SYSTEMATICS AND PHYLOGENY

Reestablishment of *Protium cordatum* (Burseraceae) based on integrative taxonomy

Gabriel Damasco,[1] Douglas C. Daly,[2] Alberto Vicentini[3] & Paul V.A. Fine[1]

1 *Department of Integrative Biology and University and Jepson Herbaria, University of California, Berkeley, California 94720, U.S.A*
2 *Institute of Systematic Botany, The New York Botanical Garden, Bronx, New York 10458, U.S.A*
3 *Coordenação de Dinâmica Ambiental e Programa de Pós-Graduação em Botânica, Instituto Nacional de Pesquisas da Amazônia, Manaus, Amazonas 70390-095, Brazil*
Address for correspondence: *Gabriel Damasco, gdamasco@berkeley.edu*

DOI https://doi.org/10.1002/tax.12022

Abstract Species delimitation remains a challenge worldwide, but especially in biodiversity hotspots such as the Amazon. Here, we use an integrative taxonomic approach that combines data from morphology, phylogenomics, and leaf spectroscopy to clarify the species limits within the *Protium heptaphyllum* species complex, which includes subsp. *cordatum*, subsp. *heptaphyllum*, and subsp. *ulei*. Molecular phylogeny indicates that populations of subsp. *cordatum* do not belong to the *P. heptaphyllum* clade, while morphology and near-infrared spectroscopy data provide additional support for the recognition of a separate taxon. *Protium cordatum* (Burseraceae) is reinstated at species rank and described in detail. As circumscribed here, *P. cordatum* is endemic to white-sand savannas located in the Faro and Tucuruí Districts, Pará State, Brazil, whereas *P. heptaphyllum* is a dominant and widespread plant lineage found in Amazonia, the Cerrado, and the Brazilian Atlantic Forest. We present an identification key to *P. cordatum* and closely related lineages and a detailed taxonomic description of *P. cordatum*, including habitat and distribution, a list and images of diagnostic features. This study demonstrates the importance of using multiple tools to characterize and distinguish plant species in highly diverse tropical regions.

Keywords Amazon; *campina*; ddRAD; plant systematics; *Protium heptaphyllum*; near-infrared spectroscopy; white-sand forest

Supporting Information may be found online in the Supporting Information section at the end of the article.

■ INTRODUCTION

The efficiency and accuracy of taxonomy have greatly increased with new technologies and analytical advances (Bik, 2017). Many species have recently been described based on different kinds of data that range from comparative morphology and anatomy to phylogeography, population genetics, and functional ecology (Dayrat, 2005; Schlick-Steiner & al., 2010). The potential for such integrative taxonomic approaches has not yet been fully embraced in botany, particularly in the tropics, where biodiversity studies are especially needed. In highly diverse forests such as the Amazon, morphological overlap among closely related lineages complicates taxonomic delimitation. However, new data from molecular phylogenetics and DNA barcoding are contributing to the delimitation of species (Kress & al., 2005; Gonzalez & al., 2009). Furthermore, high-throughput genome sequencing has enhanced the resolution of population phylogeography and species delimitation analysis (e.g., Leaché & al., 2014; Fišer & al., 2018), leading to more accurate species limits.

In addition to molecular phylogenetics, Fourier-transformed near-infrared spectroscopy (FT-NIR) has demonstrated great potential for discriminating tropical plant species (Durgante & al., 2013; Féret & Asner, 2013; Lang & al., 2017). The spectrometer generates an absorbance response that is a function of the chemical composition and internal anatomy of leaves, which has been shown to be conserved within populations of the same species in a large study with more than 1000 tropical tree species (Asner & al., 2014). Analyzing the spectral signatures of near-infrared reflectance of herbarium specimens can allow the discrimination of morphotypes even when there is a high phenotypic overlap among samples, and informative vegetative and reproductive traits are unknown (Durgante & al., 2013; Prata & al., 2018).

Protium (Burseraceae) has undergone rapid diversification in the Neotropics (Fine & al., 2014) and includes many cryptic species. A good example is *P. heptaphyllum* (Aubl.) Marchand, one of the most dominant trees in the Neotropics (Ter Steege & al., 2013). *Protium heptaphyllum* sensu lato (s.l.) is also one of the most widespread plant taxa in the Neotropics and inhabits different biomes and related ecosystems (i.e., Amazonia – terra-firme forest and white-sand ecosystems; Cerrado and Pantanal – gallery forests and seasonally dry tropical forests; Guiana Shield – rocky savannas; and Brazilian Atlantic Forests – coastal white-sands and rain forests). The morphology of *P. heptaphyllum* can be quite

Article history: Received: 5 March 2018 | returned for (first) revision: 24 June 2018 | (last) revision received: 18 Oct 2018 | accepted: 26 October 2018 ||
Associate Editor: Jessica Mary Prebble ||

variable, which may result from acclimation or adaptive divergence of populations inhabiting different soil types or climatic niches (e.g., Fine & al., 2013). For instance, individuals found in sandy and nutrient-poor soils appear as shrubs or stunted trees with bifurcated trunks and coriaceous leaves, while individuals that inhabit more nutrient-rich clay soils can be tall canopy trees with chartaceous leaves.

Many infraspecific taxa have been assigned to *Protium heptaphyllum* since Jean Baptiste C. Fusée Aublet first published *Icica heptaphylla* as part of the *Histoire des plantes de la Guiane françoise* in 1775 (Swart, 1942). Currently, three valid subspecies have been recognized in the *P. heptaphyllum* species complex (subsp. *cordatum*, subsp. *heptaphyllum*, subsp. *ulei*) yet no consensus has been reached regarding the degree to which phenotypic variation in the species complex corresponds to taxonomic entities (Daly, 1992).

Protium heptaphyllum subsp. *cordatum* was first described as *P. cordatum* (at species rank) by Huber (1909) and subsequently treated as an infraspecific taxon within *P. heptaphyllum* due to quantitative overlap in measurements of the reproductive characters (Daly, 1992). Here, we use a multidisciplinary approach that combines next-generation sequencing, morphological analyses, and NIR spectral data and sample 24 populations across the geographic range of the *P. heptaphyllum* species complex to evaluate species limits within the group. Since taxonomy has moved towards being an integrative science, we believe that formal species should represent evolutionarily diverged populations that: (1) form highly supported monophyletic clades according to molecular evidence, (2) have limited gene flow among closely related lineages or sister groups, and (3) exhibit conserved morphological features that enable their recognition. We present evidence here that subsp. *cordatum* must be reinstated at species rank, as treated by Huber (1909).

■ MATERIALS AND METHODS

Taxon sampling. — We sampled eight individuals of *P. cordatum* at the lectotype locality (Faro, Pará, Brazil) as well as 23 individuals of *P. heptaphyllum* s.l. from throughout its range (Fig. 1A). Since *P. cordatum* was previously treated as a subspecies of *P. heptaphyllum* (Daly, 1992), we also sampled 8 individuals that co-occurred with *P. cordatum* in the type locality (*P. cordatum* shrubs inhabiting the white-sand savanna and *P. heptaphyllum* adult trees inhabiting the adjacent forest; Fig. 1B). We aimed to test if these morphologically distinct populations from adjacent habitats represented genetically diverged populations. The morphological, molecular and spectral data were collected from the same samples. In addition, ten closely related outgroup species were selected based on a molecular phylogeny of the Protieae tribe (Fine & al., 2014), i.e., *P. brasiliense*, *P. dawsonii*, *P. icicariba*, *P. kleinii*, *P. krukoffii*, *P. ovatum*, *P. pillosum*, *P. trifoliolatum*, *P. unifoliolatum*, and *P. widgrenii* (Appendix 1; supplemental Table S1).

Morphological analyses. — First, we generated a character matrix with 59 continuous and 98 discrete traits (supplemental Table S2) to examine the morphological variability in multidimensional space. Non-informative characters and missing data were excluded. We used the R package clustvarsel v.2.3.3 (Scrucca & Raftery, 2014) to reduce the dimensionality of the data by selecting the set of principal components most useful for discrimination without *a priori* information about groups. Vegetative traits were measured on 56 specimens of *P. heptaphyllum* s.l. and 12 specimens of *P. cordatum*. Reproductive traits were measured on 23 specimens of *P. heptaphyllum* and 12 specimens of *P. cordatum*. Some specimens of the latter did not bear flowers or fruits during the sampling period. Therefore, descriptions of reproductive structures (e.g., corolla length, flower density and petal indumenta) were obtained from herbarium specimens collected in the same biogeographic domain (e.g., Cerrado, Amazonia, etc.).

To test the hypothesis that *P. cordatum* does not belong within *P. heptaphyllum* as an infraspecific taxon, we fit the number of morphological clusters using the normal mixture models (NMMs) implemented in the R package mclust v.5.0 (Scrucca & al., 2016). The Bayesian information criterion (BIC; Schwarz, 1978) was used to evaluate the best-fit number of morphological groups according to each NMM (Cadena & al., 2018). The vegetative characters with high loading values in the principal component analysis (PCA) were (1) leaf petiole length, (2) the maximum number of leaflets, (3) plant height and (4) specific leaf area. The reproductive traits with high loading PCA values were (1) flower density per inflorescence, (2) corolla length, and (3) petal indumenta density.

DNA library preparation. — We extracted high-quality genomic DNA from eight samples of *P. cordatum*, 23 samples of *P. heptaphyllum* s.l. widely distributed throughout the Amazon, Atlantic Forest, and the Cerrado, and 10 outgroup species (see supplemental Table S1 for details about the samples included in the phylogenetic analysis). DNA was extracted from ca. 100 mg of leaf tissue preserved in silica or from herbarium specimens when silica-dried leaves were not available. Extractions followed a modified version of the DNEasy Plant mini kit protocol (Qiagen, Crawley, U.K.). Double-digest RAD-seq libraries were prepared for high-throughput sequencing following Peterson & al. (2012). Detailed information on the library preparation procedures is available as supplemental Appendix S1. DNA was digested with *Sph*I-HF and *Eco*RI-HF enzymes. DNA libraries were sequenced on five lanes of an Illumina HiSeq 4000 at the University of Berkeley QB3 facility.

