Diversification in species complexes: Tests of species origin and delimitation in the *Bursera simaruba* clade of tropical trees (Burseraceae)

Julieta A. Rosella, Mark E. Olsona, Andrea Weeksb, J. Arturo De-Novaa, Rosalinda Medina Lemosa, Jacqueline Pérez Camachoc,1, Teresa P. Feriad, Roberto Gómez-Bermejoa, Juan C. Monteroe, Luis E. Eguiartec

1. Introduction

Species are the working units in most of biology, yet delimiting species is fraught with complexity (Barraclough and Nee, 2001; 1. Introduction

Plasticity

Paraphyletic species

Parallelism

Mexico

Convergence

Antilles

Keywords:

Available online 17 August 2010

Accepted 9 August 2010

Revised 4 August 2010

Received 1 April 2010

A B S T R A C T

Molecular phylogenies are invaluable for testing morphology-based species delimitation in species complexes, as well as for examining hypotheses regarding the origination of species in these groups. Using five nucleotide markers, we reconstructed the phylogeny of the *Bursera simaruba* species complex of neotropical trees to test the notion that four “satellite” species originated from populations of the most widely distributed member of the genus, *B. simaruba*, which the satellites strongly resemble. In addition to molecular phylogenetic reconstruction, we tested species delimitation of *B. simaruba* and the satellites using multivariate analyses of morphological and ecological characters. The analyses evaluated the taxonomic value of these traditional characters and pinpointed those in need of further study, such as the expression of pubescence. Phylogenetic data rejected the origin of three satellite species from their purported ancestor, *B. simaruba*, and we ascribe their morphological similarity to convergence or parallelism. The fourth satellite species likely represents one end of a spectrum of inflorescence length variation within *B. simaruba* and is conspecific. Despite its marked morphological variability, we recovered *B. simaruba* as a single valid species, which implies that it maintains genetic cohesion among distant populations throughout its vast range.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

Species are the working units in most of biology, yet delimiting species is fraught with complexity (Barraclough and Nee, 2001; Sites and Marshall, 2003, 2004). In a phylogenetic framework, the species of traditional taxonomy can be viewed as hypotheses tested by recovering either non-polyphyletic or polyphyletic groupings. Most species concepts view monophyly as congruent with a species hypothesis (e.g. Baum and Shaw, 1995; Shaw, 1998), and some also recognize the possibility of paraphyletic species (Coyne and Orr, 2004; Crisp and Chandler, 1996; Harrison, 1998; Templeton, 1989). However, virtually all concepts view polyphly as a rejection of a species hypothesis. Because molecular phylogenetic hypotheses can readily identify polyphyletic taxa, they are an excellent aid in delimiting species (Knowles and Carstens, 2007; Wiens and Penkrot, 2002), especially in the case of species complexes, whose continuously varying morphological characters often confound morphology-based taxonomies (e.g. Campbell et al., 2004; Lellaert et al., 2009). In this work, we test species hypotheses in the *Bursera simaruba* complex, a group of neotropical trees with a knotty taxonomic history resulting from the strong morphological similarity and overlapping geographic distributions of its species (Rzedowski et al., 2007).

The history of species delimitation in the *simaruba* complex exemplifies central issues in species complex systematics. Species complexes may contain widespread, morphologically variable species, as in the painted turtle (Starkey et al., 2003), or multiple ecologically distinct species that are morphologically similar, as in the *Rana pipsiens* complex (Hillis, 1988). Both the single and the multiple species scenarios have been proposed by taxonomists for entities of the *simaruba* complex (Bullock, 1936; McVaugh and Rzedowski, 1965; Rzedowski et al., 2007; Standley, 1923), especially for *B. simaruba*, the species from which the complex derives its name (Fig. 1A). Commonly known as the gumbo-limbo tree or palo mulato, *B. simaruba* is a conspicuous element in virtually all...
tropical forests below 1400 m from Mexico to northern South America, Florida, and the West Indies (Fig. 2). It varies from small trees of subdeciduous forest to towering rainforest emergents, and in the amount and distribution of leaf pubescence, leaflet shape and number, and bark color (Rzedowski et al., 2007; Rzedowski and Kruse, 1979). Although this variation has often been inter-

Fig. 1. Habit diversity in the *Bursera simaruba* species complex. This figure shows the red to copper colored peeling outer bark that characterizes the group, and that, for its size, the clade spans an exceptional array of life forms. (A–C) *Bursera simaruba* and “satellites.” (A) *B. simaruba* remnant in a pasture cut from lowland rainforest. Scale = 10 m. (B) *B. attenuata* in a subdeciduous tropical forest canyon. Scale = 1.5 m. (C) *B. ovalifolia* resembles a smaller, intricately branched *B. simaruba*. Person = 1.90 m. (D–H) Representatives of the *simaruba* complex from tropical deciduous forests. (D) *B. longipes*, one of the smallest members of the clade, growing in very dry tropical deciduous forest. Scale = 2.5 m. (E) *B. grandifolia* is a widespread dryland shrub to medium tree, distinguished by its waxy green trunk shedding ragged copper sheets of outer bark. Scale = 6 m. (F) *B. laurifluus* is endemic to the low, dense woods of the southern Isthmus of Tehuantepec. Scale = 1.5 m. (G) *B. instabilis* is unique for its long, slender branches, which in large individuals become lianescent. Scale = 1.5 m. (H) *B. cinerea*, a water-storing tree from tropical deciduous forest canyons. Scale = 0.75 m. (I) *B. standleyana*, the hemiepiphytic member of the genus, growing at 20 m above the floor of the rainforest. The lower arrow indicates the long lianescent roots that bring water from the soil; the upper arrow singles out arching branches. Scale = 1.5 m.
1.1. Species delimitation in the simaruba complex: genetics, phenetics, and ecology

The simaruba complex is part of an abundant, species-rich American genus of trees and shrubs whose center of diversity is Mexico. The complex comprises 15 species spanning a remarkable array of life form and ecology (Fig. 1, Table 1). They are distinguished from other Bursera by their reddish peeling phellem (outer bark), compound leaves without wings on the rachis, ovate to broadly elliptic entire leaflets with mostly acuminate apices and brochidodromous venation, and three-valved fruits with reddish pseudarils covering the endocarp completely. Although the complex has long been recognized as distinct (McVaugh and Rzedowski, 1965; Rzedowski and Kruse, 1979), and is well supported in molecular phylogenies (Becerra, 2003; Becerra and Venable, 1999), species limits and relationships within the group have always been regarded as problematic (Daly, 1993; Rzedowski et al., 2005, 2007). Molecular phylogenies have included only a few species of the complex and provide little resolution within the group (Becerra, 2003; Becerra and Venable, 1999).

