

1 Dense Sampling of Taxa and Genomes Untangles the Phylogenetic Backbone of a Non-
2 model Plant Lineage Rife with Deep Hybridization and Allopolyploidy

3 CHAO XU^{1,2†}, ZE-TAO JIN^{1,2,3†}, HUI WANG^{1,2,4†}, SI-YU XIE^{1,2,5†}, XIAO-HUA LIN^{1,2,5},
4 RICHARD G.J. HODEL⁶, YU ZHANG^{1,2,5}, DAI-KUN MA^{1,2,7}, BING LIU^{1,2}, GUANG-NING
5 LIU², SHUI-HU JIN⁴, LIANG ZHAO⁵, JUN WU³, CHEN REN⁸, DE-YUAN HONG^{1,2}, BIN-BIN
6 LIU^{1,2*}

7 ¹*State Key Laboratory of Plant Diversity and Specialty Crops, Institute of Botany,*
8 *Chinese Academy of Sciences, Beijing 100093, China;*

9 ²*China National Botanical Garden, Beijing 100093, China;*

10 ³*State Key Laboratory of Crop Genetics and Germplasm Enhancement, College of*
11 *Horticulture, Nanjing Agricultural University, Nanjing 210095, China;*

12 ⁴*College of Forestry and Biotechnology, Zhejiang Agriculture and Forestry University,*
13 *Hangzhou 311300, China;*

14 ⁵*College of Life Sciences & Herbarium of Northwest A&F University, Northwest A&F*
15 *University, Yangling 727100, China;*

16 ⁶*Department of Botany, National Museum of Natural History, Smithsonian Institution,*
17 *PO Box 37012, Washington, DC 20013-7012, USA;*

18 ⁷*University of Chinese Academy of Sciences, Beijing 100049, China;*

19 ⁸*Key Laboratory of Plant Resources Conservation and Sustainable Utilization &*
20 *Guangdong Provincial Key Laboratory of Applied Botany, South China Botanical*

21 *Garden, Chinese Academy of Sciences, Guangzhou 510650, China.*

22 *Chao Xu, Ze-Tao Jin, Hui Wang, and Si-Yu Xie contributed equally to this article.*

23 **Correspondence to be sent to: State Key Laboratory of Plant Diversity and Specialty*

24 *Crops, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China;*

25 *Email: liubinbin@ibcas.ac.cn*

26

27 **Abstract:**

28 Phylogenetic networks, rather than purely bifurcating trees, more accurately depict the
29 intricate evolutionary dynamics of most lineages, especially those characterized by
30 extensive hybridization and allopolyploidization events. However, the challenges of
31 achieving complete taxon sampling, and limited financial resources for studying non-
32 model plant lineages, have hindered comprehensive and robust estimation of
33 phylogenetic backbones with guidance from networks. The bellflower tribe,
34 Campanuleae, characterized by a reticulate evolutionary history, serves as an ideal
35 model to investigate how to diagnose nested ancient reticulation events. Here, by
36 integrating multiple genomic data sources and a range of phylogenetic inference
37 methods, we produced a robust phylogenetic backbone for the tribe Campanuleae. Our
38 investigation of reticulate evolution indicates that hybridization and
39 allopolyploidization were instrumental in shaping the diversity of the bellflower tribe,
40 particularly during the initial diversification of the subtribe Phytematinae. Additionally,
41 we ascertained that conflicting topologies resulting from distinct genomic datasets and
42 inference methodologies significantly impact downstream estimates of divergence
43 dating, ancestral area construction, and diversification rates. This study offers a
44 universally relevant framework for deciphering how to use network-based phylogenetic
45 structures using various genomic sources and inference methods. [Campanulaceae,
46 Campanuleae, Cytonuclear discordance, paralog, phylogenomics, reticulate evolution]

47

48 Building on the legacy of Darwin's Origin of Species (Darwin 1859), the Tree of
49 Life (ToL) has been used as a model and a research tool to explore the evolution and
50 relationships between living and extinct organisms with an assumption of bifurcating
51 phylogeny (Mindell 2013). Given the prevalence of reticulation via incomplete lineage
52 sorting (ILS), hybridization, polyploidization, and introgression, modeling the
53 evolutionary connectivity of all life using a bifurcating phylogeny is problematic
54 biologically and unrealistic (Rothfels 2021; Stull et al. 2023). A growing body of
55 genomic and/or phylogenomic studies have provided mounting evidence supporting a
56 network-like structure of life, such as the Bacteria and Archaea lineages (Dagan and
57 Martin 2009; Gontier 2015), the vascular plant lineages (Leebens-Mack et al. 2019;
58 Stull et al. 2021), birds (Jarvis et al. 2014), and mammals (Upham et al. 2019). Over the
59 past decades, improved bioinformatic methods have been developed for teasing apart
60 the various mechanisms underlying complex reticulate evolutionary histories, and
61 multiple software programs have also been developed for resolving the same process,
62 e.g., PhyloNet (Wen et al. 2018) and SNaQ (Solís-Lemus et al. 2017). However, the
63 lack of a standardized procedure for untangling weblike relationships has impeded our
64 understanding of evolutionary patterns and phylogenetic relationships.