Assembly and phylogenetic analysis. — We used a bioinformatics pipeline implemented in custom Perl scripts that integrate various external programs for processing ddRAD-seq data. The pipelines are available in https://github.com/CGRL-QB3-UCBerkeley/RAD. Paired-end raw fastq reads were first de-multiplexed based on the sequences of internal barcodes with a tolerance of one mismatch. The reads were then filtered to trim adapter contaminations and

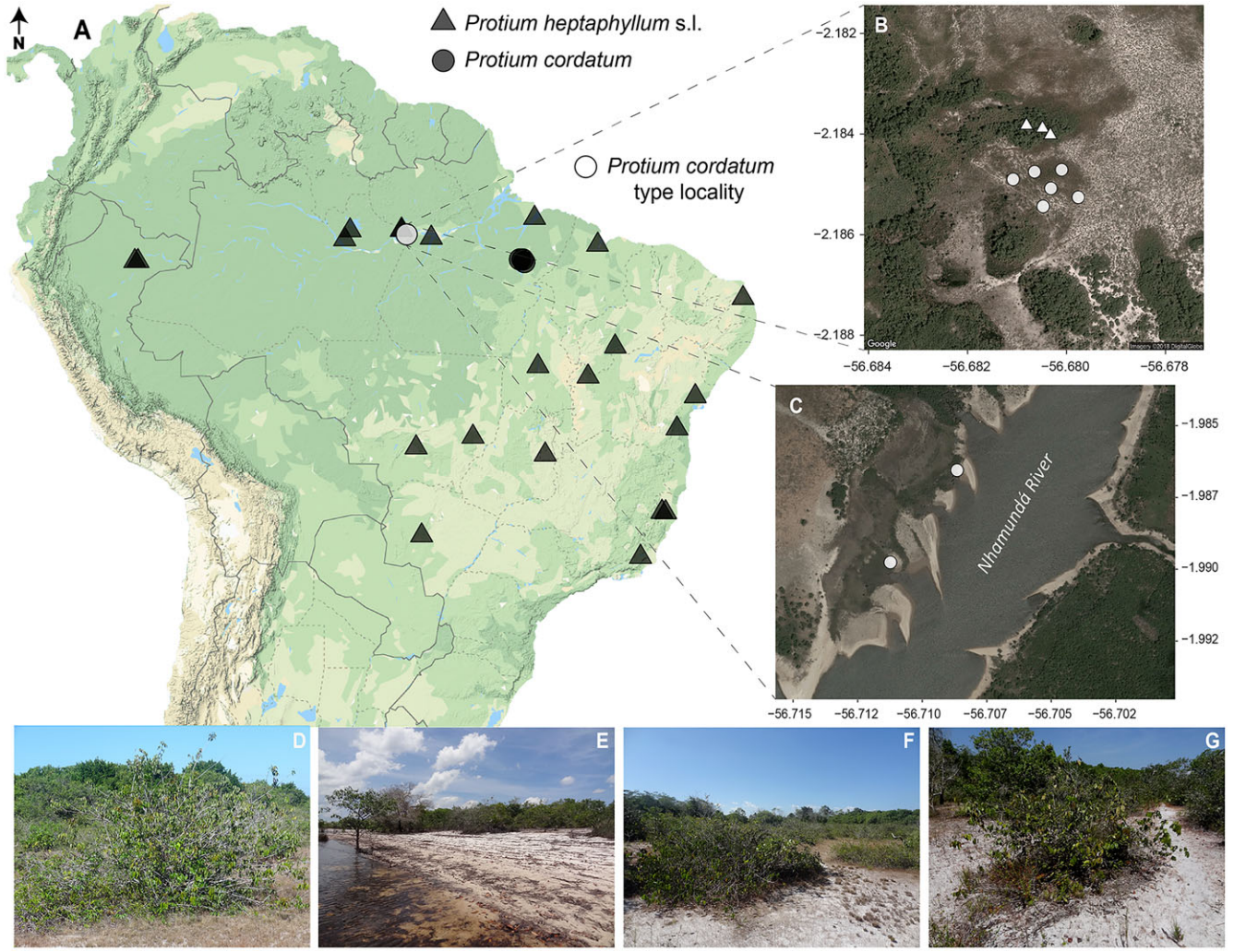

Fig. 1. A, Distribution of the specimens included in this study. Circles represent *Protium cordatum* (white circle: *P. cordatum* samples from Faro, Pará, Brazil; black circle: additional populations of *P. cordatum* from Tucuruí, Pará, Brazil not sampled in this study), and triangles represent *P. heptaphyllum* s.l. **B,** Populations sampled along a parapatric habitat ecotone (white circles: *P. cordatum* individuals; white triangles: *P. heptaphyllum* s.l. individuals) at the exact location where the lectotype was collected. **C,** Individuals of *P. cordatum* sampled at the margin of the Nhamundá River, Faro, Pará, Brazil. **D–G,** Examples of habitat and photos of *P. cordatum*, a native and endemic shrub of Amazonian white-sand savannas.

low-quality reads. The resulting cleaned reads were clustered with a sequence identity threshold of 0.95, and potential paralogs, loci containing repeats, and/or loci that were likely derived from incomplete restriction enzyme digestion were removed. The resulting RAD loci from each individual were then combined and collapsed into a non-redundant master reference set. Cleaned paired-end reads from each sample were aligned to the reference using Novoalign v.2 (http://www.novocraft.com/products/novoalign). Details about raw sequence treatment and data assembly are provided as supplemental Appendix S1.

Phylogenetic inference was based on a maximum likelihood criterion. We used the GTRGAMMA nucleotide substitution mode with 1000 bootstrap replicates in RAxML v.8.1.16 (Stamatakis, 2014). The molecular dataset consists of a concatenated matrix with 1387 filtered loci and 7762 informative SNPs. Our analysis tested for the monophyly of *P. cordatum* and aimed at reconstructing phylogenetic relationships of *P. cordatum* with respect to *P. heptaphyllum* s.l.

Near-infrared spectroscopy (NIR). — NIR technology was used to test the hypothesis that *P. cordatum* represents a distinct entity from *P. heptaphyllum* s.l. For each specimen, a single spectrum was collected from three different dried leaflets using a desktop Spectroscopy Analyzer from Thermo Fisher Scientific, model Antaris II (Antaris, Waltham, Massachusetts, U.S.A.). A total of ten specimens of each taxon were included. All of these specimens were also included in the morphological and molecular phylogenetics analyses described above. Each spectrum represents the average of 16 scans including the absorbance of 1557 values sampled at intervals of 8 cm^{-1} within wavelengths of 4000–10,000 nm. An opaque black lid was placed over the reading area to avoid light scattering. A background calibration was performed automatically during every other

reading. In total, 12 individuals of *P. cordatum* and 15 individuals of *P. heptaphyllum* s.l. were analyzed.

We used the Kolmogorov-Smirnov (KS) test to determine if the spectral curves of *P. cordatum* and *P. heptaphyllum* s.l. show significant differences. The KS-test is a non-parametric test that does not require any prior assumption about the distribution of the data (Lopes & al., 2009). We ran the KS-test comparing all curves based on *D*-values within and across populations of *P. cordatum* and *P. heptaphyllum* s.l.

■ RESULTS

Morphological analyses. — The PCA based on vegetative and reproductive characters supports the hypothesis that *P. cordatum* is morphologically distinct from *P. heptaphyllum* s.l. The first two principal components were most useful for group discrimination (Fig. 2A). NMMs ignoring both principal components explained only 3% of the morphological variance. All NMMs indicated two distinct morphological groups based on the high BIC-values and the plot shows the highest empirical support (ordinate) and the optimum number of morphological groups (abscissa) supporting the hypothesis of two distinct species based on vegetative and reproductive characters. (Fig. 2B). Regarding vegetative traits, *P. heptaphyllum* s.l. has more pairs of leaflets (juga), longer petioles (Fig. 2C), higher specific leaf area (a proxy of leaf thickness), and higher plant height (Fig. 2D). *Protium cordatum* has shorter petals and lower flower density along the inflorescence axes (Fig. 2E) and a denser petal indumentum (Fig. 2F).

Phylogenomics. — *Protium cordatum* is strongly supported as monophyletic (BS = 100%; Fig. 3) and is sister to a clade that includes *P. ovatum* and *P. dawsonii* (BS = 100%) and the monophyletic *Protium heptaphyllum* species complex (BS = 100%). The sister-group relationship between *P. heptaphyllum* s.l. and *P. ovatum*+*P. dawsonii* is poorly supported (BS = 64%), meaning that there is uncertainty regarding the exact position of the clade *P. ovatum*+*P. dawsonii* with regard to *P. heptaphyllum* and *P. cordatum*. The population of *P. heptaphyllum* sampled in the neighboring forest across the habitat ecotone is distantly related to *P. cordatum* (samples from Fig. 1B are bolded in Fig. 3), indicating that both populations are genetically highly distinct despite the large potential for gene flow.

Leaf spectroscopy. — The KS-test showed significant (p-value $< 2.2^{-16}$) dissimilarity between the spectra of *P. cordatum* and neighboring *P. heptaphyllum* populations (Fig. 4A), which corresponds to the spectral discontinuity observed in the PCA ordination space (Fig. 4B). According to *D*-values, the spectral variability within populations of *P. cordatum* and *P. heptaphyllum* s.l. is significantly lower (p-value $< 1.26^{-11}$) than across populations (Fig. 4C). Interestingly, the spectral region between 4000 and 5300 nm wavelength is more variable in terms of absorbance readings in comparison to 5300–7000 nm. In the latter, the variation within populations can be as high as among populations. Although the spectral signatures of *P. cordatum* and *P. heptaphyllum* s.l. are significantly different in the 7000–10,000 nm wavelength region, the overall absorbance variation within this spectral interval is lower than in the 4000–5300 nm region (coefficient of variation: 4000–5300 nm = 0.81 to 0.83; 7000–10,000 nm = 0.10 to 0.32).