As in many plant groups, the taxonomy of the simaruba complex employs combinations of plastic and potentially evolutionarily labile characters, rather than unique synapomorphies (Daly, 1993; Appendix B, Supplementary material). Years of taxonomic confusion have resulted, reaching a zenith in the satellite species, in which barely diagnostic combinations of leaflet number, pubescence patterns, and fruit size are used to delimit species, with taxonomic realignments continuing to the present day (Bullock, 1936; McVaugh and Rzedowski, 1965; Rzedowski et al., 2007; Standley, 1923). To test morphologically-based species delimitations, we include in our phylogenetic reconstruction two or more samples of eight of the species in the complex, including the four satellite species. In addition, we performed the first phenetic study of the morphological characters used in the simaruba complex taxonomy and compare it to our molecular phylogeny, an approach that we use to rank a wide array of traditional characters in their diagnostic effectiveness.

In addition to morphological and molecular characters, ecology can also help delimit species (Ruiz-Sanchez and Sosa, 2010). Therefore, we test for ecological differentiation between B. simaruba and its satellites through multivariate analyses of environmental variables (Stockman and Bond, 2007). We assume that different niches between species would reject the hypothesis of ecological exchangeability, which, along with genetic exchangeability, is one of two main aspects of the cohesion species concept, which we follow here (Templeton, 1989).

![Fig. 2. Distribution of B. simaruba (in gray) and its satellite species B. attenuata, B. itzae, B. ovalifolia, and B. roseana.](image-url)
Finally, we test the membership in the simaruba complex of some six species from Mexico and Central America, which have never been included in prior phylogenetic analyses (Table 1). In our phylogenetic reconstruction, we also include eight mostly narrow endemic Antillean species. These species are closely related to the simaruba complex, with which they share the tree habit, red-brown bark, and compound leaves with entire leaflets, but differ in having tougher, smaller leaflets with a thicker waxy cuticle (Daly, 1993). However, it is not clear whether these species form a clade, perhaps sister to or within the simaruba complex (Weeks and Simpson, 2004), or, for example, a polyphyletic group derived from repeated invasions of the Antilles from the mainland. In summary, our phylogenetic reconstruction of the complex and the Antillean group includes DNA sequence data from five chloroplast and nuclear markers for 71 samples; 50% of the sampled species and over 90% of the DNA sequences are new contributions to Bursera phylogeny reconstructions (Appendix A).

2. Materials and methods

2.1. Taxon sampling

We sampled all known species in the simaruba complex except for B. inversa (Table 1). From the related Antillean group, we included all species except B. hollickii and B. aromatica. Sampling for B. simaruba included morphologically and environmentally contrasting individuals from Mexico, Central America, and Cuba (Appendix A), along with published sequences from the Dominican Republic and Florida (Weeks and Simpson, 2004). The samples we collected at the type locality of B. itzae were sterile, so they could not be assigned to species with certainty and are thus referred to as B. “itzae”. Of the Antillean species group, we included an undescribed Cuban species, three poorly known Cuban endemics, which had been transferred to Commiphora based on pollen morphology (Moncada-Ferrara, 1989), and four previously sequenced Caribbean species (Weeks and Simpson, 2004; Appendix A). Our outgroup sampling included species of the two subgenera of Bursera, Bursera and Elaphrium, as well as three species of Commiphora, the sister taxon to Bursera subg. Bursera (Weeks et al., 2005) or perhaps to all of Bursera (Becerra, 2003). We downloaded outgroup sequences from GenBank and sequenced missing markers from our collections. GenBank accession numbers are listed in Appendix A.

2.2. DNA extraction, PCR amplification, sequencing, and alignment

We extracted DNA with DNeasy Plant Mini Kits (Qiagen, Valencia, CA) from living and herbarium material (Appendix A). We sequenced five markers, two low-copy nuclear regions (the fourth intron of the phosphoenolpyruvate carboxylase gene, PEPC, and the third intron of the nuclear nitrate reductase gene, NIA1), two multiple-copy regions (the internal and external transcribed spacer regions of the nuclear ribosomal DNA, ITS, and ETS), and one from the chloroplast (the psbA-trnH intergenic spacer). We used published primers and protocols to amplify psbA-trnH (Sang et al., 1997; Weeks and Simpson, 2004), PEPC (Olson, 2002), and ETS (Baldwin and Markos, 1998; Weeks and Simpson, 2004). We increased specificity for PEPC by raising the annealing temperature to 62°C for eight samples. We amplified ITS with the external and internal primers of Fine et al. (2005), with the following protocol: 97°C/3 min; nine cycles of 97°C/1 min, with a touchdown starting at 56°C/1 min and decreasing 1°C every cycle, 72°C/45 s + 4 s/cycle; 30 additional cycles with 48°C as annealing temperature and a final extension at 72°C/7 min. Finally, the protocol for NIA1 (Howarth and Baum, 2002) started at 94°C/3 min; 14 cycles of 94°C/1 min, with a touchdown starting at 62°C/2 min and decreasing 1°C each cycle, 72°C/3 min; 21 additional cycles with
48 °C as annealing temperature and a final extension at 72 °C / 7 min. PCR volumes of 25 μL included 10–100 ng of template DNA, 2.5 μL of 10x PCR buffer, 0.5 μL of 10 μM dNTPs in an equimolar ratio, 0.75 μL of 50 μM MgCl₂, 0.63 μL of each of the 10 μM primers, 5 μL of Q solution (Qiagen), and 1 U of Taq polymerase added after the initial denaturation step.

