A Phylogenomic Analysis of Genipa (Rubiaceae) Using Target Sequence Capture Data

Authors: Ridley, Rhonda, Persson, Claes, Oxelman, Bengt, Andermann, Tobias, and Bacon, Christine D.

Source: Systematic Botany, 49(3) : 617-625

Published By: The American Society of Plant Taxonomists

URL: https://doi.org/10.1600/036364424X17267811220489

Systematic Botany (2024), 49(3): pp. 617–625
ISSN 0363-6445 (print); ISSN 1548-2324 (online)

DOI 10.1600/036364424X17267811220489
Date of publication November 19, 2024

A Phylogenomic Analysis of *Genipa* (Rubiaceae) Using Target Sequence Capture Data

Rhonda Ridley,[1,2,3,5] Claes Persson,[1,2] Bengt Oxelman,[1,2] Tobias Andermann,[4] and Christine D. Bacon[1,2]

[1]Department of Biological and Environmental Sciences, University of Gothenburg, Göteborg, Sweden; claes.persson@bioenv.gu.se; bengt.oxelman@bioenv.gu.se; christine.bacon@bioenv.gu.se
[2]Gothenburg Global Biodiversity Centre, Göteborg, Sweden
[3]National Museums Northern Ireland
[4]Department of Organismal Biology, SciLifeLab, Uppsala University, Uppsala, Sweden; tobias.andermann@ebc.uu.se
[5]Author for correspondence (rhonda.ridley@nationalmuseumsni.org)

Communicating Editor: Shawn E. Krosnick

Abstract—The genus *Genipa* is a widespread, lowland, Neotropical lineage of trees in the coffee family, Rubiaceae. There is long-standing disagreement on the delimitation of species in the genus and how broadly *Genipa* is circumscribed. Here, we use genomic data to resolve the classification within *Genipa*. Using target sequence capture we generated a high-resolution 245-locus dataset to produce a comprehensive species phylogeny under the multi-species coalescent model. The phylogenomic results strongly support *Genipa spruceana*, often synonymised with *Genipa americana*, as a distinct monophyletic species. Similarly, the monophyly of *Genipa infundibuliformis*, a recently recognized species, is also strongly supported. The phylogeny also shows three distinct, well-supported clades within the widespread species, *Genipa americana*. These clades are interpreted as three independently evolving lineages in contrast to the two varieties most commonly recognized in *G. americana* based on previous morphological studies.

Keywords—Angiosperms 353, Bayesian inference, high-throughput sequencing, maximum likelihood, multi-species coalescent model, MSC, phylogenomics, STACEY.

Genipa L. is a widespread, common Neotropical genus in the coffee family Rubiaceae, tribe Gardenieae DC, consisting of small to large trees, 8–20 m in height, rarely up to 30 m. It is found in a variety of tropical and subtropical lowland (0–900 m) habitats (Pittier and Mell 1931; Steyermark 1972; Burger and Taylor 1993; Zappi et al. 1995). The genus is well known due to its economic and cultural significance.

The most well-known species in the genus is *Genipa americana* L. It has many uses, for example the fruit is eaten or made into beverages and it is used as a natural blue food colorant. It is important to several indigenous groups who extract an ink from the unripe fruit, which is used as body paint (Steyermark 1972). This practice has been commercialised and the ink is marketed as a henna alternative: so called jagua tattoos. It is also important for its medicinal uses and its timber. The tree is found in gardens, it is cultivated in and around Amazonian villages (Milliken et al. 1992), and it has been proposed as a potential shade tolerant tree crop by the United Nations Conference on Trade and Development (ProFound 2005). Despite its seeming ubiquity, many uses and importance, the taxonomy of this genus is not well resolved.

Previous classifications are based on morphological data and existing phylogenetic studies have been restricted to a few loci and only for one species, *G. americana* (Persson 2000, 2003; Andersson and Antonelli 2005; Rakotonasolo and Davis 2006; Kainulainen et al. 2013; Kainulainen and Bremer 2014; Mouly et al. 2014; Borges et al. 2021). *Genipa* has been through several taxonomic expansions and contractions over the years. According to the International Plant Names Index (IPNI 2022) 76 specific names exist in the genus plus a further five infraspecific names. Early circumscriptions of *Genipa*, for example by Baillon (1880) or Drake del Castillo (1897), encompassed species also occurring in Africa and Asia. Molecular studies demonstrated that this broad circumscription resulted in paraphyly and its size has since been gradually reduced to a solely Neotropical genus (Persson 2000, 2003; Rakotonasolo and Davis 2006). Previous *Genipa* species have been found to be congeneric with a diversity of Rubiaceae genera including: *Agouticarpa* C.H.Perss., *Aidia* Lour., *Alibertia* A.Rich. ex DC., *Benkara* Adans., *Bertiera* Aubl., *Burchellia* R. Br., *Casasia* A.Rich., *Catunaregam* Wolf, *Ceriscoides* (Hook.f.) Tirveng., *Duroia* L.f., *Gardenia* J.Ellis, *Glossostipula* Lorence, *Hyperacanthus* E.Mey. ex Bridson, *Randia* Houst. ex L., *Rosenbergiodendron* Fagerl., *Rothmannia* Thunb., *Sphinctanthus* Benth., and *Tocoyena* Aubl.

The most detailed taxonomic studies of the genus have been undertaken by Steyermark in the 1970s (Steyermark 1972, 1974) and more recently by Zappi et al. (1995). Few treatments of the genus exist outside of the floras of Central and South American countries (Dwyer 1980; Burger and Taylor 1993; Mendoza et al. 2004; Steyermark and Persson 2004; Delprete and Cortes 2005; Bernal et al. 2019; Gomes 2020). These treatments recognise a different number of species and infraspecific taxa without consensus, summarised in Table 1. Indumentum in particular is considered as an important diagnostic character in *Genipa*. The different taxonomic hypotheses center on the treatment of *G. americana* as a single highly phenotypically variable species (Burger and Taylor 1993; Zappi et al. 1995) versus treating its phenotypic variation to be of taxonomic merit (Steyermark and Persson 2004). Steyermark and Persson (2004) recognise two varieties in *G. americana* and provide a key detailing the morphological differences between the two varieties. *G. americana* var. *americana* is distinguished by its glabrous (or nearly so) lower leaf blade and upper leaf surface not rugulose, in contrast with *G. americana* var. *caruto* which has a densely soft-pubescent lower leaf blade surface and upper leaf surface sometimes rugulose.

Genipa spruceana Steyerm. was first described by Steyermark (1972) in The Botany of Guayana Highlands. *Genipa spruceana* is recognized in a number of treatments (Mendoza et al. 2004; Steyermark and Persson 2004; Bernal et al. 2019) but it is not universally recognized, for example Zappi et al. (1995) states that indumentum, the main distinguishing character, are quite variable and perhaps linked to environmental conditions. *Genipa caruto* Kunth, commonly known as the hairy genip, was first described by Kunth (1820), and

TABLE 1. Summary of *Genipa* taxa recognized in different works. N/A denotes that it is outside the known distribution of the species.

