Original Article

Target Enrichment Data Uncovers Rapid Radiation, Whole Genome Duplication, and Extensive Hybridization in Slipper Orchid Genus *Cypripedium* L.

Loudmila Jelinscaia Lagou[1], Gudrun Kadereit[1,2] and Diego F. Morales-Briones[1,*]

[1]*Princess Therese von Bayern chair of Systematics, Biodiversity and Evolution of Plants, Ludwig-Maximilians-Universität München, Menzinger Str. 67, Munich 80638, Germany*

[2]*Botanical Garden Munich and Botanical State Collection Munich, Bavarian Natural History Collections, Menzinger Str. 65-67, Munich 80638, Germany*

[*]*For correspondence. E-mail dfmoralesb@gmail.com*

- **Background and Aims:** *Cypripedium* is the most widespread and morphologically diverse genus of slipper orchids. Despite several published phylogenies based on Sanger sequencing data, the topology and monophyly of its infrageneric taxa remained uncertain. Here, we aimed to reconstruct a robust section-level phylogeny of *Cypripedium* and explore its evolutionary history using target capture data for the first time.
- **Methods:** We used the orchid-specific bait set "Orchidaceae963" to reconstruct the phylogeny of *Cypripedium* based on 614 nuclear loci, covering 11 out of 13 sections. Subsequently, we investigated tree discordance, estimated divergence times and ancestral ranges, searched for anomaly zones, polytomies, and diversification rate shifts, and identified gene duplication and hybridization events.
- **Key Results:** All sections were recovered as monophyletic, contrary to the subsections within sect. *Cypripedium*. Although the two subclades within this section did not correspond to its two subsections, they matched the geographic distribution of their species. Additionally, we discovered high levels of discordance in the short backbone branches of the genus and within sect. *Cypripedium*, which can be attributed to gene duplication and hybridization events, a potential whole genome duplication, and incomplete lineage sorting caused by rapid radiation. Our biogeographic analysis suggested a Neotropical origin of the genus during the Early Miocene (~20 Ma). The rapid radiations at the backbone likely occurred in Southeast Asia around the Middle Miocene Climatic Transition (~15-13 Ma), followed by several independent dispersals back to the New World. Moreover, the Pliocene-Quaternary glacial cycles may have contributed to further speciation and reticulate evolution, giving rise to a hybrid swarm within sect. *Cypripedium*.

- **Conclusions:** Our study provided novel insights into the evolutionary history of *Cypripedium* based on high-throughput molecular data, shedding light on the dynamics of its distribution and diversity patterns from its origin to the present.

Key words: *Cypripedium*, slipper orchids, phylogenomics, target enrichment, historical biogeography, anomaly zone, rapid radiation, whole genome duplication, reticulate evolution, hybridization.

INTRODUCTION

The family Orchidaceae comprises the most species-rich family of vascular plants, with c. 28,000 species in five subfamilies and ~750 genera (Chase *et al.*, 2015; Christenhusz *et al.*, 2017). Their great diversity has fascinated and puzzled scientists for centuries, including the father of evolutionary theory, Charles Darwin, who once wrote, "I never was more interested in any subject in my life, than in this of Orchids" (*Darwin Correspondence Project*, n.d.). Unfortunately, today, their diversity is highly threatened mainly due to habitat destruction and unsustainable harvesting (DL Roberts and Dixon, 2008), prompting their protection by local and national laws, as well as the Convention on International Trade in Endangered Species (aka CITES; *Appendices I, II, and III*, 2023). In efforts to describe their diversity and facilitate informed conservation measures, a variety of molecular data, including high-throughput genomic and transcriptomic data, has been used to reveal the relationships between orchid subfamilies in recent decades (Cameron *et al.*, 1999; Freudenstein *et al.*, 2004; Givnish *et al.*, 2015; Kim *et al.*, 2020; Pérez-Escobar *et al.*, 2021; Serna-Sánchez *et al.*, 2021). However, phylogenetic support at lower taxonomic ranks in Orchidaceae has been low primarily due to the limited genetic variation in the commonly used markers (i.e., *rbcL*, *matK*, ITS, chloroplast intergenic spacers). The genus *Cypripedium* L. is one such orchid taxon whose internal phylogenetic relationships have yet to be resolved.

Cypripedium is a genus of temperate perennial herbs in the subfamily of lady's slipper orchids, Cypripedioideae, and it currently consists of approximately 50 accepted species (Frosch and Cribb, 2012; SC Chen *et al.*, 2013; Christenhusz *et al.*, 2017; POWO, 2023). Although *Cypripedium* has about half the species number of the largest slipper orchid genus, *Paphiopedilum* Pfitzer, it is the most morphologically diverse (Fig. 1) and widespread (Supplementary Data Fig. S1) of all five cypripedioid genera. Its distribution is mainly

circumboreal, but its range extends from the Arctic Circle to Central America (~14º–70º North; J Li *et al.*, 2011; Frosch and Cribb, 2012). Eastern Asia, especially temperate China, constitutes the genus' main center of diversity, harboring approximately 70% of all *Cypripedium* species (J Li *et al.*, 2011). They occur in various habitats and altitudes, from forests to wetlands and grasslands and from sea level to 4,900 m in the Himalayas (Frosch and Cribb, 2012).

Figure 1(a): Pictures of the *Cypripedium* taxa per section included in the final phylogeny.
(**A**) *C. irapeanum*,
(**B**) *C. plectrochilum*,
(**C**) *C. subtropicum* (= *C. singchii*; Frosch & Cribb, 2012),
(**D**) *C. wardii*,
(**E**) *C. margaritaceum*,
(**F**) *C. fargesii*,
(**G**) *C. lichiangense*,
(**H**) *C. lentiginosum*,
(**I**) *C. sichuanense*,
(**J**) *C. passerinum*,
(**K**) *C. flavum*,
(**L**) *C. reginae*,
(**M**) *C. reginae* var. *alba*,
(**N**) *C. californicum*,
(**O**) *C. acaule*,
(**P**) *C. yatabeanum*,
(**Q**) *C.* × *alaskanum*,
(**R**) *C. guttatum*,
(**S**) *C. japonicum*,
(**T**) *C. formosanum*,
(**U**) *C. bardolphianum*,
(**V**) *C. micranthum*.

Credits: (A) by M. Béhar; (B), (K), (P), (Q), (T), and (V) by J.-B. Chazalon; (C), (D), (H), (J), and (O) by W. Frosch; (R) and (S) by L. Chen; (E)-(G), (I), (N), and (U) by S. Urban; (L) by B. Isaac; (M) by M. Sunouchi (see **Acknowledgements** for more details).

(b) *Cypripedium*

Figure 1(b): Pictures of the taxa of *Cypripedium* sect. *Cypripedium* included in the final phylogeny. (**A**) *C. macranthos* var. *rebunense*, (**B**) *C. macranthos* var. *alba*, (**C**) *C. macranthos* var. *taiwanianum*, (**D**) *C. macranthos* var. *macranthos*, (**E**) *C. macranthos* var. *hotei-atsumorianum*, (**F**) *C. macranthos* var. *speciosum*, (**G**) *C. franchetii*, (**H**) *C. yunnanense*, (**I**) *C. calcicola*, (**J**) *C. tibeticum*, (**K**) *C. amesianum*, (**L**) *C. froschii*, (**M**) *C. himalaicum*, (**N**) *C. fasciolatum*, (**O**) *C. farreri*, (**P**) *C. calceolus*, (**Q**) *C.* × *ventricosum*, (**R**) *C. shanxiense*, (**S**) *C. segawai*, (**T**) *C. henryi*, (**U**) *C. cordigerum*, (**V**) *C. candidum*, (**W**) *C. kentuckiense*, (**X**) *C. parviflorum* var. *pubescens* forma *planipetalum*, (**Y**) *C. parviflorum* var. *makasin* (= *C. parviflorum* var. *parviflorum*; Frosch & Cribb, 2012), (**Z**) *C. parviflorum* var. *pubescens*, (**AA**) *C.* × *columbianum*, (**BB**) *C. montanum*.

Credits: (A), (L), (Q), and (Y) by V. Steindl; (D), (E), (G), (J), (S)-(U), (X), (Z), and (AA) by J.-B. Chazalon; (P) by L. Chen; (W) by Orchi; (B), (F), (H), (I), (K),(M), (O), and (R) by S. Urban; (C), (N), (V), and (BB) by W. Frosch (see **Acknowledgements** for more details).

Like other slipper orchids, flowers of *Cypripedium* species have a profoundly inflated, slipper-shaped lip (i.e., labellum) that gives them their distinctive morphology. The lip traps pollinators that enter through the upward-facing opening thanks to its incurved, glabrous, slippery margins, with the only escape routes passing through its basal orifices under the two anthers at each side of the column (Cribb, 1997; Frosch and Cribb, 2012). Unlike other slipper orchids, *Cypripedium* species are traditionally recognized by their (usually) plicate leaves and unilocular ovaries with parietal placentation (Cox *et al.*, 1997; Cribb, 1997). Although the reliability of these distinctive characters has been questioned (Atwood, 1984), phylogenetic studies consistently support the monophyly of the genus (Fatihah *et al.*, 2011; J Li *et al.*, 2011; Guo *et al.*, 2012; H Liu *et al.*, 2021*a*; Szlachetko *et al.*, 2021; J-Y Zhang *et al.*, 2022). On the other hand, its infrageneric classification has constantly changed during the last two centuries.

Following *Cypripedium*'s description, the great interest in Cypripedioideae led to numerous taxonomic revisions in the subfamily with often incongruent results (Linnaeus, 1753; Rafinesque, 1836; Lindley, 1840; Reichenbach, 1854; Pfitzer, 1888, 1894; Rolfe, 1896; Atwood, 1984; Cox *et al.*, 1997; Eccarius, 2009; Perner, 2008; Supplementary Data Table S1). To name a few, Linnaeus (1753) initially recognized only one species of *Cypripedium* (i.e., *C. calceolus* L.) and a few varieties currently holding a species status. Lindley (1840) described 22 species within the genus classified in a number of subgeneric groups. The classifications of Pfitzer (1903) taxonomically expanded *Cypripedium* with 28 species and numerous subgeneric categories, including four sections. In Cribb's (1997) taxonomic treatment, the number of species increased to 45 and the sections to 11, while Eccarius (2009) divided *Cypripedium* into two subgenera, 13 sections, and 37 species, lowering the rank of multiple species to subspecies or varieties.

Recent molecular phylogenies based on nrDNA ITS and five cpDNA markers by J Li *et al.* (2011) indicated that, among the non-monotypic groups, eight sections are monophyletic

[*Arietinum* C. Morren, *Bifolia* (Lindl.) S. C. Chen, *Cypripedium*, *Flabellinervia* (Pfitzer) Hennessy ex P. J. Cribb, *Obtusipetala* (Pfitzer) P. J. Cribb, *Sinopedilum* Perner, *Subtropica* S. C. Chen & K. Y. Lang, and *Trigonopedia* Franch.] while two sections [*Irapeana* P. J. Cribb and *Retinervia* (Pfitzer) S. C. Chen] and the two subsections of sect. *Cypripedium* [*Cypripedium* and *Macrantha* (Kraenzl) P. J. Cribb] are non-monophyletic, following the classification by Cribb (1997) and Perner (2008). These results prompted further infrageneric treatments by Frosch and Cribb (2012) and SC Chen *et al.* (2013), producing the two currently used classification systems of *Cypripedium*. Although based on the same phylogenies by J Li et al. (2011), Frosch and Cribb (2012) proposed 13 sections with 48 species, whereas SC Chen *et al.* (2013) increased these numbers to 15 and 51, respectively, adding two new monotypic sections: *Palangshanensia* S. C. Chen & Z. J. Liu and *Wardiana* S. C. Chen & Z. J. Liu (Supplementary Data Table S1).

After the publication of J Li *et al.* (2011), several studies included molecular phylogenies with *Cypripedium* species, four of which specifically focused on the relationships of the infrageneric taxa of *Cypripedium* (Fatihah *et al.*, 2011; Guo *et al.*, 2012; H Liu *et al.*, 2021*a*; Szlachetko *et al.*, 2021; J-Y Zhang *et al.*, 2022). These studies used a multilocus approach with up to eight Sanger-sequenced nuclear and chloroplast DNA markers in different combinations and with different phylogenetic reconstruction methodologies (i.e., Parsimony, Maximum Likelihood, and Bayesian Inference). The topologies and the monophyly of some subgeneric taxa were congruent among the produced phylogenies (e.g., sect. *Irapeana* being sister to the rest; monophyly of sect. *Arietinum, Bifolia, Cypripedium, Flabellinervia, Obtusipetala, Sinopedilum*, and *Trigonopedia*). However, the topology and monophyletic status of other taxa (e.g., the monophyly of the two subsections within sect. *Cypripedium*) and the topology at the backbone of the phylogeny remain uncertain.

The unresolved phylogeny of the genus *Cypripedium* not only prevents the accurate evaluation of

the relationships between the currently established subgeneric groups but also our understanding of their evolutionary history. A well-resolved and robust phylogeny is fundamental for addressing further questions regarding their divergence time, diversification rate shifts, ancestral spatial distribution patterns, and hybridization events. Furthermore, it will provide a solid foundation for the efficient management of their conservation, especially as their continuous human-driven population decline is predicted to exacerbate due to climate change (Nicolè *et al*., 2005; Izawa *et al.*, 2007; Minasiewicz *et al.*, 2018; Kolanowska and Jakubska-Busse, 2020; H Liu *et al.*, 2021*b*; Chandra *et al.*, 2023; Yamashita *et al.*, 2023).

It is widely recognized that the use of multiple genes can improve the accuracy of phylogenetic reconstruction, and single- or low-copy nuclear genes are increasingly used due to their rapid evolutionary rates and biparental inheritance (Guo *et al.*, 2012; N Zhang *et al.*, 2012; Z Li *et al.*, 2017). A target enrichment approach would allow the sequencing of hundreds of low-copy markers via high-throughput sequencing methods and, therefore, more robust estimates of relationships with greater support. Moreover, the use of the recently designed orchid-specific baits "Orchidaceae963" by Eserman *et al.* (2021) could provide sufficient information to resolve recent and rapid radiations in deep and shallow phylogenetic scales, allowing for the characterization of species-level relationships and the resolution of long-debated polytomies within Orchidaceae.

In this study, we used a target enrichment approach using the "Orchidaceae963" baits to reconstruct a well-supported phylogeny of the genus *Cypripedium* at the section level. Based on our results, we evaluated the two most recently published classification systems of the genus by Frosch and Cribb (2012) and SC Chen *et al.* (2013) and the congruence of the recovered relationships with published phylogenies based on Sanger data. Additionally, we aimed to gain new insights into the evolution of *Cypripedium* by answering the following questions: (1) Does

the current classification stand up to phylogenetic reconstructions based on genomic data? (2) Which biological processes explain higher levels of gene tree discordance in some parts of the phylogeny? (3) When and where did *Cypripedium* originate and diversify and how did this diversification relate to geographic expansion of the lineages and paleoclimate? (4) Is the hybrid status of some taxa supported by molecular data? To answer these questions, we explored tree discordance, estimated divergence times and ancestral ranges, searched for anomaly zones and diversification rate shifts, and identified gene duplication and hybridization events.

MATERIALS AND METHODS

Taxon Sampling

We sampled leaf tissue from 53 specimens representing 36 species, eight varieties, and three natural hybrids of the genus *Cypripedium* (following the taxonomy of Frosch and Cribb, 2012; Supplementary Data Table S2). Fifty of the sampled individuals came from the Botanical Collection at Oberhof, Eurasburg, associated with the Botanical Garden Munich-Nymphenburg (BGM), and three from the Botanische Staatssammlung München herbarium (BSM-SNSB, herbarium acronym M). The material from the living collection was stored in silica-gel to be dried immediately after collection.

The sequence data from the above-collected tissue samples was combined with publicly available orchid genomes and transcriptomes. These represented species from all slipper orchid genera (incl. ten *Cypripedium* species, three of which are new to our sampling) as well as orchids from three outgroup subfamilies (i.e., Apostasioideae, Epidendroideae, and Vanilloideae). As a result, our final dataset included *Cypripedium* species from all sections except sect. *Enantiopedilum*

Pfitzer, consisting of *C. fasciculatum* and *C. palangshanense*, and sect. *Retinervia*, consisting of *C. elegans* and *C. debile* (following Frosch and Cribb, 2012; Supplementary Data Table S3).

Library Preparation, Target Enrichment, and Sequencing

We isolated total genomic DNA from silica-dried or herbarium leaf tissue using the NucleoSpin Plant II kit: Genomic DNA from plants (Macherey-Nagel, Düren, Germany) following a modified version of the manufacturer's manual (Supplementary Data Table S4). Next, we quantified the DNA concentration with a Qubit 4 fluorometer using a Broad Range (BR) or High Sensitivity (HS) assay kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA). DNA was sheared to an average fragment size of 350 bp with a Covaris M220 Focused-ultrasonicator (Covaris, Woburn, Massachusetts, USA). We assessed the DNA fragment size distribution using a High Sensitivity DNA ScreenTape on a 4150 TapeStation System (Agilent Technologies, Santa Clara, California, USA).

We prepared dual indexed libraries according to the instruction manual using the NEBNext Ultra II DNA Library Prep Kit for Illumina and the NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 1, New England Biolabs, Ipswich, Massachusetts, USA) and following the recommended conditions of bead-based size selection according to distribution of DNA fragments per sample. Next, we amplified the adaptor-ligated libraries with eight PCR cycles and measured DNA concentration using the Qubit. The average fragment size of the libraries was assessed with the TapeStation. Prior to hybridization, the libraries were pooled in equal concentrations to include 250 ng of each library, with a maximum of 15 libraries per pooled library.