When necessary, we cloned PCR products by ligation into pGEM-T vector (Promega, Madison, WI) and transformation of competent Escherichia coli JM109. We sequenced three positively transformed colonies for each sample cloned. Due to the presence of more than one product, we cloned the NIaI3 PCR products of six samples, one of B. permollis (J. Linares 7326), two samples of B. grandifolia (Rosell 50 and Olson 1026) and three of B. simaruba (Rosell 46, Rosell 51, and Olson 1063). The other nuclear markers ETS, ITS, and PEPC were not cloned due to the observed congruence between their tree topologies; previous studies have shown coalescence within samples in closely related species for ETS (Weeks and Simpson, 2004).

PCR products were purified with QIAquick PCR purification kit (Qiagen), and sequenced bidirectionally using the BigDye Terminator cycle sequencing kit (Applied Biosystems, Foster City, CA) in final volumes of 10 μL. Sequencing products were cleaned with Sephadex G-50 (GE Healthcare Bio-Sciences, Piscataway, NJ) and run on an ABI 3100 automated DNA sequencer (Applied Biosystems). Sequences were edited using Sequencher v.4.6 (Gene Codes, Ann Arbor, MI) and aligned using Se-Al v.2.0a11 (Rambaut, 2002).

2.3. Maximum parsimony analyses (MP)

We analyzed data matrices by marker, in combined analyses with congruent markers as assessed by the incongruence length difference test (ILD, Farris et al., 1995), and also the five loci together. We included potentially informative characters only, which were unordered and weighted equally. We treated gaps as missing data and performed analyses in NONA 2.0 (Goloboff, 1999) using the parsimony ratchet (Nixon, 1999) with WinClada v.1.00.08 (Nixon, 2002) as a shell program. We conducted three searches with different starting seeds using 300 iterations (100 trees held per iteration). We sampled 10% of the characters for reweighting during the parsimony ratchet and calculated a strict consensus from most parsimonious trees. To evaluate support, we used TNT (Goloboff et al., 2008) running 1000 replicates with the “traditional search” approach with TBR set to 100 replications holding 50 trees, and saving the consensus of resampling matrix. We assessed congruence between each marker using ILD tests as implemented in WinClada v.1.00.08, running 1000 replicates.

2.4. Bayesian analyses (BA)

We used ModelTest v.3.7 (Posada and Crandall, 1998) to determine the model of evolution that best fit each of the five markers and, as for parsimony, performed Bayesian analyses of each marker, combined data of congruent markers, and finally of the five loci combined. We used MrBayes v.3.2 (Huelsenbeck and Ronquist, 2001) run on the CIPRES portal v.2.2 (Miller et al., 2009). For combined datasets, we performed partitioned analyses, applying a different model of evolution to each marker (Table 2), and ran four Markov chains simultaneously for 15 million generations, starting with a random tree, setting the heating parameter to 0.01 to improve the probability of exchange between chains, which was very low with the default value of 0.2, and sampling trees every 200 generations. After 15 million generations, average split deviations were <0.01 and the plot of the likelihood score vs. MCMC generation suggested that stationarity had been reached after 3.75 million generations. Majority-rule consensus trees were derived from 56,250 trees after a 25% burn-in. Analyses by marker were run for five million generations under the same conditions as combined analyses, reaching average split deviations <0.007.

2.5. Species delimitation using molecular characters

Using combined analyses and analyses by marker derived from MP and BA, we assessed whether species were mono-, para-, or polyphyletic, or whether monophyly or paraphyly could not be rejected because of polytomies in the phylogenetic tree. In assessing the poly-, para-, or mono-phyletic status of a species, we considered nodes with parsimony bootstrap values (PB) ≥70% (Hillis and Bull, 1993) or posterior probabilities (PP) ≥0.95 to be strongly supported.

2.6. Species delimitation using morphological characters

We measured 476 herbarium specimens of eleven Mexican species of the simaruba complex, including all satellites except B. itzae, for which material was not available (Table 1). We measured 24 morphological variables of taxonomic importance (Table 3) on specimens deposited in the National Herbarium of Mexico (MEXU) and the Herbarium of the Centro Regional del Bajío (IEB). We included minimum and maximum values per specimen for continuous variables to reflect intraspecific variation. To evaluate the effectiveness and relative importance of these variables in separating species, we performed discriminant analyses in R v.2.9.2 (R Development Core Team, 2009). We log-transformed some variables to achieve normality and assessed homoscedasticity graphically; no variables had correlations >0.9. We centered and scaled all variables and there were no missing data. We repeated the discriminant analysis including only B. simaruba and its satellites to assess which variables were useful in discriminating these four morphologically similar species.

2.7. Species delimitation using ecological characters

To evaluate ecological differentiation between all satellites, except for B. itzae, and B. simaruba, we performed a principal
components analysis (PCA) on 19 climatic variables extracted from WorldClim 1.4 (Table 4; Hijmans et al., 2005) based on 462 locality data from specimens at MEXU and IEB. Following Stockman and Bond (2007), we performed a multiple analysis of variance (MANOVA) on PC scores to assess differences between those species applying the MANOVA, and when this assumption was violated, appearing as sister taxa according to our phylogenetic results.

We performed a non-parametric MANOVA (Anderson, 2001) on PC scores to assess differences between those species and that including only B. simaruba and satellites. Only the first two discriminant functions (DF) are presented. L, length; min, minimum; max, maximum; W, width. Variables with high contribution to the DF in bold.