■ DISCUSSION

In this study, we combined morphology, phylogenomics and spectroscopy to improve species delimitation within the *P. heptaphyllum* species complex. Data from different sources provide a consistent picture of the ideal taxonomic placement for *P. cordatum* within the Protieae tribe of Burseraceae (Fine & al., 2014). Based on these multidisciplinary results, we conclude that *P. cordatum* should be reinstated as a formal species as initially described by Huber (1909), and not treated as an infraspecific taxon within *P. heptaphyllum*.

Taxonomic implications. — In 1909, *P. cordatum* was first described by Huber as a habitat specialist shrub in the white-sand savannas in Amazonia (also known as *campinas*). In 1992, D.C. Daly proposed a new status and a new combination of *P. cordatum* as a subspecies of *P. heptaphyllum*. He justified this decision by stating "there is a distinct geographic component to the differences between them, but they can be distinguished only by the rather quantitative characters" and concluded that "further material of subsp. *cordatum* is needed before their differences can be defined adequately and the transfer made" (Daly, 1992: 298).

We analyzed additional material of *P. cordatum*, including additional samples from the type locality, and present a key with the most relevant differences of discrete and continuous morphological traits among *P. cordatum* and *P. heptaphyllum* s.l. We also found that vegetative (leaf petiole, number of juga, and specific leaf area) and reproductive characters (flower density, petal indumenta, corolla length) are discontinuous with minimal phenotypic overlap. Our phylogenetic results showed that *P. cordatum* is not closely related to *P. heptaphyllum*, and thus is likely genetically distinct from populations of *P. heptaphyllum* sampled at parapatric habitat ecotones. Individuals of *P. cordatum* sampled from the exact type locality represent a monophyletic clade to the exclusion of all other lineages of *P. heptaphyllum* s.l., and the *P. ovatum*+*P. dawsonii* clade.

We found that *P. cordatum* was not closely related to the other white-sand specialist taxon within the *P. heptaphyllum* complex, *P. heptaphyllum* subsp. *ulei*, which presents consistent morphological differences (i.e., habit, leaf shape, floral and fruit traits) and a broader geographic distribution over the Amazon basin (Peru, Venezuela, Guyana, Central and Northern Amazonia). A study based on gene flow estimates and hybridization tests including a larger sample size was conducted to investigate the species limits within the *P. heptaphyllum* clade,

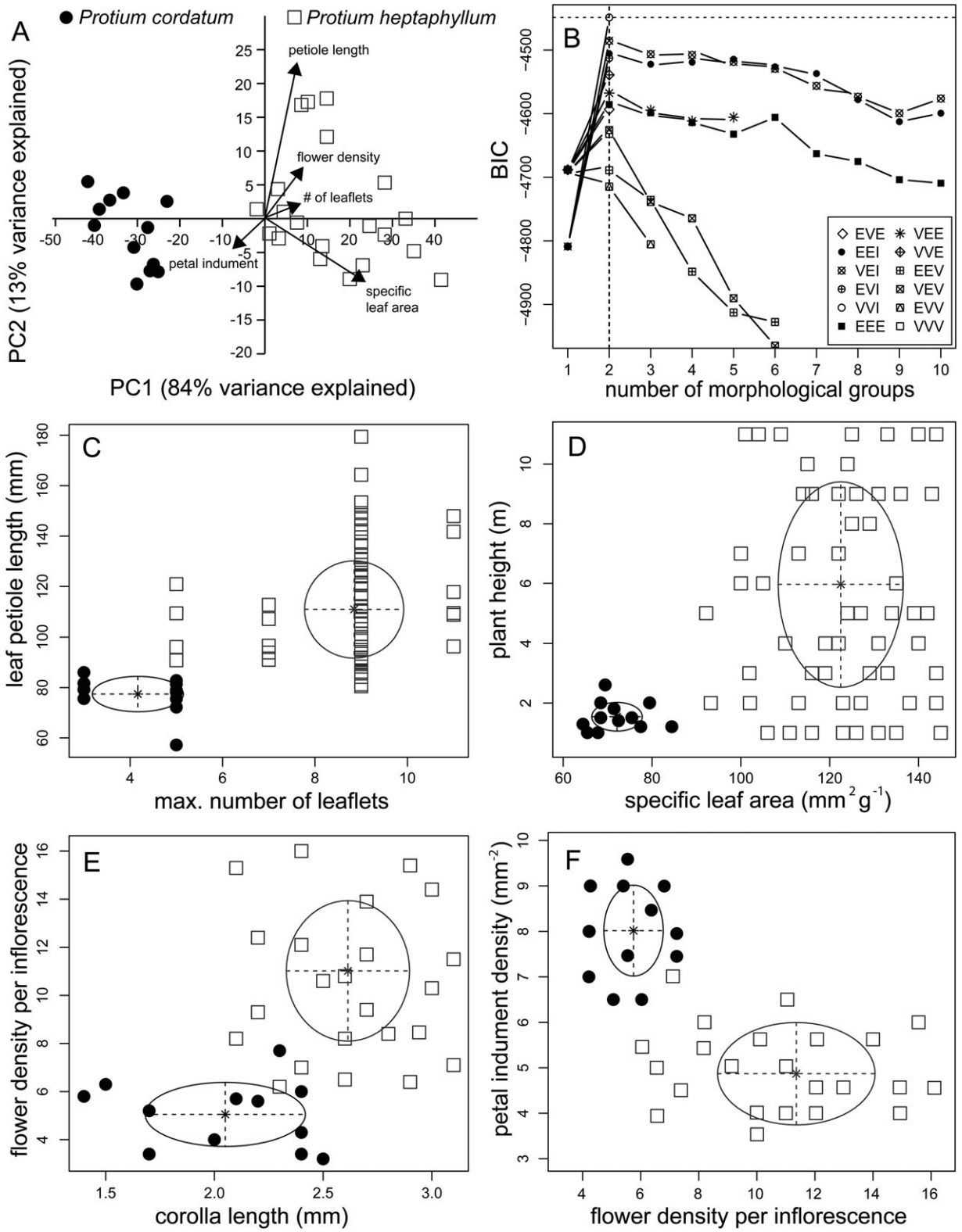

Fig. 2. A, Principal component analysis (PCA) of morphological characters of *Protium cordatum* and *P. heptaphyllum* s.l. Arrows correspond to PCA loadings of most informative morphological characters. **B,** Bayesian information criterion (BIC) from normal mixture model (NMM) analysis using 12 model parameterizations and up to 10 morphological groups. Different abbreviations (e.g., EVE, EEI, VEI, etc.) with respective symbols and line types encode different model parameterizations. **C–F,** A projection of the morphological characters, with different symbols indicating the classification corresponding to the best model as determined by the NMM analysis. The component means are marked, and ellipses with axes are drawn corresponding to their covariances. Vegetative traits were measured on 56 individuals of *P.heptaphyllum* s.l. and 12 individuals of *P. cordatum*. Reproductive traits were measured on 23 individuals of *P. heptaphyllum* and 12 individuals of *P. cordatum*.

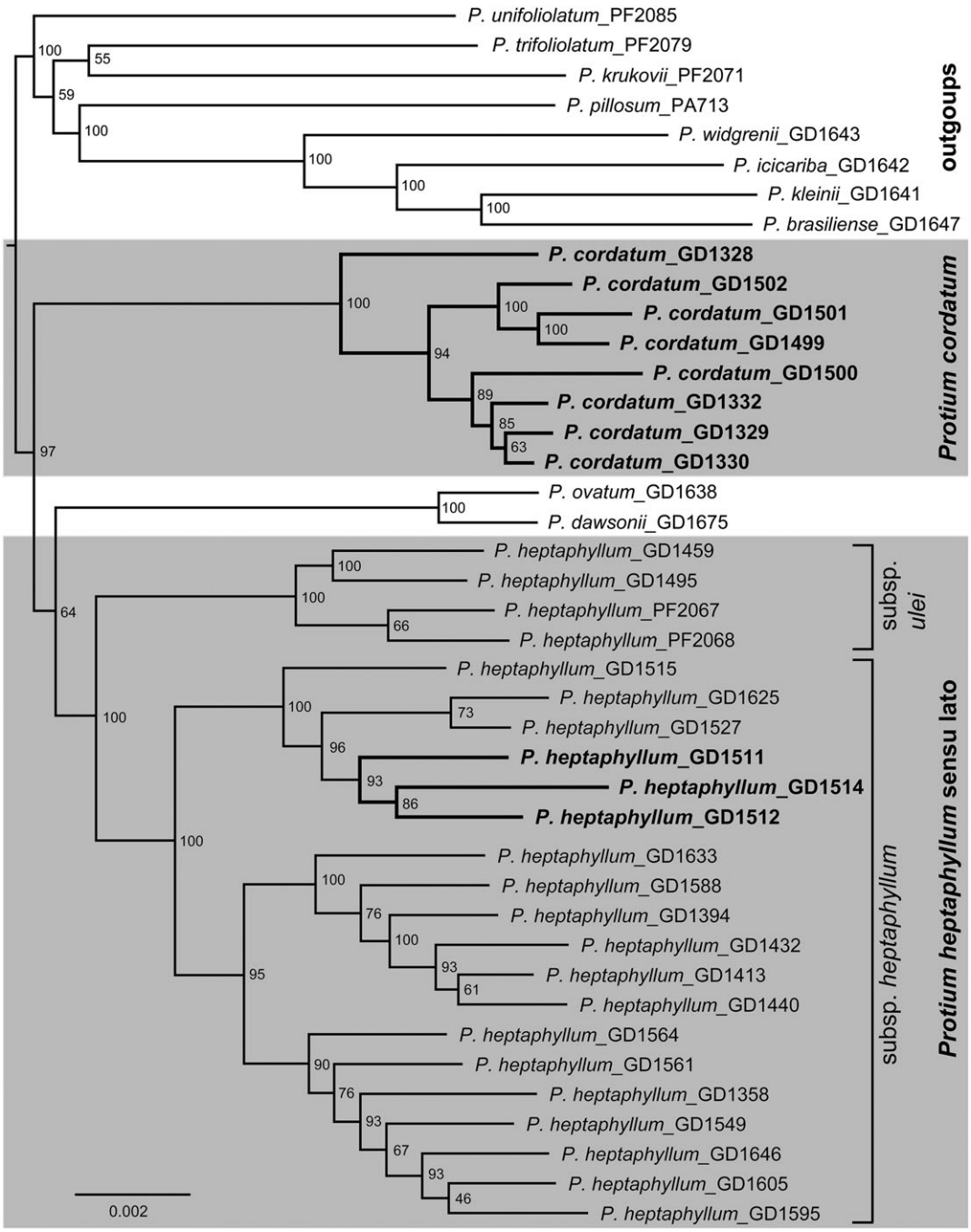

Fig. 3. Maximum likelihood phylogeny based on a genome-wide ddRAD-seq dataset with 1333 filtered loci and 21,359 informative SNPs. Grey boxes highlight the monophyletic clade of *Protium cordatum* as a new outgroup of *P. heptaphyllum* s.l. Bolded branches in the phylogeny correspond to individuals sampled in the parapatric habitat ecotone shown in Fig. 1B.

including subsp. *ulei*, and an updated taxonomic treatment is in preparation as a monograph (Damasco & al., in prep.). Here, we focus solely on the reestablishment of the *P. cordatum* lineage at species rank due to the strong support for it being morphologically distinct and outside the clade corresponding to *P. heptaphyllum* s.l.