Taxonomic publication	*G. americana*	*G. americana var. caruto*	*G. caruto*	*G. infundibuliformis*	*G. spruceana*
Bolivia	✓	✓		N/A	
French Guiana				N/A	✓
Costaricensis	✓			N/A	Na
Flora Guatemala			✓	N/A	Na
Flora Panama	✓			N/A	Na
Guayana Highlands	✓	✓		N/A	✓
Mato Grosso	✓			N/A	✓
Plants of Colombia	✓			N/A	✓
Rubiaceae Colombia	✓	✓		N/A	✓
Venezuelan Guayana	✓	✓		N/A	✓
Zappi	✓			✓	
Data Aggregator or Database					
Name	*G. americana*	*G. americana var. caruto*	*G. caruto*	*G. infundibuliformis*	*G. spruceana*
Flora do Brasil	✓			✓	
LCVP	✓			✓	✓
PoWO	✓			✓	✓
Tropicos	✓			✓	
WFO	✓			✓	✓

Taxonomic Publication and Data Aggregator or Database full title or reference

Bolivia: Guia de Arboles de Bolivia (Killeen et al. 1993). French Guiana: Guide to the Vascular Plants of Central French Guiana (Mori 1997). Costaricensis: Flora Costaricensis (Burger and Taylor 1993). Flora Guatemala: Flora of Guatemala (Standley and Steyermark 1949). Flora Panama: Flora of Panama (Woodson et al. 1980). Guayana Highlands: The Botany of the Guayana Highlands (Steyermark 1972). Mato Grosso: A Synopsis of the Rubiaceae of the States of Mato Grosso and Mato Grosso do Sul (Delprete and Cortes 2005). Plants and Lichens of Colombia: Catalogue of the Plants and Lichens of Colombia (Bernal et al. 2019). Rubiaceae Colombia: Rubiaceae de Colombia. Guía Ilustrada de Generos (Mendoza et al. 2004). Venezuelan Guayana: Flora of the Venezuelan Guayana (Steyermark 1988). Zappi: Zappi et al. (1995). Flora do Brasil: Flora e Funga do Brasil (2020). LCVP: Leipzig Catalogue of Vascular Plants (Freiberg et al. 2020). PoWO: Plants of the World Online, Royal Botanic Gardens, Kew. Tropicos: Tropicos.org, Missouri Botanical Garden (Tropicos 2023). WFO: World Flora Online (WFO 2023).

since the publication of Flora Brasiliensis (1889) it is often demoted to *G. americana* var. *caruto* (Kunth) K. Schum. or not recognized at all (Zappi et al. 1995). The most recently described species in the genus is *G. infundibuliformis* Zappi and Semir. This species has a more restricted distribution than other members of *Genipa*, having only been recorded from the Atlantic Forest of Brazil. It is easily distinguished by its distinct flower and leaf morphology.

The global botanical databases and taxonomic data aggregators reflect various taxonomic hypotheses. Kew's The World Checklist of Vascular Plants (2023) lists three accepted species: *G. americana*, *G. infundibuliformis*, and *G. spruceana*. The Missouri Botanical Garden database, Tropicos (2023) lists three species: *G. americana*, *G. chapelieri* (A. Rich.) Drake and *G. infundibuliformis*. In the Tropicos database *G. spruceana* is treated as a synonym of *G. americana*. The Leipzig Catalogue of Vascular Plants (Freiberg et al. 2020) and World Flora Online (WFO 2023) list four species in *Genipa: G. americana*, *G. infundibuliformis*, *G. spruceana,* and *G. chapelieri*. *Genipa chapelieri* is a Madagascan species (Bridson and Robbrecht 1985) synonymous with *G. talangninia* (DC.) Drake, recently moved to *Hyperacanthus talangninia* (DC.) Rakotonas. and A.P. Davis in the *Aidia* clade (sensu Mouly et al. 2014) and therefore excluded from this study. The Flora e Funga do Brasil (2020) follows the classification of Zappi et al. (1995) and treats *G. spruceana* as conspecific with *G. americana*. The entry in the Checklist of the Plants of the Guiana Shield (Funk et al. 2007) is "*G. spruceana* = *G. americana*?", indicating that it is a species of unknown certainty.

The tendency of adopting broad taxon concepts or lumping extends to the infraspecific taxa in *Genipa*. Five infraspecific names are listed in IPNI (2022) *G. americana* var. *caruto*, *G. americana* f. *grandifolia* Chodat and Hassl., *G. americana* f. *jorgensenii* Steyerm., *G. americana* f. *parvifolia* Chodat and Hassl., and *G. americana* var. *riobranquensis* Kuhlm. Most of the botanical works listed in Table 1 do not recognize these infraspecific taxa (Dwyer 1980; Burger and Taylor 1993; Zappi et al. 1995; Gomes 2020).

Genipa is widely distributed from Mexico and the Caribbean to Argentina (Fig. 1). The distribution shown in Fig. 1 is based on records from GBIF (GBIF 2020), cleaned using the package CoordinateCleaner (Zizka et al. 2019) in R (R Core Team 2020). Given the variation in the taxonomic treatments of *Genipa* (Table 1), the distribution of species (Fig. 1) reflects an approximate distribution based on GBIF data (GBIF 2020), as we do not know how the determination of each record was reached. Given the known differences in taxonomic classification in the genus it is likely that *G. spruceana*, *G. americana* var. *caruto,* and *G. infundibuliformis* are under-recorded and have been recorded as *G. americana*. Considering the economic and cultural significance of the genus, the distribution shown in Fig. 1 may result from human cultivation; this remains to be tested.

Many habitats where *Genipa* grows are undergoing drastically increased rates of deforestation and land conversion to agriculture (Hansen et al. 2013). This is especially critical in the Atlantic Forest, where *G. infundibuliformis* is distributed, where around 85% of the original area has been deforested (Ribeiro et al. 2009). While the IUCN threat status has not been assessed for *Genipa* in this work, some species are considered to be endangered (*G. americana*) and vulnerable (*G. spruceana*) (Ter Steege et al. 2015).

Here, we infer the phylogeny of *Genipa* using phylogenomic data. Target sequence capture is a genome reduction approach which allows researchers to select and specifically amplify a set of target loci across the genome, using Illumina sequencing (Andermann et al. 2020). This approach has been readily adopted for evolutionary studies as it balances cost, data scale, and computational requirements (Jones and Good 2016; Hale et al. 2020). It is suitable for DNA of limited quality that is more fragmented, such as herbarium specimens or degraded silica dried plant material (Brewer et al. 2019). A major benefit of target sequence capture is the existence of pre-designed bait kits that target known regions of the genome. One such kit is the Angiosperms 353 bait kit which