### Table 3

<table>
<thead>
<tr>
<th>Character</th>
<th>All species</th>
<th>R. simaruba + satellites</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DF1</td>
<td>DF2</td>
</tr>
<tr>
<td>Leaf L min</td>
<td>-0.048</td>
<td>0.26</td>
</tr>
<tr>
<td>Leaf L max</td>
<td>-0.115</td>
<td>-0.138</td>
</tr>
<tr>
<td>Leaf W min</td>
<td>0.172</td>
<td>-0.242</td>
</tr>
<tr>
<td>Leaf W max</td>
<td>-0.234</td>
<td>0.176</td>
</tr>
<tr>
<td>Petiole base pubescencea</td>
<td>-0.075</td>
<td>-0.691</td>
</tr>
<tr>
<td>Petiole shaft pubescencea</td>
<td>-0.014</td>
<td>0.349</td>
</tr>
<tr>
<td>Petiole L min</td>
<td>-0.116</td>
<td>-0.135</td>
</tr>
<tr>
<td>Petiole L max</td>
<td>0.197</td>
<td>0.225</td>
</tr>
<tr>
<td>Petiole L terminal leaflet min</td>
<td>0.133</td>
<td>-0.134</td>
</tr>
<tr>
<td>Petiole L terminal leaflet max</td>
<td>0.041</td>
<td>0.122</td>
</tr>
<tr>
<td>Petiole L lateral leaflet min</td>
<td>0.089</td>
<td>-0.231</td>
</tr>
<tr>
<td>Petiole L lateral leaflet max</td>
<td>-0.048</td>
<td>-0.308</td>
</tr>
<tr>
<td>Petiole pubescencea</td>
<td>-0.206</td>
<td>-0.424</td>
</tr>
<tr>
<td>Leaflet number min</td>
<td>-0.097</td>
<td>-1.172</td>
</tr>
<tr>
<td>Leaflet number max</td>
<td>0.507</td>
<td>-0.557</td>
</tr>
<tr>
<td>Leaflet upper surface pubescenceb</td>
<td>1.474</td>
<td>0.005</td>
</tr>
<tr>
<td>Leaflet lower surface pubescencec</td>
<td>0.193</td>
<td>0.56</td>
</tr>
<tr>
<td>Infructescence L min</td>
<td>0.103</td>
<td>-0.118</td>
</tr>
<tr>
<td>Infructescence L max</td>
<td>0.11</td>
<td>0.226</td>
</tr>
<tr>
<td>Infructescence robustnessd</td>
<td>-0.037</td>
<td>-0.628</td>
</tr>
<tr>
<td>Fruit L min</td>
<td>-0.092</td>
<td>-0.339</td>
</tr>
<tr>
<td>Fruit L max</td>
<td>0.126</td>
<td>0.039</td>
</tr>
<tr>
<td>Fruit apexe</td>
<td>0.103</td>
<td>0.19</td>
</tr>
<tr>
<td>Fruit pubescenced</td>
<td>-2.663</td>
<td>-0.717</td>
</tr>
</tbody>
</table>

All characters were continuous, except a absent, scarce, abundant; b absent, present only on midvein, scarce on main veins, abundant on main veins; c absent, present only on midvein, present in two basal tufts, scarce on main veins, abundant on main veins; d slender, medium, robust; e obtuse, acute, acuminate; f present, absent; g excluded character.

### Table 4

Results for the PCA of climatic variables. T, temperature; qtr, quarter; ppt, precipitation; Max, maximum; Min, minimum. Variables with high loadings in bold.

<table>
<thead>
<tr>
<th>Variable</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual mean T</td>
<td>0.366</td>
<td>-0.174</td>
<td>0.075</td>
</tr>
<tr>
<td>Mean diurnal range</td>
<td>-0.194</td>
<td>0.194</td>
<td>0.402</td>
</tr>
<tr>
<td>Isothermality</td>
<td>0.167</td>
<td>0.194</td>
<td>0.402</td>
</tr>
<tr>
<td>T seasonality</td>
<td>-0.194</td>
<td>-0.118</td>
<td>-0.481</td>
</tr>
<tr>
<td>Max T warmest month</td>
<td>0.254</td>
<td>0.026</td>
<td>-0.243</td>
</tr>
<tr>
<td>Min T coldest month</td>
<td>0.359</td>
<td>-0.074</td>
<td>0.063</td>
</tr>
<tr>
<td>T annual range</td>
<td>-0.278</td>
<td>0.102</td>
<td>-0.212</td>
</tr>
<tr>
<td>Mean T wettest qtr</td>
<td>0.332</td>
<td>-0.04</td>
<td>-0.173</td>
</tr>
<tr>
<td>Mean T driest qtr</td>
<td>0.352</td>
<td>-0.044</td>
<td>0.04</td>
</tr>
<tr>
<td>Mean T warmest qtr</td>
<td>0.331</td>
<td>-0.064</td>
<td>-0.205</td>
</tr>
<tr>
<td>Mean T coldest qtr</td>
<td>0.362</td>
<td>0.024</td>
<td>0.139</td>
</tr>
<tr>
<td>Annual ppt</td>
<td>-0.045</td>
<td>-0.37</td>
<td>0.188</td>
</tr>
<tr>
<td>Ppt wettest month</td>
<td>-0.078</td>
<td>-0.317</td>
<td>0.277</td>
</tr>
<tr>
<td>Ppt driest month</td>
<td>-0.008</td>
<td>-0.368</td>
<td>-0.115</td>
</tr>
<tr>
<td>Ppt seasonality</td>
<td>-0.037</td>
<td>0.282</td>
<td>0.288</td>
</tr>
<tr>
<td>Ppt wettest qtr</td>
<td>-0.086</td>
<td>-0.305</td>
<td>0.309</td>
</tr>
<tr>
<td>Ppt driest qtr</td>
<td>-0.004</td>
<td>-0.374</td>
<td>-0.104</td>
</tr>
<tr>
<td>Ppt warmest qtr</td>
<td>-0.001</td>
<td>-0.222</td>
<td>0.289</td>
</tr>
<tr>
<td>Ppt coldest qtr</td>
<td>-0.004</td>
<td>-0.369</td>
<td>-0.031</td>
</tr>
</tbody>
</table>

3. Results

Alignment was unambiguous, except for a portion of PEPC, which we excluded from the analyses (positions 836–886 of the matrix with concatenated markers). Alignments are available from TreeBASE (http://purl.org/phylo/treebase/phylows/study/TB2:S10354).