The importance of in-depth integrative studies of plant lineages. — Our results directly address concerns regarding the "hyperdominance phenomenon" in the Neotropics (Ter Steege & al., 2013; Cardoso & al., 2017). Based on the plot-dataset published by the Amazon Tree Diversity Network, *P. heptaphyllum* s.l. is classified as the eleventh most dominant taxon in the Amazon, especially common in the white-sands in the upper Rio Negro basin and the Guiana Shield. But, as we demonstrate here, a taxon that was considered part of a hyperdominant clade represents at least two independent lineages. If we ignore the possibility that dominant clades may include different putative species, we

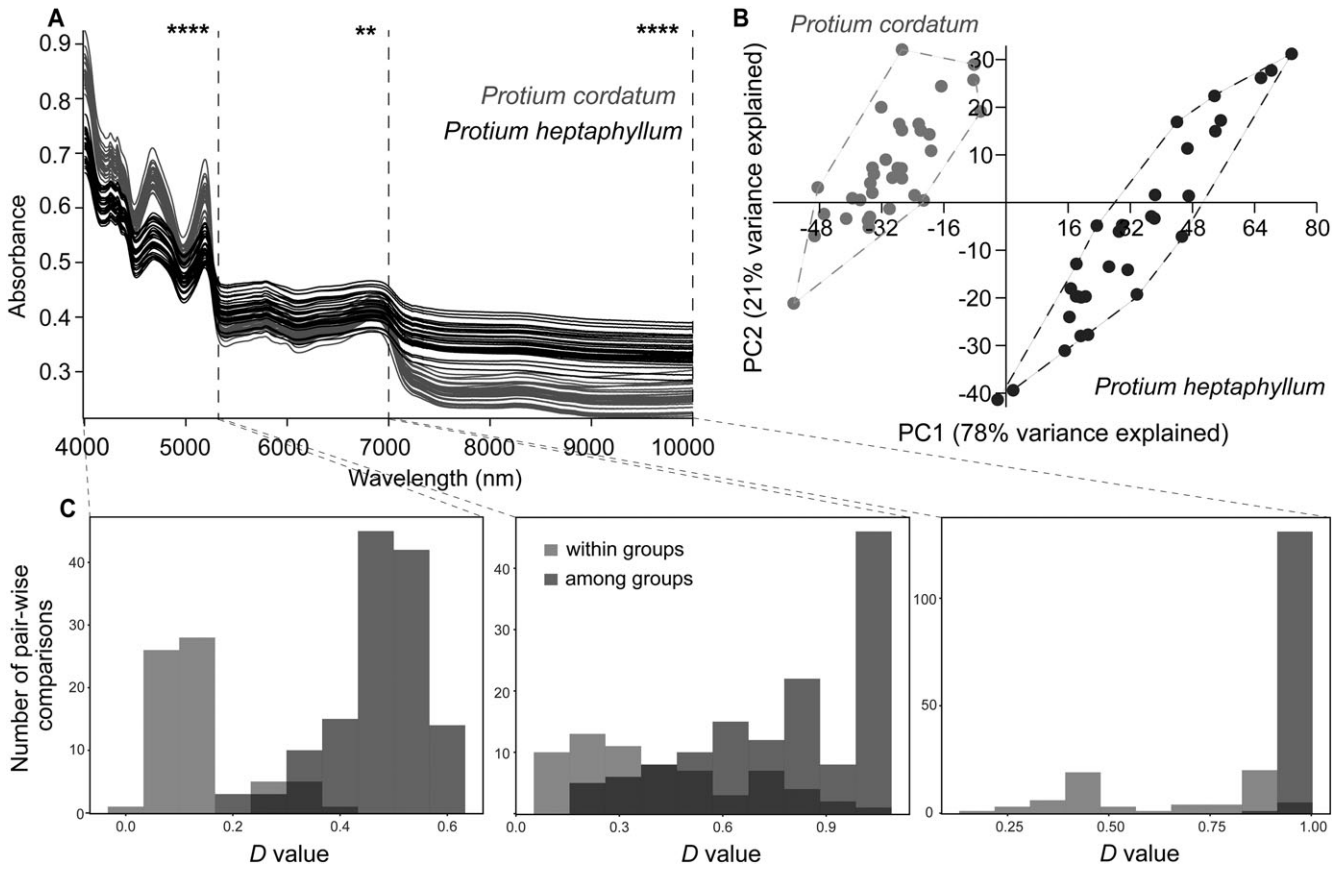

Fig. 4. A, Representation of full near infra-red (FT-NIR) spectra for ten specimens of *Protium cordatum* and *P. heptaphyllum* s.l., respectively. The spectral readings are expressed as absorbance values between 4000 and 10,000 nm wavelength and each spectrum consists of an average of 1557 absorbance values for three different leaflets per specimen. The full spectra were subdivided into three spectral regions due to different absorbance variation within and across populations. **B,** Two-dimensional principal component analysis based on FT-NIR spectral data. **C,** Histograms of pair-wise comparisons indicating the absorbance similarity based on *D*-values estimates within and across populations and compared among three spectral regions (4000–5300 nm; 5300–7000 nm; 7000–10,000 nm).

could seriously underestimate the diversity of tree species and make errors predicting the relative abundances of plant communities in the Neotropics.

Although there has been a great effort to estimate the accurate diversity of plants in the Amazon and the Americas (Cardoso & al., 2017; Ulloa & al., 2017), as many as 10%–20% of species may remain undescribed in the tropics (Pimm & Joppa, 2015). While it is possible that the majority of these unknown species are located in areas that have yet to be visited by botanists, we suggest that the number of undescribed species is likely to be even higher if we consider that dominant and widespread plant species, like *P. heptaphyllum* s.l., may contain many hidden lineages that should not be considered conspecific. Our findings highlight the importance of additional in-depth studies of individual Neotropical species, as well as the importance of using multiple lines of evidence in taxonomy to delimit taxa which will, in turn, lead to more accurate estimates of species diversity in the tropics.

Implications for the near-infrared (NIR) technology. — Near-infrared technology has been used by plant taxonomists to discriminate several Neotropical plant groups. Recent studies of two diverse Amazonian genera, *Eschweilera* (Durgante & al., 2013) and *Protium* (Lang & al., 2015), showed that NIR spectra could accurately discriminate distinct species with over 96% success. Regarding closely related plant groups, this technology has been effective to discriminate cryptic species within *Pagamea* (Rubiaceae) (Prata & al., 2018). In addition to the morphological and phylogenetic results, the NIR data indicate that *P. cordatum* and *P. heptaphyllum* have distinct spectral signatures. Even though they are closely related lineages, the spectral variation within each taxon was significantly lower than the spectral variation between taxa.

The leaf spectra can be correlated with the chemical composition and the anatomical structure inside the leaves (Asner & Martin, 2008; Féret & Asner, 2013). More specifically, changes in cell wall composition, such as polysaccharides, proteins, and phenolic compounds are believed to be evolutionarily conserved among different species (Asner & al., 2014). As NIR application to tropical botany advances, more research is needed to understand better the characteristics behind the spectral values and which factors might have a significant

impact on leaf absorbance signatures along the spectrum. For instance, we found little variation across a large section of the wavelength spectrum, as well as high redundancy of absorbance values among similar spectral regions or neighboring wavelength sites (as also noticed by Durgante & al., 2013). More studies are needed to optimize the usage of spectral data in plant taxonomy, and we believe that future research using NIR technology should investigate what the most informative regions of the wavelength spectrum are and whether there is an optimum set of wavelength bands that may increase the accuracy of discriminant models in taxonomy.

■ TAXONOMY

Protium cordatum Huber in Bol. Mus. Goeldi Hist. Nat. Ethnogr. 5: 433. 1909 ≡ *P. heptaphyllum* subsp. *cordatum* (Huber) Daly in Brittonia 44: 298. 1992 – Lectotype (designated by Swart in Recueil Trav. Bot. Néerl. 39: 330. 1942): Brazil. Pará: Faro, Campo do Tigre, 21 Aug 1907 (m fl), *Ducke s.n*. (B [destroyed]; isolectotypes: F [photo!], MG No. 8463!, NY barcode 00345733!, RB No. 20522!).

Description. — *Shrubs* ca. 1.5–3 m tall, crown open. *Stems* highly branched from base; outer bark light to dark gray, thin, often rough from high density of lenticels and light-colored lichens, inner-bark white or light yellow; branchlets, striate and brown towards apices, lenticels sparse. *Resin* flammable, transparent, and viscous when fresh, dark grey with crystalline texture when dry. *Leaves* glabrous, ca. 6–12 cm long, often 1–3-jugate; petiole ca. 2–7 cm long, 1–2 mm diam. near base, often striate with appressed fine hairs to 0.05 mm long; interjuga 1–1.5 cm long; basal petiolules 3–9 mm long, other lateral petiolules 2–6.5 mm long, terminal petiolule 4–15 mm long; lateral and distal pulvinuli inconspicuous; leaflet blades ca. 5–9.5 cm long, 1.5–5 cm wide, elliptic to ovate, highly coriaceous, drying dark green to reddish brown abaxially, grey to green or light brown adaxially, faces dull, apex cuspidate or rarely acute, the acumen to 10 mm long, base cordate or occasionally rounded, often asymmetric, margin entire, secondary vein framework festooned-brochidodromous, costal secondaries in 6–14 pairs, the spacing irregular, decreasing toward apex and base, the angle slightly decreasing toward base, course essentially straight, occasionally one intersecondary vein per pair of costal secondaries and parallel to them, adaxial face with midvein narrowly prominulous, secondary veins mostly flat, tertiaries flat, irregular-polygonal, quaternaries flat, irregular-reticulate, abaxial face with midvein and secondaries prominulous, tertiaries mostly prominulous, irregular-polygonal, quaternaries flat, irregular-reticulate.

