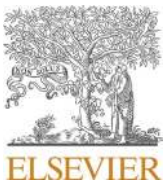
Molecular Phylogenetics and Evolution 208 (2025) 108338



Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev





Phylogeny of Palicoureeae (Rubiaceae) based on 353 low-copy nuclear genes – with particular focus on *Hymenocoleus* Robbr.

Lovisa Thilén [a,*], Olivier Lachenaud [b,c], Olle Thureborn [a], Sylvain G. Razafimandimbison [d], Catarina Rydin [a]

[a] Department of Ecology, Environment and Plant Sciences, Stockholm University, 106 91 Stockholm, Sweden
[b] Meise Botanic Garden, Nieuwelaan 38, 1860 Meise, Belgium
[c] Herbarium et Bibliothèque de Botanique Africaine, CP 265, Université Libre de Bruxelles, Boulevard du Triomphe, 1050 Brussels, Belgium
[d] Department of Botany, The Swedish Museum of Natural History, Box 50007, 104 05 Stockholm, Sweden

ARTICLE INFO

Keywords:
Angiosperms353
Incomplete lineage sorting
Phylogenomics
Puffia
Pyrenes
Species tree estimation
Target capture

ABSTRACT

Members of the tribe Palicoureeae of the coffee family (Rubiaceae) have a complex taxonomic history and have been the focus of few modern systematic studies. The tribe comprises about 1,100 tropical species in ten genera. To investigate phylogeny, we used a target capture approach and the angiosperm-wide Angiosperms353 bait set to produce genomic data for a representative taxon sample of Palicoureeae, with particular focus on the African genus *Hymenocoleus*. Using coalescent-based inference methods, we find that *Puffia gerrardii* (recently separated from *Geophila*) is sister to *Hymenocoleus*. The deepest split in *Hymenocoleus* is highly affected by incomplete lineage sorting, possibly as a consequence of rapid speciation during the early evolution of the clade. Remaining interspecific relationships in *Hymenocoleus* could be confidently resolved and while Robbrecht's traditional infrageneric classification scheme based on floral features is not supported as reflecting evolution in the group, we find that several other features do, e.g. characters of pyrenes and involucral cups. Although not free of challenges, a strong advantage with our analytical approach is that gene tree heterogeneity can be taken into account. Including flanking regions yielded data sets that had the strongest power to reject polytomies and produced less gene tree error, resulting in species trees with higher normalised quartet scores and higher average support compared to trees inferred only from exon data. Presumably paralogous loci are often filtered out prior to species tree estimation but we find that they may contribute important phylogenetic information when using an inference method that actively accounts for them.

1. Introduction

Palicoureeae Robbr. & Manen (Figs. 1–2) are a pantropical tribe in the coffee family, Rubiaceae Juss. They belong to subfamily Rubioideae Verdc., and within this group to the informal "Psychotrieae alliance" (Razafimandimbison and Rydin, 2024b) where they are sister to Psychotrieae Cham. & Schltdl. The distinction between the two tribes is mainly based on molecular data, although Palicoureeae usually differ from Psychotrieae (with some exceptions) by their persistent or marcescent (vs. usually caducous) stipules, the presence of preformed germination slits in their pyrenes (which are mostly absent in Psychotrieae), and the absence of a red pigment in the seed coat (Robbrecht and Manen, 2006). Based on morphology, the genera now included in these sister tribes were previously included in a very broadly defined Psychotrieae, together with genera such as *Gaertnera* Lam. (now in Gaertnereae Bremek. ex S.P.Darwin) and *Declieuxia* Kunth (now in Coussareeae Hook.f.) (e.g. Robbrecht, 1988). A more narrowly delimited Psychotrieae sensu Bremer and Manen (2000) was found to consist of two large lineages in several early molecular phylogenetic studies (e.g. Andersson and Rova, 1999; Nepokroeff et al., 1999), which were later separated as Palicoureeae and Psychotrieae s.str. (Robbrecht and Manen, 2006), a conclusion supported in subsequent work (e.g. Razafimandimbison et al., 2008; Rydin et al., 2009; Razafimandimbison et al., 2014; Thureborn et al., 2022).

Palicoureeae encompass at least 1,100 species, which are classified in ten genera: *Carapichea* Aubl. (Fig. 1F), *Chassalia* Comm. ex Poir. (Fig. 1C), *Eumachia* DC. (Fig. 1A), *Geophila* D.Don (Fig. 1D), *Hymenocoleus* Robbr. (Fig. 2), *Notopleura* (Benth.) Bremek., *Palicourea* Aubl.

* Corresponding author.
E-mail address: lovisa.thilen@su.se (L. Thilén).

https://doi.org/10.1016/j.ympev.2025.108338
Received 13 January 2025; Received in revised form 18 March 2025; Accepted 20 March 2025
Available online 28 March 2025
1055-7903/© 2025 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).



Fig. 1. Examples of diversity in Palicoureeae. A. *Eumachia abrupta* (Hiern) J.H.Kirkbr., photo: Bart Würsten. B. *Puffia gerrardii* (Baker) Razafim. & B.Bremer., photo: Sylvain G. Razafimandimbison. C. *Chassalia zenkeri* K.Schum. & K.Krause., photo: Vincent Droissart. D. *Geophila obvallata* (Schumach. & Thonn.) Didr., photo: Jérôme Degreef. E. *Palicourea adusta* Standl., photo: Jan Thomas Johansson. F. *Carapichea guianensis* Aubl., photo: Olivier Lachenaud.



Fig. 2. Examples of morphological diversity in *Hymenocoleus* Robbr. A. *H. nervopilosus* Robbr., photo: Olivier Lachenaud. B. Fruits of *H. globulifer* Robbr., photo: Ehoarn Bidault. C. Axillary flowers of *H. axillaris* Robbr., photo: Ehoarn Bidault. D. Terminal flowers of *H. melanochlora* (K.Schum.) O.Lachenaud, photo: Steven Dessein. E. Terminal flowers of *H. libericus* (A.Chev. ex Hutch. & Dalziel) Robbr., photo: Olivier Lachenaud. F. Involucral cup and flowers of *H. scaphus* (K.Schum.) Robbr., photo: Bart Würsten.

(Fig. 1E), *Puffia* Razafim. & B.Bremer (Fig. 1B), *Rudgea* Salisb., and *Tromlyca* Borhidi (Razafimandimbison and Rydin, 2024b, a). The tribe includes many former *Psychotria* L. species; the heterogeneous nature of this genus had long been recognised (e.g. Nepokroeff et al., 1999; Andersson, 2002; Robbrecht and Manen, 2006). *Carapichea*, *Notopleura*, *Palicourea*, *Rudgea*, and *Tromlyca* are exclusively neotropical, *Eumachia* and *Geophila* are pantropical, *Chassalia* has a wide paleotropical distribution, while *Puffia* and *Hymenocoleus* are restricted to Madagascar and continental Africa, respectively (POWO, 2024). Phylogenetic studies based on molecular data have found *Chassalia*, *Geophila, Hymenocoleus,* and *Puffia* to be closely related (e.g. Andersson, 2001, 2002; Barrabé et al., 2012; Razafimandimbison et al., 2014).

The latter two genera, *Hymenocoleus* and *Puffia*, are the taxa of major interest in this study. Both have often been confused with *Geophila* due to their usually creeping and herbaceous habit (although some species of *Hymenocoleus* may be erect and somewhat woody at the base), capitate inflorescences, and small white flowers with straight corolla tube.

Puffia is monospecific and endemic to Madagascar. It was recently described to accommodate the Malagasy species *Geophila gerrardii* Baker, which was shown to be distantly related to the remaining *Geophila* (Razafimandimbison et al., 2014). It differs from *Geophila* in fruit colour, in being heterodistylous (Vanthournout, 2002), and in the position of the stamens and style (Table 1). *Puffia* was resolved as sister to *Hymenocoleus* in Razafimandimbison et al. (2014), but the limited sampling of the latter genus (two samples representing two species) prevented unambiguous conclusions.

Hymenocoleus (Fig. 2) is a small genus comprising 19 species (Lachenaud, in press) of herbs or undershrubs growing on the forest floor of African tropical rainforests. It is mostly found in the Guineo-Congolian region (following White, 1993) of west and central Africa (including the Gulf of Guinea islands) although a few species extend further east into Uganda and Tanzania in east Africa. The highest species diversity occurs in the Lower Guinea subregion (Lachenaud, in press). *Hymenocoleus* differs from *Geophila* and *Puffia* in several vegetative and reproductive features (Table 1). A diagnostic character of the genus (from which the generic name is derived) is the presence of a membranous sheath inside the stipules, which surrounds the stem and often persists even when the stipules are caducous by fragmentation (as is frequently the case). In *Puffia* and *Geophila* this sheath is absent; instead, they present a row of colleters inside the stipules. The inflorescences are borne terminally with a few exceptions (axillary in two species) and are often surrounded by an involucre of free or fused bracts. The corolla is white (or rarely pale yellow) with a floral tube either funnel-shaped or cylindrical, and stigmas are always bifid. The fruit is a drupe, usually orange in colour (sometimes red in *H. brevicalyx* O.Lachenaud (Lachenaud, in press)), a character that is shared with *Puffia* (Razafimandimbison et al., 2014), while fruits of *Geophila* are red, blue, or black. It includes two pyrenes, which are dorsally smooth (unlike *Geophila* and *Puffia*) and have one or two ventral grooves.

Table 1
Morphological comparison between *Geophila*, *Puffia* and *Hymenocoleus*.

Characters	*Geophila*	*Puffia*	*Hymenocoleus*
sheath inside stipules	absent	absent	present
leaf venation	(sub)palmate	palmate	pinnate (rarely subpalmate)
flowers*	homostylous with A + S both included	heterodistylous with either A or S exserted	heterodistylous with A or S exserted (*H.* sect. *Hymenocoleus*), or both included (*H.* sect. *Gynandrocryptus*), or homostylous with A + S both included (*H.* sect. *Solenocalyx*)**
fruit colour	red, blue, or black	orange	orange (rarely red)
dorsal side of pyrenes	ridged and/or verrucose	ridged and slightly verrucose	smooth
distribution	pantropical	Madagascar	continental Africa

* Abbreviations: A = anthers; S = style.
** Sections according to Robbrecht (1977).