For the hybridization enrichment reaction, we combined the pooled libraries with the custom orchid-specific bait set “Orchidaceae963” (Daicel Arbor Biosciences myBaits Target Capture

Kit, Ann Arbor, MI, USA) and incubated at 60 °C (hybridization temperature, TH) for 16 hours overnight, following the Standard Protocol and the Blockers Mix setup designed for plants (myBaits Hybridization Capture for Targeted NGS, User Manual v. 5.02). The bead-based cleanup of the bait-target hybrids was performed at a wash temperature (TW) of 60 °C, and the hybridized libraries were subsequently amplified for 14 PCR cycles at 60 °C (annealing temperature, TA). Then, we purified the amplification reaction following the PCR clean-up protocol of the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel, Düren, Germany). Finally, we checked the concentration and fragment size distribution of the libraries as before, using the Qubit and the TapeStation. The enriched pooled libraries were sequenced on an Illumina NovaSeq 6000 sequencing system (SP flow cell) at the Core Facility Genomics (CF-GEN) of the Helmholtz Zentrum München, Germany (Deutsches Forschungszentrum für Gesundheit und Umwelt, GmbH).

Read Processing and Assembly

We created a set of references from orchid genomes and transcriptomes available on the Sequence Read Archive (SRA) of NCBI to improve gene extractions (Supplementary Data Table S4; Sayers *et al.*, 2022). Original target exon sequences from the Orchidaceae963 bait set (https://github.com/laeserman/Orchidaceae963/blob/main/Orchidaceae963-targets.fa) were concatenated into 'genes' and used to identify the corresponding complete CDS from the *Phalaenopsis equestris* (Schauer) Rchb. f. genome using BLAST. When no hits were produced, we used the genome of *Dendrobium catenatum* Lindl. instead. In the end, 950 out of the original 963 genes were extracted. Then, we used the raw transcriptome assembly from 18 orchids mentioned in the *Taxon Sampling* section, including *Vanilla shenzhenica* Z. J. Liu & S. C. Chen, and 17 species of slipper orchids to extend the genome references. RNAseq data processing and

transcriptome assembly followed Morales-Briones *et al.* (2021). We used CAPTUS v.0.9.90 (Ortiz *et al.*, 2023) to extract the corresponding loci from the 18 transcriptomes and the genomes of *Apostasia shenzhenica* Z. J. Liu & L. J. Chen, *D. catenatum*, *P. equestris*, and *Vanilla planifolia* Andrews. The extracted loci were used as the extended reference dataset for loci extraction in our own generated target enrichment data of *Cypripedium*.

We checked the quality of the raw reads using FastQC v0.11.9 (Andrews, 2010) and MultiQC v1.14 (Ewels *et al.*, 2016). Then, we used CAPTUS to trim the sequencing adaptors and low-quality bases, assemble the reads, and extract the nuclear loci based on the reference dataset. To decrease the retention of contigs resulting from potential erroneous reads, we set the minimum contig depth to eight for the assembly step and the minimum percentages of identity and coverage to 75 and 50, respectively, for the extraction step. Similarly, we assembled and extracted the nuclear loci from the genomes and transcriptomes used as references to combine them with our data for further analysis. We then extracted the coding sequences from the combined dataset using CAPTUS, with the removal of paralogs disabled.

Orthology Inference

To infer orthologs for the phylogenetic reconstruction, we followed a modified version of the methods described in Morales-Briones *et al.* (2022; https://bitbucket.org/dfmoralesb/target_enrichment_orthology). First, we re-aligned the untrimmed and unfiltered output alignments from CAPTUS using MACSE v2.07 (Ranwez *et al.*, 2018) with default parameters. Next, we replaced "!" with gaps at the frameshifts and used Phyx (JW Brown *et al.*, 2017) to remove aligned columns with more than 90% missing data. We inferred maximum likelihood (ML) homolog gene trees using IQ-TREE 2 v.2.0.7 using extended model selection (Kalyaanamoorthy *et al.*, 2017) and no clade support (Minh *et al.*, 2020). Then,

we masked mono- and paraphyletic tips that belong to the same taxon, keeping the tips with the most unambiguous characters in the trimmed loci alignments for each taxon as described in (Yang and Smith, 2014). Spurious tips with unusually long branches were removed by reducing the tree diameter with TreeShrink v.1.3.9 (Mai and Mirarab, 2018). We ran TreeShrink twice with the quantile set to 0.01, using the output of the first run for the second run to avoid over-trimming. We wrote FASTA files from the output homolog trees and followed the same steps as for the output alignments from CAPTUS, aligning them using MACSE with the default parameters, replacing the "!" with gaps at frameshifts, and removing aligned columns with >90% missing data with Phyx. To infer the final homolog gene trees, we used IQ-TREE 2, extended model selection, and assessed the clade support with 1,000 ultrafast bootstrap (BS) replicates.

Orthology inference was carried out with the tree-based "monophyletic outgroup" (MO) method described in Yang and Smith (2014). The MO method searches for clusters with monophyletic ingroups rooted at the outgroups in the homolog trees, discarding those with duplicated taxa in the outgroups. Subsequently, it infers the orthologs from root to tip, keeping the ortholog subtree with the most taxa. To infer the orthologs, we set all Cypripedioideae members as ingroup, and the remaining taxa (i.e., *A. shenzhenica*, *D. catenatum*, *Phalaenopsis equestris*, *Vanilla planifolia*, *V. shenzhenica*) as outgroups. In an initial MO run, we kept only ortholog groups with at least 20 ingroup taxa. *Paphiopedilum malipoense* S. C. Chen & Z. H. Tsi, was causing the outgroups of most loci trees to be "non-monophyletic", due to low quality of the transcriptome or a mislabeling of the raw data sample submission; therefore, it was removed, and the MO ortholog inference was repeated. Following the second MO run, we removed the samples present in ≤35% of the ortholog trees. Finally, we repeated the orthology inference, keeping only ortholog groups with at least 35 ingroup taxa (~50% of retained samples), retaining a dataset of 74 samples and 792 ortholog trees (326–789 ortholog trees per sample; Supplementary Data Table S5).

Phylogenetic Reconstruction

We used concatenation and coalescent-based methods to reconstruct the phylogeny of *Cypripedium*. First, a concatenated alignment was produced using the clean ortholog alignments (following the Phyx step), retaining 614 orthologs with at least 500 characters and 50 taxa. We estimated a ML tree of the concatenated matrix with IQ-TREE 2. We searched for the best partition scheme using ModelFinder implemented within IQ-TREE (Kalyaanamoorthy *et al.*, 2017) and 1,000 ultrafast BS replicates to assess clade support. Regarding the coalescent-based approach, we first inferred ML trees from the same 614 individual orthologs used for the concatenation-based phylogeny. Individual ortholog ML trees were inferred as previously described for the final homolog trees. Then we used the quartet-based species-tree inference method ASTRAL v1.15.2.4 (wASTRAL-unweighted), which is statistically consistent under the multispecies coalescent (MSC) model and thus useful for handling incomplete lineage sorting (C Zhang *et al.*, 2018; C Zhang and Mirarab, 2022). We used the 614 individual ML ortholog trees, default ASTRAL parameters, and branch support was assessed using local posterior probabilities (LPP; Sayyari and Mirarab, 2016). Due to the similarity in the topologies recovered between the concatenation and coalescent-based approaches, all subsequent analyses were carried out using the ASTRAL species tree unless stated otherwise.

Gene Tree Discordance Estimation

We quantified the conflict among gene trees on each node of the inferred species tree by estimating the number of conflicting and concordant bipartitions with Phyparts (Smith *et al.*, 2015). To do this, we used the individual ML ortholog trees and set a threshold of at least 50% BS support for a node to be considered informative. We plot the Phyparts result using the

"missing and uninformative" script (i.e., "phypartspiecharts_missing_uninformative.py; https://bitbucket.org/dfmoralesb/target_enrichment_orthology) with Python v3.10.10 to add pie charts at the nodes while taking into consideration missing data (i.e., when input trees do not have the same number of tips).

We also used Quartet Sampling (QS; Pease *et al.*, 2018) to differentiate between lack of support and conflicting nodes on the species tree. QS estimates branch support and conflict by sampling quartets from the species tree and the corresponding concatenated alignment and calculating the proportion of the three possible topologies at each node. As a result, it simultaneously evaluates the consistency of information (Quartet Concordance, QC), the presence of secondary evolutionary histories (Quartet Differential, QD), and the amount of information (Quartet Informativeness, QI) of internal nodes. We ran with 1,000 QS replicates with RAxML-NG (Kozlov *et al.*, 2019) as ML inference tool. The results were plotted using R by color-coding the values of QC on each node and annotating them with the rest of the estimated values (https://bitbucket.org/yanglab/conflict-analysis/src/master/).

Anomaly Zone Test

The anomaly zone, characterized by the presence of a set of short internal branches in the species tree, occurs when gene tree topologies that are discordant with the species tree topology are observed more frequently than those that are concordant (Linkem *et al.*, 2016). It arises from consecutive rapid diversification events leading to incomplete lineage sorting (ILS).

We estimated the boundaries of the anomaly zone for the internal nodes of our species tree following the calculations in Linkem *et al.* (2016) (https://github.com/cwlinkem/anomaly_zone) to investigate whether the high amount of gene tree discordance observed in numerous short branches of the tree could be explained by ILS. The calculations are based on equation 4 of

Degnan and Rosenberg (2006), which defines the boundaries of the anomaly zone, $\alpha(x)$. In this equation, x is the length of an internal branch in the species tree, and its descendant internal branch has a length y (in coalescent units). If y is $< \alpha(x)$, then the internode pair is considered to be in the anomaly zone.

Polytomy Test

Additionally, due to the presence of short branches with low support, we tested whether we could reject the null hypothesis that any branch in the species tree has a length equal to 0, or in other words, is a polytomy. We used the polytomy test (-t 10) option in ASTRAL version 5.7.8 with default parameters (Sayyari and Mirarab, 2018). The ASTRAL polytomy test relies on the distribution of the quartet frequencies of gene trees around each branch of the species tree to test this hypothesis, annotating the branches of the output tree with the resulting p-values. Under the null hypothesis, the three unrooted quartet topologies defined around the branch are expected to have equal frequencies. Although failure to reject the null hypothesis may indicate a real (i.e., hard) polytomy, it might also be caused by lack of power or signal (i.e., soft polytomy).

Mapping Whole Genome Duplications

We mapped gene duplication events on our species tree based on the subclade orthogroup tree topology method described in Yang *et al.* (2018; https://bitbucket.org/blackrim/clustering/src/master/). This method extracts the rooted orthogroups from each homolog tree. Then, it detects gene duplication events when the orthogroup subclades share two or more ingroup taxa and maps the percentage of duplicated genes to the corresponding branch of the species tree. Alternatively, the duplications are mapped on the most recent common ancestor branch if the gene tree has missing taxa or if its topology is

incongruent with the species tree. To avoid the overestimation of the duplication percentages due to nested duplications, each branch of the species tree is restricted to one duplication event for each extracted clade (Yang *et al.,* 2015).

Similarly to Yang *et al.* (2018), we tested two filters to map the gene duplications: a bootstrap and a local topology filter. The bootstrap filter requires orthogroups to have an average bootstrap percentage of ≥ 50% (Z Li *et al.,* 2015), while the local topology filter requires the sister clade of the gene duplication branch in the orthogroup to include a subset of the taxa in the corresponding sister clade in the species tree (Cannon *et al.,* 2015). We plotted the percentages of gene duplications per number of branches in R (R Core Team, 2023). A WGD will produce a large burst (i.e., an outlier percentage, usually ≥ 20%) of shared duplications across taxa and loci (Yang *et al.,* 2018).

Testing Hybridization Events

Our dataset included three taxa described as hybrids by both Frosch and Cribb (2012) and SC Chen *et al.* (2013); namely, *C.* × *alaskanum* P. M. Br. (*C. guttatum* Sw. × *C. yatabeanum* Makino), *C.* × *columbianum* Sheviak [*C. montanum* Douglas ex Lindl. × *C. parviflorum* Salisb. var. *pubescens* (Willd.) O. W. Knight] and *C.* × *ventricosum* Sw. (*C. calceolus* × *C. macranthos* Sw.). We tested whether our data supports their status as hybrids of their putative parent taxa using explicit phylogenetic networks in PhyloNet (Wen *et al.,* 2018). Phylonet allows for horizontal edges that visualize the genetic inheritance through gene flow, mapping the inheritance probabilities (γ) for each parent hybrid edge to estimate the percentage of loci a hybrid inherited from each parent.

To reconstruct the phylogenetic networks, we first rooted the final ortholog trees and extracted the three subclades containing each hybrid, along with the putative parents and other taxa sharing

their MRCA. We reduced computational load by removing duplicated taxa, leaving a single representative for each monophyletic taxon. In the case of paraphyletic taxa, one representative taxon was left from each conspecific monophyletic subclade or from a group of consecutively diverging conspecific varieties. The taxa present in most orthologs were favored to maximize the final number of loci used for Phylonet. Similarly, *Phalaenopsis equestris* was chosen as an outgroup taxon because it had the highest amount of retained loci. Gene trees missing any of these selected taxa were excluded from the analysis.

Since calculating the likelihood of a phylogenetic network is computationally intensive in PhyloNet, we inferred the phylogenetic networks based on a maximum pseudo-likelihood (MPL) measure via the InferNetwork_MPL command (Yu and Nakhleh, 2015). We set the number of maximum reticulation events to one and the number of optimal output networks to ten for all three tests. The option "po" was specified to optimize the branch lengths and inheritance probabilities under full likelihood for the inferred *C.* × *alaskanum* networks. This optimization was only performed for the *C.* × *alaskanum* networks, which contain only four taxa, as it gets more time-consuming with an increasing number of taxa.

Furthermore, since our sampling included other taxa that have been proposed—but not widely accepted—as hybrids (e.g., *C. froschii* Perner, also known as *C.* × *froschii* Perner) and since hybridization is considered to be pervasive within *Cypripedium* in nature (Klier *et al.*, 1991; S-J Hu *et al.*, 2011; Frosch and Cribb, 2012; Szlachetko *et al.*, 2017; Pupulin and Díaz-Morales, 2018), we ran further tests setting the number of maximum reticulation events from two to ten for the networks containing *C.* × *columbianum* and *C.* × *ventricosum*.

The phylogenetic networks with the highest total log probability were visualized in Dendroscope v3.8.8 (Huson and Scornavacca, 2012) and the inheritance probabilities were mapped with

PhyloNetworks' v0.16.2 (Solís-Lemus *et al.,* 2017) companion package PhyloPlots v1.0.0 (https://github.com/cecileane/PhyloPlots.jl) in Julia v1.9.2 (Bezanson *et al.,* 2012).

Divergence Time Estimation

We used a Bayesian Inference approach for divergence time estimation. To decrease computational resources, we reduced the volume of the datasets to a subset of genes providing the most useful information relevant to time calibration via a "gene shopping" method. Specifically, we used the SortaDate package developed by Smith *et al.* (2018) to filter the 20 best ortholog genes based on the (a) least topological conflict with a focal species tree (i.e., bipartition calculation), (b) clock-likeness (i.e., root-to-tip variance statistic calculation), and (c) discernible information content (i.e., total tree length), sorting the genes in the respective order of these properties (i.e., a, b, c).

We concatenated the resulting subset of genes and defined the positions of the 20 loci as data blocks to find the best partitioning schemes and models of nucleotide evolution with PartitionFinder 2 (Lanfear *et al.,* 2017) to inform our site model selection for the molecular calibration. The branch lengths were estimated independently for each subset (i.e., unlinked), the corrected Akaike's Information Criterion (AICc) was used to select the best-fit nucleotide substitution models among those available in BEAST 2, while the "greedy" algorithm was used to search for a good partitioning scheme. The results suggested 18 partitioning sets, with GTR+I+G4+X as the best-fit model for 17 sets and GTR+G4+X as the best-fit model for the remaining one. For this reason and to reduce computational time, we decided to carry out molecular dating with the nucleotide substitution model GTR+I+G4+X without partitioning the selected loci.

In detail, the analysis was performed using an unpartitioned scheme in BEAST v2.7.4 (Bouckaert *et al.*, 2019) under a relaxed uncorrelated lognormal clock model (Optimized Relaxed Clock, ORC; Mean clock rate: 1.0), a GTR+I+G4+X site model and using a random starting tree for 200 million generations, sampling every 10,000 generations. Since no available fossil of slipper orchids can be used for calibration and our sampling of orchids apart from Cypripedioideae is scarce, we used a secondary calibration point based on the age estimates by Givnish *et al.* (2015). The study above reconstructed a broad-scale phylogeny with species representing all orchid subfamilies with 75 plastid genes and calibrated against 17 angiosperm fossils using BEAST v. 1.8.0 (Drummond *et al.*, 2012). Based on their estimates, we set a uniform distribution for the crown age of Orchidaceae with the lower and upper bounds equal to 79.7 Ma and 99.5 Ma, respectively, and used a Birth-Death tree model (Gernhard, 2008). The rest of the priors were not modified from their default values. Four identical runs with distinct seed numbers were performed simultaneously to determine whether they converged on the same stationary distribution. A fifth run was performed, sampling from the prior to examine whether the results were significantly skewed by the prior assumptions or informed by our data.

Convergence of the Markov Chain Monte Carlo (MCMC) chains was checked with Tracer v1.7.2 (Rambaut *et al.*, 2018) by checking that the Effective Sample Size (ESS) of the combined runs was >200 for all trace statistics, and that the trace plots of the individual runs converged on the same posterior distribution. The tree files from the four independent runs were combined after removing 10% as burn-in using LogCombiner v1.8.2, and the maximum clade credibility chronogram was reconstructed using TreeAnnotator v1.8.2 with (maximum clade median node height and 95% highest posterior density (HPD) intervals.