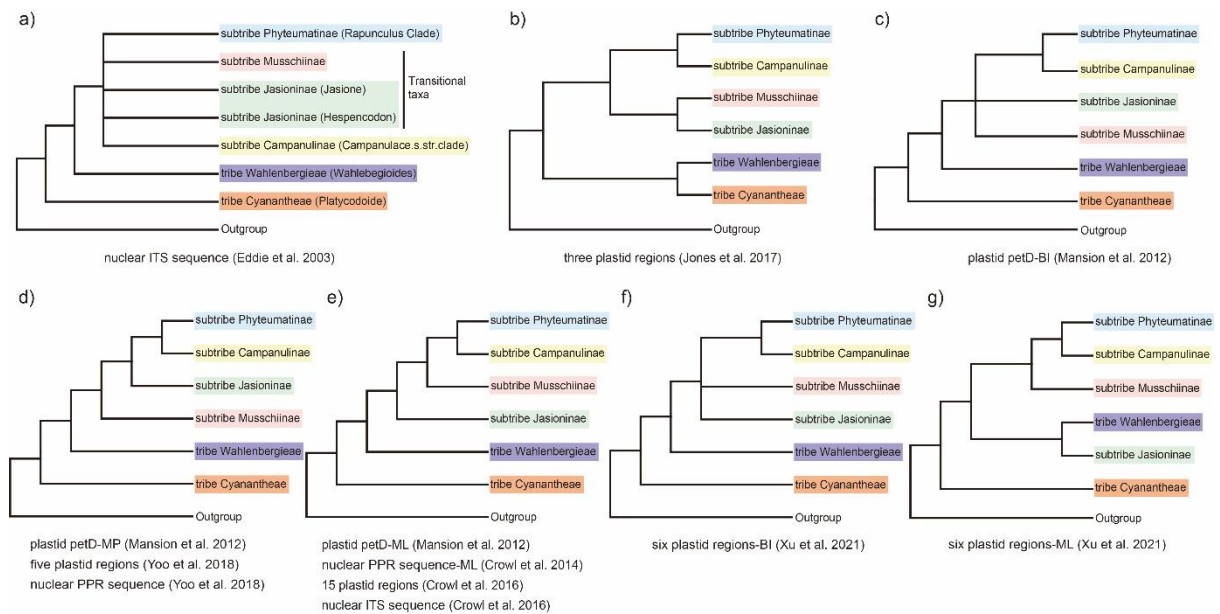
65 In contrast to model plants such as *Arabidopsis thaliana* in Brassicaceae, rice
66 (*Oryza sativa*) and maize (*Zea mays*) in Poaceae, and the tobacco plant (*Nicotiana*
67 *tabacum*) in Solanaceae, non-model plants represent the vast majority of plant diversity
68 on Earth, and most of them have significant ecological, agricultural, or medicinal

69 importance. Studying non-model plants is critical for gaining a broader understanding
70 of plant biology, evolution, and adaptation, especially given the tremendous diversity of
71 plants in nature. However, the understudied background knowledge, cost and resource
72 constraints, and unavailable biological samplings greatly challenged the phylogenomic
73 studies of non-model plants. Critically, when species diversity in non-model lineages is
74 not thoroughly sampled using multiple sources of genomic data, it can be impossible to
75 tease apart ancient evolutionarily significant events such as hybridization. The
76 decreased High-Throughput Sequencing (HTS) cost, especially in China (Liu et al.
77 2021), promoted the genome-level sequencing of non-model plants. It is important to
78 investigate in depth a non-model plant lineage characterized by pervasive hybridization
79 and allopolyploidy, and employ a multi-source genomic approach to explore its
80 phylogenetic backbone and deep reticulation.

81 The Campanuleae tribe, the largest lineage in the Campanulaceae family with over
82 620 species, has undergone extensive hybridization and polyploidy events, as noted in
83 previous studies (e.g., Lammers 2007a, 2007b; Crowl et al. 2017). The tribe
84 Campanuleae, along with other two tribes, Cyanantheae and Wahlenbergieae, forms the
85 Campanuloideae subfamily (or Campanulaceae sensu stricto), which features radial
86 floral symmetry and has a center of diversity in the Holarctic region (Hong and Wang
87 2015). Since the description of *Campanula* L., numerous taxonomists have dedicated to
88 proposing a “natural” infra-tribal classification based on morphological and
89 karyological evidence, e.g., Candolle et al. (1830), Boissier (1875), Fedorov (1957),

90 and Damboldt (1976, 1978). While these investigations provided essential clues for
91 understanding evolutionary relationships, convergent evolution, cryptic species, and
92 frequent reticulate evolution events can hinder comprehensive and accurate taxonomic
93 classification (Crowl et al. 2016, 2017). The genetic age enabled the clarification of
94 some recalcitrant evolutionary relationships. For example, Eddie et al. (2003) estimated
95 the phylogeny of Campanulaceae and diagnosed the polyphyly of *Campanula*, with
96 *Edraianthus* and *Phyteuma* nested within *Campanula*, using nuclear ribosomal internal
97 transcribed spacer (ITS) sequences (Fig. 1a). Subsequently, a series of phylogenetic
98 studies inferred the maternally phylogenetic backbone of Campanulaceae using plastid
99 regions (Fig. 1b-g; Mansion et al. 2012; Crowl et al. 2016; Jones et al. 2017; Yoo et al.
100 2018; Xu and Hong 2021); these studies confirmed the polyphyly of *Campanula*.
101 Recently, Xu and Hong (2021) generated a data matrix with extensive taxon sampling
102 and six plastid regions, identifying four major clades (Fig. 1f, g; the Campanuleae I, II,
103 III, and IV clades) and 24 subclades within the bellflower tribe, including 18 clades
104 confirmed by Mansion et al. (2012) (further subdividing Cam 04 to two clades) and six
105 separate genera: *Favratia*, *Feeria*, *Homocodon*, *Jasione*, *Peracarpa*, and *Trachelium*
106 (Xu and Hong 2021). Because plastid markers have limited variability and maternal
107 inheritance (Gitzendanner et al. 2018), these data painted an incomplete picture of the
108 evolutionary relationships among the bellflower tribe. With the development of next-
109 generation sequencing (NGS) technology and accompanying phylogenomic inference
110 programs, we can now utilize hundreds or thousands of biparentally inherited nuclear

