

A target enrichment probe set for resolving phylogenetic relationships in the coffee family, Rubiaceae

Laymon D. Ball¹  | Ana M. Bedoya¹  | Charlotte M. Taylor² | Laura P. Lagomarsino¹ 

¹Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana, 70803, USA

²Missouri Botanical Garden, 4344 Shaw Blvd., Saint Louis, Missouri, 63110, USA

Correspondence

Laymon D. Ball, Louisiana State University, 202 Life Science Building, Baton Rouge, Louisiana 70803, USA.
Email: lball4@lsu.edu

Abstract

Premise: Rubiaceae is among the most species-rich plant families, as well as one of the most morphologically and geographically diverse. Currently available phylogenies have mostly relied on few genomic and plastid loci, as opposed to large-scale genomic data. Target enrichment provides the ability to generate sequence data for hundreds to thousands of phylogenetically informative, single-copy loci, which often leads to improved phylogenetic resolution at both shallow and deep taxonomic scales; however, a publicly accessible Rubiaceae-specific probe set that allows for comparable phylogenetic inference across clades is lacking.

Methods: Here, we use publicly accessible genomic resources to identify putatively single-copy nuclear loci for target enrichment in two Rubiaceae groups: tribe Hillieae (Cinchonoideae) and tribal complex Palicoureeae+Psychotrieae (Rubioideae). We sequenced 2270 exonic regions corresponding to 1059 loci in our target clades and generated in silico target enrichment sequences for other Rubiaceae taxa using our designed probe set. To test the utility of our probe set for phylogenetic inference across Rubiaceae, we performed a coalescent-aware phylogenetic analysis using a subset of 27 Rubiaceae taxa from 10 different tribes and three subfamilies, and one outgroup in Apocynaceae.

Results: We recovered an average of 75% and 84% of targeted exons and loci, respectively, per Rubiaceae sample. Probes designed using genomic resources from a particular subfamily were most efficient at targeting sequences from taxa in that subfamily. The number of paralogs recovered during assembly varied for each clade. Phylogenetic inference of Rubiaceae with our target regions resolves relationships at various scales. Relationships are largely consistent with previous studies of relationships in the family with high support (≥ 0.98 local posterior probability) at nearly all nodes and evidence of gene tree discordance.

Discussion: Our probe set, which we call Rubiaceae2270x, was effective for targeting loci in species across and even outside of Rubiaceae. This probe set will facilitate phylogenomic studies in Rubiaceae and advance systematics and macroevolutionary studies in the family.

KEYWORDS

phylogenetics, probe set, Rubiaceae, target enrichment

Target enrichment has emerged as a relevant and widely applicable tool for generating large-scale genomic data for resolving phylogenies across the tree of life. This reduced-representation sequencing strategy is characterized by

targeting pre-selected single-copy loci from across the genome, which are isolated using RNA probes before sequencing. The resulting data set consists of hundreds to thousands of individual nuclear loci sequenced across taxa

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2023 The Authors. *Applications in Plant Sciences* published by Wiley Periodicals LLC on behalf of Botanical Society of America.

(Andermann et al., 2019). These data sets are ideal for applying complex phylogenetic models, including species tree methods that incorporate incomplete lineage sorting and model reticulate evolution. This method also has the benefit of being robust to low-quality, low-quantity input DNA, meaning that specimens from natural history collections can serve as a DNA input source (Hart et al., 2016).

Studies employing target enrichment data can rely either on universal probe sets, which target loci that are highly conserved across major clades (e.g., angiosperms [Johnson et al., 2019] or flagellate plants [Breinholt et al., 2021]), or on lineage-specific probe sets. The relative utility of lineage-specific vs. universal probe sets depends on the genetic distance between the probe sequences and the target regions in the group of interest, as well as the scope of the phylogenetic study (Kadlec et al., 2017; Chau et al., 2018; Soto Gomez et al., 2019; Straub et al., 2020; Yardeni et al., 2022). While universal probe sets such as Angiosperms353 are an important community resource and have been successfully used to resolve evolutionary relationships at both macro- and micro-evolutionary timescales (Smith et al., 2014; Slimp et al., 2021; Le et al., 2022), they also present challenges. For example, universal probes are developed to recover primarily highly conserved loci and may not be useful for resolving relationships at or below the species level, especially when applied to rapid radiations or very shallow phylogenetic splits. Difficulty with sequence assembly can arise due to the negative correlation between enrichment success and the degree of divergence between focal taxa and the taxa used for probe design (Liu et al., 2019). This can result in fewer and less complete loci when using a universal probe set as compared to a family-specific one (Siniscalchi et al., 2021; Ufimov et al., 2021; Yardeni et al., 2022). This also decreases the number of informative characters available for phylogenetic inference. Furthermore, as the degree of ploidy is extremely variable across angiosperms, using a universal probe set can result in a high proportion of paralogs that complicate phylogenetic inference, particularly in groups where polyploidy is common (Frost et al., 2022).

Multiple studies have found that lineage-specific probe sets outperform universal probe sets in species-level phylogenetic studies (Ufimov et al., 2021; Yardeni et al., 2022). An advantage of using lineage-specific probe sets is that they can be designed to be as taxon-specific as the user wants, mitigating some of the issues with universal probe sets. As a result, loci from lineage-specific probe sets tend to have more variable sites (Soto Gomez et al., 2019; Shah et al., 2021; Ufimov et al., 2021; Yardeni et al., 2022). Fortunately, predesigned, lineage-specific probe sets exist for many major clades of plants, including Asteraceae (Mandel et al., 2014), Orchidaceae (Eserman et al., 2021), Brassicaceae (Nikolov et al., 2019), and Melastomataceae (Jantzen et al., 2020). For other groups, custom lineage-specific probe sets can easily be

designed for specific projects using a variety of existing pipelines (Chamala et al., 2015; Schmickl et al., 2016; Jantzen et al., 2020), as long as there are genomic resources (such as draft genomes or transcriptome references) available. Despite their shortcomings, the utility of universal probe sets cannot be refuted, and when financially feasible it may be advantageous to incorporate both universal and lineage-specific probes in target enrichment studies (Hendriks et al., 2021).

Rubiaceae is the fourth largest flowering plant family, with ~13,500 species in ~620 genera (Plants of the World Online, 2023). While most of its genera and species are tropical, Rubiaceae species occur in nearly all habitats, ranging from taiga to rainforest, and on all continents including Antarctica. Along with geographical diversity, the family exhibits high variation in floral and fruit morphology, habit, and ecological interactions. This variation has inspired multiple studies to understand patterns of trait evolution (Bremer and Eriksson, 1992; Ferrero et al., 2012; Razafimandimbison et al., 2014; Ehrendorfer et al., 2019), and the family has been used as a model to characterize angiosperm macroevolution (Antonelli et al., 2009). While many phylogenetic studies have been performed for different clades in Rubiaceae (Löfstrand et al., 2019; Borges et al., 2021; Razafimandimbison et al., 2021; Amenu et al., 2022), as well for the entire family (Bremer and Eriksson, 2009), these have mostly relied on a few loci. Only a handful of studies have used genomic-scale phylogenetic data to resolve relationships within Rubiaceae, with most of these using the Angiosperms353 data (Antonelli et al., 2021; Canales et al., 2022; Thureborn et al., 2022) and another using microarray technology (Prata et al., 2018).

Here, we present the first lineage-specific probe set for target enrichment phylogenetics of Rubiaceae. Using publicly accessible genomic resources for the family, we identified a set of 2270 exonic regions across 1059 low-copy nuclear loci. We explored the performance of these loci using both *in silico* and *in vitro* methods, the latter targeting two distantly related and understudied groups: tribe Hillieae (Cinchonoideae) and tribal complex Palicoureae+Psychotrieae (Rubioidae). To assess the strength of our probe set in recovering intertribal and subfamilial relationships, we also inferred phylogenetic relationships for a subset of 27 Rubiaceae taxa and an outgroup (Apocynaceae) using a coalescent-aware approach. Our probe set was successful in recovering loci in Hillieae and Palicoureae+Psychotrieae, as well as in other tribes of Rubiaceae, and inferred phylogenetic relationships were largely consistent with previous studies in the family. Therefore, we anticipate that this probe set will be useful for resolving relationships across the family and can complement studies employing universal probe sets. This resource will facilitate multi-locus phylogenetic studies by the community of researchers studying the coffee family and improve our understanding of evolutionary relationships and macroevolutionary dynamics in this important plant group.

METHODS

Locus selection

Putatively single- to low-copy nuclear loci were identified from existing Rubiaceae genomic data: six transcriptomes (*Mapouria douarrei* Beauvis., *Psychotria marginata* Sw., *Carapichea ipecacuanha* (Brot.) L. Andersson, *Cinchona pubescens* Vahl, *Hamelia patens* Jacq., and *Neolamarckia cadamba* (Roxb.) Bosser) and four genome-skimming paired-end reads (*Corynanthe mayumbensis* (R. D. Good) Raym.-Hamet ex N. Hallé, *Mitragyna speciosa* (Korth.) Havil., *Gardenia jasminoides* J. Ellis, and *Galium odoratum* (L.) Scop.). Transcriptomes were retrieved from the National Center for Biotechnology Information (NCBI) TSA database (<https://www.ncbi.nlm.nih.gov/genbank/tsa/>), the 1000 Plant Transcriptomes project (1KP) database (Matasci et al., 2014), and MedPlant RNA Seq Database (<https://medplantnrnaseq.org>). Genome skim data were retrieved from the NCBI SRA database (<https://www.ncbi.nlm.nih.gov/sra>). The chloroplast plastome from *Gardenia jasminoides*, available from the NCBI Genome database (<https://www.ncbi.nlm.nih.gov/genome/>), was used to identify plastome sequences. No mitochondrial reference was used in probe design.

Sondovač version 1.3 was used to identify putative orthologous loci from the publicly mined genomic resources (Schmickl et al., 2016). Briefly, Sondovač identifies single-copy nuclear loci from low-coverage genome-skimming data that are mapped to a transcriptome while discarding non-coding reads and those that map to organellar genomes. To reduce the presence of multi-copy regions in our data set, duplicated transcripts with a BLAT score above 1000 were removed in Sondovač. The filtered genome reads were subsequently de novo assembled into exonic regions (ERs) using Geneious (version 2021.1.1; Biomatters, Auckland, New Zealand). Recovery of complete exons is not guaranteed, as this is determined by the quality of the input transcriptomes and genome skim data sets. ERs were filtered to maintain contigs ≥ 120 bp, and these were subsequently filtered to

keep those that belonged to transcripts with a total length ≥ 360 bp across all contigs. Lastly, we removed ERs with $\geq 90\%$ sequence similarity from the data set.

While Sondovač uses genome-skimming data for one sample and one transcriptome at a time, previous research has found that the number of identified loci can be maximized if multiple genome-skimming data sets are combined and paired with each available transcriptome (Uribe-Converse, 2016; Bagley et al., 2020). Adopting this strategy, we combined all genome-skimming reads belonging to the same subfamily, and the resulting composite genome-skimming read sets were paired with each same-subfamily transcriptome. Six separate Sondovač runs were performed in total, following the design in Table 1. This resulted in three sets of loci derived from Rubioideae genomic resources (RUB) and three sets derived from Cinchonoideae genomic resources (CIN). To identify loci shared among taxa, ERs with $\geq 90\%$ sequence similarity across the six sets were clustered using cd-hit-est (V4.8.1; Li and Godzik, 2006), and the longest sequence in each cluster was retained. We subsequently removed unique ERs from the set to avoid including sequences that are not shared across taxa and were left with 2270 remaining final ERs, each of which may be a complete or partial exon.

To determine whether our developed locus data set shared sequences with the Angiosperms353 data set, we ran a BLAST search of our 2270 ERs against the Angiosperms353 representative sequences (Johnson et al., 2019; <https://github.com/mossmatters/Angiosperms353/tree/master/Probes>).