The genus *Hymenocoleus* – first tentatively delimited under the provisional name *Geobenthamia* (Petit, 1963) – was described by Robbrecht (1975) when he separated it from *Geophila.* He then recognised eight species, of which three were newly described and the remainder transferred from *Geophila*, *Psychotria*, and *Uragoga* Baill. (Robbrecht, 1975). Shortly afterwards, Robbrecht (1977) presented a new revision of the genus, now recognising 12 species, with four new species and two new combinations, and recognised three sections in the genus, mostly based on characters of floral organisation and pollen structure. The typical section, *Hymenocoleus* sect. *Hymenocoleus*, includes *H. axillaris* Robbr., *H. glaber* Robbr., *H. hirsutus* (Benth.) Robbr., *H. nervopilosus* Robbr., and *H. neurodictyon* (K. Schum.) Robbr. These species have heterodistylous flowers with either anthers or style exserted, and their pollen is dimorphic, being nonaperturate and granulate in brevistylous flowers, and 3 (–5)-porate and reticulate in longistylous flowers (when known). *Hymenocoleus* sect. *Gynandrocryptus* Robbr. includes *H. libericus* (A.Chev. ex Hutch. & Dalziel) Robbr., *H. rotundifolius* (A.Chev. ex Hepper) Robbr., *H. subipecacuanha* (K. Schum.) Robbr., and (with some doubt) *H. multinervis* Robbr. They have heterodistylous flowers with anthers and style both included; their pollen is also dimorphic, nonaperturate and reticulate in brevistylous flowers, and 3–4-porate and reticulate in longistylous flowers. Finally, species of *H.* sect. *Solenocalyx* Robbr. – which includes *H. barbatus* Robbr., *H. globulifer* Robbr., and *H. scaphus* (K. Schum.) Robbr. – have homostylous flowers with anthers and style both included, and (2–)3(–5)-porate and reticulate pollen. Features of the involucre, the calyx and the pyrenes vary within the sections (Robbrecht, 1977).

The African genera of Palicoureeae, including *Hymenocoleus*, have recently undergone a taxonomic revision (Lachenaud, in press), which includes the description of six new *Hymenocoleus* species and the transfer of one species to *Hymenocoleus* from *Geophila* (*H. bracteosus* O.Lachenaud, *H. brevicalyx* O.Lachenaud, *H. campanulatus* O.Lachenaud, *H. emarginatus* (K.Krause) O.Lachenaud, *H. pseudoscaphus* O.Lachenaud, *H. robbrechtianus* O.Lachenaud, and *H. villosissimus* O.Lachenaud), leaving the total number of species in the genus at 19. For priority reasons and based on the basionym *Uragoga melanochlora* K.Schum., Lachenaud (in press) chose to replace the name *Hymenocoleus rotundifolius* with *Hymenocoleus melanochlora* in his revision. The sections proposed by Robbrecht (1977) are not maintained in this revision, in part because the flowers of several species are insufficiently known, and in part because two species of *H.* sect. *Gynandrocryptus* share a putatively important morphological character with two species in *H.* sect. *Solenocalyx*: an involucral cup surrounding the inflorescence (Lachenaud, in press).

Since the work of Robbrecht (1975; 1977) the genus has received very limited attention until the current revision (Lachenaud, in press), and its infrageneric phylogenetic relationships have never been investigated. Molecular data, restricted to a few markers of a single species (*H. hirsutus*), have been used in a number of studies that investigated phylogenetic relationships in different parts of Rubiaceae (e.g. Nepokroeff et al., 1999; Andersson, 2002; Robbrecht and Manen, 2006; Barrabé et al., 2012) and a few studies included data of *H. scaphus* in addition to *H. hirsutus* (Andersson, 2001; Razafimandimbison et al., 2014; Razafimandimbison et al., 2017). Target-captured nuclear genomic data have only been generated for a single species, *H. neurodictyon,* within the Plant and Fungal Trees of Life project (PAFTOL; Baker et al., 2022). Since no previous study has included more than two species of *Hymenocoleus* in its taxon sample, it has not been possible to assess relationships within the genus.

Here we produced sequences for hundreds of nuclear genes using a target capture approach and the Angiosperms353 bait set (Johnson

et al., 2019) for 16 out of the above-mentioned 19 species of *Hymenocoleus,* and in addition for *Puffia gerrardii* and selected species in the remaining Palicoureeae, including *Geophila* and *Chassalia*. From these data we inferred a species tree using coalescent-based inference methods, and evaluated the robustness of the results to the use of different types of sequence data (protein-coding and protein-coding + flanking non-coding regions) and different data filtering strategies. The main objective of the present study was to analyse phylogenetic relationships in Palicoureeae based on molecular data, with particular focus on the *Hymenocoleus-Puffia* clade, and 1) rigorously assess the monophyly of *Hymenocoleus* and elucidate the hitherto unknown infrageneric relationships in the genus, 2) test the monophyly of the sections defined by Robbrecht (1977), 3) assess the previously indicated sister relationship between *Hymenocoleus* and *Puffia*, and 4) discuss some morphological characters in *Hymenocoleus* in relation to the phylogenetic results.

2. Materials and methods

2.1. Taxon sampling

We sampled 11 out of the 12 *Hymenocoleus* species described by Robbrecht (Robbrecht, 1975, 1977) and five of the seven additional species recognised by Lachenaud in his revision (Lachenaud, in press), a total of 16 species. Further, we included two of the three varieties of *H. neurodictyon* recognised by Lachenaud (in press). Species missing from the analysis are discussed in 4.2.2. Where plant material was available, we sampled two individuals per taxon, and four samples were included for the highly variable taxon *H. neurodictyon* var. *orientalis* (Verdc.) Robbr. The final sampling comprised 25 *Hymenocoleus* samples. When sampling multiple individuals per taxon we aimed to sample from different parts of the distribution range. The putative sister of *Hymenocoleus*, *Puffia gerrardii,* was represented by two samples. The sampling of the rest of Palicoureeae included at least one sample of seven of the remaining eight genera. As outgroups we selected one representative from the sister tribe Psychotrieae and one from the tribe Schradereae, also in the Psychotrieae alliance (Razafimandimbison and Rydin, 2024b). In total 44 samples representing 34 species were included in the study.

For five samples raw sequence data files were downloaded from the European Nucleotide Archive (ENA) at EMBL-EBI. Remaining samples were silica-dried samples from previous field collections or extracted from herbarium specimens. Sample information, including taxon names, herbarium voucher information, collection country and year, and ENA accession numbers, is provided in Table S1.

2.2. DNA extraction

DNA was extracted from up to 25 mg of dried leaf material (silica-dried or herbarium material) using a CTAB protocol (Doyle and Doyle, 1987). For some silica-dried samples, the DNeasy® Plant Mini Kit (Qiagen, Hilden, Germany) was used according to the manufacturer's protocol with the following modifications: the lysis incubation step at 65 °C was extended to 30 min and the elution step was divided into two with 2 x 50 μL elution buffer and 2 x 10–15 min incubation.

2.3. Library preparation

Prior to library preparation the samples were run on an agarose gel to assess the degree of fragmentation, and samples with a large proportion of fragments longer than 350 bp were sonicated to circa 350 bp on a Covaris E220 focused-ultrasonicator (Covaris, Woburn, MA, USA) at the Science for Life Laboratory (Solna, Sweden). Libraries were prepared using a modified version of the Meyer and Kircher (2010) protocol, following Thureborn et al. (2022) with the following modifications: 5 μL library template was used per amplification reaction and the amplified libraries were cleaned using a 0.8–1:1 ratio of AMPure XP beads (Beckman Coulter, Indianapolis, IN, United States) to reaction volume. Non-sonicated samples with high initial concentration were cleaned twice before library preparation with AMPure XP beads using a 0.9:1 ratio of AMPure XP beads to sample volume.

2.4. Target enrichment

Prior to target enrichment libraries were combined in pools of 8–18 libraries. Libraries for pooling were primarily selected based on length, with one pool consisting of libraries from very degraded herbarium material much shorter than the rest. The remainder were pooled so as to minimise the taxonomic breadth within each pool. The input in nanograms was the same for each library within a pool, pooling up to 100 ng per library. The resulting pools of 600 to 1,800 ng DNA were dried completely, resuspended in 7.5 μL water and then enriched with the myBaits Expert Predesigned Panel (Arbor Biosciences, Ann Arbor, MI, USA) Angiosperms353 v1 (Catalog no 308196; Johnson et al. 2019) with v4 chemistry following the MYBaits Standard protocol v5 (https://arborbiosci.com/mybaits-manual/). The blockers mix was assembled according to v4 of the manual since we were working with v4 chemistry. Hybridisation took place at 62 °C for 24 h. The enriched libraries were reamplified for 12–14 cycles with the KAPA HiFi (2X) HotStart ReadyMix PCR Kit (Roche, Basel, Switzerland) and IS5_reamp.P5 and IS6_reamp.P7 primers (Meyer and Kircher, 2010) and then cleaned with AMPure XP beads using a 0.9:1 ratio of beads to sample volume.

2.5. Sequencing and sequence data pre-processing

The samples were multiplexed with 78 samples in one pool and sequenced on both lanes of the flow cell on a NovaSeq 6000 with a 151 bp paired-end read setup using the 'NovaSeqStandard' workflow in 'SP' mode at the Science for Life Laboratory (Solna, Sweden). 39 of these 78 samples were used in the present study. The Bcl to FastQ conversion was performed using bcl2fastq_v2.20.0.422 from the CASAVA software suite, also at the Science for Life Laboratory. Duplicate reads were removed using fastp v0.23.4 (Chen et al., 2018; Chen, 2023), setting the level of accuracy of calculating duplication to 4 and disabling all trimming and filtering enabled by default. The deduplicated raw reads were trimmed using fastp v0.23.4, with adapter sequence auto-detection enabled, base correction enabled with default parameters, and quality filtering and poly-G trimming enabled by default. Bases were trimmed from the front of reads if they had a quality $Q < 15$, and a sliding window approach (fastp's –cut_right option) was used to trim the ends of reads when the mean quality in the window dropped below Q15, with a window size of 4 bp. Reads shorter than 36 bp after trimming were discarded.