111 genes, whole plastomes, and mitochondrial genes for estimating phylogenies. Newly
 112 developed approaches such as Deep Genome Skimming (DGS; Liu et al. 2021, 2022)
 113 and target enrichment sequencing (Hyb-Seq; Weitemier et al. 2014; Baker et al. 2022)
 114 are getting us closer to the goal of accurate phylogenies, which can be used to detect
 115 underlying mechanisms for gene tree and cytonuclear discordance (Guo et al. 2021). It
 116 is clear that multiple sources of genomic data can overcome some of the limitations and
 117 idiosyncracieaes involved with using single markers types.



118 **Figure 1.** Phylogenetic hypotheses estimated in previous studies among the tribes
 119 Campanuleae, Cyanantheae, and Wahlenbergieae, particularly emphasizing the
 120 relationships among the four subtribes within the tribe Campanuleae, i.e.,
 121 Campanulinae, Jasioninae, Musschiinae, and Phyteumatinae (referring to Fig. 2a). **a)** a
 122 combined nuclear ITS1 and ITS2 sequences (MP tree; Eddie et al. 2003). **b)** three
 123 plastid regions (*petD*, *rpl16*, and *trnK/matK*; MP, MO, and BI trees; Jones et al. 2017).
 124 **c)** plastid *petD* sequence (BI tree; Mansion et al. 2012). **d)** plastid *petD* sequence (MP

125 tree; Mansion et al. 2012); five plastid regions (*atpB*, *matK*, *petD*, *rbcL*, and *trnL-F*)
126 and nuclear PPR70 sequence (ML and BI trees; Yoo et al. 2018). **e)** plastid *petD*
127 sequence (ML tree; Mansion et al. 2012); nuclear PPR70 sequence (ML tree; Crowl et
128 al. 2014); 15 plastid regions (ML tree; Crowl et al. 2016); nuclear ITS sequence (ML
129 tree; Crowl et al. 2016). **f)** six plastid regions (*atpB-rbcL*, *matK*, *petD*-intron, *rbcL*,
130 *rpl16*, and *trnL-F*; BI tree; Xu and Hong 2021). **g)** six plastid regions (ML tree; Xu and
131 Hong 2021).

132 Complementary lines of evidence have shown that reticulate evolution played a
133 significant role in the diversification of the bellflower tribe, especially via
134 hybridization, polyploidization, and ILS (Lammers 2007b; Crowl et al. 2017).
135 Integrating plastomes and 130 nuclear loci, Crowl et al. (2017) uncovered cryptic
136 tetraploid and octoploid *Campanula*, a lineage with four species from the
137 Mediterranean, and revealed that morphological traits failed to distinguish polyploid
138 lineages because only one parental morphology is retained. Previous cytological
139 evidence also showed that nearly 13% of the Campanuloideae are presumed polyploid
140 derivatives (Lammers 2007a), indicating a substantial role of polyploidization in the
141 evolutionary history of bellflowers and their relatives.

142 Increasing genomic data resources in public databases, including highly reusable
143 data types (Guo et al. 2021), such as DGS, Whole-Genome Sequencing (WGS), and
144 transcriptomic sequencing (RNA-Seq) enable using multiple lines of genomic evidence
145 to investigate reticulation histories. In this study, we integrate multiple genomic data

146 sources into a phylogenomic study of Campanulaceae. In total, 659 single-copy nuclear
147 (SCN) genes, derived from Hyb-Seq, RNA-Seq, DGS, as well as plastid protein-coding
148 sequences (CDSs) will be used for phylogenomic analysis. Using this genomic dataset
149 with different gene histories, we assessed cytonuclear discordance to identify potential
150 ILS and hybridization events throughout the deep nodes along the phylogenetic
151 backbone. Molecular dating and biogeographic analysis were used to infer when and
152 where both recent and ancient hybridization and polyploidization occurred, promoting
153 the diversification of Campanuleae. Explicitly, we aim to explore the utility of multi-
154 source genomic data for 1) building a well-supported phylogenetic network backbone
155 for a non-model plant lineage, the tribe Campanuleae, and 2) elucidating hybridization
156 and polyploidization events deep in phylogeny based on phylogenomic and
157 biogeographic analyses.

158 MATERIALS AND METHODS

159 *Taxon Sampling, DNA Extraction, and Sequencing*

160 In this study, we adopted the taxonomic system as described by Lammers (2007a,
161 2007b), which offers a comprehensive synopsis of generic and species delimitation, and
162 has gained extensive acceptance within the Campanulaceae community. Later, Mansion
163 et al. (2012) subdivided *Campanula* sensu lato into 17 distinct clades, relying on
164 chloroplast *petD* group II intron sequences. Our sampling was meticulously designed to
165 encompass all 17 major clades of Campanuleae, to optimize our ability to obtain a well-

166 supported nuclear and plastid backbone for the bellflower tribe. Specifically, our
167 samples comprised 134 accessions, which included 110 ingroup species (116
168 individuals) spanning all 17 clades in Campanuleae and 18 outgroup species (Mansion
169 et al. 2012; see Supplementary Table S1 for details).