3.1. Analyses of single markers and ILD tests

The clones of ETS downloaded from GenBank (Appendix A), and the clones of NIA3 of B. permollis (J. Linares 7326), and two of the three samples of B. simaruba (Rosell 46 and Olson 1063) coalesced per sample under MP, so consensus sequences were used for terminal taxa. The remaining three samples for which we cloned NIA3 did not coalesce. Two sequences of NIA3, differing in five nucleotides, were present in two samples of B. grandifolia (Rosell 50 and Olson 1026) and grouped with other B. grandifolia collections but not with each other. To streamline combined analyses, whose main purpose was to identify relationships among simaruba complex members rather than within B. grandifolia, we generated a consensus sequence. One sample of B. simaruba (Rosell 51) had two different sequences differing mainly in a 26 bp deletion and 14 nucleotides. The other 27 B. simaruba samples had only one or the other of these two sequences. Stop codons in the coding portion (the third exon), which would indicate loss of function, were absent in both sequences, so we decided to include the two NIA3 sequences of B. simaruba as separate terminal taxa (∗ in Figs. 3 and 4) duplicating all the other markers.

We include the lengths of the datasets and results from MP by marker in Table 2. Each marker separately recovered only certain groups of species. Fig. 3 includes the results of psbA-trnH and NIA3 in isolation, and of the combined analyses of the other nuclear markers, which were congruent according to ILD tests. Pairwise ILD tests showed that significant incongruence was mostly restricted to psbA-trnH and NIA3 with respect to all the other markers. No significant incongruence was found between PEPC.
and ETS (p = 0.393), PEPC and ITS (p = 0.114), ETS and ITS (p = 0.173), and between psbA-trnH and ITS (p = 0.08).

The chloroplast marker mostly informed regarding the deepest relationships, poorly resolving the complex itself (Fig. 3). All markers showed maximum phylogenetic support (PP = 1.0, PB = 100) for the genus Bursera and mostly recovered clades of its two subgenera. A well-supported clade including the simaruba complex species and the Antillean species was recovered by all markers with BA (PP > 0.98) and by PEPC, ETS and NIAi3 under MP (PB > 82). All markers recovered a well-supported clade including all Antillean species (PP = 1.0, PB > 85), but with ETS and PEPC the clade excluded B. spinescens. With most markers, this Antillean clade appeared in a polytomy with clades of the simaruba complex, but with NIAi3 it was recovered well within the simaruba complex as sister to the grandifolia-longipes clade with strong Bayesian support (Fig. 3). Two well-supported clades of B. simaruba samples were recovered by NIAi3, each composed of one of the two different sequences found in the cloned B. simaruba sample (Fig. 3).

3.2. Analyses from combined-marker datasets

The aligned combined dataset had a length of 2974 bp. In combination, the five loci strongly supported a clade including the species of the B. simaruba complex and also of the Antillean group (Fig. 4). Although major clades in the phylogeny had good support, deeper relationships were not fully resolved. The Antillean clade was maximally supported, but its position was unresolved, as was the case also for B. permollis and B. standleyana (Fig. 4). Satellite species B. attenuata and B. roseana grouped in a clade that did not include B. simaruba, whereas B. “itzae” samples appeared intercalated among B. simaruba samples. Finally, B. ovalifolia appeared as sister of B. permollis, but this relationship had low support (Fig. 4).

3.3. Species delimitation using molecular characters

In the five-marker analyses, B. simaruba samples formed a clade with strong support, which also included B. “itzae” samples (Fig. 4). Well-supported groups with geographic affinity were recovered within the clade, including a Costa Rican Pacific coast group, a Caribbean group (with collections from the Dominican Republic, and most of those from the Yucatan and Cuba), and the Mexican Pacific coast group (Nayarit, Colima, and all but one Jalisco sample; Fig. 4). NIAi3 recovered two well-supported clades of B. simaruba, but left out one sample (Fig. 3). A clade with all but one sample of B. simaruba and B. “itzae” was recovered by ITS (PP = 0.94) and...
also with ETS (PP = 0.98) but included B. arborea and B. krusei.
When ITS, PEPC, and ETS were combined, a well-supported clade of B. simaruba samples was recovered (Fig. 3).

Two of the four satellites were non-mono-phyletic, while the other two were recovered as strongly supported clades. In the five-marker analyses, B. roseana was paraphyletic in a clade that also included a strongly-supported group of B. attenuata samples (Fig. 4). The B. attenuata group was recovered by ITS (PP > 0.97, PB = 54) and NIAi3 (Fig. 3), but included also a B. roseana accession in the case of PEPC and psbA-trnH (Fig. 3). The other B. roseana collection usually appeared in a polytomy at the next highest level in the tree. The two B. ovalifolia samples were recovered as a clade with

Fig. 4. Phylogeny of the Bursera simaruba species complex and its related Antillean species based on five markers. Bayesian 50% majority-rule consensus cladogram is shown with posterior probabilities above branches, and parsimony bootstrap values > 50 below branches. Labels on terminals indicate voucher locality (Appendix A); gray boxes highlight simaruba satellite species; line clones of the B. simaruba sample with two different versions of NIAi3.
strong support in the combined analyses (Fig. 4) and by ITS (PP = 1.0, PB = 70). The other markers did not reject its monophyly, except for psbA-trnH, which recovered a B. ovalifolia sample in a strongly supported clade with B. simaruba samples (Fig. 3). In contrast, the two samples from the type locality of B. itzeu were recovered as polyphyletic within the clade of B. simaruba samples in the combined analyses and with NIAI2 and psbA-trnH (Fig. 3), and in a polytomy including these samples with all other individual markers.

Finally, B. inaguensis appeared polyphyletic, with Cuban samples included with other Cuban species and B. nashii, whereas B. inaguensis from the Bahamas was sister to B. freenangi (Fig. 4). Under BA, ETS, ITS, and NIAI2 also supported the polyphyletic status of B. inaguensis, but its potential monophyly was not rejected under MP or by the remaining markers. All other species represented by two or more samples appeared monophyletic in the combined analysis (Fig. 4) or their monophyly could not be rejected by individual markers (Fig. 3).