Staminate inflorescences, 4–15 mm long, 4–10 mm diam. near base, secondary axes 2–9 mm long, all axes with dense malpighiaceous hairs to 0.1 mm long, bristles (short, fine, erect white hairs) also present; bracts 0.3–0.7 mm long on primary axes and 0.2–0.6 mm on secondary axes, elliptic to deltate, apex acute; bracteoles 0.1–0.3 mm, coriaceous, with dense, thick, white malpighiaceous hairs; pedicel 0.5–1.5 mm long, 0.2–0.6 mm diam., cylindrical, with pubescence as on inflorescence axes. *Staminate flowers* 4-merous, 2–3 mm long; calyx 0.4–0.6 mm long, 1.5–2.5 mm diam., exceeding disk or nearly equal, not divided to base, the lobes mostly inconspicuous and separated by a flat sinus, few flowers with visible lobes 0.2–0.5 mm long, 0.7–1 mm wide with occasional acute to acuminate apex, abaxial pubescence as on inflorescence axes; corolla ovate to urceolate, 1.3–2.2 mm long, 0.5–1 mm wide, mainly light yellow with occasional orange-reddish tonality (especially in bud), apiculum 0.1–0.3 mm long, broadly ovate, somewhat coriaceous, inflexed, abaxial pubescence as on calyx except trichomes longer and denser toward apex, adaxially mostly glabrous or with sparse bristles, margin sparsely papillate; stamens 8, equal, inserted on outer edge of disk, 1–1.8 mm long, anthers 0.4–0.6 mm long, oblong-ovate in dorsiventral view, lanceolate in profile, filaments cylindrical to compressed with dense bristles; annular disk globose, 0.5–0.9 tall, glabrous or with sparse bristles, essentially discoid with narrowly conical center; pistillode 0.2–0.4 mm, exceeding disk; pistillode with high density of long malpighiaceous hairs ca. 0.3 mm (longer relative to corolla indumentum).

Pistillate inflorescences 5–15 mm long, 5–8 mm diam. near base, secondary axes 4–9 cm long; bracts 0.4–0.7 mm long on primary axes and ca. 0.4 mm long on secondary axes, ovate to rarely deltate, apex acute to acuminate; bracteoles 0.1–0.3 mm, coriaceous with dense, thick white malpighiaceous hairs; pedicel 0.5–1.5 mm long, 0.2–0.6 mm diam., cylindrical, with pubescence as on inflorescence axes. *Pistillate flowers* 2.1–2.8 mm long; calyx 0.3–0.6 mm, 1.1–1.8 mm diam., height relative to disk as on staminate flowers, pubescence as on inflorescence axes; corolla 1.5–2 mm long, 0.8–1.2 mm wide, ovate to urceolate, somewhat coriaceous, apiculum 0.2–0.3 mm long; staminode insertion and shape as on staminate flowers, 0.8–1.2 mm long, anthers 0.4–0.5 mm long; annular disk 1–1.2 mm long, 0.9–1.1 mm diam., glabrous or with sparse malpighiaceous hairs; style 1.1–1.5 mm long, stigma 0.2–0.4 mm long, sessile, erect, depressed-globose; ovary globose-ovoid, glabrous or with sparse malpighiaceous hairs.

Fruit maturing red, mostly globose to slightly oblique-ovoid, ventricose, dry size ca. 6–8 mm long, 4–6 mm diam. (1 locule), 8–10 mm long, 7–9 mm diam. (2–4 locules), glabrous with sparse bristle hairs near the base, smooth, drying slightly wrinkled, apex mostly obtuse to round (in globose fruits), occasionally acute (in ovoid fruits), base slightly substipitate (stipe ca. 0.8–1 mm long), rounded to truncate above stipe. *Fruiting pedicel* 0.5–1.5 mm long, 0.4–0.6 mm diam., cylindrical.

Distribution and habitat. — *Protium cordatum* is a rare shrub endemic to white-sand savannas and sandy riverbanks. This species is most common in seasonally flooded areas of white-sand ecosystems (G. Damasco, pers. obs.). It occurs in white-sand areas of the Nhamundá River (Fig. 1C), Faro municipality, and was previously reported from the

white-sand savannas near the Tocantins River, Tucuruí municipality, Pará, Brazil. We did visit the Tucuruí region and looked for populations of *P. cordatum* in areas where the specimens have been collected before but we could not find them. We examined the specimens collected in Tucuruí, and they were morphologically identical to populations collected in Faro (the type locality). There is a good chance that additional populations of *P. cordatum* could be found in white-sand patches located nearby Oriximiná or savannas near Santarém (both in Pará State, Brazil).

Uses. — No uses are reported, but *P. cordatum* contains copious amounts of resin like many species of Burseraceae. In many species of *Protium*, these resins are often burned as light sources, incense or used as medicines (e.g., Siani & al., 2012).

Diagnostic features. — *Protium cordatum* is morphologically similar to *P. heptaphyllum* s.l. but differs with its shrubby habit (vs. stunted to tall tree habit in *P. heptaphyllum* s.l.) and coriaceous leaflets that are disposed perpendicular to the leaf branch axis (vs. chartaceous and often smooth leaflets with a non-perpendicular disposition in *P. heptaphyllum* s.l.) (Fig. 5M). Furthermore, the leaf petiole is shorter in *P. cordatum* (77.4 ± 7.3 mm) than in *P. heptaphyllum* s.l. (110.8 ± 19.3 mm), as is the terminal petiolule (26.9 ± 3.5 mm and 53.8 ± 15 mm, in *P. cordatum* and *P. heptaphyllum* s.l. respectively). The secondary veins of *P. cordatum* are usually impressed rather than prominent in *P. heptaphyllum* s.l. and tertiary venation is barely visible in *P. cordatum* due to darker coloration and flatness at the adaxial face compared to the abaxial face (Fig. 5K,L). The inflorescence has fewer flowers (4–8) than *P. heptaphyllum* s.l. (6–16) and the corolla length is usually shorter in *P. cordatum* (up to 15 mm) than in *P. heptaphyllum* s.l. (up to ca. 70 mm). The calyx and corolla of *P. cordatum* have a high density of malpighiaceous hairs on the abaxial surfaces, while the calyx and corolla of *P. heptaphyllum* s.l. are glabrous or sparsely bristly. In addition, the pistillode in the staminate flowers of *P. cordatum* has long malpighiaceous hairs while the pistillode is glabrous or sparsely bristly in *P. heptaphyllum* s.l. Fruits of *P. cordatum* are often globose, whereas *P. heptaphyllum* s.l. often has obliquely ovoid fruit with an occasionally obtuse apex. The resin is more viscous and darker when dry in *P. cordatum* in comparison to that of *P. heptaphyllum* s.l., which is more transparent and more abundant and waterier. Diagnostic features for *P. ovatum* and *P. dawsonii* will be covered in a future publication because more samples are needed to review their taxonomy. Both taxa are savanna specialists inhabiting nutrient-scarce soils in the Brazilian Cerrado (habitat also known as cerrado sensu stricto). *Protium ovatum* and *P. dawsonii* are found as stunted shrubs, and their leaflets are usually ovate and often have a serrate margin. According to Jose Cuatrecasas (author that described *P. dawsonii*), *P. ovatum* differs from *P. dawsonii* by having hairs present on the adaxial leaflet surface. However, after examining several herbarium specimens in NY, this morphological character is not consistent in *P. dawsonii*. A detailed taxonomic revision including more samples is necessary to resolve the taxonomy of the clade *P. ovatum*+*P. dawsonii*.

Additional specimens examined. — Brazil. Pará, Campina de Santa Rosa, lat −3.7661, long −49.6725, *J. Ramos 626* (INPA); Campina de Santa Rosa, ramal da BR-422, lat −3.7661, long −49.6725, *J. Ramos 1147* (INPA); Campinas de Santa Rosa, lat −3.7661, long −49.6725, *J. Revilla 8502* (INPA); Campos a Leste de Faro, *A. Ducke s.n*. (MG); Margem direita da BR-263, km 16, lat −3.75105, long −49.5473, *M.G. Silva 5503* (INPA, MG); BR-263, km 16, lat −3.75105, long −49.5473, *M.G. Silva 5806* (INPA, MG); Approx. 25 km S of Tucuruí, just off the old BR-422 at the junction with an abandoned railroad bed, lat −3.99065, long −49.6736, *D.C. Daly 1080* (INPA, NY); Margem direita da PA-149, km 35, lat −3.76457, long −49.6736, *J. Ramos 883* (INPA); PA-149, lat −3.76457, long −49.6736, *F.E.L. Miranda 395* (INPA); PA-149, lat −3.76457, long −49.6736, *F.E.L. Miranda 408* (INPA).

