UNIVERSITY OF GOTHENBURG

DEPARTMENT OF BIOLOGICAL AND ENVIRONMENTAL SCIENCES

SYSTEMATICS AND THE EVOLUTION OF FLOWER SYMMETRY OF *POSOQUERIA* (RUBIACEAE)

Sebastián Giraldo Gómez

Degree project for Master of Science (120 hec) with a major in Biodiversity & Systematics

BIO707, Degree project in Biodiversity and Systematics (60 hec)

Second cycle

Semester/year: Autumn 2020- Spring 2023

Supervisors: Claes Persson, Christine Bacon. Department of Biological and Environmental Sciences

Examiner: Bengt Oxelman. Department of Biological and Environmental Sciences

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The genus *Posoqueria* is a Neotropical group of shrubs and trees in the coffee family Rubiaceae. *Posoqueria* is phylogenetically poorly understood and presents several internal taxonomical difficulties. Previous studies using limited molecular data and taxon sampling suggest the species in *Posoqueria* could be split into two monophyletic groups based on the presence or absence of a unique, specialized pollination system known as the Pollen Catapult Mechanism (PCM). This split implies the need for the resurrection of the genus *Stannia*, which groups the species lacking the PCM. Here, 15 species were sampled with target sequence capture and a phylogeny of *Posoqueria* was generated based on 177 genes and the multispecies coalescent model. The study confirms the monophyly of most *Posoqueria* taxa and provides evidence of the phylogenomic distinctiveness of species such as *P. mutisii, P. longiflora, P. maxima, P. grandiflora,* and *P. williamsii,* but questions the current circumscriptions of *P. chocoana* and *P. costaricensis.* Particularly, the phylogeny supports the re-circumscription of *P. latifolia*, suggesting that populations from Central America and Colombia belong to a yet undescribed species in the genus. Furthermore, this phylogenomic study indicates that the unique catapult mechanism linked to zygomorphic flowers in *Posoqueria* is the product of several evolutionary transitions and is, therefore, a non-homologous trait. The results challenge the hypothesis that this trait played a significant role in the diversification of the genus and reject the recognition of *Stannia* as a valid taxon. This novel view of the phylogenetic relationships in *Posoqueria* expands the horizons for future research in the group, especially from the alpha-taxonomy, developmental genetics, pollination biology, ecology, and biogeography points of view.

Keywords: Posoquerieae, Cinchonoideae, Ixoroideae, Dialypetalanthoideae, Angiosperms 353, target capture, Hybpiper, Pollen Catapult Mechanism.

INTRODUCTION

Rubiaceae is one of the largest families of flowering plants, with about 610 genera and up to 14.000 species (POWO, 2023; Wikström et al., 2020). During 2014–2018, Rubiaceae ranked in the top five for new species reported, and in 2019 it ranked second, with 157 new species (Cheek et al., 2020). These statistics indicate that the diversity of Rubiaceae is not well known, and that research is needed, particularly in the megadiverse region of the Neotropics, where this family is one of the most important components of vegetation (Delprete & Jardim, 2012).

One particularly poorly understood Neotropical genus in the family is *Posoqueria* Aubl. This genus has been previously classified within the subfamily Dialypetalanthoideae (formerly Ixoroideae) (Delprete, 2009; Reveal, 2012; Wikström et al., 2020), but Antonelli et al. (2021) only recognized two subfamilies in Rubiaceae, namely Rubioideae and Cinchonoideae, with *Posoqueria* belonging to the latter. *Posoqueria* and the monospecific genus *Molopanthera* Turcz. constitute the tribe Posoquerieae (Delprete, 2009). Some species present taxonomic uncertainties, like *P. coriacea*, *P. williamsii*, and the widely distributed *P. latifolia*, which exhibits considerable morphological and ecological variability, deserving a circumscription re-evaluation (Taylor, 2021).

Posoqueria is distributed from Central America to southern Brazil (Figure 1) and includes around 15-20 species of shrubs and trees (Taylor, 2021) with large, showy flowers pollinated by long-tongued sphingid moths (Delprete, 2009). Actinomorphic flowers are present in some *Posoqueria* species, whereas zygomorphic flowers are found in others, and only the latter exhibit a remarkable specialized pollination system characterized by an asymmetric androecium with a Pollen Catapult Mechanism (PCM) (Figure 2). When a zygomorphic flower opens, all anthers are united into an ellipsoid structure held a few millimeters away from the throat of the corolla, and as a pollinator approaches to insert its proboscis into the corolla tube, it may touch the anther filaments, which triggers the mechanism and causes the anther head to separate explosively and to throw the pollen at the pollinator (Puff et al., 1995). The PCM also occurs in *Molopanthera* and is considered a unique trait within the Rubiaceae family, and possibly even among angiosperms (Cortes-B & Motley, 2015).

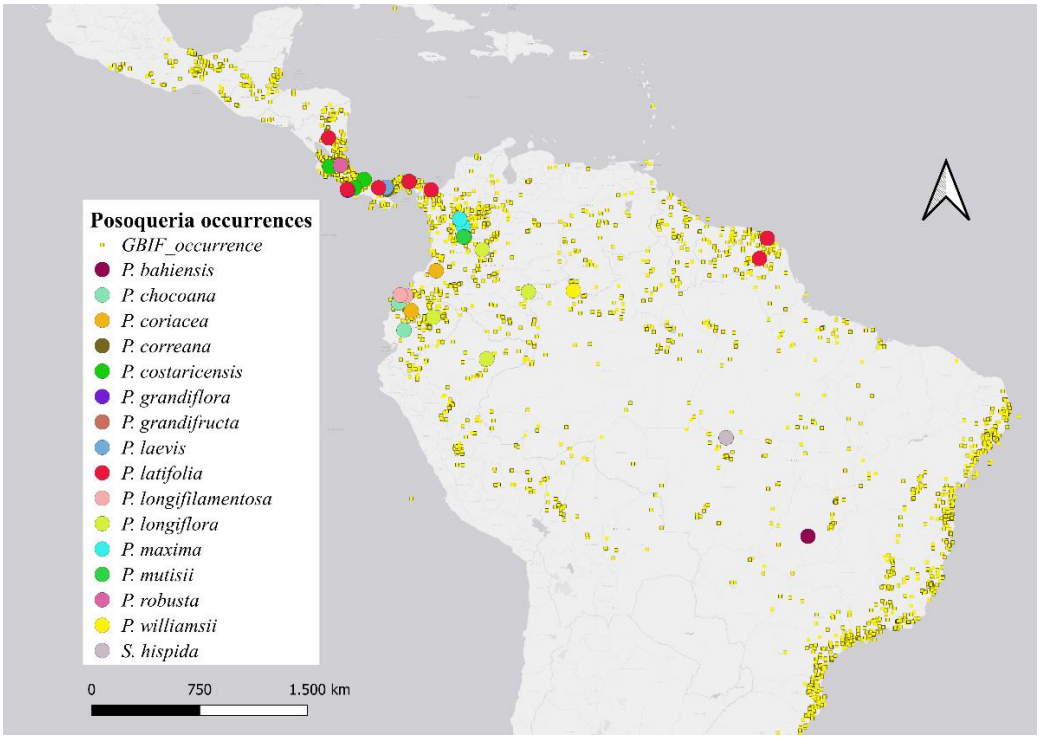

Figure 1. Distribution of *Posoqueria* and voucher localities. [*Posoqueria* occurrences in GBIF.org (09 May 2023) GBIF Occurrence Download https://doi.org/10.15468/dl.h2xgh4].

Figure 2. Flower symmetry in *Posoqueria.* A) Actinomorphic flower of *P. costaricensis*. B) Zygomorphic flower of *P. latifolia* with the anthers united into an ellipsoid structure before the triggering of the Pollen Catapult Mechanism (PCM). C-D) Zygomorphic flower of *P. latifolia* after the triggering of the PCM. Photos: S. Giraldo.

Not all species of *Posoqueria* exhibit the PCM, and for this reason, some taxonomists have previously classified taxa with the PCM in a different genus, *Stannia* Karsten, which is currently considered a synonym of *Posoqueria* (Delprete, 2009). Nevertheless, Cortes-B & Motley (2015) found that the species lacking the PCM formed a monophyletic group, suggesting that the resurrection of *Stannia* may be justified, but this needs to be tested with increased sampling and phylogenetic evidence.

In general, flowers are a morphological innovation that promoted the rapid diversification of angiosperms, with flower symmetry playing a key role in this process (Busch & Zachgo, 2009). In addition, it has been reported that species with zygomorphic flowers have a greater risk of extinction, as they have fewer potential pollinators and there is a decrease in global pollinator populations (Yoder et al., 2020). Particularly, zygomorphic flowers with the PCM have been hypothesized to have played a central role in the initial divergence that led to the formation of *Posoqueria* and *Molopanthera*, as well as in the early radiation of the *Posoqueria* genus (Cortes-B & Motley, 2015). Taken together, investigating the evolution of the PCM and flower symmetry in *Posoqueria* is crucial for understanding the patterns of floral symmetry expression in a predominantly actinomorphic family such as Rubiaceae, as well as in angiosperms in general, and can contribute to defining if the conservation concern in *Posoqueria* species should consider flower symmetry as a key trait.

In the present study, a phylogeny of *Posoqueria* was generated, including representatives of 15 species and using the target sequence capture method (Gnirke et al., 2009), which is a high-throughput DNA sequencing technique that instead of sequencing whole genomes, allows to concentrate sequencing efforts on sets of pre-selected loci, reducing costs and bioinformatic challenges while increasing sequencing depth (Andermann et al., 2020). This method is also appropriate for degraded DNA, common in museum and herbarium specimens (Brewer et al., 2019). Next-generation sequencing technologies, in conjunction with advances in high-performance computing, have revolutionized molecular biology (Andermann et al., 2020), but only recently began to be used in Rubiaceae with promising results (Antonelli et al., 2021; Ly et al., 2020; Ridley, 2022; Wikström et al., 2020).

The universal sequence capture probe set "Angiosperms353" (Johnson et al., 2019) was employed to target 353 nuclear loci, and after performing a gene selection methodology and an orthology inference analysis, 177 genes were chosen to generate a phylogeny of *Posoqueria* with the multispecies coalescent (MSC) method implemented by the "Accurate Species TRee Algorithm" ASTRAL-III (Zhang et al., 2018). This method is a powerful tool and theoretical framework for inferring species phylogenies while accounting for ancestral polymorphism and gene tree - species tree conflicts (Xu & Yang, 2016). The MSC has been proven to outperform the concatenation model, especially for large phylogenomic data sets on which more than 10 loci are analyzed (Jiang et al., 2020). It also has been demonstrated the inconsistency of concatenation methods under the anomaly zone (Mendes & Hahn, 2018), on which coalescent models of tree building are still consistent (Jiang et al., 2020).