2.6. Sequence assembly

We used HybPiper v2.1.2 (Johnson et al., 2016) to assemble deduplicated and quality-trimmed reads into sequences. Unpaired reads were included (provided with the –unpaired flag to the HybPiper assemble command). We used the target file from Thureborn et al. (2022) and through the HybPiper pipeline reads were mapped to the nucleotide target sequences using bwa v0.7.17 (Li and Durbin, 2009), assembled using Spades v3.15.5 (Bankevich et al., 2012), and coding sequences extracted with exonerate v2.4.0 (Slater and Birney, 2005). The HybPiper assemble command was run with the –run_intronerate flag in order to extract so-called 'supercontigs' in addition to the exon sequences (constructed from all assembled contigs that overlap, and potentially comprising intronic and intergenic sequences in addition to the exons (Johnson et al., 2016)). HybPiper was used to retrieve exon sequences of the targeted genes and corresponding supercontigs for all samples included in the study.

The HybPiper command paralog_retriever was run to get information

about paralog warnings across all samples and to retrieve all assembled copies of target regions including potential paralogs. HybPiper gives two kinds of paralog warnings: for 'long paralogs' when there are multiple contigs covering at least 75 % of the target sequence, and when the sequence depth is more than one for at least 75 % of the target sequence (but the assembled contigs not necessarily assembled into long contigs, at least 75 % of the target length). Only sequences flagged as long paralogs are outputted by the paralog_retriever command, and this in two sets; one comprising all assembled sequences and one without putatively chimeric sequences (i.e. sequences for which it cannot be ruled out that they are derived from contigs of different paralogs). When using HybPiper to retrieve only single copies of the targeted genes, by default the copy with a coverage depth much deeper (10x) than the others is chosen. If there is no such difference in coverage the copy with the greatest similarity to the reference sequence is chosen (HybPiper wiki documentation: https://github.com/mossmatters/HybPiper/wiki/Paralogs). Three samples with extremely poor sequence recovery (sequences recovered for fewer genes than 25 % of the average number across all samples, with sequences at 50 % of the target length recovered for fewer than 4 % of the average number across all samples) were excluded from further downstream analyses.

2.7. Data set generation

Three main sets of alignments were created for each target locus: exons, supercontigs, and exons including all potentially paralogous copies without putatively chimeric sequences (we will refer to this as the 'paralog' set). All sequences were aligned with mafft v7.520 (Katoh and Standley, 2013) (–thread -1 –localpair –maxiterate 1000). These sets of alignments were treated in two different ways: 1) the resulting alignments were used in downstream analyses without any modifications (except removal of fragmentary sequences, see below), we refer to these as 'untrimmed', and 2) the alignments were trimmed, using the program trimAl v1.4.1 (Capella-Gutiérrez et al., 2009) with the gap threshold 0.1, resulting in columns with less than 10 % non-gap characters being removed. We refer to these alignments as 'trimmed'. From the resulting six sets of alignments we removed all sequences consisting of less than 25 % non-gap characters as it has been shown that fragmentary data can negatively impact gene tree accuracy and hence species tree accuracy (Sayyari et al., 2017). Alignment statistics were summarised with AMAS v1.0 (Borowiec, 2016).

2.8. Tree inference

For each of the six data sets a gene tree was inferred for every single-locus alignment using IQ-TREE v2.2.2.7 (Minh et al., 2020), for each alignment using the substitution model selected by ModelFinder (Kalyaanamoorthy et al., 2017) (default behaviour in IQ-TREE 2). Branch support was assessed with 1,000 ultrafast bootstrap (UFBoot) pseudoreplicates (Minh et al., 2013; Hoang et al., 2018). Gene tree inference was parallelised using GNU Parallel (Tange, 2021). A species tree was estimated from each set of gene trees (the maximum likelihood trees; see Mirarab et al. (2016)) under the multispecies coalescent (MSC) model (Pamilo and Nei, 1988; Mirarab et al., 2021), using the quartet-based summary method ASTRAL (Mirarab et al., 2014). We used ASTRAL-III v5.7.8 (Zhang et al., 2018) for the exon and supercontig data sets and ASTRAL-Pro 2 v1.15.1.3 (Zhang et al., 2020; Zhang and Mirarab, 2022) for the paralog data sets. Prior to running ASTRAL very low support branches in the gene trees were contracted using Newick Utilities (Junier and Zdobnov, 2010) as it has been shown that this can improve species tree accuracy (Zhang et al., 2018). We contracted branches at or less than 20 % (UFboot) support. Trees inferred with ASTRAL-III were annotated using the -t 2 option and trees inferred with ASTRAL-Pro 2 were annotated using the -u 2 option. This produces extensive branch annotations, including quartet support (percentage of quartets in the gene trees that support a branch) for all three possible topologies and support values in the form of local posterior probabilities (LPP; Sayyari and Mirarab, 2016). The normalised quartet score (NQS; part of the output from ASTRAL-III) of the species tree was used to evaluate gene tree discordance (ASTRAL-III analyses only). This score represents the proportion of quartets in the gene trees that are found in the species tree (Mirarab, 2023). In addition to inferring species trees for the full sets of exon and supercontig trees, for each of these data sets we also inferred a species tree using only trees built from loci for which HybPiper did not give any paralog warnings. For the final eight supercontig and exon data sets (untrimmed and trimmed, and for both of these including and excluding potentially paralogous loci) we performed a polytomy test in ASTRAL-III (-t 10 option), which tests if the null hypothesis that a branch is of length zero can be rejected (Sayyari and Mirarab, 2018).

To complement the species tree estimation using summary methods we used SVDquartets (Chifman and Kubatko, 2014), a site-based method, to infer a species tree under the MSC model. SVDquartets is quartet-based like ASTRAL, but estimates a species tree directly from site pattern frequencies without first estimating gene trees. In simulations, SVDquartets has been shown to perform better than summary methods under certain conditions involving gene flow (Long and Kubatko, 2018). Two data sets were analysed using SVDquartets as implemented in PAUP* v4.0a168 (Swofford, 2003): the untrimmed exon and supercontig alignments excluding loci with paralog warnings. These data sets did not have fragmentary sequences removed as this practice aims to decrease gene tree error, which is not an issue with site-based methods. The developers caution against filtering the input (Swofford and Kubatko, 2023), and as recommended by them, we used the entire alignments as input rather than extracting SNPs. All possible quartets were evaluated and branch support was assessed with 1,000 multilocus bootstrap (Seo, 2008) replicates.

Prior to topology comparisons and visualisation of the inferred trees, trees were rooted using Phyx v1.1 (Brown et al., 2017), with the function pxrr. Trees were visualised in R (R Core Team, 2023) and compared for topological incongruences, including calculation of Robinson-Foulds distances (Robinson and Foulds, 1981), using the packages ape v5.7.1 (Paradis and Schliep, 2019), phytools v1.5.1 (Revell, 2012), treeio v1.23.1 (Wang et al., 2020), distory v1.4.4 (Chakerian and Holmes, 2020), and TreeDist v2.7.0 (Smith, 2020).

3. Results

3.1. Sequencing, assembly, and data set characteristics

Across the 44 samples that were included in the phylogenetic analyses in this study, the number of reads per sample ranged from 5.9 to 52.2 million, with an average of 27 million reads (28.4 million for the 39 newly sequenced samples, range 10–52.2 million). The average duplication rate across these 44 samples was 22 % (same for newly sequenced samples), and for the average sample 19.5 million reads remained after deduplication and trimming (20.5 million for newly sequenced). Sequencing and read trimming results for all samples can be found in Table S2.

Gene recovery was quite uniform and very high across all 44 samples; for the average sample sequences were assembled for 98 % (345 out of 353) of the targeted genes, totalling on average 255 kbp. For the average sample 810 kbp of supercontig sequences (i.e. protein-coding sequences plus flanking non-coding sequences) were assembled. On average samples had sequences at least 25, 50, and 75 % of the target length for 341 (97 %), 323 (92 %), and 274 (78 %) of the targeted loci. Out of the targeted 353 loci 351 had sequences for at least 4 samples and thus could be used for quartet-based phylogeny reconstruction. For 99 loci HybPiper gave a paralog warning for at least one sample. Assembly statistics for all samples can be found in Table S3.

Characteristics of the ten main data sets are summarised in Table 2. Across all the prepared data sets, the total number of aligned sites

Table 2
Selected statistics characterising the different data sets analysed in this study.

Data set		No. loci	Average alignment length	Average no. PIS per locus	Average % missing data	Total no. sites	Average taxon occupancy
supercontigs all	untrimmed	351	4567	793	46	1,603,364	41/44
	trimmed	351	3133	804	27	1,099747	42/44
supercontigs paralogs excl.	untrimmed	252	5026	853	47	1,266,531	40/44
	trimmed	252	3452	864	27	869,990	42/44
exons all	untrimmed	351	952	118	21	334,186	43/44
	trimmed	351	845	118	13	296,514	43/44
exons paralogs excl.	untrimmed	252	1030	122	23	259,586	42/44
	trimmed	252	916	122	15	230,885	42/44
paralogs	untrimmed	351	950	123	21	333,545	43/44
	trimmed	351	845	123	13	296,585	43/44

PIS: parsimony informative sites.

ranged from 1,603,364 in the data set comprising the untrimmed supercontig sequences for all 351 loci, to 230,885 in the trimmed data set comprising exon sequences for loci without paralog warnings (252 loci). The average alignment length in data sets without potential paralogs was consistently longer than in the corresponding full data sets, and the average number of parsimony informative sites per alignment was also consistently bigger in these data sets. In the data sets comprising supercontigs the average alignment length was about five (untrimmed) to four (trimmed) times longer than in the corresponding exon data sets, and the supercontig alignments had on average about seven times more parsimony informative sites.

The data sets that we used for the SVDquartets analyses (untrimmed supercontigs and exons excluding potential paralogs) did not have fragmentary sequences removed and differed slightly from the corresponding data sets used in the ASTRAL-III analyses: the exon data set had on average 123 parsimony informative sites per locus, 24 % missing data, and a taxon occupancy of 43/44. The supercontig data set had on average 874 parsimony informative sites per locus, 47 % missing data, and a taxon occupancy of 43/44.