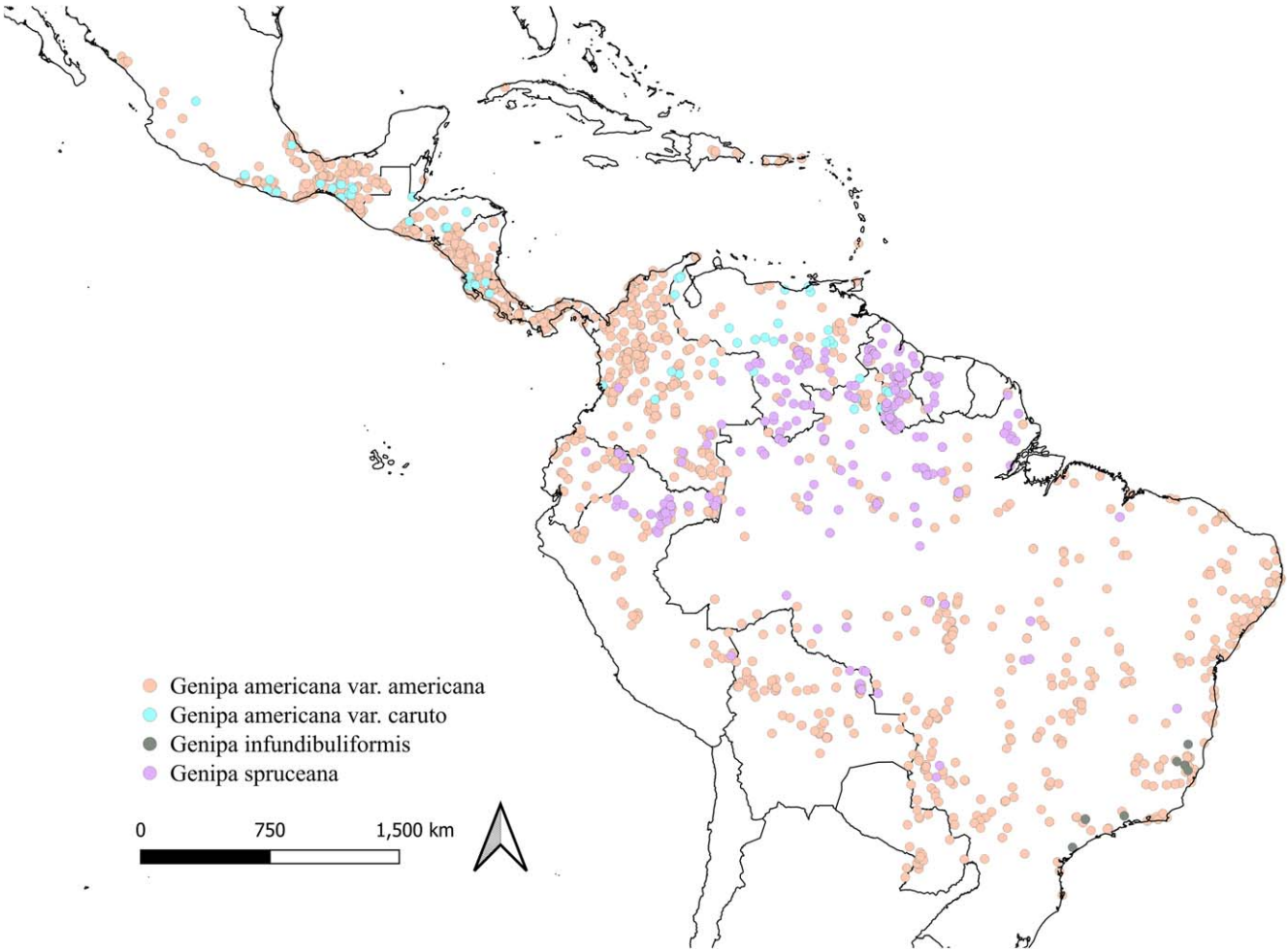

FIG. 1. Putative distribution of *Genipa* based on cleaned GBIF occurrence records.

targets 353 single-copy protein-coding genes and works across all angiosperms (Johnson et al. 2019).

We use two phylogenetic inference methods to identify independently evolving lineages from genomic data i) a heuristic two-step approach where gene trees are created first independently and then combined to infer a species tree, and ii) using Bayesian inference where gene trees and the species tree are co-estimated. Both methods implement the multispecies coalescent model (Rannala and Yang 2003; Degnan and Rosenberg 2009; Liu et al. 2009) for phylogeny construction. The model applies probabilistic theory to explain the evolution of alleles and accounts for the incongruence between gene trees and species trees because of incomplete lineage sorting.

MATERIALS AND METHODS

Taxon Sampling—Twenty-eight *Genipa* samples (Appendix 1) representing all four putative taxa in the genus were sampled. The taxonomic concept of *Genipa* in this study follows Steyermark and Persson (2004): *G. americana* is recognised with two varieties; *G. spruceana* is treated as a separate species and we follow (Zappi et al. 1995) in the recognition of the species *G. infundibuliformis*. The samples represent interspecific and infraspecific variation in the genus and comprise of: 11 individuals of *G. americana* var. *americana*, eight of *G. americana* var. *caruto*, two of *G. infundibuliformis*, and seven of *G. spruceana*. *Tocoyena pittieri* (Standl.) Standl, also in Rubiaceae, was included as the outgroup. All specimens in this study were collected legally and the permits can be presented on request.

Methodology—Total genomic DNA was extracted from silica dried plant material using the NucleoSpin Plant II Kit (Macherey-Nagel, Düren, Germany) or DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). The protocol followed manufacturer's instructions apart from the cell lysis time, which was increased to overnight to maximise DNA yield. DNA quality was assessed using a NanoDrop 2000 spectrophotometer and quantified using the Qubit 2.0. The NanoDrop 2000 and Qubit 2.0 results were used to determine samples that needed concentration by vacuum centrifugation. Gel electrophoresis was also carried out to assess DNA fragment size. Multiple extraction rounds were pooled as necessary when initial DNA quantity was low, in order to meet the minimum concentration requirements of Rapid Genomics, Florida, USA who performed target capture library preparation and sequencing. The minimum DNA sample concentration was 8.52 ng/uL. The DNA was mechanically sheared to a size of 200–500 base pairs (bp). Illumina libraries were constructed and barcode adapters for the Illumina Sequencing platform were ligated to the libraries then PCR-amplified using standard cycling protocols. Samples were pooled into 16 barcoded libraries with equimolar amounts to a total of 500 ng for hybridization. Target enrichment was performed using the Angiosperms 353 bait set (Johnson et al. 2019) targeting 353 putatively orthologous genes. After enrichment, samples were re-amplified for an additional 6–12 PCR cycles and sequenced using an Illumina NovaSeq 6000 with paired-end 250 bp reads.

The Illumina raw read data was processed using the bioinformatic pipeline SECAPR 2.2.5 (Andermann et al. 2018). The bioinformatic pipeline was run on the Sigma2 High-Performance Computing cluster at NTNU, Norway. Raw sequence data was quality checked using FastQC (Andrews 2010) and MultiQC (Ewels et al. 2016) to gain an overview of sequence quality and determine cleaning parameters. Illumina adapters were removed and cleaning of sequences was carried out using FastP 0.23 (Chen et al. 2018). FastP default settings implemented in SECAPR were: i) the read was cut if the accuracy between adapter and read Phred quality

score was below 20; ii) the maximum percent of low-quality nucleotides allowed was set to 40 and reads with a higher percentage of unqualified (low quality) nucleotides were discarded; iii) size of sliding window for quality trimming was set to 5 nucleotides; iv) trimming from front and tail if quality value was lower than 10; v) reads below complexity threshold of 10 removed; vi) trim poly repeats at end of read of length 7; vii) low complexity filtering was enabled; and viii) length filtering was disabled. Quality of cleaned reads was checked, using FastQC, MultiQC, and the plotting function in SECAPR.

De novo contig assembly was performed on cleaned reads using Spades 3.15.2 (Bankevich et al. 2012). Overlapping sequences were combined into contig sequences using kmer values 21, 33, 55, 77, 99, and 127. The minimum contig length was set to 200 and contigs under this threshold were discarded. Contigs belonging to target loci were identified by using Blastn (Camacho et al. 2009) to match the contig sequences with a set of reference sequences for each locus. The reference sequences used were the *Gardenia philastrei* Pierre ex Pit. Davis, A.P. 4055 (K) sequences from the Royal Botanic Gardens Kew PAFTOL project (Baker et al. 2022). A custom target file from *Gardenia philastrei*, a species close to *Genipa,* was used as this method has been shown to maximise gene recovery (McLay et al. 2021). A sequence-match was identified if the sequence matched with at least 80% identity across at least 80% of the contig length. Loci with multiple contig matches were discarded as they may represent paralogous sequences. A multiple species alignment (MSA) was created from the contig data using MAFFT 7.490 (Katoh et al. 2019) for each locus that was recovered across at least three samples with the addition of the "no trim" parameter to keep full contig sequence length. Next we repeated the read assembly using the consensus sequence of each locus' MSA as a genus-specific reference library. This additional reference assembly leads in general to a more efficient and less biased retrieval of DNA reads across all samples for each locus (Andermann et al. 2018), as opposed to using the recovered contig sequences for each sample. The minimum coverage parameter was set at four reads. Consensus sequences were generated from the reads mapping to the genus-specific reference at each locus for each sample and from these consensus sequences multiple sequence alignments were computed for each locus using MAFFT 7.490 (Katoh et al. 2019).