Here, I developed a phylogenomic framework for *Posoqueria* to conduct a systematic evaluation of the genus. This framework allowed the assessment of species circumscriptions, reconstruction of the flower symmetry evolution in the genus, and testing whether actinomorphic species lacking the PCM form a monophyletic group. Additionally, the phylogeny was used to investigate the monophyly of the widespread species *Posoqueria latifolia* and to address some taxonomic problems within the genus. In this way, in a time in which Neotropical biodiversity is rapidly declining due to anthropogenic factors (Antonelli, 2022), increasing the knowledge about plants with high sensitivity to habitat degradation due to their specialized pollination mechanism (Amorim et al., 2014; Yoder et al., 2020), serves as an important contribution to solving, describing, and disseminating the systematics of the threatened Neotropical flora, information that is essential for both the conservation of ecosystems and the species they harbor.

MATERIALS AND METHODS

Taxon Sampling and Data Collection

Material was sourced from silica-dried samples and herbarium specimens from ALCB, GB, HUA, MO, and NY (herbaria abbreviations following Index Herbariorum; Thiers, 2023). Additional specimens for this project were collected during a field trip in 2022 in Colombia, Costa Rica, and Panama, prioritizing collections nearby the type localities. For three supplementary samples, target-enriched by Antonelli et al. (2021) with the Angiosperms353 probe set (Johnson et al., 2019), sequences were obtained from the European Nucleotide Archive (ENA) in a custom project area (https://www.ebi.ac.uk/ena/browser/view/PRJEB35285). The 42 voucher samples analyzed in this study are listed in Table 1.

Specimens were identified by reviewing both physical and online type collections, protologues, and representative pressed herbarium samples from GB, HUA, JAUM, MEDEL, and MO. Supplementary botanical keys and descriptions were also consulted (Burger & Taylor, 1993; Lorence & Taylor, 2012; Macias, 1998). Species classification followed the accepted species names in the Tropicos Rubiaceae Project (Taylor, 2021).

Molecular Protocols

DNA extraction, quantification, and qualification: Total genomic DNA was extracted from 30 mg of herbarium material and 15 mg of silica gel-dried material, using the DNeasy Plant Pro Kit (Qiagen, Hilden, Germany). Leaf tissue was pulverized using a Tissuelyser II (Qiagen, Venlo, Netherlands). The manufacturer's extraction protocol was followed, except for the homogenization and cell lysis step. An initial homogenization was carried out by running the Tissuelyser at 24 Hz for 2 min with the plant tissue in dry condition. The Tissuelyser adapters were reoriented so that the side that was closest to the machine body became furthest from it, and then the Tissuelyser was run again at 24 Hz for 2 min. Subsequently, 450 µL of Solution CD1 and 50 µL of Solution PS were added to each tube, placing the tubes in a dry block heater for 2 h at 60 ºC, to maximize DNA yield. After incubation, the manufacturer's protocol was followed, starting with the centrifugation

process, step 3. The final step of the protocol consisting of DNA elution in Buffer EB solution was carried out twice to improve DNA yield, using 50 µL of Buffer EB in each consecutive step, obtaining a final volume of 100 µL. Extracted DNA was evaluated with a Nanodrop 2000c spectrophotometer, quantified using a Qubit 3.0 fluorometer (Thermo Fisher Scientific, Waltham, US), and then gel electrophoresis was run in a 1% agarose gel to assess the average DNA fragment size.

Table 1. Voucher specimens. ** Tissue sourced from pressed herbarium samples. * Sequences generated by Antonelli et al. (2021). The rest of the collections were sourced from silica-dried leaves.

N	Species	Collector	Collector No.	Country	Herbarium ID
1	*Posoqueria bahiensis*	Silva	199**	Brazil	ALCB
2	*Posoqueria chocoana*	Persson, Nordenhäll, Aulestia	412	Ecuador	GB
3	*Posoqueria chocoana*	Persson, Nordenhäll	426	Ecuador	GB
4	*Posoqueria chocoana*	Persson, Andersson, Lilja	1341	Ecuador	GB
5	*Posoqueria chocoana*	Antonio	2528**	Panama	MO
6	*Posoqueria chocoana*	McPherson	9775**	Panama	MO
7	*Posoqueria coriacea*	Persson, Nordenhäll, Tapia	459	Ecuador	GB
8	*Posoqueria coriacea*	Persson *et al.*	2134	Colombia	GB
9	*Posoqueria correana*	Persson, Giraldo, Rova	3870	Panama	GB
10	*Posoqueria correana*	Persson, Giraldo, Rova	3871	Panama	GB
11	*Posoqueria costaricensis*	Persson, Antonelli, González	861	Panama	GB
12	*Posoqueria costaricensis*	Persson, Giraldo, Rova	4002	Costa Rica	GB
13	*Posoqueria costaricensis*	Persson, Giraldo, Rova	4040	Costa Rica	GB
14	*Posoqueria grandiflora*	Persson, Giraldo, Rova	4018	Costa Rica	GB
15	*Posoqueria grandifructa*	Persson, Giraldo, Rova	3977	Costa Rica	GB
16	*Posoqueria grandifructa*	Persson, Giraldo, Rova	3984	Costa Rica	GB
17	*Posoqueria laevis*	McPherson	19861**	Panama	MO
18	*Posoqueria latifolia*	Giraldo	735	Colombia	HUA
19	*Posoqueria latifolia*	Persson *et al.*	1950	French Guiana	GB
20	*Posoqueria latifolia*	Persson, Giraldo, Rova	3801	Panama	GB
21	*Posoqueria latifolia*	Persson, Giraldo, Rova	3963	Panama	GB
22	*Posoqueria latifolia*	Persson, Giraldo, Rova	3966	Costa Rica	GB
23	*Posoqueria latifolia*	Persson, Giraldo, Rova	4006	Costa Rica	GB
24	*Posoqueria latifolia*	Mori	24280	French Guiana	NY
25	*Posoqueria latifolia*	Stevens	34378**	Nicaragua	MO
26	*Posoqueria latifolia*	Maurin	4390*	Panama	K
27	*Posoqueria longifilamentosa*	Persson *et al.*	1864	Ecuador	GB
28	*Posoqueria longifilamentosa*	Luteyn	15022	Ecuador	MO
29	*Posoqueria longiflora*	Medina *et al.*	10	Colombia	HUA
30	*Posoqueria longiflora*	Persson, Grández	681	Peru	GB
31	*Posoqueria longiflora*	Giraldo	719	Colombia	HUA
32	*Posoqueria longiflora*	Giraldo	720	Colombia	HUA
33	*Posoqueria longiflora*	Perez *et al.*	10797	Ecuador	GB
34	*Posoqueria maxima*	Alzate *et al.*	228	Colombia	GB
35	*Posoqueria maxima*	Tobón	4163	Colombia	HUA
36	*Posoqueria mutisii*	Cortés *et al.*	2298	Colombia	HUA
37	*Posoqueria mutisii*	Cortés *et al.*	2310	Colombia	HUA
38	*Posoqueria robusta*	Persson, Giraldo, Rova	3975	Costa Rica	GB
39	*Posoqueria robusta*	Persson, Giraldo, Rova	3982	Costa Rica	GB
40	*Posoqueria williamsii*	Cortés, Rivas, Evangelista	2203	Colombia	HUA
41	*Emmenopterys henryi*	s.n.	1069437*	s.l.	K
42	*Sipanea hispida*	Zappi	990*	Brazil	K

***Library preparation, enrichment, and sequencing*:** DNA extracts were sent to Rapid Genomics (Florida, USA). Library preparation was performed for Illumina sequencing utilizing their high-throughput workflow with proprietary chemistry. DNA was sheared to a mean fragment length of 500 bp, fragments were end-repaired and A- tailed, followed by the incorporation of unique dual-indexed Illumina adaptors and PCR enrichment. Samples were pooled equimolar, target-enriched with the Angiosperms353 probe set (Johnson et al., 2019), and sequenced on an Illumina NovaSeq 6000 S4 flow cell, producing 2x150 bp paired-end reads.

Read Processing and Assembly

***Read mapping*:** The quality of the target sequence capture raw reads (.fastq files) was checked with FastQC (Andrews, 2010). Raw reads were then trimmed and cleaned using Fastp (Chen et al., 2018) in order to obtain high-quality and high-confidence data in the downstream analysis. Fastp was run using default parameters for paired end (PE) data, which includes filtering based on phred quality (>=Q15), polyG tail trimming, reads length filtering (L≥15), maximum percent of low-quality bases allowed 40%, read/pair discarded when the number of N bases greater than limit (N>5), and although adapter sequences for PE data can be automatically detected by per-read overlap analysis, specific adapter sequences (.fasta files) were provided for each accession (--adapter_fasta) to trim the forward and reverse adapters, which results in cleaner output data, since the overlap analysis may fail due to sequencing errors or adapter dimers (Chen et al., 2018). Raw reads were also cleaned with Trimmomatic (Bolger et al., 2014), but better results were obtained with Fastp. The quality of cleaned reads was assessed using FastQC, and both Fastp and FastQC reports were summarized with MultiQC (Ewels et al., 2016), visualizing global results and trends.

Cleaned paired reads were used to recover target sequences with HybPiper v2.1.2 (Johnson et al., 2016) using default settings. The “assemble and sequence extraction” command was run (hybpiper assemble), assembling de novo each gene using SPAdes v3.15.5 (Bankevich et al., 2012) (coverage cutoff for SPAdes is by default 8x) and extracting coding sequences using Exonerate v2.4.0 (Slater & Birney, 2005). Recovery statistics, available at Zenodo (https://doi.org/10.5281/zenodo.7896478), were obtained using the HybPiper commands for “statistics” (hybpiper stats) and “gene recovery heatmap” (hybpiper recovery_heatmap).

Hybpiper was run three times with different methods for aligning reads to targets and with different target files, in order to find the approach that best recovered the gene sequences: i) Hybpiper was run with the BWA method (Li & Durbin, 2009) and the “Angiosperms353” target file (Johnson et al., 2019), available at GitHub (https://github.com/mossmatters/Angiosperms353). BWA aligns reads to targets using nucleotide sequences; ii) Hybpiper was run with the BWA method and a filtered version of the "mega353" target file (McLay et al., 2021). To maximize computational efficiency, the "mega353" target file was subsampled to include the existing sequences of the family Rubiaceae in the 1KP project (https://sites.google.com/a/ualberta.ca/onekp/), using the script filter_mega353.py. Both, the script and the expanded "mega353" target file, are available at GitHub (https://github.com/chrisjackson-pellicle/NewTargets). The final filtered version of the "mega353" target file, included the standard “Angiosperms353” target sequences and sequences from 4 Rubiaceae taxa (*Galium boreale, Morinda citrifolia, Psychotria ipecacuanha, and Psychotria marginata*). This expanded version (posoq.fasta) of the “Angiosperms353” target file is available at Zenodo (https://doi.org/10.5281/zenodo.7896478); iii) Hybpiper was run with the BLASTx method (Camacho et al., 2009) and the filtered version of the "mega353" target file (posoq.fasta), which is a nucleotide target file. As BLASTx uses amino-acid sequences as a reference, HybPiper translated the nucleotide target file into a protein target, and then mapped reads to targets.

***Selection of genes and paralogy resolution*:** HybPiper run statistics were used to select only genes for which more than 90% of the sequences (≥38/42 sample sequences) were recovered with a

sequence length longer than 75% of the mean target length, as the presence of fragmentary sequences can increase gene tree error (Hosner et al., 2016; Sayyari et al., 2017).