Sampling and DNA extraction for target enrichment

We included 132 species of Rubiaceae, representing 164 samples. Sampling was mostly from two distantly related clades: (1) Hillieae (Cinchonoideae), represented by 26 species and 44 accessions in the genera *Hillia*, *Cosmibuena*, and *Balmea* (90% of all species in the clade); and (2) Palicoureeae and the closely related Psychotriaceae (Rubioideae), represented by 102 species and 109 accessions in

TABLE 1 Sondovač design and output. Sources of genomic resources are included in Appendix 2.

| Locus set | Transcriptome input | Genome skim input | Plastome input | Unfiltered ERs | Final ERs |
|-----------|-------------------------------|---|-----------------------------|----------------|-----------|
| RUB1 | <i>Mapouria douarrei</i> | Combined <i>Galium</i> paired-end reads | <i>Gardenia jasminoides</i> | 273 | 38 |
| RUB2 | <i>Carapichea ipecacuanha</i> | | | 1930 | 429 |
| RUB3 | <i>Psychotria marginata</i> | | | 1676 | 201 |
| CIN1 | <i>Hamelia patens</i> | Combined <i>Neolamarckia</i> and <i>Corynanthe</i> paired-end reads | | 2904 | 462 |
| CIN2 | <i>Neolamarckia cadamba</i> | | | 12,209 | 834 |
| CIN3 | <i>Cinchona pubescens</i> | | | 1108 | 306 |

Note: ER = exonic region.

Palicourea, *Notopleura*, *Rudgea*, *Carapichea*, and *Eumachia* in Palicoureae and *Psychotria* in Psychotriaceae (~5% of all species in the clade; Appendix 1). Our sampling also included additional members of Cinchonoideae: one *Hamelia patens* (Hameliaceae), one *Hoffmannia phoenicopoda* K. Schum. (Hameliaceae), one *Coutarea hexandra* (Jacq.) K. Schum. (Chiococceae), and one *Ferdinandusa paraensis* Ducke (Dialypetalanthaeae) (Appendix 1). From the total samples included in this study, 79% came from herbarium specimens deposited at MO or LSU (acronyms following Thiers, 2023), including all Hillieae accessions (Appendix 1).

We extracted total genomic DNA from leaf tissue using a sorbitol extraction protocol adapted from Štorchová et al. (2000). DNA concentration was assessed with a Qubit fluorometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA), and fragment size distribution was visualized with gel electrophoresis (1% agarose). Sample extractions that yielded low DNA concentrations were repeated and then pooled together in order to isolate enough DNA for sequencing (i.e., ≥ 280 ng).

Target enrichment probe design, sequencing, and assembly

Probe design, library preparation, and sequencing were conducted by RAPID Genomics (Gainesville, Florida, USA). Using the 2270 identified ERs as templates and a proprietary workflow, RAPID Genomics synthesized 14,429 biotinylated 120-mer RNA probes in an overlapping strategy to reach full coverage of each locus. Following library preparation and target enrichment, 150-bp paired-end read sequencing was conducted in one lane of an Illumina NovaSeq S4 (Illumina, San Diego, California, USA).

The quality of demultiplexed raw paired-end sequenced reads was assessed in FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), and adapter sequences were trimmed with Trimmomatic (version 0.39; Bolger et al., 2014) using default settings. Cleaned reads were assembled with HybPiper (version 2.0.1; Johnson et al., 2016). For each clade of interest (i.e., Hillieae and Palicoureae+Psychotriaceae),

we applied two approaches to data assembly to assess performance of (1) individual ERs or single-exonic regions (SERs) and (2) ERs concatenated into multi-exonic regions (MERs). With the first approach, we were able to more directly assess the utility and performance of our target sequences. The second approach allows us to recover longer sequences that likely include more informative sites useful for gene tree inference. Analyzing exons individually may be desirable for data sets that have a high proportion of multi-copy loci and paralogs in which the assembly of supercontigs (i.e., sequences containing both coding and non-coding flanking sequences) that span multiple exons can result in chimeric sequences (Morales-Briones et al., 2022).

In the first approach, the individual 2270 ERs were added to the HybPiper target file and used as references for assembly. In the second approach, all ERs within each gene were concatenated within the target file. Same-gene ERs were identified by mapping the 2270 ER sequences to multiple sets of assembled CapSim captured fragments (see section below) in an iterative fashion (Appendix S1; see Supporting Information with this article) using the Geneious mapper function in Geneious (version 2021.1.1). This approach allowed us to concatenate ERs into loci without knowledge of their synteny within each locus.

In both the SER and MER assemblies, we used the HybPiper intronerate.py script to extract supercontigs. We calculated summary statistics for all assemblies (Table 2, Appendices S2 and S3) and identified putative paralogs with HybPiper.

Testing loci and probe performance across Rubiaceae using simulations

We performed multiple in silico target enrichment experiments using CapSim to (1) assess whether loci that were identified via Sondovač would be captured in other, distantly related Rubiaceae taxa prior to probe design and (2) assess how well the final probes would perform on taxa from across Rubiaceae (Cao et al., 2018). Briefly, CapSim performs in silico targeted enrichment experiments by

TABLE 2 HybPiper single-exonic region (SER) and multi-exonic region (MER) assembly statistics averaged for each clade. Statistics are presented as SER/MER. Additional HybPiper statistics are provided in Appendix S6.

| Statistics | Hillieae | Other CIN | Palicoureae+Psychotriaceae | Other RUB | Ixoroideae | Apocynaceae |
|----------------------|----------|-----------|----------------------------|-----------|------------|-------------|
| PctOnTarget | 85/84 | 58/61 | 74/74 | 39/39 | 32/32 | 48/48 |
| GenesWithSeqs | 1736/883 | 1377/685 | 1786/943 | 895/433 | 1204/577 | 403/224 |
| GenesAt25pct | 1712/819 | 1355/625 | 1759/865 | 869/393 | 1168/520 | 391/199 |
| GenesAt50pct | 1701/794 | 1340/593 | 1754/851 | 865/384 | 1162/510 | 384/189 |
| GenesAt75pct | 1669/764 | 1294/557 | 1733/823 | 847/370 | 1139/496 | 365/175 |
| GenesAt150pct | 1341/549 | 882/417 | 1513/640 | 688/333 | 910/443 | 242/117 |
| ParalogWarningsDepth | 596/274 | 383/153 | 336/124 | 37/17 | 65/27 | 10/5 |

Note: CIN = Cinchonoideae; RUB = Rubioideae.

taking a set of sequences and an assembled genome as input and simulates probe hybridization and Illumina sequencing, resulting in a set of captured fragments representing paired-end reads (Cao et al., 2018).

First, we ran CapSim using five Rubiaceae assembled genomes from GenBank (Appendix 2): three Ixoroideae (*Coffea canephora* Pierre ex A. Froehner, *Coffea eugenioides* S. Moore, and *Gardenia jasminoides*) and two Rubioideae (*Leptodermis oblonga* Bunge and *Ophiorrhiza pumila* Champ. ex Benth.). We also included one outgroup Apocynaceae, also in Gentianales (*Rhazya stricta* Decne.). The classification of subfamily Ixoroideae has been in disagreement, with mixed evidence for the monophyly of Ixoroideae and its sister subfamily, Cinchonoideae (Rydin et al., 2017; Antonelli et al., 2021). Throughout this paper, we recognize both traditional subfamilies: Ixoroideae and Cinchonoideae *sensu stricto*. In these runs, the full length of the 2270 filtered ERs served as probes. While using CapSim in this way is not directly comparable to an *in vitro* target enrichment experiment that uses 120-bp probe sequences, it allows us to determine whether our loci are conserved across Rubiaceae, and thus, their likely empirical utility. In each CapSim run, we simulated Illumina sequencing and chose to produce between 10 and 40 million captured fragments (see Appendix S4), approximating real sequencing effort. Each resulting set of captured fragments was mapped to its original reference genome assembly and contigs were *de novo* assembled while trimming paired-end read overhangs in Geneious (version 2021.1.1), using the ‘Geneious mapper’ function under default settings. After assembling contigs from the CapSim output for these six taxa, we used the total number of contigs assembled as a proxy for sequence recovery from our probe set, with a higher number indicating greater utility within that clade.

After validating the performance of ERs across Rubiaceae using the first CapSim approach described above, we assessed how well the final designed probes would perform on other taxa from across Rubiaceae. We reran CapSim, using the 14,429 120-mer RNA probe sequences, rather than the full-length ERs as above. We simulated 10 million

captured fragments from all assembled genomes listed in Appendix S5. This approach approximates how the probes designed by RAPiD Genomics from our target loci will perform on other Rubiaceae taxa, especially those distantly related from our focal subfamilies Rubioideae and Cinchonoideae for which we did not generate empirical target enrichment data. We followed the same downstream methodology as stated in the previous section: HybPiper was used to assemble captured fragments using both approaches (i.e., using SERs and MERs), calculate summary statistics, and identify potential paralogs.

To assess the informativeness of the sequence data captured across taxonomic scales in Rubiaceae, we created multiple sequence alignments of the ER assemblies (Table 3) with MAFFT (Katoh and Standley, 2013). Summary statistics were calculated for each alignment set with AMAS (Borowiec, 2016). No cleaning was performed on these alignments prior to calculating summary statistics.

Phylogenetic inference

To assess the phylogenetic utility of our target sequences at different taxonomic scales and to compare our SER and MER data sets, we inferred two coalescent-aware phylogenetic trees using each data type and a subset of our sampled taxa. We included a small group of 10 closely related *Palicourea* in subfamily Rubioideae, three species from tribe Hillieae in subfamily Cinchonoideae (*Hillia triflora* var. *triflora* (Oerst.) C. M. Taylor, *Balmea stormiae* Martínez, and *Cosmibuena grandiflora* (Ruiz & Pav.) Rusby), and two species from Hamelieae in subfamily Cinchonoideae (*Hamelia patens* and *Hoffmannia phoenicopoda*). We also included all taxa from the CapSim experiments: four Rubioideae from four tribes (Morindeae, Paederieae, Rubieae, and Ophiorrhizeae), four Ixoroideae from two tribes (Coffeeae and Gardenieae), two Cinchonoideae from two tribes (Cinchoneae and Naucleaeae), and outgroup *Rhazya stricta* from Apocynaceae (Appendix 2). We excluded all loci flagged as paralogous by HybPiper from

TABLE 3 AMAS single-exonic region alignment statistics, showing average values for each statistic with ranges in parentheses. Additional AMAS statistics are provided in Appendix S8.

| Statistics | Hillieae | Palicoureeae+Psychotrieae | All combined ^a |
|----------------------------------|-------------------|---------------------------|---------------------------|
| No_of_taxa | 32 (2–41) | 90 (2–109) | 123 (2–164) |
| Alignment_length (bp) | 1510 (282–12,591) | 2199 (227–32,805) | 4184 (389–41,669) |
| Missing_percent (%) | 52 (7–80) | 56 (1.5–90) | 73 (14–95) |
| No_variable_sites | 247 (1–1237) | 808 (2–9764) | 1473 (11–13,407) |
| Proportion_variable_sites | 0.2 (0.002–0.5) | 0.4 (0.003–0.7) | 0.4 (0.01–0.6) |
| Parsimony_informative_sites | 85 (0–674) | 386 (0–5462) | 710 (0–7080) |
| Proportion_parsimony_informative | 0.06 (0–0.3) | 0.2 (0–0.5) | 0.2 (0–0.5) |

^a“All combined” alignments include Hillieae, Palicoureeae+Psychotrieae, and all outgroups (including CapSim samples).

the analysis, leaving a remainder of 356 out of 2270 SERs and 428 out of 1059 MERs to be used for downstream analyses. In analyses of both data sets, sequences were aligned using MAFFT (Katoh and Standley, 2013). Columns with more than 20% missing data were removed from each alignment using Phyx version 1.3 (Brown et al., 2017), and spurious sequences, here defined as sequences shorter than 20% of the total alignment length, were removed using trimAl (Capella-Gutiérrez et al., 2009). We also removed erroneous sequences from alignments using TAPER (Zhang et al., 2021).

After cleaning the alignments, gene trees were inferred using RAxML-NG (Kozlov et al., 2019) under the GTR+G substitution model. This method automatically excludes alignments with fewer than four taxa. All other gene trees were kept, but nodes with less than 20% support were collapsed using Newick Utilities version 1.6 (Junier and Zdobnov, 2010). Species trees were inferred using ASTRAL III (Zhang et al., 2018) with default settings. A total of 322 and 406 gene trees (with four or more taxa) from the SER and MER data sets, respectively, were inferred and used as input for each ASTRAL analysis. We used the local posterior probability scores (LPP) and normalized quartet scores provided by ASTRAL to estimate branch support and explore gene tree discordance.