Phylogenetic Analysis—Two different phylogenetic methods were used. The first method employed was ASTRAL-III (Zhang et al. 2018), which produces a species tree that shares the maximum number of quartet topologies with the input gene trees and the lengths of the internal branches are inversely proportional to the number of quartets concordant to the split. Gene trees were created using IQ-TREE 2 (Minh et al. 2020). A set of 245 bootstrap consensus maximum likelihood gene trees were created using 1000 bootstrap replicates with UFBoot2 (Hoang et al. 2018) and automatic substitution model selection with ModelFinder (Kalyaanamoorthy et al. 2017) implemented in the IQ-TREE 2 software package. The tree was visualised using Figtree v.1.4.3 (Rambaut 2017).

The second species phylogeny was produced using Bayesian inference, created with Species Tree And Classification Estimation, Yarely (STACEY; Jones 2017) in BEAST2 (Bouckaert et al. 2019) on the CIPRES Science Gateway web portal (Miller et al. 2012). This method simultaneously estimates gene trees and species trees using a birth-death collapse model. The input data was a subset of six loci from the de novo contig assembly dataset. The subset selection consisted of the first six loci in the de novo assembly dataset (5, 9, 20, 43, 55, and 62), with the exception of locus 59, it was excluded from the analysis as it only had seven out of 29 samples. The xml input was generated in BEAUTi 2.6 (Bouckaert et al. 2019). The samples were not preassigned to species and no partitions were selected. The following parameters and priors were selected: species tree model collapse height: $1e^{-5}$; strict clock model: each locus was set as relative to each other; JC69 substitution model; bdcGrowthRate: lognormal (M = 5, S = 2); collapseWeight: beta (alpha = 2, beta = 2); population prior log normal (M = -7, S = 2); relativeDeathRate: beta (alpha = 1, beta = 1). The MCMC was run for 100 million generations and Tracer version v1.7.1 (Rambaut et al. 2018) was used to explore convergence of parameters and effective sample size (ESS). The species tree was generated using TreeAnnotator 2.6.3 (Drummond and Rambaut 2007), after discarding 10% as burn-in, and then visualised using Figtree v. 1.4.3 (Rambaut 2017).

All sequence data generated for this study are available at the GenBank Sequence Read Archive (SRA) under BioProject ID PRJNA1029819 and individual sample accession numbers can be found in Appendix 1. The assemblies and individual gene trees generated in this work are deposited in Dryad (Ridley et al. 2024). The scripts used to create the data in this paper are available on Github at https://github.com/AntonelliLab/seqcap_processor.

Geographic Distribution—An analysis of the geographic distance between samples was undertaken to determine to what degree geographic isolation assists in the interpretation of our phylogenies. The distance between samples was calculated using the open source GIS software QGIS. Those samples found within 50 km of each other and 51–100 km of each other were noted.

Results

Phylogenomic Analyses—The mean number of raw reads for the samples was 1,126,098, the maximum was 2,183,270 and the minimum 535,602. After cleaning, the average number of raw reads per sample was 1,108,523. The maximum percentage reduction after cleaning was a reduction of 4.48% and the minimum 0.57%. The mean number of de novo contigs that could be identified as being part of the set of target loci was 198 of which 28 loci had matching de novo contig sequences in all samples. An additional step of remapping the reads to a *Genipa* specific reference library created from the MSAs of the recovered contigs led to the recovery of more loci for more samples. In this approach, sample-specific sequences for 245 loci were recovered on average, 240 of which contained all 29 samples (28 *Genipa* and one *Tocoyena pittieri* outgroup). The following alignment summary statistics were calculated using the AMAS tool (Borowiec 2016), mean alignment length was 1508 bp, the maximum was 5674 and the minimum 190, the missing data per alignment mean was 21.55% (min = 2%, max = 96%), the mean proportion of variable sites was 0.12 (min = 0, max = 0.58) and the mean number of parsimony informative sites was 81 (min = 0, max = 764).

The ASTRAL-III phylogeny for the MSA from the reference-based phylogeny containing all 245 loci is shown in Fig. 2. It identified *G. infundibuliformis* and *G. spruceana* as fully-supported clades with a local posterior probability (LPP) of one. Within *G. americana* there are three subclades that are fully supported with a LPP of one: clade A, which contains eight samples: three from Colombia, four from Ecuador, and one from Panama; clade B is comprised of three Bolivian, one Peruvian, and one Colombian sample; and clade C is comprised of six *G. americana* var. *caruto* samples: one from Costa Rica, two from Panama, two from Guyana, and one from French Guiana. One *G. americana* var. *caruto* from Ecuador is found in clade A and one *G. americana* var. *caruto* from Bolivia is found in clade B. The clades within *G. americana* received full support.

The ESS for all parameters in the STACEY anlaysis was > 241. The STACEY phylogeny, shown in Fig. 3, supports *G. infundibuliformis* and *G. spruceana* as monophyletic clades. Three clades are present within *G. americana* A, B, and C. However, one Peruvian sample (G_am6) was placed within *G. americana* clade A, whereas in the ASTRAL-III tree it is in clade B. *Genipa americana* clades B and C received maximum posterior probability scores in STACEY and 0.98 for clade A. The node bars shown on the tree are the height posterior density which represents the 95% central posterior distribution of species tree split times.

Geographic Distribution—The sample distribution shows that several taxa grow in geographical sympatry. Samples of different taxa that were collected within 50 km of each other are: *G. americana* var. *caruto* G_car27 and *G. spruceana* G_spru18; *G. spruceana* G_spru11 and *G. americana* G_am9 clade A; *G. spruceana* G_spru11 and *G. americana* G_am14 clade A. In addition the sample pairs: *G. americana* G_am30