3.4. Species delimitation using morphological characters

The first two discriminant functions (DF) of the analysis including all species explained 69% of the observed variance (DF1 = 55%, DF2 = 14%). DF1 indicated that fruit and upper leaflet surface pubescence and the maximum number of leaflets are the most important characters for discriminating between species, followed by the minimum number of leaflets, petiole base and lower leaflet surface pubescence, and infructescence robustness, characters with high coefficients for DF2 (Table 3). A plot of these two DFs shows B. grandifolia and B. krusei close to one another due to their pubescent fruits and leaflets (Fig. 5A), but because of its more numerous leaflets, B. grandifolia was lower on the vertical axis. Specimens of B. cinerea, a species with pubescent leaflets but glabrous fruits, appeared to the right and close to B. arborea, which has pubescent to glabrous leaflets. Another cluster was formed by B. instabilis and B. laurihuertae, which share the glabrous one to three-foliolate condition, although some specimens of B. instabilis showed scant pubescence and are higher on the vertical axis. A large group formed by B. simaruba and its satellites fell to the right because of their glabrous fruits and upper leaflet surfaces, and higher maximum leaflet number. Bursera longipes overlapped with this group, but is readily distinguished by its bluish waxy upper cuticle and small habit. Some specimens fell far from the species they belong to, e.g. the pubescent-fruited specimens of B. simaruba (Fig. 5A). Leave-one-out classification tables revealed that species were classified correctly >92% of the time except for B. simaruba, which was misclassified 10% of the time, B. attenuata 12%, B. ovalifolia 12%, B. roseana 28%, B. arborea 18%, and B. instabilis 38% (Appendix C, Supplementary material). Misclassifications of B. simaruba and its satellites were mostly due to assignment to another of these four taxonomically difficult species. Specimens of B. instabilis were often misclassified as B. laurihuertae, a situation that could be avoided by including leaflet apex and shape, which readily distinguish these species.

In the discriminant analysis of B. simaruba and its satellites, fruit pubescence was removed because it was present in only four specimens. The first two DF explained 93% (DF1 = 72%, DF2 = 21%) of total observed variance. To discriminate between B. simaruba and its satellites, infructescence robustness, leaflet lower surface and petiole base pubescence, minimum number of leaflets, and maximum fruit length were useful characters, followed by leaflet upper surface and petiolule pubescence, minimum lateral leaflet petiolule length, and minimum leaf length (Table 3). Although these characters tended to separate satellites, some specimens overlapped, highlighting the continuous ranges of some of these characters (Fig. 5B). DF1 separated B. simaruba with its robust infructescences, more numerous leaflets, large fruits, and pubescent petiole bases. On the left side of the graph, B. attenuata has the opposite characteristics. The other two satellites, B. ovalifolia and B. roseana, could not be separated on the basis of DF1 alone because of their wide ranges in leaflet lower surface pubescence and infructescence robustness (Appendix B). In contrast, DF2 helped separate B. ovalifolia because of its long lateral leaflet petiolules. Classification of B. simaruba and B. ovalifolia tended to be correct (>92%), but not of B. attenuata (88%) and B. roseana (62%, Appendix C).

3.5. Satellite species delimitation using ecological characters

The first three PCs of environmental variables accounted for 82% of the variability (PC1 = 38%, PC2 = 33%, PC3 = 11%). Six temperature variables had high loadings on PC1, six precipitation variables on
PC2, and isothermality and temperature seasonality on PC3 (Table 4). We compared ecologically B. attenuata and B. roseana, which appeared as sister taxa in our phylogeny, and also B. simaruba and B. ovalifolia, a relationship resulting from one of the possible locations of B. ovalifolia, given its unresolved position in the phylogeny, but the only one that would be of interest for this comparison given our goal of using ecological information to delimit satellite species from B. simaruba. The MANOVA of the first three PCs rejected the null hypothesis of no ecological differences between B. attenuata and B. roseana (F1,78 = 11.07, p < 0.001), but only PC2, which was related to precipitation, showed significant differences (t = –4.717, p < 0.001). A non-parametric MANOVA of the first three PCs rejected the hypothesis of no differences between B. ovalifolia and B. simaruba (F1,378 = 52.17, p < 0.001). Mann–Whitney tests found differences between these two species regarding temperature (PC1: W = 18360.5, p < 0.001), precipitation (PC2: W = 22111.5, p < 0.001), and seasonality (PC3: W = 19012.5, p < 0.001).

4. Discussion

4.1. Species delimitation and the taxonomy of the simaruba complex

We recovered the two most widespread species of the complex as monophyletic. Given its morphological variability and wide range, along with the confusing taxonomy of the complex, we fully expected to recover a paraphyletic B. simaruba. However, not only was the species monophyletic but the apparently weak geographic structuring of its populations (Fig. 4) implies genetic cohesion between distant populations via pollen and seed dispersal (Dunphy and Hamrick, 2007; Greenberg et al., 1995; Sousa, 1969). Bursera simaruba monophyly is based on considering B. itzae a likely synonym of B. simaruba, an issue discussed below. Like B. simaruba, and despite a long history of nomenclatural changes and wide distribution (Rzedowski et al., 2007; Table 1, Fig. 2), B. ovalifolia collections from distant localities formed a single clade. Although one of the possible positions of B. ovalifolia is as sister to B. simaruba and their ranges overlap (Fig. 2), their ecological conditions differ in temperature, precipitation, and seasonality, suggesting they are not ecologically interchangeable and supporting the validity of B. ovalifolia as a separate species.