3.2. The effect of data type and filtering on species tree inference with ASTRAL

Some statistics of the species trees inferred with ASTRAL and the gene trees they were inferred from are summarised in Table 3. The inclusion of flanking non-coding sequences (i.e. using supercontigs rather than exon sequences only) resulted in species trees with higher normalised quartet scores (Table 3). Supercontig trees also obtained a higher average support than corresponding exon trees. The average UFBoot support in the sets of gene trees used as input to ASTRAL was also higher for supercontig trees than for exon trees and the input supercontig trees had fewer branches contracted due to very low support. In addition, the topologies of the inferred species trees were more stable across the different supercontig data sets than across the exon data sets; the mean pairwise Robinson-Foulds (RF) distance between all supercontig trees was nine, and 94 between all exon trees. The average RF distance between corresponding supercontig and exon trees was 65.

Trimming did not noticeably affect the NQS values for exon data sets but for supercontigs, higher NQS values were obtained when using untrimmed data. The average RF distance between trees inferred from corresponding untrimmed and trimmed data sets was 33.

Species trees inferred from data including potential paralogs had higher average support and lower NQS than trees inferred from corresponding data sets without potential paralogs, although the effect was less pronounced compared to the inclusion of flanking regions (i.e. supercontigs versus exons). When excluding potential paralogs, the input gene trees had fewer branches collapsed. The topologies of the supercontig trees did not change with the exclusion of potential paralogs (i.e. the topologies inferred from untrimmed and trimmed data, respectively); for exons the RF distance between trees was 109 and 151 for untrimmed and trimmed data, respectively.

The multi-copy data sets that include all potential paralogs (referred to here as 'paralog data') performed similarly when analysed with ASTRAL-Pro as did the full set of exons analysed with ASTRAL-III. The topology inferred from the untrimmed paralog data was identical to those inferred with ASTRAL-III from the full set of exons (untrimmed and trimmed). The RF distance between these exon trees and the one inferred from trimmed paralog data was 95. Compared to exons without paralog warnings analysed with ASTRAL-III the paralog data sets performed better in terms of average the support, the average LPP was 0.03 higher in the ASTRAL-Pro analyses for both untrimmed and trimmed data (Table 3). The RF distance between topologies inferred with ASTRAL-Pro and those inferred with ASTRAL-III from the paralog-filtered set of exons was 109 and 56, for untrimmed and trimmed data respectively.

The power to reject polytomies was highest in the supercontig data sets. In the trees built using these data sets, a polytomy could be rejected for all but three–four branches ($\alpha = 0.05$; Table 3). The null hypothesis of a polytomy could not be rejected for the branch that places *H. libericus* as sister to all other *Hymenocoleus* species (Fig. 3) in any of the supercontig trees. The other branches in the supercontig trees that could not be rejected as polytomies are intraspecific, i.e. within *H. neurodictyon*,

Table 3
Selected statistics for species trees inferred with ASTRAL in this study.

Data set		NQS	Average LPP	Polytomy not rejected (no. branches)	No. branches < 0.95 LPP	Average UFBoot in input trees	Average no. contracted branches in input trees
supercontigs all	untrimmed	0.850	0.98	3	5	87	0.26
	trimmed	0.847	0.98	4	4	87	0.34
supercontigs paralogs excl.	untrimmed	0.864	0.97	3	5	87	0.2
	trimmed	0.860	0.97	3	5	87	0.25
exons all	untrimmed	0.783	0.95	8	8	75	3.2
	trimmed	0.783	0.95	8	7	75	3.6
exons paralogs excl.	untrimmed	0.790	0.92	9	9	75	3.0
	trimmed	0.791	0.92	10	9	75	3.3
paralogs	untrimmed	NA	0.95	NA	7	75	3.4
	trimmed	NA	0.95	NA	8	75	3.8

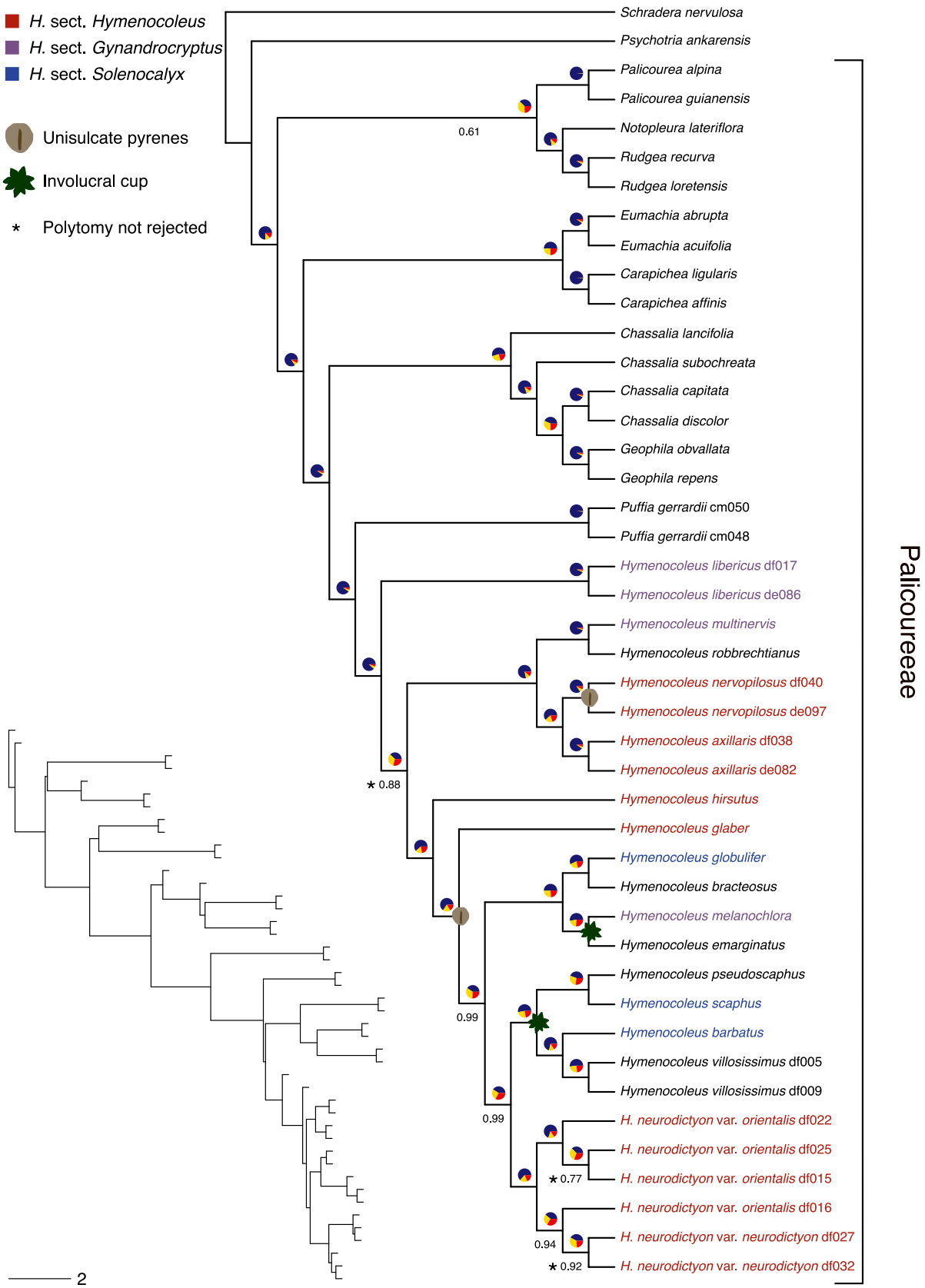


Fig. 3. Species tree inferred with ASTRAL-III from the data set comprising all untrimmed supercontig sequences. Branch support values, local posterior probabilities (LPP), are shown below branches if smaller than 1. Branches without numbers are supported by an LPP of 1. Pie charts show quartet support for the three possible quartet topologies (blue – main topology, yellow – first alternative, red – second alternative). The sections indicated on the phylogeny are according to Robbrecht (1977). Inset shows internal branch lengths in coalescent units (CU). Terminal branch lengths are not calculated by ASTRAL; terminal branch lengths of 0.2 CU have been added for ease of interpretation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

and different branches are affected across the trees. It is also within this species that most of the (overall very few) topological incongruences between the ten species trees inferred with ASTRAL are found. Across all ASTRAL trees, a total of five different topologies were observed (Fig. 3, Figs. S1–S9) and there were no supported (LPP $\geq$ 0.95) conflicts.

3.3. Species tree inference with SVDquartets

The topology of the bootstrap consensus tree from the SVDquartets analysis of supercontig data (Fig. S10) largely agrees with those inferred using ASTRAL; only infraspecific relationships (within *H. neurodictyon*) differ from ASTRAL supercontig trees. The mean pairwise RF distance between the SVDquartets supercontig tree and all the ASTRAL supercontig trees was 49, the mean pairwise distance between this tree and all ASTRAL trees was 77. The average bootstrap support percentage (BS) for all clades in the supercontig consensus tree was 93. The bootstrap consensus tree from the SVDquartets analysis of exon data (Fig. S11) presented the most divergent topology across all analyses in this study: the mean pairwise RF distance to all other trees was 281. Most of these deviating results were unsupported (BS < 70; see Hillis and Bull, 1993), but a small number were supported (Fig. S11). In view of the consistent topology resulting from all other analyses, we will not further consider the results from the SVDquartets analysis of exon data.

3.4. Phylogenetic relationships

Across all analyses the inferred species tree topologies were highly congruent and, especially in analyses of supercontig data, highly supported. In all analyses *Puffia* and *Hymenocoleus* were resolved as sisters with strong support. Across all trees there were only two incongruences concerning relationships between taxa at the level of species or above, and neither was supported. One of these incongruences occurred within *Chassalia* s.l. (including *Geophila*) sensu Razafimandimbison et al. (2014), and concerned the position of *C. subochreata*. In trees inferred from exons excluding potential paralogs this species was placed as sister to *C. discolor* and *C. capitata* (Figs. S6–S7), whereas in all other trees it was found as sister to these species and *Geophila*. The other concerned the deepest split within *Hymenocoleus*, where either *H. libericus* or a clade comprising *H. multinervis*, *H. robbrechtianus*, *H. nervopilosus*, and *H. axillaris* was sister to the remaining species. The former relationship was observed in all supercontig trees (ASTRAL and SVDquartets), in the exon trees with potential paralogs excluded, and in the results of the ASTRAL-Pro analysis of trimmed data (Fig. 3, Figs. S1–S3, S6–S7, S9–S10), and the latter in all remaining trees (Figs. S4–S5, S8). A polytomy could not be rejected for the deepest branch, regardless of divergence order, in any of the trees (polytomy test not implemented in ASTRAL-Pro or SVDquartets), and support was always low (between 0.88 and 0.38 LPP; BS 46) for this branch. Apart from this, all interspecific relationships were identically resolved across all trees, and supported in all supercontig trees (both ASTRAL and SVDquartets). All species, for which we included multiple samples, were supported as monophyletic in all analyses. Relationships among the varieties of *H. neurodictyon* were not robustly resolved, but *H. neurodictyon* var. *orientalis* was not monophyletic.