The HybPiper post-processing command (hybpiper paralog_retriever) was run to recover coding sequences of alternative long paralogs from loci flagged by HybPiper as having putative paralogs. Subsequently, an orthology inference analysis, critical for phylogenomic reconstruction, was run with the paragone-nf paralogy resolution pipeline, with commands and instructions available at GitHub (https://github.com/chrisjackson-pellicle/paragone-nf). The paragone-nf nextflow pipeline was run with default parameters, working over sequences not flagged as a putative chimera (output folder paralogs_no_chimeras) by the hybpiper paralog_retriever command, selecting as outgroups (--internal_outgroups) the accessions 1069437 and 990 (*Emmenopterys henryi* and *Sipanea hispida*). The output folder of the "Monophyletic outgroup" (MO) method (19_alignments_stripped_names_MO_realigned), containing the .fasta files for the inferred orthologous target genes, was used in downstream analyses.

Phylogenetic Analysis

***Sequence alignments*:** Gene matrices recovered using the orthology inference analysis were aligned independently using MAFFT v7.515 (Katoh & Standley, 2013) with the high accuracy method, incorporating iterative refinement with local pairwise alignment (--maxiterate 1000 --localpair) and using the default gap scoring scheme. Multiple Sequence Alignments (MSA) for each gene were then trimmed using Phyutility v2.7.3 (Smith & Dunn, 2008) (https://github.com/blackrim/phyutility), removing sites that were missing 50% data (-clean 0.5).

***Gene trees*:** Gene trees from trimmed MSA were obtained using IQ-TREE multicore v2.2.2.2 COVID-edition (Minh et al., 2020) with the ModelFinder option (-m MFP) (Kalyaanamoorthy et al., 2017), which determined the best-fit substitution model for each gene alignment. ModelFinder chooses the model that minimizes the Bayesian Information Criterion (BIC) score. A thorough and more accurate analysis that invokes a full tree search for each model considered (-mtree) was also performed in IQ-TREE. Generated gene trees were then analyzed using TreeShrink v1.3.9 (Mai & Mirarab, 2018) with default settings, excluding branches that increased the diameter of each gene tree by more than 20% (-b 20), filtering outliers from both trees and original trimmed MSA. These filtered alignments were used to generate the final gene trees with IQ-TREE, executing again ModelFinder (-m MFP) and (-mtree) options. Ultrafast bootstrapping (UFBoot2) option in IQ-TREE was also run with 1000 replicates (-B 1000) (Hoang et al., 2018), including the option to reduce the risk of overestimating branch supports with UFBoot due to severe model violations (-bnni). Finally, branches with support values below 10% were collapsed using Newick Utilities v1.6 (Junier & Zdobnov, 2010).

Multispecies coalescent-based species tree: A species tree was generated using the maximum-likelihood (ML) gene trees obtained with IQ-TREE and contracted with Newick Utilities, thus acknowledging gene tree uncertainty. The resulting multifurcating gene trees were used as input to the Accurate Species TRee Algorithm ASTRAL-III v5.7.8 (Zhang et al., 2018), on which the final species tree was reconstructed with full branch annotations (quartet support, quartet frequency, and posterior probability for all three alternatives, plus total number of quartets around the branch and effective number of genes) using the (-t 2) branch annotation level. The final tree was visualized and manually edited in FigTree v1.4.4 (Rambaut, 2018) and Adobe Photoshop CS6 13.1 Portable.

Ancestral state reconstruction of the flower symmetry: The ASTRAL-III coalescent tree was used as input for an ancestral state reconstruction of the flower symmetry in the software "Reconstruct Ancestral State in Phylogenies" RASP 4 v4.3 (Yu et al., 2020), on which a MultiState reconstruction was implemented using BayesTraits v3.0.1 (Meade & Pagel, 2018) with default

values (Iterations 5.050.000, Sample 10.000, BurnIn 50.000, MLTries 100) for two different approaches: Markov chain Monte Carlo (MCMC) method and maximum likelihood (ML) method.

RESULTS

Taxon Sampling

In total, 42 samples corresponding to 15 species of *Posoqueria* and two outgroups, *Emmenopterys henryi* and *Sipanea hispida*, were included (Table 1). Selected specimens represented to the best extent possible the morphological diversity and geographic distribution of *Posoqueria* (Figure 1).

Sequences Recovery and Orthology Inference

After cleaning the raw reads with Fastp, an average of 98% of the reads passed the filtering process. Figure 3 provides a summary of the filtering analysis for the forward reads (R1), with similar results observed for the reverse reads (R2). The Phred quality scores of the nucleotide base calls in the read sequences were high and consistent across the read length, representing a good quality of the Illumina sequencing output. As shown in Figure 4, the Phred quality scores ranged from 35 to 40 after filtering with Fastp, indicating accuracy of approximately 99.9% to 99.99% (Negi et al., 2022).

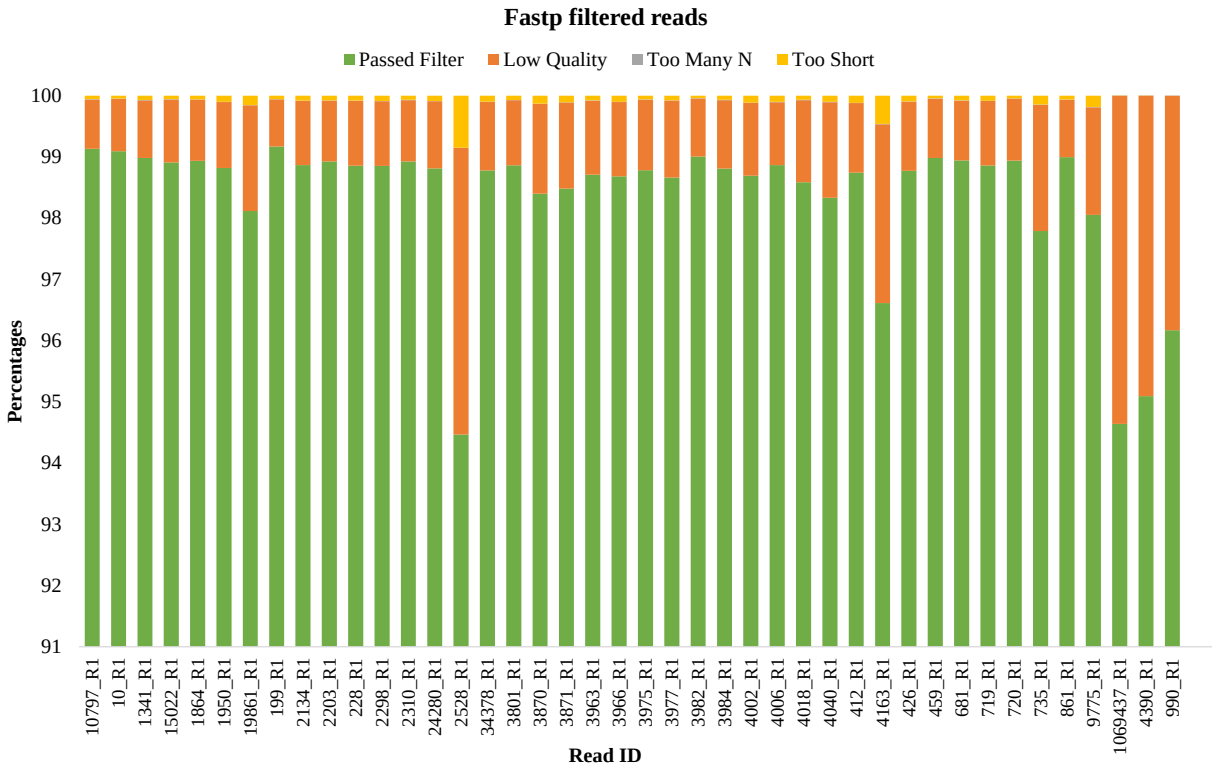

Figure 3. Percentages of reads in the forward read (R1) .fastq files that passed the Fastp filtering process.

Based on the comparison of the three approaches implemented in Hybpiper, it was observed that the BWA method using the "Angiosperms353" target file yielded suboptimal results, while the BWA method coupled with the expanded target file "posoq.fasta" produced intermediate levels of genes recovery. Most notably, the BLASTx method in conjunction with the expanded target file "posoq.fasta" outperformed both BWA methods, and enabled the most effective recovery strategy of gene sequences (Figure 5).

In terms of general recovery efficiency, the three Hybpiper methods performed similarly, recovering about 340 genes with sequences. However, when comparing non-fragmentary genes, i.e., genes recovered with a sequence length longer than 75% of the mean target length, the differences were noteworthy (Figure 5). At the 75% threshold, the BLASTx method with the

expanded target file "posoq.fasta" generated about 12% and 34% more genes sequences than BWA with the expanded "posoq.fasta" target file and with the "Angiosperms353" target file, respectively. Therefore, only gene sequences recovered by the BLASTx method with the "posoq.fasta" target file were chosen for downstream analyses, and the following results and discussion sections are exclusively based on the outcomes of this approach.

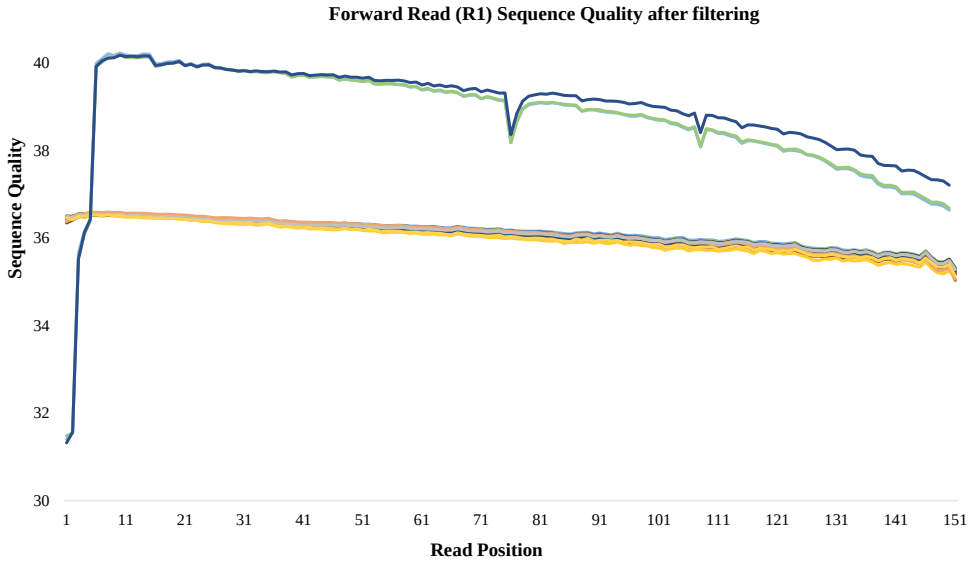

Figure 4. Phred quality scores of the forward reads (R1) after the Fastp filtering process. Each line represents an (R1) .fastq file.