The three species recovered as non-monophyletic highlight the need for taxonomic revision of these taxa. The paraphyletic pattern of B. roseana could suggest that B. attenuata is derived from this species and reciprocal monophyly has not been attained, or that both species are in fact the same taxon. Bursera attenuata is conceived of as a lowland entity and B. roseana its highland counterpart, so we expected their environmental conditions to differ strongly. Surprisingly, only precipitation differed (PC2). Given that their pubescence patterns overlap considerably, and to a lesser extent their distributions (Fig. 2), it may be that the two species are a single entity that varies with elevation (McVaugh and Rzedowski, 1965). With regard to B. itzae, it was polyphyletically embedded within B. simaruba (Fig. 4). Bursera itzae is distinguished from B. simaruba by its relatively long inflorescences (15–20 cm; Lundell, 1968). Not known when B. itzae was described, however, is that this characteristic varies considerably within B. simaruba, especially in Guatemala (M. Veliz, pers. comm.) and adjacent Mexico, where we have observed inflorescences up to 22 cm long. Although our results await confirmation based on unequivocally identified fertile material, we suspect that B. itzae simply represents a variant of B. simaruba with long inflorescences. Finally, Cuban and Bahamian collections of B. inaguensis appeared in two different places in the phylogeny (Fig. 4). Our Cuban collections differ markedly from Bahamian material in leaf and inflorescence features, making it likely that Cuban B. inaguensis represents a distinct species.

As was the case with molecular characters, species of the simaruba complex could for the most part be reliably distinguished in discriminant analyses of morphological characters (Fig. 5, Appendix C), despite their pronounced morphological plasticity (Appendix B). Although pubescence patterns were crucial for discriminating species (Table 3), the ranges and patterns of expression of indumentum are poorly understood in the complex. Nevertheless, pubescence may vary between populations, individuals, and even leaflets of the same leaf. In addition, the expression of pubescence in one part of the leaf may not be independent of its production in other parts. For example, the development of trichomes on the lower surface of leaflets seems necessary for production on the upper surface. Ascertaining the independence or lack thereof of pubescence states will determine how they can be coded and interpreted in future analyses. Because of high levels of variation, accurate identification of some species in the clade, particularly the satellites, would seem to require extensive collections from different parts of individual trees, as well as from numerous individuals of the same population. In cases of unusual combinations of morphological characters, identification via molecular characters may be the only alternative at present.

4.2. The origin of satellite species

Our results do not support the origin of the satellites B. attenuata, B. ovalifolia, and B. roseana from B. simaruba, suggesting instead that their similarity is due to convergence or parallelism. Distant B. attenuata and B. roseana are clearly not derived from B. simaruba (Fig. 4). As for B. ovalifolia, support for its position as sister to B. per mollis in the analyses with all markers was low, so one of its possible positions is as sister to B. simaruba (Fig. 4), making the origin of B. ovalifolia from B. simaruba a possibility. Nevertheless, support for a ovalifolia-permollis clade becomes strong in the absence of the chloroplast marker (see Section 4.3 and Fig. 7, Supplementary material), a topology that would translate into rejecting the origin of B. ovalifolia from B. simaruba. The similarity between B. simaruba and the satellites is in a suite of features, many of which occur in other species but all of which tend to be found together only in these four species (Fig. 6, Appendix B). Given that most species in the genus are trees of tropical deciduous forests or shrubs of xerophytic scrubs, it is likely that the massive size of some populations of B. simaruba and its satellites derives from their repeated invasions of moister forests (Fig. 6).

4.3. Congruence and testing membership in the simaruba complex

Although the recovery of some species as members of the simaruba complex clade reaffirmed previous results (Becerra, 2003), the placement of others was heretofore uncertain. This was the case of the unifoliolate Oaxacan endemic B. laurhuertae (Rzedowski and Calderón, 2000), whose position among species with several leaflets makes it another example of the changes in leaflet number widespread in the genus (Rzedowski and Kruse, 1979). Likewise, the Central American B. standleyana and B. permollis were confirmed as belonging to the complex. Although unresolved, the distant placement of B. permollis from B. grandifolia suggests that the synonymy proposed for these two species (Daly, 1993; Porter and Pool, 2001) is unnecessary.

Although we recovered a single clade including all Antillean species (Figs. 3 and 4), its position as sister to or part of the simaruba complex could not be resolved. Based on our five marker reconstruction, the placement of the Antillean clade as sister to the whole simaruba complex (Weeks and Simpson, 2004) or to the grandifolia-longipes clade, as suggested by NIAI3, results in a tree with the same number of steps. Because of the numerous synapomorphies supporting the Antillean clade itself, it seems
unlikely that its ambiguous placement is due to hybridization or lineage sorting, but rather that more characters are needed. This lack of information could also explain the unresolved position of B. standleyana, whose molecular changes were also mostly autapomorphic. The position of B. permollis within a strongly supported clade of B. simaruba samples in the analyses of psbA-trnH contrasts with its unresolved position in the analyses by other markers (Fig. 3), and also with its strongly supported location as sister to B. ovalifolia when psbA-trnH is excluded from analysis (Fig. 7, Supplementary material). This incongruence between the chloroplast and nuclear markers could suggest genetic influence of B. simaruba in this sample.

The two main clades within B. simaruba reflect two NIAI3 sequence variants that differ mainly in a characteristic deletion. In general, each B. simaruba sample had one or the other of these variants, but both occurred together in one sample. These two sequence types may represent different alleles or paralogues. Differences between the sequences are comparable to those found between alleles in Scaevola (Howarth and Baum, 2005), which, together with the absence of stop codons, would suggest that the versions are alleles rather than pseudogenes. At the same time, duplications of NIAI3 have been found in the tribe Canarieae within Burseraceae (Weeks, 2009), and in other families (Hamman-Khalifa et al., 2007; Yi et al., 2008), although the sequences found in B. simaruba did not correspond to those in Canarieae. Whether they represent alleles or paralogues with potential associated lineage sorting, in the genus these sequences seem to be restricted to B. simaruba (J.A. De-Nova, unpublished results), and they would not affect the phylogenetic reconstruction of the simaruba complex.