4. Discussion

We present here the first phylogeny of *Hymenocoleus* using representation from more than two species. Our results show that the genus is monophyletic and includes several distinct subgroups, many of which are supported by morphology (as discussed in 4.2 below) (Fig. 3). The phylogeny is inferred from several hundred target-captured nuclear loci analysed using coalescent-based methods of species tree inference. In addition, we report some noteworthy results on relationships in the remaining Palicoureeae; for example, the present study is the first to confidently demonstrate the sister relationship between *Puffia* and *Hymenocoleus*. The results are consistent; there is topological agreement between the results of nearly all the analyses performed in our study and results were mostly resolved with support (LPP $\geq$ 0.95), especially when using all available data. The only interspecific relationship within *Hymenocoleus* that could not be confidently inferred was that of the deepest split. Not even when using the overall highly informative supercontig data sets could a polytomy be rejected for the deepest branch within the genus. The tree with the highest support for the deepest split in *Hymenocoleus* (that from the analysis of all untrimmed supercontigs; Fig. 3) resolves *H. libericus* as sister to the remaining genus but with low statistical support (LPP of 0.88). Otherwise, interspecific relationships in *Hymenocoleus* could be confidently resolved, with relationships above the infraspecific level identically inferred in all trees built from supercontig sequences and from exon sequences (albeit not always supported in analyses of only exon sequences).

The use of a data set of hundreds of putatively independently evolving genes from the nuclear genome offers an important advantage over phylogenetic studies that rely solely on a handful of genes (e.g. Ruane et al., 2015) – especially if these are plastid genes, as the chloroplast genome is typically considered to functionally represent a single gene (e.g. Gitzendanner et al., 2018; Doyle, 2022). For phylogenomic nuclear data the use of coalescent-based species tree estimation is more appropriate than concatenation methods (e.g. Jiang et al., 2020), and based on the data set of low-copy nuclear genes used here, we are able to make a species tree estimate that is less likely to be flawed by processes such as incomplete lineage sorting (ILS; retention of ancestral gene polymorphism through speciation events). It is further interesting to note that the angiosperm-wide bait set we used (Johnson et al., 2019) produced data useful also at the generic and specific levels in Palicoureeae. Clearly, however, coalescent-based species tree estimation based on independently evolving low-copy nuclear genes is not free of challenges. Gene flow (Long and Kubatko, 2018), poorly resolved gene trees (Zhang et al., 2018), inclusion/exclusion of paralogous genes (Yang and Smith, 2014), inclusion/exclusion of non-coding data, and different data filtering strategies (Molloy and Warnow, 2018) are but a few examples that may affect the accuracy of species tree estimation.

4.1. Methodological considerations

4.1.1. The effect of inclusion of putative paralogs on phylogenetic inference with ASTRAL

A recent study by Yan et al. (2022) showed using both simulation and empirical analyses that several methods of species tree inference designed to handle ILS, among them ASTRAL-III (Zhang et al., 2018), are robust to the presence of paralogs in the analysed data sets. Common practice in phylogeny reconstruction has been to filter out paralogous loci, potentially leading to loss of vast amounts of data in large phylogenomic studies (Smith and Hahn, 2021), but given that data sets with paralogous sequences yield highly accurate species tree estimates this could be unnecessary (Yan et al., 2022). In the present study we compared both approaches, excluding loci with paralog warnings from HybPiper or including one copy per loci (the one 'selected' by HybPiper in the case of multiple copies being assembled, see Materials and methods for details). It is important to note, however, that not all paralog warnings may be due to true paralogy. Multiple sequences could be recovered also due to e.g. allelic variation and assembly of reads with low similarity to the target (Johnson et al., 2016).

We also used ASTRAL-Pro (Zhang et al., 2020), which is designed to handle paralogous multicopy data and allowed us to make use of all assembled copies of all loci, thus increasing the amount of data. The results of all analyses were in striking agreement, indicating that the inclusion of putatively paralogous loci did not have a big impact on species tree inference. Inclusion of putative paralogs resulted, for both exon and supercontig data, in overall slightly increased discordance and increased support (Table 3). The same pattern was observed also by Thureborn et al. (2022) and is not entirely surprising; paralogy

introduces gene tree heterogeneity (Smith and Hahn, 2021) and a larger amount of genes can give strong support even to branches with high levels of discordance (Sayyari and Mirarab, 2016).

We note one branch, however, where inclusion of putative paralogs had a big impact on support values in the analyses of supercontig data, leading to significantly lower support. This concerns the branch placing *Palicourea* as sister to *Notopleura* + *Rudgea*. In the analyses excluding putative paralogs, the LPP was 0.90 and 0.91 using untrimmed and trimmed data, respectively (Figs. S2 and S3, respectively). In the analyses including putative paralogs, the LPP for this branch was 0.61 and 0.75 using untrimmed and trimmed data, respectively (Fig. 3 and Fig. S1, respectively). We also note with interest that in the ASTRAL-Pro analyses of untrimmed and trimmed data, the LPP of this branch was 0.92 and 0.95, respectively (Figs. S8 and S9, respectively). This indicates that in this particular case, paralogs should preferably be handled in some way, either by removing them or using an inference method that actively accounts for them. It is also worth to note, however, that the inclusion of potentially paralogous loci did not impact the topology, even in this case.

4.1.2. The effect of data type and filtering on phylogenetic results

Including flanking regions (i.e. using supercontigs rather than exon sequences only) is important for phylogeny reconstruction at this shallow level, but also at deeper levels (e.g. Thureborn et al., 2022). As reported previously (e.g. Gardner et al., 2021; Thomas et al., 2021), it clearly leads to less gene tree error and/or more informative gene trees, true also in the present study where the inclusion of flanking sequences resulted in species trees with higher normalised quartet scores and trees with a higher average support than corresponding exon trees, and gene trees with fewer branches contracted due to low support (Table 3).

When it comes to the effect of trimming, it did not noticeably influence the normalised quartet scores for the exon data sets. However, for the supercontig data sets, higher NQS was obtained when using untrimmed data. The higher NQS values are likely the result of reduced gene tree error, indicating that the removed data contain valuable information and a consistent phylogenetic signal rather than introducing noise. This conclusion is further supported by the fact that input gene trees built from untrimmed alignments had fewer branches contracted due to low support. Portik and Wiens (2021) extensively tested the effect of trimming on phylogenomic inference using ultraconserved elements and found negligible difference on topology and support values between untrimmed and trimmed data. However, both no trimming and light trimming performed considerably better than aggressive trimming (Portik and Wiens, 2021).

4.2. Phylogenetic relationships

4.2.1. Palicoureeae

Our phylogenetic results are mostly in agreement with previous molecular phylogenetic studies of the tribe (Barrabé et al., 2012; Barrabé et al., 2014; Razafimandimbison et al., 2014; Razafimandimbison et al., 2017). However, our analyses indicated substantial gene tree heterogeneity regarding the relationship of *Palicourea*, the clade comprising *Notopleura* and *Rudgea*, and the remaining genera of Palicoureeae. In most of our analyses the relationship among these groups could not be confidently resolved. Interestingly though, when analysing the trimmed multi-copy data set including all potential paralogs with a method designed to handle paralogs (ASTRAL-Pro), their interrelationship is resolved; *Palicourea* and *Notopleura* + *Rudgea* were supported as sisters (0.95 LPP; Fig. S9), this clade being sister to the remaining tribe. This result is in conflict with the topology inferred by Razafimandimbison et al. (2014), where *Palicourea* was supported (0.98 Bayesian posterior probability (BPP)) as sister to the remaining tribe, a result also found but without support in several other studies as well (Barrabé et al., 2012; Barrabé et al., 2014; Razafimandimbison et al., 2017). While it is worth noticing that including (potential) paralog information and analysing it with an appropriate method here seems to dissolve phylogenetic uncertainty, the gene tree heterogeneity around this branch indicates that the question should be analysed further in the future, using genomic data from an extended sample of species and investigating the presence of true paralogy.

Apart from this deepest split in the tribe all intergeneric relationships within Palicoureeae could be resolved with an LPP of 1 in all our ASTRAL trees based on supercontig data (Fig. 3, Figs. S1–S3). We find *Eumachia* and *Carapichea* robustly confirmed as sisters, a relationship also observed by Barrabé et al. (2012; 2014) but unsupported. This clade is sister to a clade comprising *Geophila*, *Chassalia*, *Puffia*, and *Hymenocoleus*. Razafimandimbison et al. (2014; 2017) instead found *Carapichea* and *Eumachia* as successive sisters to the latter clade, but with low support (0.94 BPP). The analyses performed by Razafimandimbison et al. (2014; 2017) were primarily based on plastid data, and the results of their analyses and ours thus indicate cytonuclear discordance regarding the relationship of *Carapichea* and *Eumachia*. Furthermore, from the quartet frequencies we can clearly see that the branch placing *Carapichea* and *Eumachia* as sisters is heavily affected by ILS (Fig. 3). A pattern where the two alternative topologies occur at similar frequencies is consistent with ILS causing the observed gene tree discordance (e.g. Degnan and Rosenberg, 2009). It has previously been observed that cytonuclear conflicts can sometimes be explained by high levels of ILS (e.g. García et al., 2017; Stull et al., 2020; Thureborn et al., 2024), and the high levels of ILS around the branch we detect here could thus potentially explain the discordant results so far obtained using plastid data (e.g. Razafimandimbison et al., 2014) and low-copy nuclear data (the present study).