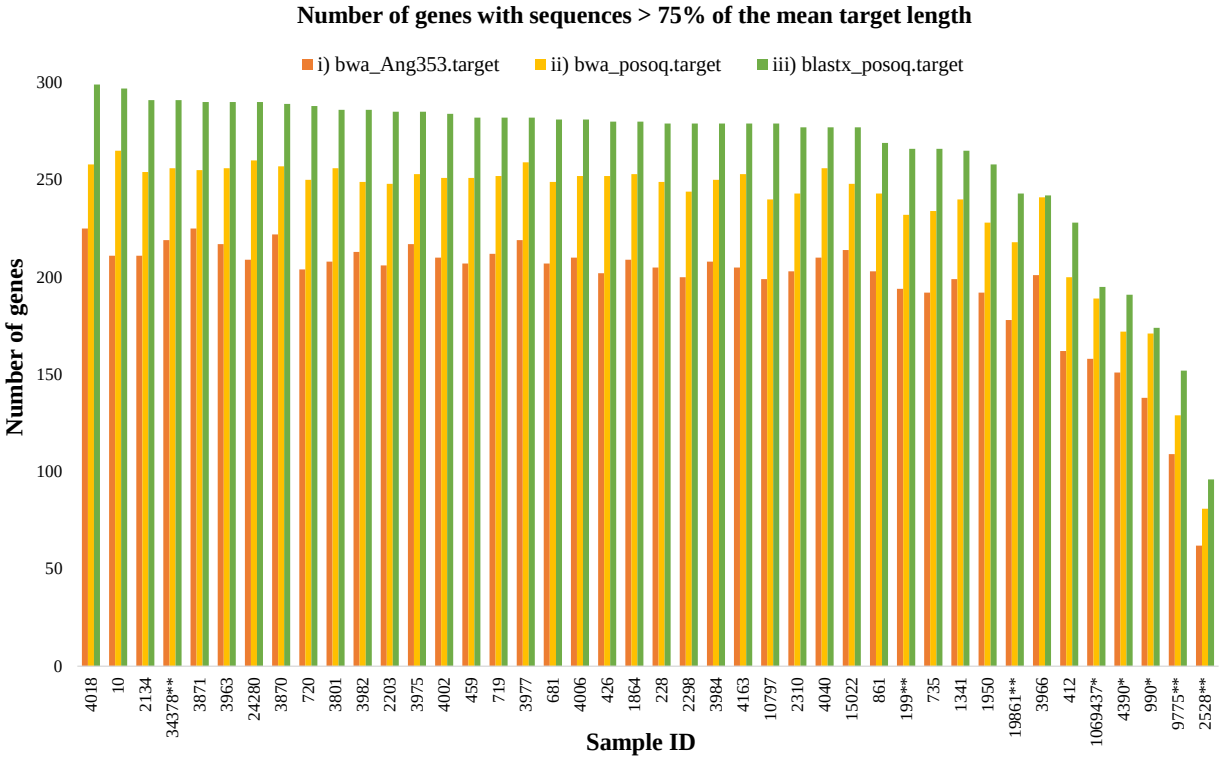

Figure 5. Number of genes with sequences longer than 75% of the mean target length, recovered with three different methods implemented in Hybpiper: i) BWA method with the "Angiosperms353" target file (orange); ii) BWA method with the expanded target file "posoq.fasta" (yellow); iii) BLASTx method with the expanded target file "posoq.fasta" (green). **Tissue sourced from pressed herbarium samples. *Sequences generated by Antonelli et al. (2021). The rest of the samples were sourced from silica-dried leaves.

The average number of input reads that mapped to sequences in the target file was 3.607.220 reads per sample. The percentage of reads in the input .fastq files that mapped to sequences in the target file was, on average, 36.5% per sample, with a maximum of 41.3% (sample 3801), and a minimum of 10.2% (sample 4390). The total number of genes with sequences, i.e., the number of genes that had coding sequences extracted after the Exonerate analysis, was on average 344 genes per sample. These genes with sequences were recovered with sequence lengths ranging from more than 0% to 100% relative to the mean target length, as shown in the recovery heatmap in Figure 6 (high resolution at Zenodo https://doi.org/10.5281/zenodo.7896478) for each gene-sample pair and summarized in Figure 7 for the total number of genes per sample.

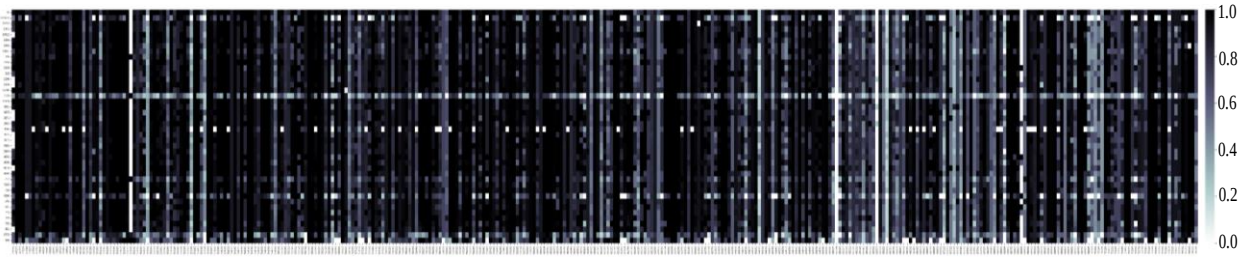

Figure 6. Hybpiper Recovery Heatmap. Each row shows a sample, and each column represents a recovered gene. The shading intensity in each box reflects the length of the gene recovered for that sample, relative to the length of the reference target sequence, as indicated by the scale bar on the right.

At the 75% target length threshold, an average of 264 genes per sample were recovered (Figure 7**),** and after applying a selection criterion to exclude fragmentary data, choosing only genes for which more than 90% of the sequences (≥38/42 sample sequences) were recovered with a sequence length longer than 75% of the target length, 197 genes were selected for subsequent analyses.

After running the Hybpiper paralog_retriever command, it was found that all 42 samples were reported as having one putative paralog sequence at least for one of the genes recovered. In the most extreme cases, a sample could have up to 6 putative paralog sequences recovered for a single gene. In total, 77 genes out of the 197 genes previously selected, had one or more putative paralogue sequences retrieved. The number of paralogue sequences for each gene and sample is graphically represented in Figure 8 (high resolution at Zenodo https://doi.org/10.5281/zenodo.7896478).

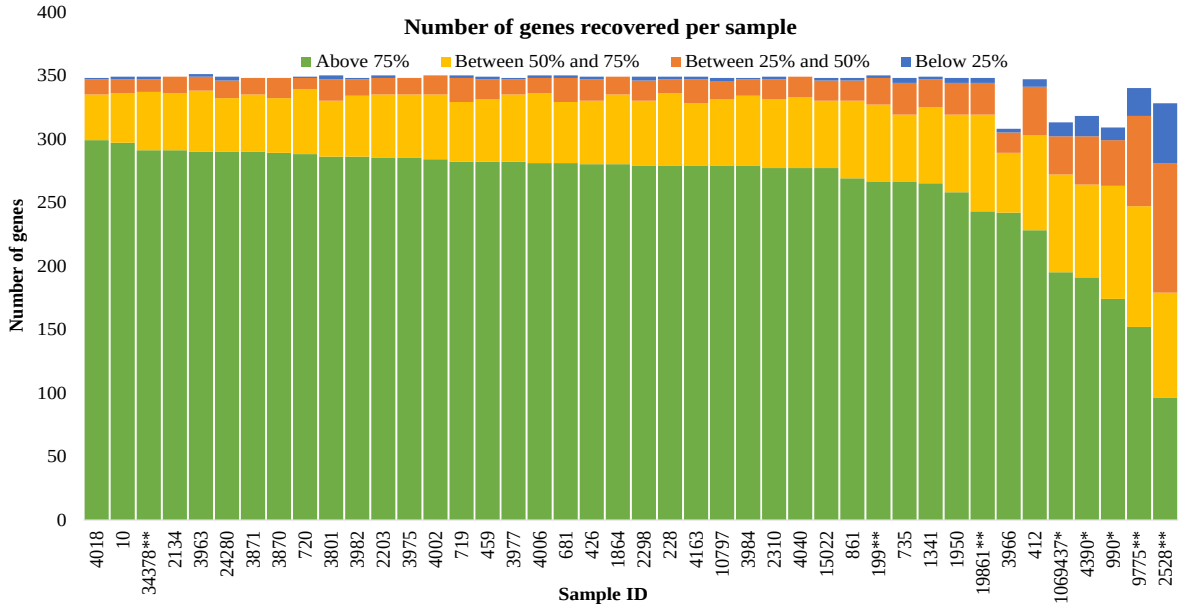

Figure 7. Number of genes with sequences per sample, categorized by percentage of target length recovered.

The output folder of the Hybpiper paralog_retriever command with putative chimeric sequences removed (paralogs_no_chimeras) was used as input for the paragone-nf pipeline, including only the 197 selected non-fragmentary genes. The paragone-nf analysis using the "Monophyletic outgroup" (MO) method inferred 177 orthologs, effectively recovering 57 orthologs from the initial 77 genes with putative paralogs, and definitely excluding 20 genes for which the orthologue copies could not be identified. Paragone-nf removed 633 sequences from the complete data set, and the detail per gene is provided at Zenodo (https://doi.org/10.5281/zenodo.7896478).

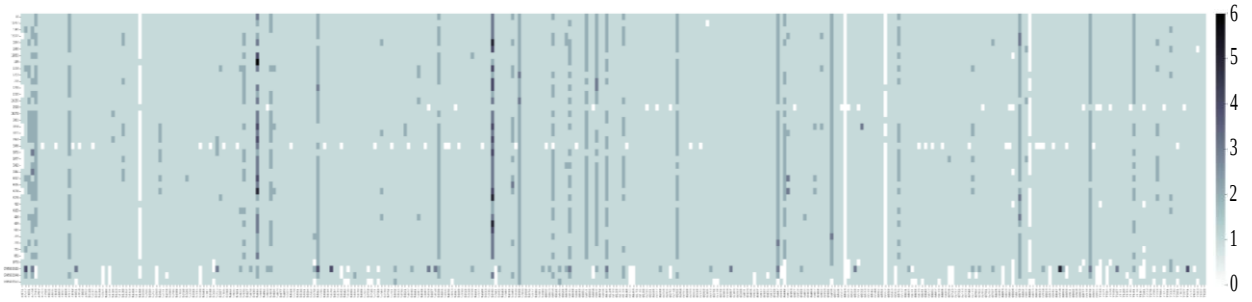

Figure 8. Hybpiper Paralog Heatmap. Each row shows a sample, and each column represents a recovered gene. The amount of shading in each box corresponds to the number of retrieved sequences for each gene and sample (potential paralogs if number >1), as indicated by the scale bar on the right.