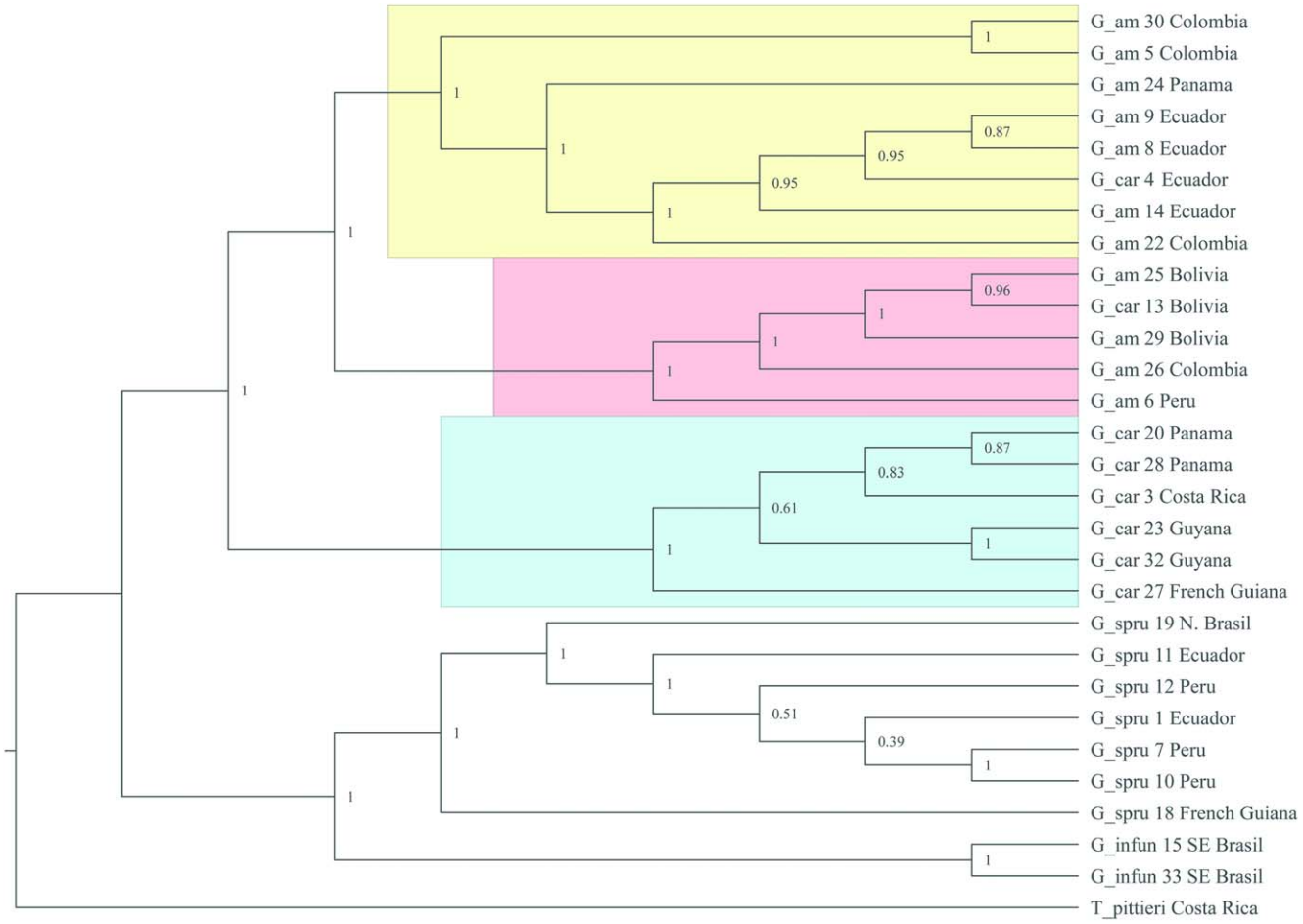

FIG. 2. Cladogram produced using ASTRAL-III, of 28 *Genipa* samples, based on 245 nuclear loci, with ASTRAL local posterior branch support shown. Color is used to highlight the three separate clades in *G. americana*: clade A yellow, clade B pink, and clade C blue. The tip labels show species abbreviation, sample number, and country. Species abbreviations: G_am = *G. americana* var. *americana*; G_car = *G. americana* var. *caruto*; G_spru = *G. spruceana*; G_infun = *G. infundibuliformis*.

clade A and *G. americana* G_am26 clade B; *G. americana* G_am8 clade A and *G. spruceana* G_spru1; *G. americana* G_am24 clade A and *G. americana* var. *caruto* G_car28 were collected within 100 km of each other.

DISCUSSION

We produced two well-resolved phylogenies from Angiosperms 353 target capture data using two methods, both based on the multi-species coalescent model that are consistent in topology. The data support the monophyly of *G. americana*, *G. infundibuliformis*, and *G. spruceana*. This study identifies genomic support for the recognition of *G. infundibuliformis* and *G. spruceana* as sister species separate from *G. americana*. We have considered the effect that sampling regimes can have on species delimitation. In this study several sample locations show sufficient distributional overlap to suggest that phylogenetic structure is not solely the result of the sample locations or geographic distance. Our interspecific sampling is especially high in the north of South America where *G. americana* s.l. and *G. spruceana* are sympatric. These samples could potentially interbreed in these contact zones but they form independently evolving lineages even when there is sympatry.

Genipa americana, a widespread species distributed from southern Mexico and the Caribbean to northern Argentina, is divided into three well-supported clades. This pattern of separate lineages within one species is common to many species and it is likely to be an understudied but frequent occurrence in the Neotropics (Antonelli et al. 2018; Finch et al. 2022). The phylogenies show that most *G. americana* var. *caruto* samples are found in a single clade however two *G. americana* var. *caruto* samples fall outside this clade. Samples determined as *Genipa americana* var. *americana* are found in clades A and B, and samples determined as *G. americana* var. *caruto* are found in clades B and C. This indicates that the current morphological infraspecific classification dividing *G. americana* into two varieties (Steyermark and Persson 2004) is not supported by the current study. The geographic analysis of sample locations did not find that the distribution distinguished any of the three *G. americana* clades.

In an attempt to increase taxonomic stability and not add to the already lengthy list of synonyms in this genus, no taxonomic changes are recommended in the genus until diagnostic evidence other than genomic data, such as morphological differentiation, is acquired. Current phylogenomic species delimitation methods do not readily distinguish between population structure and species (Carstens et al. 2013; Sukumaran and Knowles 2017). This can result in taxonomic

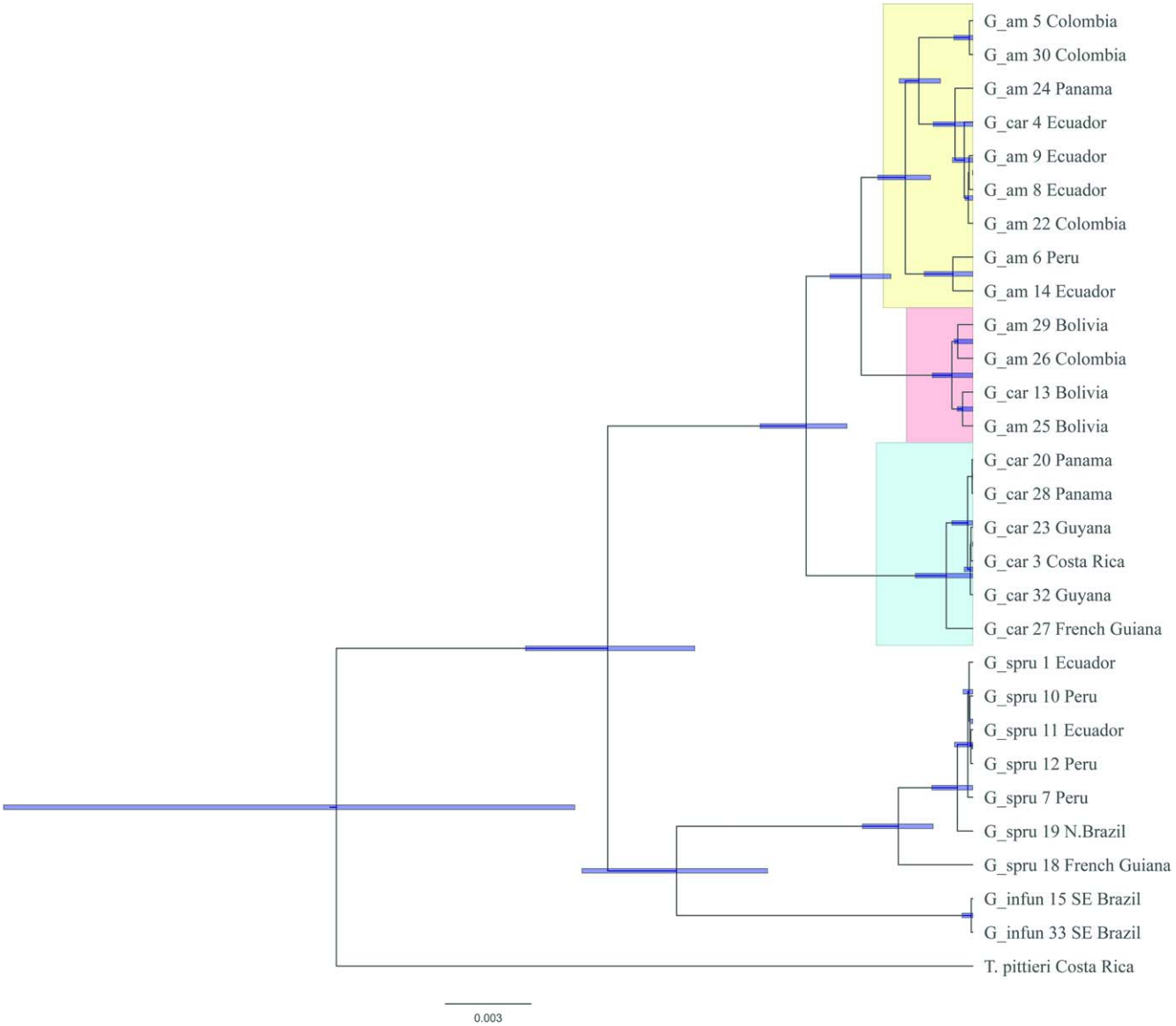

FIG. 3. Phylogeny from STACEY (BEAST2 plugin) analysis of six locus dataset, the units of branch length are the number of nucleotide substitutions per site, and node bars show 95% height posterior density. Color is used to highlight the three separate clades in *G. americana*: clade A yellow, clade B pink, and clade C blue. The tip labels show species abbreviation, sample number, and country. Species abbreviations: G_am = *G. americana* var. *americana*; G_car = *G. americana* var. *caruto*; G_spru = *G. spruceana*; G_infun = *G. infundibuliformis*.

inflation whereby previously identified infraspecific taxa or new clades are erroneously recognized as new species (Isaac et al. 2004; Sukumaran and Knowles 2017). The use of genomic data as the only means to delimit angiosperm species is not desirable and should only be considered for truly cryptic taxa, as defined by Struck et al. (2018). Here, further evidence is required to determine if the clades in *G. americana* warrant species status or if the infraspecific rank variety is more appropriate or if they should be recognized at all.