Detection of Diversification Rate Shifts

We investigated if diversification rates changed throughout the evolutionary history of *Cypripedium* and whether there were significant rate shifts. To achieve this, we used BAMM v2.5 (Rabosky *et al.,* 2013), a program developed to model the dynamics of speciation and extinction on phylogenetic trees. It considers time-dependent (e.g., a lineage's age) and diversity-dependent (e.g., the number of lineages in a clade) effects to quantify diversification rates using a reversible-jump MCMC approach.

Regarding the input tree used for this analysis, we modified the time-calibrated maximum clade credibility tree obtained from the divergence time estimation analysis by removing all non-*Cypripedium* species, as well as taxon duplicates, hybrids, and varieties to avoid inflating the diversification rates. To account for non-random taxon sampling between the included *Cypripedium* sections, section-specific sampling fractions were calculated based on the classification by Frosch and Cribb (2012). The expected number of shifts was set to one, following the recommendation for small trees with less than 500 tips. The priors on the initial lambda, the lambda shift parameter, and the time mode for the speciation rate were calculated with the R package BAMMtools v2.1.10 (lambdaInitPrior = muInitPrior = 1.35, lambdaShiftPrior = 0.05; Rabosky *et al.*, 2014), the segment length (segLength) was set to 0.1, and the rest of the parameters were left as default. We ran four MCMC chains for 50 million generations and sampled every 1,000 generations.

Subsequently, we used BAMMtools to check whether the MCMC runs converged (ESS >200) and discarded the first 25% of samples as burn-in. Then we estimated the prior and posterior distributions, plotted the speciation rate through time and the set of distinct shift configurations that account for 95% of the probability of the data (i.e., 95% credible set, threshold = 5),

checking which of these configurations had the maximum a posteriori (MAP) probability (aka best shift configuration).

Ancestral Range Estimation

We used the R package BioGeoBEARS (Matzke, 2013) to infer the biogeographic history of *Cypripedium*. BioGeoBEARS reconstructs the ancestral geographic distributions on phylogenies while testing for the best-fit model of range evolution. It replicates the basic assumptions of three widely used models in historical biogeography: DEC (Dispersal-Extinction-Cladogenesis; Ree and Smith, 2008), DIVA (Dispersal-Vicariance Analysis; Ronquist, 1997) and BayArea (Bayesian Inference of Historical Biogeography for Discrete Areas; Landis *et al.*, 2013), implementing them in a Maximum Likelihood framework to allow for direct comparison. Together, these models allow for a broad range of processes, such as vicariance, sympatric speciation, range expansion, and contraction. They can also be combined with a founder-event ("jump") speciation model specified with the parameter "j" (Matzke, 2014).

We conducted different BioGeoBEARS runs using DEC, DEC+J, DIVALIKE, DIVALIKE+J, BAYAREALIKE, and BAYAREALIKE+J to find the best-fitting model. Again, we used a modified version of the time-calibrated maximum clade credibility tree obtained from the divergence time estimation analysis for the input tree. However, this time, hybrids and duplicated species were removed, but when multiple varieties were present, a single specimen per each accepted variety, according to Frosch and Cribb (2012), was kept due to distinct distributions. *Cypripedium amesianum* Schltr. and the ambiguous *C. macranthos* var. *alba* were also kept since they were not monophyletic with their presumably synonymous taxa [i.e., *C. yunnanense* Franch. and *C. macranthos* var. *albiflorum* Makino (now synonym of *C. macranthos* Sw. var. *macranthos*), respectively], and therefore considered distinct taxonomic units for this analysis.

We also removed non-slipper orchid taxa because of scarce sampling. The taxon distributions were based on Eccarius (2009), Frosch and Cribb (2012), SC Chen *et al.* (2013), and Walid *et al.* (2019). The distributions of *C. amesianum* and *C. macranthos* var. *alba* were considered the same as their synonyms. A distance matrix was also used in the analysis to adjust the dispersal probabilities. The matrix included the distances between the closest points at the perimeters of every area pair combination in kilometers, measured in Google Maps. When two areas were adjacent, we set the distance between them to 1 km as recommended by the guidelines. To reduce computational time, we decided on nine total areas and allowed up to five areas to be combined in an ancestral range.

The areas were divided based on the current distribution of the taxa included in the analysis, their proximity, and their distinct floristic and topoclimatic characteristics (e.g., climate, precipitation, elevation). South America (area A) was specified as a large distinct area since only a few outgroup slipper orchids are restricted to its Northern part [i.e., *Selenipedium aequinoctiale* Garay and *Phragmipedium lindleyanum* (R. H. Schomb. ex Lindl.); POWO, 2023]. We separated Central America and Mexico (area B, containing section Irapeana) from South America at the Isthmus of Panama and from North America at the deserts to its north (i.e., Baja, Mojave, Sonoran, and Chihuahuan deserts). North America was divided into three areas: the Northern area E (colder to polar, humid climate), the Eastern area D (lower altitudes, higher humidity and precipitation than the Western area), and the Western area C (higher altitudes, lower humidity and precipitation than Eastern area; Kottek *et al.*, 2006; Schmidt, 2018; Xiao *et al.*, 2020). All three areas match the distributions of different *Cypripedium* species, with only five species found in area E, while the Great Plains in the middle of North America seemingly create a distribution boundary for multiple *Cypripedium* species (Supplementary Data Fig. S1). The Mediterranean and Scandinavian regions were grouped with Western and Central Europe (area F) because only

C. calceolus occurs in all three areas (Eccarius, 2009; Frosch and Cribb, 2012; SC Chen *et al.*, 2013; Walid *et al.*, 2019). We split area F from Eastern Europe and Russia (area G) at the boundaries of the Sarmatic and Pontic-South Siberian floristic provinces according to Schroeder (1998), as they have a more continental climate and match the limit of *C. guttatum's* distribution in the European continent (Pfadenhauer and Klötzli, 2020; Supplementary Data Fig. S1). Area G was separated from the two Asiatic areas (namely, the Southeast Asian area I and the Northeast Asian area H) due to the higher number of unique *Cypripedium* species occurring there, as well as their different climates and floristic provinces (Kottek *et al.*, 2006; Fridley, 2008). The Southeast and Northeast Asian areas are split around the Qinling Mountains–Huaihe River Line (aka Qinling–Huaihe line), a natural topographic boundary that separates North temperate from South tropical China (Y Hu *et al.*, 2020), which also seems to create a boundary for the distribution of several *Cypripedium* species. To reduce the state space and thus the computational time of the analysis, Taiwan was included in the same area as Southeast Asia (area I), and Japan in the same area as Northeast Asia and Eastern Russia (area H) due to their proximity.

In addition to these runs, a separate analysis with only two specified areas—the Old and the New World—was performed for comparison. In this case, the same six models were tested, setting the maximum number of areas to two, but a distance matrix was not included. Finally, the best-fit models of both analyses were plotted on the maximum clade credibility tree.

RESULTS

Assembly and Orthology Inference

The number of total extracted loci per species (with ≥75% and ≥50% of the target length and identity, respectively) ranged from 489 (*C. parviflorum* Salisb. var. *pubescens* (Willd.) O. W.

Knight, herbarium sample Nr. 69) to 906 (*C. irapeanum* La Llave & Lex., Oberhof sample Nr. 14) out of the 950 loci from the extended reference dataset. We recovered ~804 loci on average, which is notably higher than the corresponding proportion of loci reported for slipper orchids in the original publication by Eserman *et al.* [2021; i.e., 430 loci for *Paphiopedilum exul* (Ridl.) Rolfe and 533 loci for *Phragmipedium longifolium* (Warsz. & Rchb. f.) Rolfe]. Paralogs were found in all samples, from 57 (*C. irapeanum*, Oberhof sample Nr. 14) to 1,845 (*C. micranthum* Franch., Oberhof sample Nr. 22) with ~771 paralogs per sample on average. The orthology inference resulted in 792 MO orthologs available for ≥ 35 ingroup taxa. Further filtering for loci with at least 50 out of the 74 retained taxa and at least 500 sites per locus reduced the number of orthologs to 614. Unfortunately, most herbarium samples were removed as they generally yielded fewer loci with shorter length coverage than the living specimens. In total, 63 *Cypripedium* specimens representing 52 different taxa and 11 sections (Frosch and Cribb, 2012) were included in the final dataset (Fig. 1).

Inferred Species Phylogeny and Discordance

Both concatenation and coalescent-based phylogenetic reconstruction analyses recovered the genus *Cypripedium* as monophyletic with maximum support (i.e., BS = 100, LPP =1; Fig. 2; Supplementary Data Fig. S2). Additionally, both employed gene tree concordance analysis approaches for the ASTRAL species tree showed high concordance for the MRCA of *Cypripedium*, with Phyparts identifying 521 informative concordant genes out of 554 and QS giving full support (i.e., 1.0/–/1.0), indicating that all sampled quartet replicates support the node (Supplementary Data Fig. S3 and S4). The phylogenetic relationships among the slipper orchid genera were congruent between the ASTRAL and IQ-TREE trees, with *Cypripedium* being the sister to the rest. Within its sister clade, the plicate-leaved genus *Selenipedium* Rchb. f. was

recovered as sister to the clade of the conduplicate-leaved genera *Mexipedium* V. A. Albert & M. W. Chase, *Phragmipedium* Rolfe, and *Paphiopedilum*, with the two New World genera *Mexipedium* and *Phragmipedium* more closely related to each other than to the Old World *Paphiopedilum*.

All 11 sections were monophyletic within *Cypripedium*. On the other hand, the subsections *Cypripedium* and *Macrantha* within sect. *Cypripedium* were non-monophyletic based on the classification of Frosch and Cribb (2012). However, when considering the classification of SC Chen *et al.* (2013), subsect. *Macrantha* was monophyletic whereas subsect. *Cypripedium* was paraphyletic, and these results were consistent between the two phylogenies. Although the species grouped in the two largest sister clades within sect. *Cypripedium* (clades I and II) did not correspond to the species compositions of its two subsections, they matched the distribution of the species within them, with clade I only found in the New World and clade II in the Old World (Supplementary Data Fig. S1).

Maximum support was recovered for most branches in both inferred phylogenies, except for some branches along the backbone and within the sections *Trigonopedia*, *Sinopedilum*, *Bifolia*, and *Cypripedium*. Among the topologies with maximum support, high concordance, and congruence between the two phylogenies was the placement of the Mesoamerican sect. *Irapeana* being sister to the rest (LPP = 1; BS = 100; Phyparts: 521/554; QS: 1/–/1), followed by sect. *Arietinum* (LPP = 1; BS =100; Phyparts: 454/569; QS: 1/–/1). Additionally, sect. *Subtropica*, following the classification by Frosch and Cribb (2012), was recovered as monophyletic (LPP = 1; BS = 100; Phyparts: 431/550; QS: 1/–/1), including both sect. *Wardiana* and sect. *Subtropica* as described by SC Chen *et al.* (2013). The two fly-pollinated sections *Sinopedilum* and *Trigonopedia* were also supported as most closely related to each other, which was congruent

between the gene trees and between the ASTRAL and IQ-TREE trees (LPP = 1; BS = 100; Phyparts: 489/571; QS: 1/–/1).

Nevertheless, the remaining inter-sectional relationships showed decreased gene tree concordance based on Phyparts and QS. For instance, although both inferred phylogenies supported the sister relationship between sect. *Subtropica* and the clade containing the sections *Californica* Z. J. Liu, P. J. Cribb & Y. B. Luo and *Flabellinervia* (LPP = 1; BS = 100), both gene tree concordance analyses indicated elevated levels of discordance for these nodes (Phyparts: 68/531; QS: 0.15/0.77/0.99). The gene tree disagreement regarding the rest of the section-level relationships is largely attributed to the five short-length backbone (Fig. 3a; branches 3 to 7; Supplementary Data Fig. S5), with the proportion of informative concordant genes estimated by Phyparts falling between 2.4–22%, and QS indicating either weak or counter support for these branches. This is likely to be the cause of the mismatched topology of sect. *Acaulia* (Lindl.) C. Morren between the ASTRAL and the IQ-TREE trees, as it was recovered as sister to the (*Obtusipetala*, (*Bifolia*, *Cypripedium*)) clade in the former and sister to the (*Sinopedilum*, *Trigonopedia*) clade in the latter.

Gene tree heterogeneity was higher for nodes within sections than between sections, except for the sect. *Cypripedium* (LPP = 1; BS = 100; Phyparts: 455/568; QS: 1/–/1). Although this section was highly supported as monophyletic with high concordance, the number of informative concordant genes was lower than the number of discordant genes for almost all nodes within it. Only two nodes within this clade showed substantial concordance; namely, the (*C. henryi* Rolfe, *C. segawai* Masam.) clade (LPP = 1; BS =100; Phyparts: 321/383; QS: 1/–/1) and the node corresponding to the MRCA of clade II within the section (LPP = 1; BS =100; Phyparts: 314/501; QS: 1/–/1).

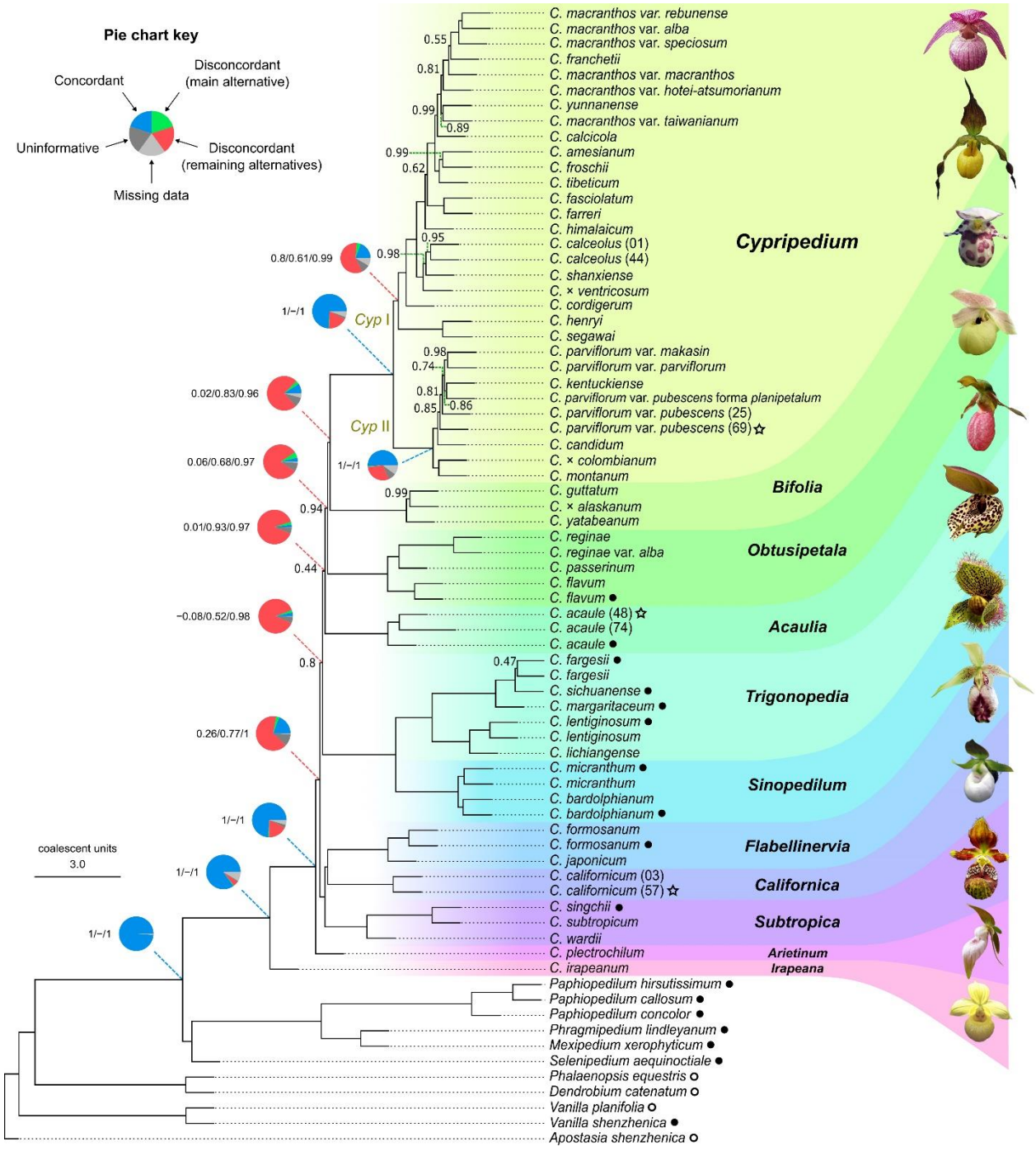

Figure 2: The coalescence-based phylogeny of the genus *Cypripedium*, inferred from 614 nuclear loci using ASTRAL. Local posterior probabilities are shown above the branches when <1. Branches are annotated according to their section-level classification following Frosch and Cribb (2012). Quartet Sampling support results (i.e., Quartet Concordance/Quartet Differential/Quartet Informativeness, in the same order) are shown to the left of the Phyparts piecharts for the backbone nodes, as well as the nodes leading to the two large clades within sect. *Cypripedium*, "*Cyp* I" and "*Cyp* II". Colors in the pie charts: blue denotes the proportion of concordant gene tree topologies, green denotes the proportion of gene trees with the main alternative topology, red denotes the proportion of gene trees with

the remaining discordant topologies, light grey denotes the proportion of gene trees with missing taxa, dark grey denotes the proportion of uninformative gene trees. Tip symbols: filled circles "●" denote transcriptomes, unfilled circles "○" denote genomes, unfilled stars "☆" denote specimens from herbarium M. Tips without symbols come from living specimens of the Botanical Collection at Oberhof. Flowers of representative species from each section are displayed to the right. Species names corresponding to the flower pictures, from top to bottom: *C. franchetii*, *C. parviflorum* var. *pubescens*, *C. guttatum*, *C. flavum*, *C. acaule*, *C. lentiginosum*, *C. micranthum*, *C. japonicum*, *C. californicum*, *C. subtropicum*, *C. plectrochilum*, and *C. irapeanum*. See the legends of **Figure 1** and the **Acknowledgements** section for the credits of the flower pictures.