Phylogenetic Analysis and Ancestral State Reconstruction

Following the generation of 177 gene trees using IQ-TREE, the trees were processed with TreeShrink (Mai & Mirarab, 2018), which removed 265 sequences from the whole data set. The list of sequences excluded from each locus is available at Zenodo (https://doi.org/10.5281/zenodo.7896478). Gene trees were analyzed using TreeShrink to find suspicious patterns of branch lengths and then shrink the diameter (i.e., the maximum total branch length between any two leaves) of each gene tree by removing species, as it has been shown that erroneous sequences often appear as unexpectedly long branches in the inferred tree (Mai & Mirarab, 2018). About 4% of the total number of sequences were removed from the complete data set utilizing this method. TreeShrink filtered alignments were subsequently used to generate the final gene trees with IQ-TREE and branches with support values below 10% were collapsed with Newick Utilities, as it has been demonstrated that collapsing branches with extremely low support (BS below 5-20%) can considerably improve accuracy of the subsequent ASTRAL analysis (Mirarab, 2019; Zhang et al., 2018). The species tree inferred in ASTRAL-III along with the results of the flower symmetry ancestral reconstruction in RASP 4 is presented as an ultrametric tree in Figure 9. The original coalescent species tree inferred in ASTRAL-III can be checked in Appendix 2 and at Zenodo (https://doi.org/10.5281/zenodo.7896478). Statistical support of relationships is reported as Local Posterior Probabilities (LPP) and the following terms are used to discuss this support in the species tree: (i) nodes with LPP 100 are described as fully supported; (ii) nodes with LPP > 85 are described as highly supported; (iii) nodes with 60 < LPP < 85 are described as moderately supported; and (iv) nodes with LPP < 60 are described as weakly supported.

The ASTRAL-III species tree reconstructed the phylogenetic relationships among species in *Posoqueria* with different levels of support, with a trend to higher local posterior probabilities at deeper nodes. Reviewing the phylogeny in Figure 9 from bottom to top, it can be observed that *P. bahiensis, P. latifolia* (from French Guiana), and *P. williamsii* form a fully supported clade, but relationships within this clade are not well resolved, being weakly to moderately supported. In this group, only *P. bahiensis* has actinomorphic flowers. *Posoqueria longiflora* is fully supported as a monophyletic group and sister to a clade formed by *P. mutisii* + *P. latifolia* (from Central America and Colombia). In *P. longiflora*, intraspecific phylogenomic structure was observed, with Colombian samples forming a clade that is distinct from Peruvian and Ecuadorian individuals.

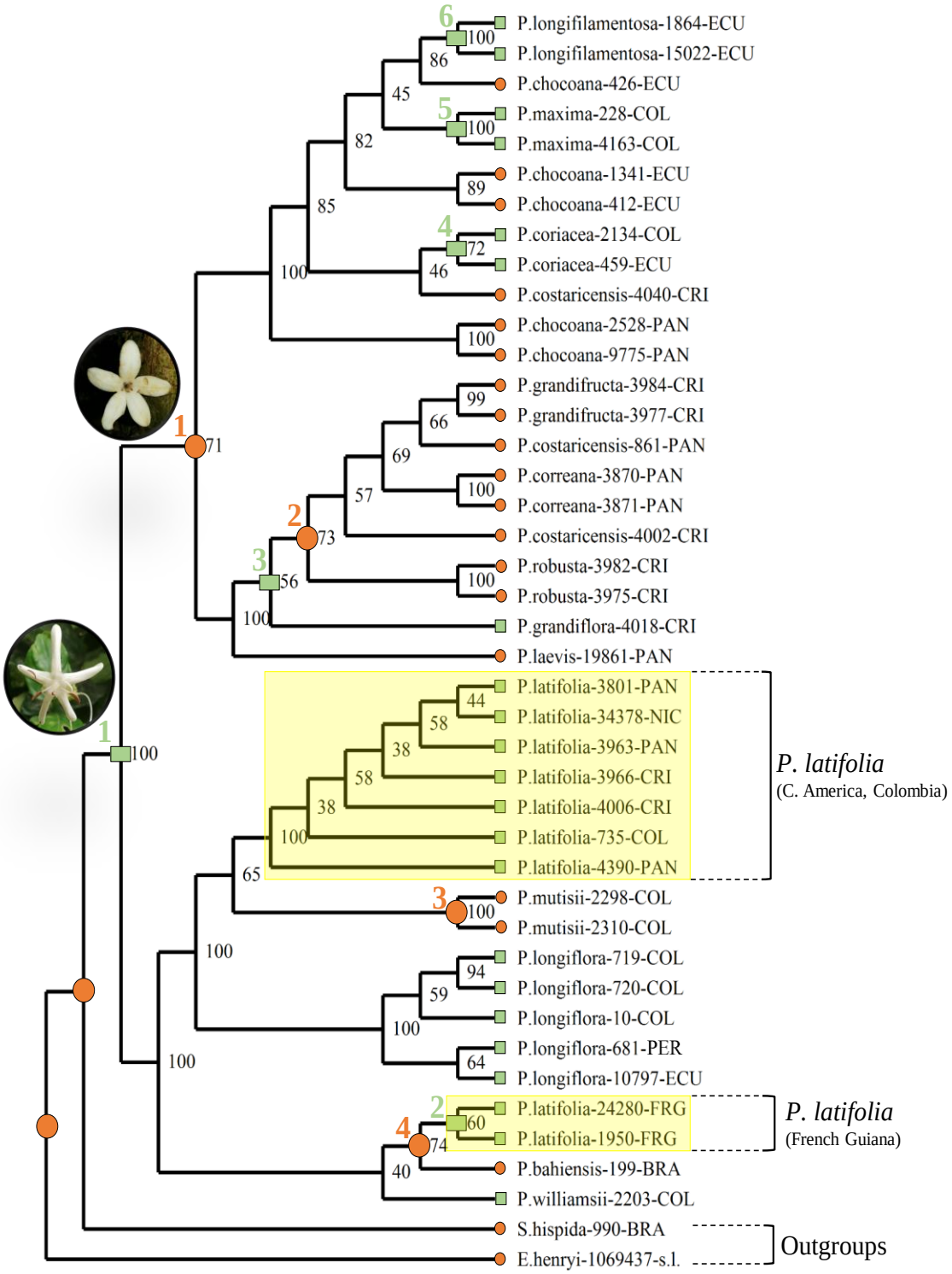

Figure 9. Phylogeny of *Posoqueria* inferred in ASTRAL-III from 177 gene trees, with the results of the ancestral reconstruction of flower symmetry generated in RASP 4. Local posterior probability values are presented in the nodes. Only key nodes are highlighted with the most likely ancestral states, with orange circles (●) representing actinomorphic flowers and green rectangles (■) representing zygomorphic flowers with the Pollen Catapult Mechanism (PCM). Numbers in green denote evolutionary transition events from actinomorphic to zygomorphic flowers, while numbers in orange represent reversal events from zygomorphic to actinomorphic flowers. Each terminal is labeled with the species name, sample ID, and acronym of collection country (BRA:Brazil, COL:Colombia, CRI:Costa Rica, ECU:Ecuador, FRG:French Guiana, NIC:Nicaragua, PAN:Panama, PER:Peru, s.l:unknown locality)

The division between *P. mutisii* and *P. latifolia* is moderately supported (LPP 65) and these species are morphologically different, the former being actinomorphic, while the latter presents zygomorphic flowers with the Pollen Catapult Mechanism (PCM). It is noteworthy that *P. latifolia* was recovered as polyphyletic, with one clade formed by specimens from French Guiana and a second one including only the plants from Central America and Colombia.

A second clade in *Posoqueria*, starting at the node labeled in Figure 9 as 1 (orange), was recovered and is further divided into two groups. In the first group, *P. laevis* is fully supported as sister to (*P. grandiflora* + *P. robusta* + *P. costaricensis* + *P. correana* + *P. grandifructa*). The rest of relationships within the clade are moderately to weakly supported, with *P. costaricensis* being particularly problematic, as it was recovered as a polyphyletic taxon. *P. grandiflora* is the only zygomorphic species in this clade. In the second group, *P. chocoana* (from Panama) is fully supported as sister to (*P. costaricensis* + *P. coriacea* + *P. chocoana* (from Ecuador) + *P. maxima* + *P. longifilamentosa*) (Figure 9). Although most of the nodes in this group were inferred as highly to moderately supported, the positions of the specimens *P. costaricensis* (4040) and *P. chocoana* (426) contradict the current taxonomic classification.

The ancestral state reconstruction of the flower symmetry implemented by BayesTraits in RASP 4 with the Markov chain Monte Carlo (MCMC) method to derive posterior distributions and with the maximum likelihood (ML) method to derive point estimates of log-likelihoods (Meade & Pagel, 2018), yielded the same patterns of inferred flower symmetry states at the ancestral nodes in the phylogeny (Figure 9). The ancestral state reconstruction infers at least six transitions from actinomorphic to zygomorphic flowers, and four reversals from zygomorphic to actinomorphic flowers, indicating that the Pollen Catapult Mechanism, linked to the zygomorphic flowers, is a non-monophyletic trait.

DISCUSSION

The phylogenomic study presented here is the product of an extensive taxonomic and molecular sampling of *Posoqueria*, providing a new perspective on the interspecific relationships in the genus. The results and insights about the phylogenomic methods implemented, the systematics of *Posoqueria,* and the floral evolution, are discussed in the following paragraphs.

Phylogenomics

The results indicate the BLASTx method outperformed the BWA approach in terms of generating longer contigs for numerous genes and successfully generating contigs for some genes that had none when utilizing the BWA option, despite its slower processing speed. These findings are consistent with prior reports from Johnson et al., (2016) and Murphy et al., (2020), showing the BLASTx approach is more accommodating to nucleotide substitutions between the target sequences and sample reads, as alignments are conducted at the peptide level.

The use of the "mega353" target file has been found to facilitate optimal locus recovery from “Angiosperms353” capture data with the Hybpiper pipeline (McLay et al., 2021). A filtered version of the "mega353" target file, including the standard “Angiosperms353” target data and sequences from four Rubiaceae taxa, namely *Galium boreale, Morinda citrifolia, Psychotria ipecacuanha,* and *Psychotria marginata*, was used, confirming the improvement in locus recovery by using this expanded target file. This approach enabled a superior recovery of target capture loci with both BWA and BLASTx methods for aligning reads to targets. The combination of BLASTx with the filtered version of the "mega353" target file (posoq.fasta) yielded better quality of sequences and more genes recovered.

To mitigate gene tree and species tree errors that can result from fragmentary sequences (Hosner et al., 2016; Sayyari et al., 2017), HybPiper run statistics were used to select only genes for which more than 90% of the sequences were recovered with a sequence length longer than 75% of the mean target length. Sayyari et al. (2017) defined species with less than 50% of the total alignment

length as fragmentary, while Wickett et al. (2014) defined their limit as less than 33% of the total alignment length. Here, a more stringent approach was implemented, with a threshold of 75%.

The number of gene sequences recovered at the 75% mean target length threshold was similar for most samples, with an average of 264, except for two samples with comparatively poor results: sample 9775 with 152 genes recovered, and sample 2528 with 96 genes at this threshold (Figure 5). Tissue for DNA extraction from 9775 and 2528 samples was sourced from herbarium pressed material, originally collected in 1986 and 1979, respectively, making them the oldest samples included here. Despite their relatively low number of recovered sequences, about 100 genes still provided useful phylogenetic information, demonstrating the potential of high-throughput sequencing techniques, such as target capture, to study valuable museum specimens and herbarium samples with possibly degraded DNA. These results support the findings of Brewer et al. (2019) about the high potential of using herbarium specimens for novel phylogenetic studies.

