Molecular Phylogenetics and Evolution 175 (2022) 107581

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Out of chaos: Phylogenomics of Asian Sonerileae

Qiu-Jie Zhou [a,b], Jin-Hong Dai [a], Che-Wei Lin [c], Wei-Lun Ng [d], Truong Van Do [e,f], Jarearnsak Sae Wai [g], Fabián A. Michelangeli [h], Marcelo Reginato [i], Ren-Chao Zhou [a,*], Ying Liu [a,*]

[a] State Key Laboratory of Biocontrol and Guangdong Key Laboratory of Plant Resources, School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China
[b] Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden
[c] Herbarium of Taiwan Forestry Research Institute, Taipei 100, Taiwan
[d] China-ASEAN College of Marine Sciences, Xiamen University Malaysia, Sepang, Malaysia
[e] Vietnam National Museum of Nature, Vietnam Academy of Science and Technology, 18th Hoang Quoc Viet Road, Cau Giay, Hanoi, Viet Nam
[f] Graduate University of Science and Technology, Vietnam Academy of Science and Technology, 18th Hoang Quoc Viet Road, Cau Giay, Hanoi, Viet Nam
[g] Department of Biology, Faculty of Science, Prince of Songkla University, Hat Yai, Songkhla 90110, Thailand
[h] Institute of Systematic Botany, The New York Botanical Garden, Bronx, NY 10458–5126, USA
[i] Departamento de Botânica, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Campus do Vale Agronomia, RS, 91509900, Porto Alegre, Brazil

ARTICLE INFO

Keywords:
Sonerileae
Melastomataceae
Phylogeny
Discordance
Generic circumscription

ABSTRACT

Sonerileae is a diverse Melastomataceae lineage comprising ca. 1000 species in 44 genera, with >70% of genera and species distributed in Asia. Asian Sonerileae are taxonomically intractable with obscure generic circumscriptions. The backbone phylogeny of this group remains poorly resolved, possibly due to complexity caused by rapid species radiation in early and middle Miocene, which hampers further systematic study. Here, we used genome resequencing data to reconstruct the phylogeny of Asian Sonerileae. Three parallel datasets, viz. single-copy ortholog (SCO), genomic SNPs, and whole plastome, were assembled from genome resequencing data of 205 species for this purpose. Based on these genome-scale data, we provided the first well resolved phylogeny of Asian Sonerileae, with 34 major clades identified and 74% of the interclade relationships consistently resolved by both SCO and genomic data. Meanwhile, widespread phylogenetic discordance was detected among SCO gene trees as well as species trees reconstructed using different tree estimation methods (concatenation/site-based coalescent method/summary method) or different datasets (SCO/genomic/plastome). We explored sources of discordance using multiple approaches and found that the observed discordance in Asian Sonerileae was mainly caused by a combination of biased distribution of missing data, random noise from uninformative genes, incomplete lineage sorting, and hybridization/introgression. Exploration of these sources can enable us to generate hypotheses for future testing, which is the first step towards understanding the evolution of Asian Sonerileae. We also detected high levels of homoplasy for some characters traditionally used in taxonomy, which explains current chaotic generic delimitations. The backbone phylogeny of Asian Sonerileae revealed in this study offers a solid basis for future taxonomic revision at the generic level.

1. Introduction

Melastomataceae are among the most species-rich families of flowering plants, containing ca. 176 genera and ca. 5857 species distributed in subtropical and tropical regions worldwide (Renner, 1993; Clausing and Renner, 2001a; Goldenberg et al., 2015; Bacci et al., 2019; Michelangeli et al., 2020; Ulloa Ulloa et al., in press). In many wet tropical areas, Melastomataceae constitute a large portion of the local flora. Some genera are valued for their showy inflorescence (e.g., *Medinilla* Gaudich.), bright-colored flowers (e.g., *Tibouchina* Aubl.) and striking leaf coloration (e.g., *Sonerila* Roxb.). Despite their ecological importance and ornamental potential, Melastomataceae, especially the palaeotropical groups, remain largely understudied.

Sonerileae is a pantropical clade comprising species from the traditionally circumscribed tribes Sonerileae *s.s.*, Oxysporeae, and some genera recently excluded from Dissochaeteae and Bertolonieae

* Corresponding authors.
E-mail addresses: zhrench@mail.sysu.edu.cn (R.-C. Zhou), liuyng73@mail.sysu.edu.cn (Y. Liu).

https://doi.org/10.1016/j.ympev.2022.107581
Received 5 January 2022; Received in revised form 23 May 2022; Accepted 26 May 2022
Available online 8 July 2022

(Clausing and Renner, 2001a; Renner et al., 2001; Bacci et al., 2019; Zhou et al., 2019a, 2019b; Reginato et al., 2020; Baker et al., 2021; Kartonegoro et al., 2021; Maurin et al., 2021). It contains ca. 1080 species in 44 genera, two thirds of which occur in Southeast Asia (SEA) (Liu et al., in press). The Asian groups are taxonomically intractable. Different choice of diagnostic characters had led to controversial generic circumscriptions, and in turn, conflicting generic placement of the same species. For example, the long-standing dispute over the generic limit of *Phyllagathis* Blume (Diels, 1932; Li, 1944; Chen, 1984a; Hansen, 1992; Cellinese, 1997, 2002, 2003; Chen and Renner, 2007) has resulted in frequent transfer of species between *Phyllagathis* and various genera such as *Anerincleistus* Korth. (Hansen, 1982; Maxwell, 1982; Cellinese, 2002), *Cyphotheca* Diels (Hu, 1952), *Plagiopetalum* Rehder (Chen, 1984b), and *Scorpiothyrsus* H.L.Li (Li, 1944) etc. In addition, many new species of Sonerileae have been described from SEA in recent years (e.g., Lin et al., 2015; Mathew et al., 2016; Lin et al., 2017; Pham et al., 2017; Lin and Lee, 2018; Lin, 2019), and more discoveries are expected as poorly explored areas become better botanized. The problematic generic limits of the Asian groups would undoubtedly cause serious problems in generic placement of the new species and hinder further systematic and evolutionary studies.

Molecular phylogenetics has drastically improved our understanding of the phylogeny and evolution of Melastomataceae (e.g., Clausing, 1999; Clausing and Renner, 2001a; Renner et al., 2001; Michelangeli et al., 2011, 2013; Penneys and Judd, 2013; Veranso-Libalah et al., 2017; Bacci et al., 2019; Reginato et al., 2020; Kartonegoro et al., 2021; Maurin et al., 2021). Species of Asian Sonerileae had been included in a handful of phylogenetic studies, but limited sampling of genera or species in most analyses (e.g., Renner et al., 2001; Zeng et al., 2016; Zhou et al., 2018; Bacci et al., 2019; Maurin et al., 2021) prevented discussion of generic limits. Zhou et al. (2019a) first addressed this issue using sequence data of nrITS and chloroplast *trnV-trnM* from 96 species (18 genera) of Sonerileae. They subsequently explored the use of fully sequenced plastomes to gain better resolution, expanding taxon sampling to include 131 species (21 genera) (Zhou et al., 2019b). These analyses clearly demonstrated the chaotic generic circumscriptions and succeeded in identifying a dozen major lineages in Asian Sonerileae (Zhou et al., 2019a, 2019b). Although plastid genomic data provided better supported trees than nrITS, the backbone phylogeny of the tribe has remained poorly resolved in both studies. Divergence time estimation together with poorly supported short internodes along the backbone indicated putative rapid radiation of the Asian groups around early (20.25 Mya) and middle Miocene (13.22 Mya) (Zhou et al., 2019b). A robust and accurate phylogeny is needed to function as the basis for devising a solid taxonomic framework of Asian Sonerileae as well as addressing character evolution, diversification, and biogeographic history.

Disentangling the relationships of lineages that have undergone rapid radiations (both ancient and recent) is especially challenging (Patel et al., 2013; McLean et al., 2019). During radiation, many lineages evolve from a common ancestor in a short period, leaving very few nucleotide differences reflecting shared ancestry. In the case of Asian Sonerileae, nrITS resolved only 55% of the nodes (Bayesian posterior probabilities $\geq$ 0.99; bootstrap support values $\geq$ 70%) (Zhou et al., 2019a). Although plastid data resolved 87% of the nodes (Zhou et al., 2019b), these are uniparentally inherited, providing only one perspective of evolutionary history (Gitzendanner et al., 2018) and thus have a limited use in resolving groups with complex history (Zhang et al., 2014). The biparentally inherited nuclear genome contains multiple independent (unlinked) loci with faster rates of sequence evolution than chloroplast genes in plants (Gaut, 1998; Small et al., 2004), which provides a promising option for tackling difficult phylogenetic questions. Recent development of next-generation sequencing (NGS) technology has enabled the use of hundreds of nuclear genes in the study of previously intractable plant lineages (e.g., Huang et al., 2016; Pouchon et al., 2018; Nikolov et al., 2019; Maurin et al., 2021). However, only 23 species of Asian Sonerileae had been sampled in such datasets (Baker et al., 2021; Maurin et al., 2021).

Gene tree discordance is ubiquitous in phylogenomic studies. Multiple factors can be the cause of discordance, such as random noise from uninformative genes, systematic error, incomplete lineage sorting (ILS), hybridization/introgression, and paralogy (Schrempf and Szöllösi, 2020). These factors can act alone or, more often, in combination (Meyer et al., 2017; Knowles et al., 2018; Morales-Briones et al., 2021), especially in clades that have undergone rapid radiation. A recent simulation study (Molloy and Warnow, 2018) showed that the level of ILS and gene tree estimation error (GTEE) impacted the accuracy of species tree estimation methods (concatenation method, site-based coalescent method, and summary method) differently. Analytical methods accounting for some of the above factors have been developed in recent years (e.g., Edwards et al., 2016; Mirarab et al., 2016; Solís-Lemus and Ané, 2016; Xu and Yang, 2016; Wen and Nakhleh, 2018; Zhu et al., 2019), however, an efficient method modeling all factors simultaneously is not yet available. Therefore, application of multiple methods accounting for these factors is of fundamental importance for understanding the evolutionary processes as well as accurate phylogenetic reconstruction.

This study explores the use of NGS data for resolving phylogenetic relationships within Asian Sonerileae, a lineage that has undergone rapid radiation. To this end, we identified 332 single-copy genes based on the draft genome of *Tigridiopalma magnifica* C.Chen (Sonerileae) and 10 transcriptomes sampled across Asian Sonerileae. Three parallel datasets, viz. single-copy ortholog (SCO) dataset, nuclear genomic dataset, and plastome dataset, were assembled from genome resequencing data of 205 species. For phylogenetic reconstruction, partitioned concatenation was used for the genomic and plastome dataset, and three methods (concatenation, site-based coalescent, and summary methods) were used for the SCO dataset. We then assessed the phylogenetic discordance among SCO gene trees and that among species trees reconstructed using different methods or datasets. The main sources of discordance were explored using multiple approaches [i.e., coalescent-based species tree, Quartet Sampling, long branch attraction (LBA) test, simplex plot of quartet concordance factors, plastid gene tree simulations, and Patterson's *D*-statistic]. Finally, we discussed morphological homoplasy and generic circumscriptions in Asian Sonerileae based on the phylogenetic framework.

2. Materials and methods

2.1. Taxon sampling, DNA/RNA extraction, and sequencing

Ten species of Asian Sonerileae representing different lineages were collected for transcriptome sequencing and subsequent ortholog identification, viz. *Bredia hirsuta* Blume, *Fordiophyton peperomiifolium* (Oliv.) C.Hansen, *Oxyspora paniculata* DC., *Phyllagathis cavaleriei* Guillaumin, *P. setotheca* H.L.Li, *P. rufa* (Stapf) Cellin., *P. tetrandra* Diels, *Sarcopyramis nepalensis* Wall., *Scorpiothyrsus shangszeensis* C.Chen and *Tashiroea amoena* (Diels) R.Zhou & Ying Liu. For phylogenetic analyses, we sampled 205 species (223 accessions) from Sonerileae (185 Asian, 4 African and 1 American), Dissochaeteae (11 spp.), Melastomateae (2 spp.), Blakeeae (1 sp.) and Bertolonieae (1 sp.). For Asian Sonerileae, 74% of the genera (24) were included. *Bertolonia acuminata* Gardner (Bertolonieae) was chosen as an outgroup following Bacci et al. (2019), Zhou et al. (2019b) and Kartonegoro et al. (2021). A complete list of the sampled taxa and their voucher information is provided in Supplementary Table S1.

Total DNA was extracted using HiPure Plant DNA Mini Kit (Magen, Guangzhou, China) or the modified CTAB protocol (Doyle and Doyle, 1987). RNA was extracted from fresh leaf tissues with NucleoSpin RNA Plant and Fungi Kit (Gene Company Limited, Guangzhou) according to the manufacturer's protocol. RNA quality was assessed using agarose gel electrophoresis. Paired-end Illumina genomic libraries were prepared

and then sequenced on the HiSeq™ 2500 platform at Vazyme (Nanjing, China)/Novogene (Beijing, China). Raw reads were first subjected to quality control using Trimmomatic (Bolger et al., 2014). The first seven nucleotides of each read were trimmed for transcriptome sequencing data. Nucleotides on both ends with quality score < 5 and 4-bp sliding windows with average quality score < 15 were trimmed for reads of both transcriptome and re-sequencing data. Reads ≥ 10% unidentified nucleotides and those < 130 bp after trimming were then removed. We also removed reads with a phred quality < 13 for > 40% of read length using the IlluQC_PRLL.pl script in NGSQCToolkit v2.3.3 (Patel and Jain, 2012). Finally, duplicated reads were removed for genome resequencing data using FastUniq (Xu et al., 2012). The resulting reads were assessed using FastQC (Andrews, 2010).

2.2. Transcriptome assembly and identification of single-copy orthologs (SCOs)

Clean reads were de novo assembled using Trinity v2.1.1 (Grabherr et al., 2011) with default parameters, except that the minimum kmer coverage was set to 2. Redundantly assembled sequences were removed using cd-hit v4.6 algorithm (Li and Godzik, 2006) with a sequence identity threshold of 0.90 (-c 0.90) and minimum coverage of 0.8 (-aS 0.8). The longest isoform within a cluster was retained to construct the unigene set. The assembled transcriptomes were used as input into Orthofinder v1.1.4 (Emms and Kelly, 2015) to identify single-copy orthologs (SCOs) for phylogenetic reconstruction, with e-value set to 1e-5. The resulting SCOs were then aligned to the draft genome of *Tigridiopalma magnifica* (available at https://doi.org/10.17632/4bmxjhxv4m.1) using Blastn (Mcginnis and Madden, 2004). Genes aligned to a single locus of the draft genome were retained. The MITObim (Hahn et al., 2013) approach was then adopted to retrieve the intron sequences of the identified SCOs in *Bredia hirsuta* using cDNA sequences as seeds and resequencing data (>100×) of the same species as readpool. Maximum iteration was set to 20. Firstly, reads partly or fully overlapping with the seed were fished from the readpool and then de novo assembled into a new seed for the next iteration. This process was repeated until the seeds resulting from two successive iterations retrieved the same number of reads from the readpool. Finally, SCOs of *B. hirsuta* with intron sequences recovered were assembled into a partial reference genome with each SCO treated as a separate scaffold.