Nomenclatural notes. — Huber described *Protium cordatum* in 1909 with a citation of only one specimen, indicated as "*Ducke 8463*", 21 Aug 1907, but with no mention of the type or the herbarium in which the specimen was deposited. Later in 1942, in a review of *Protium* and allied genera in Burseraceae, J.J. Swart included two specimens under the name *P. cordatum*, one of which is the specimen ("*Ducke 8463*") cited by Huber (1909). In his review, Swart stated that both specimens were deposited in B. In 1992, upon making the new combination *P. heptaphyllum* subsp. *cordatum*, Daly indicated that the number "8463" of Ducke's collection cited by Huber (1909) was actually the catalog number of the MG Herbarium, where Huber was working at the time, and that the correct collection number of this specimen should be "*Ducke 20522*". Daly also noted three isolectotypes deposited in MG, NY, and RB and declared that the lectotype designated by Swart deposited in B was almost certainly destroyed. After a new careful inspection of all isolectotypes mentioned in Daly (1992), we conclude that there is no collection number associated with Ducke's specimen. The numbers "8463" and "20522" correspond to specimen catalog numbers at MG and RB, respectively. In addition, both Swart (1942) and Daly (1992) mentioned in the protologue that the publication year of *P. cordatum* was 1908. Volume 5 of the *Boletim do Museu Goeldi de Historia Natural e Ethnographia* was organized in two fascicles, the first one dated February 1908, and the second one March 1909. After examining the original manuscript, we noticed that the description of *P. cordatum* was published in the second fascicle. Therefore, the correct year of Huber's publication is 1909, rather than 1908.

Key to the identification of *Protium cordatum* and members of the *P. heptaphyllum* species complex

1. Shrubs ca. 1.5–3 m tall, terminal petiolules 4–15 mm, lateral and terminal pulvinuli inconspicuous; leaflet blades often coriaceous, base cordate, rarely truncate and rounded; tertiary veins flat on adaxial surface; petals to 2.5 mm long with dense malpighiaceous hairs towards the apex; anthers

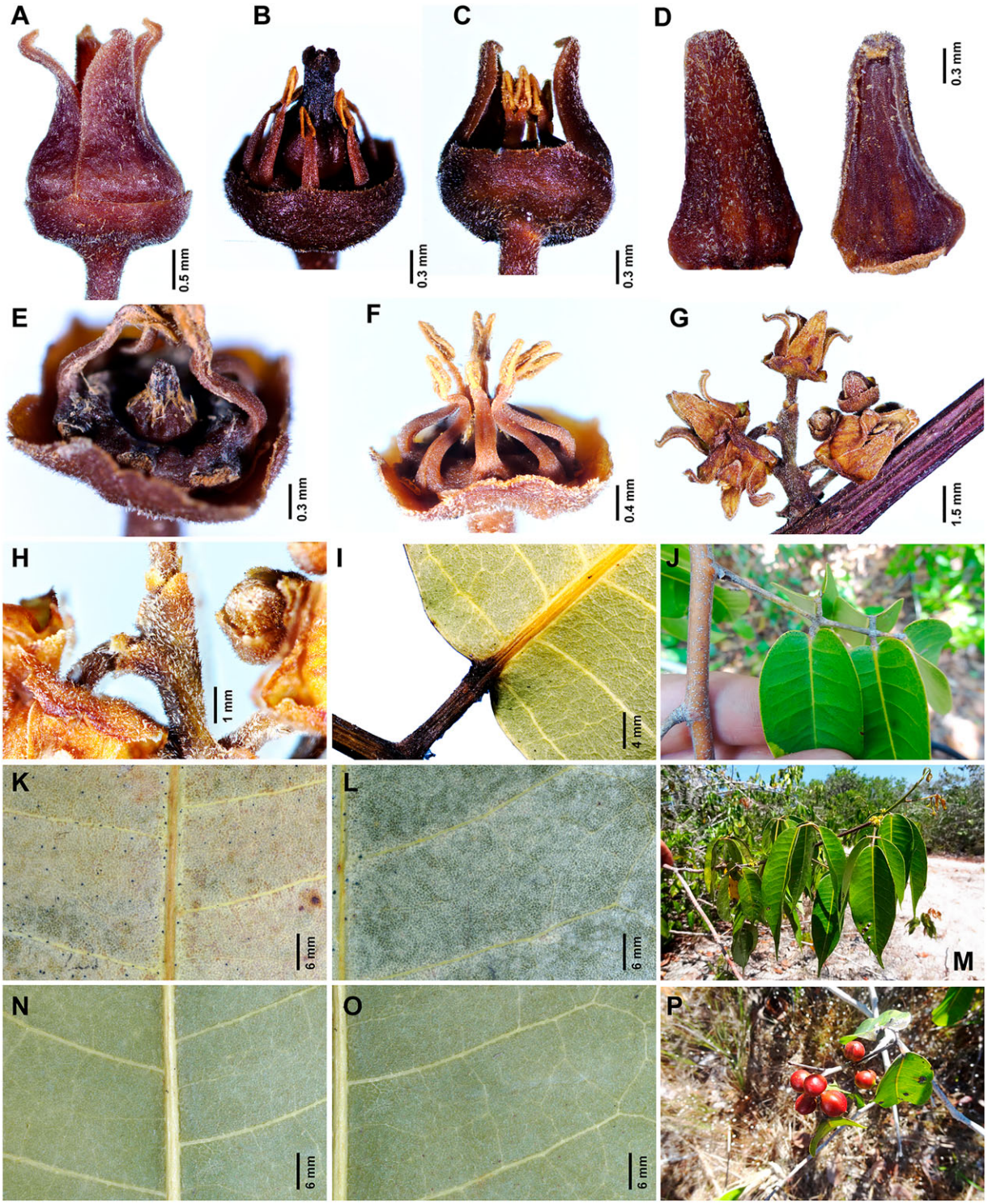

Fig. 5. *Protium cordatum*. **A,** Flower 4-merous with corolla molding a urceolate shape. **B,** Pistillate flower with globose ovary and stigma erect and sessile. **C,** Staminate flower with calyx exceeding disk or nearly equal, not divided to base, the lobes mostly inconspicuous and separated by a flat sinus. **D,** Corolla with dense, thick, white malpighiaceous hairs abaxially and mostly glabrous or with sparse bristle hairs adaxially. **E,** Pistillode with high density of long malpighiaceous hairs. **F,** Anthers lanceolate in profile and filaments cylindrical to compressed with dense bristles. **G,** Inflorescence with all bracts and bracteoles coriaceous and dense pubescence as on the corolla. **H,** Bracteoles coriaceous with dense, thick, white malpighiaceous hairs. **I,** Dry leaflet with cordate base or occasionally rounded and often asymmetric. **J,** Fresh leaflets collected in the field with cordate base. **K–L,** The secondary veins are usually impressed rather than prominent and tertiary venation is barely visible due to darker coloration and flatness at the adaxial face compared to the abaxial face (N and O). **M,** Fresh stem and leaves with 3 juga collected in the field. **N–O,** Leaflet abaxial face. **P,** Mature fruits mostly globose to slightly oblique-ovoid.

0.45–0.6 mm; pistillode with malpighiaceous hairs to 0.3 mm; fruit to 1.5 cm ..... ***Protium cordatum***

1. Treelets and trees to 15 m tall; terminal petiolules 5–45 mm, lateral and terminal pulvinuli often present; leaflet blades chartaceous to moderately coriaceous, base acute to obtuse; tertiary veins prominent to prominulous on adaxial surface; petals to 4.2 mm long with glabrous or with scattered appressed hairs; anthers 0.45–1.2 mm; pistillode glabrous or less often with scattered ascending hairs; fruit to 2.3 cm 2
2. Leaflets drying greenish-tan or brown, chartaceous, margin flat and entire near base; leaflet tertiary veins prominulous on abaxial surface; petals lanceolate, 2.8–4.2 mm; anthers 0.7–1.2 mm; pistil 1.7–2.5 mm, style 0.8–1.1 mm; fruit 1.4–2.3 cm..... ***P. heptaphyllum*** subsp. ***heptaphyllum***
2. Leaflets drying reddish with slightly caudate to acuminate apex, moderately coriaceous, margin often revolute near base; leaflet tertiary veins prominulous to flat on abaxial surface; petals lanceolate to ovate, 2–2.6 mm; anthers 0.45–0.6 mm; pistil 1.2–1.7 mm, style 0.5–0.75 mm; fruit 1–1.4 cm..... ***P. heptaphyllum*** subsp. ***ulei***

■ AUTHOR CONTRIBUTIONS

GD, DCD and PVAF conceived the presented idea; GD developed the fieldwork, morphological measurements, molecular laboratory work, leaf spectral readings and all the analysis in this manuscript; DCD, AV and PVAF verified the analytical methods and supervised the findings of this work; AV and PVAF funded part of the research cost. All authors discussed the results and contributed to the final manuscript. — GD, https://orcid.org/0000-0001-9768-520X; DCD, https://orcid.org/0000-0003-1205-9491; AV, https://orcid.org/0000-0002-5906-9358; PVAF, https://orcid.org/0000-0002-0550-5628

■ ACKNOWLEDGMENTS

We thank Bruce Baldwin and Lúcia Lohmann for the insightful comments that enhanced the quality of the manuscript, and Ke Bi, Chodon Sass and Diego Salazar for the supervision on the molecular laboratory work. We also thank Ricardo Perdiz for double checking the historical taxonomic notes, Clarissa Fontes, Ricardo Haidar and Daniel Carneiro for the assistance during fieldwork and botanical collection, and the undergraduate students Colby Gekko, Keana Richmond and Maria Juarez for lab work assistance. We are thankful for the substantial support offered by the New York Botanical Garden (NYBG), the Brazilian Institute for Amazonian Research (INPA) Herbarium and the University and Jepson Herbaria at UC Berkeley. This research is part of the Ph.D. thesis of the first author (GD), and the work was funded by the Science Without Borders Fellowship (202437/2011-6), the Universal/CNPq Grant (459966/2014-6), the National Science Foundation (Awards DEB 1254214 and DEB 1311117), the Center for Latin American Studies at UC Berkeley, SigmaXi, the International Society of Plant Taxonomists and the International Association for Plant Taxonomy.