Genipa infundibuliformis is fully supported as a separate species in both our phylogenomic analyses. It can be readily determined by its morphology, namely the long corolla tube, reflexed petal lobes, lobed juvenile leaves, and spherical fruit with a smooth surface, which are all distinct characters only found in this species of *Genipa*. It has a restricted distribution in southeastern Brazil, however, it is not geographically isolated as *G. americana* s.l. is also present in the Brazilian Atlantic Forest.

Currently *G. spruceana* is not universally recognized; for example the recent Flora e Funga do Brasil (2020) treats *G. spruceana* as a synonym of *G. americana*. This is an example where species circumscription can have considerable consequences for conservation. The taxonomist's decision of broad versus narrow species circumscription can impede efforts to halt biodiversity loss (for example May 1990; Mace 2004; Garnett and Christidis 2017). In this case lumping *G. spruceana* in *G. americana* falsely inflates the abundance and possibly the distribution of *G. americana*, while *G. spruceana* goes unrecorded, which likely has considerable conservation implications for both species (Bickford et al. 2007; Adams et al. 2014).

In order to ensure that species in the genus are diagnostic beyond the genomic level, more field studies are required, particularly in Brazil, Guatemala, and Paraguay as it would provide further morphological and ecological data. The results of this phylogeny can be combined with additional

lines of evidence such as morphology in an integrated approach. This may elucidate diagnostic features for the three clades in *G. americana*. Once the above avenues are investigated a decision can be made on the taxonomic rank applicable (if any) to the clades within *G. americana.* By applying the multi-species coalescence model to detect independently evolving lineages in *Genipa*, we show support for three species and evidence of infraspecific genomic structure within *G. americana* s.l. A stable systematic framework for *Genipa* based on an integrative taxonomy approach is important for conservation of species in areas undergoing unprecedented rates of habitat modification, putting species and its interspecific variation at risk of extinction.

Acknowledgments

We thank Adrian Hill, Weston Testo, and Maria Fernández Torres for assistance and support on this work in the laboratory and with the bioinformatic pipeline. Thank you to Rodrigo Borges for donation of genomic material and Fiona McCrory for assistance in the production of figures. We also thank the staff and contributors of the herbaria (GB, NY) that provided material for study. Finally, we thank the journal editors and reviewers for valuable comments on the manuscript. This research was supported by funding to CDB from Swedish Research Council (2017-04980) and the Biodiversity and Ecosystem Services in a Changing Climate Strategic Research Area at the University of Gothenburg, Sweden, and TA received financial support from the SciLifeLab and Wallenberg Data Driven Life Science Program (grant: KAW 2020.0239).

Author Contributions

RR collected the data, performed the analysis and wrote the manuscript. CP designed the study, provided comments on the manuscript, and provided the sample material. BO designed the study, performed phylogenomic analysis, and provided comments on the manuscript. TA provided support with the bioinformatic analyses and provided comments on the manuscript. CDB designed the study, provided comments on the manuscript, and provided funding.

Literature Cited

Adams, M., T. A. Raadik, C. P. Burridge, and A. Georges. 2014. Global biodiversity assessment and hyper-cryptic species complexes: More than one species of elephant in the room? *Systematic Biology* 63: 518–533.

Andermann, T., Á. Cano, A. Zizka, C. Bacon, and A. Antonelli. 2018. SECAPR—A bioinformatics pipeline for the rapid and user-friendly processing of targeted enriched Illumina sequences, from raw reads to alignments. *PeerJ* 6: e5175.

Andermann, T., M. F. Torres Jiménez, P. Matos-Maraví, R. Batista, J. L. Blanco-Pastor, A. L. S. Gustafsson, L. Kistler, et al. 2020. A guide to carrying out a phylogenomic target sequence capture project. *Frontiers in Genetics* 10: 1407.

Andersson, L. and A. Antonelli. 2005. Phylogeny of the tribe Cinchoneae (Rubiaceae), its position in Cinchonoideae, and description of a new genus, *Ciliosemina. Taxon* 54: 17–28.

Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc (last accessed 19 April 2022).

Antonelli, A., M. Ariza, J. Albert, T. Andermann, J. Azevedo, C. Bacon, S. Faurby, et al. 2018. Conceptual and empirical advances in Neotropical biodiversity research. *PeerJ* 6: e5644.

Baillon, H. 1880. Rubiaceae. *Historie des Plantes* 7: 257–501. Paris: Librairie Hachette.

Baker, W. J., P. Bailey, V. Barber, A. Barker, S. Bellot, D. Bishop, L. R. Botigué, et al. 2022. A comprehensive phylogenomic platform for exploring the angiosperm tree of life. *Systematic Biology* 71: 301–319.

Bankevich, A., S. Nurk, D. Antipov, A. Gurevich, M. Dvorkin, A. Kulikov, V. Lesin, et al. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19: 455–477.

Bernal, R., S. R. Gradstein, and M. Celis. 2019. Catalog of plants and lichens of Colombia. http://catalogoplantasdecolombia.unal.edu.co.

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

Borges, R. L., S. G. Razafimandimbison, N. Roque, and C. Rydin. 2021. Phylogeny of the Neotropical element of the Randia clade (Gardenieae, Rubiaceae, Gentianales). *Plant Ecology and Evolution* 154: 458–469.

Borowiec, M. L. 2016. AMAS: A fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4: e1660.

Bouckaert, R., T. G. Vaughan, J. Barido-Sottani, S. Duchêne, M. Fourment, A. Gavryushkina, J. Heled, et al. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 15: e1006650.

Brewer, G. E., J. J. Clarkson, O. Maurin, A. R. Zuntini, V. Barber, S. Bellot, N. Biggs, et al. 2019. Factors affecting targeted sequencing of 353 nuclear genes from herbarium specimens spanning the diversity of angiosperms. *Frontiers in Plant Science* 10: 1102.

Bridson, D. and E. Robbrecht. 1985. Validation of the African genus *Hyperacanthus* E. Mey. (Rubiaceae Tribe Gardenieae). *Kew Bulletin* 40: 273.

Burger, W. and C. W. Taylor. 1993. Flora Costaricensis Family #202 Rubiaceae. *Fieldiana* 33.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden. 2009. BLAST+: Architecture and applications. *BMC Bioinformatics* 10: 421.

Carstens, B. C., T. A. Pelletier, N. M. Reid, and J. D. Satler. 2013. How to fail at species delimitation. *Molecular Ecology* 22: 4369–4383.

Chen, S., Y. Zhou, Y. Chen, and J. Gu. 2018. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34: i884–i890.

Degnan, J. H. and N. A. Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* 24: 332–340.

Delprete, P. G. and R. Cortes. 2005. A synopsis of the Rubiaceae of the states of Mato Grosso and Mato Grosso do Sul, central-western Brazil, with a key to genera, and a preliminary species list. *Revista de Biologia Neotropical* 1: 4–10.