4.2.2. Puffia is the sister of Hymenocoleus

Puffia was robustly supported as sister to *Hymenocoleus* in all our analyses. One morphological character uniting *Puffia gerrardii* and *Hymenocoleus* is their inflated, watery fruits of various shades of orange. This combination of fruit characters is not seen in any other Palicoureeae (Robbrecht, 1975). Consequently, one taxonomic option could be to include the monospecific genus *Puffia* in *Hymenocoleus*. However, *Puffia gerrardii* differs from *Hymenocoleus* in having pyrenes dorsally ridged and slightly verrucose (smooth in *Hymenocoleus*), and the sheaths present inside the stipules of *Hymenocoleus* are absent in *Puffia* (Table 1). There is also a geographic distinction between the two: while *Hymenocoleus* is more broadly distributed in tropical continental Africa and the Gulf of Guinea islands, *Puffia* is restricted to eastern Madagascar. Based on this, and their phylogenetic results in which *Geophila gerrardii* was sister to the two sampled species of *Hymenocoleus*, Razafimandimbison et al. (2014) chose to describe the new genus *Puffia* to accommodate *Geophila gerrardii*, which was clearly shown to be distantly related to the remaining genus *Geophila* in that same study. We agree with these conclusions and have chosen to endorse the taxonomic choice made by Razafimandimbison et al. (2014).

However, while the sister relationship between *Puffia* and *Hymenocoleus* was found with strong statistical support also in a subsequent study (Razafimandimbison et al., 2017), the taxon sampling of *Hymenocoleus* in these studies was restricted to only two samples representing two species, meaning that the possibility of *Puffia* being nested in *Hymenocoleus* could not be dismissed. Here, with a much more extensive taxon sample comprising 16 out of the 19 recognised species of *Hymenocoleus* (Lachenaud, in press), we confirm that *Hymenocoleus* is monophyletic and sister to *Puffia*. Of the three missing species (*H. brevicalyx*, *H. campanulatus*, and *H. subipecacuanha*), at least two show clear morphological affinities to taxa well nested within the genus; *H. campanulatus* and *H. brevicalyx* are very similar to *H. scaphus* and allied species (i.e. *H. barbatus*, *H. pseudoscaphus*, and *H. villosissimus*) (Lachenaud, in press).

4.2.3. Incomplete lineage sorting affects the deepest branch in *Hymenocoleus*

Incomplete lineage sorting is clearly indicated to affect several branches in our phylogeny, and one of the most striking examples is the deepest divergence in *Hymenocoleus*, i.e. the divergence order of the two clades that are successive sisters to the remaining species of the genus. A polytomy could not be rejected for this branch in any of our analyses. While most loci contributed information about this branch (the effective number of genes was 345 as reported by ASTRAL in the analysis of all untrimmed supercontigs), it is clear that gene tree heterogeneity is strong with the three possible topologies appearing at nearly identical fractions (Fig. 3). This pattern of similar quartet frequencies indicates high levels of ILS, since the frequencies of the two alternative topologies are expected to be equal under the multispecies coalescent model (e.g. Degnan and Rosenberg, 2009). ILS is likely to cause discordance among gene trees when the time between speciation events is short (e.g. Degnan and Rosenberg, 2009), and the high levels of ILS observed here could possibly be attributed to rapid speciation during the earliest evolution of the genus. Future studies based on organellar genomic data will show if there is cytonuclear discordance regarding the deepest splits in *Hymenocoleus*, which could indicate alternative explanations for the discordance such as ancient introgression, but based on current information ILS is the most probable explanation for the observed gene tree heterogeneity.

4.2.4. The infrageneric classification of Hymenocoleus is untenable

Robbrecht (1977) focussed primarily on the floral organisation and pollen apertures and ultrastructure when he divided *Hymenocoleus* into three sections. Species of *H.* sect. *Hymenocoleus* (Fig. 2A, C) have heterodistylous flowers with either anthers or style exserted (Robbrecht, 1977) and a corolla rather thin in texture with a funnel-shaped tube (OL pers. obs.). The flowers of *H.* sect. *Gynandrocryptus* (Fig. 2D, E) and *H.* sect. *Solenocalyx* (Fig. 2F) are similar in having both anthers and style included (Robbrecht, 1977) and a corolla thick and waxy with a (sub) cylindrical tube (OL pers. obs.) but while flowers of *H.* sect *Gynandrocryptus* are heterodistylous, those of *H.* sect. *Solenocalyx* are homostylous. Here we do not find any consistent phylogenetic pattern in pollen features, and also not in floral characters, neither in shape nor organisation. In our analyses none of Robbrecht's three sections was found to be monophyletic, rather, they were all highly paraphyletic (Fig. 3). Species of *H.* sect. *Gynandrocryptus* (*H. libericus*, *H. melanochlora, H. multinervis*) are found in three different clades, species of *H.* sect. *Hymenocoleus* (*H. axillaris*, *H. glaber*, *H. hirsutus*, *H. nervopilosus*, *H. neurodictyon*) are found in four different clades, and species of *H.* sect. *Solenocalyx* (*H. barbatus*, *H. globulifer*, *H. scaphus*) are found in three different clades. Flowers in *Hymenocoleus* are small, white (very rarely yellow), and rather inconspicuous (Fig. 2), which is rather common within Psychotrieae and Palicoureeae, where few genera can be distinguished by floral characters (Robbrecht, 1989). On the other hand, Robbrecht (1989) pointed out that there is a significantly higher diversity of fruit characters within these tribes, which is of potential taxonomic value, something we exemplify here for *Hymenocoleus*.

Relationships within *H. neurodictyon* were considered uncertain and problematic already when the two varieties of the species were first described in 1975, with variable morphological features that sometimes appear to overlap between the varieties (Robbrecht, 1975). Also Lachenaud (in press) considers this species highly variable, which prompted him to recognise a third variety intermediate between the other two in many characters, and he notes that there is substantial variation within *H. neurodictyon* var. *orientalis*. Our results, in which *H. neurodictyon* var. *orientalis* is polyphyletic (Fig. 3), support these conclusions and we agree with both Robbrecht (1975) and Lachenaud (in press) who stress the need of further studies of the morphologically complex species *H. neurodictyon*.

4.2.5. More details on phylogenetic relationships and morphology within *Hymenocoleus*

The taxonomic value of pyrene characters within African genera now classified in Palicoureeae was already noted by Petit (1964) and has since been confirmed and used in several taxonomic studies on Palicoureeae genera (Robbrecht, 1975, 1989; Piesschaert, 2001; Piesschaert et al., 2001). Robbrecht (1975) mentioned pyrene characters as diagnostic when he described *Hymenocoleus*; members of the genus have pyrenes with a smooth dorsal side and one or two lateral grooves on the ventral side. The characters of the ventral surface have been considered most useful in taxonomic work (Piesschaert, 2001). For example, Piesschaert et al. (2001) could distinguish four types of pyrenes within *Chassalia* based on the structure of the ventral surface, and they transferred a couple of species to the genus based on them possessing one of these pyrene types.

Also in *Hymenocoleus* one of the most important morphological characters is the ventral surface of the pyrenes. As Robbrecht (1975; 1977) noted the pyrenes are either uni- or bisulcate, i.e. they have either one or two grooves on the ventral side, a feature that is specifically constant, also in the new species discovered since (Lachenaud, in press). Supported by phylogeny (Fig. 3) we can further confirm the taxonomic usefulness of these characters. It seems likely that the bisulcate condition is ancestral in the genus, since pyrenes of the sister genus *Puffia* have two (weakly marked) ventral grooves (Vanthournout, 2002; confirmed by OL, pers. obs.). The presence of only one ventral groove is mostly limited to a core clade of the genus, but has also evolved in parallel in *H. nervopilosus* (Fig. 3). Interestingly, the three closest relatives of *H. nervopilosus* (*H. axillaris*, *H. multinervis*, and *H. robbrechtianus*; Fig. 3) share the trait that the two ventral grooves are positioned very close to each other, and it is tempting to speculate that the single groove in *H. nervopilosus* is the result of two merged grooves.

Another striking morphological character in *Hymenocoleus* is the presence in some species of a campanulate involucral cup, a feature that is unique to *Hymenocoleus* within African Rubiaceae (Lachenaud, in press). We find such an involucral cup distributed in two clades (Fig. 3), and there are clear differences between its structure in these two clades. In *H. scaphus* and allied species (*H. barbatus, H. brevicalyx, H. campanulatus, H. pseudoscaphus, H. villosissimus*), the involucral cup is star-shaped and bicoloured (Fig. 2F), pale green on the outside and dark green on the inside, with patent lobes. In the *H. emarginatus-H. melanochlora* clade (Fig. 2D), the involucral cup closely surrounds the inflorescence (instead of being much broader, as in *H. scaphus* and allied species) and its lobes are unicoloured (pale green) and erect. One additional species not sampled here, *H. subipecacuanha*, has an involucral cup somewhat intermediate between the two types described: the lobes are patent, as in *H. scaphus* and allies, but the inside of the involucre is only slightly darker than the outside. Robbrecht (1977) describes more variation in the structure supporting the inflorescence, where some species (i.e. *H. globulifer, H. multinervis, H. nervopilosus*) have an involucre of free bracts, other species have bracts that may be connate at the base (e.g. *H. axillaris, H. glaber, H. libericus, H. neurodictyon*), whereas yet other species (e.g. *H. barbatus, H. rotundifolius* (accepted name *H. melanochlora*; Lachenaud, in press), *H. scaphus*) have fused bracts forming a cup-shaped structure. This variation can be seen also in newly described species, where e.g. *H. bracteosus* has an involucre of free bracts (Lachenaud, in press). This could potentially indicate an evolutionary trend in *Hymenocoleus*, towards or away from fused bracts forming an involucral cup. We can, however, not find any indications of such a trend in our phylogenetic results (Fig. 3), and the notable morphological differences between involucral cups in species of different clades suggest that the feature referred to as an involucral cup has evolved in parallel at least twice in *Hymenocoleus*.

There are a few additional interesting morphological characters within the genus that are consistent with phylogeny and seem to represent synapomorphies of smaller clades. *Hymenocoleus axillaris* (Fig. 2C) and *H. nervopilosus* (Fig. 2A), which grouped together (Fig. 3),

are the only two species with paired inflorescences in an axillary position (rarely terminal in the latter species). *Hymenocoleus multinervis* and *H. robbrechtianus*, which formed a clade that is sister to *H. axillaris* and *H. nervopilosus*, are united by having a caducous calyx and most notably by having the uppermost pair of stipules surrounding the inflorescence; their leaves are also very similar. *Hymenocoleus bracteosus* and *H. globulifer* (Fig. 2B) form a clade and show many similarities, e.g. the leaf shape and a long calyx tube. However, there are also clades for which we find no obvious morphological support, e.g. the clade comprising *H. multinervis*, *H. robbrechtianus*, *H. nervopilosus*, and *H. axillaris*, and our phylogeny can provide a framework for further studies of morphological evolution in *Hymenocoleus*.