Regarding the monophyly at the species level, two species within sect. *Cypripedium* that included numerous infraspecific taxa (i.e., varieties and one form), *C. macranthos* and *C. parviflorum* Salisb., were consistently recovered as paraphyletic in both phylogenies, with *C. franchetii* Wilson and *C. yunnanense* nested in the former, and *C. kentuckiense* C. F. Reed nested in the latter. Two pairs of synonyms (according to Frosch and Cribb, 2012) formed monophyletic clades: that is, (*C. parviflorum* Salisb. var. *makasin* (Farw.) C. J. Sheviak, *C. parviflorum* Salisb. var. *parviflorum*), and (*C. subtropicum* S. C. Chen & K. Y. Lang, *C. singchii* Z. J. Liu & L. J. Chen). However, *C. amesianum* was more closely related to *C. froschii* rather than its synonymous species, *C. yunnanense*. Additionally, the ambiguous taxon *C. macranthos* var. *alba*, which was presumed to be either *C. macranthos* var. *albiflorum* Makino (now a synonym of *C. macranthos* var. *macranthos*) or *C. macranthos* var. *album* Mandl, was more closely related to the equally white-flowered *C. macranthos* var. *rebunense* (Kudo) Ohwi. Regarding the rest of the species, only ASTRAL recovered them as monophyletic, grouping all conspecific specimens retrieved from different sources (i.e., Botanical Collection at Oberhof, herbarium M, and SRA). In contrast, IQ-TREE also produced non-monophyletic groupings of *C. micranthum* and *C. bardolphianum* W. W. Sm. & Farrer within sect. *Sinopedilum*, which matched their sequence type (i.e., transcriptome or target enrichment data). Notably, the three included hybrids following

Frosch and Cribb (2012) were placed in the same clades as one or both of their putative parent taxa.

Anomaly Zone and Polytomy Test

The anomaly zone boundary estimations detected four pairs of nodes at the backbone of *Cypripedium* that are in the anomaly zone [i.e., $y < \alpha(x)$], as well as the internode between the MRCAs of the (*Subtropica*, (*Californica*, *Flabellinervia*)) clade and the (*Californica*, *Flabellinervia*) clade (Fig. 3a; Supplementary Data Table S6 and Fig. S5). Additionally, several internode pairs within the two clades of sect. *Cypripedium* also fell into the anomaly zone (Supplementary Data Table S6 and Fig. S5). This suggests that most of the gene tree incongruence at these nodes could be explained by ILS. Interestingly, the nodes between the MRCA of sect. *Cypripedium* and the MRCAs of its subclades I and II were not found in the anomaly zone despite elevated gene tree heterogeneity at these nodes, especially for the former subclade, which could indicate that other factors are playing a role in this discordance, such as hybridization.

The ASTRAL polytomy test failed to reject the null hypothesis that the branch length is 0 (i.e., p-value > α; α = 0.05) for branches 5 and 6 (Fig. 3b). These branches were also found to be in the anomaly zone, while branch 5 also received a low LPP support (LPP = 0.44). This could explain the topology of sect. *Acaulia*, which varies between the ASTRAL and the IQ-TREE phylogenies, switching places within the potential polytomy. Additionally, branches within the subclades I and II of sect. *Cypripedium* found in the anomaly zone were also identified as soft polytomies (Supplementary Data Fig. S6).

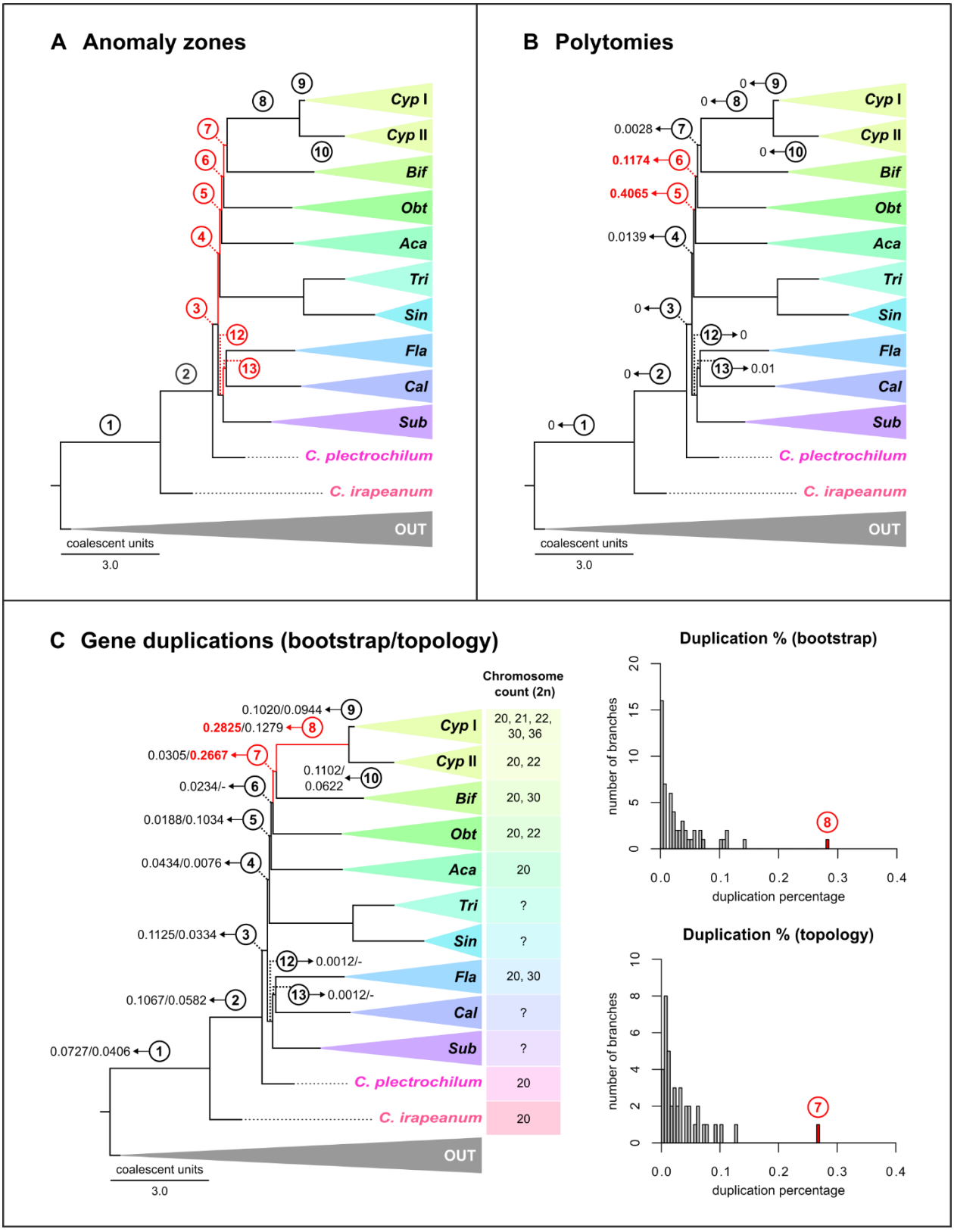

Figure 3: Results of the anomaly zone, polytomy, and gene duplication analyses annotated on the backbone of the collapsed ASTRAL species tree (branch numbers are shown in circles). (**A**) **Anomaly zone test**: Branches shown in

red are in the anomaly zone. (**B**) **Polytomy test**: Resulting p-values are annotated on the backbone branches. Polytomies based on α = 0.05 are shown in red. (**C**) **Gene duplications**: (Left) Percentages of duplicated genes are annotated on the backbone nodes based on the two used filtering methods (i.e., min 50% bootstrap and local topology). Branches with potential WGDs are shown in red (outlier duplication percentages > 20%). Known diploid chromosome counts from species within each clade are shown to the right of the collapsed clades, with "?" marking sections with unknown chromosome numbers [based on SC Chen *et al.* (2013) and Eccarius (2009)]. (Right) The number of branches is plotted against the gene duplication percentages per filtering method. Outlier gene duplication percentages correspond to branches 7 and 8 with potential WGDs. Key to collapsed clades: *Cyp* I = clade I of sect. *Cypripedium* (see **Fig. 2**); *Cyp* II = clade II of sect. *Cypripedium* (see **Fig. 2**); *Bif* = *Bifolia*; *Obt* = *Obtusipetala*; *Aca* = *Acaulia*; *Tri* = *Trigonopedia*; *Sin* = *Sinopedilum*; *Fla* = *Flabellinervia*; *Cal* = *Californica*; *Sub* = *Subtropica*. *C. plectrochilum* and *C. irapeanum* represent sections *Arietinum* and *Irapeana*, respectively. OUT = Outgroups.

Gene Duplication and Whole Genome Duplication Events

We mapped gene duplications on the ASTRAL species tree using two filtering methods (i.e., bootstrap and local topology filters; Fig. 3c; Supplementary Data Fig. S7). Most percentages of duplicated genes reached a maximum of ~14%, from 0.12–14.3% using the bootstrap filter and from 0.2–12.79% using the local topology filter. However, an outlier percentage of elevated gene duplications was identified by both methods (28.25% by the bootstrap filter and 26.67% by the local topology filter), although it was mapped to two different nodes, namely, the MRCA of sect. *Cypripedium*, or on the MRCA of the (*Cypripedium*, *Bifolia*) clade. The observed outliers could be caused by a WGD event, which could be further supported by the higher 2n chromosome number count found in species from both sections compared to other sections (i.e., > 20 chromosomes).

Hybridization Networks

In our PhyloNet analysis, we tested for a maximum of one reticulation event for the phylogenetic network containing *C.* × *alaskanum*, with the "po" option enabled, one to ten events for the network with *C.* × *columbianum*, and one to nine events for the network with *C.* × *ventricosum* as

the run for ten events has proven to be overly time-consuming. Each test produced ten optimal output networks, and we plotted the total log probabilities of the most likely network from each run for comparison (Supplementary Data Fig. S8).

To address our first question of whether the three known hybrid species are indeed supported as products of hybridization between their putative parent taxa by our molecular data, we plotted the most likely phylogenetic networks that tested for one reticulation event (Fig. 4). Regarding the test with *C.* × *alaskanum*, the most likely network indicated one hybridization event (Fig. 4a; total log probability ~ -550.38). However, our results suggested that *C. guttatum* is most probably a hybrid between *C.* × *alaskanum* (γ = 0.84) and an unsampled or extinct species that shares a common ancestor with the (*C.* × *alaskanum*, *C. yatabeanum*) clade (γ = 0.16). The most likely hybridization network of *C.* × *columbianum* on the other hand (total log probability = ~ -17545.08) supports its status as a hybrid between *C. parviflorum* var. *pubescens* and *C. montanum* as described in Frosch and Cribb (2012), and SC Chen *et al.* (2013), with relatively even inheritance probabilities (i.e., γ = 0.48 and 0.52, respectively; Fig. 4b). As for the network that includes *C.* × *ventricosum* (total log probability = ~ -88871.73), the results indicate that both *C.* × *ventricosum* and *C. calceolus* are sister taxa that arose from a hybridization event between *C. macranthos* var. *rebunense* (γ = 40) and *C. shanxiense* S. C. Chen (γ = 60; Fig. 4c).

To address our second question of whether other hybrids could be detected within the *Cypripedium* section, we plotted the most likely overall phylogenetic network from all runs that tested from one to nine or ten reticulation events for the networks including *C.* × *ventricosum* and *C.* × *columbianum*, respectively (Supplementary Data Fig. S8). The analyses that tested for up to nine hybridization events produced phylogenetic networks with the highest total log probabilities for both tests, identifying nine reticulation events for the network with *C.* × *ventricosum* (total log probability = ~ -88508.91) and seven reticulation events for the network with *C.* × *columbianum*

(total log probability = ~ -17501.44; Supplementary Data Fig. S8d and S8e). These phylogenetic networks indicate that in addition to *C.* × *columbianum*, *C.* × *ventricosum*, and *C. calceolus*, which were already identified as hybrids in our previous tests, *C. froschii*, *C. calcicola* Schltr., *C. amesianum*, *C. shanxiense*, *C. macranthos* var. *macranthos*, *C. franchetii*, *C. candidum* Muehl. ex Willd., *C. kentuckiense*, all sampled taxa of *C. parviflorum*, and some unsampled ancestral species also constitute potential hybrids (Supplementary Data Fig. S8a and S8b).

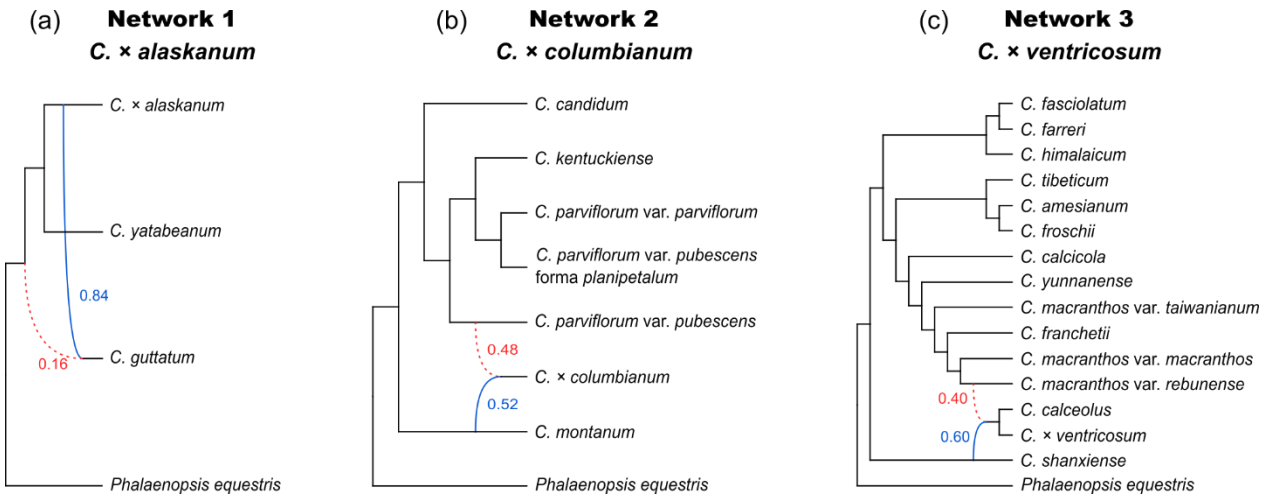

Figure 4: Phylogenetic networks with the highest total log probabilities resulting from the PhyloNet analysis testing for one hybridization event for the extracted subclades with each of the three hybrids: (**A**) *C.* × *alaskanum*, (**B**) *C.* × *columbianum*, (**C**) *C.* × *ventricosum*. The branch lengths and inheritance probabilities of Network 1 were optimized under full likelihood (option "po"). The inheritance probabilities are shown for each parent hybrid edge (blue, solid = major hybrid edge; red, dotted = minor hybrid edge).

Divergence Times

The time-calibrated maximum clade credibility tree, which was produced with BEAST 2 using 20 nuclear genes amounting to 59,202 sites, supported that the subfamily Cypripedioideae diverged from the rest of the slipper orchids close to the K-Pg boundary (64.84 Ma; 95% HPD 89.09–41.18 Ma) while genus *Cypripedium* split from the rest of the slipper orchids in the

Oligocene (30.26 Ma; 95% HPD 46.05–17.24 Ma). The Mesoamerican section *Irapeana* was the first to diverge within the genus, originating in the Early Miocene (20.27 Ma; 95% HPD 31.63–11.57 Ma). Following this split, rapid diversification occurred in the middle Miocene (15.93–13.74 Ma), giving rise to most *Cypripedium* sections or lineages from which the sections diverged (15.93–9.29 Ma). Section *Cypripedium* bifurcated during the late Miocene (8.52 Ma; 95% HPD 13.29–4.78 Ma), producing its two subclades.

Diversification Rates

The BAMM analysis illustrated a pattern of an initial elevated net diversification rate during the early stages of *Cypripedium*'s evolution, which steadily declined through time (from 0.23 to 0.13) due to decreasing speciation rate (from 0.28 to 0.17; Fig. 5a). No significant rate shifts have been detected in the maximum a posteriori probability shift configuration (f = 0.9; Fig. 5b; Supplementary Data Fig. S10), suggesting that a single macroevolutionary rate better explains the diversification within *Cypripedium* over time. The second most frequently sampled configuration in the 95% credible set indicated one significant rate shift on the MRCA branch of the sister clade of sect. *Irapeana* (f = 0.1), with speciation rate increasing sharply and slowing down towards the tips. In both cases, the analysis showed that the internode pairs previously identified to be in the anomaly zone indeed had higher diversification rates compared to the rate at later time points in the phylogeny, providing support to the hypothesis of ILS playing a role in their elevated levels of discordance.