2.3. Assembly of SCO, genomic, and plastome datasets

2.3.1. SCO dataset

High-quality reads from genome re-sequencing data (223 accessions) were mapped to the partial reference genome of *Bredia hirsuta* using BWA-MEM (Li and Durbin, 2010). Single nucleotide polymorphisms (SNPs) and short InDels were called with HaplotypeCaller in GATK v3.7.0 (McKenna et al., 2010) for each sample separately under the GVCF mode. Hard filtering was performed to reduce false positives with the following parameters: (1) QD < 2.5; (2) FS > 15.0; (3) MQ < 25.0; (4) SOR > 10.0; (5) ReadPosRankSum < -1.0; (6) InbreedingCoeff < -0.5. Finally, the consensus option in BCFtools (https://github.com/samtools/bcftools) was used to replace corresponding positions of the partial reference genome with SNP information of each sample, resulting in SCO sequences of the 223 samples. Sequences were then aligned using MAFFT v7.042 (Katoh and Standley, 2013) with default settings.

2.3.2. Genomic dataset

We assembled a genomic dataset by mapping the reads to the draft genome of *Tigridiopalma magnifica*. Mapping, SNP calling, and filtering were carried out as described above. The resulting SNPs were further pruned based on their linkage disequilibrium (LD) patterns using the –indep-pairwise option in PLINK (Purcell et al., 2007). For each SNP pair with an r^2 value above 0.25 within a 50-SNP sliding window (advanced by 5 SNPs each time), only one SNP was retained.

2.3.3. Plastome dataset

Plastomes were assembled in NOVOPlasty v1.2.4 (Dierckxsens et al., 2017) and aligned as described in Zhou et al. (2019b). Only one copy of the inverted repeat (IR) region was used in the final alignment. Dubiously aligned regions were removed using trimAl v1.2 with "-gappyout" mode (Capella-Gutiérrez et al., 2009) as they may bias phylogenetic inferences (Misof and Misof, 2009; Reginato et al., 2016).

2.4. Phylogenetic analyses

2.4.1. SCO dataset

The concatenation approach, summary method, and site-based coalescent method are among the most widely used methods of species tree estimation in recent phylogenomic studies. These approaches differ in methodology of multilocus phylogenetics (total evidence/coalescent-based), the type of phylogenetic data they use (sites/gene tree), and their statistical consistence under the Multi-Species Coalescent (MSC) model and perform differently for empirical datasets with various levels of ILS and gene tree estimation error (GTEE) (Kubatko and Degnan, 2007; Huang et al., 2010; Leaché and Rannala, 2010; Patel et al., 2013; Mirarab et al., 2014a; Mirarab and Warnow, 2015; Roch and Steel, 2015; Mirarab et al., 2016; Molloy and Warnow, 2018). Therefore, we analyzed the SCO dataset using concatenation, a two-step summary method, and a site-based coalescent method. Unpartitioned and partitioned maximum likelihood analyses were conducted in IQ-TREE v2.1.2 (Nguyen et al., 2015) and the results were compared to evaluate the impact of data partitioning on phylogenetic reconstruction. In the former analysis, TVMe + R6 model was selected, in the latter, optimal partitioning scheme and best fitting model for each partition (Table S2) was selected, both using ModelFinder (Kalyaanamoorthy et al., 2017) under the Bayesian Information Criterion (BIC). Node support was evaluated by 1000 replicates of ultrafast bootstrap (UFBS) (Minh et al., 2013) and SH-aLRT test. For a two-step coalescent-based method, gene trees were estimated in RAxML v8.2.10 (Stamatakis, 2014) using the GTR + G model, with nodes below 10% bootstrap support collapsed using NEWICK Utilities v1.6 (Junier and Zdobnov, 2010). A coalescence species tree was then reconstructed in ASTRAL v5.7.4 (Mirarab et al., 2014b) with quartet support values computed to evaluate the level of gene tree incongruence. For site-based method, SVDquartets (Chifman and Kubatko, 2014) implemented in PAUP* v4.0a159 (Swofford, 2003) was used to estimate a coalescent-based species tree from all possible quartet trees using concatenated SCO alignment as input. Finally, node support was measured by nonparametric bootstrapping with 500 replicates.

2.4.2. Genomic dataset

To evaluate whether genomic SNPs data could resolve the phylogeny better than the SCO dataset, maximum likelihood analysis of the genomic dataset was conducted in IQ-TREE using a partitioned approach. The dataset was partitioned into bins of equal length (2000 SNPs). Substitution model GTR + F + ASC + G4 was selected for all partitions using ModelFinder under BIC. Node support was evaluated by 1000 replicates of UFBS and SH-aLRT test.

2.4.3. Plastome dataset

Maximum likelihood analysis of the plastome dataset was conducted in RAxML v8.2.10. Zhou et al. (2019b) tested five partitioning schemes for the plastome dataset of Sonerileae. Here we followed their optimal partitioning scheme, viz. three partitions corresponding to large single copy (LSC) region, small single copy (SSC) region, and IR region. The substitution model GTR + G was selected for all three partitions as recommended by the author. We evaluated node support with 1000 bootstrap replicates using a fast bootstrapping algorithm in RAxML (Stamatakis et al., 2008).

2.5. Incongruence assessment and visualization

The plastome, genomic, and SCO trees were first manually inspected to identify incongruent topologies. We then used the Quartet Sampling (QS) method (Pease et al., 2018) to assess congruence and identify the strong phylogenetic conflicts. This method is a quartet-based evaluation system that evaluates the confidence, consistency, and informativeness for each branch in the given phylogeny (Pease et al., 2018). The plastome, genomic, and SCO datasets and the species trees were used as input, with the number of replicate quartet searches per branch set to 500 and the log-likelihood cut-off to 2. Discordance among species trees was visualized using a R script (Kong et al., 2021). We also used the multidimensional scaling (MDS) of tree space to visualize the extent of discordance among gene trees and species trees. Pairwise Robinson–Foulds (RF) distances among trees (Duchêne et al., 2018) were calculated using the R package APE (Paradis and Schliep, 2019). MDS was then used to plot the RF distances with different dimensions. MDS points of individual SCO gene trees were colored by average bootstrap support values. Pearson's correlation coefficient was used to assess the relationship between information content of each SCO and per-gene tree distance from the ASTRAL species tree.

2.6. Test for long branch attraction (LBA)

We tested the occurrence of long-branch attraction in the SCO species trees from maximum likelihood analyses as well as gene trees used in ASTRAL. Analyses were carried out using TreeShrink (Mai and Mirarab, 2018), with a false-positive error rate (α) of 0.05. Output gene trees with the suggested tips removed were used to reconstruct a species tree in ASTRAL to assess possible topological changes with strong support.

2.7. Assessing incomplete lineage sorting

We addressed the extent of ILS, using an R package MSCquartet (Allman et al., 2021; Rhodes et al., 2021). Given a set of gene trees, MSCquartet calculates frequencies of quartets displayed on these trees (quartet count concordance factor, qcCF) for all choices of four taxa. All qcCFs in a gene tree collection can be visualized using the simplex plot (Allman et al., 2021), which summarizes all gene tree discord and provides a visual comparison to the expected discord under Multispecies Coalescent model (MSC) of ILS on a given species tree, indicating how closely the model fits the data. To assess ILS, 332 SCO gene trees were used as input. The 223 tips were divided into three subsets. Tree hypothesis test (T1) for all quartet counts were performed using the function quartetTreeTest under rejection levels of $\alpha = 0.01$, 0.001, and 0.0001. Finally, quartetTestPlot was used to generate a simplex plot for each level of rejection with color-coded symbols indicating qcCFs consistent with MSC of ILS or not. Symbols close to the vertex indicate little ILS, those near and along the centroid indicate substantial ILS, while those rejected should be explained by processes other than ILS.

2.8. Plastid gene tree simulations

To test whether ILS alone can cause the observed cytonuclear discordance, we simulated plastid gene trees under the coalescent model and summarized the clade frequencies on the plastid topology. If cytonuclear discord is caused by ILS alone, topologies from the empirical plastid tree are expected to be present in the simulated trees with high frequency. On the other hand, if hybridization/introgression is involved instead, the frequencies are expected to be very low or zero. Three selected subsets with strong cytonuclear conflict were tested: (1) *Bredia* Blume; (2) *Blastus* Lour., abridged *Bredia*, the *Kerriothyrsus* C. Hansen clade, *Fordiophyton breviscapum* (C.Chen) Y.F.Deng & T.L.Wu, *Phyllagathis erecta* (S.Y.Hu) C.Y.Wu ex C.Chen, and the *Plagiopetalum* clade; (3) *Medinilla*, *Tigridiopalma* C.Chen, *Phyllagathis prostrata* C. Hansen, *Heteroblemma* (Blume) Cámara-Leret, Ridd.-Num. & Veldkamp, *Driessenia* Korth. (isomorphic) clade, *Phyllagathis* (thyrsoid) clade, and *Phyllagathis* (pentamerous) clade. Five thousand plastid gene trees were simulated for each subset with DendroPy v4.5.2 package (Sukumaran and Holder, 2010) under the coalescent scenario, using the SCO ASTRAL species tree as the guide tree. Branch lengths were scaled by four as the effective population size of plastid genes is fourfold smaller comparing to nuclear genes. Simulated clade frequencies were summarized and plotted on the ASTRAL species trees and plastid tree using phyparts v0.0.1 (https://bitbucket.org/blackrim/phyparts).

2.9. Test for reticulation

Patterson's *D*-statistic (or ABBA-BABA test), also known as the ABBA-BABA statistic, and the related admixture fraction f_4-ratio (Patterson et al., 2012) are often used to detect gene flow between populations or closely related species based on genomic biallelic SNPs across four populations/taxa. However, interpretation of these results is not always straightforward as a single gene flow may cause multiple elevated D and f_4-ratio when the taxa share branches on the phylogeny (Malinsky et al., 2021). Dsuite (Malinsky et al., 2021) allows direct calculations of the *D* and f_4 statistics across all trios of species from genomic SNPs in a variant call format (VCF) file, and a program Fbranch is implemented to disentangle correlated f_4-ratio results and assign gene flow evidence to specific branches on a phylogeny. Here, we used Dsuite to assess the level of gene flow and introgression based on the genomic dataset (VCF files containing biallelic genomic SNPs). Fourteen selected data subsets were analyzed, each containing 6 to 38 terminals (Table S3). The *D* and f_4-ratio statistics (Patterson et al., 2012) were calculated for all trios using Dtrios. Fbranch statistics were then calculated using subtrees extracted from the ASTRAL species tree as input. The output *f*-branch statistics in matrix-like format were then displayed using the plotting function.

2.10. Morphological characters and coding

Fourteen characters were coded (Table S4), including growth form (not climbing/climbing), floral merosity (isomerous/anisomerous; 3-merous/4-merous/5-merous/6-merous/4- and 5-merous), stamen arrangement (two whorls, fertile/two whorls with one sterile/one whorl), stamen morphology (isomorphic/dimorphic), scorpioid cymes (absence/presence), fruit type (berry/capsule), enlarged ovary crown (absence/presence), shape of ovary top in mature fruit (humped or rounded/inverted frustum/obpyramidal), persistence of placental column and placentas upon capsule dehiscence (disintegrating/persistent), crystal form (druses/raphides), and the absence/presence of horned placenta column, thready placenta, and minute hyaline glands. Morphological data were gathered via examination of fresh and herbarium materials and literature survey (Chen, 1984a; Hansen, 1984, 1985a, 1987a, 1987b, 1988; Maxwell, 1989; Regalado, 1990; Hansen, 1990, 1992; Mentink and Baas, 1992; Cellinese, 2002, 2003; Chen and Renner, 2007; Penneys and Judd, 2011, 2013; Yeh et al., 2008; Kartonegoro and Veldkamp, 2010; Cámara-Leret et al., 2013; Lin et al., 2015; Lin et al., 2017). Crystal form was examined using either silica-gel dried samples or fresh leaves. Leaves were cleaned, softened (for dried samples) and sectioned using a double-edge razor blade. Sections were mounted in distilled water under a cover-slip and examined using Motic BA410E microscope. The morphological matrix was edited and reconstructed onto the ASTRAL species tree inferred based on SCO dataset, using both maximum parsimony (MP) and maximum likelihood (ML) in Mesquite v3.50 (Maddison and Maddison, 2018). For MP analyses, characters were treated as unordered and with equal weight, and multistate characters were treated as non-additive. The Markov k-state 1 parameter model (Mk1 model) of evolution was used for ML reconstruction, with equal probability for any character change. The consistency index (CI) and retention index (RI) were calculated to evaluate the

level of homoplasy.

3. Results

3.1. Transcriptome assembly and identification of single-copy orthologs

Ten transcriptomes of Asian Sonerileae were generated in this study. Raw reads were deposited in NCBI Sequence Read Archive (PRJNA794323). After assembly and redundancy reduction, 112,091 to 285,464 unigenes were obtained for each sample with total length ranging from 124,915,590 to 232,635,681 nt. Voucher information and statistics of the transcriptomes are summarized in Supplementary Table S5. Of the 1265 SCOs identified using ten transcriptomes, 332 aligned to a single locus of the draft genome of *Tigridiopalma magnifica* and were thus retained for subsequent analyses.

3.2. Assembly of SCO, genomic, and plastome datasets

The partial reference of *Bredia hirsuta* was 898 kb long, containing 332 SCOs with their intron sequences recovered. The final SCO dataset contained 332 genes from 223 samples (available at https://doi.org/10.17632/4bmxjhxv4m.1). The concatenated matrix had 941,176 aligned bp with 89,612 parsimony informative sites and 11.75% of missing data. After SNP filtering and pruning, the genomic dataset contained 378,459 SNPs, from which 308,984 were parsimony informative, with 33.79% of missing data (available at https://doi.org/10.17632/4bmxjhxv4m.1). Eighty-four complete chloroplast genomes in Sonerileae were newly sequenced and assembled for this study (Table S1). The final alignment of the plastome dataset was 129,573 bp with 24,435 parsimony informative sites.