CRediT authorship contribution statement

Lovisa Thilén: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Olivier Lachenaud:** Writing – review & editing, Validation, Investigation, Data curation. **Olle Thureborn:** Writing – review & editing, Validation, Methodology, Conceptualization. **Sylvain G. Razafimandimbison:** Writing – review & editing, Validation, Supervision, Investigation, Data curation, Conceptualization. **Catarina Rydin:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank the herbaria BR, BRLU, GB, MO, S, UPS, and WU for access to their collections. Many of the *Hymenocoleus* samples used in this paper derive from field work in Gabon by OL and colleagues, conducted under the Memorandum of Understanding between the Missouri Botanical Garden (MBG) and the Centre National de la Recherche Scientifique et Technologique (CENAREST) (permit AR0045/19/MESRSTT/CENAREST/CG/CST/CSAR). The director of IPHAMETRA (Institut de Pharmacopée et de Médecine Traditionelle), Nestor Engone Obiang, and the curator of the National Herbarium of Gabon, Archange Boupoya, are thanked for their collaboration. We thank Eric Akouangou, Ehoarn Bidault, Archange Boupoya, Steven Dessein, Yves Issembé, Steven Janssens, Jean de Dieu Kaparidi, Diosdado Nguema, Thomas Nzabi (†) and Tariq Stévart for their assistance in the field and/or in organising the expeditions. The DGF (Direction Générale des Forêts) in Madagascar is thanked for issuing collecting and exportation permits for SGR; the Missouri Botanical Garden, Madagascar Program for logistical support; and the Parc Botanique et Zoologique de Tsimbazaza and the Missouri Botanical Garden, Madagascar Program (F. Lantoarisoa) for arranging collecting and exportation permits for SGR. We also thank Niklas Wikström and Tanja Slotte (Stockholm University) for fruitful discussions and Bodil Cronholm (Swedish Museum of Natural History) and Anbar Khodabandeh (Royal Swedish Academy of Sciences) for technical assistance, and we acknowledge support from the National Genomics Infrastructure in Stockholm funded by Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council, and SNIC/NAISS/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. The project was funded by grants from the Royal Swedish Academy of Sciences and Stockholm University to CR, the Royal Swedish Academy of Sciences and Swedish Museum of Natural History to SGR, and from WWF-Gabon, the Agence Nationale des Parcs Nationaux (ANPN) and the National Geographic Society to OL.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2025.108338.

Data availability

The raw sequencing data generated for the present study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under project accession number PRJEB86169 (https://www.ebi.ac.uk/ena/browser/view/PRJEB86169). ENA run accession numbers for individual samples are available in Table S1.

References

Andersson, L., 2001. *Margaritopsis* (Rubiaceae, Psychotrieae) is a pantropical genus. Syst. & Geogr. Pl. 71, 73–85. https://doi.org/10.2307/3668754.
Andersson, L., 2002. Relationships and generic circumscription in the *Psychotria* complex (Rubiaceae, Psychotrieae). Syst. & Geogr. Pl. 72, 167–202.
Andersson, L., Rova, J.H.E., 1999. The *rps16* intron and the phylogeny of the Rubioideae (Rubiaceae). Pl. Syst. Evol. 214, 161–186. https://doi.org/10.1007/BF00985737.
Baker, W.J., Bailey, P., Barber, V., Barker, A., Bellot, S., Bishop, D., Botigué, L.R., Brewer, G., Carruthers, T., Clarkson, J.J., Cook, J., Cowan, R.S., Dodsworth, S., Epitawalage, N., Françoso, E., Gallego, B., Johnson, M.G., Kim, J.T., Leempoel, K., Maurin, O., McGinnie, C., Pokorny, L., Roy, S., Stone, M., Toledo, E., Wickett, N.J., Zuntini, A.R., Eiserhardt, W.L., Kersey, P.J., Leitch, I.J., Forest, F., 2022. A comprehensive phylogenomic platform for exploring the angiosperm tree of life. Syst. Biol. 71, 301–319. https://doi.org/10.1093/sysbio/syab035.
Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V. M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477. https://doi.org/10.1089/cmb.2012.0021.
Barrabé, L., Buerki, S., Mouly, A., Davis, A.P., Munzinger, J., Maggia, L., 2012. Delimitation of the genus *Margaritopsis* (Rubiaceae) in the Asian, Australasian and Pacific region, based on molecular phylogenetic inference and morphology. Taxon 61, 1251–1268. https://doi.org/10.1002/tax.616007.
Barrabé, L., Maggia, L., Pillon, Y., Rigault, F., Mouly, A., Davis, A.P., Buerki, S., 2014. New Caledonian lineages of *Psychotria* (Rubiaceae) reveal different evolutionary histories and the largest documented plant radiation for the archipelago. Molec. Phylogen. Evol. 71, 15–35. https://doi.org/10.1016/j.ympev.2013.10.020.
Borowiec, M.L., 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ 4, e1660. https://doi.org/10.7717/peerj.1660.
Bremer, B., Manen, J.F., 2000. Phylogeny and classification of the subfamily Rubioideae (Rubiaceae). Pl. Syst. Evol. 225, 43–72. https://doi.org/10.1007/BF00985458.
Brown, J.W., Walker, J.F., Smith, S.A., 2017. Phyx: phylogenetic tools for unix. Bioinformatics 33, 1886–1888. https://doi.org/10.1093/bioinformatics/btx063.
Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.
Chakerian, J., Holmes, S., 2020. distory: distance between phylogenetic histories (Version 1.4.4) [R package]. Comprehensive R Archive Network. https://doi.org/10.32614/CRAN.package.distory.
Chen, S., 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. iMeta 2, e107. https://doi.org/10.1002/imt2.107.
Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890. https://doi.org/10.1093/bioinformatics/bty560.
Chifman, J., Kubatko, L.S., 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics 30, 3317–3324. https://doi.org/10.1093/bioinformatics/btu530.
Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24, 332–340. https://doi.org/10.1016/j.tree.2009.01.009.
Doyle, J.J., 2022. Defining coalescent genes: theory meets practice in organelle phylogenomics. Syst. Biol. 71, 476–489. https://doi.org/10.1093/sysbio/syab053.
Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19, 11–15.
García, N., Folk, R.A., Meerow, A.W., Chamala, S., Gitzendanner, M.A., de Oliveira, R.S., Soltis, D.E., Soltis, P.S., 2017. Deep reticulation and incomplete lineage sorting obscure the diploid phylogeny of rain-lilies and allies (Amaryllidaceae tribe Hippeastreae). Molec. Phylogen. Evol. 111, 231–247. https://doi.org/10.1016/j.ympev.2017.04.003.
Gardner, E.M., Johnson, M.G., Pereira, J.T., Puad, A.S.A., Arifiani, D., Sahromi, Wickett, N.J., Zerega, N.J.C., 2021. Paralogs and off-target sequences improve phylogenetic resolution in a densely sampled study of the breadfruit genus (*Artocarpus*, Moraceae). Syst. Biol. 70, 558–575. https://doi.org/10.1093/sysbio/syaa073.

Gitzendanner, M.A., Soltis, P.S., Wong, G.K.-S., Ruhfel, B.R., Soltis, D.E., 2018. Plastid phylogenomic analysis of green plants: a billion years of evolutionary history. Amer. J. Bot. 105, 291–301. https://doi.org/10.1002/ajb2.1048.

Hillis, D.M., Bull, J.J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42, 182–192. https://doi.org/10.1093/sysbio/42.2.182.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35, 518–522. https://doi.org/10.1093/molbev/msx281.

Jiang, X., Edwards, S.V., Liu, L., 2020. The multispecies coalescent model outperforms concatenation across diverse phylogenomic data sets. Syst. Biol. 69, 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. Appl. Plant Sci. 4, 1600016. 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.H., 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. Syst. Biol. 68, 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, 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. Nat. Methods 14, 587–589. 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. Mol. Biol. Evol. 30, 772–780. https://doi.org/10.1093/molbev/mst010.

Lachenaud, O., in press. Les genres *Chassalia*, *Eumachia*, *Geophila* et *Hymenocoleus* (Rubiaceae) en Afrique tropicale. Opera Bot. Belg. 18.

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

Long, C., Kubatko, L., 2018. The effect of gene flow on coalescent-based species-tree inference. Syst. Biol. 67, 770–785. https://doi.org/10.1093/sysbio/syy020.

Meyer, M., Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. 2010, pdb. prot5448. https://doi.org/10.1101/pdb.prot5448.

Minh, B.Q., Nguyen, M.A.T., von Haeseler, A., 2013. Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. 30, 1188–1195. https://doi.org/10.1093/molbev/mst024.

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. Mol. Biol. Evol. 37, 1530–1534. https://doi.org/10.1093/molbev/msaa015.

Mirarab, S., 2023. Species tree estimation using ASTRAL: practical considerations. In: Kubatko, L.S., Knowles, L.L. (Eds.), Species Tree Inference: a Guide to Methods and Applications. Princeton University Press, Princeton, pp. 43–67.

Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., Warnow, T., 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30, i541–i548. https://doi.org/10.1093/bioinformatics/btu462.

Mirarab, S., Bayzid, M.S., Warnow, T., 2016. Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. Syst. Biol. 65, 366–380. https://doi.org/10.1093/sysbio/syu063.

Mirarab, S., Nakhleh, L., Warnow, T., 2021. Multispecies coalescent: theory and applications in phylogenetics. Annu. Rev. Ecol. Evol. Syst. 52, 247–268. https://doi.org/10.1146/annurev-ecolsys-012121-095340.

Molloy, E.K., Warnow, T., 2018. To include or not to include: the impact of gene filtering on species tree estimation methods. Syst. Biol. 67, 285–303. https://doi.org/10.1093/sysbio/syx077.

Nepokroeff, M., Bremer, B., Sytsma, K.J., 1999. Reorganization of the genus *Psychotria* and tribe Psychotrieae (Rubiaceae) inferred from ITS and *rbcL* sequence data. Syst. Bot. 24, 5–27. https://doi.org/10.2307/2419383.