170 Our approach integrated data from diverse sequencing strategies, to optimize the
171 data available to guide phylogenomic inference, as demonstrated by Liu et al. (2021,
172 2022). For the phylogenomic analyses of Campanuleae, we harnessed data from DGS,
173 WGS, Hyb-Seq, and RNA-Seq (detailed in Supplementary Table S1). For DGS, WGS,
174 and Hyb-Seq sequencing, we extracted total genomic DNAs from silica-gel dried
175 leaves and, where applicable, herbarium/museum specimens. The extraction was
176 performed using the modified CTAB (mCTAB) method (Li et al. 2013) and carried out
177 at the Institute of Botany, Chinese Academy of Science (IBCAS). For DGS and WGS,
178 we utilized the NEB Next Ultra DNA Library Prep Kit for Illumina (NEB, USA),
179 following the manufacturer's guidelines. Sequencing was performed on the DNBSEQ-
180 T7 and BGISEQ-500 Sequencing System (Novogene, Beijing), yielding paired-end
181 reads of 2×150 bp for 35 accessions and 2×100 bp for five accessions, respectively
182 (see Supplementary Table S1 for more information). For Hyb-Seq, library preparation
183 was completed using the Fast Library Prep Kit (iGeneTech, Beijing). Subsequent
184 solution-based hybridization and target enrichment were carried out with the TargetSeq
185 One[®] Kit at the iGeneTech facility in Zhejiang, China. We used the Illumina NovaSeq
186 6000 platform to generate paired-end reads (2×150 bp) for 82 accessions (refer to

187 Supplementary Table S1 for details). For RNA-Seq, we extracted total genomic RNAs
188 from silica-gel dried leaves via the mCTAB method. The libraries were prepared using
189 the NEBNext[®] Ultra RNA Library Prep Kit for Illumina (USA). This process resulted
190 in paired-end reads of 2×150 bp, which were sequenced on the Illumina NovaSeq
191 6000 Sequencing System for a total of eight accessions (further details in
192 Supplementary Table S1). All raw reads (130 accessions) sequenced for this study have
193 been deposited in the NCBI Sequence Read Archive (SRA) under the BioProject
194 PRJNA895940.

195 *Raw Reads Processing, and Nuclear SCN and Plastid Sequence Assembly*

196 In a recent series of studies, we have developed approaches to integrate multi-
197 source genomic data for phylogenomic analyses (Liu et al. 2021, 2022; Jin et al. 2023).
198 Here, the processing and assembly of raw reads follow the workflow established in
199 previous studies. After sequencing, low-quality reads and base calls were trimmed, and
200 adaptor sequences were removed using Trimmomatic v. 0.39 (Bolger et al. 2014). The
201 quality of the results was subsequently checked using FastQC v. 0.11.9 (Andrews
202 2018). Leveraging the transcriptome of *Adenophora polyantha* Nakai (SRA accession:
203 SRX8528008), we screened putative single-copy genes with MarkerMiner v. 1.0
204 (Chamala et al. 2015), which yielded 659 SCN genes. The clean reads were then used
205 to assemble the SCN genes via the HybPiper v. 2.0 pipeline (Johnson et al. 2016).
206 Specifically, the “hybpiper assemble” command was executed to assemble contigs and

207 extract sequences with the 659 SCN gene sequences as the references. We summarized
208 gene recovery statistics using the “hybpiper stats” command. The gene recovery
209 statistics were used to generate a visual representation of recovery efficiency through
210 the “hybpiper recovery_heatmap” command, enabling us to assess the assembly quality
211 of each gene. Finally, the “hybpiper paralog_retriever” command was run to obtain the
212 sequences, potentially containing paralogs, for all recovered genes. This process
213 generated an unaligned multi-FASTA file for each gene.

214 Given the history of gene rearrangements and diverse repeat sequences in the
215 Campanulaceae plastome (Li et al. 2020), we focused on assembling only the protein-
216 coding sequences (hereafter referred to as plastid CDSs) for plastid phylogenetic
217 inference. We extracted 79 CDS sequences from three chloroplast genomes using
218 Geneious Prime (Kearse et al. 2012); they are *Adenophora remotiflora* (GenBank
219 accession: KP889213), *Campanula takesimana* (GenBank accession: KP006497), and
220 *Trachelium caeruleum* (GenBank accession: EU090187). These 79 CDS sequences
221 served as the reference for assembly. For the assembly of the plastid CDSs, we
222 employed the HybPiper v. 2.0 pipeline (Johnson et al. 2016), mainly following the
223 approach described in our earlier nuclear SCN genes assembly, except the final multi-
224 FASTA file for each CDS sequence were retrieved using “hybpiper retrieve_sequence”
225 command.

226 *Orthology Inference and Data Matrices Generation for Nuclear SCN Genes*

227 Considering the prevalence of allopolyploid species documented in the Index to
228 Plant Chromosome Numbers (IPCN), we applied several methods to identify paralogs
229 and differentiate potential orthologs from homologs. We adopted the orthology
230 inference methodology developed by the Ya Yang Group (Yang and Smith 2014;
231 Morales-Briones et al. 2022). This strategy includes the Monophyletic Outgroups
232 (MO), Rooted Ingroups (RT), and one-to-one orthologs (1to1) approaches. These
233 methods are effective for minimizing the influence of paralogs in phylogenetic
234 inference. The 1to1 method retains only homologs with no duplicated taxa, which avoid
235 the introduction of potential paralogs. In the MO approach, paralogs were identified
236 and pruned using homologs with monophyletic, non-repeating outgroups, and this
237 process involved rerooting and pruning paralogs from root to tip. When the outgroup
238 was absent, only those without duplicated taxa were used. Meanwhile, the RT method
239 removed paralogs by extracting ingroup clades and cutting paralogs from root to tip.
240 Similarly, when the outgroup was missing, only those sequences free from duplicated
241 taxa were considered. For both methods, we set a minimum threshold of 25 ingroup
242 taxa. The full details of paralog assessment and orthology inference are available online
243 (https://bitbucket.org/dfmoralesb/target_enrichment_orthology/src/master/). We
244 generated three datasets from these analyses to use as the basis for phylogenetic
245 inference, i.e., the 1to1, MO, and RT datasets.