Historical Biogeography

Our comparison of biogeographic models in BioGeoBEARS showed that allowing founder-event speciation (parameter "j") provided the best fit to our data. Namely, the DEC+J model (d =

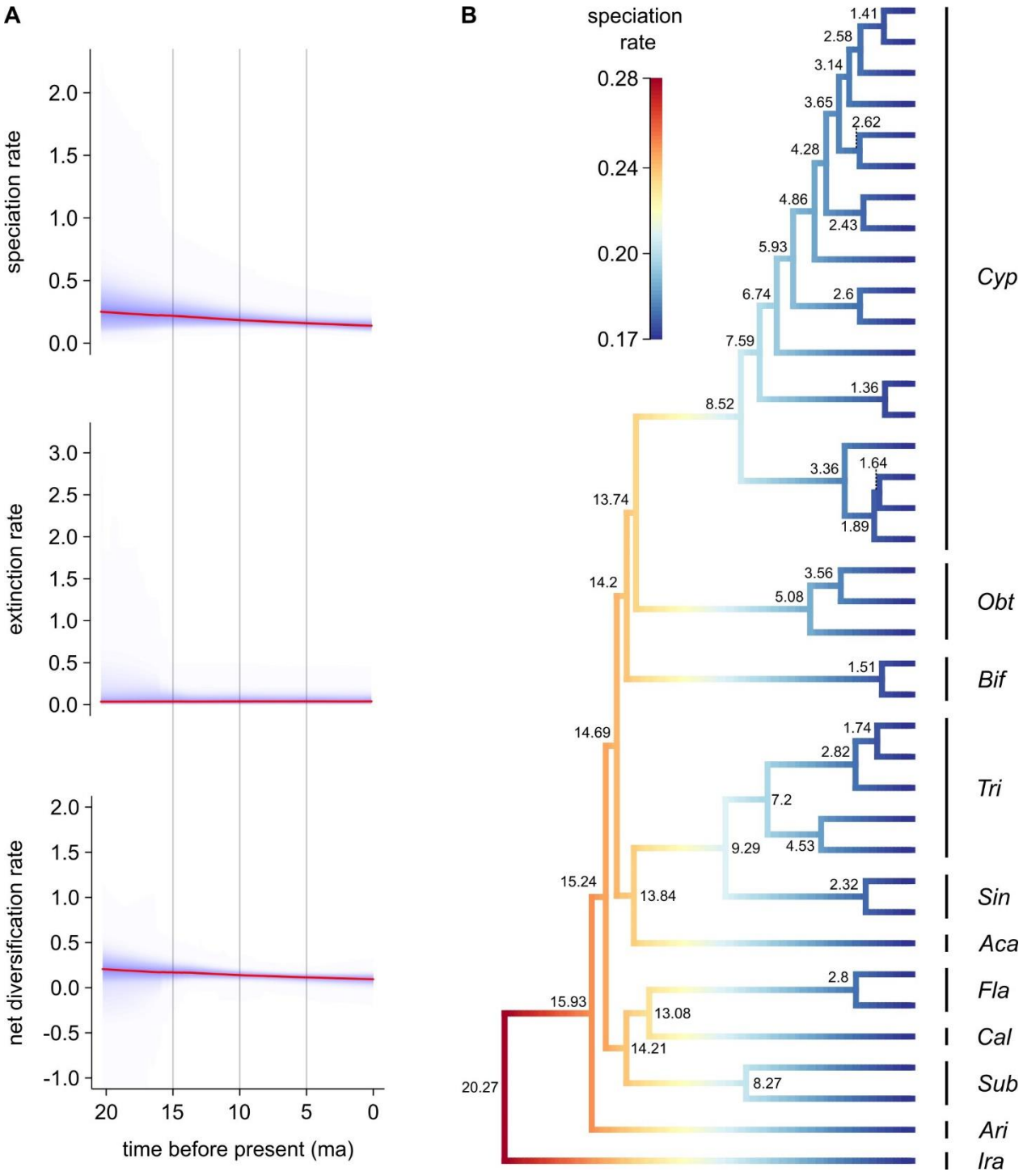

Figure 5: Results of BAMM analysis in *Cypripedium*. (**A**) Rate variations in (top) speciation, (middle) extinction, and (bottom) net diversification through time, based on all samples in the posterior distribution (density shading on confidence regions). (**B**) Maximum a posteriori probability shift configuration represented as a phylorate plot showing variations in speciation rates (cooler colors = slow, warmer colors = fast) along each branch of the dated *Cypripedium* phylogeny (posterior median node height estimates of the divergence times in Ma are shown on the nodes). The clades are annotated with the first three letters of the name of each section.

0.0201; e = 0; x = −0.2711; j = 0.0299; LnL = −119.27) had the lowest AIC (246.5) and AICc (247.4) scores for the first test with nine defined areas, while the BAYAREALIKE+J model (d = 0.0026; e = 0; j = 0.0747; LnL = −25.97) had the lowest AIC (57.94) and AICc (58.44) scores for the second test with the New and Old World areas. The relative probabilities of the ancestral geographical ranges are illustrated with pie charts in Supplementary Data Fig. S11 and S12 for the first and second test, respectively, while Fig. 6 combines the single most probable ancestral ranges from both analyses with the results of the latter shown only for nodes where they disagree. Regarding the results of the first test, the DEC+J model indicated that the ancestors of the Cypripedioideae and *Cypripedium* clades were more widespread, distributed across the Old and the New World (ranges ABI and BI, respectively; Fig. 6) and that potential long-distance dispersals took place when *Cypripedium* diverged from the rest of the slipper orchids in the Oligocene, as well as when the Mesoamerican sect. *Irapeana* split from the Southeastern ancestor of its sister clade in the Miocene. On the other hand, the BAYAREALIKE+J model implemented in the second test supported the New World being the ancestral range of both Cypripedioideae and *Cypripedium*, with the ancestor of sect. *Irapeana*'s sister clade speciating after its dispersal to the Old World. In both cases, the models suggested that the sister clade of sect. *Irapeana* rapidly diversified in the Old World in the Middle Miocene, specifically in Southeast Asia (i.e., area "I"), where most *Cypripedium* species are still found today.

Many of the lineages produced during these rapid diversification events dispersed and speciated in other Old World regions, such as in Northeast Asia and the nearby islands of Japan and Taiwan (e.g., *C. japonicum* Thunb. and *C. formosanum* Hayata). There were also multiple independent dispersals back to the New World between the Miocene and the Pliocene (e.g., sect. *Acaulia*, sect. *Californica*; and MRCA of *C. reginae* Walter and *C. passerinum* Richardson), while the MRCA of sect. *Bifolia* spread both Eastwards and Westwards, acquiring a wide

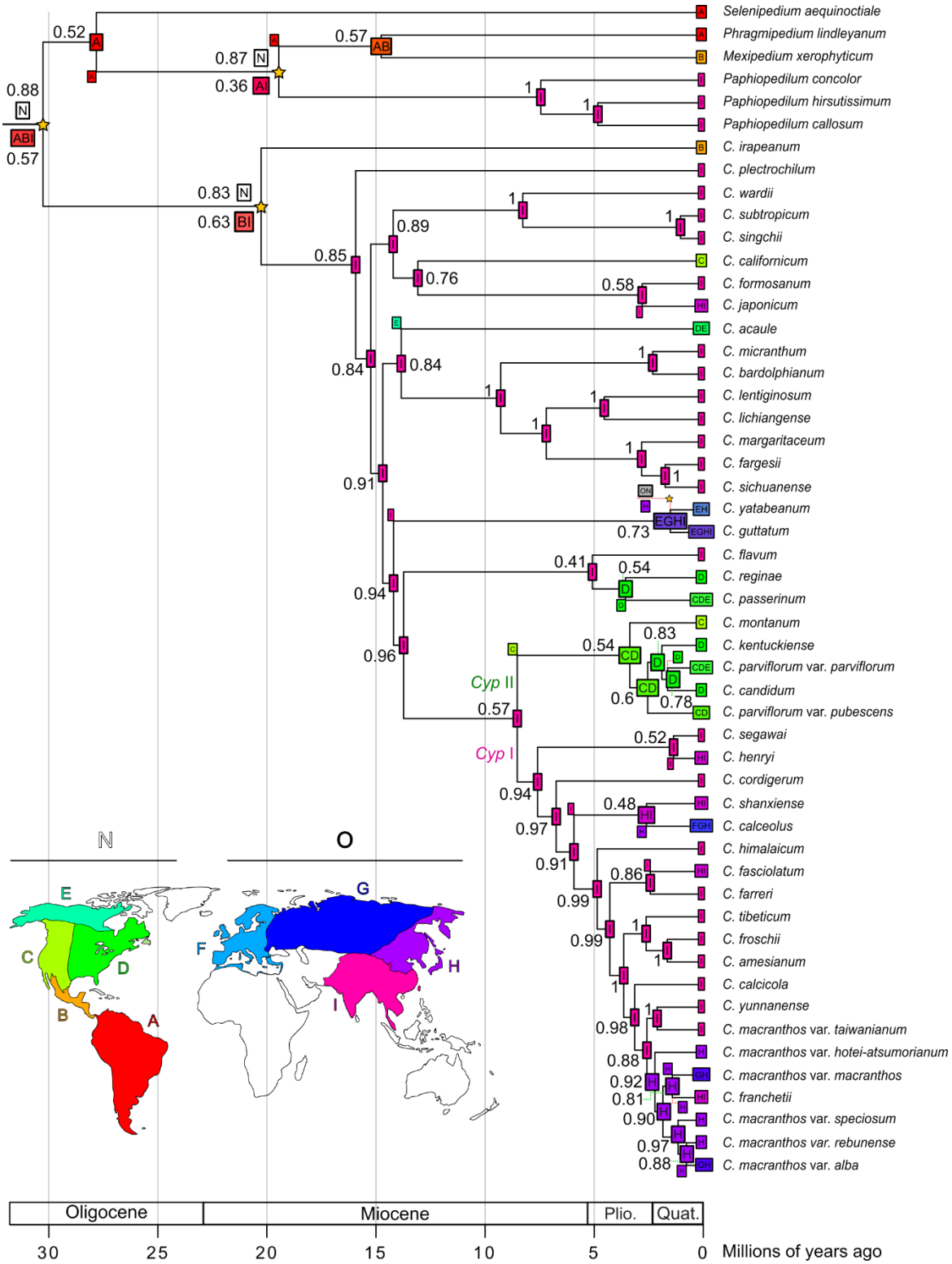

Figure 6: Estimations of ancestral ranges with the highest likelihood, plotted on the dated maximum clade credibility tree of slipper orchids and their state probabilities at the nodes. Results from both BioGeoBEARS runs (i.e., run with nine areas and with only New and Old World areas) are shown together when they disagree (below and above the

branch, respectively; the corresponding nodes/corners are marked with yellow stars); otherwise, only the results of the former test are shown. Distribution on the nodes and branch corners are right before and right after cladogenesis, respectively (the latter is not shown if it shares the same distribution with the following node). Key to area codes: A = South America; B = Central America and Mexico; C = Western North America; D = Eastern North America; E = Northern North America; F = Western and Central Europe, the Mediterranean, and Scandinavia; G = Eastern Europe and Eurasia; H = Eastern Russia and Northeast Asia; I = Southeast Asia; N = New World; O = Old World. Plio. = Pliocene; Quat. = Quaternary. *Cyp* I = clade I of sect. *Cypripedium* (see **Fig. 2**); *Cyp* II = clade II of sect. *Cypripedium* (see **Fig. 2**).

distribution in both the Old (i.e., Eastern Europe, Eurasia, Northeast Asia, Japan, East Russia, Southeast Asia) and the New World (i.e., Alaska), with the two species evolving sympatrically in the broader sense.

The clades I and II of sect. *Cypripedium* diverged with the long-distance dispersal of the MRCA of clade II to the New World during the Late Miocene. Following this, the two clades diversified allopatrically, with subclade II spreading throughout North America, evolving several closely related taxa within the *C. parviflorum* complex. Within subclade I, some lineages and species expanded their distribution from Southeast Asia to the adjacent area of continental Northeast Asia [e.g., *C. henryi*; MRCA of (*C. shanxiense*, *C. calceolus*); *C. fasciolatum* Franch.] as well as the neighboring island of Taiwan [e.g., *C. segawai*, *C. macranthos* var. *taiwanianum* (Masam.) Maekwa]. *C. shanxiense* also spread to the island of Japan, while *C. calceolus* speciated in Northeast Asia, establishing populations in large latitudinal ranges as the only known species in Western Europe, Scandinavia, and the Mediterranean. The ancestor of the sister clade of (*C. yunnanense*, *C. macranthos* var. *taiwanianum*) switched its distribution to Northeast Asia, from where it probably further dispersed to Japan, producing multiple endemic varieties [e.g., *C. macranthos* var. *rebunense*, *C. macranthos* var. *speciosum* (Rolfe) Koidz., and *C. macranthos* var. *hotei-atsumorianum* Sadovsky] while *C. macranthos* var. *macranthos* also expanded to East, Asiatic, and European Russia.

DISCUSSION

Monophyly and Topology of Established Infrageneric Taxa

In the present study, we reconstructed the first robust phylogeny of the genus *Cypripedium* based on high-throughput target enrichment data of 614 nuclear loci using the "Orchidaceae963" bait set (Eserman *et al.,* 2021). The inferred phylogenetic tree showed that *Cypripedium* is sister to the clade of the other four slipper orchid genera, (*Selenipedium*, (*Paphiopedilum*, (*Phragmipedium*, *Mexipedium*))), which agrees with the topologies recovered by the supra-generic phylogenies of Guo *et al.* (2012), Wong and Peakall (2022) and Pérez-Escobar *et al.* (2023) based on Sanger sequences (cpDNA and nDNA markers), transcriptomic data, and a combination of target enrichment (low-copy nuclear loci) and Sanger sequences (matK and ITS), respectively. All sections were monophyletic based on the classification system proposed by Frosch and Cribb (2012) and SC Chen *et al.* (2013). However, subsect. *Macrantha*, which included the clade of *C. farreri* W. W. Sm. and *C. fasciolatum* as proposed by SC Chen *et al.* (2013), was nested within subsect. *Cypripedium*.

Although the incongruence between the taxonomy and the monophyly of these subsections may be affected by the elevated gene tree discordance within sect. *Cypripedium*, our results match the findings of Szlachetko *et al.* (2021), who also evaluated the monophyly of these subsections based on the same two classification systems. Indeed, most published *Cypripedium* phylogenies supported that one or both subsections may be non-monophyletic and that the (*C. farreri*, *C. fasciolatum*) clade is more closely related to subsect. *Macrantha* rather than subsect. *Cypripedium* (Fatihah *et al.*, 2011; J Li *et al.*, 2011; H Liu *et al.*, 2021*a*), as suggested by SC Chen *et al.* (2013). Moreover, the species composition of the two clades within sect. *Cypripedium*

in our phylogeny matched their distributions, with clades I and II only found in the Old World and the New World, respectively. A similar trend was observed in the phylogenies by Fatihah *et al.* (2011), J Li *et al.* (2011), and Szlachetko *et al.* (2021), where the division between the two groups of species seemed to be based on their distribution rather than the traditionally used morphological characteristics (i.e., the floral coloration and the shape of the labellum and the lateral petals).

Within these two subclades, we recovered the two morphologically diverse species, *C. parviflorum* and *C. macranthos*, as paraphyletic. Other authors also found that the latter is paraphyletic, with *C. kentuckiense* embedded in the clade, similar to our phylogeny (Fatihah *et al.*, 2011; J Li *et al.*, 2011; H Liu *et al.*, 2021*a*; Szlachetko *et al.*, 2021). However, even though *C. yunnanense* and *C. franchetii* were shown to be closely related to *C. macranthos*, with the latter recovered monophyletic (Fatihah *et al.*, 2011; J Li *et al.*, 2011; H Liu *et al.*, 2021*a*; Szlachetko *et al.*, 2021), our results showed that they are nested within the *C. macranthos* group, which could have resulted due to the inclusion of different *C. macranthos* varieties (i.e., *C. macranthos* var. *alba* and all five accepted natural varieties described by Frosch and Cribb, 2012). *Cypripedium parviflorum* and *C. macranthos* are widespread and morphologically variable species, traditionally distinguished mainly based on flower size and coloration (SC Chen *et al.*, 2013; Cribb, 1997). Due to high morphological variation within their infraspecific taxa, as well as the existence of intermediate forms, they have been historically difficult to classify. This difficulty could be attributed to the recent divergence of these varieties and forms within each species that could cause ILS, in addition to hybridization, gene duplication, and WGD events, all occurring within sect. *Cypripedium*, according to our analyses.

Besides sect. *Cypripedium*, several supported topologies with high gene tree concordance recovered in our phylogeny, are consistent with those in other studies. For instance, sect.

Irapeana, the only neotropical section of *Cypripedium*, was recovered as the sister to the rest in the majority of published molecular phylogenies (Cox *et al.*, 1997; Fatihah *et al.*, 2011; J Li *et al.*, 2011; Guo *et al.*, 2012; H Liu *et al.*, 2021*a*; Szlachetko *et al.*, 2021) as well as in the present study. The placement of sect. *Irapeana* is further supported by morphological and biogeographical data, as it is considered to share "ancestral" morphological features with the plicate-leaved genus *Selenipedium*, also found in the Neotropics (Cox *et al.*, 1997; Szlachetko *et al.*, 2021). Similarly to the results of Cox *et al.* (1997), Fatihah *et al.* (2011), J Li *et al.* (2011), and Szlachetko *et al.* (2021), sections *Arietinum* and *Irapeana* form a grade that is sister to the remaining *Cypripedium* included in our phylogeny. Furthermore, sect. *Trigonopedia* and sect. *Sinopedilum*, the two Southeast Asian fly-pollinated sections, previously grouped as a single section (Cribb, 1997), form a clade in our phylogeny. This sister relationship is consistently well-supported by other studies (Fatihah *et al.*, 2011; J Li *et al.*, 2011; H Liu *et al.*, 2021*a*; Szlachetko *et al.*, 2021), while the infra-sectional topology we recovered for *Trigonopedia* matches that of H Liu *et al.* (2021*a*).