3.3. Phylogenetic analyses

3.3.1. SCO dataset and genomic dataset

We reconstructed the phylogeny of Sonerileae based on the SCO dataset using three different methods. The unpartitioned (Fig. S1) and partitioned (Fig. S2) ML tree recovered nearly identical topology, with 91.3% and 90.5% of the nodes well resolved respectively (SH-aLRT test $\geq$ 80%, UFBS $\geq$ 95%). The ASTRAL species tree (Fig. S3), with 88% of the nodes strongly supported (LPP $\geq$ 0.7), was highly similar to the ML trees in overall topology except for some weakly supported nodes. The species tree from SVDquartets was the least resolved (Fig. S4) with 80% of the nodes well supported (nonparametric BS $\geq$ 80%). The genomic ML tree (Fig. S5) showed similar overall topology and phylogenetic resolution to SCO ML trees, with 92.2% of the nodes well supported (SH-aLRT test $\geq$ 80%, UFBS $\geq$ 95%). The SCO trees recovered Sonerileae as monophyletic with full support, with the Neotropical lineage *Opisthocentra* Hook.f. as sister to the remaining members of the tribe, followed by a split between the Afrotropical clade (SH-aLRT test = 100%, UFBS = 100%; LPP = 1; nonparametric BS = 99.7%) and the mainly Asian superclade (SH-aLRT test = 100%, UFBS = 100%; LPP = 1; nonparametric BS = 100%). In the genomic tree, however, Sonerileae was nonmonophyletic. *Opisthocentra* was resolved as sister to *Blakea* P. Browne (SH-aLRT test = 59%, UFBS = 97%), while the Afrotropical clade + mainly Asian superclade of Sonerileae was sister to Dissochaeteae + Melastomateae (SH-aLRT test = 100%, UFBS = 100%).

Both genomic and SCO trees recovered thirty-four morphologically cohesive lineages in the Asian Sonerileae superclade which were summarized in Supplementary Table S6 (also see Fig. 1). Of the 34 nodes concerning relationships among these lineages, 25 (74%) were supported by both SCO and genomic datasets (Fig. 1). Some lineages consistently formed close allies, viz. clade A [*Heteroblemma*, *Phyllagathis* (thyrsoid) clade, *Driessenia* (isomorphic) clade, and *Phyllagathis* (pentamerous) clade], clade B (*Tashiroea* Matsum. ex T.Itô & Matsum., *Poilannammia* C.Hansen clade, *Scorpiothyrsus* clade, and *Driessenia axantha* Korth.), clade C (*Medinilla*, *Tigridiopalma*, and *Phyllagathis prostrata* clade), clade D (*Sarcopyramis* Wall. and *Barthea* Hook.f.), clade E [*Anerincleistus* clade, *Phyllagathis scortechinii* King, *Phyllagathis griffithii* King-*Oxyspora bullata* (Griff.) J.F.Maxwell, and *Phyllagathis tuberculata* King-*P. hispida* King], clade F [*Aschistanthera* C.Hansen, *Styrophyton* S.Y. Hu clade, and *Oxyspora* DC. (Bonean) clade], clade G (*Sporoxeia* W.W. Sm. clade, *Phyllagathis cavaleriei*, *Cyphotheca* clade, and *Plagiopetalum* clade), clade H (*Kerriothyrsus* clade, *Phyllagathis erecta*, and *Fordiophyton breviscapum*), and clade I (*Bredia*, *Blastus*, and clade H).

3.3.2. Plastome dataset

As shown in Supplementary Fig. S6, 86% of the nodes in the partitioned concatenated ML tree from RAxML received strong support (BSML $\geq$ 80%). Sonerileae was fully resolved as monophyletic, with *Opisthocentra* sister to Afrotropical clade + Asian superclade. Most of the major lineages recovered from SCO/genomic trees were also strongly supported by plastome tree, except that *Blastus*, *Driessenia* (isomorphic) clade, *Heteroblemma*, *Phyllagathis* (pentamerous) clade, *Phyllagathis* (thyrsoid) clade, *Phyllagathis tuberculata-P. hispida* were resolved as nonmonophyletic (Table S6).

3.4. Incongruence assessment and visualization

Discordance among SCO, genomic, and plastome species trees were visualized in Supplementary Fig. S7. Incongruences were identified among species trees at both interspecific and interclade level. Quartet Sampling results were presented in Supplementary Figs. S8–12. Nodes discordant among genomic and SCO trees had an average QI value of 0.88 and 70% of them gained weak support ($0 < QC < 0.2$) or counter-support ($QC < 0$), while in the plastome tree, nodes discordant with SCO/genomic trees had an average QI value of 0.93 and 70% of them were strongly supported ($QC > 0.2$). A tree space based on pairwise RF distances among all tree topologies (SCO, genomic, plastome datasets, and individual SCOs) was presented in Fig. 2A. The average BSML value of individual nuclear gene tree ranged from 29% to 87% (after collapsing nodes with BSML < 10%). The pairwise RF distances between individual gene trees and the SCO species trees were significantly negatively correlated to the average BSML values (Pearson's $r = 0.93$, $P < 2e{-}16$): gene trees with the highest values were closest to the species trees while those with lowest values were furthest (Fig. 2A). Among SCO species trees inferred using different approaches, the ML tree and ASTRAL tree were more congruent (distance of 46) with each other than SVDquartets tree (72 and 80) (Table S7) and therefore were retained for subsequent discussion. Among all species trees, the SCO ML tree was the congruent (average distance of 93), whereas the plastome tree was the most controversial (average distance of 214) (Table S7).

3.5. Test for long branch attraction (LBA)

The ASTRAL species tree reconstructed using gene trees with possible outlier tips removed was shown in Supplementary Fig. S13. Only two topological changes were detected comparing to the original ASTRAL tree, but both with weak support (LPP < 0.5): (1) in *Phyllagathis* (raphides) clade, the *P. ovalifolia* H.L.Li subclade was sister to the remaining of *Phyllagathis* (raphides) clade instead of the *P. stenophylla* (Merr. & Chun) H.L.Li subclade; (2) the sister relationship of *P. prostrata* clade and *Tigridiopalma* collapsed.

3.6. Assessing incomplete lineage sorting

Three data subsets were analyzed using MSCquartets, containing 77, 69, and 77 tips respectively. The quartet number of the three subsets were 1353275, 864501, and 1353275. As presented in Fig. 2B and Supplementary Fig. S14, most quartet trees were consistent with ILS under the MSC (blue circles) and located near and along the centroid, indicating substantial ILS. Nevertheless, 417, 521, and 206 quartet trees were still rejected for the three subsets under rejection level of α =

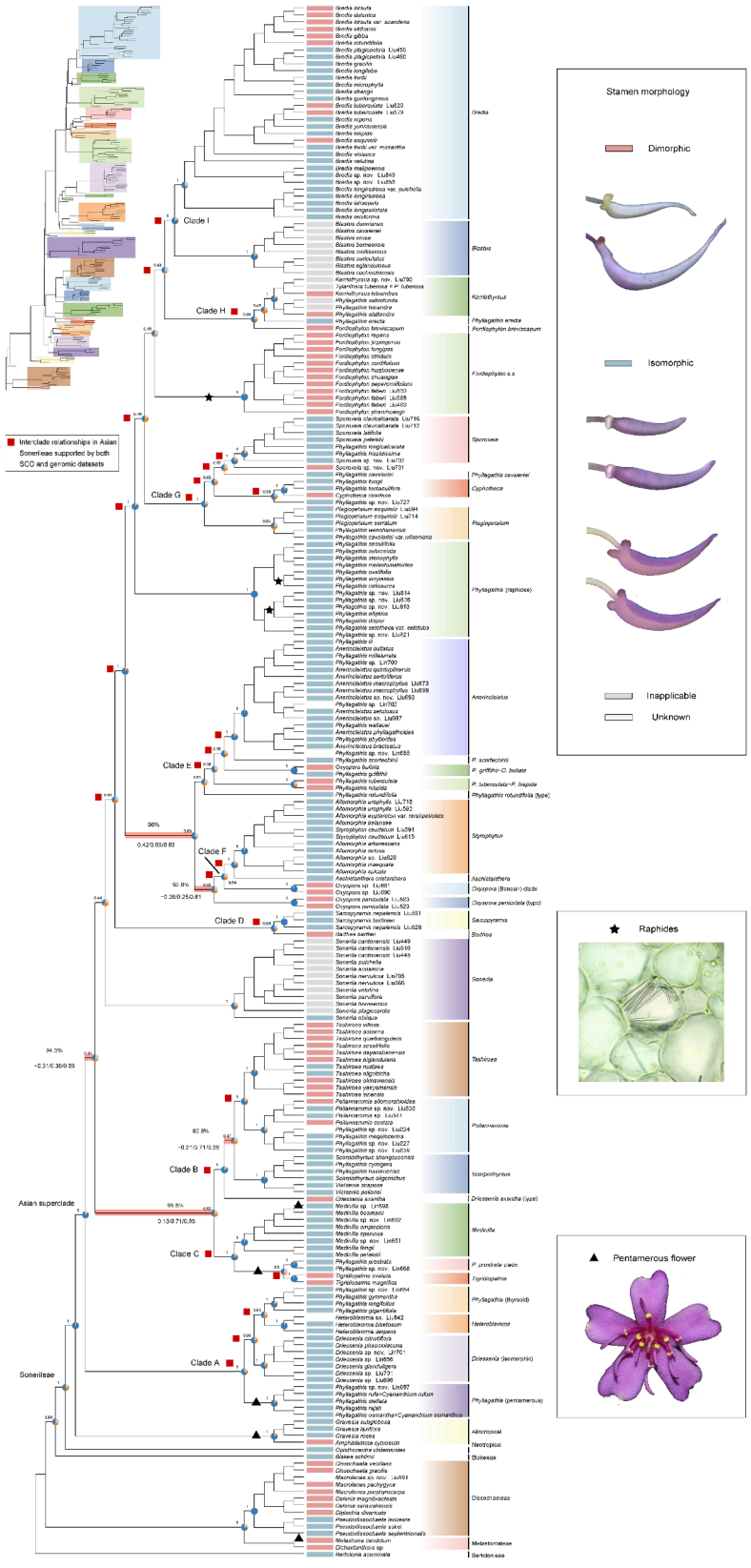

Fig. 1. Coalescence species tree of Sonerileae inferred from 332 single copy orthologs (SCOs) shown as a cladogram (phylogram inset). For major clades and interclade relationships, support values (local posterior probability) are indicated above the branches. Pie charts indicate proportion of quartet trees supporting that clade (blue), the main alternative bifurcation (orange), and the remaining alternative (grey). Red squares indicate interclade relationships supported by both SCO and genomic datasets. The shaded branches indicate interclade relationships incongruent between SCO ASTRAL species tree and the partitioned maximum likelihood (ML) tree, with proportion of uninformative gene trees and Quartet Sampling (QS) scores (quartet concordance/quartet differential/quartet informativeness) noted above and below the branches, respectively. The highly homoplasious stamen morphology is indicated at the tips using color bars. Two characters showing low homoplasy were noted on the branches, viz. the presence of raphides (star) and pentamerous flowers (triangle). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

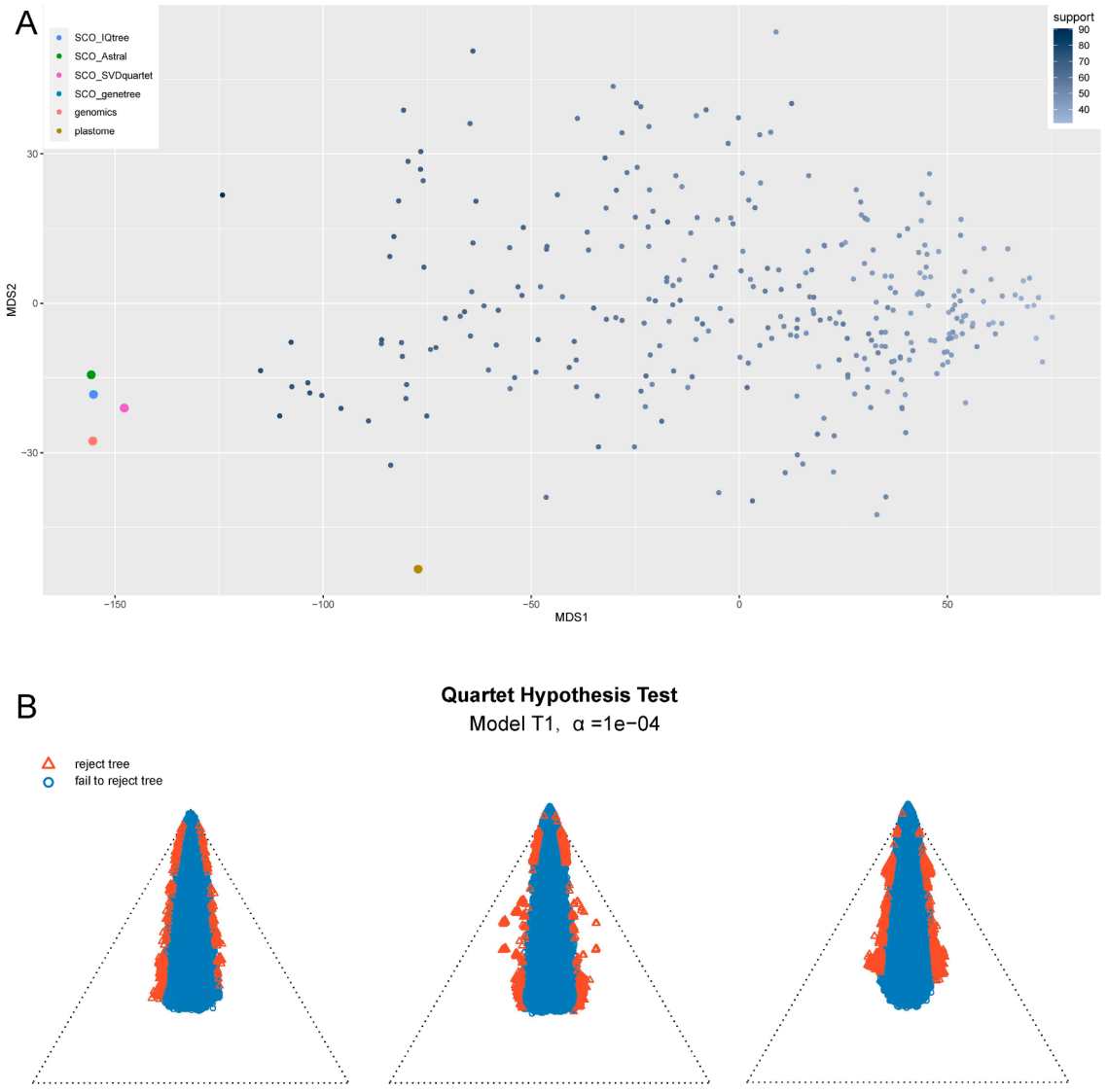

Fig. 2. (A) Multidimensional scaling (MDS) of pairwise Robinson–Foulds distance among 332 single copy ortholog (SCO) gene trees, SCO species trees inferred using three different approaches, genomic tree, and plastome tree. Points of individual SCO gene trees were colored by average bootstrap support values. (B) Simplex plots of quartet concordance factors (qcCFs) under the multi-species coalescent (MSC) model of incomplete lineage sorting (ILS) with T1 model and a rejection level of $\alpha = 0.0001$ based on the SCO ASTRAL species tree, dividing the 223 tips into three data subsets. Blue circles in the plots indicate qcCFs consistent with ILS under the MSC on the species tree, while orange triangles indicate those rejected. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

0.0001, indicating that mechanisms other than ILS should also be considered as potential causes of discordance.