When the SPAdes assembler generates multiple contigs that contain coding sequences representing more than 75% (default value) of the reference sequence length, HybPiper flags the locus, indicating multiple long-length matches (putative paralogous sequences) to the reference sequence have been found, and it then chooses one contig among these multiple contigs, based on coverage depth and percent identity with the reference sequence (Johnson et al., 2016). Nonetheless, the criteria used to select the main sequence may not be ideal in all situations and in some cases, the multiple long-length contigs might even represent recent polyploidy, allelic variation, contamination, etc. (Johnson et al., 2016). Therefore, choosing the appropriate gene copy for subsequent phylogenetic analyses needed further attention. Accordingly, the HybPiper post-processing command (hybpiper paralog_retriever) was run, to recover coding sequences from alternative long paralogs. As can be examined in the paralog heat map (Figure 8), when paralog sequences were recovered, usually several samples had more than one copy for the same gene (vertical shaded lines), which may indicate ancient gene duplications. In contrast, no tendency for one sample to have many copies was observed (absent horizontal shaded lines), which may have indicated polyploidy taxa.

HybPiper also performs a read-mapping approach to detect cases where chimeric locus sequences are created (i.e., a sequence derived from stitching together coding sequences from different paralogs) (Johnson et al., 2016). Non-fragmentary genes that were free from putative chimeric sequences were utilized as input for the paragone-nf analysis, which employs the tree-based orthology inference methods originally described by Yang & Smith (2014). These methods have been reported to significantly increase the completeness and accuracy of the inferred orthologs. The orthology inference was carried out using the outgroup-aware strategy "Monophyletic outgroup" (MO), which explicitly accommodates gene and genome duplications events among the ingroups, being especially suitable for clades that have many gene/genome duplications and requires high-quality outgroup taxa that are phylogenetically distinct from the ingroup (Yang & Smith, 2014). The paralog heat map suggests several duplication events have occurred in *Posoqueria* (Figure 8). In addition, previous research have confirmed *Posoqueria* is phylogenetically distinct from the genera used here as outgroups (*Emmenopterys* and *Sipanea*) (Antonelli et al., 2021; Cortes-B & Motley, 2015), so the MO method was regarded appropriate for this study, effectively recovering 177 orthologs.

Systematics

In most cases, each species in the phylogenetic analysis included two or more samples. The gene trees and species tree reconstructions were not constrained to cluster together samples of the same species. Nonetheless, many samples of the same species were clustered in the same clades, thus

suggesting the monophyly of most of the taxa included in this study and supporting the previous morphological alfa-taxonomical classification in the group.

In the original publication of *P. mutisii*, Standley (1936) described the species as resembling *P. longiflora,* but the latter having much longer flowers and broader calyx lobes. The identity of *P. mutisii* has been considered by Taylor (2021) as not entirely clear and maybe not distinct from *P. longiflora*. The type collections of *P. mutisii* "Mutis, 2257" at US and particularly "Mutis, 4947" at MA, unlike *P. longiflora* and *P. latifolia* specimens, have straight flowers with apparently equal filaments and buds lacking the characteristic bend of zygomorphic flowers in the lobe part at the top of the corolla. For these reasons and considering the fully supported relationship of *P. longiflora* as sister to *P. mutisii* + *P. latifolia* (from Central America and Colombia; Figure 9), *P. mutisii* is regarded here as a valid name and as a species different from both *P. latifolia* and *P. longiflora*.

The current circumscription of *P. latifolia* is non-monophyletic, with the French Guianese specimens forming an independent and distant clade from the Central American and Colombian samples (Figure 9). *Posoqueria latifolia* was originally described from French Guiana but is distributed throughout the range of the genus and is variable both in morphology and habitat (Taylor, 2021; Taylor et al., 2011). Based on this new phylogenomic evidence, and considering its morphological and ecological variability, the circumscription of *P. latifolia* should be reexamined, as the plants from Central America and Colombia appear to belong to an undescribed species in the genus.

The clade formed by *P. laevis* (*P. grandiflora* + *P. robusta* + *P. costaricensis* + *P. correana* + *P. grandifructa*) constitutes most of the Central American species in the genus. *Posoqueria laevis* inferred as sister to the rest of the species was the only fully supported bifurcation in this group. The other relationships within the clade are not well resolved, having moderate to weak support, likely indicating a recent diversification that could not be characterized using the universal "Angiosperms353" probe set.

The clade formed by *P. chocoana* (from Panama) as sister to (*P. costaricensis* + *P. coriacea* + *P. chocoana* (from Ecuador) + *P. maxima* + *P. longifilamentosa*) has relationships in conflict with the current taxonomic classification in the group. For instance, the current circumscription of *P. chocoana* is challenged, as the specimens from Panama form a fully supported group, separated from all the plants of this species from Ecuador. In addition, the position of the specimen *P. chocoana* (426) is particularly tricky, as it is not clustered with any of the other samples in the species. Therefore, the classification of *P. chocoana* should be revised, including increased sampling to gather more morphological and molecular evidence to test whether this species is monophyletic. Similarly, all the specimens from *P. costaricensis* are scattered in the phylogeny, indicating the classification of the species should be revisited. Regarding *P. coriacea*, this species could be difficult to separate from *P. latifolia*, especially when these grow in the intermediate areas of their ranges (Taylor, 2021), but the phylogeny in Figure 9 shows these species are phylogenomically distinct. Likewise, *P. maxima* is considered morphologically similar to *P. grandiflora* and *P. williamsii*, and the taxonomy of these large-flowered plants is not fully resolved (Taylor, 2021); here is shown that *P. maxima*, *P. grandiflora*, and *P. williamsii* are phylogenomically different, and *P. maxima* is inferred to be closer to the zygomorphic large-flowered *P. longifilamentosa*.

The evolutionary relationships inferred in this study depended on the phylogenomic resolution allowed by the "Angiosperms353" probe set, which as a universal bait panel is designed for resolving deep phylogenetic nodes, usually at the order or family level (Boer et al., 2022). In this research, several relationships within *Posoqueria* were confidently resolved using these baits, while others were recovered with moderate to weak support. The unresolved relationships could indicate clades with more recent diversification events, which could not be differentiated by employing the relatively conserved genes captured by universal panels. In the future, this limitation

could be solved by designing and using a customized bait panel for *Posoqueria*, or at least for Rubiaceae, which will allow higher phylogenomic resolution through the capture of more variable sequences, higher on-target read ratios, and superior target recovery rates (Boer et al., 2022). Nevertheless, the use of the "Angiosperms353" probes in this research was a cost-effective strategy that yielded interesting insights into the evolutionary patterns in *Posoqueria*, demonstrating the potential of this universal bait set to conduct standardized phylogenomic studies in angiosperms and confirming its reported utility to resolve infrageneric relationships (Frost et al., 2021; Murphy et al., 2020; Ridley, 2022) and even infraspecific variability (Beck et al., 2021).

Floral Evolution

Concerning the flower symmetry in *Posoqueria*, Cortes-B & Motley (2015) reconstructed the actinomorphic species, i.e. species lacking the Pollen Catapult Mechanism (PCM), as a monophyletic group, which would justify the re-circumscription of *Posoqueria* and resurrection of the genus *Stannia,* but the limited sampling and weak phylogenetic support prevented them from drawing conclusions. In the current study, about 80% of the diversity of the genus was included and the phylogenomic evidence showed with good support that the flower symmetry and the Pollen Catapult Mechanism (PCM) linked to it, is a non-monophyletic trait. Therefore, the resurrection of *Stannia,* which in the past included only the species presenting the PCM (Delprete, 2009), is not justified, confirming the synonymization of the two genera. It is worth stating that the genus *Molopanthera,* which could not be included in this study due to problems with DNA extraction, has been recovered as the sister taxon to *Posoqueria*, with phylogenetic evidence showing the PCM mechanism originated in the *Molopanthera-Posoqueria* clade (Cortes-B & Motley, 2015). Consequently, the first transition from actinomorphic to zygomorphic flowers, represented in the phylogeny in Figure 9 as the number 1 (green), likely occurred instead in a common ancestor to the *Molopanthera-Posoqueria* clade.

The frequent occurrence of evolutionary transitions and reversals between actinomorphic and zygomorphic forms linked to the absence and presence of the PCM raises questions about the homology of the underlying developmental processes. These results are similar to what has been found in studies of symmetry expression in Angiosperms as a whole, on which multiple transitions between actinomorphic and zygomorphic flowers are inferred (Naghiloo, 2020). As the number of species with zygomorphic flowers presenting the PCM is similar to the number of species lacking the mechanism and considering that several actinomorphic species were recovered as sister to zygomorphic species (Figure 9), it seems this trait has not been a factor that particularly influenced diversification in the group.

Cortes-B & Motley (2015) extensively described pollination mechanisms in angiosperms similar to the PCM, on which potential mechanical energy is liberated through quick fast movements. They highlighted that these specialized mechanisms for pollen release are commonly recovered as synapomorphies for the groups displaying the trait. As an example, they reported the cases of *Moreae* with inflexed stamens (Zerega et al., 2005), *Stylidium* with the triggered position of the flower column (Laurent et al., 1998), and *Cornus* with the flowers opening explosively and catapulting pollen upwards (Fan & Xiang, 2001). Considering these patterns in angiosperms, the uniqueness of the PCM, and their phylogenetic results, Cortes-B & Motley (2015) regarded the PCM as a synapomorphy for the *Molopanthera-Posoqueria* clade, and the lack of the PCM in some *Posoqueria* species was interpreted as evolutionary loses. Under this alternative explanation, the flower symmetry patterns in the phylogeny in Figure 9 should be understood as the product of one single PCM origin in an ancestor of the *Molopanthera-Posoqueria* clade and at least eight subsequent independent evolutionary loses within *Posoqueria.*

The way in which the symmetry expression patterns in *Posoqueria* are understood also has ecological implications. It has been documented that zygomorphic flowers have fewer potential pollinators, which could cause species with this kind of flowers to face a greater risk of extinction,

as pollinator populations are decreasing globally due to habitat degradation, pesticide use, infections, among other factors (Yoder et al., 2020). For example, Amorim et al. (2014) showed that a low abundance of long-tongued pollinators such as hawkmoths, which have been reported to pollinate some *Posoqueria* species, leads to pollen limitation, making the plants completely dependent on these long-tongued pollinators to set fruits. Future pollination biology research could help to better understand the role of the PCM in the efficiency of pollination in *Posoqueria* and to test whether this specialized pollination mechanism could compensate for the fewer potential pollinators in the zygomorphic species. These kinds of studies could also contribute to the assessment of the endangerment category of species in the group.

CONCLUSIONS

This study constitutes the most widely and densely sampled phylogenomic analysis of *Posoqueria* to date, including 15 species representing about 80% of its diversity, samples from eight countries, and extensive phylogenomic data from 177 molecular markers. A novel view of the phylogenetic relationships within the genus has been generated, opening opportunities for future research in the group, especially from the alpha-taxonomy, developmental genetics, pollination biology, ecology, and biogeography points of view.