Our molecular data also provides support for the classification of sect. *Subtropica* following Frosch and Cribb (2012), as well as sect. *Subtropica* and *Wardiana* following SC Chen *et al.* (2013). Specifically, two species have been described to belong to sect. *Subtropica* in the former monograph, namely, *C. subtropicum* and *C. wardii*, with *C. singchii* being a synonym of *C. subtropicum*. However, SC Chen *et al.* (2013) redefined *C. singchii* as a distinct species and transferred *C. wardii* to its own monotypic section, *Wardiana*. We showed that all three species share an MRCA based on our molecular data, with *C. subtropicum* and *C. singchii* being more closely related to each other than to *C. wardii*, matching the corresponding topology in J Li *et al.* (2011) and Szlachetko *et al.* (2021)

Concerning the remaining intersectional phylogenetic relationships, we uncovered high gene tree discordance at several backbone nodes. Low backbone branch support has been observed in all molecular phylogenies focusing on the section-level relationships within *Cypripedium* (Cox *et al.*, 1997; Fatihah *et al.*, 2011; J Li *et al.*, 2011; Szlachetko *et al.*, 2021), except H Liu *et al.* (2021*a*), where four plastid regions were used for phylogenetic reconstruction. Our results suggest that gene tree heterogeneity could explain why the evolutionary relationships between the majority of the sections within *Cypripedium* remain unresolved, with different sets of loci and phylogenetic inference methods producing incongruent topologies, such as the changing placement of sect. *Acaulia* between the concatenation- and coalescent-based phylogenies in our study.

Rapid Radiation, Whole Genome Duplication, and Hybridization Promoted Diversification

The concordance analyses indicated that most gene tree topologies disagree at certain nodes of our phylogeny. High levels of discordance were particularly observed at nodes along the backbone, the MRCA between sect. *Subtropica* and the (*Californica*, *Flabellinervia*) clade, and within the two subclades of sect. *Cypripedium*. Other authors (Fatihah *et al.*, 2011; J Li *et al.*, 2011; Szlachetko *et al.*, 2021) have previously speculated that potential rapid radiation events could interpret the low support at the branches along the backbone of their *Cypripedium* phylogenies, as well as the high morphological differentiation between the sections. Here, we provided supporting evidence for this hypothesis based on our analyses of ~600 nuclear loci. Specifically, our anomaly zone test showed that this incongruence largely owed to ILS caused by rapid radiation. This is further corroborated by the increased diversification rates at the corresponding nodes and their chronologically close placement at the geological time scale.

The only intra-sectional nodes detected in the anomaly zone belonged to sect. *Cypripedium*, indicating that rapid and recent diversification took place within its two subclades, leading to ILS and producing multiple closely related species and infra-specific taxa with high morphological variation (e.g., within the *C. macranthos* and *C. parviflorum* complexes). However, we showed that other factors could have also contributed to the incongruence between the gene and the species trees observed within this section and throughout the phylogeny. For instance, another source of mixed gene tree signals could be reticulation.

The heterogeneous gene topologies at the MRCA nodes of the two subclades I and II of sect. *Cypripedium* could be mainly explained by hybridization and WGD, as they have not been shown to fall within the anomaly zone, contrary to the nodes within these subclades. As suggested by Unruh *et al.* (2018), an increased taxon sampling in *Cypripedium*, which is the genus with the widest genome size range (4.1–43.1 pg/C) among all slipper orchids, revealed a potential WGD event. Specifically, we located a WGD either at the MRCA branch of sect. *Cypripedium*, or of the (*Cypripedium*, *Bifolia*) clade, and widespread gene duplications within the phylogeny, using the subclade orthogroup tree topology method. These factors could contribute to gene tree discordance (Górecki and Eulenstein, 2014) and diversification by producing paralogous genes with the potential to develop novel functions. WGDs, in particular, are considered one of the main driving forces of species-, and in some cases, niche-diversification, as polyploids may occupy new or wider ranges of niches in contrast to their diploid relatives (Dodsworth *et al.*, 2016; Ren *et al.*, 2018). Thus, a WGD event could have promoted the higher diversity observed within sect. *Cypripedium*, as well as the widespread distribution of both sect. *Cypripedium* and *Bifolia*, which contain species with the widest latitudinal (i.e., *C. guttatum*) and longitudinal (i.e., *C. calceolus*) ranges in the genus, found in a variety of different altitudes (i.e.,

50 to 4100 m and sea level to 2000 m, respectively; Frosch and Cribb, 2012; SC Chen *et al.*, 2013).

The time interval between the WGD and the diversification of sect. *Cypripedium* and sect. *Bifolia* could be interpreted by the 'WGD Radiation Lag-Time' model, where some lineages within a clade sharing an ancestral WGD expand millions of years following the event. This has been proposed to occur not only due to the time needed for the subfunctionalization or neofunctionalization but also for the right ecological conditions to uncover their adaptive advantage (e.g., migration or climatic changes) or for re-diploidization to occur, as neopolyploids are more likely to go extinct and less likely to speciate than diploids (Schranz *et al.*, 2012; Tank *et al.*, 2015; Dodsworth *et al.*, 2016; Robertson *et al.*, 2017). On the other hand, this lag time could also be a consequence of extinction events taking place between the divergence of these sections and their diversification, which we also observed in other sections of our phylogeny (e.g., sect. *Flabellinervia* and sect. *Obtusipetala*).

Additionally, the increased number of chromosomes within sections *Cypripedium* (2n = 20-36) and *Bifolia* (2n = 20-30), compared to the usual diploid count in the genus (2n = 20), provides further evidence to support the existence of a potential WGD at either of the identified branches (Eccarius, 2009; SC Chen *et al.*, 2013). In particular, the species *C. macranthos* has the widest range of recorded chromosome counts (i.e., 20, 21, 30, 36) in the genus. These numbers nearly match different ploidy levels, from the usual 2n (20) to less than 4n (36).

The elevated chromosome count in these two sections may have also resulted from allopolyploidization, as hybrids are observed in both clades. All nodes leading to the hybrids identified through our phylogenetic network analyses had decreased concordance levels, including the MRCA node of the (*C. guttatum*, *C.* × *alaskanum*) clade. Although *C.* × *alaskanum* has been previously described as a natural hybrid between *C. yatabeanum* and *C. guttatum* by

PM Brown (1995), it was difficult to assess its validity because no analysis accompanied the description (Cribb, 1997). We found that, most likely, *C. guttatum* is a hybrid between *C.* × *alaskanum* and an unsampled species sharing an MRCA with the (*C.* × *alaskanum*, *C. yatabeanum*) clade.

As for the hybrids within sect. *Cypripedium*, testing for a maximum of one hybridization event yielded both expected and unexpected results. Sheviak (1992) proposed that *C.* × *columbianum* is a hybrid between *C. montanum* and *C. parviflorum* var. *pubescens* based on an extensive survey and analysis of wild and herbarium specimens, with hybrids having intermediate morphological characteristics between the presumed parent taxa. The phylogenetic network containing *C.* × *columbianum* agreed with Sheviak's taxon description, with both proposed parent taxa having almost equal inheritance probabilities.

Nonetheless, the results of a similar test on the dataset containing *C.* × *ventricosum* supported that the taxon did not constitute a hybrid from a cross between *C. calceolus* and *C. macranthos* as it was previously suggested based on the resemblance of its flower with an artificial hybrid between the presumed parent taxa (Rolfe, 1904, 1910; Cribb, 1997). Instead, the results indicated that *C.* × *ventricosum* and *C. calceolus* are products of a hybridization event between *C. shanxiense* and *C. macranthos*. Another hybrid, *C.* × *catherinae*, which occurs in Far East Russia (Siberia) and possibly in Korea and Northeast China, where the parent taxa's distributions overlap, was already described from the hybridization between these two taxa (Frosch and Cribb, 2012; SC Chen *et al.*, 2013). *Cypripedium calceolus* and *C.* × *ventricosum* can also be found sympatrically with the parent taxa identified in our analysis, namely, in Far Eastern Russia (Siberia) and Northeastern China, while *C. calceolus* is also found in Japan and on Sakhalin Island (Cribb, 1997; Frosch and Cribb, 2012; SC Chen *et al.*, 2013). Interestingly, a sister relationship is consistently recovered between *C. calceolus* and *C. shanxiense* (Fatihah *et al.*,

2011; J Li *et al.*, 2011; H Liu *et al.*, 2021*a*; Szlachetko *et al.*, 2021) or between *C. calceolus* and *C. macranthos* in the absence of *C. shanxiense* (Cox *et al.*, 1997). Additionally, *C.* × *ventricosum* and *C. macranthos* formed a clade closely related to the (*C. calceolus*, *C. shanxiense*) clade in the phylogeny of H Liu *et al.* (2021*a*).

In support of Rolfe's description, a statistical analysis of morphological and allozyme data from *C.* × *ventricosum* corroborated its status as a hybrid between *C. calceolus* and *C. macranthos* (Knyasev *et al.*, 2000). However, the fact that *C. calceolus* and *C.* × *ventricosum* may share the same parent taxa based on our results (i.e., *C. shanxiense* and *C. macranthos*) and that introgressive hybridization has been previously reported between *C. calceolus* and *C.* × *ventricosum* (Knyasev *et al.*, 2000) could elucidate the relationship between these species. Furthermore, it has been shown that the genetic structure of *C. shanxiense* was similar to *C. calceolus* from the eastern part of the range where they are sympatric, with the authors arguing that it may have been a result of introgressive hybridization, making the classification of these taxa even more challenging (Filippov and Andronova, 2011). At the same time, a hybrid has also been described between these two species (i.e., *C.* × *microsaccos* Kraenzl; Frosch and Cribb, 2012; SC Chen *et al.*, 2013).

When looking at the most probable phylogenetic networks produced for our second test, the results suggested that extensive hybridization events between sampled and unsampled taxa occurred within both subclades of sect. *Cypripedium*. Specifically, the patterns of reticulate evolution in these networks indicated that the hybrids that persisted interbred with other taxa (incl. other hybrids) and backcrossed with their parent taxa, producing a hybrid swarm. The mixed genetic and morphological signals created by these extensive and overlapping reticulation events could have significantly contributed to the difficulty in specific delimitation and taxonomic classification within the sect. *Cypripedium*, with the same taxa receiving a species,

variety, or hybrid rank by different taxonomists. For example, a self-pollinating taxon that has been reported as a synonym of *C. yunnanense*, *C. amesianum* (Frosch and Cribb, 2012; SC Chen *et al.*, 2013; P Cribb, Royal Botanical Garden, Kew, UK, 'pers. comm.'), was more closely related to *C. froschii* and *C. tibeticum* King ex Rolfe in our phylogeny. Our results show that *C. amesianum* may constitute a hybrid between unsampled parent taxa, one of which is closely related to *C. tibeticum* and the other to *C. yunnanense*. Additionally, *C. froschii*, the nature of which is still strongly debated (e.g., variety of *C. tibeticum* according to Eccarius, 2009; species status according to Frosch and Cribb, 2012; *C. tibeticum* × *C. yunnanense* hybrid according to SC Chen *et al.*, 2013) seems also to be a result of hybridization between the latter parent taxon of *C. amesianum*, and a taxon within an unsampled sister clade of *C. franchetii*. However, an ancestor of the latter parent taxon also experienced hybridization between an unsampled taxon and *C. macranthos*.

Further studies are necessary to confirm the hybridization events that we identified in the evolutionary history of sections *Cypripedium* and *Bifolia*. Unfortunately, we could not test for reticulation events at a broader scope within *Cypripedium* as the phylogenetic network analysis becomes highly computationally intensive and time-consuming with increasing taxa and hybridization events. However, as hybridization is considered to be common in nature between sympatric *Cypripedium* species, we could assume that it might have also played a part in the discordance observed at other nodes of the genus' phylogeny.

Climatic Fluctuations Influenced Current Distribution and Diversity Patterns

Previous primary molecular calibrations placed the stem age of Cypripedioideae at 77 Ma (Givnish *et al.*, 2016) or between ca. ~75–82 Ma (based on the youngest and oldest fossil ages, respectively; Guo *et al.*, 2012). Our estimation places the divergence time of the slipper orchids

at a more recent time point around the K-Pg boundary (64.84 Ma), a time that was followed by rapid radiations of flowering plants at the genus and family level (e.g., Koenen *et al.*, 2021; Morales-Briones *et al.*, 2021), as they replaced some of the old lineages and filled empty niches left by the mass extinction (Berendse and Scheffer, 2009). Similarly to the fossil-assisted calibration of (Givnish *et al.*, 2016) and a secondary calibration based on the divergence time estimates above (H Liu *et al.*, 2021*a*), we dated the split of *Cypripedium* from its sister clade to ~30 Ma in the Oligocene.

During that time, according to our biogeographic analysis, the ancestors of the slipper orchids either had a wider distribution across the New and the Old World or were initially limited to the New World. Such uncertainty was also noticed in the genus-level Cypripedioideae phylogeny of Guo *et al.* (2012), who identified either the New World or the Old World + the New World as the ancestral regions of slipper orchids using both S-DIVA and Lagrange methods. A biogeographic analysis of a broader Orchidaceae phylogeny that included placeholders from all five orchid subfamilies (representing 18/19 tribes and 40/43 subtribes), all genera of slipper orchids, and 96 outgroup angiosperms suggested that the ancestors of Cypripedioideae were in the Neotropics with >75% probability. This result agreed with the analysis of H Liu *et al.* (2021*a*), whose estimates pinpointed Central America as the origin of slipper orchids despite a different topology recovered at the genus level. The more extensive taxon-sampling of outgroup taxa within non-*Cypripedium* slipper orchids, as well as other orchids and angiosperms in Givnish *et al.* (2016), would have likely better informed the estimations of the ancestral region of the subfamily Cypripedioideae compared to our outgroup sampling, therefore, providing support to our second biogeographic analysis indicating the New World as the MRCA.

A similar dichotomy was observed in the results of our two biogeographic tests regarding the most likely ancestral regions of the MRCA of *Cypripedium* following its divergence from the rest

of the Cypripedioideae (i.e., either Central America + Southeast Asia or the New World). Again, the results of Guo *et al.* (2012) resembled ours at the corresponding node, which either recognized both the Old World + the New World (S-DIVA) or the Old or New World (Lagrange) as the ancestral regions of *Cypripedium*. This ambivalence was also reflected in the ancestral region estimations of Givnish *et al.* (2016), supporting that Neotropics + Eurasia was the most probable distribution of the ancestor of *Cypripedium*, followed by the Neotropics. Contrarily, H Liu *et al.* (2021*a*) found that the most probable ancestral area of slipper orchids was Central America, while the second most probable was Central America + South Asia. Thus, despite evidence favoring both scenarios, a wider ancestral distribution across the Old and the New World received higher support in most studies. However, it is important to note that due to incomplete taxon sampling (i.e., 11 out of 13 *Cypripedium* sections following Frosch & Cribb, 2012, and few outgroups) as well as the discordance in the backbone of the phylogeny, these results should be interpreted with caution.

A wider ancestral distribution could have been facilitated by the Bering Land Bridge acting as a dispersal corridor between the New and the Old World, as it was intermittently exposed throughout most of the Cenozoic (Hopkins, 1959). Additionally, a significantly warmer climate during the Early to Middle Miocene Climatic Optimum (MMCO; ca. 17–15 Ma), especially at higher latitudes where mixed deciduous-evergreen forests dominated the vegetation type (Dutton and Barron, 1997; Herold *et al.*, 2010; Steinthorsdottir *et al.*, 2021), coupled with adaptations to colder climates after diverging from the neotropical sister clade (H Liu *et al.*, 2021*a*), could have allowed ancestors of the genus *Cypripedium* to spread northwards and disperse via the Bering Land Bridge or over the strait. The current distribution of sect. *Bifolia* at both sides of the Bering Strait (i.e., Eastern Siberia and Alaska) supports the potential of *Cypripedium* to disperse via this corridor.

As global cooling took place following the MMCO (Frigola *et al.*, 2018; Methner *et al.*, 2020; RM Brown *et al.*, 2022), a climatic dispersal barrier may have formed between the Old and the New World at higher latitudes, leading to a vicariance and allopatric diversification. The East Asian monsoon system, whose development was associated with the global cooling and the uplift of the Tibetan Plateau in the Early Miocene, may have promoted *Cypripedium*'s dispersal to South Asia (H Lu and Guo, 2014; Hui *et al.*, 2021; Ye *et al.*, 2022). The Middle Miocene Climate Transition (MMCT; 15–13 Ma), a period of climate transition between the MMCO and the Middle Miocene Glaciation (MMG) associated with major Antarctic ice sheet expansion and global cooling (Frigola *et al.*, 2018), coincided with the rapid radiation of *Cypripedium* in Southeast Asia during the Middle Miocene (~16–14 Ma), according to our results. Interestingly, our dating of the divergence time between sect. *Irapeana* and its sister clade (~21 Ma) coincides with the stem ages of most seed plant clades originating in the Sino-Japanese floristic region (Y-S Chen *et al.*, 2018), while the identified period of the rapid diversification events in the mountains of China (~16–14 Ma) fits the observed trend of large-scale plant diversifications in W. China during the Miocene (L-M Lu *et al.*, 2018).