Drake del Castillo, M. E. 1897. Plantes nouvelles de Madagascar. *Bulletin Mensuel de la Societe Linneenne de Paris*: 2.

Drummond, A. J. and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7: 214.

Dwyer, J. D. 1980. Flora of Panama, part IX. Family 179. Rubiaceae, part 1. *Annals of the Missouri Botanical Garden* 67: 1–256.

Ewels, P., M. Magnusson, S. Lundin, and M. Käller. 2016. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32: 3047–3048.

Finch, K. N., F. A. Jones, and R. C. Cronn. 2022. Cryptic species diversity in a widespread neotropical tree genus: The case of *Cedrela odorata. American Journal of Botany* 109: 1622–1640.

Flora e Funga do Brasil. 2020. Jardim Botânico do Rio de Janeiro. http://floradobrasil.jbrj.gov.br/ (last accessed 23 October 2023).

Freiberg, M., M. Winter, A. Gentile, A. Zizka, A. N. Muellner-Riehl, A. Weigelt, and C. Wirth. 2020. LCVP, The Leipzig catalogue of vascular plants, a new taxonomic reference list for all known vascular plants. *Scientific Data* 7: 416, doi: 10.1038/s41597-020-00702-z.

Funk, V., T. Hollowell, P. Berry, C. Kelloff, and S. N. Alexander. 2007. Checklist of the Plants of the Guiana Shield (Venezuela: Amazonas, Bolivar, Delta Amacuro; Guyana, Surinam, French Guiana). *Contributions from the United States National Herbarium* 55: 1–584.

Garnett, S. T. and L. Christidis. 2017. Taxonomy anarchy hampers conservation. *Nature* 546: 25–27.

GBIF. 2020. GBIF.org. GBIF Occurrence Download https://doi.org/10.15468/dl.ueyckq (last accessed 22 December 2020).

Gomes, M. 2020. *Genipa* in Flora do Brasil. Jardim Botânico do Rio de Janeiro. http://floradobrasil.jbrj.gov.br/reflora/floradobrasil/FB14044 (last accessed 23 October 2023).

Hale, H., E. M. Gardner, J. Viruel, L. Pokorny, and M. G. Johnson. 2020. Strategies for reducing per-sample costs in target capture sequencing for phylogenomics and population genomics in plants. *Applications in Plant Sciences* 8: e11337.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend. 2013. High-resolution global maps of 21st-century forest cover change. *Science* 342: 850–853, doi: 10.1126/science.1244693.

Hoang, D. T., O. Chernomor, A. von Haeseler, B. Q. Minh, and L. S. Vinh. 2018. UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* 35: 518–522.

IPNI. 2022. International Plant Names Index. http://www.ipni.org. The Royal Botanic Gardens, Kew, Harvard University Herbaria and Libraries, and Australia National Herbarium (last accessed 27 July 2023).

Isaac, N. J. B., J. Mallet, and G. M. Mace. 2004. Taxonomic inflation: Its influence on macroecology and conservation. *Trends in Ecology & Evolution* 19: 464–469.

Johnson, M. G., L. Pokorny, S. Dodsworth, L. R. Botigué, R. S. Cowan, A. Devault, W. L. Eiserhardt, et al. 2019. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic Biology* 68: 594–606.

Jones, G. 2017. Algorithmic improvements to species delimitation and phylogeny estimation under the multispecies coalescent. *Journal of Mathematical Biology* 74: 447–467.

Jones, M. R. and J. M. Good. 2016. Targeted capture in evolutionary and ecological genomics. *Molecular Ecology* 25: 185–202.

Kainulainen, K. and B. Bremer. 2014. Phylogeny of *Euclinia* and allied genera of Gardenieae (Rubiaceae), and description of *Melanoxerus*, an endemic genus of Madagascar. *Taxon* 63: 819–830.

Kainulainen, K., S. G. Razafimandimbison, and B. Bremer. 2013. Phylogenetic relationships and new tribal delimitations in subfamily Ixoroideae (Rubiaceae): Phylogeny of Ixoroideae. *Botanical Journal of the Linnean Society* 173: 387–406.

Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. von Haeseler, and L. S. Jermiin. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* 14: 587–589.

Katoh, K., J. Rozewicki, and K. D. Yamada. 2019. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics* 20: 1160–1166.

Killeen, T. J., E. García E., S. Beck, Missouri Botanical Garden, and Herbario Nacional de Bolivia. 1993. *Guía de Arboles de Bolivia.* La Paz, Bolivia: Editorial del Instituto de Ecologia and Saint Louis: Missouri Botanical Garden.

Liu, L., L. Yu, L. Kubatko, D. K. Pearl, and S. V. Edwards. 2009. Coalescent methods for estimating phylogenetic trees. *Molecular Phylogenetics and Evolution* 53: 320–328.

Mace, G. M. 2004. The role of taxonomy in species conservation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 359: 711–719.

May, R. M. 1990. Taxonomy as destiny. *Nature* 347: 129–130.

McLay, T. G. B., J. L. Birch, B. F. Gunn, W. Ning, J. A. Tate, L. Nauheimer, E. M. Joyce, L. Simpson, A. N. Schmidt-Lebuhn, W. J. Baker, F. Forest, and C. J. Jackson. 2021. New targets acquired: Improving locus recovery from the Angiosperms353 probe set. *Applications in Plant Science* 9: 10.1002/aps3.11420. doi: 10.1002/aps3.11420. PMID: 34336399; PMCID: PMC8312740.

Mendoza, H., B. Ramírez, and L. C. Jiménez. 2004. *Rubiaceae de Colombia. Guía Ilustrada de Géneros.* Bogota, Colombia: Instituto de Investigacion de Recursos Biologicos Alexander von Humboldt.

Miller, M. A., W. Pfeiffer, and T. Schwartz. 2012. The CIPRES science gateway: Enabling high-impact science for phylogenetics researchers with limited resources. Pp. 1–8 in *Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the eXtreme to the Campus and Beyond, XSEDE '*12. Chicago: Association for Computing Machinery.

Milliken, W., R. P. Miller, and P. S. Pollard. (1992). *Ethnobotany of the Waimiri Atroari Indians of Brazil.* Kew: Royal Botanic Gardens Publishing.

Minh, B. Q., H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler, and R. Lanfear. 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37: 1530–1534.

Mori, S. A. 1997. *Guide to the Vascular Plants of Central French Guiana.* Bronx, New York: New York Botanical Garden.

Mouly, A., K. Kainulainen, C. Persson, A. Davis, K. Wong, S. Razafimandimbison, and B. Bremer. 2014. Phylogenetic structure and clade circumscriptions in the Gardenieae complex (Rubiaceae). *Taxon* 63: 801–818.

Persson, C. 2000. Phylogeny of Gardenieae (Rubiaceae) based on chloroplast DNA sequences from the rps16 intron and trnL(UAA)-F(GAA) intergenic spacer. *Nordic Journal of Botany* 20: 257–270.

Persson, C. 2003. Agouticarpa, a new Neotropical genus of Tribe Gardenieae (Rubiaceae). *Brittonia* 55: 176–201.

Pittier, H. and C. D. Mell. 1931. *A Century of Trees of Panama*. United States Department of Agriculture. Biodiversity Heritage Library. Details - A century of trees of Panama - Biodiversity Heritage Library (biodiversitylibrary.org) (last accessed 23 October 2023).

ProFound. 2005. *Market Brief in the European Union for Selected Natural Ingredients Derived from Native Species:* Genipa americana *Jagua, huito.* The United Nations Conference on Trade and Development (UNCTAD).

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rakotonasolo, F. and A. Davis. 2006. Six species of Madagascan *Genipa* transferred to *Hyperacanthus* (Rubiaceae–Gardenieae) and new data on general morphology, placentation and ovary structure in *Hyperacanthus. Taxon* 55: 387–396.