3.7. Plastid gene tree simulations

Five thousand trees were simulated for each of the three subsets showing strong cytonuclear discordance. The ASTRAL species trees and plastome ML trees with simulated clade frequencies were presented in Supplementary Figs. S15–17. In subset 1 (*Bredia*), three incongruent nodes from the plastome tree had near zero frequencies, while the remaining had lower frequencies than those of the ASTRAL tree (Fig. S15). In subset 2, all incongruent nodes concerning the phylogenetic position of the two *Blastus* lineages in the plastome tree were absent in the simulated trees (Fig. S16). In subset 3, the phylogenetic position of *Medinilla* and *Phyllagathis prostrata* in the plastome tree had zero or near zero simulated clade frequencies, while the relationships within *Heteroblemma* + *Phyllagathis* (thyrsoid) clade and *Driessenia* (isomorphic) clade + *Phyllagathis* (pentamerous) clade had similar or lower frequencies than topologies from the ASTRAL tree (Fig. S17). These results showed that ILS cannot explain many of the strong cytonuclear conflicts detected in this study.

3.8. Test for reticulation

Heat maps from Fbranch statistics were presented in Supplementary Figs. S18–31. Introgression was detected among closely related species within some major clades, viz. *Blastus*, *Bredia*, *Fordiophyton* Stapf clade, *Kerriothyrsus* clade, *Medinilla*, *Phyllagathis* (raphides) clade, *Phyllagathis* (thyrsoid) clade, *Poilannammia* clade, *Scorpiothyrsus* clade, *Sonerila*, *Sporoxeia* clade, *Styrophyton* clade, and *Tashiroea*. Introgression signals among closely related clades were also detected: (1) *Blastus*, *Bredia*, *Kerriothrysus* clade, *Fordiophyton breviscapum* and *Phyllagathis erecta*; (2) *Cyphotheca* clade, *Plagiopetalum* clade, and *Sporoxeia* clade; (3) *Phyllagathis rotundifolia* (Jack) Blume (type) and *Oxyspora* (Bornean) clade; (4) *Oxyspora paniculata* (type) and *Styrophyton* clade; (5) four lineages in clade A, namely *Driessenia* (isomorphic) clade, *Heteroblemma*,

Phyllagathis (pentamerous) clade, and *Phyllagathis* (thyrsoid) clade; (6) clades A, B, and C.

3.9. Morphological inference

Fourteen morphological characters were reconstructed onto the ASTRAL species tree based on SCO dataset (Figs. S32–45). Reconstruction of characters using two different methods yielded overall similar results, but ML generally returned more nodes with equivocal character states than MP. Of the 14 characters tested, growth form (CI = 0.33, RI = 0.82), floral merosity [isomerous/anisomerous (CI = 1.0, RI = 0); 3-merous/4-merous/5-merous/6-merous/4- and 5-merous (CI = 0.4, RI = 0.77)], stamen arrangement (CI = 0.25, RI = 0.77), fruit type (CI = 0.25, RI = 0.86), crystal form (CI = 0.33, RI = 0.89), and presence/absence of hyaline glands (CI = 0.20, RI = 0.82) showed relatively low homoplasy, while stamen morphology (CI = 0.05, RI = 0.63), and presence of scorpioid cymes (CI = 0.05, RI = 0.56), enlarged ovary crown (CI = 0.04, RI = 0.70), horned placental column (CI = 0.06, RI = 0.70), and thready placentas (CI = 0.06, RI = 0.58) were highly homoplasious (Table S8).

4. Discussion

4.1. Causes of phylogenetic discordance

In the present study, we used three datasets (SCO, genomic, and plastome) to reconstruct the phylogeny of Asian Sonerileae. Conflicting phylogenetic signals were detected at various taxonomic levels among SCO gene trees and SCO/genomic/plastome species trees. Phylogenetic discordance can be the product of one or multiple factors, such as stochastic error caused by genes with low information content, missing data, rate variation across lineages, hidden paralogy, incomplete lineage sorting, and hybridization/introgression (Rieseberg et al., 1996; Wendel and Doyle, 1998; Stefanovic et al., 2004; Philippe et al., 2005; Edwards, 2008; Galtier and Daubin, 2008). Exploring these factors is the first step towards understanding the processes shaping the evolution of Asian Sonerileae.

Random noise from uninformative genes is no doubt an important source of discordance. Among 332 nuclear gene trees, 41% had average supporting value below 50% (Fig. 2A), suggesting high GTEE which may reduce the accuracy of summary methods (e.g., Huang et al., 2010; Patel et al., 2013; Lanier and Knowles, 2015; Xi et al., 2015; Molloy and Warnow, 2018). As shown in Fig. 1, for the weakly supported and incongruent nodes connecting some of the major lineages in the tribe (e.g., relationships among clades A, B, and C), the proportion of uninformative quartet gene trees had reached 92%–96%. Low information in most genes for resolving these relationships had led to conflicting but weakly supported topologies.

Missing data or its biased distribution may also impact the accuracy of species tree estimation (Driskell et al., 2004; Hovmöller et al., 2013; Vachaspati and Warnow, 2015; Xi et al., 2016; Molloy and Warnow, 2018). In the present study, we detected one case of discordance unequivocally related to the amount of missing data. Sonerileae was resolved to be monophyletic in SCO and plastome trees, whereas it was diphyletic in the genomic tree, with the Neotropical *Opisthocentra* sister to Blakeeae instead of the remaining Sonerileae. The missing rate of this sample was 94% in the genomic dataset after SNP pruning based on LD patterns, much higher than that of 39% in the SCO dataset. Reanalysis of the genomic dataset without pruning based on LD patterns had reverted to recovering Sonerileae as monophyletic, indicating that this discordance was caused by high missing data for *Opisthocentra* in the pruned genomic dataset.

Variation of evolutionary rate across lineages often leads to LBA (divergent but unrelated taxa erroneously clustering together in the phylogeny), a tree reconstruction artifact not confined to specific tree building method (Philippe and Laurent, 1998; Bergsten, 2005). The SCO ASTRAL trees reconstructed using gene trees with or without the suggested tips removed showed only two topological changes with weak support. Therefore, LBA can be excluded as the main source of discordance.

Paralogy is yet another cause of phylogenetic discordance. When paralogous copies are included in the dataset, the phylogeny may reflect duplication history of the gene instead of speciation (Wendel and Doyle, 1998). The genomic dataset almost surely contained paralogs. While SCOs used in this study were identified based on a draft genome and 10 transcriptomes representing different clades of Sonerileae, the possibility of hidden paralogs resulted from gene duplication/gene loss cannot be completely ruled out. If paralogy is one of the main sources of discordance, phylogeny based on genomic dataset (which contained more paralogs) would be expected to deviate considerably from the SCO phylogeny. Nevertheless, the RF distance between the genomic tree and SCO trees (ML and ASTRAL) were 42 and 54 respectively, similar to the distance between the latter two trees (46), indicating that phylogenies based on SCO and genomic datasets were no more incongruent than between the SCO trees. Paralogy is therefore rejected as a main source of discordance.

ILS and hybridization/introgression can result in similar patterns of discordance that are often difficult to distinguish (Holder et al., 2001; Buckley et al., 2006; Holland et al., 2008; Joly et al., 2009). In the simplex plot, most nuclear quartet trees consistent with MSC of ILS were located near or along the centroid (Fig. 2B), indicating substantial ILS. Nevertheless, 206 to 521 quartet trees were still rejected for each of the subsets under rejection level of $\alpha = 0.0001$, suggesting that ILS alone was unable to explain all nuclear discordance. Simulation of gene trees under the coalescent model for selected subsets showed that some clades had zero or near zero simulated clade frequencies and therefore should be explained by mechanisms other than ILS, such as introgression. QS results, viz. low QD value (<0.3) on some incongruent nodes, also indicated possible introgression. Fbranch statistics revealed signals of introgression between closely related species in the same clade as well as between closely related clades, many of which were strongly corroborated by cytonuclear discordance, showing that hybridization/introgression was yet another source of the discordance observed.

Based on the above data, conflicting phylogenetic signals in Sonerileae were mainly caused by a combination of biased distribution of missing data, random noise from uninformative genes, ILS, and hybridization/introgression. This is expected, as all the incongruent clades had rather short branch lengths, suggesting periods of rapid radiation. When multiple lineages evolved within a short period, there are very few nucleotide differences reflecting their shared ancestry, while short intervals between speciation events and close relationships facilitate widespread ILS and hybridization, which simultaneously confound accurate phylogenetic reconstruction.

4.2. Understanding recalcitrant relationships: an example

Resolving phylogenetic relationships is difficult when multiple sources of discordance are present. Nevertheless, the use of different tools enables us to generate hypotheses which can be further tested in future studies. A case of discordance in *Bredia*, a genus with complete taxon sampling in this study, is discussed below as an example.

Both nuclear and cytonuclear discordance were detected regarding the relationships among five subclades in *Bredia*, namely *B. gracilis* (Hand.-Mazz.) Diels, *B. hirsuta*, *B. yunnanensis* (H.Lév.) Diels, *B. fordii* (Hance) Diels, and *B. guidongensis* (K.M.Liu & J.Tian) R.Zhou & Ying Liu subclades. *Bredia hirsuta* subclade + *B. gracilis* subclade was resolved as sister to *B. yunnanensis* subclade with weak and full support in SCO and genomic ML trees respectively (topology A, Fig. 3A), whereas it was sister to *B. fordii* subclade in the SCO ASTRAL tree with weak support (topology B, Fig. 3B). Low major quartet scores (q1 = 0.35, 0.37) of the two internodes in topology B indicated high level of discordance among nuclear gene trees. The simplex plot for this subset indicated that ILS

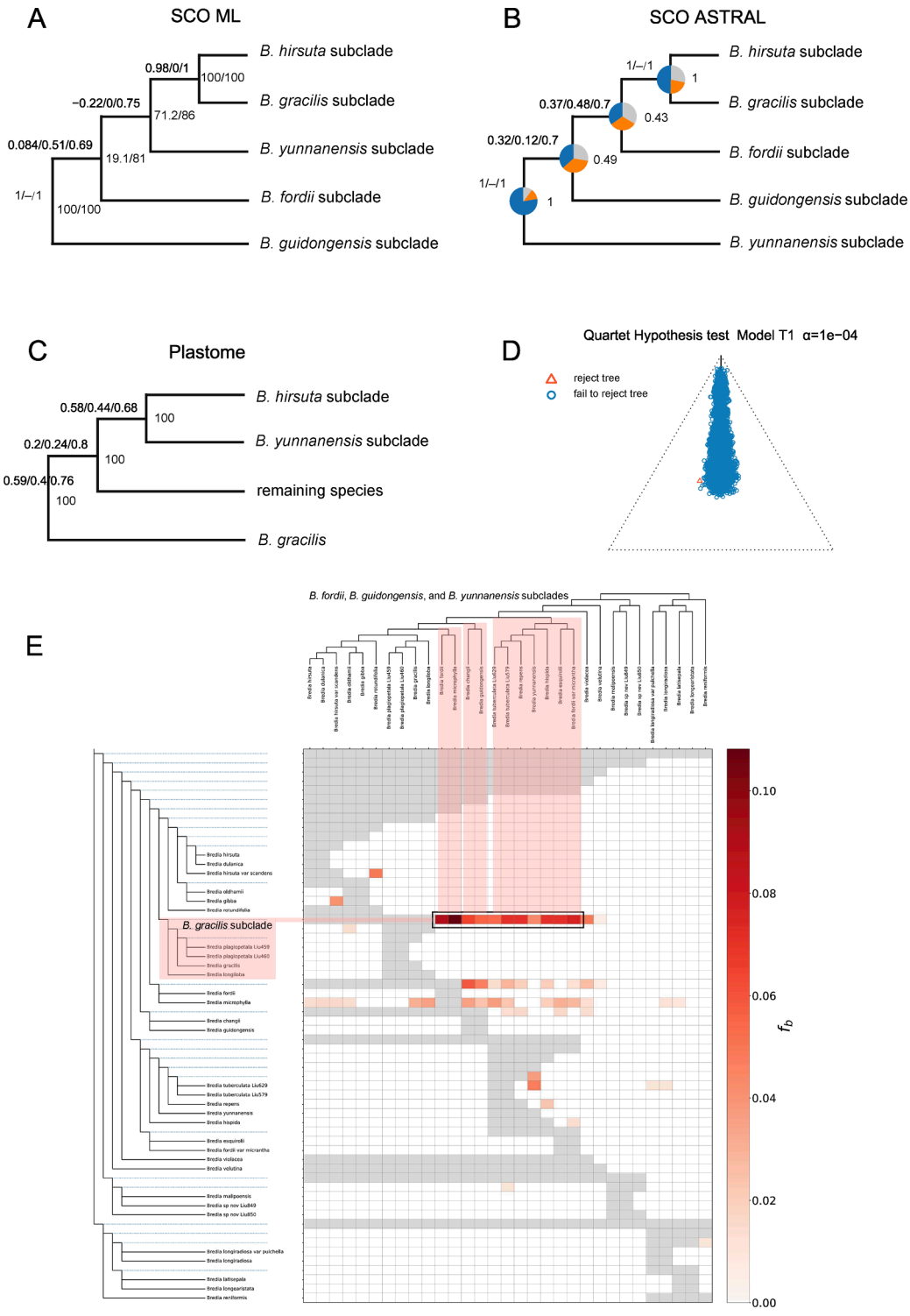

Fig. 3. An example of phylogenetic discordance in *Bredia*. (A–C) Topology A from single copy ortholog (SCO) ML tree, topology B from SCO ASTRAL species tree, and topology C from plastome tree, with Quartet Sampling scores (quartet concordance/quartet differential/quartet informativeness) and support values (SH-aLRT test/ultrafast bootstrap/local posterior probability/bootstrap) indicated at the nodes. Pie charts in topology B indicate proportion of quartet trees supporting that clade (blue), the main alternative bifurcation (orange), and the remaining alternatives (grey). (D) Simplex plot of the quartet concordance factors (qcCFs) under the multi-species coalescent (MSC) model of incomplete lineage sorting (ILS) with T1 model and a rejection level of $\alpha = 0.0001$ based on the SCO ASTRAL subtree containing the five subclades discussed. (E) A heat map from Fbranch statistics of *Bredia*, boxes indicated signals of introgression detected among four subclades. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

alone was able to explain most discordance among nuclear genes, with only one tree rejected under $\alpha = 0.0001$ (Fig. 3D), which favored topology B from ASTRAL, as summary method is usually more accurate than concatenation under substantial ILS. Quartet Sampling results, i.e. counter-support for topology A (QC values < 0) and moderate to strong support (QC values > 0.3) for topology B, also favored the latter. In the plastome tree, however, *B. hirsuta* subclade was sister to *B. yunnanensis* subclade, while species of *B. gracilis*, *B. fordii*, and *B. guidongensis* subclades were nested forming two lineages (topology C, Fig. 3C). Of the five nodes concerning the above relationships in the plastome tree, plastid gene tree simulation returned 1 (0.02%), 1 (0.02%), 3 (0.06%), 9 (0.18%), and 14 (0.28%) branch frequencies (Fig. S15), suggesting that ILS, or at least ILS alone, is unable to explain all cytonuclear conflicts. Fbranch statistics based on genomic data detected distinct signals of introgression between *B. gracilis* subclade and three subclades, viz. *B. fordii*, *B. guidongensis*, and *B. yunnanensis* subclades (Fig. 3E). The range of *B. gracilis* subclade overlaps with that of the latter three subclades in some parts of southern China, which makes natural hybridization possible. The nested position of *B. gracilis* subclade with species of *B. fordii* and *B. guidongensis* subclades in the plastome tree strongly

corroborated the introgression signal detected and vice versa. In brief, simplex plot of qcCFs in 332 SCO gene trees indicated ILS as the main theme. This scenario, together with QS results, better supported topology B (SCO ASTRAL tree). Nevertheless, Fbranch statistics based on genomic data suggested an introgression scenario. This discrepancy is probably caused by the different information content between SCO and genomic datasets. Relationships among the subclades should be further tested using a larger number of SCOs identified specifically for *Bredia*, as the informative sites per gene and the number of genes used are known to impact the accuracy of summary method (Roch and Warnow, 2015; Mirarab et al., 2016).