After dispersing to Southeast Asia, the time interval between the origin and the diversification of this new clade (i.e., stem age: 20.27 Ma; crown age: 15.93 Ma) indicates that extinctions might have taken place, which could explain why the ancestors of this lineage were limited to that region according to our analysis. The southeastern margins of the Tibetan Plateau may have provided potential refugia to *Cypripedium* from the cooling climate of the MMCT due to their warmer and more humid conditions owing to the intensification of the Asian monsoons (Xing and Ree, 2017; S Li *et al.*, 2020; Zuo *et al.*, 2022). Specifically, the isolation of the *Cypripedium* population in deep valleys between mountains, such as the Hengduan mountains, a biodiversity and endemism hotspot harboring the most species of *Cypripedium* (H Liu *et al.*, 2021*a*; J Liu *et*

al., 2022), could have promoted allopatric diversification at a highly dissected landscape from which several sections or lineages leading to sections diverged, producing the observed pattern of rapid radiations along the backbone of our phylogeny (Favre *et al.*, 2015; Xing and Ree, 2017; Spicer *et al.*, 2020). Additionally, the continuous uplift post-Middle Miocene and the orogeny of the mountains at the southeastern margins of the Tibetan Plateau may have provided a diverse variety of habitats throughout its wide altitudinal range, promoting adaptive radiation (López-Pujol *et al.*, 2011; Wang *et al.*, 2012; Xing and Ree, 2017), while the relatively stable conditions of the Hegduan mountains, which glaciated minimally, could have allowed time for speciation to take place (Spicer *et al.*, 2020; H Liu *et al.*, 2021*a*).

As mentioned before, the cooling climate during the Middle to Late Miocene glacial events (Hansen *et al.*, 2013; Frigola *et al.*, 2018) may have prevented the dispersal of many lineages back to the New World through the Bering Land Bridge, besides the most cold tolerant lineages such as sect. *Bifolia* and sect. *Acaulia* (H Liu *et al.*, 2021*a*). In this case, "jump" dispersals across the Pacific Ocean may have been a more likely route back to the New World for some lineages, as suggested by our biogeographic analysis (e.g., sect. *Californica*, and New World clades of sections *Obtusipetala* and *Cypripedium*). Such long-distance dispersals could be attributed to their small, dust-like seeds that lack endosperm, as their seedlings solely depend on their mycorrhizal associates for nutrition until they become photosynthetically active (Arditti and Ghani, 2000). Due to the size of the orchid seeds, dispersal models indicated that they could travel a distance of up to 2,000 km by wind (Arditti and Ghani, 2000). However, evidence of endozoochory was also found for some orchid species (Suetsugu *et al.*, 2015; Suetsugu, 2020; Y Zhang *et al.*, 2021), while dispersal via epizoochory (e.g., on bodies of mammals and birds) and water currents has been suggested, although it is still uncertain whether orchids seeds could germinate after long exposure to salt water (Arditti and Ghani, 2000).

The glacial cycles of the Late Pliocene and the Quaternary may have promoted the infra-sectional diversification of *Cypripedium*. The isolation of populations in refugia, such as in the mountains at the margins of the Tibetan Plateau, the islands of Japan and Taiwan (Tang *et al.*, 2018), as well as numerous areas in North America, including the Californian coast and the Klamath Mountains (DR Roberts and Hamann, 2015) would have allowed survival and allopatric diversification throughout glacial periods. During the interglacial periods, population expansions resulting in overlapping distributions could have led to hybridization and introgression, further increasing genetic and morphological diversity (e.g., in the *C. macranthos*/*C. tibeticum* complex; P Cribb, Royal Botanical Garden, Kew, UK, 'pers. comm.'). The hybrid swarms and the high number of taxa endemic to Japan and Taiwan within sect. *Cypripedium* underpin this hypothesis.

Conclusion

The phylogenetic relationships within the genus *Cypripedium* have remained generally unresolved despite numerous phylogenies based on different molecular markers and phylogenetic reconstruction methods, largely due to the low support and conflicting topologies at nodes between and within sections. Our study provided a new perspective on the causes of this phenomenon, using high-throughput sequence data from over 600 nuclear loci to gain more insight into the evolution of this genus. Although all established sections were supported as monophyletic, we uncovered potential rapid diversification events that led to incomplete lineage sorting at the backbone of the tree. A similar pattern was also observed within the two subclades of the largest section, *Cypripedium*, which did not correspond to its two subsections following Frosch and Cribb (2012) and SC Chen *et al.* (2013) but matched their respective geographic distributions. Additionally, a whole genome duplication and multiple hybridization events were

linked to the sections *Cypripedium* and *Bifolia*. All these sources of gene tree discordance likely produced mixed phylogenetic signals, preventing the accurate resolution of *Cypripedium*'s phylogeny.

The events above were associated with climatic transitions that led to major ancestral distribution changes. After the slipper orchids likely originated in the Neotropics around the K-Pg boundary, the milder climate of the Early Miocene could have allowed them to disperse to the Old World via the Bering Land Bridge. The Middle Miocene Climatic Transition coincided with the rapid radiations following *Cypripedium*'s dispersal to Southeast Asia, possibly due to isolation in local refugia. Several independent dispersals back to the New World via the Bering Land Bridge or across the Pacific Ocean likely took place, while the glacial-interglacial cycles probably played a role in the further speciation and reticulate evolution observed in *Cypripedium*.

SUPPLEMENTARY DATA

Table S1: Selected examples of taxonomic revisions for the infrageneric classification of *Cypripedium* L. Table S2: *Cypripedium* specimens sampled from the Botanical Collection at Oberhof associated with the BGM and the herbarium M. Table S3: List of publicly available orchid genomes and transcriptomes used in this study. Table S4: Modifications to the Macherey-Nagel NucleoSpin Plant II kit: Genomic DNA from plant (Macherey-Nagel – 07/2014, Rev.09) protocol. Table S5: Final ortholog occupancy statistics per specimen after MO orthology inference. Table S6: The $\alpha(x)$ calculations of the anomaly zone test with the corresponding branch numbers and their branch lengths. Figure S1: Distribution of *Cypripedium* species per section. Figure S2: The concatenation-based phylogeny of the genus *Cypripedium*, inferred with 614 nuclear loci using IQ-TREE 2. Figure S3: The output phylogeny of *Cypripedium* from the Phyparts analysis. Figure S4: The output *Cypripedium* phylogeny from the Quartet Sampling

analysis. Fig. S5: The ASTRAL phylogeny of *Cypripedium* annotated with the corresponding branch numbers referred to in Supplementary Table S6 for the results of the anomaly zone test calculations and in Figure 3. Figure S6: Results of the ASTRAL polytomy test on the ASTRAL phylogeny of *Cypripedium*. Figure S7: Results of the gene duplication test based on the subclade orthogroup tree topology method implemented with the ASTRAL phylogeny of *Cypripedium*, using the (A) bootstrap and the (B) local topology filters. Fig. S8: Supplementary figure for the PhyloNet analysis. Figure S9: The maximum clade credibility tree of *Cypripedium* resulting from the molecular analysis with BEAST 2. Figure S10: Supplementary figures from the BAMM analysis of the *Cypripedium* phylogeny. Figure S11: The relative proportions of the ancestral ranges estimated by BioGeoBEARS (nine area test, DEC+J model) and plotted as pie charts on the corresponding nodes (i.e., right before cladogenesis) or branch corners (i.e., right after cladogenesis) on the dated maximum clade credibility tree of slipper orchids. Figure S12: The relative proportions of the ancestral ranges estimated by BioGeoBEARS (two area test, BAYAREALIKE+J model) and plotted as pie charts on the corresponding nodes (i.e., right before cladogenesis) or branch corners (i.e., right after cladogenesis) on the dated maximum clade credibility tree of slipper orchids.

FUNDING

This work was supported by the Carl Friedrich von Siemens Foundation, Germany.

DATA AVAILABILITY

Target enrichment data generated for this study can be found in the NCBI BioProject XXXXX (please refer to Table XXX for SRA accession numbers). Analyses files are available from the Dryad repository https://doi.org/10.5061/dryad.XXXX

ACKNOWLEDGEMENTS

We thank the curator of the BSM herbarium (acronym M), Dr. Hans-Joachim Esser, and Mr. Sebastian Urban from the Botanical Collection at Oberhof associated with BGM, for helping with the sampling of the herbarium and living *Cypripedium* specimens, respectively. We would also like to thank Dr. Elizabeth Joyce for providing technical support for the biogeographical analysis, as well as Dr. Phillip Cribb and Mr. Tom Velardi, who took the time to share some of their knowledge about *Cypripedium* with us.

Lastly, we would like to acknowledge the people who provided images of the *Cypripedium* taxa included in Figure 1 and 2, namely, Jean-Baptiste Chazalon, Verena Steindl, Luo Chen, Sebastian Urban, Werner Frosch, and the authors of four pictures who made their work available in the Openverse (https://openverse.org/): Figure 1 (a), picture (A): "*Cypripedium irapeanum*" by Moises Béhar is licensed under CC BY-SA 3.0; Figure 1 (a), picture (L): "*Cypripedium reginae*" by Bonnie Isaac is marked with CC0 1.0; Figure 1 (a), picture (M): "[North America] *Cypripedium reginae* fma. *album* '#190405' Walter, Fl. Carol.: 222 (1788)" by Motohiro Sunouchi is licensed under CC BY 2.0; Figure 1 (b), picture (W): "*Cypripedium kentuckiense* Orchi 2012-05-21 009" by Orchi is licensed under CC BY-SA 3.0.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

LITERATURE CITED

Andrews S. **2010**. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online].

Appendices I, II and III. **2023**. https://cites.org/eng/app/appendices.php. 10 Feb. 2023.

Arditti J, Ghani AKA. **2000**. Tansley Review No. 110. *New Phytologist* **145**: 367–421.

Atwood JT. **1984**. The relationships of the slipper orchids (subfamily Cypripedioideae, Orchidaceae). *Selbyana* **7**: 129–247.

Berendse F, Scheffer M. **2009**. The angiosperm radiation revisited, an ecological explanation for Darwin's 'abominable mystery.' *Ecology Letters* **12**: 865–872.

Bezanson J, Karpinski S, Shah V, Edelman A. **2012**. *Julia: A Fast Dynamic Language for Technical Computing*.

Bouckaert R, Vaughan TG, Barido-Sottani J, *et al.* 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology* **15**: e1006650.

Brown PM. **1995**. New taxa and taxonomic notes. **1**: 195–200.

Brown RM, Chalk TB, Crocker AJ, Wilson PA, Foster GL. **2022**. Late Miocene cooling coupled to carbon dioxide with Pleistocene-like climate sensitivity. *Nature Geoscience* **15**: 664–670.

Brown JW, Walker JF, Smith SA. **2017**. Phyx: phylogenetic tools for unix. *Bioinformatics* **33**: 1886–1888.

Cai J, Liu X, Vanneste K, *et al.* 2015. The genome sequence of the orchid Phalaenopsis equestris. *Nature Genetics* **47**: 65–72.

Cameron KM, Chase MW, Whitten WM, *et al.* 1999. A phylogenetic analysis of the Orchidaceae: evidence from rbcL nucleotide. *American Journal of Botany* **86**: 208–224.

Cannon SB, McKain MR, Harkess A, *et al.* 2015. Multiple Polyploidy Events in the Early Radiation of Nodulating and Nonnodulating Legumes. *Molecular Biology and Evolution* **32**: 193–210.

Chandra N, Singh G, Rai ID, *et al.* 2023. Predicting Distribution and Range Dynamics of Three Threatened Cypripedium Species under Climate Change Scenario in Western Himalaya. *Forests* **14**: 633.

Chase MW, Cameron KM, Freudenstein JV, *et al.* 2015. An updated classification of Orchidaceae. *Botanical Journal of the Linnean Society* **177**: 151–174.

Chen Y-S, Deng T, Zhou Z, Sun H. **2018**. Is the East Asian flora ancient or not? *National Science Review* **5**: 920–932.

Chen SC, Liu ZJ, Chen LJ, Li LQ. **2013**. *The Genus Cypripedium in China*. Peking: Science Press.

Choi B, Crisp MD, Cook LG, *et al.* 2019. Identifying genetic markers for a range of phylogenetic utility–From species to family level. *PLoS ONE* **14**: e0218995.

Christenhusz MJM, Fay MF, Chase MW. **2017**. *Plants of the World: An Illustrated Encyclopedia of Vascular Plants*. University of Chicago Press.

Cox AV, Pridgeon AM, Albert VA, Chase MW. **1997**. Phylogenetics of the slipper orchids (Cypripedioideae, Orchidaceae): Nuclear rDNA ITS sequences. *Plant Systematics and Evolution* **208**: 197–223.

Cribb P. **1997**. *The Genus Cypripedium*. Portland: Timber Press.

Degnan JH, Rosenberg NA. **2006**. Discordance of Species Trees with Their Most Likely Gene Trees. *PLOS Genetics* **2**: e68.

Dodsworth S, Chase MW, Leitch AR. **2016**. Is post-polyploidization diploidization the key to the evolutionary success of angiosperms? *Botanical Journal of the Linnean Society* **180**: 1–5.

Drummond AJ, Suchard MA, Xie D, Rambaut A. **2012**. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**: 1969–1973.

Dutton JF, Barron EJ. **1997**. Miocene to present vegetation changes: A possible piece of the Cenozoic cooling puzzle. *Geology* **25**: 39–41.

Eccarius W. **2009**. *Orchideengattung Cypripedium*. EchinoMedia.

Eric Schranz M, Mohammadin S, Edger PP. **2012**. Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. *Current Opinion in Plant Biology* **15**: 147–153.

Eserman LA, Thomas SK, Coffey EED, Leebens-Mack JH. **2021**. Target sequence capture in orchids: Developing a kit to sequence hundreds of single-copy loci. *Applications in Plant Sciences* **9**: e11416.

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

Favre A, Päckert M, Pauls SU, *et al.* 2015. The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. *Biological Reviews* **90**: 236–253.

Filippov E, Andronova E. **2011**. Genetic differentiation in plants of the genus Cypripedium from Russia inferred from allozyme data. *Genetika* **47**: 615–23.

Freudenstein JV, van den Berg C, Goldman DH, Kores PJ, Molvray M, Chase MW. **2004**. An expanded plastid DNA phylogeny of Orchidaceae and analysis of jackknife branch support strategy. *American Journal of Botany* **91**: 149–157.

Fridley J. **2008**. Of Asian Forests and European Fields: Eastern U.S. Plant Invasions in a Global Floristic Context. *PloS one* **3**: e3630.

Frigola A, Prange M, Schulz M. **2018**. Boundary conditions for the Middle Miocene Climate Transition (MMCT v1.0). *Geoscientific Model Development* **11**: 1607–1626.

Frosch W, Cribb P. **2012**. *Hardy Cypripedium: Species, hybrids and cultivation*. Kew Publishing Kew.

Gernhard T. **2008**. The conditioned reconstructed process. *Journal of Theoretical Biology* **253**: 769–778.

Givnish TJ, Spalink D, Ames M, *et al.* 2015. Orchid phylogenomics and multiple drivers of their extraordinary diversification. *Proceedings of the Royal Society B: Biological Sciences* **282**: 20151553.

Givnish TJ, Spalink D, Ames M, *et al.* 2016. Orchid historical biogeography, diversification, Antarctica and the paradox of orchid dispersal. *Journal of Biogeography* **43**: 1905–1916.

Górecki P, Eulenstein O. **2014**. Refining discordant gene trees. *BMC Bioinformatics* **15**: S3.

Guo Y-Y, Luo Y-B, Liu Z-J, Wang X-Q. **2012**. Evolution and Biogeography of the Slipper Orchids: Eocene Vicariance of the Conduplicate Genera in the Old and New World Tropics. *PLOS ONE* **7**: e38788.

Hansen J, Sato M, Russell G, Kharecha P. **2013**. Climate sensitivity, sea level and atmospheric carbon dioxide. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **371**: 20120294.

Herold N, Müller RD, Seton M. **2010**. Comparing early to middle Miocene terrestrial climate simulations with geological data. *Geosphere* **6**: 952–961.

Hopkins DM. **1959**. Cenozoic History of the Bering Land Bridge. *Science* **129**: 1519–1528.

"Letter no. 3285". https://www.darwinproject.ac.uk/letter/?docId=letters/DCP-LETT-3285.xml. 10 Feb. 2023.

Hu S-J, Hu H, Yan N, Huang J-L, Li S-Y. **2011**. Hybridization and asymmetric introgression between Cypripedium tibeticum and C. yunnanense in Shangrila County, Yunnan Province, China. *Nordic Journal of Botany* **29**: 625–631.

Hu Y, Yao Y, Kou Z. **2020**. Exploring on the climate regionalization of Qinling-Daba mountains based on Geodetector-SVM model. *PLoS ONE* **15**: e0241047.

Hui Z, Zhou X, Chevalier M, Wei X, Pan Y, Chen Y. **2021**. Miocene East Asia summer monsoon precipitation variability and its possible driving forces. *Palaeogeography, Palaeoclimatology, Palaeoecology* **581**: 110609.

Huson DH, Scornavacca C. **2012**. Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees and Networks. *Systematic Biology* **61**: 1061–1067.

Izawa T, Kawahara T, Takahashi H. **2007**. Genetic diversity of an endangered plant, Cypripedium macranthos var. rebunense (Orchidaceae): background genetic research for future conservation. *Conservation Genetics* **8**: 1369–1376.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. **2017**. ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods 14: 587–589.

Kim Y-K, Jo S, Cheon S-H, *et al.* 2020. Plastome Evolution and Phylogeny of Orchidaceae, With 24 New Sequences. *Frontiers in Plant Science* **11**.

Klier K, Leoschke MJ, Wendel JF. **1991**. Hybridization and Introgression in White and Yellow Ladyslipper Orchids (Cypripedium candidum and C. pubescens). *Journal of Heredity* **82**: 305–318.

