-00148-09) and CNPq (471915/2009-2).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2011.08.026.

References


Harpke, D., Peterson, A., 2008. 5.8S motifs for the identification of pseudogenic ITS regions. Botany 86, 300–305.