RESULTS

Locus selection and testing

We identified a total of 20,100 ERs across the six runs with Sondovač. The initial run of cd-hit-est resulted in 17,464 clusters, although most (15,194 = 87%) included only a single ER. These unique ERs predominantly originated from the CIN1 and CIN2 ER sets (Table 1), likely due to the relatively high quality and large size of the *Hamelia patens* and *Neolamarckia cadamba* transcriptomes. After removal of unique ERs, the set was reduced to 2270 clusters; 662 consisted of sequences uniquely mined from Rubioideae, 1589 consisted of sequences uniquely mined from Cinchonoideae, and 19 included sequences from both Rubioideae and Cinchonoideae. Of the representative sequences selected to represent the 19 clusters shared between subfamilies, six came from Rubioideae and 13 from Cinchonoideae. Below, we refer to loci that are derived from Rubioideae sequences as RUB and those derived from Cinchonoideae as CIN. ER length was 120–4841 bp (\bar{x} = 323 bp); concatenating same-gene ERs resulted in a set of 1059 MER loci 125–4841 bp long (\bar{x} = 669 bp). Only 21 ERs (0.9%) had hits longer than 120 bp corresponding to Angiosperms353 loci.

The first round of CapSim target enrichment simulations to assess the performance of our probe set consistently resulted in sets of >500 assembled contigs despite variation in the size of the input assembled genome and number of reads simulated (Appendix S4). While paralogs and contigs

assembled from off-target reads may add to the total number of contigs assembled per input genome, the relatively high number of contigs assembled for all Rubiaceae taxa included suggests phylogenetic utility of our selected loci across the family.

Sequencing, assembly, and performance of probes across Rubiaceae

Silica-dried specimens tended to yield DNA with higher molecular weight than herbarium specimens (25 samples had DNA with a high enough molecular weight to be sheared prior to sequencing and only one of these was derived from a herbarium specimen); however, total DNA quantity varied considerably among these samples. While the samples with the lowest DNA quantities were derived from herbarium specimens, the sample with the most DNA also came from a herbarium specimen with degraded DNA. Of the 164 samples for which empirical target enrichment was attempted, 153 (~93%) were successfully sequenced (Appendix 1). All of the samples that failed sequencing were derived from herbarium specimens. Sequenced samples include at least one sample representing each of 25 Hillieae species (86% of all species), three other Cinchonoideae species (*Hamelia patens*, *Hoffmannia phoenicopoda*, and *Coutarea hexandra*), 97 Palicoureeae (~10% of all species), and 12 Psychotrieae (~2% of all species). On average, 4,684,470 reads were sequenced per sample.

HybPiper assembly statistics for SERs and MERs averaged for each clade (Rubioideae, Cinchonoideae, Ixoroideae, Apocynaceae) are summarized in Table 2 (additional statistics are provided in Appendix S6). HybPiper statistics for each sample are provided in Appendices S2 and S3 for SERs and MERs, respectively. Both newly sequenced samples and those that underwent CapSim simulation using the final designed probes are included in the assembly results (Table 2). The percentage of reads that mapped to sequences in the HybPiper target file (PctOnTarget) varied from 30% to 96% across all samples. CapSim-derived samples had a lower PctOnTarget, averaging 36% for SER assemblies, while all other samples had an average of 77%. We recovered an average of 1712 SERs and 892 MERs per Rubiaceae sample, amounting to 75% and 84% of SER and MER targets, respectively. Only six samples (five Hillieae and one Apocynaceae) had a recovery of less than 25% of SERs and MERs. Most SERs (71%) had both $\geq 75\%$ of the locus length recovered and $\geq 70\%$ taxon coverage.

In both the SER and MER assemblies, Palicoureeae+Psychotrieae samples had the greatest number of genes with assembled sequences on average, followed by Hillieae, other Cinchonoideae, Ixoroideae, other Rubioideae, and Apocynaceae (Table 2). Generally, the number and completeness of genes assembled for a given sample was greater when the sample originated from the same clade as

the genomic source from which the probe sequence was derived (Figure 1, Appendix S7). For Rubioideae assemblies, the percentage of the locus length recovered by HybPiper was higher for loci captured by probes derived from Rubioideae (RUB) vs. the probes derived from Cinchonoideae (CIN). Within this clade, 91% of SERs captured by RUB probes had both $\geq 75\%$ of the locus length recovered and $\geq 70\%$ taxon coverage, compared to 67% of SERs captured by CIN probes. Similarly for Cinchonoideae, the percentage of the locus length recovered is higher for loci captured by CIN probes. In this clade, 92% of SERs captured by CIN probes had both $\geq 75\%$ of the locus length recovered and $\geq 70\%$ taxon coverage, compared to 62% of SERs captured by RUB probes. For the Ixoroideae assemblies, completeness (in terms of taxon coverage and the percentage of the locus length recovered) was also higher for loci captured by CIN probes. Genes assembled for *Rhazya stricta* (Apocynaceae) had similar capture for both RUB and CIN probes. We observed no significant differences in the number of sequenced reads (two-tailed *t*-test *P* value = 0.14, $\alpha = 0.05$) or locus completeness (two-tailed *t*-test *P* value = 0.054, $\alpha = 0.05$) between samples derived from herbarium and silica-dried tissue.

The number of potential paralogs flagged by HybPiper, as identified by the number of loci for which the coverage depth of coding sequences extracted by Exonerate (Slater and Birney, 2005) within HybPiper was >1 for 75% of the length of the reference sequence (ParalogWarningsDepth in HybPiper), varied widely from clade to clade. To a lesser extent, the number of potential paralogs also varied according to

whether SERs or MERs were used as references during assembly. At the subfamily level, Cinchonoideae had the highest average proportion of paralogs (32% of SERs and 26% of MERs), followed by Rubioideae (18% of SERs and 13% of MERs), Ixoroideae (5% of both SERs and MERs), and Apocynaceae (1% of SERs and 2% of MERs; Table 2).

AMAS alignment statistics are summarized in Table 3 (additional statistics are provided in Appendix S8). Alignment statistics varied depending on which Rubiaceae taxa were included in the alignments (Figure 2A). For the set of alignments including all Rubiaceae sampled in this study, the number of variable sites ranged from 11–13,407 per locus ($\bar{x} = 1473$ sites) and the number of parsimony informative sites ranged from 0–7080 per locus ($\bar{x} = 710$ sites; Figure 2B). As expected, the average proportion of parsimony informative sites across loci was lower for within-clade (Hillieae and Palicoureaeae+Psychotriaceae) assemblies than for the across-clade assemblies (Figure 2A). In both cases, long loci composed of hundreds of informative sites were recovered.

Phylogenetic inference

Species tree inference from SERs (Figure 3) and MERs (Appendix S9) recovered the same phylogenetic relationships and varied in support at certain internal nodes. Notably, the placement of *Cinchona pubescens* as sister to the rest of Cinchonoideae is highly supported (LPP = 1.00) in the tree inferred from SERs but is poorly supported (LPP = 0.22) in

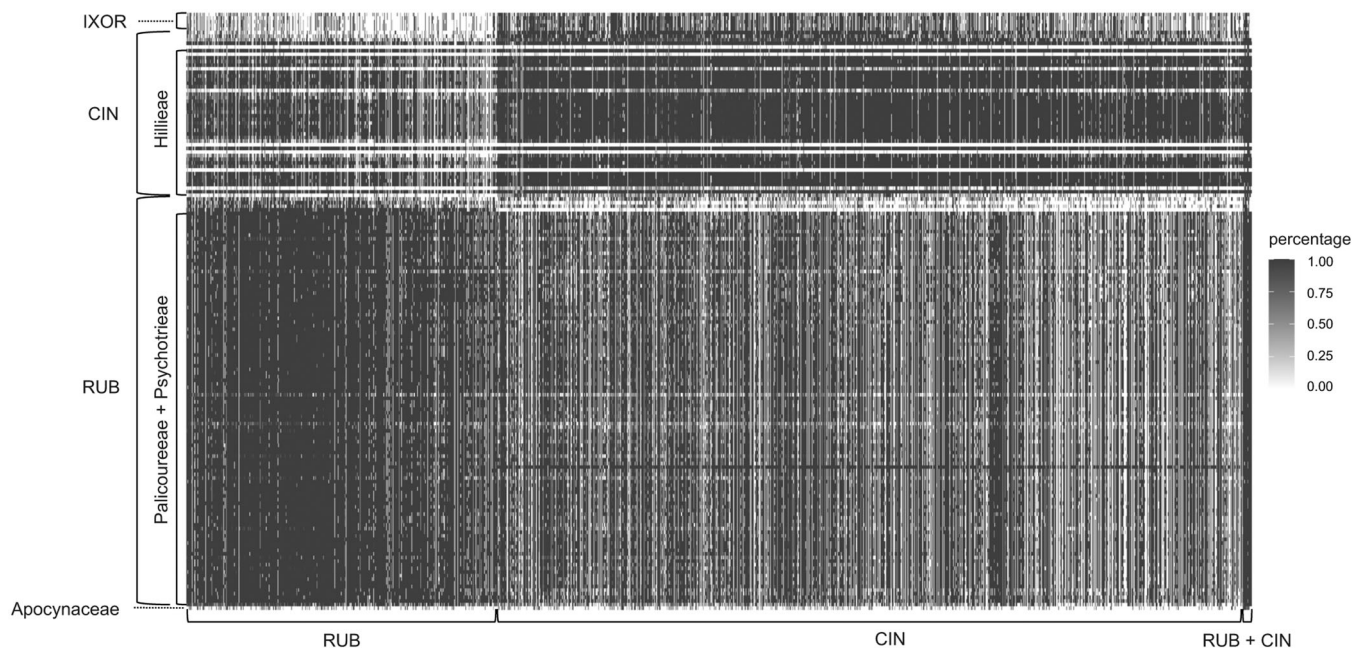


FIGURE 1 Heatmap showing the results of HybPiper single-exonic region (SER) assemblies for Cinchonoideae (CIN), Rubioideae (RUB), and Ixoroideae (IXOR) samples, and one Apocynaceae sample. Target SERs are on the *x*-axis and are grouped by the taxonomic source of the reference sequence. “RUB+CIN” indicates the 19 SERs from either RUB or CIN that shared $\geq 90\%$ sequence similarity with one or more SERs from other clades during cd-hit-est clustering. Samples are on the *y*-axis and are grouped by taxonomy. Darker shades represent higher percentages of the locus length recovered by HybPiper. Locus completeness tends to be higher when target species are from the same clade as the probe sequences.