4.3. Morphological homoplasy

Reconstruction of morphological characters onto the SCO ASTRAL tree detected high level of homoplasy for some of the characters traditionally emphasized in taxonomy. As ML generally returned more nodes with equivocal character states, we based the following discussion on results from MP reconstruction. Stamen morphology (Fig. S36), presence of scorpioid cymes (Fig. S37), and capsule morphology [crowned/uncrowned capsule (Fig. S39), horned/unhorned placental column (Fig. S42), and thready/non-thready placentas (Fig. S43)] were important characters used in generic circumscription (Chen, 1984a, b; Hansen, 1992), all of which are highly homoplasious in Asian Sonerileae. For these characters, shift of character states occurred 19, 15, 22, 15, and 16 times across Asian Sonerileae respectively. These shifts occurred not only across closely related major lineages but also within lineages. For example, shifts between dimorphic and isomorphic stamens were found in *Bredia*, *Tashiroea*, *Cyphotheca* clade, and *Poilannammia* clade (Figs. 1, S36), sometimes even among closely related species. The over-emphasis of such characters in generic circumscription had led to non-monophyletic genera and thus chaotic generic limits of Asian Sonerileae (also see Zhou et al., 2019c). Of the characters studied, growth form, floral merosity, stamen arrangement, fruit type, crystal form, and the presence of hyaline glands are the least homoplasious. For example, the shift from 4-merous to 5-merous flowers (Fig. S34) and druses to raphides (Fig. S44) only occurred three times each across Asian Sonerileae, which makes them useful diagnostic characters for some major lineages. Future taxonomic reshuffling at the generic level should be based on robust phylogenetic relationships and character combinations across the major lineages.

4.4. Taxonomic implications

Despite widespread phylogenetic discordance, the present analyses provided the first well resolved phylogeny of Asian Sonerileae sampling 74% of the Asian genera, which serves as a framework for future taxonomic revision at the generic level. Thirty-four well supported and morphologically cohesive lineages in Asian Sonerileae (Fig. 1) were identified and relationships among most lineages were well resolved. Among these lineages, *Aschistanthera* and *Barthea* are monotypic. *Blastus*, *Bredia*, *Heteroblemma*, *Sarcopyramis*, *Tashiroea*, *Tigridiopalma*, and *Sonerila* represent monophyletic genera. *Medinilla*, nonmonophyletic according to a previous study (Kartonegoro et al., 2021), was recovered as monophyletic in the present analyses due to limited sampling without many key taxa (e.g., *Boerlagea* Cogn., *Carionia* Naudin, *Pachycentria* Blume, and *Plethiandra* Hook.f.). The remaining 24 lineages needing to be revised are discussed below, followed by a summary of the highly polyphyletic *Phyllagathis* and *Oxyspora* (=*Allomorphia* Blume). A comparison of the 34 major lineages of Asian Sonerileae is provided in Supplementary Table S9.

4.4.1. Phyllagathis (thyrsoid) clade

This is a Bornean clade comprising *Phyllagathis longifolia* (Cogn.) J.F. Maxwell, *P. gymnantha* Korth., *P. gigantifolia* M.P.Nayar, and a new species. The species usually have a spike-like thyrse (rarely reduce to pleiochasium) and narrowly campanulate to cylindric hypanthium. Judging from morphology and distribution, *P. atroviolacea* C.Hansen ex Cellin., *P. longispicata* (Cogn.) J.F.Maxwell, and *P. pulcherrima* M.P. Nayar are also possible members of *Phyllagathis* (thyrsoid) clade. This clade was recovered as sister to *Heteroblemma*, a peculiar genus endemic to Vietnam and Malesia (Cámara-Leret et al., 2013), which is somewhat surprising, as the two lineages differ strikingly in habit (erect herbs vs. woody root climber), inflorescence morphology (thyrse vs. cymose bundles), and fruit type (capsule vs. berry). *Phyllagathis* (thyrsoid) clade does not include the generic type and should be established as a genus of its own.

4.4.2. Phyllagathis (pentamerous) clade

This clade contains Bornean species of *Phyllagathis* with pentamerous flowers, viz. *P. osmantha* (M.P.Nayar) Cellin. (=*Cyanandrium osmanthus* M.P.Nayar), *P. rajah* C.W.Lin, Chien F.Chen & T.Y.A.Yang, *P. rufa* (=*C. rufum* Stapf), *P. stellata* C.W.Lin & C.H.Lee and a new species. They are subacaulescent herbs with pentamerous flowers, campanulate hypanthium, crowned capsule, and non-thready placentas, inflorescence a congested pleiochasium or umbel. Based on close resemblance in morphology and distribution, other pentamerous species of *Phyllagathis*, viz. *P. guttata* (Stapf) Cellin. (=*Cyanandrium guttatum* Stapf), *P. jacobsiana* (M.P.Nayar) Cellin. (=*C. jacobsianum* M.P.Nayar), *P. subacaulis* (Cogn.) Cellin. (=*Brittenia subacaulis* Cogn.), *P. steenisii* Cellin. (=*Enaulophyton lanceolatum* Steenis), and *P. penrissenensis* Cellin. (=*E. sarawakense* M.P.Nayar), are also potential members of *Phyllagathis* (pentamerous) clade. This clade was sister to *Heteroblemma* + *Phyllagathis* (thyrsoid) clade + *Driessenia* (isomorphic) clade, but morphologically distinct from the latter three (Table S9). As the type of *Phyllagathis* is not included, *Phyllagathis* (pentamerous) clade should be segregated from the former and treated as a distinct genus.

4.4.3. Driessenia axantha and Driessenia (isomorphic) clade

Driessenia consists of ca. 15 species and 2 varieties endemic to Borneo, Sumatra, Java, and Sulawesi (Hansen, 1985a; Lin, 2019). This genus is recognized by its three staminal appendages, with the two ventral ones at least about half as long as the anthers and longer than the dorsal one. Hansen (1985a) pointed out that *D. axantha* (the generic type) is morphologically isolated from the rest of *Driessenia*. Hansen's suspicion is confirmed by the present analyses. Eight accessions of *Driessenia* formed two distantly related lineages. The generic type *D. axantha* consistently constituted a branch (clade B) with *Tashiroea*, *Scorpiothyrsus* clade, and *Poilannammia* clade, while the remaining accessions formed a strongly supported clade [*Driessenia* (isomorphic) clade] sister to *Heteroblemma* + *Phyllagathis* (thyrsoid) clade. *Driessenia* (isomorphic) clade differs from *D. axantha* in the inflorescence a dichasium or thyrse (vs. scorpioid inflorescence) and isomorphic stamens (vs. dimorphic). Both morphological and molecular data indicate that *Driessenia* (isomorphic) clade should be excluded from *Driessenia* and treated as a distinct genus.

4.4.4. Scorpiothyrsus clade

The *Scorpiothyrsus* clade comprises *Scorpiothyrsus* (Yunnan, Guangxi, Hainan, China), *Vietsenia* C.Hansen (Vietnam), *Phyllagathis cymigera* C. Chen (Yunnan) and *P. hainanensis* (Merr. & Chun) C.Chen (Hainan). They are shrublets or herbs characterized by terminal scorpioid cymes, isomorphic stamens with 2-setose ventral appendages, and cup-shaped capsules. The placental column and placentas are often disintegrating upon dehiscence. The *Scorpiothyrsus* clade constituted clade B together with *Tashiroea*, *Poilannammia* clade, and the type of *Driessenia* (*D. axantha*), but is morphologically quite distinct from the latter groups (Table S9). This clade may be better treated as one genus, with *Scorpiothyrsus* adopted as the generic name based on its priority over *Vietsenia*.

4.4.5. Poilannammia clade

This is a Vietnamese clade comprising two subclades, one containing *Poilannammia* and the other containing *Phyllagathis megalocentra* C. Hansen and three new species. Despite their diverse habit (herb to small trees), these species are characterized by somewhat coriaceous, glabrous leaf blade (when mature), terminal inflorescence, dorsally spurred and ventrally 2-lobed anther base, crowned capsule, and 4-horned placental column. The *Poilannammia* clade was sister to *Tashiroea* but differs in the crowned capsule (vs. ovary crown often evanescent), 4-horned placental column (vs. not horned) and distribution (Vietnam vs. southeastern mainland China, Taiwan, and Ryukyu Islands). Two unsampled species, *Phyllagathis driessenioides* C.Hansen and *P. truncata* C.Hansen are also possible members of the *Poilannammia* clade, based on morphology and distribution. *Poilannammia* could be expanded to accommodate additional species.

4.4.6. Phyllagathis prostrata clade

This clade was sister to *Tigridiopalma* and consists of *Phyllagathis prostrata* and a new species. The *P. prostrata* clade and *Tigridiopalma* share obvious resemblance in habit, pentamerous flowers, purple anther sac, and yellow connective appendages, but differ markedly in morphology of inflorescence (umbel with large, persistent bracts vs. scorpioid cymes with small, caducous bracts), stamen (isomorphic vs. dimorphic), placenta column (horned vs. not horned), and placenta (thready vs. unthready). Geographically, their ranges are adjacent but not overlapping (Guangdong, China for *Tigridiopalma*; Hainan, China and Vietnam for *P. prostrata* clade). As the sister relationship of the two clades collapsed in the plastome tree as well as the SCO ASTRAL tree resulted from LBA test, they are better treated as separate genera. *Phyllagathis brevipedunculata* C.Hansen unsampled in this study is a potential member of the *P. prostrata* clade.

4.4.7. Anerincleistus clade and its allies (=clade E)

The *Anerincleistus* clade (Borneo) and its allies, i.e. *Phyllagathis tuberculata-P. hispida* (Malay Peninsula), *P. griffithii-Oxyspora bullata* (=*Allomorphia malaccensis* Ridl.) (Malay Peninsula), and *P. scortechinii* (Malay Peninsula) consistently constituted a well-resolved lineage (clade E in Fig. 1). The four clades are morphologically distinct. *Phyllagathis tuberculata-P. hispida* are herbs or erect, single-stemmed shrubs with a compound or simple umbel, dimorphic stamens, and crowned capsule. *Phyllagathis griffithii-O. bullata* (=*A. malaccensis*) are characterized by a thyrse, tubular or slightly urceolate, narrow hypanthium, and isomorphic stamens. *Phyllagathis scortechinii* is a caulescent herb with a dense umbel displaced onto a phyllophore, isomorphic stamens, and crowned capsule. The *Anerincleistus* clade comprises Bornean species of *Anerincleistus* and *Phyllagathis*, often recognized by panduriform-obovate, basal or plinerved leaf blades, campanulate hypanthium, and isomorphic stamens with minute connective appendages. Based on their phylogenetic relationships, there are two options of generic treatment: lumping the above four lineages in one genus or treating them as separate genera. As these lineages are morphologically diverse, lumping them will result in a genus difficult (if not impossible) to define. Treating them separately seems more feasible. If the second option is adopted, *Anerincleistus* will include additional species currently placed in *Phyllagathis*, viz. the species published in Lin et al. (2017).

4.4.8. Phyllagathis rotundifolia

The type of *Phyllagathis* was weakly supported as sister to clade E (*Anerincleistus* clade + alliances) in the SCO specie trees, or to *Oxyspora paniculata* (the type of *Oxyspora*) in the genomic tree. This species occurs in peninsular Thailand, peninsular Malaysia, and Sumatra. It has a unique contracted head-like thyrse consisting of scorpioid short branches and large subtending bracts, showing no close relationship to other species of *Phyllagathis*. *Phyllagathis* may have to be redefined as monotypic.

4.4.9. Oxyspora paniculata

The type of *Oxyspora* was weakly supported as sister to *Phyllagathis rotundifolia* (genomic dataset) or to the lineage containing *Styrophyton* clade, *Aschistanthera*, and *Oxyspora* (Bornean) clade (SCO dataset). This species is endemic to Indo-Burma and has large, spreading inflorescence, large flowers, dimorphic stamens, and large capsules. Based on these characters, the Chinese species *O. cernua* Hook.f. & Thomson ex Triana, *O. vagans* Hook., and *O. yunnanensis* H.L.Li are also members of this clade.

4.4.10. Oxyspora (Bornean) clade

The *Oxyspora* (Bornean) clade contains two accessions of *Oxyspora* endemic to Borneo. It was sister to *Styrophyton* clade + *Aschistanthera*, forming clade F. The two accessions closely resemble the *Styrophyton* clade but differ from the latter in the dimorphic stamens (vs. isomorphic) and absence of hyaline glands (vs. presence). As the generic type is not included, *Oxyspora* (Bornean) clade will have to be segregated from *Oxyspora*.

4.4.11. Styrophyton clade

The *Styrophyton* clade was sister to *Aschistanthera*. It consists of *Styrophyton*, *Oxyspora teretipetiolata* (C.Y.Wu & C.Chen) W.H.Chen & Y. M.Shui (=*Allomorphia eupteroton* var. *teretipetiolata* C.Y.Wu & C.Chen), and species of *Allomorphia* Blume from southwestern China and Vietnam. *Styrophyton caudatum* (Diels) S.Y.Hu was originally published in *Anerincleistus* (Diels 1932), later transferred to *Allomorphia* (Li 1944), and eventually established as a monotypic genus (Hu 1952). Upon the publication of *Styrophyton*, Hu (1952) emphasized the short calyx tube and spicate inflorescence of this species as distinct from *Allomorphia*. However, a short calyx tube is also found in other species of the *Styrophyton* clade, such as *Allomorphia sulcata* C.Hansen, and the shift from narrowly thyrsoid to spicate inflorescence only involves shortening of the pedicels and lateral peduncles. The *Styrophyton* clade closely resembles *Blastus* in the shrubby habit, the presence of hyaline glands, often uncrowned capsule, unhorned placental column and non-thready placentas, but differs in stamen number (8, isomorphic vs. 4). As the type of *Allomorphia* is not included in this clade, *Allomorphia* from Vietnam and China are better accommodated in *Styrophyton*.