Pamilo, P., Nei, M., 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5, 568–583. https://doi.org/10.1093/oxfordjournals.molbev.a040517.

Paradis, E., Schliep, K., 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35, 526–528. https://doi.org/10.1093/bioinformatics/bty633.

Petit, E. 1963. Bijdrage tot de kennis van het geslacht *Psychotria* L. (Rubiaceae) in Afrika. Doctoral Thesis, Katholieke Universiteit Leuven, Leuven.

Petit, E., 1964. Les espèces africaines du genre *Psychotria* L. (Rubiaceae): I. Bull. Jard. Bot. État Bruxelles 34, 1–160. https://doi.org/10.2307/3667373.

Piesschaert, F. 2001. Carpology and pollen morphology of the Psychotrieae (Rubiaceae-Rubioideae): towards a new tribal and generic delimitation. Doctoral Thesis, Katholieke Universiteit Leuven, Leuven.

Piesschaert, F., Robbrecht, E., Smets, E., 2001. Patterns in pyrenes: the systematic significance of pyrene morphology in *Chassalia* (Rubiaceae-Psychotrieae) and related genera. Flora 196, 121–131. https://doi.org/10.1016/S0367-2530(17)30027-0.

Portik, D.M., Wiens, J.J., 2021. Do alignment and trimming methods matter for phylogenomic (UCE) analyses? Syst. Biol. 70, 440–462. https://doi.org/10.1093/sysbio/syaa064.

POWO, 2024. Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. https://powo.science.kew.org/ (accessed 14 May 2024).

Razafimandimbison, S.G., Rydin, C., 2024a. Addendum and corrigendum to: Phylogeny and classification of the coffee family (Rubiaceae, Gentianales): overview and outlook [in Taxon 73(3): 673–717. 2024]. Taxon 73, 1556–1557. https://doi.org/10.1002/tax.13293.

Razafimandimbison, S.G., Rydin, C., 2024b. Phylogeny and classification of the coffee family (Rubiaceae, Gentianales): overview and outlook. Taxon 73, 673–717. https://doi.org/10.1002/tax.13167.

Razafimandimbison, S.G., Rydin, C., Bremer, B., 2008. Evolution and trends in the Psychotrieae alliance (Rubiaceae) – a rarely reported evolutionary change of many-seeded carpels from one-seeded carpels. Molec. Phylogen. Evol. 48, 207–223. https://doi.org/10.1016/j.ympev.2008.03.034.

Razafimandimbison, S.G., Taylor, C.M., Wikström, N., Pailler, T., Khodabandeh, A., Bremer, B., 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. Amer. J. Bot. 101, 1102–1126. https://doi.org/10.3732/ajb.1400076.

Razafimandimbison, S.G., Kainulainen, K., Wikström, N., Bremer, B., 2017. Historical biogeography and phylogeny of the pantropical Psychotrieae alliance (Rubiaceae), with particular emphasis on the Western Indian Ocean Region. Amer. J. Bot. 104, 1407–1423. https://doi.org/10.3732/ajb.1700116.

R Core Team, 2023. R: a language and environment for statistical computing (Version 4.3.0) [software]. R Foundation for Statistical Computing. https://www.R-project.org/.

Revell, L.J., 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3, 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x.

Robbrecht, E., 1975. *Hymenocoleus*, a new genus of Psychotrieae (Rubiaceae) from tropical Africa. Bull. Jard. Bot. Natl. Belg. 45, 273–300. https://doi.org/10.2307/3667482.

Robbrecht, E., 1977. The tropical African genus *Hymenocoleus* (Rubiaceae-Psychotrieae): additions. Bull. Jard. Bot. Natl. Belg. 47, 3–29. https://doi.org/10.2307/3667978.

Robbrecht, E., 1988. Tropical woody Rubiaceae. Opera Bot. Belg. 1, 1–271.

Robbrecht, E., 1989. A remarkable new *Chazaliella* (African Psychotrieae), exemplifying the taxonomic value of pyrene characters in the Rubiaceae. Bull. Mus. Natl. Hist. Nat., B, Adansonia 11, 341–349.

Robbrecht, E., Manen, J.-F., 2006. The major evolutionary lineages of the coffee family (Rubiaceae, angiosperms). Combined analysis (nDNA and cpDNA) to infer the position of *Coptosapelta* and *Luculia*, and supertree construction based on *rbcL*, *rps16*, *trnL-trnF* and *atpB-rbcL* data. A new classification in two subfamilies, Cinchonoideae and Rubioideae. Syst. & Geogr. Pl. 76, 85–146.

Robinson, D.F., Foulds, L.R., 1981. Comparison of phylogenetic trees. Math. Biosci. 53, 131–147. https://doi.org/10.1016/0025-5564(81)90043-2.

Ruane, S., Raxworthy, C.J., Lemmon, A.R., Lemmon, E.M., Burbrink, F.T., 2015. Comparing species tree estimation with large anchored phylogenomic and small Sanger-sequenced molecular datasets: an empirical study on Malagasy pseudoxyrhophiine snakes. BMC Evol. Biol. 15, 221. https://doi.org/10.1186/s12862-015-0503-1.

Rydin, C., Razafimandimbison, S.G., Khodabandeh, A., Bremer, B., 2009. Evolutionary relationships in the Spermacoceae alliance (Rubiaceae) using information from six molecular loci: insights into systematic affinities of *Neohymenopogon* and *Mouretia*. Taxon 58, 793–810. https://doi.org/10.1002/tax.583009.

Sayyari, E., Mirarab, S., 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Mol. Biol. Evol. 33, 1654–1668. https://doi.org/10.1093/molbev/msw079.

Sayyari, E., Mirarab, S., 2018. Testing for polytomies in phylogenetic species trees using quartet frequencies. Genes 9, 132. https://doi.org/10.3390/genes9030132.

Sayyari, E., Whitfield, J.B., Mirarab, S., 2017. Fragmentary gene sequences negatively impact gene tree and species tree reconstruction. Mol. Biol. Evol. 34, 3279–3291. https://doi.org/10.1093/molbev/msx261.

Seo, T.-K., 2008. Calculating bootstrap probabilities of phylogeny using multilocus sequence data. Mol. Biol. Evol. 25, 960–971. https://doi.org/10.1093/molbev/msn043.

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

Smith, M.R., 2020. TreeDist: distances between phylogenetic trees (Version 2.7.0) [R package]. Comprehensive R Archive Network. https://doi.org/10.32614/CRAN.package.TreeDist.

Smith, M.L., Hahn, M.W., 2021. New approaches for inferring phylogenies in the presence of paralogs. Trends Genet. 37, 174–187. https://doi.org/10.1016/j.tig.2020.08.012.

Stull, G.W., Soltis, P.S., Soltis, D.E., Gitzendanner, M.A., Smith, S.A., 2020. Nuclear phylogenomic analyses of asterids conflict with plastome trees and support novel relationships among major lineages. Amer. J. Bot. 107, 790–805. https://doi.org/10.1002/ajb2.1468.

Swofford, D.L., 2003. PAUP*. Phylogenetic analysis using parsimony (*and other methods) (Version 4.0a168) [software]. Sinauer Associates. https://paup.phylosolutions.com/.

Swofford, D.L., Kubatko, L.S., 2023. Species tree estimation using site pattern frequencies. In: Kubatko, L.S., Knowles, L.L. (Eds.), Species Tree Inference: a Guide to Methods and Applications. Princeton University Press, Princeton, pp. 68–88.

Tange, O., 2021. GNU Parallel 20210822 ('Kabul') (Version 1) [software]. Zenodo. https://doi.org/10.5281/zenodo.5233953.

Thomas, A.E., Igea, J., Meudt, H.M., Albach, D.C., Lee, W.G., Tanentzap, A.J., 2021. Using target sequence capture to improve the phylogenetic resolution of a rapid radiation in New Zealand *Veronica*. Amer. J. Bot. 108, 1289–1306. https://doi.org/10.1002/ajb2.1678.

Thureborn, O., Razafimandimbison, S.G., Wikström, N., Rydin, C., 2022. Target capture data resolve recalcitrant relationships in the coffee family (Rubioideae, Rubiaceae). Front. Pl. Sci. 13, 967456. https://doi.org/10.3389/fpls.2022.967456.

Thureborn, O., Wikström, N., Razafimandimbison, S.G., Rydin, C., 2024. Plastid phylogenomics and cytonuclear discordance in Rubioideae, Rubiaceae. PLOS ONE 19, e0302365. https://doi.org/10.1371/journal.pone.0302365.

Vanthournout, S. 2002. *Geophila*. Revisie van de paleotropische taxa. Licentiate Thesis, Katholieke Universiteit Leuven, Leuven.

Wang, L.-G., Lam, T.T.-Y., Xu, S., Dai, Z., Zhou, L., Feng, T., Guo, P., Dunn, C.W., Jones, B.R., Bradley, T., Zhu, H., Guan, Y., Jiang, Y., Yu, G., 2020. Treeio: an R package for phylogenetic tree input and output with richly annotated and associated data. Mol. Biol. Evol. 37, 599–603. https://doi.org/10.1093/molbev/msz240.

White, F., 1993. The AETFAT chorological classification of Africa: history, methods and applications. Bull. Jard. Bot. Natl. Belg. 62, 225–281. https://doi.org/10.2307/3668279.

Yan, Z., Smith, M.L., Du, P., Hahn, M.W., Nakhleh, L., 2022. Species tree inference methods intended to deal with incomplete lineage sorting are robust to the presence of paralogs. Syst. Biol. 71, 367–381. https://doi.org/10.1093/sysbio/syab056.

Yang, Y., Smith, S.A., 2014. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. Mol. Biol. Evol. 31, 3081–3092. https://doi.org/10.1093/molbev/msu245.

Zhang, C., Mirarab, S., 2022. ASTRAL-Pro 2: ultrafast species tree reconstruction from multi-copy gene family trees. Bioinformatics 38, 4949–4950. https://doi.org/10.1093/bioinformatics/btac620.

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

Zhang, C., Scornavacca, C., Molloy, E.K., Mirarab, S., 2020. ASTRAL-Pro: quartet-based species-tree inference despite paralogy. Mol. Biol. Evol. 37, 3292–3307. https://doi.org/10.1093/molbev/msaa139.