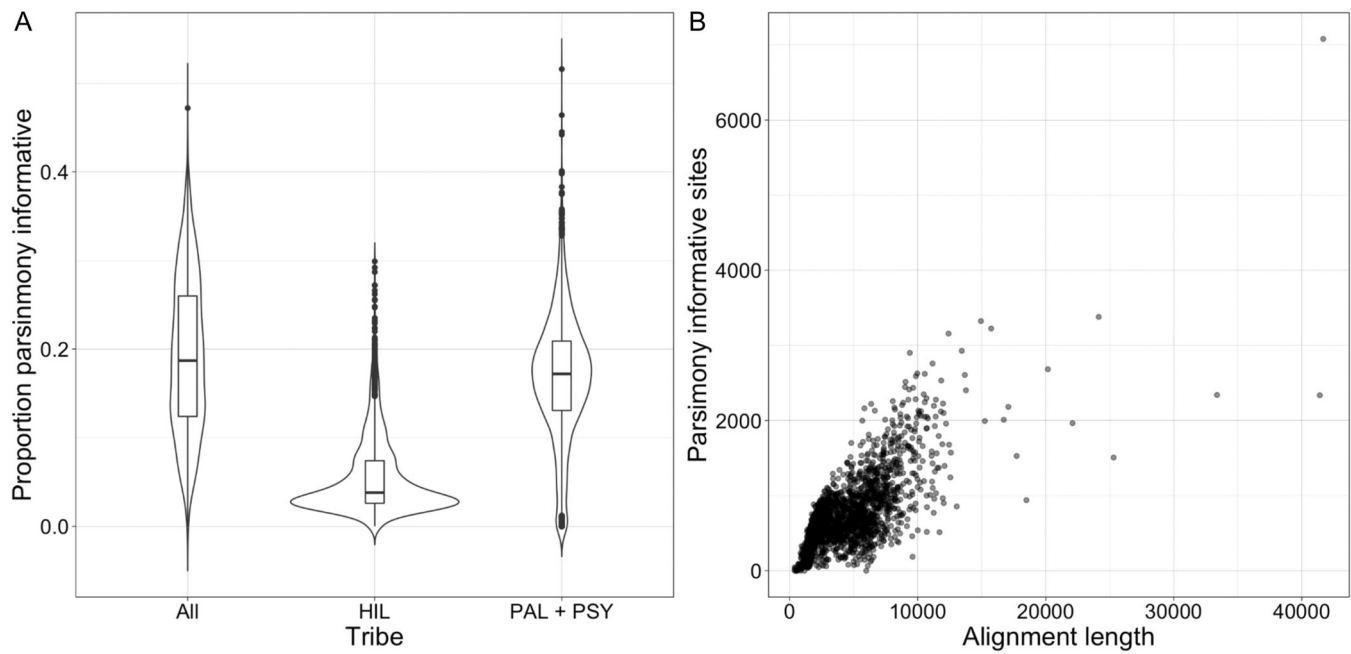


FIGURE 2 AMAS summary statistics for the single-exonic region alignments. (A) Violin plots showing the proportion of parsimony informative sites for within-clade (“HIL” and “PAL+PSY”) and across-clade (“All”) alignments. (B) Scatter plot of alignment length against the number of parsimony informative sites, including the set of alignments with all Rubiaceae combined.

the tree inferred from MERs. Additionally, two internal branches in Palicoureeae have lower support in the tree inferred from MERs compared to the tree inferred from SERs.

In analyses of both SERs and MERs, quartet topologies at most branches tended to be congruent with the species tree; in both analyses, the main quartet topology accounted for $\geq 50\%$ of possible topologies at 84% of branches. Quartet support for the main topology on internal branches of the SER tree ranged from 0.36–0.98 ($\bar{x} = 0.67$), support for the second alternative topology ranged from 0.01–0.4 ($\bar{x} = 0.17$), and support for the third alternative topology ranged from 0.01–0.4 ($\bar{x} = 0.16$). In the MER tree, quartet support for the main topology ranged from 0.37–0.97 ($\bar{x} = 0.65$), support for the second alternative topology ranged from 0.02–0.38 ($\bar{x} = 0.17$), and support for the third alternative topology ranged from 0.01–0.39 ($\bar{x} = 0.17$). In both species trees, Cinchonoideae is inferred as paraphyletic with respect to Ixoroideae and Hamelieae is inferred as paraphyletic with respect to Hillieae.

DISCUSSION

Despite the growing importance of target enrichment as a strategy to estimate phylogenies across many different plant groups, the method has rarely been used in Rubiaceae. To date, most published target enrichment data from the family were obtained using universal probe sets (Antonelli et al., 2021; Thureborn et al., 2022) or with microarray technology (Prata et al., 2018), and Rubiaceae lacked a

publicly accessible set of family-specific probes for target enrichment. Filling this gap, we used publicly accessible genomic resources for various Rubioideae and Cinchonoideae taxa to isolate a set of 2270 ERs for target enrichment with phylogenetic utility at various scales across Rubiaceae. We tested the performance of these probes, called Rubiaceae2270x, via *in vitro* target enrichment in two distantly related clades—Hillieae and Palicoureeae+Psychotrieae—and with *in silico* simulations for taxa outside of our clades of interest. In general, performance of this probe set was high in both empirical data generation and simulated target enrichment in taxa across the family.

Although low-depth whole genome sequences may not yield high coverage across single-copy nuclear loci, by mapping genome skimming reads to transcriptomes (the basis of our target selection strategy) we were able to identify thousands of exonic regions, useful for phylogenetic inference. Our Rubiaceae2270x probe set performed well in targeting loci from species across Rubiaceae, despite probe sequences being entirely derived from Rubioideae and Cinchonoideae. There was little overlap between the loci identified from these two subfamilies: less than 1% of the CIN and RUB final ERs mapped to the same genes. Despite this, probes designed from one clade's genomic resources were efficient at targeting loci in other clades. Illustrating this, although ca. 70% of the probe sequences originated from a single clade (Cinchonoideae), an average of 1733 SERs (76.3%) were recovered with at least 75% of the target sequence for sequenced Rubioideae taxa in Palicoureeae+Psychotrieae (Table 2). Similarly, simulated target enrichment was high for Rubiaceae outside these two clades,

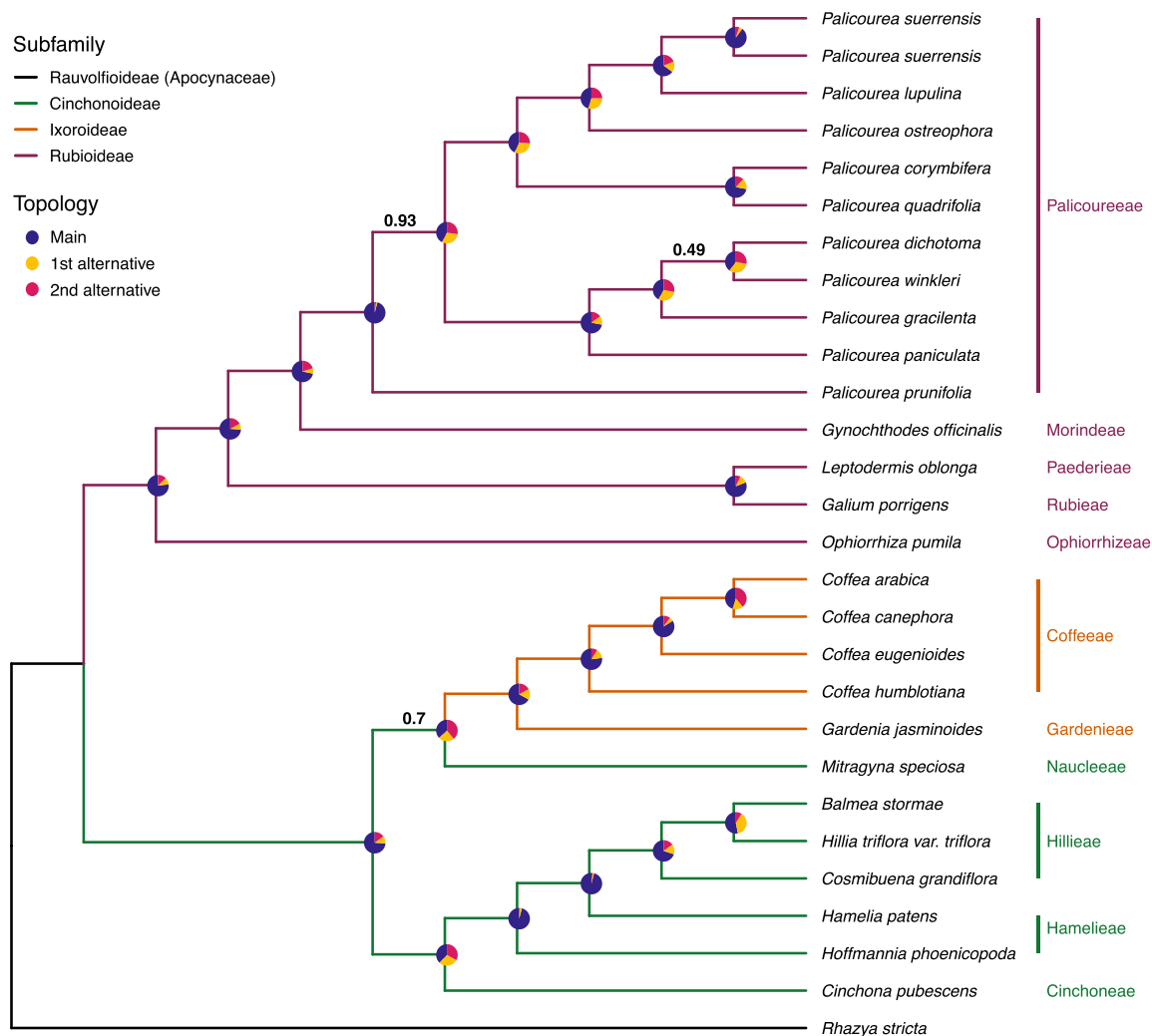


FIGURE 3 ASTRAL species tree estimated using the single-exonic region data set. Numbers above branches indicate local posterior probability support values; only values <0.98 are shown. Pie charts at internal nodes indicate quartet support (i.e., the percentage of quartets in gene trees that agree with the branch) for the main topology (blue), the first alternate topology (yellow), and the second alternate topology (pink). The two *Palicourea suerrensis* samples included originate from different populations.

resulting in 668–1168 SERs with at least 75% of the target sequence; simulated capture even resulted in 345 SERs for *Rhazya stricta* (Apocynaceae), the outgroup taxon (Figure 2). However, probes did tend to perform better (at least in terms of percentage of the locus length recovered) when the target species were from the same clade from which probe sequences were derived. Overall, we demonstrate the utility of Rubiaceae2270x by generating dense empirical data sets for Cinchonoideae and Palicoureeae+Psychotriaceae using a high proportion of herbarium specimens, while success of in silico simulations suggests that this locus data set can be successfully captured across Rubiaceae.

The loci isolated using Rubiaceae2270x have properties that are desirable for phylogenomics, including high variability (Figure 2), long average length (Table 3), and a relatively low proportion of paralogs (the greatest average proportion of potential paralogs observed in a clade was

27% [MERs] for Hillieae). Comparing our Palicoureeae+Psychotriaceae (Rubiaceae) data set to Thureborn et al. (2022), which used Angiosperms353 to sequence taxa across Rubioideae, Rubiaceae2270x resulted in a similar proportion of total loci with paralog warnings (55% in our study vs. 53% in Thureborn et al., 2022). Here, these percentages refer to loci for which >1 contig covering a specified percentage of the length of the reference sequence was assembled by HybPiper (our study uses a 75% cutoff and Thureborn et al. uses a less conservative cutoff of 85%). While informative for assessing the number of potentially paralogous genes present in a data set, this metric does not distinguish allelic variation from potential paralogs. Our probe set will facilitate future work needed to determine the nature of the loci that receive this paralog warning.

Both family-wide and within-clade alignment sets showed sufficient sequence divergence and informativeness for phylogenetic analysis, with an average proportion of

parsimony informative sites varying from 6% in Hillieae to nearly 20% across Rubiaceae (Figure 2A). Palicoureeae +Psychotriaceae alignments tended to be more variable than Hillieae alignments (Figure 2), a pattern that is expected when sequences for more species are included. An increase in the proportion of informative sites across alignments may also be explained by the relative ages of the clades in this study—18.7 Ma for Hillieae vs. 63.0 Ma for Palicoureeae +Psychotriaceae (Bremer and Eriksson, 2009). By representing an older clade, Palicoureeae+Psychotriaceae may have accumulated more substitutions through time. The variation present in these loci suggests that Rubiaceae2270x has utility across phylogenetic scales within Rubiaceae.

In the ASTRAL analyses of a subset of our sampled Rubiaceae, most nodes were recovered with high support ($LPP \geq 0.98$) in both species trees, including relationships between closely related *Palicourea* species, despite only including the small proportion of loci with no paralog warnings (14%) in the analyses. Relationships between tribes and subfamilies were mostly consistent with previous studies (Bremer and Eriksson, 2009; Wikström et al., 2015). One major distinction is that our results reject the monophyly of subfamily Cinchonoideae. In our SER (Figure 3) and MER (Appendix S9) species trees, *Mitragyna speciosa* (Cinchonoideae) was recovered as sister to Ixoroideae, although placement of this species was not strongly supported ($LPP = 0.70$ in the SER tree and 0.75 in the MER tree) and a second alternative topology is supported by a large proportion of loci. While *Cinchona pubescens* was sister to Hillieae+Hamelieae in both trees, its placement was poorly supported ($LPP = 0.22$) in the MER tree. Other recent phylogenetic studies of Rubiaceae and Gentianales more broadly have also shown Cinchonoideae to be paraphyletic with respect to Ixoroideae (Rydin et al., 2017; Wikström et al., 2020; Antonelli et al., 2021). However, these studies also demonstrate that tree topology and support for relationships can vary depending on the nature of the data as well as the methods being used for phylogenetic inference (e.g., plastid vs. nuclear genomic data, or concatenation-based vs. coalescent-aware phylogenetic inference). Other lower-order phylogenetic relationships inferred (e.g., relationships within Hillieae and Rubioideae) were more consistent with our expectations based on previous studies (Bremer and Eriksson, 2009; Sedio et al., 2013; Wikström et al., 2015; Razafimandimbison et al., 2017; Thureborn et al., 2022). These phylogenetic results provide additional support for the efficacy of our probe set in recovering informative loci in taxa from across Rubiaceae, and for their applicability in both shallow- and deep-level phylogenomic studies. By providing thousands of variable loci specific to Rubiaceae, our probe set could facilitate the study of sources of gene tree discordance (i.e., incomplete lineage sorting, gene flow, whole-genome duplication), identify nodes with higher conflict, and infer non-bifurcating relationships in Rubiaceae.