4.4.12. Phyllagathis (raphides) clade

The *Phyllagathis* (raphides) clade consists of four new species from Vietnam and some species currently treated in *Phyllagathis*, viz. *P. calisaurea* C.Chen, *P. melastomatoides* (Merr. & Chun) W.C.Ko, *P. ovalifolia*, *P. setotheca*, *P. stenophylla*, and *P. xinyiensis* Z.J.Feng from China, *P. sessilifolia* C.Hansen and *P. suberalata* C.Hansen from Vietnam, and *P. dispar* (Cogn.) C.Hansen and *P. elliptica* Stapf from Borneo. Phylogenetic analyses placed it sister to an Indo-Burmese branch containing *Fordiophyton s.s.* clade and clades G, H, and I (Fig. 1). Species of the *Phyllagathis* (raphides) clade are shrubs to caulescent herbs with cuneate to rounded leaf bases, isomorphic stamens, dorsally spurred connectives, crowned capsules, horned placental column, thready placentas, and presence of raphide crystals in some of the species (8 out of 14). Among species of *Phyllagathis* unsampled in the present study, *P. elegans* Hai L.Chen, Yan Liu & Ying Liu and *P. deltoda* C.Chen from China, *P. guillauminii* H.L.Li and *P. marumiaetricha* (Guillaumin) C. Hansen from Vietnam, *P. fruticosa* (Ridl.) C.Hansen ex Cellin. from Malay Peninsula, and *P. brookei* M.P.Nayar and *P. rupicola* (M.P.Nayar) C.Hansen ex Cellin. from Borneo can be listed as plausible members of this clade based on strong morphological resemblance in crystal type, habit, leaf shape, inflorescence, stamen, and capsule morphology. Therefore, this clade should at least contain the aforementioned 20 species from southernmost China, Vietnam, Malay Peninsula to Borneo. The *Phyllagathis* (raphides) clade should be treated as a genus of its own as the type of *Phyllagathis* is not included. *Perilimnastes* Rild. is a name published based on one potential member of this clade, i.e. *P. fruticosa* (Ridley, 1918). This name is derived from Greek and means near water,

which fits well with the habitat preference of many species in this clade. *Perilimnastes* can be resurrected as the generic name.

4.4.13. *Plagiopetalum clade, Cyphotheca clade, Sporoxeia clade, and Phyllagathis cavaleriei (=clade G)*

The *Plagiopetalum* clade, *Cyphotheca* clade, *Sporoxeia* clade, and *Phyllagathis cavaleriei* constantly formed a well-supported clade (clade G in Fig. 1), with *Plagiopetalum* clade the basal branch and *Cyphotheca* clade sister to *Sporoxeia* clade + *Phyllagathis cavaleriei*. The *Plagiopetalum* clade includes *Plagiopetalum*, *P. wenshanensis* S.Y.Hu, and *Phyllagathis cavaleriei* var. *wilsoniana* Guillaum. (=*P. longipes* H.L.Li), all occurring in southwestern China. Diagnostic characters of the *Plagiopetalum* clade include small calyx lobes dorsally ridged, 8 isomorphic stamens, yellow to whitish anther, dorsally minute spurred connectives, placental column not horned and placenta non-thready. Two subgroups are resolved, one contains erect shrubs or shrublets (*Plagiopetalum*), and the other prostrate herbs (two species of *Phyllagathis*). If *Plagiopetalum* clade is treated as one genus, *Phyllagathis wenshanensis* and *Phyllagathis cavaleriei* var. *wilsoniana* should be transferred to *Plagiopetalum*. The *Cyphotheca* clade contains *Cyphotheca*, *Phyllagathis fengii* C.Hansen, and *Phyllagathis tentaculifera* C.Hansen. The similarity between *Cyphotheca* and species of *Phyllagathis* in stamen and capsule morphology have long been noted by previous authors (Li, 1944; Hansen, 1990). These species are characterized by terminal cymose umbellate inflorescences, thickened connectives usually not spurred dorsally, crowned capsule, horned placental column, and thready placentas. *Cyphotheca* should also include *Phyllagathis fengii* and *Phyllagathis tentaculifera*. The *Sporoxeia* clade encompasses *Sporoxeia*, *Phyllagathis longicalcarata* C.Hansen and *Phyllagathis hispidissima* (C.Chen) C.Chen from Yunnan, one new species (732) from Guangxi, China, and possibly also an unsampled species *Phyllagathis scorpiothyrsoides* C.Chen. Morphological synapomorphies may include low rim inside the sepal, connectives with conspicuous dorsal spurs, and horned placental column. In this clade, some species have axillary inflorescence and hyaline glands, while the remaining have terminal inflorescence and without hyaline glands. *Sporoxeia* should be expanded to include additional species with terminal inflorescence. *Phyllagathis cavaleriei* is an herb widely distributed in southern China. Its sister relationship to the *Sporoxeia* clade was strongly supported. However, *Phyllagathis cavaleriei* shows no obvious resemblance to the *Sporoxeia* clade, expect for the crowned capsule and horned placental column which are highly homoplasious in Asian Sonerileae. Therefore, transferring *Phyllagathis cavaleriei* to *Sporoxeia* seems inappropriate. Two new species, viz. *Sporoxeia* sp. nov. 731 and *Phyllagathis* sp. nov. 727 are morphologically distinct from the above four lineages. The phylogenetic positions of both species show strong cytonuclear discordance, which will be dealt with in a separate study.

4.4.14. *Fordiophyton s.s. clade*

The *Fordiophyton s.s.* clade was sister to *Bredia* + *Blastus* + clade H. It contains most species of *Fordiophyton*, including the type. These species are characterized by the presence of raphides, eight fertile, distinctly dimorphic stamens, and inflated but not decurrent connectives and is easily distinguished from closely related lineages, i.e., *Bredia*, *Blastus*, *Kerriothyrsus* clade, *Fordiophyton breviscapum*, and *Phyllagathis erecta*.

4.4.15. *Kerriothyrsus clade, Fordiophyton breviscapum, and Phyllagathis erecta (=clade H)*

The *Kerriothyrsus* clade, together with *Phyllagathis erecta*, and *Fordiophyton breviscapum*, formed a lineage (clade H in Fig. 1) sister to clade I (*Bredia* + *Blastus*), with *P. erecta* sister to the *Kerriothyrsus* clade and *F. breviscapum* sister to *P. erecta* + *Kerriothyrsus* clade. The *Kerriothyrsus* clade comprises *Kerriothyrsus* (Vietnam and Laos), *Phyllagathis elattandra* Diels and *P. tetrandra* (China), *P. subrotunda* C.Hansen (Vitenam), and *P. tuberosa* (C.Hansen) Cellin. & S.S.Renner (=*Tylanthera tuberosa* C. Hansen) (Thailand). These species are acaulescent to subacaulescent herbs with 4 fertile stamens (another 4 sterile or undeveloped), often crowned capsule, unhorned placental column and non-thready placenta. Two subclades are recognized, one contains *Kerriothyrsus* and *P. tuberosa* (=*T. tuberosa*), and the other contains three species of *Phyllagathis*. If they are treated as one genus, *Kerriothyrsus* has priority over *Tylanthera* and should be adopted as the generic name. Based on strong morphological evidence, *P. indica* J.Mathew, Yohannan & Kad.V.George (India), *P. siamensis* Cellin. & S.S.Renner, and *P. nanakorniana* Wangwasit, Norsaengsri & Cellin. (Thailand) also belong in this clade. *Fordiophyton breviscapum* (Guangdong, China) most closely resembles the unsampled *F. degeneratum* (C.Chen) Y.F.Deng & T.L.Wu (Guangxi, China). The two species are characterized by the caulescent habit, winged stem and hypanthium, and sterile inner whorl of stamens, differing from the *Fordiophyton s.s.* clade in the number of fertile stamens and the absence of raphides. They should be excluded from *Fordiophyton* and treated as a separate genus. *Phyllagathis erecta* (Guangxi and Yunnan, China) is a shrub with terminal umbellate cymose panicles, eight isomorphic, fertile stamens, and horned placenta column. It is morphologically isolated from both *F. breviscapum* and the *Kerriothyrsus* clade. A recently published species *Phyllagathis chongzuoensis* Zhe Zhang, S.W.Yao & S.Jin Zeng is no doubt a close relative of *P. erecta* based on their strong resemblance in morphology of inflorescence and capsule. Discovery of more close relatives is expected when remote areas of southwestern China become better botanized.

4.4.16. *Taxonomic fate of Phyllagthis*

Phyllagathis as currently defined comprises >70 species distributed from southern China, Indo-Burma, Malay Peninsula, to Sumatra and Borneo. The highly polyphyletic nature of *Phyllagathis* is confirmed in the present study, with members of the genus nested in 17 lineages scattered across Asian Sonerileae. As the generic type *P. rotundifolia* is quite isolated (both morphologically and phylogenetically) from other members, *Phyllagathis* may have to be recircumscribed as monotypic. Members nested in seven of the 17 clades may have to be transferred to the following genera accordingly, i.e., *Anerincleistus*, *Cyphotheca*, *Kerriothyrsus*, *Plagiopetalum*, *Poilannammia*, *Scorpiothyrsus*, and *Sporoxeia*. Four clades, *Phyllagathis* (pentamerous), *Phyllagathis* (raphides), *Phyllagathis* (thyrsoid), and *Phyllagathis prostrata* clade should be treated as distinct genera. The remaining species, viz. *P. cavaleriei*, *P. erecta*, *P. griffithii*, *P. scortechinii*, and *P. tuberculata-P. hispida*, are morphologically isolated from their close relatives and difficult to treat. In addition, although many unsampled species can be attributed to the above clades based on strong morphological evidence, some species with unique morphology still need to be phylogenetically analyzed, such as *P. cordata* Ridl., *P. maxwellii* B.C.Stone & A.Weber, and *P. stolonifera* Kiew.

4.4.17. *Oxyspora and Allomorphia*

Oxyspora and *Allomorphia* contain about 37 species from southwestern China, India, Indo-Burma, Malay Peninsula, Sumatra, and Borneo. *Oxyspora* was traditionally envisioned to have large, spreading inflorescences, large flowers, dimorphic anthers, and large ellipsoid capsules, while *Allomorphia* was considered as having smaller inflorescences and flowers, equal or subequal anthers, and smaller urceolate capsules (Maxwell, 1982). However, their generic limit remains problematic, and *Allomorphia* is often treated as synonym of *Oxyspora* (Maxwell, 1982). The present analyses recovered four lineages in *Oxyspora* and *Allomorphia*. Indo-Burmese species, viz. the generic type *Oxyspora paniculata*, possibly also *O. cernua*, *O. vagans*, and *O. yunnanensis* should be treated in the redefined *Oxyspora*. Chinese and Vietnamese species of *Allomorphia* nested in *Styrophyton* clade will have to be transferred to the latter. *Oxyspora* (Bornean) clade and *Oxyspora bullata* (Malay Peninsula) should be excluded from *Oxyspora*, but further analyses sampling more species from Malesia (including the type of *Allomorphia*) are needed to reach a taxonomic decision for these species.

4.4.18. *Unsampled genera*

Eight Asian genera were unsampled in the present study, viz.

Boerlagea, *Catanthera* F.Muell., *Kendrickia* Hook.f., *Neodriessenia* M.P. Nayar, *Pachycentria*, *Plethiandra*, *Poikilogyne* Baker f., and *Stussenia* C. Hansen. *Catanthera* is endemic to Malesia and *Kendrickia* to Sri Lanka. According to a previous phylogenetic study based on plastid *ndhF* (Clausing and Renner, 2001b), *Catanthera* was sister to *Kendrickia* and the two genera constituted the sister group to *Heteroblemma*. Both *Pachycentria* and *Plethiandra* have been included in recent studies (Kartonegoro et al., 2021; Maurin et al., 2021). Maurin et al. (2021) resolved *Plethiandra* as sister to *Medinilla*, whereas Kartonegoro et al. (2021) recovered *Pachycentria* as nested within *Medinilla*. However, the relationships among *Pachycentria*, *Plethiandra*, and *Medinilla* remains to be further tested because of sparse sampling of taxa and lack of support. *Boerlagea*, *Neodriessenia*, *Poikilogyne*, and *Stussenia* have never been phylogenetically tested. *Boerlagea* is a monotypic shrubby genus with axillary trimerous flowers. It was traditionally allied with *Medinilla* based on its fruit an urceolate berry (Maxwell, 1984). *Neodriessenia* shares some general resemblance with *Blastus* and *Driessenia* but differs from the former in stamen number (eight vs. four) and from the latter in the presence of hyaline glands and ligulate, apically two-lobed ventral connective appendages (Hansen, 1985b). *Stussenia* contains only one species. It was originally described in *Blastus* (Li, 1945), later transferred to *Neodriessenia* (Hansen, 1980), and subsequently treated as a distinct genus based on its two-celled hyaline glands, trimerous flowers, and dimorphic stamens with two ventral connectives lobes (Hansen, 1985c). *Poikilogyne* is an enigmatic genus of 28 species endemic to New Guinea, the Bismarck Archipelago, and the Solomon Islands. They are shrubs, treelets or vines with pentamerous flowers, and their capsules bear prominent vascular strands, showing no close resemblance to other Asian genera of Sonerileae.

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 Dr. Qiang Fan for providing the materials of *Dissochaeta gracilis* and *D. vacillans*. We are deeply grateful to Dr. Huang-hui Kong for his advice during data analysis.

Funding

This work was supported by the National Natural Science Foundation of China (32170220, 31770214) and Natural Science Foundation of Guangdong Province (grant 2021A1515011214). Sample collection in northern Vietnam was partly supported by the Ministry of Planning and Investment, Vietnam, and the Vietnam Academy of Science and Technology under the project code UQĐTCB.06/22–23 to TVD.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2022.107581.

References

Allman, E.S., Mitchell, J.D., Rhodes, J.A., 2021. Gene tree discord, simplex plots, and statistical tests under the coalescent. Syst. Biol. syab008 https://doi.org/10.1093/sysbio/syab008.

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Bacci, L.F., Michelangeli, F.A., Goldenberg, R., 2019. Revisiting the classification of Melastomataceae: implications for habit and fruit evolution. Bot. J. Linn. Soc. 190, 1–24. https://doi.org/10.1093/botlinnean/boz006.