■ LITERATURE CITED

Asner, G.P. & Martin, R.E. 2008. Spectral and chemical analysis of tropical forests: Scaling from leaf to canopy levels. *Remote Sensing Environm.* 112: 3958–3970. https://doi.org/10.1016/j.rse.2008.07.003

Asner, G.P., Martin, R.E., Carranza-Jiménez, L., Sinca, F., Tupayachi, R., Anderson, C.B. & Martinez, P. 2014. Functional and biological diversity of foliar spectra in tree canopies throughout the Andes to Amazon region. *New Phytol.* 204: 127–139. https://doi.org/10.1111/nph.12895

Bik, H.M. 2017. Let's rise up to unite taxonomy and technology. *PLOS Biol.* 15: e2002231. https://doi.org/10.1371/journal.pbio.2002231

Cadena, C.D., Zapata, F. & Jiménez, I. 2018. Issues and perspectives in species delimitation using phenotypic data: Atlantean evolution in Darwin's finches. *Syst. Biol.* 67: 181–194. https://doi.org/10.1093/sysbio/syx071

Cardoso, D., Särkinen, T., Alexander, S., Amorim, A.M., Bittrich, V., Celis, M., Daly, D.C., Fiaschi, P., Funk, V.A. & Giacomin, L.L. 2017. Amazon plant diversity revealed by a taxonomically verified species list. *Proc. Natl. Acad. Sci. U.S.A.* 114: 10695–10700. https://doi.org/10.1073/pnas.1706756114

Daly, D.C. 1992. New taxa and combinations in *Protium* Burm. f. Studies in neotropical Burseraceae VI. *Brittonia* 44: 280–299. https://doi.org/10.2307/2806927

Dayrat, B. 2005. Towards integrative taxonomy. *Biol. J. Linn. Soc.* 85: 407–415. https://doi.org/10.1111/j.1095-8312.2005.00503.x

Durgante, F.M., Higuchi, N., Almeida, A. & Vicentini, A. 2013. Species spectral signature: Discriminating closely related plant species in the Amazon with near-infrared leaf-spectroscopy. *Forest Ecol. Managem.* 291: 240–248. https://doi.org/10.1016/j.foreco.2012.10.045

Féret, J.-B. & Asner, G.P. 2013. Tree species discrimination in tropical forests using airborne imaging spectroscopy. *IEEE Trans. Geosci. Remote Sensing* 51: 73–84. https://doi.org/10.1109/TGRS.2012.2199323

Fine, P.V.A., Zapata, F., Daly, D.C., Mesones, I., Misiewicz, T.M., Cooper, H.F. & Barbosa, C. 2013. The importance of environmental heterogeneity and spatial distance in generating phylogeographic structure in edaphic specialist and generalist tree species of *Protium* (Burseraceae) across the Amazon basin. *J. Biogeogr.* 40: 646–661. https://doi.org/10.1111/j.1365-2699.2011.02645.x

Fine, P.V.A., Zapata, F. & Daly, D.C. 2014. Investigating processes of neotropical rain forest tree diversification by examining the evolution and historical biogeography of the Protieae (Burseraceae). *Evolution* 68: 1988–2004. https://doi.org/10.1111/evo.12414

Fišer, C., Robinson, C.T. & Malard, F. 2018. Cryptic species as a window into the paradigm shift of the species concept. *Molec. Ecol.* 27: 613–635. https://doi.org/10.1111/mec.14486

Gonzalez, M.A., Baraloto, C., Engel, J., Mori, S.A., Pétronelli, P., Riéra, B., Roger, A., Thébaud, C. & Chave, J. 2009. Identification of Amazonian trees with DNA barcodes. *PLOS ONE* 4: e7483. https://doi.org/10.1371/journal.pone.0007483

Huber, J. 1909. Materiaes para a Flora amazonica VII. Plantae Duckeanae austro-guyanenses. *Bol. Mus. Goeldi Hist. Nat. Ethnogr.* 5(2): 294–436.

Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A. & Janzen, D.H. 2005. Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. U.S.A.* 102: 8369–8374. https://doi.org/10.1073/pnas.0503123102

Lang, C., Costa, F.R.C., Camargo, J.L.C., Durgante, F.M. & Vicentini, A. 2015. Near Infrared Spectroscopy facilitates rapid identification of both young and mature Amazonian tree species. *PLOS ONE* 10: e0134521. https://doi.org/10.1371/journal.pone.0134521

Lang, C., Almeida, D.R. & Costa, F.R. 2017. Discrimination of taxonomic identity at species, genus and family levels using Fourier Transformed Near-Infrared Spectroscopy (FT-NIR). *Forest Ecol.*

Managem. 406: 219–227. https://doi.org/10.1016/j.foreco.2017.09.003

Leaché, A.D., Fujita, M.K., Minin, V.N. & Bouckaert, R.R. 2014. Species delimitation using genome-wide SNP data. *Syst. Biol.* 63: 534–542. https://doi.org/10.1093/sysbio/syu018

Lopes, R.H., Reid, I. & Hobson, P.R. 2009. The two-dimensional Kolmogorov-Smirnov test. In: XI International Workshop on Advanced Computing and Analysis Techniques in Physics Research, April 23–27 2007, Amsterdam, the Netherlands. *PoS (ACAT) 045*. https://doi.org/10.22323/1.050.0045

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S. & Hoekstra, H.E. 2012. Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLOS ONE* 7: e37135. https://doi.org/10.1371/journal.pone.0037135

Pimm, S.L. & Joppa, L.N. 2015. How many plant species are there, where are they, and at what rate are they going extinct? *Ann. Missouri Bot. Gard.* 100: 170–176. https://doi.org/10.3417/2012018

Prata, E.M.B., Sass, C., Rodrigues, D.P., Domingos, F.M.C.B., Specht, C.D., Damasco, G., Ribas, C.C., Fine, P.V.A. & Vicentini, A. 2018. Towards integrative taxonomy in Neotropical botany: Disentangling the *Pagamea guianensis* species complex (Rubiaceae). *Bot. J. Linn. Soc.* 188: 213–231. https://doi.org/10.1093/botlinnean/boy051

Schlick-Steiner, B.C., Steiner, F.M., Seifert, B., Stauffer, C., Christian, E. & Crozier, R.H. 2010. Integrative taxonomy: A multisource approach to exploring biodiversity. *Annual Rev. Entomol.* 55: 421–438. https://doi.org/10.1146/annurev-ento-112408-085432

Schwarz, G. 1978. Estimating the dimension of a model. *Ann. Statist.* 6: 461–464. https://doi.org/10.1214/aos/1176344136

Scrucca, L. & Raftery, A.E. 2014. clustvarsel: A package implementing variable selection for model-based clustering in R. *arXiv: 1411.0606*.

Scrucca, L., Fop, M., Murphy, T.B. & Raftery, A.E. 2016. mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *R Journal* 8: 289–317. https://journal.r-project.org/archive/2016/RJ-2016-021/

Siani, A., Nakamura, M.J., Tappin, M., Monteiro, S., Guimarães, A. & Ramos, M. 2012. Chemical composition of South American Burseraceae non-volatile oleoresins and preliminary solubility assessment of their commercial blend. *Phytochem. Analysis* 23: 529–539. https://doi.org/10.1002/pca.2351

Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Swart, J.J. 1942. A monograph of the genus *Protium* and some allied genera (Burseraceae). *Recueil Trav. Bot. Néerl.* 39: 211–446.

Ter Steege, H., Pitman, N.C., Sabatier, D., Baraloto, C., Salomão, R.P., Guevara, J.E., Phillips, O.L., Castilho, C.V., Magnusson, W.E., Molino, J., Monteaugo, A., Vargas, P.N., Montero, J.C., Feldpausch, T.R., Coronado, E.N.H., Kileen, T.J., Mostacedo, B., Vasquez, R., Assis, R.L., Terborgh, J., Wittmann, F., Andrade, A., Laurance, W.F., Laurance, S.G. W., Marimon, B.S., Marimon, B., Jr., Vieira, I.C.G., Amaral, I.L., Brienen, R., Castellanos, H., López, D.C., Duivenvoorden, J.F., Mongollón, H.F., Matos, F.D.A., Dávila, N., García-Villacorta, R., Diaz, P.R.S., Costa, F., Emilio, T., Levis, C., Schietti, J., Souza, P., Alonso, A., Dallmeier, F., Montoya, A.J.D., Piedade, M.T.F., Araujo-Murakami, A., Arroyo, L., Gribel, R., Fine, P.V.A., Peres, C.A., Toledo, M., Aymard, G.A., Baker, T.R., Cerón, C., Engel, J., Henkel, T.W., Maas, P., Petronelli, P., Stropp, J., Zartman, C.E., Daly, D., Neil, D., Silveira, M., Paredes, M.R., Chave, J., Filho, D.A.L., Jørgensen, P.M., Fuentes, A., Schöngart, J., Valverde, F.C., Di Fiore, A., Jimenez, E.M., Mora, M.C.P., Phillips, J.F., Rivas, G., Van Andel, T.R., von Hildebrand, P., Hoffman, B., Zent, E.L., Mahli, Y., Prieto, A., Rudas, A., Ruschell, A.R., Silva, N., Vos, V., Zent, S., Oliveira, A. A., Schutz, A.C., Gonzales, T., Nascimento, M.T., Ramirez-Angulo, H., Sierra, H., Tirado, M., Medina, M.N.U., Van der Heijden, G., Vela, C.I.A., Torre, E.V., Vriesendorp, C., Wang, O., Young, K.R., Baider, C., Baslev, H., Ferreira, C., Mesones, I., Torres-Lezana, A., Giraldo, L.E.U., Zagt, R., Alexiades, M.N., Hernandez, L., Huamantupa-Chuquimaco, I., Milliken, W., Cuenca, W.P., Pauletto, D., Sandoval, E.V., Gamarra, L.V., Dexter, K.G., Feeley, K., Lopez-Gonzales, G. & Silman, M.R. 2013. Hyperdominance in the Amazonian tree flora. *Science* 342: 1243092. https://doi.org/10.1126/science.1243092

Ulloa, C., Acevedo-Rodríguez, P., Beck, S., Belgrano, M.J., Bernal, R., Berry, P.E., Brako, L., Celis, M., Davidse, G., Forzza, R.C., Gradstein, S.R., Hokche, O., León, B., León-Yánez, S., Magill, R.E., Neill, D.A., Nee, M., Raven, P.H., Stimmel, H., Strong, M.T., Villaseñor, J.L., Zarucchi, J.L., Zuloaga, F.O. & Jørgensen, P.M. 2017. An integrated assessment of the vascular plant species of the Americas. *Science* 358: 1614–1617. https://doi.org/10.1126/science.aao0398

Appendix 1. List of specimens included in this study. An extended version is available in supplemental Table S1 for samples used in the phylogenetic analysis.