Knyasev MS, Kulikov PV, Knyaseva OI, Semerikov VL. **2000**. Interspecific hybridization in northern Eurasian Cypripedium: morphometric and genetic evidence of the hybrid origin of C. ventricosum. *Lindleyana* **15**: 10–20.

Koenen EJM, Ojeda DI, Bakker FT, *et al.* 2021. The Origin of the Legumes is a Complex Paleopolyploid Phylogenomic Tangle Closely Associated with the Cretaceous–Paleogene (K–Pg) Mass Extinction Event. *Systematic Biology* **70**: 508–526.

Kolanowska M, Jakubska-Busse A. **2020**. Is the lady's-slipper orchid (Cypripedium calceolus) likely to shortly become extinct in Europe?—Insights based on ecological niche modelling. *PLoS ONE* **15**: e0228420.

Kottek M, Grieser J, Beck C, Rudolf B, Rubel F. **2006**. World Map of the Köppen-Geiger Climate Classification Updated. *Meteorologische Zeitschrift* **15**: 259–263.

Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. **2019**. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**: 4453–4455.

Landis MJ, Matzke NJ, Moore BR, Huelsenbeck JP. **2013**. Bayesian Analysis of Biogeography when the Number of Areas is Large. *Systematic Biology* **62**: 789–804.

Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. **2017**. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Molecular Biology and Evolution* **34**: 772–773.

Li Z, Baniaga AE, Sessa EB, *et al.* 2015. Early genome duplications in conifers and other seed plants. *Science Advances* **1**: e1501084.

Li Z, De La Torre AR, Sterck L, *et al.* 2017. Single-Copy Genes as Molecular Markers for Phylogenomic Studies in Seed Plants. *Genome Biology and Evolution* **9**: 1130–1147.

Li S, Ji X, Harrison T, *et al.* 2020. Uplift of the Hengduan Mountains on the southeastern margin of the Tibetan Plateau in the late Miocene and its paleoenvironmental impact on hominoid diversity. *Palaeogeography, Palaeoclimatology, Palaeoecology* **553**: 109794.

Li J, Liu Z, Salazar GA, *et al.* 2011. Molecular phylogeny of Cypripedium (Orchidaceae: Cypripedioideae) inferred from multiple nuclear and chloroplast regions. *Molecular Phylogenetics and Evolution* **61**: 308–320.

Lindley J. **1840**. *The genera and species of orchidaceous plants*. London: Ridgways.

Linkem CW, Minin VN, Leaché AD. **2016**. Detecting the Anomaly Zone in Species Trees and Evidence for a Misleading Signal in Higher-Level Skink Phylogeny (Squamata: Scincidae). *Systematic Biology* **65**: 465–477.

Linnaeus C. **1753**. *Species plantarum : exhibentes plantas rite cognitas ad genera relatas, cum diferentiis specificis, nominibus trivialibus, synonymis selectis, locis natalibus, secundum systema sexuale digestas.* Stockholm: Holmiae, Impensis Laurentii Salvii.

Liu ZJ, Cribb PJ, Lou YB. **2012**. Cypripedium Section Californica, a New Section of Cypripedium (Orchidaceae). *Journal of Fairylake Botanical Garden* **11**: 4.

Liu H, Jacquemyn H, Chen W, *et al.* 2021*a*. Niche evolution and historical biogeography of lady slipper orchids in North America and Eurasia. *Journal of Biogeography* **48**: 2727–2741.

Liu H, Jacquemyn H, He X, *et al.* 2021*b*. The Impact of Human Pressure and Climate Change on the Habitat Availability and Protection of Cypripedium (Orchidaceae) in Northeast China. *Plants* **10**: 84.

Liu J, Milne RI, Zhu G-F, *et al.* 2022. Name and scale matter: Clarifying the geography of Tibetan Plateau and adjacent mountain regions. *Global and Planetary Change* **215**: 103893.

López-Pujol J, Zhang F-M, Sun H-Q, Ying T-S, Ge S. **2011**. Centres of plant endemism in China: places for survival or for speciation? *Journal of Biogeography* **38**: 1267–1280.

Lu H, Guo Z. **2014**. Evolution of the monsoon and dry climate in East Asia during late Cenozoic: A review. *Science China Earth Sciences* **57**: 70–79.

Lu L-M, Mao L-F, Yang T, *et al.* 2018. Evolutionary history of the angiosperm flora of China. *Nature* **554**: 234–238.

Mai U, Mirarab S. **2018**. TreeShrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC genomics* **19**: 23–40.

Matzke NJ. **2013**. BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis in R Scripts. *R package, version 0.2* **1**: 2013.

Matzke NJ. **2014**. Model Selection in Historical Biogeography Reveals that Founder-Event Speciation Is a Crucial Process in Island Clades. *Systematic Biology* **63**: 951–970.

Methner K, Campani M, Fiebig J, Löffler N, Kempf O, Mulch A. **2020**. Middle Miocene long-term continental temperature change in and out of pace with marine climate records. *Scientific Reports* **10**: 7989.

Minasiewicz J, Znaniecka JM, Górniak M, Kawiński A. **2018**. Spatial genetic structure of an endangered orchid Cypripedium calceolus (Orchidaceae) at a regional scale: limited gene flow in a fragmented landscape. *Conservation Genetics* **19**: 1449–1460.

Minh BQ, Schmidt HA, Chernomor O, *et al.* 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **37**: 1530–1534.

Morales-Briones DF, Kadereit G, Tefarikis DT, *et al.* 2021. Disentangling Sources of Gene Tree Discordance in Phylogenomic Data Sets: Testing Ancient Hybridizations in Amaranthaceae s.l. *Systematic Biology* **70**: 219–235.

Morales-Briones DF, Gehrke B, Huang C-H, *et al.* 2022. Analysis of paralogs in target enrichment data pinpoints multiple ancient polyploidy events in Alchemilla sl (Rosaceae). *Systematic Biology* **71**: 190–207.

Neubig K, Williams N, Whitten W, Pupulin F. **2009**. Molecular phylogenetics and the evolution of fruit and leaf morphology of Dichaea (Orchidaceae: Zygopetalinae). *Annals of botany* **104**: 457–67.

Nicolè F, Brzosko E, Till-Bottraud I. **2005**. Population viability analysis of Cypripedium calceolus in a protected area: longevity, stability and persistence. *Journal of Ecology* **93**: 716–726.

Nur Fatihah HN, Fay M, Maxted N. **2011**. Molecular Phylogenetics of Cypripedium L. (Cypripedioideae: Orchidaceae) Based on Plastid and Nuclear DNA Sequences. *Journal of Agrobiotechnology* **2**: 111–118.

Ortiz EM, Höwener A, Shigita G, et al. 2023. A novel phylogenomics pipeline reveals complex pattern of reticulate evolution in Cucurbitales. bioRxiv: 2023.10.27.564367.

Pease JB, Brown JW, Walker JF, Hinchliff CE, Smith SA. **2018**. Quartet Sampling distinguishes lack of support from conflicting support in the green plant tree of life. *American journal of botany* **105**: 385–403.

Pérez-Escobar OA, Dodsworth S, Bogarín D, *et al.* 2021. Hundreds of nuclear and plastid loci yield novel insights into orchid relationships. *American Journal of Botany* **108**: 1166–1180.

Pérez-Escobar OA, Bogarín D, Przelomska N, *et al.* 2023. *The Origin And Speciation Of Orchids*.

Perner H. **2008**. Sinopedilum—a new section of the genus Cypripedium. *Die Orchidee* **59**: 35–51.

Pfadenhauer JS, Klötzli FA. **2020**. Fundamentals towards Understanding Global Vegetation In: Pfadenhauer JS, Klötzli FA, eds. *Global Vegetation: Fundamentals, Ecology and Distribution*. Cham: Springer International Publishing, 1–120.

Pfitzer EHH. **1888**. Die Orchideen. In: *Engler, A., Prantl, K., (Eds): Die Natiirlichen Pflanzen-familien.* Leipzig: Engelmann, 52–96.

Pfitzer EHH. **1894**. Beitraege zur Systematik der Orchideen. **19**: 1–42.

Piet Q, Droc G, Marande W, *et al.* 2022. A chromosome-level, haplotype-phased *Vanilla planifolia* genome highlights the challenge of partial endoreplication for accurate whole-genome assembly. *Plant Communications* **3**: 100330.

POWO. **2023**. *Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. Published on the Internet.* http://www.plantsoftheworldonline.org/. 12 Feb. 2023.

Pupulin F, Díaz-Morales M. **2018**. On the meaning of Cypripedium × grande (Orchidaceae) and its taxonomic history, with a new name for the nothospecies occurring in Costa Rica and Panama. *Phytotaxa* **382**: 167.

R Core Team. **2023**. R: A language and environment for statistical computing.

Rabosky DL, Santini F, Eastman J, *et al.* 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nature Communications* **4**: 1958.

Rabosky DL, Grundler M, Anderson C, *et al.* 2014. BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution* **5**: 701–707.

Rafinesque CS. **1836**. *Flora Telluriana*. Philadelphia [Printed for the author by H. Probasco].

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. **2018**. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology* **67**: 901–904.

Ranwez V, Douzery EJ, Cambon C, Chantret N, Delsuc F. **2018**. MACSE v2: toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. *Molecular biology and evolution* **35**: 2582–2584.

Ree RH, Smith SA. **2008**. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology* **57**: 4–14.

Reichenbach HG. **1854**. *Xenia orchidacea 1*. Leipzig: Brockhaus.

Ren R, Wang H, Guo C, *et al.* 2018. Widespread Whole Genome Duplications Contribute to Genome Complexity and Species Diversity in Angiosperms. *Molecular Plant* **11**: 414–428.

Roberts DL, Dixon KW. **2008**. Orchids: status survey and conservation action plan. *Current Biology* **18**: R325–R329.

Roberts DR, Hamann A. **2015**. Glacial refugia and modern genetic diversity of 22 western North American tree species. *Proceedings of the Royal Society B: Biological Sciences* **282**: 20142903.

Robertson FM, Gundappa MK, Grammes F, *et al.* 2017. Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification. *Genome Biology* **18**: 111.

Rolfe RA. **1896**. The Cypripedium group. *Orchid Review* **4**: 327–334.

Rolfe RA. **1904**. Cypripedium calceolus × macranthos. **12**: 185.

Rolfe RA. **1910**. Cypripedium × ventricosum. **18**: 215.

Ronquist F. **1997**. Dispersal-Vicariance Analysis: A New Approach to the Quantification of Historical Biogeography. *Systematic Biology* **46**: 195–203.

Sayers EW, Bolton EE, Brister JR, *et al.* 2022. Database resources of the national center for biotechnology information. *Nucleic Acids Research* **50**: D20–D26.

Sayyari E, Mirarab S. **2016**. Fast coalescent-based computation of local branch support from quartet frequencies. *Molecular biology and evolution* **33**: 1654–1668.

Sayyari E, Mirarab S. **2018**. Testing for Polytomies in Phylogenetic Species Trees Using Quartet Frequencies. *Genes* **9**: 132.

Schmidt G. **2018**. *Climate Classification Outline*.

Schroeder FG. **1998**. *Lehrbuch der Pflanzengeographie*. Quelle & Meyer.

Serna-Sánchez MA, Pérez-Escobar OA, Bogarín D, *et al.* 2021. Plastid phylogenomics resolves ambiguous relationships within the orchid family and provides a solid timeframe for biogeography and macroevolution. *Scientific Reports* **11**: 6858.

Sheviak C. 1992. Natural hybridisation between Cypripedium montanum and its yellow-lipped relatives. **61**: 558.

Smith SA, Moore MJ, Brown JW, Yang Y. 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC evolutionary biology* **15**: 1–15.

Smith SA, Brown JW, Walker JF. 2018. So many genes, so little time: A practical approach to divergence-time estimation in the genomic era. *PLOS ONE* **13**: e0197433.

Solís-Lemus C, Bastide P, Ané C. 2017. PhyloNetworks: A Package for Phylogenetic Networks. *Molecular Biology and Evolution* **34**: 3292–3298.

Spicer RA, Farnsworth A, Su T. 2020. Cenozoic topography, monsoons and biodiversity conservation within the Tibetan Region: An evolving story. *Plant Diversity* **42**: 229–254.

Steinthorsdottir M, Coxall HK, de Boer AM, *et al.* 2021. The Miocene: The Future of the Past. *Paleoceanography and Paleoclimatology* **36**: e2020PA004037.

Suetsugu K, Kawakita A, Kato M. 2015. Avian seed dispersal in a mycoheterotrophic orchid Cyrtosia septentrionalis. *Nature Plants* **1**: 1–2.

Suetsugu K. 2020. A novel seed dispersal mode of Apostasia nipponica could provide some clues to the early evolution of the seed dispersal system in Orchidaceae. *Evolution Letters* **4**: 457–464.

Szlachetko DL, Kolanowska M, Muller F, Vannini J, Rojek J, Górniak M. 2017. First Guatemalan record of natural hybridisation between Neotropical species of the lady's slipper orchid (Orchidaceae, Cypripedioideae). *PeerJ* **5**: e4162.

Szlachetko DL, Górniak M, Kowalkowska AK, Kolanowska M, Jurczak-Kurek A, Archila Morales F. **2021**. The natural history of the genus Cypripedium (Orchidaceae). *Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology* **155**: 772–796.

Tang CQ, Matsui T, Ohashi H, *et al.* 2018. Identifying long-term stable refugia for relict plant species in East Asia. *Nature Communications* **9**: 4488.

Tank DC, Eastman JM, Pennell MW, *et al.* 2015. Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. *New Phytologist* **207**: 454–467.

Unruh SA, McKain MR, Lee Y-I, *et al.* 2018. Phylotranscriptomic analysis and genome evolution of the Cypripedioideae (Orchidaceae). *American Journal of Botany* **105**: 631–640.

Walid N, Rebbas K, Krouchi F. **2019**. Découverte de Cypripedium calceolus (Orchidaceae) au Djurdjura (Algérie), nouvelle pour l'Afrique du Nord. *Flora Mediterranea* **29**: 207–214.

Wang Y, Zheng J, Zhang W, *et al.* 2012. Cenozoic uplift of the Tibetan Plateau: Evidence from the tectonic–sedimentary evolution of the western Qaidam Basin. *Geoscience Frontiers* **3**: 175–187.

Wen D, Yu Y, Zhu J, Nakhleh L. **2018**. Inferring Phylogenetic Networks Using PhyloNet. *Systematic Biology* **67**: 735–740.

Wong DCJ, Peakall R. **2022**. Orchid Phylotranscriptomics: The Prospects of Repurposing Multi-Tissue Transcriptomes for Phylogenetic Analysis and Beyond. *Frontiers in Plant Science* **13**.

Xiao X, Liang S, He T, Wu D, Pei C, Gong J. **2020**. *Estimating fractional snow cover from passive microwave brightness temperature data using MODIS snow cover product over North America*.

Xing Y, Ree RH. **2017**. Uplift-driven diversification in the Hengduan Mountains, a temperate biodiversity hotspot. *Proceedings of the National Academy of Sciences* **114**: E3444–E3451.

Yamashita Y, Satoh N, Kurosawa T, Kaneko S. **2023**. Genetic diversity and structure of the endangered lady's slipper orchid Cypripedium japonicum Thunb. (Orchidaceae) in Japan. *Population Ecology* **65**: 54–63.

Yang Y, Smith SA. **2014**. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. *Molecular biology and evolution* **31**: 3081–3092.

Yang Y, Moore MJ, Brockington SF, *et al.* 2015. Dissecting Molecular Evolution in the Highly Diverse Plant Clade Caryophyllales Using Transcriptome Sequencing. *Molecular Biology and Evolution* **32**: 2001–2014.

Yang Y, Moore MJ, Brockington SF, *et al.* 2018. Improved transcriptome sampling pinpoints 26 ancient and more recent polyploidy events in Caryophyllales, including two allopolyploidy events. *New Phytologist* **217**: 855–870.

Ye J-W, Tian B, Li D-Z. **2022**. Monsoon intensification in East Asia triggered the evolution of its flora. *Frontiers in Plant Science* **13**.

Yu Y, Nakhleh L. **2015**. A maximum pseudo-likelihood approach for phylogenetic networks. *BMC Genomics* **16**: S10.

Zhang C, Rabiee M, Sayyari E, Mirarab S. **2018**. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC bioinformatics* **19**: 15–30.

Zhang C, Mirarab S. **2022**. Weighting by Gene Tree Uncertainty Improves Accuracy of Quartet-based Species Trees. *Molecular Biology and Evolution* **39**: msac215.

Zhang G-Q, Xu Q, Bian C, *et al.* 2016. The Dendrobium catenatum Lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution. *Scientific Reports* **6**: 19029.

Zhang G-Q, Liu K-W, Li Z, *et al.* 2017. The Apostasia genome and the evolution of orchids. *Nature* **549**: 379–383.

Zhang J-Y, Liao M, Cheng Y-H, *et al.* 2022. Comparative chloroplast genomics of seven endangered Cypripedium species and phylogenetic relationships of Orchidaceae. *Frontiers in Plant Science*: 2029.

Zhang N, Zeng L, Shan H, Ma H. **2012**. Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. *New Phytologist* **195**: 923–937.

Zhang Y, Li Y-Y, Wang M, Liu J, Luo F, Lee Y-I. **2021**. Seed dispersal in Neuwiedia singapureana: novel evidence for avian endozoochory in the earliest diverging clade in Orchidaceae. *Botanical Studies* **62**: 3.

Zuo M, Sun Y, Zhao Y, Ramstein G, Ding L, Zhou T. **2022**. South Asian summer monsoon enhanced by the uplift of Iranian Plateau in Middle Miocene. *Climate of the Past Discussions*: 1–31.