Rubiaceae2270x has very little overlap with the Angiosperms353 loci, with fewer than 1% of loci mapping to Angiosperms353 for more than 120 bp in length. Therefore,

our data set can be combined with Angiosperms353, or other universal locus data sets, in future target enrichment-based studies. Angiosperms353 has already been used successfully to infer highly supported phylogenetic relationships in the Rubioideae clade (Thureborn et al., 2022). Combining universal probe set data and lineage-specific data is a promising future direction for resolving recalcitrant relationships in Rubiaceae (particularly at shallow nodes). Studies in other families have already successfully adopted this approach (Chau et al., 2018; Hendriks et al., 2021; Shah et al., 2021; Siniscalchi et al., 2021; Ufimov et al., 2021). Combining probe sets allows for the selection of loci that have the most desirable properties for phylogenetic inference within the taxon of interest (Baker et al., 2022; Yardeni et al., 2022) and has the added benefit of drawing from and adding to a large, publicly available data set of shared taxa across plant groups. The Rubiaceae2270x target enrichment probe set is a tool for researchers to generate hundreds of single-copy loci for phylogenomic inference in Rubiaceae. Similar resources have been developed for other species-rich angiosperm families, including Asteraceae (Mandel et al., 2014), Melastomataceae (Jantzen et al., 2020), Orchidaceae (Eserman et al., 2021), and Bromeliaceae (Yardeni et al., 2022). We hope that Rubiaceae2270x will facilitate an increase in the amount of genome-scale data available for this large and ecologically important plant family, leading to improved resolution of relationships among major clades of Rubiaceae, which remain to be fully understood (Bremer and Eriksson, 2009).

AUTHOR CONTRIBUTIONS

L.D.B. designed the probe set, performed all Hillieae DNA extractions, and wrote the first draft of this manuscript. L.P.L. designed the sampling strategy for the target clades and funded the project. L.D.B. and A.M.B. performed bioinformatic analyses. C.M.T. validated the taxonomic status of all studied specimens and designed the sampling strategy for Rubioideae. All authors collaborated on writing, editing, and approving the final version of the document ahead of submission.

ACKNOWLEDGMENTS

Financial support was provided by start-up funds from the Louisiana State University (LSU) College of Science, LSU Office of Research and Economic Development, and National Science Foundation (NSF) Award #2055525 to L.P.L., and graduate student grants from the Botanical Society of America, Garden Club of America, American Society of Plant Taxonomists, and Society of Systematic Biologists to L.D.B. L.D.B. was further supported by an NSF Graduate Research Fellowship. Feedback from Janet Mansaray, Diego Paredes-Burneo, Katherin Arango-Gómez, Aislinn Mumford, and Laura Frost significantly improved previous drafts of this manuscript. We thank LSU and the Missouri Botanical Garden for providing access to their herbaria, and for granting permission to destructively sample specimens.

DATA AVAILABILITY STATEMENT

The Rubiaceae2270x sequences are available on Figshare (DOI: 10.6084/m9.figshare.22776512; Li, 2018). Raw reads from the in vitro target enrichment study are available on the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (BioProject ID PRJNA970862).

ORCID

Laymon D. Ball  <http://orcid.org/0000-0001-6427-6274>

Ana M. Bedoya  <https://orcid.org/0000-0002-6245-3481>

Laura P. Lagomarsino  <http://orcid.org/0000-0003-4537-0761>

REFERENCES

- Amenu, S. G., N. Wei, L. Wu, O. Oyebanji, G. Hu, Y. Zhou, and Q. Wang. 2022. Phylogenomic and comparative analyses of Coffeae alliance (Rubiaceae): Deep insights into phylogenetic relationships and plastome evolution. *BMC Plant Biology* 22(1): 88.
- Andermann, T., M. F. Torres Jiménez, P. Matos-Maraví, R. Batista, J. L. Blanco-Pastor, A. L. S. Gustafsson, L. Kistler, et al. 2019. A guide to carrying out a phylogenomic target sequence capture project. *Frontiers in Genetics* 10: 1407.
- Antonelli, A., J. A. A. Nylander, C. Persson, and I. Sanmartín. 2009. Tracing the impact of the Andean uplift on Neotropical plant evolution. *Proceedings of the National Academy of Sciences, USA* 106(24): 9749–9754.
- Antonelli, A., J. J. Clarkson, K. Kainulainen, O. Maurin, G. E. Brewer, A. P. Davis, N. Epiawalage, et al. 2021. Settling a family feud: A high-level phylogenomic framework for the Gentianales based on 353 nuclear genes and partial plastomes. *American Journal of Botany* 108(7): 1143–1165.
- Bagley, J. C., S. Uribe-Convers, M. M. Carlsen, and N. Muchhala. 2020. Utility of targeted sequence capture for phylogenomics in rapid, recent angiosperm radiations: Neotropical *Burmeistera* bellflowers as a case study. *Molecular Phylogenetics and Evolution* 152(November): 106769.
- Baker, W. J., P. Bailey, V. Barber, and A. Barker. 2022. A comprehensive phylogenomic platform for exploring the angiosperm tree of life. *Systematic Biology* 71(2): 301–319. <https://academic.oup.com/sysbio/article-abstract/71/2/301/6275244>
- Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15): 2114–2120.
- 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(3): 458–469.
- Borowiec, M. L. 2016. AMAS: A fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4(January): e1660.
- Breinholt, J. W., S. B. Carey, G. P. Tiley, E. C. Davis, L. Endara, S. F. McDaniel, L. G. Neves, et al. 2021. A target enrichment probe set for resolving the flagellate land plant tree of life. *Applications in Plant Sciences* 9(1): e11406.
- Bremer, B., and O. Eriksson. 1992. Evolution of fruit characters and dispersal modes in the tropical family Rubiaceae. *Biological Journal of the Linnean Society* 47(1): 79–95.
- Bremer, B., and T. Eriksson. 2009. Time tree of Rubiaceae: Phylogeny and dating the family, subfamilies, and tribes. *International Journal of Plant Sciences* 170(6): 766–793.
- Brown, J. W., J. F. Walker, and S. A. Smith. 2017. Phyx: Phylogenetic tools for Unix. *Bioinformatics* 33(12): 1886–1888.
- Canales, N. A., E. M. Gardner, T. Gress, K. Walker, V. Bieker, M. D. Martin, M. Nesbitt, et al. 2022. Museomic approaches to genotype historic *Cinchona* barks. *BioRxiv* [Preprint]. Available at: <https://doi.org/10.1101/2022.04.26.489609> [posted 28 April 2022; accessed 5 October 2023].
- Cao, M. D., D. Ganesamoorthy, C. Zhou, and L. J. M. Coin. 2018. Simulating the dynamics of targeted capture sequencing with CapSim. *Bioinformatics* 34(5): 873–874.
- Capella-Gutiérrez, S., J. M. Silla-Martínez, and T. Gabaldón. 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15): 1972–1973.
- Chamala, S., N. García, G. T. Godden, V. Krishnakumar, I. E. Jordon-Thaden, R. De Smet, W. B. Barbazuk, et al. 2015. MarkerMiner 1.0: A new application for phylogenetic marker development using angiosperm transcriptomes. *Applications in Plant Sciences* 3(4): e1400115. <https://doi.org/10.3732/apps.1400115>
- Chau, J. H., W. A. Rahfeldt, and R. G. Olmstead. 2018. Comparison of taxon-specific versus general locus sets for targeted sequence capture in plant phylogenomics. *Applications in Plant Sciences* 6(3): e1032.
- Ehrendorfer, F., M. H. J. Barfuss, J.-F. Manen, and G. M. Schneeweiss. 2019. Correction: Phylogeny, character evolution and spatiotemporal diversification of the species-rich and world-wide distributed tribe Rubieae (Rubiaceae). *PLoS ONE* 14(1): e0211589.
- Eserman, L. A., S. K. Thomas, E. E. D. Coffey, and J. H. Leebens-Mack. 2021. Target sequence capture in orchids: Developing a kit to sequence hundreds of single-copy loci. *Applications in Plant Sciences* 9(7): e11416.
- Ferrero, V., D. Rojas, A. Vale, and L. Navarro. 2012. Delving into the loss of heterostyly in Rubiaceae: Is there a similar trend in tropical and non-tropical climate zones? *Perspectives in Plant Ecology, Evolution and Systematics* 14(3): 161–167.
- Frost, L., A. M. Bedoya, and L. Lagomarsino. 2022. Strong phylogenetic signal despite high phylogenomic complexity in an Andean plant radiation (*Freziera*, Pentaphragaceae). *BioRxiv* [Preprint]. Available at: <https://doi.org/10.1101/2021.07.01.450750> [posted 12 September 2023; accessed 5 October 2023].
- Hart, M. L., L. L. Forrest, J. A. Nicholls, and C. A. Kidner. 2016. Retrieval of hundreds of nuclear loci from herbarium specimens. *Taxon* 65(5): 1081–1092.
- Hendriks, K. P., T. Mandáková, N. M. Hay, E. Ly, A. Hooft van Huysduynen, R. Tamrakar, S. K. Thomas, et al. 2021. The best of both worlds: Combining lineage-specific and universal bait sets in target-enrichment hybridization reactions. *Applications in Plant Sciences* 9(7): e11438. <https://doi.org/10.1002/aps3.11438>
- Jantzen, J. R., P. Amarasinghe, R. A. Folk, M. Reginato, F. A. Michelangeli, D. E. Soltis, N. Cellinese, and P. S. Soltis. 2020. A two-tier bioinformatic pipeline to develop probes for target capture of nuclear loci with applications in Melastomataceae. *Applications in Plant Sciences* 8(5): e11345.
- Johnson, M. G., E. M. Gardner, Y. Liu, R. Medina, B. Goffinet, A. J. Shaw, N. J. C. Zerega, and N. J. Wickett. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences* 4(7): e160016.
- 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(4): 594–606.
- Junier, T., and E. M. Zdobnov. 2010. The Newick utilities: High-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics* 26(13): 1669–1670.
- Kadlec, M., D. U. Bellstedt, N. C. Le Maitre, and M. D. Pirie. 2017. Targeted NGS for species level phylogenomics: ‘Made to measure’ or ‘one size fits all?’ *PeerJ* 5(July): e3569.
- Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30(4): 772–780.