Taxon, country of origin, state, municipality, *collector(s)* and *collection number* (herbarium code). Collection numbers followed by "†" are samples included only in the phylogenetic analysis, collection numbers followed by "§" are samples included in both morphological and spectral analysis, and collection numbers followed by "⁑" are samples included in all analysis (phylogenetic, morphological, spectral). Samples used in the phylogenetic analysis († and ⁑) correspond to newly generated ddRAD-seq sequences. Full sequences and SNP files are available in supplemental Appendices S2 & S3 (NEXUS and PHYLIP formats). AM = Amazonas, BA = Bahia, DF = Distrito Federal, ES = Espírito Santo, GO = Goiás, LO = Loreto, MA = Maranhão, MG = Minas Gerais, MS = Mato Grosso do Sul, PA = Pará, PI = Piauí, RJ = Rio de Janeiro, RN = Rio Grande do Norte.

Protium brasiliense (Spreng.) Engl., Brazil, MG, Diamantina, *Damasco, G. 1647*† (NY), ***P. cordatum*** Huber, Brazil, PA, Faro, *Damasco, G. 1328*⁑ (NY), ***P. cordatum***, Brazil, PA, Faro, *Damasco, G. 1329*⁑ (NY), ***P. cordatum***, Brazil, PA, Faro, *Damasco, G. 1330*⁑ (NY), ***P. cordatum***, Brazil, PA, Faro, *Damasco, G. 1331*§ (NY), ***P. cordatum***, Brazil, PA, Faro, *Damasco, G. 1332*⁑ (NY), ***P. cordatum***, Brazil, PA, Faro, *Damasco, G. 1333*§ (NY), ***P. cordatum***, Brazil, PA, Faro, *Damasco, G. 1334*§ (NY), ***P. cordatum***, Brazil, PA, Faro, *Damasco, G. 1499*⁑ (NY), ***P. cordatum***, Brazil, PA, Faro, *Damasco, G. 1500*⁑ (NY), ***P. cordatum***, Brazil, PA, Faro, *Damasco, G. 1501*⁑ (NY), ***P. cordatum***, Brazil, PA, Faro, *Damasco, G. 1502*⁑ (NY), ***P. cordatum***, Brazil, PA, Faro, *Damasco, G. 1503*§ (NY), ***P. dawsonii*** Cuatrec., Brazil, GO, São João da Aliança, *Damasco, G. 1675*† (NY), ***P. heptaphyllum*** (Aubl.) Marchand, Brazil, BA, Mata de São João, *Damasco, G. 1588*⁑ (UC), ***P. heptaphyllum***, Brazil, BA, Itacaré, *Damasco, G.* 1394⁑ (UC), ***P. heptaphyllum***, Brazil, BA, Mata de São João, *Damasco, G. 1574*§ (NY), ***P. heptaphyllum***, Brazil, BA, Una, *Perdiz, R. 2845*§ (NY), ***P. heptaphyllum***, Brazil, DF, Planaltina, *Damasco, G. 1549*⁑ (UC), ***P. heptaphyllum***, Brazil, ES, Linhares, *Damasco, G. 1413*⁑ (UC), ***P. heptaphyllum***, Brazil, ES, Pontal do Ipiranga, *Damasco, G. 1432*⁑ (UC), ***P. heptaphyllum***, Brazil, MA, São Luiz, *Damasco, G. 1625*⁑ (NY), ***P. heptaphyllum***, Brazil, MT, Chapada Guimarães, *Damasco,*

Appendix 1. Continued.

G. 1604§ (NY), ***P. heptaphyllum***, Brazil, MT, Chapada Guimarães, *Damasco, G. 1605*‡ (NY), ***P. heptaphyllum***, Brazil, MT, Chapada Guimarães, *Damasco, G. 1614*§ (NY), ***P. heptaphyllum***, Brazil, MT, Chapada Guimarães, *Damasco, G. 1615*§ (NY), ***P. heptaphyllum***, Brazil, MT, Chapada Guimarães, *Damasco, G. 1608*§ (NY), ***P. heptaphyllum***, Brazil, MT, Chapada Guimarães, *Damasco, G. 1609*§ (NY), ***P. heptaphyllum***, Brazil, MT, Chapada Guimarães, *Damasco, G. 1610*§ (NY), ***P. heptaphyllum***, Brazil, MT, Nova Xavantina, *Damasco, G. 1646*‡ (UC), ***P. heptaphyllum***, Brazil, MS, Aquidauana, *Damasco, G. 1595*‡ (UC), ***P. heptaphyllum***, Brazil, RN, Pipa, *Damasco, G. 1633*‡ (UC), ***P. heptaphyllum***, Brazil, PA, Alter do Chão, *Damasco, G. 1515*‡ (NY), ***P. heptaphyllum***, Brazil, PA, Alter do Chão, *Damasco, G. 1521*§ (NY), ***P. heptaphyllum***, Brazil, PA, Alter do Chão, *Damasco, G. 1335*§ (NY), ***P. heptaphyllum***, Brazil, PA, Alter do Chão, *Damasco, G. 1336*§ (NY), ***P. heptaphyllum***, Brazil, PA, Alter do Chão, *Damasco, G. 1337*§ (NY), ***P. heptaphyllum***, Brazil, PA, Alter do Chão, *Damasco, G. 1338*§ (NY), ***P. heptaphyllum***, Brazil, PA, Faro, *Damasco, G. 1511*‡ (NY), ***P. heptaphyllum***, Brazil, PA, Faro, *Damasco, G. 1512*‡ (UC), ***P. heptaphyllum***, Brazil, PA, Faro, *Damasco, G. 1514*‡ (UC), ***P. heptaphyllum***, Brazil, PA, Salvaterra, *Damasco, G. 1527*‡ (NY), ***P. heptaphyllum***, Brazil, PA, Salvaterra, *Damasco, G. 1339*§ (NY), ***P. heptaphyllum***, Brazil, PA, Salvaterra, *Damasco, G. 1340*§ (NY), ***P. heptaphyllum***, Brazil, PA, Salvaterra, *Damasco, G. 1341*§ (NY), ***P. heptaphyllum***, Brazil, PA, Salvaterra, *Damasco, G. 1529*§ (NY), ***P. heptaphyllum***, Brazil, PA, Salvaterra, *Damasco, G. 1530*§ (NY), ***P. heptaphyllum***, Brazil, PI, Caracol, *Damasco, G. 1564*‡ (NY), ***P. heptaphyllum***, Brazil, PI, Caracol, *Damasco, G. 1567*§ (NY), ***P. heptaphyllum***, Brazil, PI, Caracol, *Damasco, G. 1571*§ (NY), ***P. heptaphyllum***, Brazil, PI, Caracol, *Damasco, G. 1572*§ (NY), ***P. heptaphyllum***, Brazil, PI, Caracol, *Damasco, G. 1573*§ (NY), ***P. heptaphyllum***, Brazil, PI, Formoso do Rio Preto, *Damasco, G. 1554*‡ (NY), ***P. heptaphyllum***, Brazil, PI, Formoso do Rio Preto, *Damasco, G. 1556*§ (NY), ***P. heptaphyllum***, Brazil, PI, Formoso do Rio Preto, *Damasco, G. 1559*§ (NY), ***P. heptaphyllum***, Brazil, PI, Formoso do Rio Preto, *Damasco, G. 1561*§ (NY), ***P. heptaphyllum***, Brazil, RJ, Carapebus, *Damasco, G. 1440*‡ (UC), ***P. heptaphyllum***, Brazil, RJ, Carapebus, *Damasco, G. 1441*§ (NY), ***P. heptaphyllum***, Brazil, RJ, Carapebus, *Damasco, G. 1442*§ (NY), ***P. heptaphyllum***, Brazil, RJ, Carapebus, *Damasco, G. 1445*§ (NY), ***P. heptaphyllum***, Brazil, TO, Palmas, *Damasco, G. 1358*‡ (UC), ***P. heptaphyllum***, Brazil, TO, Palmas, *Damasco, G. 1364*§ (NY), ***P. heptaphyllum***, Brazil, TO, Palmas, *Damasco, G. 1377*§ (NY), ***P. heptaphyllum*** subsp. ***ulei*** (Swart.) Daly, Brazil, AM, Manaus, *Damasco, G. 1459*‡ (UC), ***P. heptaphyllum*** subsp. ***ulei***, Brazil, AM, Presidente Figueiredo, *Damasco, G. 1485*§ (NY), ***P. heptaphyllum*** subsp. ***ulei***, Brazil, AM, Presidente Figueiredo, *Damasco, G. 1495*‡ (NY), ***P. heptaphyllum*** subsp. ***ulei***, Brazil, AM, Presidente Figueiredo, *Damasco, G. 1497*§ (NY), ***P. heptaphyllum*** subsp. ***ulei***, Brazil, AM, Barcelos, *Damasco, G. 1658*§ (NY), ***P. heptaphyllum*** subsp. ***ulei***, Peru, LO, Maynas, *Fine, P. 2067*‡ (NY), ***P. heptaphyllum*** subsp. ***ulei***, Peru, LO, Maynas, *Fine, P. 2068*‡ (NY), ***P. icicariba*** (DC.) Marchand, Brazil, ES, Linhares, *Damasco, G. 1642*† (UC), ***P. kleinii*** Cuatrec., Brazil, ES, Linhares, *Damasco, G. 1641*† (UC), ***P. krukoffii*** Swart, Peru, LO, Maynas, *Fine, P. 2071*† (UC), ***P. ovatum*** Engl., Brazil, DF, Brasília, *Damasco, G. 1638*† (UC), ***P. pilosum*** (Cuatrec.) Daly, Brazil, AM, Manaus, *Assunção, P. 713*† (INPA), ***P. trifoliolatum*** Engl., Peru, LO, Maynas, *Fine, P. 2079*† (UC), ***P. unifoliolatum*** Engl., Peru, LO, Maynas, *Fine, P. 2085*† (UC), ***P. widgrenii*** Engl., Brazil, ES, Linhares, *Damasco, G. 1643*† (UC).