246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266

Cleaning of Nuclear and Plastid Sequences

We implemented a series of processing steps to clean low-quality sequences, a method previously employed with success in our studies, such as Liu et al. (2021, 2022) and Jin et al. (2023). To achieve refined alignments despite sequences with inconsistent quality, we utilized MAFFT v. 7.505 (Nakamura et al. 2018) to align each SCN sequence, employing the Smith-Waterman algorithm and the “--maxiterate 1000” parameter. The resulting multiple sequence alignments were trimmed by trimAl v. 1.2 (Capella-Gutiérrez et al. 2009) to remove spurious sequences or poorly aligned regions. Specifically, columns with gaps in over 20% of the sequences or with a similarity score below 0.001 were removed using the parameters “-gt 0.8 -st 0.001”. Further cleaning of the sequences was performed using Spruceup (Borowiec 2019), which identified, visualized, and eliminated outlier sequences, with a window size of 50 and overlap of 25. Alignments produced before and after using Spruceup were concatenated and split, respectively, using AMAS v. 1.0 (Borowiec 2016). Recognizing that exceptionally short sequences could hinder accurate phylogenetic inference for each SCN gene, sequences shorter than 250 bp in each alignment were excluded using a Python script (`exclude_short_sequences.py`) from Liu et al. (2022). The cleaned sequences then served as inputs to infer gene trees through RAxML v. 8.2.12 (Stamatakis 2014), using the option “-f a” and 200 BS replicates for clade support evaluation. To ensure the accuracy of species tree inference, TreeShrink v. 1.3.9 (Mai and Mirarab 2018) was employed to identify and remove excessively long branches in each gene tree. After the

267 above processing steps were complete, the sequences were termed ‘clean nuclear
268 genes’, with three separate data matrices: 1to1, MO, and RT approach.

269 *Accurate Phylogenetic Inference with Multiple Methods*

270 To obtain accurate phylogenies and identify the topological discordance between
271 trees, we employed both concatenated and coalescent-based methods. The gene trees,
272 refined by removing long branches via TreeShrink, served as the input trees to estimate
273 the species tree using ASTRAL-III (Zhang et al. 2018), which is statistically
274 consistency with the multi-species coalescent model. Notably, any input gene tree
275 branches with low support (≤ 10) were collapsed using phyx (Brown et al. 2017), as
276 collapsing gene tree nodes with BS support below a threshold value can enhance
277 accuracy (Zhang et al. 2018). Clean nuclear genes were used for both ML and BI tree
278 inference. The most suitable partitioning schemes and molecular evolution models were
279 identified through PartitionFinder2 (Stamatakis 2006; Lanfear et al. 2016), with default
280 settings. The resulting schemes and models were then used in subsequent ML tree
281 estimates via IQ-TREE2 v. 2.2.0.3 (Minh et al. 2020) — with 1000 SH-aLRT and
282 ultrafast bootstrap replicates — and RAxML v. 8.2.12 (Stamatakis 2014) using the
283 GTRGAMMA model for each partition and 200 rapid bootstrap (BS) replicates for
284 clade support. BI analysis was conducted using MrBayes 3.2.7a (Ronquist et al. 2012),
285 running Markov Chain Monte Carlo (MCMC) analyses for 50 million generations.
286 Stationarity was achieved when the average standard deviation of split frequencies

287 remained under 0.01. Trees were analyzed every 1,000 generations, with the initial 25%
288 of samples discarded as burn-in. Subsequent trees were used to generate a 50%
289 majority-rule consensus tree.

290 We compiled a dataset comprising 79 plastid CDS sequences, hereafter referred to
291 as the ‘plastid CDS dataset’, for phylogenetic analysis. While processing the plastid
292 sequences and performing the phylogenetic inference, we largely follow the nuclear
293 SCN methodology, excluding the steps of eliminating short sequences and long
294 branches.

295 *Gene Tree and Species Tree Discordance Analyses*

296 We used *phyparts* to calculate unique, conflicting, and concordant bipartitions
297 within individual orthologs across the phylogeny (Smith et al. 2015). We conducted
298 both quick concordance (-a 0) and full concordance (-a 1) analyses to address the issue
299 arising from orthologs with missing taxa. The conflict analysis produces a pie chart for
300 each node, segmented into five sections. These sections depict varying proportions of
301 orthologs, such as those supporting the clade (in blue), those supporting the main
302 alternative for that clade (in green), those supporting the remaining alternatives (in red),
303 the uninformative (in dark grey), and the missing ones (in light grey). Additionally, we
304 computed the ‘internode certainty all’ (ICA) scores on the input
305 concatenated/coalescent-based tree based on the set of ortholog trees (Salichos et al.
306 2014). The results from *phyparts* were illustrated using