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., 2021. A comprehensive phylogenomic platform for exploring the angiosperm tree of life. Syst. Biol. syab035 https://doi.org/10.1093/sysbio/syab035.

Bergsten, J., 2005. A review of long-branch attraction. Cladistics 21, 163–193. https://doi.org/10.1111/j.1096-0031.2005.00059.x.

Bolger, A.M., Marc, L., Bjoern, U., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Buckley, T.R., Cordeiro, M., Marshall, D.C., Simon, C., 2006. Differentiating between hypotheses of lineage sorting and introgression in New Zealand alpine cicadas (*Maoricicada* Dugdale). Syst. Biol. 55, 411–425. https://doi.org/10.1080/10635150600697283.

Cámara-Leret, R., Ridder-Numan, J.W.A., Veldkamp, J.F., 2013. Revision of *Heteroblemma* gen. nov. (Dissochaeteae–Melastomataceae) from Malesia and Vietnam. Blumea 58, 229–240. https://doi.org/10.3767/000651913X674945.

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.

Cellinese, N., 1997. Notes on the systematics and biogeography of the *Sonerila* generic alliance (Melastomataceae) with special focus on fruit characters. Trop. Biodivers. 4, 83–93.

Cellinese, N., 2002. Revision of the genus *Phyllagathis* Blume (Melastomataceae: Sonerileae) I. The species of Burma, Thailand, Peninsular Malaysia and Sumatra. Blumea 47, 463–492.

Cellinese, N., 2003. Revision of the genus *Phyllagathis* (Melastomataceae: Sonerileae) II. The species in Borneo and Natuna Island. Blumea 48, 69–97.

Chen, C., 1984a. Melastomataceae. In: Chen, C. (Ed.), Flora Reipublicae Popularis Sinicae, vol. 53. Science Press, Beijing, pp. 135–293.

Chen, C., 1984b. Materia ad flora Melastomataceae sinensium. Bull. Bot. Res. 4, 33–68.

Chen, C., Renner, S.S., 2007. Melastomataceae. In: Wu, Z.Y., Raven, P.H., Hong, D.Y. (Eds.), Flora of China, vol. 13. Science Press, Beijing; Missouri Botanical Garden Press, St. Louis, pp. 360–399.

Chifman, J., Kubatko, L., 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics 30, 3317–3324. https://doi.org/10.1093/bioinformatics/btu530.

Clausing, G., Renner, S.S., 2001a. Molecular phylogenetics of Melastomataceae and Memecylaceae: implications for character evolution. Amer. J. Bot. 88, 486–498. https://doi.org/10.2307/2657114.

Clausing, G., Renner, S.S., 2001b. Evolution of growth form in epiphytic Dissochaeteae (Melastomataceae). Org. Divers. Evol. 1, 45–60.

Clausing, G., 1999. Die Systematik der Dissochaeteae und ihre Stellung innerhalb der Melastomataceae. Ph.D. Dissertation. University of Mainz, Mainz.

Diels, L., 1932. Beiträge zur Kenntnis der Melastomataceen Ostasiens. Bot. Jahrb. Syst. 65, 97–119.

Dierckxsens, N., Mardulyn, P., Smits, G., 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucl. Acids Res. 45, e18 https://doi.org/10.1093/nar/gkw955.

Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemistry 19, 11–15.

Driskell, A.C., Ané, C., Burleigh, J.G., McMahon, MCM., O'meara, B.C., Sanderson, M.J., 2004. Prospects for building the tree of life from large sequence databases. Science 306, 1172–1174. https://doi.org/10.1126/science.1102036.

Duchêne, D.A., Bragg, J.G., Duchêne, S., Neaves, L.E., Moritz, C., Johnson, R.N., Ho, S.Y.W., Eldridge, M.D.B., 2018. Analysis of phylogenomic tree space resolves relationships among marsupial families. Syst. Biol. 67, 400–412. https://doi.org/10.1093/sysbio/syx076.

Edwards, S.V., 2008. Is a new and general theory of molecular systematics emerging? Evolution 63, 1–19. https://doi.org/10.1111/j.1558-5646.2008.00549.x.

Edwards, S.V., Xi, Z., Janke, A., Faircloth, B.C., McCormack, J.E., Glenn, T.C., Zhong, B., Wu, S., Lemmon, E.M., Lemmon, A.R., Leaché, A.D., Liu, L., Davis, C.C., 2016. Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. Mol. Phylogenet. Evol. 94, 447–462. https://doi.org/10.1016/j.ympev.2015.10.027.

Emms, D.M., Kelly, S., 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16, 157. https://doi.org/10.1186/s13059-015-0721-2.

Galtier, N., Daubin, V., 2008. Dealing with incongruence in phylogenomic analyses. Philos. Trans. Ser. B 363, 4023–4029. https://doi.org/10.1098/rstb.2008.0144.

Gaut, B.S., 1998. Molecular clocks and nucleotide substitution rates in higher plants. Evol. Biol. 30, 93–120.

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.

Goldenberg, R., Almeda, F., Sosa, K., Ribeiro, R.C., Michelangeli, F.A., 2015. *Rupestrea*: a new Brazilian genus of Melastomataceae, with anomalous seeds and dry indehiscent fruits. Syst. Biol. 40, 561–571. https://doi.org/10.1600/036364415X688862.

Grabherr, M., Haas, B., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652. https://doi.org/10.1038/nbt.1883.