- Kozlov, A. M., D. Darriba, T. Flouri, B. Morel, and A. Stamatakis. 2019. RAXML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35(21): 4453–4455.
- Le, H. T. T., L. N. Nguyen, H. L. B. Pham, H. T. M. Le, T. D. Luong, H. T. T. Huynh, V. T. Nguyen, et al. 2022. Target capture reveals the complex origin of Vietnamese ginseng. *Frontiers in Plant Science* 13(July): 814178.
- Li, W. 2018. Figshare dataset. Website: <https://doi.org/10.6084/m9.figshare.6025748.v1> [accessed 6 October 2023].
- Li, W., and A. Godzik. 2006. Cd-Hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13): 1658–1659.
- Liu, Y., M. G. Johnson, C. J. Cox, R. Medina, N. Devos, A. Vanderpoorten, L. Hedenäs, et al. 2019. Resolution of the ordinal phylogeny of mosses using targeted exons from organellar and nuclear genomes. *Nature Communications* 10(1): 1485.
- Löfstrand, S. D., S. G. Razafimandimbison, and C. Rydin. 2019. Phylogeny of Coussareeae (Rubiaceae, Rubiaceae). *Plant Systematics and Evolution* 305(4): 293–304.
- Mandel, J. R., R. B. Dikow, V. A. Funk, R. R. Masalia, S. E. Staton, A. Kozik, R. W. Michelmore, et al. 2014. A target enrichment method for gathering phylogenetic information from hundreds of loci: An example from the Compositae. *Applications in Plant Sciences* 2(2): e1300085. <https://doi.org/10.3732/apps.1300085>
- Matasci, N., L.-H. Hung, Z. Yan, E. J. Carpenter, N. J. Wickett, S. Mirarab, N. Nguyen, et al. 2014. Data access for the 1,000 Plants (1KP) Project. *GigaScience* 3(October): 17.
- Morales-Briones, D. F., B. Gehrke, and C. H. Huang. 2022. Analysis of paralogs in target enrichment data pinpoints multiple ancient polyploidy events in *Alchemilla* s.l. (Rosaceae). *Systematic Biology* 71(1): 190–207. <https://academic.oup.com/sysbio/article-abstract/71/1/190/6274658>
- Nikolov, L. A., P. Shushkov, B. Nevado, X. Gan, I. A. Al-Shehbaz, D. Filatov, C. Do. Bailey, and M. Tsiantis. 2019. Resolving the backbone of the Brassicaceae phylogeny for investigating trait diversity. *New Phytologist* 222(3): 1638–1651.
- Plants of the World Online. 2023. Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. Website: <https://powo.science.kew.org/> [accessed 4 February 2023].
- Prata, E. M. B., C. Sass, D. P. Rodrigues, F. M. C. B. Domingos, C. D. Specht, G. Damasco, C. C. Ribas, et al. 2018. Towards integrative taxonomy in Neotropical botany: Disentangling the *Pagamea guianensis* species complex (Rubiaceae). *Botanical Journal of the Linnean Society* 188(2): 213–231.
- Razafimandimbison, S. G., C. M. Taylor, N. Wikström, T. Paillet, A. Khodabandeh, and B. Bremer. 2014. Phylogeny and generic limits in the sister tribes Psychotrieae and Palicoureeae (Rubiaceae): Evolution of schizocarps in *Psychotria* and origins of bacterial leaf nodules of the Malagasy species. *American Journal of Botany* 101(7): 1102–1126.
- Razafimandimbison, S. G., K. Kainulainen, N. Wikström, and B. Bremer. 2017. Historical biogeography and phylogeny of the pantropical Psychotrieae alliance (Rubiaceae), with particular emphasis on the Western Indian Ocean region. *American Journal of Botany* 104(9): 1407–1423.
- Razafimandimbison, S. G., K.-M. Wong, and C. Rydin. 2021. Molecular systematics of the tribe Pristmatomerideae (Rubiaceae) and its taxonomic consequences, with notes on the importance of the inflorescence morphology for species-group recognition in *Rennellia*. *Taxon* 70(2): 324–338.
- Rydin, C., N. Wikström, and B. Bremer. 2017. Conflicting results from mitochondrial genomic data challenge current views of Rubiaceae phylogeny. *American Journal of Botany* 104(10): 1522–1532.
- Schmickl, R., A. Liston, V. Zeisek, K. Oberlander, K. Weitemier, S. C. K. Straub, R. C. Cronn, et al. 2016. Phylogenetic marker development for target enrichment from transcriptome and genome skim data: The pipeline and its application in Southern African *Oxalis* (Oxalidaceae). *Molecular Ecology Resources* 16(5): 1124–1135.
- Sedio, B. E., J. R. Paul, C. M. Taylor, and C. W. Dick. 2013. Fine-scale niche structure of Neotropical forests reflects a legacy of the great American biotic interchange. *Nature Communications* 4: 2317.
- Shah, T., J. V. Schneider, G. Zizka, O. Maurin, W. Baker, F. Forest, G. E. Brewer, et al. 2021. Joining forces in Ochnaceae phylogenomics: A tale of two targeted sequencing probe kits. *American Journal of Botany* 108(7): 1201–1216.
- Siniscalchi, C. M., O. Hidalgo, L. Palazzesi, J. Pellicer, L. Pokorny, O. Maurin, I. J. Leitch, et al. 2021. Lineage-specific vs. universal: A comparison of the Compositae1061 and Angiosperms353 enrichment panels in the sunflower family. *Applications in Plant Sciences* 9(7): e11422. <https://doi.org/10.1002/aps3.11422>
- Slater, G., and E. Birney. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6: 31.
- Slimp, M., L. D. Williams, H. Hale, and M. G. Johnson. 2021. On the potential of Angiosperms353 for population genomic studies. *Applications in Plant Sciences* 9(7): e11419. <https://doi.org/10.1002/aps3.11419>
- Smith, B. T., M. G. Harvey, B. C. Faircloth, T. C. Glenn, and R. T. Brumfield. 2014. Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Systematic Biology* 63(1): 83–95.
- Soto Gomez, M., L. Pokorny, M. B. Kantar, F. Forest, I. J. Leitch, B. Gravendeel, P. Wilkin, et al. 2019. A customized nuclear target enrichment approach for developing a phylogenomic baseline for *Dioscorea* yams (Dioscoreaceae). *Applications in Plant Sciences* 7(6): e11254.
- Straub, S. C. K., J. Boutte, M. Fishbein, and T. Livshultz. 2020. Enabling evolutionary studies at multiple scales in Apocynaceae through Hyb-Seq. *Applications in Plant Sciences* 8(11): e11400.
- Štorchová, H., R. Hrdličková, J. Chrtek Jr., M. Tetera, D. Fitze, and J. Fehrer. 2020. An improved method of DNA isolation from plants collected in the field and conserved in saturated NaCl/CTAB solution. *Taxon* 49(1): 79–84.
- Thiers, B. 2023. (continuously updated). Index Herbariorum. Website <http://sweetgum.nybg.org/science/ih/> [accessed 6 October 2023].
- Thureborn, O., S. G. Razafimandimbison, N. Wikström, and C. Rydin. 2022. Target capture data resolve recalcitrant relationships in the coffee family (Rubiaceae, Rubiaceae). *Frontiers in Plant Science* 13(September): 967456.
- Ufimov, R., V. Zeisek, S. Píšová, W. J. Baker, T. Fér, M. van Loo, C. Dobeš, and R. Schmickl. 2021. Relative performance of customized and universal probe sets in target enrichment: A case study in subtribe Malinae. *Applications in Plant Sciences* 9(7): e11442.
- Uribe-Converse, S. 2016. Capture probe design with Sondovac. Simon's Lab Book. Website: <https://uribeconvers.wordpress.com/> [posted 11 May 2016; accessed 14 September 2020].
- Wikström, N., K. Kainulainen, S. G. Razafimandimbison, J. E. E. Smedmark, and B. Bremer. 2015. A revised time tree of the asterids: Establishing a temporal framework for evolutionary studies of the coffee family (Rubiaceae). *PLoS ONE* 10(5): e0126690.
- Wikström, N., B. Bremer, and C. Rydin. 2020. Conflicting phylogenetic signals in genomic data of the coffee family (Rubiaceae). *Journal of Systematics and Evolution* 58(4): 440–460.
- Yardeni, G., J. Viruel, M. Paris, J. Hess, C. Groot Crego, M. de La Harpe, N. Rivera, et al. 2022. Taxon-specific or universal? Using target capture to study the evolutionary history of rapid radiations. *Molecular Ecology Resources* 22(3): 927–945.
- 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(Suppl 6): 153.
- Zhang, C., Y. Zhao, E. L. Braun, and S. Mirarab. 2021. TAPER: Pinpointing errors in multiple sequence alignments despite varying rates of evolution. *Methods in Ecology and Evolution* 12(11): 2145–2158.

SUPPORTING INFORMATION

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

Appendix S1. Exonic region concatenation protocol.

Appendix S2. HybPiper summary statistics for single-exonic region assemblies. Grayed rows indicate CapSim-derived sequencing data. Bolded rows indicate silica-dried specimens.

Appendix S3. HybPiper summary statistics for multi-exonic region assemblies. Grayed rows indicate CapSim-derived sequencing data. Bolded rows indicate silica-dried specimens.

Appendix S4. Results of initial CapSim, simulated target enrichment experiment. Contigs is the number of contigs recovered during the map to reference assembly.

Appendix S5. Statistics for downloaded genomes used in CapSim experiments. “Total length” (Mb = megabases) is the length of all contigs combined in the original genome assembly. “N50” is the scaffold N50 score provided with the NCBI accession for the genome (see Appendix 2), used here as a metric for genome quality.

Appendix S6. HybPiper single-exonic region and multi-exonic region (MER) assembly statistics averaged for each clade. Statistics for MER assemblies are in blue.

Appendix S7. Heatmap showing results of HybPiper multi-exonic region (MER) assemblies for Cinchonoideae (CIN), Rubioideae (RUB), and Ixoroideae (IXOR) samples, and one Apocynaceae sample. Target MERs are on the *x*-axis and are grouped by the taxonomic source of the reference sequence.

“RUB+CIN” indicates the 22 MERs with constituent ERs from both RUB and CIN. Samples are on the *y*-axis and are grouped by taxonomy. Darker shades represent higher percentages of the locus length recovered by HybPiper. Locus completeness tends to be higher when target species are from the same clade as the probe sequences.

Appendix S8. AMAS single-exonic region alignment statistics, showing average values for each statistic with ranges in parentheses. “All combined” alignments include Hillieae, Palicoureeae+Psychotriaceae, and all outgroups (including CapSim samples).

Appendix S9. ASTRAL species tree estimated using the multi-exonic region data set. Numbers above branches indicate local posterior probability support values; only values <0.98 are shown. Pie charts at internal nodes indicate quartet support (i.e., the percentage of quartets in gene trees that agree with the branch) for the main topology (blue), the first alternate topology (yellow), and the second alternate topology (pink). The two *Palicourea suerrensis* samples included originate from different populations.

How to cite this article: Ball, L. D., A. M. Bedoya, C. M. Taylor, and L. P. Lagomarsino. 2023. A target enrichment probe set for resolving phylogenetic relationships in the coffee family, Rubiaceae. *Applications in Plant Sciences* 11(6): e11554. <https://doi.org/10.1002/aps3.11554>

APPENDICES

Appendix 1. Voucher information for sampled specimens that underwent sequencing.