307 `phypartspiecharts_missing_uninformative.py`, a Python script developed by Morales-
308 Briones, which is available at:
309 [https://bitbucket.org/dfmoralesb/target_enrichment_orthology/src/master/phypartspiech](https://bitbucket.org/dfmoralesb/target_enrichment_orthology/src/master/phypartspiecharts_missing_uninformative.py)
310 `arts_missing_uninformative.py`

311 As a counterpart to *phyparts* in analyzing phylogenomic discordance, QS adeptly
312 identifies discordance in large-sparse and genome-wide datasets. This method addresses
313 challenges related to alignment sparsity and can differentiate between strong conflict
314 and weak support (Pease et al. 2018). For each internal branch, QS produces three
315 distinct scores: Quartet Concordance (QC), Quartet Differential (QD), and Quartet
316 Informativeness (QI). Each approach for quantifying discord offers unique yet
317 complementary insights. The results from QS are visually represented using the
318 `plot_QC_ggtree.R`, an R package developed by Shui-Yin Liu, accessible at
319 https://github.com/ShuiyinLIU/QS_visualization.

320 *Coalescence Simulation for Testing the Effect of ILS*

321 Gene tree discordance can arise from single evolutionary events, including ILS,
322 hybridization, and allopolyploidization, or some combination of these factors. To
323 distinguish among these, we implemented a series of analyses. We utilized coalescence
324 simulation to test if ILS could explain gene tree conflicts; this method has proven
325 effective in recent research (Moralis-Briones et al. 2021; He et al. 2022; Liu et al.
326 2022). First, we employed the previously described method (in the phylogenetic

327 inference section) to estimate the species tree using ASTRAL-III (Zhang et al. 2018).
328 Using this ASTRAL ultrametric species tree, we simulated 10,000 gene trees under the
329 multi-species coalescent (MSC) model with the “sim.coal.tree.sp” function in the R
330 package Phybase v. 1.5 (Liu and Yu 2010). We subsequently compared the distribution
331 of tree-to-tree distances between simulated and empirical gene trees using the
332 DendroPy v. 4.5.2 Python package (Sukumaran and Holder 2010). The result was
333 visualized in a column chart where the extent of overlap between simulated and
334 empirical gene tree bars represents the goodness-of-fit of the coalescent model,
335 indicating whether ILS is a plausible explanation for gene tree discordance.

336 *Inference of Global Split Networks*

337 The split network is a valuable tool for visualizing inconsistencies within a dataset.
338 In such a network, ancestral species are not designated by specific nodes. Instead,
339 parallel edges denote the splits that derive from the dataset, and their length indicates
340 the importance of these splits. For the tribe Campanuleae, we constructed a split
341 network using SplitsTree v. 4.19.0 (Huson and Bryant 2006), drawing on aligned SCN
342 sequences from the MO dataset. This construction involved applying split
343 decomposition to the uncorrected_P distances. Given the large divergence in this
344 dataset, we presented the split network produced by the NeighborNet method for
345 achieving higher resolution, and we then employed the EqualAngle network
346 construction algorithm to infer the split network.

347 *Phylogenetic Network Estimation*

348 In this study, we used network approaches investigate the early diversification of
349 major clades within the tribe Campanuleae, encompassing three tribes and four
350 subtribes. To reduce the computational burden of the network analysis, we selected a
351 subset of 11 species. This dataset is specifically tailored to test the potential ancient
352 reticulate evolution events that occurred among the three tribes and four subtribes.

353 Given the comprehensive taxon sampling within the subtribe Phyteumatinae, we
354 grouped all 78 samples (73 species) into six monophyletic clusters, i.e., six subclades,
355 across 12 nuclear and four plastid trees. Additionally, we generated another dataset,
356 comprising seven species, to assess the impacts of hybridization and
357 allopolyploidization during the initial diversification of the subtribe Phyteumatinae.

358 We used Species Networks applying Quartets (SNaQ) approach, as detailed by
359 Solís-Lemus and Ané (2016), for the Maximum pseudolikelihood estimation of species
360 networks. This method is integrated within the PhyloNetworks package (Solís-Lemus et
361 al. 2017) in Julia. Notably, it accounts for ILS via the coalescent model while
362 addressing horizontal gene inheritance through reticulation nodes in the network. The
363 methodology leverages pseudolikelihood, avoiding the intensive computation
364 associated with full likelihood and facilitating estimations at the quartet level. This
365 enhances computational efficiency due to its easy parallelization (Solís-Lemus and Ané
366 2016). For phylogenetic network analysis, we followed the comprehensive protocol
367 provided by Solís-Lemus, available at <https://github.com/crs14/PhyloNetworks.jl/wiki>.

368 We used h values ranging between 0 and 6, undertaking 50 runs for the network
369 inference.

370 *Allopolyploidy Analysis*

371 Gene-tree Reconciliation Algorithm with Multi-labeled trees (MUL-trees) for
372 Polyploid Analysis (GRAMPA) adapted an algorithm for topology-based gene-tree
373 reconciliation to work with MUL-trees (Thomas et al. 2017), and this program,
374 GRAMPA, can identify the parental lineages that hybridized to form
375 auto/allopolyploids. Given the similarity between processes such as hybridization and
376 allopolyploidization, we also tested potential auto/allopolyploidy events among the
377 major clades in whole tree (including tribe Wahlenbergieae, tribe Cyanantheae, and four
378 clades in tribe Campanuleae) and among the six subtribes in clade I of the tribe
379 Campanuleae. Because GRAMPA can only infer one WGD at a time, we employed the
380 method proposed by Morales-Briones et al. (2022) to successively explore WGD events
381 along the backbone of a phylogeny. We applied this approach to the backbone of the
382 tribe Campanuleae. Briefly, we specified the clades mentioned above as -h1 and -h2 and
383 perform reconciliation search using all MO ortholog trees against the MO species tree
384 estimated with ASTRAL-III (Zhang et al. 2018). The MO ortholog trees, and species
385 tree after removing the clade identified as polyploid in former GRAMPA analysis, were
386 used to run the next round of analysis on the remaining clades with the same settings
387 until no polyploid clade were detected.