Hahn, C., Bachmann, L., Chevreux, B., 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucl. Acids Res. 41, e129 https://doi.org/10.1093/nar/gkt371.
Hansen, C., 1980. *Neodriessenia membranifolia* (H.L.Li) C. Hansen, Comb. Nov. (Melastomataceae). Adansonia 20, 321–324.
Hansen, C., 1982. New species and combinations in *Anerincleistus*, *Driessenia*, *Oxyspora* and *Phyllagathis* (Melastomataceae) in Sarawak. Nord. J. Bot. 2, 557–559. https://doi.org/10.1111/j.1756-1051.1983.tb01049.x.
Hansen, C., 1984. *Vietsenia* C. Hansen, a new genus of the Melastomataceae for Vietnam. Bull. Mus. Natl. Hist. Nat. B Adansonia 6, 147–156.
Hansen, C., 1985a. Taxonomic revision of *Driessenia* (Melastomataceae). Nordic J. Bot. 5, 335–352.
Hansen, C., 1985b. Revision of the genus *Neodriessenia* Nayar (Melastomataceae). Bot. Jahrb. Syst. 106, 1–13.
Hansen, C., 1985c. *Stussenia*, a new genus in the Melastomataceae. Willdenowia 15, 175–176.
Hansen, C., 1987a. *Aschistanthera*, a monotypic new genus for Vietnam (Melastomataceae). Nordic J. Bot. 7, 653–654.
Hansen, C., 1987b. *Poilannammia* C. Hansen (Melastomataceae) a new genus of four new species endemic to Vietnam. Bull. Mus. Natl. Hist. Nat. B Adansonia 9, 263–271.
Hansen, C., 1988. *Kerriothyrsus*, a new genus of Melastomataceae. Willdenowia 17, 153–157.
Hansen, C., 1990. The monotypic genus *Cyphotheca* (Melastomataceae). Nordic J. Bot. 10, 21–23.
Hansen, C., 1992. The genus *Phyllagathis* (Melastomataceae): characteristics; delimitation; the species in Indo-China and China. Bull. Mus. Natl. Hist. Nat. B Adansonia 14, 355–428.
Holder, M.T., Anderson, J.A., Holloway, A.K., 2001. Difficulties in detecting hybridization. Syst. Biol. 50, 978–982.
Holland, B.R., Benthin, S., Lockhart, P.J., Moulton, V., Huber, K.T., 2008. Using supernetworks to distinguish hybridization from lineage-sorting. BMC Evol. Biol. 8, 1–11. https://doi.org/10.1186/1471-2148-8-202.
Hovmöller, R., Knowles, L.L., Kubatko, L.S., 2013. Effects of missing data on species tree estimation under the coalescent. Mol. Phylogenet. Evol. 69, 1057–1062. https://doi.org/10.1016/j.ympev.2013.06.004.
Hu, S.Y., 1952. Notes on the Flora of China. II. J. Arnold Arbor. 33, 166–178.
Huang, H., He, Q., Kubatko, L.S., Knowles, L.L., 2010. Sources of error inherent in species-tree estimation: impact of mutational and coalescent effects on accuracy and implications for choosing among different methods. Syst. Biol. 59, 573–583. https://doi.org/10.1093/sysbio/syq047.
Huang, C.H., Sun, R., Hu, Y., Zeng, L., Zhang, N., Cai, L., Zhang, Q., Koch, M.A., Al-Shehbaz, I., Edger, P.P., Pires, J.C., Tan, D.Y., Zhong, Y., Ma, H., 2016. Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. Mol. Biol. Evol. 33, 394–412. https://doi.org/10.1093/molbev/msv226.
Joly, S., McLenachan, P.A., Lockhart, P.J., 2009. A statistical approach for distinguishing hybridization and incomplete lineage sorting. Am. Nat. 174, E54–E70. https://doi.org/10.1086/600082.
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.
Kartonegoro, A., Veldkamp, J.F., 2010. Revision of *Dissochaeta* (Melastomataceae) in Java, Indonesia. Reinwardtia 13, 125–145. https://doi.org/10.14203/reinwardtia.v13i2.2133.
Kartonegoro, A., Veranso-Libalah, M.C., Kadereit, G., Frenger, A., Penneys, D.S., Mota de Oliveira, S., Van Welzen, P.C., 2021. Molecular phylogenetics of the *Dissochaeta* alliance (Melastomataceae): Redefining tribe Dissochaeteae. Taxon 70, 793–825. https://doi.org/10.1002/tax.12508.
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.
Knowles, L.L., Huang, H., Sukumaran, J., Smith, S.A., 2018. A matter of phylogenetic scale: Distinguishing incomplete lineage sorting from lateral gene transfer as the cause of gene tree discord in recent versus deep diversification histories. Amer. J. Bot. 105, 376–384. https://doi.org/10.1002/ajb2.1064.
Kong, H.H., Condamine, F., Yang, L.H., Harris, A.J., Wen, F., Kang, M., 2021. Phylogenomic and macroevolutionary evidence for an explosive radiation of a plant genus in the Miocene. Syst. Biol. syab068 https://doi.org/10.1093/sysbio/syab068.
Kubatko, L.S., Degnan, J.H., 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56, 17–24. https://doi.org/10.1080/10635150601146041.
Lanier, H.C., Knowles, L.L., 2015. Applying species-tree analyses to deep phylogenetic histories: Challenges and potential suggested from a survey of empirical phylogenetic studies. Mol. Phylogenet. Evol. 83, 191–199. https://doi.org/10.1016/j.ympev.2014.10.022.
Leaché, A.D., Rannala, B., 2010. The accuracy of species tree estimation under simulation: a comparison of methods. Syst. Biol. 60, 126–137. https://doi.org/10.1093/sysbio/syq073.
Li, H.L., 1944. Studies in the Melastomataceae of China. J. Arnold Arbor. 25, 1–42.
Li, H.L., 1945. Further notes on the flora of Indo-china. J. Arnold Arbor. 26, 119–121.
Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589–595. https://doi.org/10.1093/bioinformatics/btp698.
Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658–1659. https://doi.org/10.1093/bioinformatics/btl158.
Lin, C.W., Lee, C.H., 2018. *Phyllagathis stellata* (Sonerileae, Melastomataceae), a new species from southwestern Sarawak, Borneo. Phytotaxa 365, 295–300. https://doi.org/10.11646/phytotaxa.365.3.7.
Lin, C.W., Chen, C.F., Yang, T.Y.A., 2015. Two new taxa of Melastomataceae Trib. Sonerileae: *Phyllagathis rajah* and *Sonerila metallica* from Batang Ai, Sarawak. Borneo. Phytotaxa 201, 122–130. https://doi.org/10.11646/phytotaxa.201.2.2.
Lin, C.W., Chen, C.F., Yang, T.Y.A., 2017. Ten new species of *Phyllagathis* (Trib. Sonerileae, Melastomataceae) from Sarawak. Borneo. Phytotaxa 302, 201–228. https://doi.org/10.11646/phytotaxa.302.3.1.
Lin, C.W., 2019. Driessenia phasmolacuna (Sonerileae, Melastomataceae), a new species from Batang Ai, Sarawak, Borneo. Taiwania 64, 69–73. https://doi.org/10.6165/tai.2019.64.69.
Liu, Y., Veranso-Libalah, M.C., Kadereit, G., Zhou, R.C., Quakenbush, J.P., Lin, C.W., Wai, J.S., in press. Systematics of the Tribe Sonerileae. In: Goldenberg, R., Michelangeli, F.A., Almeda, F. (Eds.), Advances in Melastomataceae Systematics and Biology. Springer Nature, Cham, Switzerland. https://doi.org/10.1007/978-3-030-99742-7_15.
Maddison, W.P., Maddison, D.R., 2018. Mesquite: a modular system for evolutionary analysis. http://www.mesquiteproject.org (accessed 4 May 2018).
Mai, U., Mirarab, S., 2018. TreeShrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees. BMC Genomics 19, 23–40. https://doi.org/10.1186/s12864-018-4620-2.
Malinsky, M., Matschiner, M., Svardal, H., 2021. Dsuite - fast *D*-statistics and related admixture evidence from VCF files. Mol. Ecol. Resour. 21, 584–595. https://doi.org/10.1111/1755-0998.13265.
Mathew, J., Yohannan, R., George, K.V., 2016. *Phyllagathis* Blume (Melastomataceae: Sonerileae), a new generic record for India with a new species. Bot. Lett. 163, 175–179. https://doi.org/10.1080/23818107.2016.1166453.
Maurin, O., Anest, A., Bellot, S., Biffin, E., Brewer, G., Charles-Dominique, T., Cowan, R. S., Dodsworth, S., Epitawalage, N., Gallego, B., Giaretta, A., Goldenberg, R., Gonçalves, D.J.P., Graham, S., Hoch, P., Mazine, F., Low, Y.W., McGinnie, C., Michelangeli, F.A., Morris, S., Penneys, D.S., Pérez-Escobar, O.A., Pillon, Y., Pokorny, L., Shimizu, G., Staggemeier, V.G., Thornhill, A.H., Tomlinson, K.W., Turner, I.M., 2021. A nuclear phylogenomic study of the angiosperm order Myrtales, exploring the potential and limitations of the universal Angiosperms353 probe set. Amer. J. Bot. 108, 1087–1111. https://doi.org/10.1002/ajb2.1699.
Maxwell, J.F., 1982. Taxonomic and nomenclatural notes on *Oxyspora* DC., *Anerincleistus* Korth., *Poikilogyne* Baker f., and *Allomorphia* BL. (Melastomataceae, tribe Oxysporeae). Gard. Bull. Singapore 35, 209–226.
Maxwell, J.F., 1984. Taxonomic studies of the Melastomataceae: A revision of subtribes Diplectrinae Maxw. and Dissochaetinae (Naud.) Triana (Genera *Diplectria* (Bl.) Reichb., *Dissochaeta* Bl., *Macrolenes* Naud., *Creochiton* Bl., and *Pseudodissochaeta* Nayar), Part 1. Fed. Mus. J. 29, 45–117.
Maxwell, J.F., 1989. The genus *Anerincleistus* Korth. (Melastomataceae). Proc. Acad. Nat. Sci. Philadelphia 141, 29–72.
Mcginnis, S.D., Madden, T.L., 2004. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucl. Acids Res. 32, 20–25. https://doi.org/10.1093/nar/gkh435.
McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303. https://doi.org/10.1101/gr.107524.110.
Mclean, B.S., Bell, K.C., Allen, J.M., Helgen, K.M., Cook, J.A., 2019. Impacts of inference method and data set filtering on phylogenomic resolution in a rapid radiation of ground squirrels (Xerinae: Marmotini). Syst. Biol. 68, 298–316. https://doi.org/10.1093/sysbio/syy064.
Mentink, H., Baas, P., 1992. Leaf anatomy of the Melastomataceae, Memecylaceae, and Crypteroniaceae. Blumea 37, 189–225.
Meyer, B.S., Matschiner, M., Salzburger, W., 2017. Disentangling incomplete lineage sorting and introgression to refine species-tree estimates for lake Tanganyika cichlid fishes. Syst. Biol. 66, 531–550. https://doi.org/10.1093/sysbio/syw069.
Michelangeli, F.A., Nicolas, A., Morales, P.M.E., David, H., 2011. Phylogenetic relationships of *Allomaieta*, *Alloneuron*, *Cyphostyla* and *Wurdastom* (Melastomataceae) and the resurrection of the tribe Cyphostyleae. Int. J. Plant Sci. 172, 1165–1178. https://doi.org/10.1086/662032.
Michelangeli, F.A., Guimarães, P.J.F., Penneys, D.S., Almeda, F., Kriebel, R., 2013. Phylogenetic relationships and distribution of new world Melastomateae (Melastomataceae). Bot. J. Linn. Soc. 171, 38–60. https://doi.org/10.1111/j.1095-8339.2012.01295.x.
Michelangeli, F.A., Almeda, F., Goldenberg, R., Penneys, D.S., 2020. A guide to curating New World Melastomataceae collections with a linear generic sequence to world-wide Melastomataceae. Preprints 2020, 2020100203. https://doi.org/10.20944/preprints202010.0203.v1.
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.
Mirarab, S., Bayzid, M.S., Boussau, B., Warnow, T., 2014a. Statistical binning enables an accurate coalescent-based estimation of the avian tree. Science 346, 6215. https://doi.org/10.1126/science.1250463.
Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., Warnow, T., 2014b. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30, 541–548. 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., Warnow, T., 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31, i44–i52. https://doi.org/10.1093/bioinformatics/btv234.
Misof, B., Misof, K., 2009. A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. Syst. Biol. 58, 21–34. https://doi.org/10.1093/sysbio/syp006.
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.
Morales-Briones, D.F., Kadereit, G., Tefarikis, D.T., Moore, M.J., Smith, S.A., Brockington, S.F., Timoneda, A., Yim, W.C., Cushman, J.C., Yang, Y., 2021. Disentangling sources of gene tree discordance in phylogenomic data sets: testing ancient hybridizations in Amaranthaceae s.l. Syst. Biol. 70, 219–235. https://doi.org/10.1093/sysbio/syaa066.
Nguyen, L., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. Mol. Biol. Evol. 32, 268–274. https://doi.org/10.1093/molbev/msu300.
Nikolov, L.A., Shushkov, P., Nevado, B., Gan, X., Al-Shehbaz, I.A., Filatov, D., Bailey, C. D., Tsiantis, M., 2019. Resolving the backbone of the Brassicaceae phylogeny for investigating trait diversity. New Phytol. 222, 1638–1651. https://doi.org/10.1111/nph.15732.
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.
Patel, R.K., Jain, M., 2012. NGS QC Toolkit: A toolkit for quality control of next generation sequencing data. PLoS ONE 7, e30619. https://doi.org/10.1371/journal.pone.0030619.
Patel, S., Kimball, R., Braun, E., 2013. Error in phylogenetic estimation for bushes in the tree of life. J. Phylogen. Evol. Biol. 1, 110. https://doi.org/10.4172/2329-9002.1000110.
Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., Reich, D., 2012. Ancient admixture in human history. Genetics 192, 1065–1093. https://doi.org/10.1534/genetics.112.145037.
Pease, J.B., Brown, J.W., Walker, J.F., Hinchliff, C.E., Smith, S.A., 2018. Quartet Sampling distinguishes lack of support from conflicting support in the green plant tree of life. Amer. J. Bot. 105, 385–403. https://doi.org/10.1002/ajb2.1016.
Penneys, D.S., Judd, W.S., 2011. Phylogenetics and morphology in the Blakeeae (Melastomataceae). Int. J. Plant Sci. 172, 78–106. https://doi.org/10.1086/657284.
Penneys, D.S., Judd, W.S., 2013. Combined molecular and morphological phylogenetic analyses of the Blakeeae (Melastomataceae). Int. J. Plant Sci. 174, 802–817. https://doi.org/10.1086/670011.
Pham, V.T., Vu, T.C., Rajapaksha, R., Trinh, N.B., Averyanov, L., Nguyen, T.L.T., 2017. *Phyllagathis phamhoangii* (Sonerileae, Melastomataceae), a new species from central Vietnam. Phytotaxa 314, 140–144. https://doi.org/10.11646/phytotaxa.314.1.15.
Philippe, H., Lartillot, N., Brinkmann, H., 2005. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. Mol. Biol. Evol. 22, 1246–1253. https://doi.org/10.1093/molbev/msi111.
Philippe, H., Laurent, J., 1998. How good are deep phylogenetic trees? Curr. Opin. Genet. 8, 616–623. https://doi.org/10.1016/S0959-437X(98)80028-2.
Pouchon, C., Fernández, A., Nassar, J.M., Boyer, F., Aubert, S., Lavergne, S., Mavárez, J., 2018. Phylogenomic analysis of the explosive adaptive radiation of the *Espeletia* complex (Asteraceae) in the tropical Andes. Syst. Biol. 67, 1041–1060. https://doi.org/10.1093/sysbio/syy022.
Purcell, S., Neale, B.M., Toddbrown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575. https://doi.org/10.1086/519795.
Regalado Jr., J.C., 1990. Revision of Medinilla (Melastomataceae) of Borneo. Blumea 35, 5–70.
Reginato, M., Neubig, K.M., Majure, L.C., Michelangeli, F.A., 2016. The first complete plastid genomes of Melastomataceae are highly structurally conserved. PeerJ 4, e2715. https://doi.org/10.7717/peerj.2715.
Reginato, M., Vasconcelos, T.N., Kriebel, R., Simoes, A.O., 2020. Is dispersal mode a driver of diversification and geographical distribution in the tropical plant family Melastomataceae? Mol. Phylogenet. Evol. 148, 106815 https://doi.org/10.1016/j.ympev.2020.106815.
Renner, S.S., 1993. Phylogeny and classification of the Melastomataceae and Memecylaceae. Nordic J. Bot. 13, 519–540. https://doi.org/10.1111/j.1756-1051.1993.tb00096.x.
Renner, S.S., Clausing, G., Meyer, K., 2001. Historical biogeography of Melastomataceae: the roles of Tertiary migration and long-distance dispersal. Amer. J. Bot. 88, 1290–1300. https://doi.org/10.2307/3558340.
Rhodes, J.A., Baños, H., Mitchell, J.D., Allman, E.S., 2021. MSCquartets 1.0: Quartet methods for species trees and networks under the multispecies coalescent model in R. Bioinformatics 37, 1766–1768. https://doi.org/10.1093/bioinformatics/btaa868.
Ridley, H.N., 1918. New and rare Malayan Plants Series X. J. Straits Branch Roy. Asiat. Soc. 79, 63–100.
Rieseberg, L.H., Whitton, J., Linder, C.R., 1996. Molecular marker incongruence in plant hybrid zones and phylogenetic trees. Acta Bot. Neerl. 45, 243–262.
Roch, S., Steel, M., 2015. Likelihood-based tree reconstruction on a concatenation of alignments can be statistically inconsistent. Theor. Popul Biol. 100, 56–62. https://doi.org/10.1016/j.tpb.2014.12.005.
Roch, S., Warnow, T., 2015. On the robustness to gene tree estimation error (or lack thereof) of coalescent-based species tree methods. Syst. Biol. 64, 663–676.
Schrempf, D., Szöllösi, G., 2020. The sources of phylogenetic conflicts. In: Celine, S., Frédéric, D., Nicolas, G. (Eds.), Phylogenetics in the Genomic Era 3-1. No commercial publisher, pp. 1–23.
Small, R., Cronn, R., Wendel, J., 2004. Use of nuclear genes for phylogeny reconstruction in plants. Austral. Syst. Bot. 17, 145–170. https://doi.org/10.1071/SB03015.
Solís-Lemus, C., Ané, C., 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. PLoS Genet. 12, e1005896 https://doi.org/10.1371/journal.pgen.1005896.
Stamatakis, A., 2014. RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML web servers. Syst. Biol. 57, 758–771. https://doi.org/10.1080/10635150802429642.
Stefanovic, S., Rice, D.W., Palmer, J.D., 2004. Long branch attraction, taxon sampling, and the earliest angiosperms: *Amborella* or monocots? BMC Evol. Biol. 4, 35. https://doi.org/10.1186/1471-2148-4-35.
Sukumaran, J., Holder, M.T., 2010. DendroPy: a Python library for phylogenetic computing. Bioinformatics 26, 1569–1571. https://doi.org/10.1093/bioinformatics/btq228.
Swofford, D.L., 2003. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
Ulloa Ulloa, C., Almeda, F., Goldenberg, R., Kadereit, G., Michelangeli, F.A., Penneys, D. S., Stone, R.D., Veranso-Libalah, M.C., in press. Melastomataceae: Global diversity, distribution, and endemism. In: Goldenberg, R., Michelangeli, F.A., Almeda, F. (Eds.), Advances in Melastomataceae Systematics and Biology. Springer Nature, Cham, Switzerland. doi: 10.1007/978-3-030-99742-7_15.
Vachaspati, P., Warnow, T., 2015. ASTRID: Accurate species trees from internode distances. BMC Genomics 16, 1–13. https://doi.org/10.1186/1471-2164-16-S10-S3.
Veranso-Libalah, M.C., Stone, R.D., Fongod, A.G., Couvreur, T.L., Kadereit, G., 2017. Phylogeny and systematics of African Melastomateae (Melastomataceae). Taxon 66, 584–614. https://doi.org/10.12705/663.5.
Wen, D., Nakhleh, L., 2018. Coestimating reticulate phylogenies and gene trees from multilocus sequence data. Syst. Biol. 67, 439–457. https://doi.org/10.1093/sysbio/syx085.
Wendel, J.F., Doyle, J.J., 1998. Phylogenetic Incongruence: Window into genome history and molecular evolution. In: Soltis, D.E., Soltis, P.S., Doyle, J.J. (Eds.), Molecular Systematics of Plants II. Springer, Boston, pp. 265–296. https://doi.org/10.1007/978-1-4615-5419-6_10.
Xi, Z., Liu, L., Davis, C.C., 2015. Genes with minimal phylogenetic information are problematic for coalescent analyses when gene tree estimation is biased. Mol. Phylogenet. Evol. 92, 63–71. https://doi.org/10.1016/j.ympev.2015.06.009.
Xi, Z., Liu, L., Davis, C.C., 2016. The impact of missing data on species tree estimation. Mol. Biol. Evol. 33, 838–860. https://doi.org/10.1093/molbev/msv266.
Xu, H., Luo, X., Qian, J., Pang, X.H., Song, J.Y., Qian, G.R., Chen, J.H., Chen, S.L., 2012. FastUniq: a fast de novo duplicates removal tool for paired short reads. PLoS ONE 7, e52249. https://doi.org/10.1371/journal.pone.0052249.
Xu, B., Yang, Z., 2016. Challenges in species tree estimation under the multispecies coalescent model. Genetics 204, 1353–1368. https://doi.org/10.1534/genetics.116.190173.
Yeh, C.L., Chung, S.W., Hsu, T.C., Yeh, C.R., 2008. Two new species of *Bredia* (Melastomataceae) from Taiwan. Edinburgh J. Bot. 65, 393–405. https://doi.org/10.1017/S0960428608005064.
Zeng, S.J., Zou, L.H., Wang, P., Hong, W.J., Zhang, G.Q., Chen, L.J., Zhuang, X.Y., 2016. Preliminary phylogeny of *Fordiophyton* (Melastomataceae), with the description of two new species. Phytotaxa 247, 45–61.
Zhang, L.N., Zhang, X.Z., Zhang, Y.X., Zeng, C.X., Ma, P.F., Zhao, L., Guo, Z.H., Li, D.Z., 2014. Identification of putative orthologous genes for the phylogenetic reconstruction of temperate woody bamboos (Poaceae: B ambusoideae). Mol. Ecol. Resour. 14, 988–999. https://doi.org/10.1111/1755-0998.12248.
Zhou, Q.J., Lin, C.W., Dai, J.H., Zhou, R.C., Liu, Y., 2019a. Exploring the generic delimitation of *Phyllagathis* and *Bredia* (Melastomataceae): A combined nuclear and chloroplast DNA analysis. J. Syst. Evol. 57, 256–267. https://doi.org/10.1111/jse.12451.
Zhou, Q.J., Lin, C.W., Ng, W.L., Dai, J.H., Denda, T., Zhou, R.C., Liu, Y., 2019b. Analyses of plastome sequences improve phylogenetic resolution and provide new insight into the evolutionary history of Asian Sonerileae/Dissochaeteae. Front. Plant Sci. 10, 1477. https://doi.org/10.3389/fpls.2019.01477.
Zhou, Q.J., Dai, J.H., Lin, C.W., Denda, T., Zhou, R.C., Liu, Y., 2019c. Recircumscription of *Bredia* and resurrection of *Tashiroea* (Sonerileae, Melastomataceae) with description of a new species *T. villosa*. PhytoKeys 127, 121–150. https://doi.org/10.3897/phytokeys.127.36608.
Zhou, R.C., Zhou, Q.J., Liu, Y., 2018. *Bredia repens* (Melastomataceae), a new species from Hunan. China. Syst. Bot. 43, 544–551. https://doi.org/10.1600/036364418X697265.
Zhu, J., Liu, X., Ogilvie, H.A., Nakhleh, L.K., 2019. A divide-and-conquer method for scalable phylogenetic network inference from multilocus data. Bioinformatics 35, i370–i378. https://doi.org/10.1093/bioinformatics/btz359.