5. Conclusion

As in many species complexes, molecular tools proved invaluable in our analysis of the simaruba clade, in which high levels of intraspecific morphological plasticity coexist with marked overlap between species. Results pinpointed priority species that require further sampling for molecular and morphological work, and also characters that need to be reexamined to refine the taxonomy of the group. The molecular phylogeny of the simaruba complex allowed us to reject the origin of three satellite species from their purported ancestor, B. simaruba, and to ascribe their morphological similarity as likely being due to convergence or parallelism. Disentangling the effects of similarity due to common ancestry, convergence, or parallelism, if they can in fact be separated, is a major challenge of species complex biology, especially as evolutionary developmental biology blurs some of these distinctions (Abouheif, 2008; Kim et al., 2003). The simaruba complex appears to provide excellent examples for studies of adaptive divergence, convergence, or parallelism, and our phylogeny offers an essential framework for any such effort using the simaruba complex as a model system.

Acknowledgments

This project was supported by grants from CONACYT (46475), PAPIIT/ DGAPA (IN228207), US NSF (0919179) and the Cactus and Succulent Society of America. J.A.R. acknowledges the support of CONACYT through a graduate scholarship, the Programa de Posgrado en Ciencias Biológicas, UNAM, and the Graduate Student Grant from the American Society of Plant Taxonomists. We thank Drs. Jerzy Rzedowski and Graciela Calderón for their kind and generous orientation; Laura Márquez for her help with sequencing of the samples; Cristina Martínez for samples of Cuban species; Hilda Flores, Verónica Juárez, Sergio Zamudio, Victor Steinmann, Yocupitzia Ramírez and the rest of the staff of MEXU and IEB for facilitating the measurement of specimens; Katherine Renton for samples; René Cerritos and Martín García Varela for their help with cloning; Erika Aguirre, Laura Espinosa, and Calixto León for their help in the
Appendix A.

Voucher information, sample labels and GenBank accession numbers (psbA-trnH, PEFC, ITS, ITS, NIAI2) for the species sampled. All vouchers are deposited in MEXU, unless otherwise stated. Mx = Mexico; NA = not available. GenBank accession numbers of previously published sequences are designated by a superscript corresponding to the following publications: bBecerra (2003), aWeeks and Simpson (2004), aWeeks et al. (2005), aWeeks and Simpson (2007).

b. arborea: Jal = Chamela, Jalisco, Mx (Rosell 16), GQ377870, GQ377934, GQ378006, GQ378070, GQ378140.

b. attenuata: Nay-1 = Tepic, Nayarit, Mx (R. Medina 3412), GQ377871, GQ377935, GQ378007, GQ378071, GQ378141; Nay-2 = Santa Ma. Del Oro, Nayarit, Mx (R. Medina 3417), GQ377872, GQ377936, GQ378008, GQ378072, GQ378142; Nay-3 = Santa Ma. Del Oro, Nayarit, Mx (R. Medina 3423), GQ377873, GQ377937, GQ378009, GQ378073, GQ378143.

b. cinerea: Oax-1 = Jayacatlán, Oaxaca, Mx (Rosell 40), GQ378784, GQ378795, GQ3788010, GQ3788014; Oax-2 = Teotitlán, Oaxaca, Mx (R. Medina 1221), GQ377875, GQ377939, GQ378081, GQ378145.

b. freenngiae: Baham = Fairchild Tropical Garden, Bahamas (79410F, TEK), AY309392b, AY309366b, AY309311–AY309313b, GQ378076, GQ378152.

b. grandifolia: Mor-1 = Alpuyeca, Morelos, Mx (Rosell 50), GQ377876, GQ377940, GQ378012, GQ378077, GQ378146; Mor-2 = Atlacholoya, Morelos, Mx (Rosell 5), GQ377877, GQ377941, GQ378013, GQ378078, GQ378148; Mor-3 = Atlacholoya, Morelos, Mx (Rosell 1026), GQ377878, GQ377942, GQ378014, GQ378079, GQ378149-GQ378150; Gro = Guzmán, Oaxaca, Mx (R. Medina 3977), GQ377879, GQ377943, GQ378005, GQ378151.

b. inaguensis: Cuba-1 = Rafael Freyre, Holguín, Cuba (42572, HAC), GQ377880, GQ377944, GQ378016, GQ378081, GQ378153; Cuba-2 = Bahía de Naranjos, Holguín, Cuba (42569, HAC), GQ377881, GQ377945, GQ378017, GQ378082, GQ378154; Baham = Inagua, Fairchild Tropical Garden (64269M, TEK), AY309393b, AY309367b, AY309314–AY309316b, GQ378083, GQ378155.

b. instabilis: Jal-1 = Chamela, Jalisco, Mx (Rosell 11), GQ377882, GQ377946, GQ378018, GQ378084, GQ378156; Jal-2 = Chamela, Jalisco, Mx (Rosell 12), GQ377883, GQ377947, GQ378019, GQ3778085, GQ378157; Jal-3 = Chamela, Jalisco, Mx (Rosell 13), GQ377884, GQ377948, GQ378020, GQ378086, GQ378158; Jal-4 = Chamela, Jalisco, Mx (J.C. Montero 923), GQ377885, GQ377949, GQ378021, GQ378087, GQ378159.

b. itzca: Guat-1 = Lake Petén Itzá, Guatemala (J.C. Montero 1029), GQ377886, GQ377950, GQ378022, GQ378088, GQ378160; Guat-2 = Lake Petén Itzá, Guatemala (J.C. Montero 1030), GQ377887, GQ377951, GQ378023, GQ378089, GQ378161.

b. krusei: Gro = Rio Papagayo, Guerrero, Mx (R. Medina 4045), GQ377888, GQ377952, GQ378024, GQ378090, GQ378162.

b. laurihuertae: Oax = Tehuantepec, Oaxaca, Mx (Olson 1119), GQ377889, GQ377953, GQ378025, GQ378091, GQ378163.
Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2010.08.004.

References


Goloboff, P., 1999. NONA (NO NAME) ver. 2. Published by the author, Tucumán, Argentina.

Goloboff, P., 1999. NONA (NO NAME) ver. 2. Published by the author, Tucumán, Argentina.