388

Dating Analysis and Ancestral Area Reconstruction

389

The earliest known macrofossils of the family Campanulaceae are seeds from

390

Campanula paleopyramidalis, discovered in Miocene deposits ca. 17-16 million years

391

ago (Mya), of Nowy Sacz in the Carpathians, Poland (Oszast and Stuchlik 1977;

392

Łańcucka-Środoniowa 1979; Nemcok et al. 1998). These fossilized seeds have been a

393

pivotal calibration point in several prior studies, including Cellinese et al. (2009),

394

Mansion et al. (2012), Olesen et al. (2012), Crowl et al. (2014, 2016), and Jones et al.

395

(2017). This species is closely related to the extant *Campanula pyramidalis*. We utilized

396

this fossil as the MRCA of a combined clade comprising *Campanula carpatica*, *C.*

397

pulla, *C. rainerii*, and *Favratia zoysii*. Furthermore, we incorporated two secondary

398

calibration points to constrain the deeper nodes in Campanulaceae: the crown clade of

399

the family Campanulaceae is dated between 72.24 to 52.66 Mya, and the crown clade

400

of the subfamily Campanuloideae is dated between 63.52 to 44.58 Mya (Li et al. 2019).

401

To investigate the impact of conflicting topologies on dating analyses, we used the

402

MCMCTree package in PAML v. 4.9j to estimate divergence times of Campanulaceae

403

based on multiple distinct topologies (concatenated and coalescent-based trees) and

404

datasets (plastome and nuclear). We started by determining the nucleotide substitution

405

rate and Hessian Matrix through the MCMCTree program, with the independent rates

406

clock model and the GTR substitution model. For each dataset, we performed two

407

separate runs of the MCMCTree analysis, every time initiated with different seeds. The

408

initial 1,000,000 generations of each Markov Chain Monte Carlo (MCMC) were

409 discarded as burn-in. Subsequently, we collected samples every ten generations,
410 amounting to a total of 500,000 samples. These samples were assessed using Tracer v.
411 1.7.1 (Rambaut et al. 2018) to confirm convergence and ensure that the effective
412 sample size (ESS) of all parameters exceeded 200. Finally, we visualized the results
413 from both runs using FigTree v. 1.4.4 and cross-validated them to ascertain
414 convergence.

415 We utilized the software BioGeoBEARS v. 1.1.2 (Matzke 2018) implemented in
416 RASP v. 4.2 to reconstruct the ancestral area of the bellflower tribe. Based on the
417 current distribution of Campanulaceae, we classified its geographic range into six
418 distinct regions: (A) Europe, (B) Northern Asia, (C) Africa, (D) North America, (E)
419 Southern East Asia combined with Australasia, and (F) South West Asia. Three distinct
420 estimated trees were used as input to test the effect of topological discordance on
421 ancestral area reconstruction, in which a maximum of four areas were specified. Six
422 models provided by BioGeoBEARS v.1.1.2 were utilized to estimate the biogeographic
423 history, and the AICc values from these models were then compared to identify the
424 optimal model for reconstructing the ancestral area of Campanuleae.

425 *Chromosome Number Reconstructions*

426 We collected haploid chromosome numbers (n) from both the Chromosome Count
427 Database (<http://ccdb.tau.ac.il>) and the Campanulaceae monograph by Lammers
428 (2007a). Using the PAML ultrametric tree inferred from the nuclear concatenation-

429 based tree as input, we employed ChromEvol v. 2.2 to infer the ancestral haploid
430 chromosome number through a likelihood-based approach. Notably, species lacking
431 chromosomal data are denoted with the symbol ‘X’. When a single species exhibited
432 multiple chromosome counts, we selected the count observed most frequently for
433 subsequent analysis. Our analysis was conducted under the parameter “_mainType =
434 All_Models”, encompassing ten evolutionary models, and we utilized
435 “_simulationsNum 1000000” to enhance the precision of our analysis. The best-fit
436 model was estimated using the Akaike Information Criterion (AIC) score.

437 *Diversification Analyses*

438 We used Bayesian Analysis of Macroevolutionary Mixtures (BAMM; Rabosky et
439 al. 2014a) to estimate the diversification rate of Campanulaceae. The dated tree,
440 constructed from the aforementioned concatenated orthologs tree, served as the input
441 tree. We adopted a non-random incomplete taxon sampling strategy to avoid error
442 introduced by incomplete sampling species and computed the sampling fraction for
443 each clade within Campanulaceae. MCMC simulation with four chains was performed
444 for 10,000,000 generations under the “species-extinction” model and sampling every
445 1,000 generations. The first 1,000,000 generations were discarded as burn-in, and the
446 remaining samples were then analyzed using the R package BAMMTOOLS (Rabosky
447 et al. 2014b) to assess whether the effective sample size (ESS) exceeded 200 and to
448 generate plots.