| Species | Tribe | Voucher | Institution | Tissue type |
|--|-------------------|--------------------------|-------------|-------------|
| <i>Coutarea hexandra</i> (Jacq.) K. Schum. | Chiococceae | Callejas 4587 | LSU | Herbarium |
| <i>Ferdinandusa paraensis</i> Ducke | Dialypetalantheae | Mori 20866 | LSU | Herbarium |
| <i>Hamelia patens</i> Jacq. | Hamelieae | Lagomarsino s.n. | LSU | Silica |
| <i>Hoffmannia phoenicopoda</i> K. Schum. | Hamelieae | Wendt 3392 | LSU | Herbarium |
| <i>Balmea stormiae</i> Martínez | Hillieae | Vázquez 1081 | MO | Herbarium |
| <i>Cosmibuena grandiflora</i> (Ruiz & Pav.) Rusby | Hillieae | Haber 409 | MO | Herbarium |
| <i>Cosmibuena grandiflora</i> (Ruiz & Pav.) Rusby* | Hillieae | Pipoly 10310 | LSU | Herbarium |
| <i>Cosmibuena grandiflora</i> (Ruiz & Pav.) Rusby | Hillieae | Tyson 904 | MO | Herbarium |
| <i>Cosmibuena macrocarpa</i> (Benth.) Klotzsch ex Walp. | Hillieae | Silverstone-Sopkin 10729 | MO | Herbarium |
| <i>Cosmibuena matudae</i> (Standl.) L. O. Williams | Hillieae | Moreno 9550 | MO | Herbarium |
| <i>Cosmibuena valerii</i> Standl. | Hillieae | Haber 548 | LSU | Herbarium |
| <i>Hillia allenii</i> C. M. Taylor | Hillieae | McPherson 11676 | MO | Herbarium |
| <i>Hillia bonoi</i> Steyererm. | Hillieae | Burandt V0007 | MO | Herbarium |
| <i>Hillia chiapensis</i> Standl. | Hillieae | Stevens 11558 | LSU | Herbarium |
| <i>Hillia foldatsii</i> Steyererm. | Hillieae | Holst 3772 | MO | Herbarium |
| <i>Hillia grayumii</i> C. M. Taylor | Hillieae | Stevens 24958 | MO | Herbarium |
| <i>Hillia illustris</i> (Vell.) K. Schum. | Hillieae | Rimachi Y. 9155 | LSU | Herbarium |
| <i>Hillia illustris</i> (Vell.) K. Schum. | Hillieae | Solomon 13957 | LSU | Herbarium |
| <i>Hillia illustris</i> (Vell.) K. Schum. | Hillieae | Fuentes 4120 | MO | Herbarium |
| <i>Hillia illustris</i> (Vell.) K. Schum. | Hillieae | Vázquez 4130 | MO | Herbarium |
| <i>Hillia killipii</i> Standl. | Hillieae | Valenzuela 7574 | MO | Herbarium |
| <i>Hillia killipii</i> Standl. | Hillieae | van der Werff 25095 | MO | Herbarium |
| <i>Hillia longifilamentosa</i> (Steyererm.) C. M. Taylor | Hillieae | Gamboia R. 2206 | MO | Herbarium |
| <i>Hillia loranthoides</i> Standl. | Hillieae | Bello 844 | MO | Herbarium |
| <i>Hillia macbridei</i> Standl. | Hillieae | Zak 3788 | MO | Herbarium |
| <i>Hillia macrophylla</i> Standl. | Hillieae | Bello 767 | MO | Herbarium |
| <i>Hillia macrophylla</i> Standl. | Hillieae | Cornejo 8056 | MO | Herbarium |
| <i>Hillia macrophylla</i> Standl. | Hillieae | Rojas 9037 | MO | Herbarium |
| <i>Hillia macrophylla</i> Standl. | Hillieae | Werner 2171 | MO | Herbarium |
| <i>Hillia maxonii</i> Standl.* | Hillieae | Dwyer 7377 | MO | Herbarium |
| <i>Hillia maxonii</i> Standl. | Hillieae | Luteyn 12690 | MO | Herbarium |
| <i>Hillia maxonii</i> Standl. | Hillieae | Morales 4855 | MO | Herbarium |
| <i>Hillia palmana</i> Standl. | Hillieae | Hammel 13930 | MO | Herbarium |
| <i>Hillia parasitica</i> Jacq. | Hillieae | Jiménez 2189 | MO | Herbarium |
| <i>Hillia parasitica</i> Jacq. | Hillieae | Prance 29385 | LSU | Herbarium |
| <i>Hillia parasitica</i> Jacq.* | Hillieae | Veley 1362 | LSU | Herbarium |
| <i>Hillia parasitica</i> Jacq. | Hillieae | Zarucchi 5659 | MO | Herbarium |

| Species | Tribe | Voucher | Institution | Tissue type |
|--|--------------|------------------|-------------|-------------|
| <i>Hillia parasitica</i> Jacq. | Hillieae | Kvist192 | LSU | Herbarium |
| <i>Hillia pumila</i> C. M. Taylor | Hillieae | Vásquez 28189 | MO | Herbarium |
| <i>Hillia tetrandra</i> Sw. | Hillieae | Martínez 20613 | MO | Herbarium |
| <i>Hillia triflora</i> (Oerst.) C. M. Taylor var. <i>triflora</i> | Hillieae | Bello 5314 | MO | Herbarium |
| <i>Hillia triflora</i> (Oerst.) C. M. Taylor var. <i>triflora</i> | Hillieae | Feinsinger 626-A | MO | Herbarium |
| <i>Hillia triflora</i> var. <i>pittieri</i> (Standl.) C. M. Taylor | Hillieae | Croat 14341 | MO | Herbarium |
| <i>Hillia ulei</i> K. Schum. ex Ule | Hillieae | Foster 9667 | LSU | Herbarium |
| <i>Hillia ulei</i> K. Schum. ex Ule | Hillieae | Revilla 576 | MO | Herbarium |
| <i>Hillia wurdackii</i> Steyerl. | Hillieae | Monteagudo 15255 | MO | Herbarium |
| <i>Hillia wurdackii</i> Steyerl. | Hillieae | Valenzuela 10940 | MO | Herbarium |
| <i>Hillia wurdackii</i> Steyerl. | Hillieae | Woytkowski 8236 | MO | Herbarium |
| <i>Carapichea guianensis</i> Aubl. | Palicoureeae | González 2158 | MO | Herbarium |
| <i>Carapichea ipecacuanha</i> (Brot.) L. Andersson | Palicoureeae | Croat 15117 | MO | Herbarium |
| <i>Eumachia boliviana</i> (Standl.) Delprete & J. H. Kirkbr. | Palicoureeae | Campbell 22035 | MO | Herbarium |
| <i>Notopleura epiphytica</i> (K. Krause) C. M. Taylor | Palicoureeae | Neill 15737 | MO | Silica |
| <i>Notopleura uliginosa</i> (Sw.) Bremek. | Palicoureeae | Stevens 37138 | MO | Silica |
| <i>Palicourea acanthacea</i> (Standl. ex Steyerl.) C. M. Taylor | Palicoureeae | Monsalve 352 | MO | Herbarium |
| <i>Palicourea acuminata</i> (Benth.) Borhidi | Palicoureeae | Fonnegra 6968 | MO | Herbarium |
| <i>Palicourea acuminata</i> (Benth.) Borhidi | Palicoureeae | Lachenaud 966 | MO | Herbarium |
| <i>Palicourea acuminata</i> (Benth.) Borhidi | Palicoureeae | Suazo 4616 | MO | Herbarium |
| <i>Palicourea allenii</i> (Standl.) Borhidi | Palicoureeae | Clark 243 | MO | Herbarium |
| <i>Palicourea amethystina</i> (Ruiz & Pav.) DC. | Palicoureeae | Jaramillo 1989 | MO | Herbarium |
| <i>Palicourea andina</i> C. M. Taylor | Palicoureeae | Dziedziuch 149 | MO | Herbarium |
| <i>Palicourea angustifolia</i> Kunth | Palicoureeae | Wolff 16 | MO | Herbarium |
| <i>Palicourea apicata</i> Kunth | Palicoureeae | Stergios 2543 | MO | Herbarium |
| <i>Palicourea apoda</i> (Steyerl.) Delprete & J. H. Kirkbr.* | Palicoureeae | Pipoly 8373 | MO | Herbarium |
| <i>Palicourea attenuata</i> Rusby | Palicoureeae | Fuentes 4647 | MO | Silica |
| <i>Palicourea bangii</i> (Rusby) C. M. Taylor | Palicoureeae | Fuentes 12864 | MO | Silica |
| <i>Palicourea berteriana</i> (DC.) Borhidi | Palicoureeae | Taylor 11646 | MO | Silica |
| <i>Palicourea brachiata</i> (Sw.) Borhidi | Palicoureeae | Davidse 36906 | MO | Silica |
| <i>Palicourea brachiata</i> (Sw.) Borhidi | Palicoureeae | Taylor 11718 | MO | Silica |
| <i>Palicourea brevicollis</i> (Müll. Arg.) C. M. Taylor | Palicoureeae | Zardini 15831 | MO | Herbarium |
| <i>Palicourea callithrix</i> (Miq.) Delprete & J. H. Kirkbr. | Palicoureeae | Granville 13399 | MO | Herbarium |
| <i>Palicourea colorata</i> (Benth.) Borhidi | Palicoureeae | Liesner s.n. | MO | Herbarium |
| <i>Palicourea conephoroides</i> (Rusby) C. M. Taylor | Palicoureeae | Quizhpe 672 | MO | Herbarium |
| <i>Palicourea correae</i> (Dwyer & M. V. Hayden) Borhidi | Palicoureeae | Dwyer 1968 | MO | Herbarium |
| <i>Palicourea correae</i> (Dwyer & M. V. Hayden) Borhidi | Palicoureeae | MacDougal 6258 | MO | Silica |
| <i>Palicourea corymbifera</i> (Müll. Arg.) Standl. | Palicoureeae | Grimes 3319 | MO | Herbarium |
| <i>Palicourea croceoides</i> (Sw.) Roem. & Schult. | Palicoureeae | Merello 1713 | MO | Silica |

(Continues)

| Species | Tribe | Voucher | Institution | Tissue type |
|---|--------------|-----------------------|-------------|-------------|
| <i>Palicourea croceoides</i> Desv. ex Ham. | Palicoureeae | Taylor 11640 | MO | Silica |
| <i>Palicourea cyanococca</i> (Dombrain) Borhidi | Palicoureeae | Stevens 30807 | MO | Silica |
| <i>Palicourea deflexa</i> (DC.) Borhidi | Palicoureeae | Meave 1172 | MO | Herbarium |
| <i>Palicourea deflexa</i> (DC.) Borhidi | Palicoureeae | Taylor 11717 | MO | Silica |
| <i>Palicourea demissa</i> Standl. | Palicoureeae | Zak 3081 | MO | Herbarium |
| <i>Palicourea dichotoma</i> (Rudge) Delprete & J. H. Kirkbr. | Palicoureeae | Maceda 1475 | MO | Herbarium |
| <i>Palicourea didymocarpos</i> (A. Rich.) Griseb. | Palicoureeae | Nee 35839 | MO | Herbarium |
| <i>Palicourea divaricata</i> Schltldl. | Palicoureeae | Carvalho 6490 | MO | Herbarium |
| <i>Palicourea domingensis</i> (Jacq.) DC. | Palicoureeae | Axelrod 1028 | MO | Herbarium |
| <i>Palicourea egensis</i> (Müll. Arg.) Borhidi | Palicoureeae | Liesner 7059 | MO | Herbarium |
| <i>Palicourea elata</i> (Sw.) Borhidi | Palicoureeae | Ibarra-Manriquez 5301 | MO | Herbarium |
| <i>Palicourea elata</i> (Sw.) Borhidi | Palicoureeae | Stevens 36104 | MO | Silica |
| <i>Palicourea flavescens</i> Kunth | Palicoureeae | van derWerff 10948 | MO | Herbarium |
| <i>Palicourea flavifolia</i> (Rusby) Standl. | Palicoureeae | Maldonado 2948 | MO | Silica |
| <i>Palicourea glomerulata</i> (Donn. Sm.) Borhidi | Palicoureeae | Berger 1481 | MO | Herbarium |
| <i>Palicourea gracilentia</i> (Müll. Arg.) Delprete & J. H. Kirkbr. | Palicoureeae | Croat 102206 | MO | Herbarium |
| <i>Palicourea grandifolia</i> (Humb. & Bonpl. ex Roem. & Schult.) Standl. | Palicoureeae | Liesner 6534 | MO | Herbarium |
| <i>Palicourea guianensis</i> Aubl. | Palicoureeae | de la Quintana 257 | MO | Silica |
| <i>Palicourea guianensis</i> Aubl. | Palicoureeae | Redden 2302 | MO | Herbarium |
| <i>Palicourea guianensis</i> Aubl. | Palicoureeae | Will 83 | MO | Herbarium |
| <i>Palicourea hazenii</i> (Standl.) Borhidi | Palicoureeae | Freire 1098 | MO | Herbarium |
| <i>Palicourea hoffmannseggiana</i> (Müll. Arg.) Delprete & J. H. Kirkbr. | Palicoureeae | Torida-Marbot 177 | MO | Herbarium |
| <i>Palicourea jelskii</i> Standl. | Palicoureeae | Fuentes 12894 | MO | Silica |
| <i>Palicourea justicifolia</i> (Rudge) Delprete & J. H. Kirkbr. | Palicoureeae | Gutiérrez 534 | MO | Herbarium |
| <i>Palicourea lasiantha</i> K. Krause | Palicoureeae | Graham 199 | MO | Herbarium |
| <i>Palicourea lasiorrhachis</i> Oerst.* | Palicoureeae | Wilbur 19554 | MO | Herbarium |
| <i>Palicourea lehmannii</i> (K. Schum. & K. Krause) Standl. | Palicoureeae | Silverstone 8396 | MO | Herbarium |
| <i>Palicourea lineata</i> Benth. | Palicoureeae | Garcia 112 | MO | Herbarium |
| <i>Palicourea loxensis</i> C. M. Taylor | Palicoureeae | Neill 16912 | MO | Silica |
| <i>Palicourea luteonivea</i> C. M. Taylor | Palicoureeae | Cayola 2534 | MO | Silica |
| <i>Palicourea macrobotrys</i> (Ruiz & Pav.) DC. | Palicoureeae | Gatti 17549 | MO | Herbarium |
| <i>Palicourea marcgravii</i> A. St.-Hil. | Palicoureeae | Vasconcelos s.n. | MO | Herbarium |
| <i>Palicourea muscosa</i> (Jacq.) Delprete & J. H. Kirkbr. | Palicoureeae | Meier 3151 | MO | Herbarium |
| <i>Palicourea nitidella</i> (Müll. Arg.) Standl. | Palicoureeae | Liesner 6369 | MO | Herbarium |
| <i>Palicourea obliquinervia</i> (Müll. Arg.) Borhidi | Palicoureeae | Clarke 1268 | MO | Herbarium |
| <i>Palicourea ostreophora</i> (Wernham) Borhidi | Palicoureeae | Schunke 8366 | MO | Herbarium |
| <i>Palicourea padifolia</i> (Humb. & Bonpl. ex Roem. & Schult.) C. M. Taylor & Lorence | Palicoureeae | Dietzsch 1390 | MO | Herbarium |
| <i>Palicourea petiolaris</i> Kunth | Palicoureeae | Ortega 3101 | MO | Herbarium |