The results support the monophyly of most of the species in *Posoqueria* and contribute to clarifying the classification of taxonomically problematic taxa, for instance proving the phylogenomic distinctiveness of species such as *P. mutisii, P. longiflora, P. maxima, P. grandiflora,* and *P. williamsii.* In addition, the current circumscription of *P. chocoana* and *P. costaricensis* is challenged by this phylogenomic evidence, suggesting increased morphological and molecular data should be gathered to verify their monophyletic condition. Finally, this phylogenetic data strongly supports the re-circumscription of *P. latifolia,* confirming previous recorded morphological and ecological variability in this widespread species (Taylor, 2021), indicating that the populations from Central America and Colombia belong to an undescribed taxon in the genus.

Strikingly, the unique Pollen Catapult Mechanism (PCM) linked to zygomorphic flowers is supported by this phylogenomic study as the product of several evolutionary transitions, yielding this character as non-monophyletic and questioning its role in the diversification of the genus. These results also close a long-standing debate in the taxonomy of *Posoqueria* and its nomenclatural synonym *Stannia,* dating back to the 1850s (Delprete, 2009), demonstrating that the resurrection of the latter as a valid taxon is not justified.

Finally, this target capture research exemplifies the trade-off between cost and detail when using the universal “Angiosperms353” probe set, saving money and time by not having to develop a customized bait panel and still resolving several phylogenomic relationships in *Posoqueria,* but losing detail by recovering some relationships with moderate to weak support.

ACKNOWLEDGMENTS

I am grateful to my supervisors Claes Persson and Christine Bacon, and my examiner Bengt Oxelman for their academic guidance and for co-funding this research. I am also thankful to the Axel Adler Scholarship and the Colombian SAPIENCIA scholarship-fund for co-financing my master's studies at the University of Gothenburg, and with the Helge Ax:son Johnsons Scholarship, the Carl Skottsberg Scholarship, and the Idea Wild grant for sponsoring the fieldwork expenses. I appreciate the study facilities and the encouragement provided by the members of the Gothenburg Global Biodiversity Center (GGBC) and thank Adrian Hill, Johan Rova, Juvenal Batista, and Alejandra Carvajal for assisting with the fieldwork and for their academic support. I give thanks to the directors and curators of herbaria ALCB, GB, HUA, JAUM, MEDEL, NY, and MO, and to the Colombian researchers Rocio Cortés and Juan P. Tobón, for granting access to their collections

and donating plant tissues for this research, and particularly to the MO herbarium for the loan of valuable type collections and additional representative specimens of *Posoqueria*. Finally, I thank Felipe Cardona and Heriberto David from HUA, María de Stapf from PMA, Enrique Castro and Orlando Vargas from The Organization for Tropical Studies (OTS), Reinaldo Aguilar from Centro de Diversidad de Plantas Regionales de Osa, and Marvin Hidalgo from Estación Biológica Monteverde for their fundamental collaboration with the fieldwork logistics and with the collection and exportation permits.

Samples were collected in Colombia under the collection permit "Permiso Marco Universidad de Antioquia" No. 0524 from 2014 and exported under the permit No. 3076 from 2022, granted by the Autoridad Nacional de Licencias Ambientales (ANLA); in Costa Rica, the collection and exportation were carried out under the permits No. R-040-2022-OT-CONAGEBIO and R-043-2022-OT-CONAGEBIO, granted by the Oficina Técnica de la Comisión Nacional para la Gestión de la Biodiversidad (CONAGEBIO); in Panama, the collection was performed under the permit No. ARG-0048-2022 and the exportation under the permit No. PA-01-ARB-143-2022, granted by the Ministerio de Ambiente. The data handling and analysis were enabled by resources in projects SNIC 2022/22-467 and NAISS 2023/23-45, provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) at UPPMAX, funded by the Swedish Research Council through grant agreement no. 2022-06725.

DATA AVAILABILITY

All alignments, Newick tree files, figures in high resolution, and supplementary information of this study can be found at Zenodo (https://doi.org/10.5281/zenodo.7896478).

REFERENCES

Amorim, F. W., Wyatt, G. E., & Sazima, M. (2014). Low abundance of long-tongued pollinators leads to pollen limitation in four specialized hawkmoth-pollinated plants in the Atlantic Rain forest, Brazil. *Naturwissenschaften*, *101*(11), 893–905. https://doi.org/10.1007/s00114-014-1230-y

Andermann, T., Torres Jiménez, M. F., Matos-Maraví, P., Batista, R., Blanco-Pastor, J. L., Gustafsson, A. L. S., Kistler, L., Liberal, I. M., Oxelman, B., Bacon, C. D., & Antonelli, A. (2020). A Guide to Carrying Out a Phylogenomic Target Sequence Capture Project. *Frontiers in Genetics*, *10*. https://www.frontiersin.org/articles/10.3389/fgene.2019.01407

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: Http://www.bioinformatics.babraham.ac.uk/projects/fastqc/*.

Antonelli, A. (2022). The rise and fall of Neotropical biodiversity. *Botanical Journal of the Linnean Society*, *199*(1), 8–24. https://doi.org/10.1093/botlinnean/boab061

Antonelli, A., Clarkson, J. J., Kainulainen, K., Maurin, O., Brewer, G. E., Davis, A. P., Epitawalage, N., Goyder, D. J., Livshultz, T., Persson, C., Pokorny, L., Straub, S. C. K., Struwe, L., Zuntini, A. R., Forest, F., & Baker, W. J. (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. https://doi.org/10.1002/ajb2.1697

Bankevich, A., Nurk, S., Nurk, S., Antipov, D., Gurevich, A., Mikhail Dvorkin, Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Beck, J. B., Markley, M. L., Zielke, M. G., Thomas, J. R., Hale, H. J., Williams, L. D., & Johnson, M. G. (2021). *Is Palmer's elm leaf goldenrod real? The Angiosperms353 kit provides within-species signal in Solidago ulmifolia s.l* (p. 2021.01.07.425781). bioRxiv. https://doi.org/10.1101/2021.01.07.425781

Boer, H. de, Rydmark, M. O., Verstraete, B., & Gravendeel, B. (2022). Molecular identification of plants: From sequence to species. *Advanced Books*, *1*, e98875. https://doi.org/10.3897/ab.e98875

Bolger, A., Lohse, M., Usadel, B., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Brewer, G. E., Clarkson, J. J., Maurin, O., Zuntini, A. R., Barber, V., Bellot, S., Biggs, N., Cowan, R. S., Davies, N. M. J., Dodsworth, S., Edwards, S. L., Eiserhardt, W. L., Epitawalage, N., Frisby, S., Grall, A., Kersey, P. J., Pokorny, L., Leitch, I. J., Forest, F., & Baker, W. J. (2019). Factors Affecting Targeted Sequencing of 353 Nuclear Genes From Herbarium Specimens Spanning the Diversity of Angiosperms. *Frontiers in Plant Science*, *10*. https://www.frontiersin.org/articles/10.3389/fpls.2019.01102

Burger, W., & Taylor, C. M. (1993). *Flora Costaricensis: Family 202 Rubiaceae. Field museum of natural history.* (pp. 1–352). Field Museum of Natural History. https://www.biodiversitylibrary.org/item/20372

Busch, A., & Zachgo, S. (2009). Flower symmetry evolution: Towards understanding the abominable mystery of angiosperm radiation. *BioEssays*, *31*(11), 1181–1190. https://doi.org/10.1002/bies.200900081

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*(1), 421. https://doi.org/10.1186/1471-2105-10-421

Cheek, M., Nic Lughadha, E., Kirk, P., Lindon, H., Carretero, J., Looney, B., Douglas, B., Haelewaters, D., Gaya, E., Llewellyn, T., Ainsworth, A. M., Gafforov, Y., Hyde, K., Crous, P., Hughes, M., Walker, B. E., Campostrini Forzza, R., Wong, K. M., & Niskanen, T. (2020). New scientific discoveries: Plants and fungi. *PLANTS, PEOPLE, PLANET*, *2*(5), 371–388. https://doi.org/10.1002/ppp3.10148

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Cortes-B, R., & Motley, T. J. (2015). Phylogeny of the Henriquezieae-Posoquerieae-Sipaneeae, a Guayanan-centered clade of Rubiaceae: Implications for morphological evolution. *Phytotaxa*, *206*(1), Article 1. https://doi.org/10.11646/phytotaxa.206.1.12

Delprete, P. G. (2009). Taxonomic History, Morphology, and Reproductive Biology of the Tribe Posoquerieae (Rubiaceae, Ixoroideae)1. *Annals of the Missouri Botanical Garden*, *96*(1), 79–89. https://doi.org/10.3417/2006192

Delprete, P. G., & Jardim, J. (2012). Systematics, taxonomy and floristics of Brazilian Rubiaceae: An overview about the current status and future challenges. *Rodriguésia*, *63*, 101–128. https://doi.org/10.1590/S2175-78602012000100009

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Fan, C., & Xiang, (Jenny) Qiu-Yun. (2001). Phylogenetic relationships within Cornus (Cornaceae) based on 26S rDNA sequences. *American Journal of Botany*, *88*(6), 1131–1138. https://doi.org/10.2307/2657096

Frost, L., Santamaría-Aguilar, D. A., Singletary, D., & Lagomarsino, L. P. (2021). *Niche evolution of the Neotropical tree genus Otoba in the context of global biogeography of the nutmeg family, Myristicaceae* (p. 2020.10.02.324368). bioRxiv. https://doi.org/10.1101/2020.10.02.324368

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D. B., Lander, E. S., & Nusbaum, C. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology*, *27*(2), Article 2. https://doi.org/10.1038/nbt.1523

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

Hosner, P. A., Faircloth, B. C., Glenn, T. C., Braun, E. L., & Kimball, R. T. (2016). Avoiding Missing Data Biases in Phylogenomic Inference: An Empirical Study in the Landfowl (Aves: Galliformes). *Molecular Biology and Evolution*, *33*(4), 1110–1125. https://doi.org/10.1093/molbev/msv347

Jiang, X., Edwards, S. V., & Liu, L. (2020). The Multispecies Coalescent Model Outperforms Concatenation Across Diverse Phylogenomic Data Sets. *Systematic Biology*, *69*(4), 795–812. https://doi.org/10.1093/sysbio/syaa008

Johnson, M. G., Gardner, E. M., Liu, Y., Medina, R., Goffinet, B., Shaw, A. J., Zerega, N. J. C., & Wickett, N. J. (2016). HybPiper: Extracting Coding Sequence and Introns for Phylogenetics from High-Throughput Sequencing Reads Using Target Enrichment. *Applications in Plant Sciences*, *4*(7). https://doi.org/10.3732/apps.1600016

Johnson, M. G., Pokorny, L., Dodsworth, S., Botigué, L. R., Cowan, R. S., Devault, A., Eiserhardt, W. L., Epitawalage, N., Forest, F., Kim, J. T., Leebens-Mack, J., Leitch, I. J., Maurin, O., Soltis, D. E., Soltis, P. S., Wong, G. K.-S., Baker, W. J., & Wickett, N. J. (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. https://doi.org/10.1093/sysbio/syy086

Junier, T., & Zdobnov, E. M. (2010). The Newick utilities: High-throughput phylogenetic tree processing in the Unix shell. *Bioinformatics*, *26*(13), 1669–1670. https://doi.org/10.1093/bioinformatics/btq243