Rambaut, A. 2017. FigTree-version 1.4.3, a graphical viewer of phylogenetic trees. Computer program distributed by the author. http://tree.bio.ed.ac.uk/software/figtree.

Rambaut, A., A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard. 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* 67: 901–904, doi: 10.1093/sysbio/syy032.ac.uk/Tracer. Available at http://beast.bio.ed.

Rannala, B. and Z. Yang. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* 164: 1645–1656.

Ribeiro, M. C., J. P. Metzger, A. C. Martensen, F. J. Ponzoni, and M. M. Hirota. 2009. The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. *Biological Conservation* 142: 1141–1153.

Ridley, R., C. Persson, B. Oxelman, T. Andermann, and C. D. Bacon. 2024. Data from: A phylogenomic analysis of *Genipa* (Rubiaceae) using target sequence capture data. Dryad Digital Repository. https://doi.org/10.5061/dryad.7wm37pw0g.

Schumann, K. M. 1889. Rubiaceae. In *Flora Brasiliensis* 6(6), ed. C. F. P. Martius. Missouri Botanical Garden.

Standley, P. C. and J. A. Steyermark. 1949. *Flora of Guatemala.* Chicago, Illinois: Chicago Natural History Museum.

Steyermark, J. A. 1972. The Botany of the Guyana Highland. *Memoirs of the New York Botanical Garden.* New York Botanical Garden Press.

Steyermark, J. A. 1974. Rubiaceae, *Genipa*. Pp. 660–669 in *Flora de Venezuela 9*, eds. T. Lasser and J. A. Steyermark. Carácas: Instituto Botánico.

Steyermark, J. A. 1988. Flora of the Venezuelan Guayana -V. *Annals of the Missouri Botanical Garden* 75: 1058.

Steyermark, J. A. and C. Persson. 2004. *Flora of the Venezuelan Guayana: Poaceae - Rubiaceae.* St. Louis, Missouri: Missouri Botanical Garden Press.

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

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

Ter Steege, H., N. C. A. Pitman, T. J. Killeen, W. F. Laurance, C. A. Peres, J. E. Guevara, R. P. Salomão, et al. 2015. Estimating the global conservation status of more than 15,000 Amazonian tree species. *Science Advances* 1: e1500936.

The World Checklist of Vascular Plants. 2023. World Checklist of Vascular Plants, version 2.0. Facilitated by the Royal Botanic Gardens, Kew. http://wcvp.science.kew.org/ (last accessed 27 July 2023).

Tropicos. 2023. Tropicos.org. v. 3.4.2. Missouri Botanical Garden. https://tropicos.org (last accessed 27 July 2023).

Woodson, R. E., R. W. Schery, and J. D. Dwyer. 1980. Flora of Panama. Part IX. Family 179. Rubiaceae, Part 1. *Annals of the Missouri Botanical Garden* 67: 1–256.

WFO. 2023. World Flora Online. http://www.worldfloraonline.org (last accessed 27 July 2023).

Zappi, D. C., J. Semir, and N. I. Pierozzi. 1995. *Genipa infundibuliformis* sp. nov. and notes on *Genipa americana* (Rubiaceae). *Kew Bulletin* 50: 761–771.

Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19: 153.

Zizka, A., D. Silvestro, T. Andermann, J. Azevedo, C. D. Ritter, D. Edler, H. Farooq, et al. 2019. CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution* 10: 744–751.

Appendix 1. List of voucher information for 29 samples used in this study. Taxon name, sample ID, voucher information (collector, collection number, location, and date), herbarium code, and NCBI SRA

number, — missing information or voucher not seen. NCBI SRA BioProject ID PRJNA1029819.

***Genipa americana* var. *americana*,** G_am5, (Idarraga, 5330, Colombia, Antioquia, —), GB, SAMN37879591. ***Genipa americana* var. *americana*,** G_am6, (Persson, 612, Peru, Loreto, 19 September 2002), GB, SAMN37879592. ***Genipa americana* var. *americana*,** G_am8, (Persson, 1865, Ecuador, Pichincha, 11 May 2014), GB, SAMN37879594. ***Genipa americana* var. *americana*,** G_am9, (Persson, 1866, Ecuador, Pichincha, 11 May 2014), GB, SAMN37879595. ***Genipa americana* var. *americana*,** G_am14, (Stahl, 7529, Ecuador, Monocongo, —), GB, SAMN37879600. ***Genipa americana* var. *americana*,** G_am22, (Tuberquia, 296, Colombia, Choco, —), — SAMN37879605. ***Genipa americana* var. *americana*,** G_am24, (Rova, 2372, Panama, Colon, 1 June 1997), GB, SAMN37879607. ***Genipa americana* var. *americana*,** G_am25, (Persson, 306, Bolivia, Santa Cruz, 9 October 1996), GB, SAMN37879608. ***Genipa americana* var. *americana*,** G_am26, (Persson, 2143, Colombia, Amazonas, 19 April 1994), GB, SAMN37879609. ***Genipa americana* var. *americana*,** G_am29, (Persson, 231, Bolivia, Beni, 11 September 1996), GB, SAMN37879612. ***Genipa americana* var. *americana*,** G_am30, (Alzate, 225, Colombia, Antioquia, 20 May 1997), GB, SAMN37879613. ***Genipa americana* var. *caruto*,** G_car3, (Santamarïa, S-959, Costa Rica, Punta Arenas, 23 October 2005), GB, SAMN37879589. ***Genipa americana* var. *caruto*,** G_car4, (Ståhl, 5849, Ecuador, Los Rïos, 23 May 2002), GB, SAMN37879590. ***Genipa americana* var. *caruto*,** G_car13, (Persson, 342, Bolivia, Santa Cruz, 15 October 1996), GB, AMN37879599. ***Genipa americana* var. *caruto*,** G_car20, (Rova, 2402, Panama, Chiriqui, 12 June 1997), GB, SAMN37879604. ***Genipa americana* var. *caruto*,** G_car23, (Jansen, 3680, Guyana, Rupununi, 10 February 1994), NY, SAMN37879606. ***Genipa americana* var. *caruto*,** G_car27, (Persson, 1976, French Guiana, Ile de Cayenne, 14 March 1994), GB, SAMN37879610. ***Genipa americana* var. *caruto*,** G_car28, (Rova, 2388, Panama, Panama, 9 June 1997), GB, SAMN37879611. ***Genipa americana* var. *caruto*,** G_car32, (Jansen, 4031, Guyana, Rupununi, Dadanawa, 10 June 1995), GB, SAMN37879614. ***Genipa infundibuliformis*,** G_infun15, (Antonelli, 406, Brazil, Sao Paulo, Campinas, 14 September 2008), GB, SAMN37879601. ***Genipa infundibuliformis*,** G_infun33, (Antonelli, 327, Brazil, —), —, SAMN37879615. ***Genipa spruceana*,** G_spru1, (Persson, 1612, Ecuador, Orellana, 22 October 2010), GB, SAMN37879587. ***Genipa spruceana*,** G_spru7, (Persson, 606, Peru, Loreto, 18 September 2002), GB, SAMN37879593. ***Genipa spruceana*,** G_spru10, (Persson, 604, Peru Loreto, 18 September 2002), GB, SAMN37879596. ***Genipa spruceana*,** G_spru11, (Persson, 1802, Ecuador, Sucumbïos, 4 May 2014), GB, SAMN37879597. ***Genipa spruceana*,** G_spru12, (Persson, 674, Peru, Loreto, 4 October 2002), GB, SAMN37879598. ***Genipa spruceana*,** G_spru18, (Persson, 1959, French Guiana, Crique Tibourou, 12 March 1994), GB, SAMN37879602. ***Genipa spruceana*,** G_spru19, (Antonelli, 246, Brazil, Amazonas, 7 January 2003), GB, SAMN37879603. ***Tocoyena pittieri*,** T_pit, (Santamarïa, S-936, Costa Rica, —), —, SAMN37879588.