| Species | Tribe | Voucher | Institution | Tissue type |
|--|--------------|---------------------|-------------|-------------|
| <i>Palicourea polycephala</i> (Benth.) Delprete & J. H. Kirkbr. | Palicoureeae | Ehringhaus 56 | MO | Herbarium |
| <i>Palicourea potaroensis</i> (Sandwith) Delprete & J. H. Kirkbr.* | Palicoureeae | Henkel 1670 | MO | Herbarium |
| <i>Palicourea prunifolia</i> (Kunth) Borhidi | Palicoureeae | Nee 41304 | MO | Herbarium |
| <i>Palicourea pubescens</i> (Sw.) Borhidi | Palicoureeae | Stevens 21197 | MO | Herbarium |
| <i>Palicourea pubescens</i> Sw. | Palicoureeae | Taylor 316 | MO | Herbarium |
| <i>Palicourea pyramidalis</i> Standl. | Palicoureeae | Hurtado 1015 | MO | Herbarium |
| <i>Palicourea quadrifolia</i> (Rudge) DC. | Palicoureeae | Richard 76 | MO | Herbarium |
| <i>Palicourea quadrilateralis</i> C. M. Taylor* | Palicoureeae | Callejas 4018 | MO | Herbarium |
| <i>Palicourea quinquepyrena</i> C. M. Taylor | Palicoureeae | Rodríguez 1748 | MO | Herbarium |
| <i>Palicourea racemosa</i> (Aubl.) G. Nicholson | Palicoureeae | Rivero 265 | MO | Herbarium |
| <i>Palicourea reticulata</i> (Ruiz & Pav.) C. M. Taylor | Palicoureeae | Zambrana 5775 | MO | Silica |
| <i>Palicourea rhodothamna</i> (Standl.) C. M. Taylor | Palicoureeae | Rimachi 7296 | MO | Herbarium |
| <i>Palicourea rigida</i> Kunth | Palicoureeae | Gillespie 1715 | MO | Herbarium |
| <i>Palicourea rigida</i> Kunth | Palicoureeae | Subieta 322 | MO | Herbarium |
| <i>Palicourea seemanii</i> Standl. | Palicoureeae | Juncosa 556 | MO | Herbarium |
| <i>Palicourea sessilis</i> (Vell.) C. M. Taylor | Palicoureeae | Fiaschi 2844 | MO | Herbarium |
| <i>Palicourea solitudinum</i> (Standl.) Borhidi* | Palicoureeae | Duke 5262 | MO | Herbarium |
| <i>Palicourea standleyana</i> C. M. Taylor | Palicoureeae | Gentry 63598 | MO | Herbarium |
| <i>Palicourea stenosepala</i> Standl. | Palicoureeae | Link 13 | MO | Herbarium |
| <i>Palicourea stipularis</i> Benth. | Palicoureeae | Leiva 15000 | MO | Herbarium |
| <i>Palicourea subfusca</i> (Müll. Arg.) C. M. Taylor | Palicoureeae | Parada 996 | MO | Herbarium |
| <i>Palicourea suerrensis</i> (Donn. Sm.) Borhidi | Palicoureeae | McPherson 15864 | MO | Silica |
| <i>Palicourea suerrensis</i> (Donn. Sm.) Borhidi | Palicoureeae | Stevens 36741 | MO | Silica |
| <i>Palicourea sulphurea</i> (Ruiz & Pav.) DC. | Palicoureeae | Palacios 9594 | MO | Herbarium |
| <i>Palicourea tetragona</i> (Donn. Sm.) C. M. Taylor & Lorence | Palicoureeae | Beach 1469 | MO | Herbarium |
| <i>Palicourea tetragona</i> (Donn. Sm.) C. M. Taylor & Lorence | Palicoureeae | Davidse 36909 | MO | Silica |
| <i>Palicourea thyrsoflora</i> (Ruiz & Pav.) DC. | Palicoureeae | Homeier 5199 | MO | Herbarium |
| <i>Palicourea timbiquensis</i> (Standl.) C. M. Taylor | Palicoureeae | Hoover 4120 | MO | Herbarium |
| <i>Palicourea tinctoria</i> Roem. & Schult. | Palicoureeae | Cornejo 149 | MO | Herbarium |
| <i>Palicourea tomentosa</i> (Aubl.) Borhidi | Palicoureeae | Araujo 1546 | MO | Herbarium |
| <i>Palicourea topoensis</i> C. M. Taylor | Palicoureeae | Zak 3747 | MO | Herbarium |
| <i>Palicourea trichocephala</i> (Poepp. & Endl.) Borhidi | Palicoureeae | Schunke 7495 | MO | Herbarium |
| <i>Palicourea triphylla</i> DC. | Palicoureeae | Killeen 6548 | MO | Herbarium |
| <i>Palicourea triphylla</i> DC. | Palicoureeae | Nee 46793 | MO | Herbarium |
| <i>Palicourea triphylla</i> DC. | Palicoureeae | Stevens 36461 | MO | Silica |
| <i>Palicourea winkleri</i> Borhidi | Palicoureeae | Stevens 37485 | MO | Silica |
| <i>Palicourea woronovii</i> (Standl.) C. M. Taylor, Bruniera & Zappi | Palicoureeae | van der Werff 20100 | MO | Herbarium |
| <i>Rudgea cornifolia</i> (Kunth) Standl. | Palicoureeae | de Gracia Cruz 818 | MO | Herbarium |
| <i>Psychotria brachypoda</i> (Müll. Arg.) L. B. Sm. & Downs | Psychotrieae | Silva 1622 | MO | Herbarium |

(Continues)

| Species | Tribe | Voucher | Institution | Tissue type |
|---|--------------|---------------|-------------|-------------|
| <i>Psychotria carthagenensis</i> Jacq. | Psychotrieae | Araujo 2124 | MO | Silica |
| <i>Psychotria grandis</i> Sw. | Psychotrieae | Taylor 11745 | MO | Silica |
| <i>Psychotria guianensis</i> DC. | Psychotrieae | Merello 1711 | MO | Herbarium |
| <i>Psychotria horizontalis</i> Sw. | Psychotrieae | Stevens 32733 | MO | Silica |
| <i>Psychotria jinotegensis</i> C. Nelson, Ant. Molina & Standl. | Psychotrieae | Stevens 33549 | MO | Silica |
| <i>Psychotria limonensis</i> K. Krause | Psychotrieae | Stevens 31580 | MO | Silica |
| <i>Psychotria marginata</i> Sw. | Psychotrieae | Stevens32781 | MO | Silica |
| <i>Psychotria nervosa</i> Sw. | Psychotrieae | Stevens 32362 | MO | Silica |
| <i>Psychotria panamensis</i> Standl. | Psychotrieae | Stevens 32285 | MO | Silica |
| <i>Psychotria subsessilis</i> Benth. | Psychotrieae | Stevens 31494 | MO | Silica |
| <i>Psychotria suterella</i> Müll. Arg. | Psychotrieae | Souza 8789 | MO | Herbarium |

Note: LSU = Louisiana State University; MO = Missouri Botanical Garden.

*Samples that failed sequencing.

Appendix 2. Genomic resources used throughout the study.

| Species | Subfamily ^a | Tribe | Data type | Source | ID |
|---|------------------------|---------------|-----------|-----------------|---------------------|
| <i>Carapichea ipecacuanha</i> (Brot.) L. Andersson | RUB | Palicoureeae | T | 1KP | JOPH |
| <i>Cinchona pubescens</i> Vahl | CIN | Cinchoneae | T | MedPlant | medp_cinpu-20110618 |
| | | | AG | NCBI Genome | GCA_025175665.1 |
| <i>Coffea arabica</i> L. | IXOR | Coffeae | AG | NCBI Genome | GCA_003713225.1 |
| <i>Coffea eugenioides</i> S. Moore | IXOR | Coffeae | AG | NCBI Genome | GCA_003713205.1 |
| <i>Coffea canephora</i> Pierre ex A. Froehner | IXOR | Coffeae | AG | NCBI Genome | GCA_900059795.1 |
| <i>Coffea humblotiana</i> Baill. | IXOR | Coffeae | AG | NCBI Genome | GCA_023065734.1 |
| <i>Corynanthe mayumbensis</i> (R. D. Good) Raym.-Hamet ex N. Hallé | CIN | Naucleaeae | GS | NCBI SRA | SRX5486283 |
| <i>Galium californicum</i> Hook. & Arn. | RUB | Rubieae | GS | NCBI SRA | SRX5658933 |
| <i>Galium odoratum</i> (L.) Scop. | RUB | Rubieae | GS | NCBI SRA | SRX8928310 |
| <i>Galium porrigens</i> Dempster | RUB | Rubieae | AG | NCBI Genome | GCA_012274505.1 |
| <i>Galium verum</i> L. | RUB | Rubieae | GS | NCBI SRA | ERR3089164 |
| <i>Gardenia jasminoides</i> J. Ellis | IXOR | Gardenieae | AG | NCBI Genome | GCA_013103745.1 |
| | | | AG | NCBI Nucleotide | CM023130.1 |
| <i>Gynochthodes officinalis</i> (F. C. How) Razafim. & B. Bremer | RUB | Morindeae | AG | NCBI Genome | GCA_020080225.1 |
| <i>Hamelia patens</i> Jacq. | CIN | Hamelieae | T | NCBI TSA | SRX8873813 |
| <i>Leptodermis oblonga</i> Bunge | RUB | Paederieae | AG | NCBI Genome | GCA_016801395.1 |
| <i>Mitragyna speciosa</i> (Korth.) Havil. | CIN | Naucleaeae | AG | NCBI Genome | GCA_024721245.1 |
| | | | GS | NCBI SRA | SRR5602600 |
| <i>Neolamarckia cadamba</i> (Roxb.) Bosser | CIN | Naucleaeae | T | NCBI TSA | SRX400176 |
| <i>Ophiorrhiza pumila</i> Champ. ex Benth. | RUB | Ophiorrhizeae | AG | NCBI Genome | GCA_016586305.1 |
| <i>Mapouria douarrei</i> Beauvis. | RUB | Psychotrieae | T | 1KP | DNQA |

| Species | Subfamily ^a | Tribe | Data type | Source | ID |
|---------------------------------|------------------------|--------------|-----------|-------------|-----------------|
| <i>Psychotria marginata</i> Sw. | RUB | Psychotrieae | T | 1KP | PCNH |
| <i>Rhazya stricta</i> Decne. | RAUV | Amsonieae | AG | NCBI Genome | GCA_001752375.1 |

Note: CIN = Cinchonoideae; IXOR = Ixoroideae; RAUV = Rauvolfoideae; RUB = Rubioideae; AG = assembled genome; GS = genome skim; T = transcriptome.

^aSubfamily RAUV is in Apocynaceae.