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), Article 6. https://doi.org/10.1038/nmeth.4285

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Laurent, N., Bremer, B., & Bremer, K. (1998). Phylogeny and Generic Interrelationships of the Stylidiaceae (Asterales), with a Possible Extreme Case of Floral Paedomorphosis. *Systematic Botany*, *23*(3), 289–304. https://doi.org/10.2307/2419506

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Lorence, D. H., & Taylor, C. M. (2012). Rubiaceae. In *Flora mesoamericana: Rubiaceae a Verbenaceae. Natural History Museum.*

Ly, S. N., Garavito, A., Block, P. D., Asselman, P., Guyeux, C., Charr, J.-C., Janssens, S., Mouly, A., Hamon, P., & Guyot, R. (2020). Chloroplast genomes of Rubiaceae: Comparative genomics and molecular phylogeny in subfamily Ixoroideae. *PLOS ONE*, *15*(4), e0232295. https://doi.org/10.1371/journal.pone.0232295

Macias, L. (1998). *Revisão taxonomica do genero Posoqueria Aubl. (Rubiaceae). Master's thesis.*

Mai, U., & Mirarab, S. (2018). TreeShrink: Fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genomics*, *19*(Suppl 5), 272. https://doi.org/10.1186/s12864-018-4620-2

McLay, T. G. B., Birch, J. L., Gunn, B. F., Ning, W., Tate, J. A., Nauheimer, L., Joyce, E. M., Simpson, L., Schmidt-Lebuhn, A. N., Baker, W. J., Forest, F., & Jackson, C. J. (2021). New targets acquired: Improving locus recovery from the Angiosperms353 probe set. *Applications in Plant Sciences*, *9*(7). https://doi.org/10.1002/aps3.11420

Meade, A., & Pagel, M. (2018). *BayesTraits: A computer package for analyses of trait evolution. Available at http://www.evolution.rdg.ac.uk/BayesTraitsV3.0.1/BayesTraitsV3.0.1.html*.

Mendes, F. K., & Hahn, M. W. (2018). Why Concatenation Fails Near the Anomaly Zone. *Systematic Biology*, *67*(1), 158–169. https://doi.org/10.1093/sysbio/syx063

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, *37*(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Mirarab, S. (2019). *Species Tree Estimation Using ASTRAL: Practical Considerations*.

Murphy, B., Forest, F., Barraclough, T., Rosindell, J., Bellot, S., Cowan, R., Golos, M., Jebb, M., & Cheek, M. (2020). A phylogenomic analysis of Nepenthes (Nepenthaceae). *Molecular Phylogenetics and Evolution*, *144*, 106668. https://doi.org/10.1016/j.ympev.2019.106668

Naghiloo, S. (2020). Patterns of Symmetry Expression in Angiosperms: Developmental and Evolutionary Lability. *Frontiers in Ecology and Evolution*, *8*. https://www.frontiersin.org/articles/10.3389/fevo.2020.00104

Negi, A., Shukla, A., Jaiswar, A., Shrinet, J., & Jasrotia, R. S. (2022). Chapter 6—Applications and challenges of microarray and RNA-sequencing. In D. B. Singh & R. K. Pathak (Eds.), *Bioinformatics* (pp. 91–103). Academic Press. https://doi.org/10.1016/B978-0-323-89775-4.00016-X

POWO. (2023). *Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. Website: Https://powo.science.kew.org/ [accessed 1 May 2023]* [Data set].

Puff, C., Igersheim, A., Buchner, R., & Rohrhofer, U. (1995). The United Stamens of Rubiaceae. Morphology, Anatomy; Their Role in Pollination Ecology. *Annals of the Missouri Botanical Garden*, *82*(3), 357–382. https://doi.org/10.2307/2399888

Rambaut, A. (2018). *FigTree v1.4.4, a graphical viewer of phylogenetic trees. Computer Program Distributed by the Author, Website: Http://tree.bio.ed.ac.uk/software/figtree/*.

Reveal, J. L. (2012). *Newly required infrafamilial names mandated by changes in the Code of Nomenclature For Algae, Fungi, and Plants. Phytoneuron 2012-33: 1–32. Published 9 April 2012. ISSN 2153 733X*.

Ridley, R. (2022). *Species delimitation in Genipa (Rubiaceae) using integrative taxonomy* [University of Gothenburg]. https://gupea.ub.gu.se/bitstream/handle/2077/73476/BIO707%20Rhonda%20Ridley%20vt22.pdf?sequence=1

Sayyari, E., Whitfield, J. B., & Mirarab, S. (2017). Fragmentary Gene Sequences Negatively Impact Gene Tree and Species Tree Reconstruction. *Molecular Biology and Evolution*, *34*(12), 3279–3291. https://doi.org/10.1093/molbev/msx261

Slater, G. S. C., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, *6*(1), 31. https://doi.org/10.1186/1471-2105-6-31

Smith, S. A., & Dunn, C. W. (2008). Phyutility: A phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics*, *24*(5), 715–716. https://doi.org/10.1093/bioinformatics/btm619

Standley, P. C. (1936). *Publications of the Field Museum of Natural History, Botanical Series 11(5): 232. Studies of American plants-VI* (Vol. 11, pp. 1–140). Field Museum of Natural History. https://www.biodiversitylibrary.org/item/19793

Taylor, C. M. (2021, continuously updated). *Rubiaceae Project Selected Rubiaceae Tribes and Genera*. http://legacy.tropicos.org/projectwebportal.aspx?pagename=homeIntro&projectid=34

Taylor, C. M., Hammel, B., & Gereau, R. (2011). Rubiacearum Americanarum Magna Hama Pars XXVII: Six New Species and a New Taxonomic View of Posoqueria. *Novon*, *21*, 118–132. https://doi.org/10.2307/23018507

Thiers, B. (2023). *Index Herbariorum: A Global Directory of Public Herbaria and Associated Staff. New York Botanical Garden's Virtual Herbarium. Http://sweetgum.nybg.org/science/ih/ (Accessed Mar, 2023).*

Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S., Barker, M. S., Burleigh, J. G., Gitzendanner, M. A., Ruhfel, B. R., Wafula, E., Der, J. P., Graham, S. W., Mathews, S., Melkonian, M., Soltis, D. E., Soltis, P. S., Miles, N. W., … Leebens-Mack, J. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences*, *111*(45), E4859–E4868. https://doi.org/10.1073/pnas.1323926111

Wikström, N., Bremer, B., & Rydin, C. (2020). Conflicting phylogenetic signals in genomic data of the coffee family (Rubiaceae). *Journal of Systematics and Evolution*, *58*(4), 440–460. https://doi.org/10.1111/jse.12566

Xu, B., & Yang, Z. (2016). Challenges in Species Tree Estimation Under the Multispecies Coalescent Model. *Genetics*, *204*(4), 1353–1368. https://doi.org/10.1534/genetics.116.190173

Yang, Y., & Smith, S. A. (2014). Orthology Inference in Nonmodel Organisms Using Transcriptomes and Low-Coverage Genomes: Improving Accuracy and Matrix Occupancy for Phylogenomics. *Molecular Biology and Evolution*, *31*(11), 3081–3092. https://doi.org/10.1093/molbev/msu245

Yoder, J. B., Gomez, G., & Carlson, C. J. (2020). Zygomorphic flowers have fewer potential pollinator species. *Biology Letters*, *16*(9), 20200307. https://doi.org/10.1098/rsbl.2020.0307

Yu, Y., Blair, C., & He, X. (2020). RASP 4: Ancestral State Reconstruction Tool for Multiple Genes and Characters. *Molecular Biology and Evolution*, *37*(2), 604–606. https://doi.org/10.1093/molbev/msz257

Zerega, N. J. C., Clement, W. L., Datwyler, S. L., & Weiblen, G. D. (2005). Biogeography and divergence times in the mulberry family (Moraceae). *Molecular Phylogenetics and Evolution*, *37*(2), 402–416. https://doi.org/10.1016/j.ympev.2005.07.004

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, *19*(6), 153–153. https://doi.org/10.1186/s12859-018-2129-y

APPENDIX 1. Popular Science Summary: *Plants with A Catapult in The Flower*

Have you ever seen a plant using its flowers as a ballistic mechanism to launch the pollen as a projectile? Well, *Posoqueria*, a group of tropical trees relative to the famous coffee plant, have a unique system in its male reproductive structures, the stamens, which functions as a living catapult. Some of the species in this group of plants lack this catapult mechanism, so several experts in the past have suggested the species having the mechanism should be classified in a different group, separated from the species lacking this kind of biological catapult. Here, I reconstructed the evolutionary history of these plants using information coming from the DNA, which allowed me to test whether *Posoqueria* should be divided in two genera based on the presence-absence of the flower catapult and address other classification problems in the group.

So, Is *Posoqueria* Another Kind of Coffee Plant*?*

Posoqueria is a type of plants that grow in Central and South America. It belongs to a botanical family called Rubiaceae, which includes coffee plants! Rubiaceae is a big group of plants with lots of different types. Scientists have found over 14,000 species of Rubiaceae, but there are probably even more out there that we don't know about yet. Rubiaceae are key components of the ecosystems in the tropics, providing nectar and fruits for different kinds of animals (Figure 10).

Figure 10. *Posoqueria* and pollinator. Left: Hawk moth, ready to lick the sweet nectar (Image source: https://en.wikipedia.org/wiki/Sphingidae). Right: *Posoqueria* flower with the catapult mechanism ready to be triggered.

Splitting *Posoqueria* in Two?

From the 1850s until the modern times, several scientists have thought that some *Posoqueria* plants might be split into a different group called *Stannia*. They based this on the absence or presence of the catapult mechanism, which allows the plants to explosively throw the pollen with great velocity at the pollinators head, thus improving its pollination efficiency by ensuring that the pollinator will carry the pollen stuck in the head to the next flower it visits. Nonetheless, this new study used more advanced methods to look at the genes of 15 different *Posoqueria* species and found that them are actually closely related, so there is not evolutionary justification to split *Posoqueria* in two groups.

What Else This Study Tells Us?

The study found that some *Posoqueria* plants, as *P. mutisii*, *P. longiflora*, *P. maxima*, *P. grandiflora*, and *P. williamsii*, are genetically different from other types, removing previous doubts about their uniqueness. It also suggested that one kind of *Posoqueria* found in Central America and Colombia is mistakenly called *P. latifolia*, as this might actually be a new, unnamed species for science!

What's next for *Posoqueria*?

There is still a lot more to learn! Now we have more questions about the evolurionary relationships within *Posoqueria* and the role the catapult mechanism played on its diversification. There might even be new species of *Posoqueria* that we haven't seen before!

APPENDIX 2. Phylogeny of *Posoqueria* inferred in ASTRAL-III from 177 gene trees. Local posterior probability values are presented in the nodes. Each terminal is labeled with the species name, sample ID, and acronym of collection country (BRA:Brazil, COL:Colombia, CRI:Costa Rica, ECU:Ecuador, FRG:French Guiana, NIC:Nicaragua, PAN:Panama, PER:Peru, s.l:unknown locality)