449

RESULTS

450

SCN and Plastid Genes Assembly and Nuclear Orthology Inference

451

452

453

454

455

456

457

458

459

460

This study integrated multiple sources of genomic data, including 82 Hyb-Seq, eight RNA-Seq, and 48 DGS data, for phylogenomic analyses. The nuclear assembly from HybPiper resulted in a variable number of SCN genes, ranging from 519 to 654 (Supplementary Fig. S1). Non-chimeric sequences recovered from HybPiper have been utilized to perform orthology inference using three different methods, and this process resulted in three distinct datasets: 1to1 containing 445 genes, MO with 645 genes, and RT with 660 genes. The final alignment of these three datasets, each with 134 taxa, contained 598,993, 864,819, and 887,333 characters, respectively. We employed HybPiper for the assembly of 79 plastome protein-coding sequences (CDSs), and the recovery efficiency was visualized as a heatmap (Supplementary Fig. S2).

461

Nuclear Phylogenetic Inference and Gene Tree Discordance Analyses

462

463

464

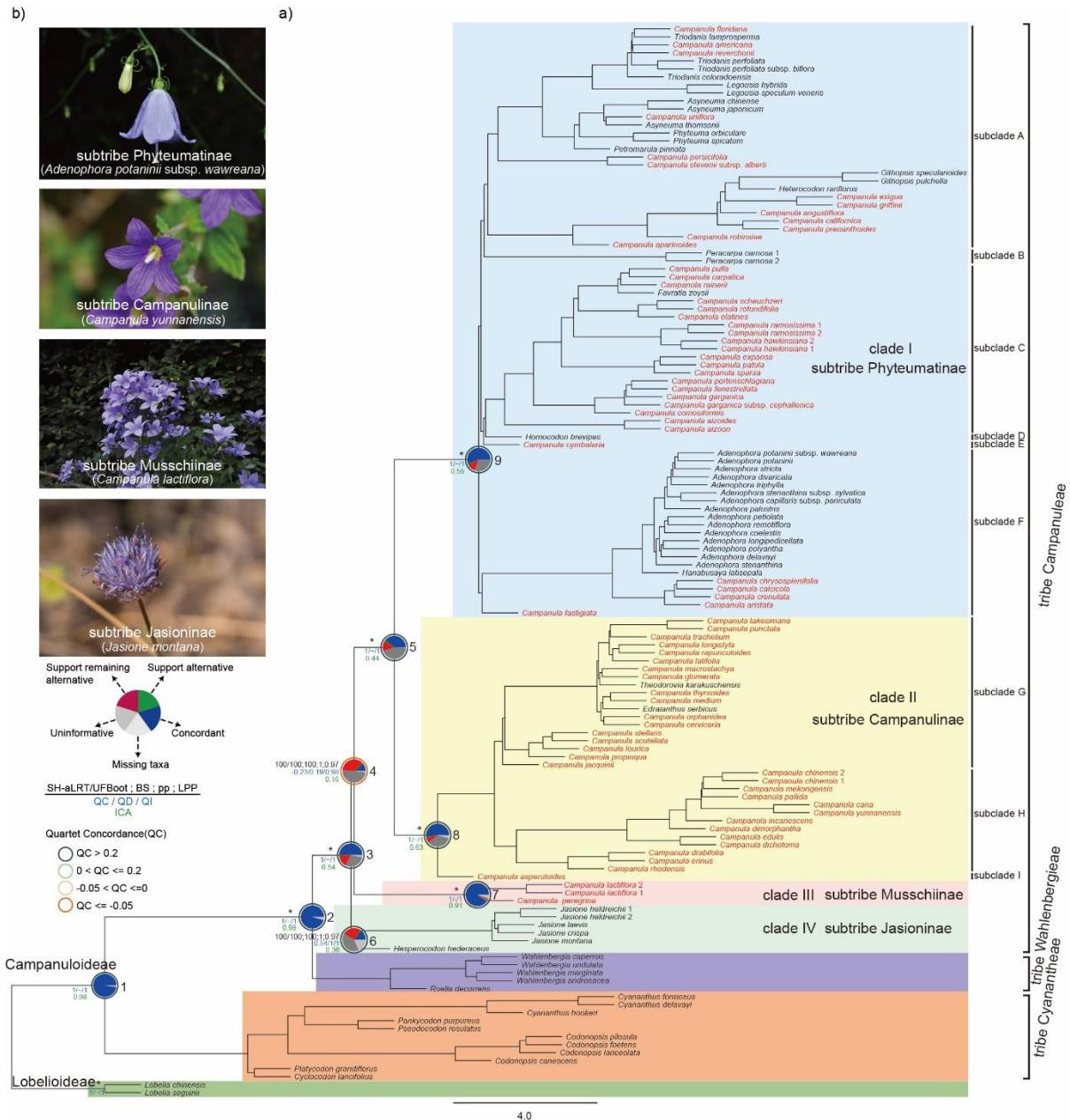
465

466

467

Integrating concatenation and coalescent-based methods, we generated four trees for each dataset based on various phylogenetic inference methods: two trees from ML (RAxML and IQ-TREE2), one from Bayesian Inference (MrBayes), and one species tree estimated based on coalescent theory (ASTRAL-III). These 12 nuclear trees confirmed the monophyly of the subfamily Campanuloideae and the three-tribe classification within Campanuloideae: tribe Campanuleae, tribe Cyanantheae (orange),

468 and tribe Wahlenbergieae (purple) (Supplementary Figs. S3-S14), almost all the
 469 informative gene trees were concordant with these nodes (node 1: 521 out of 522, ICA
 470 = 0.98; node 2: 510 out of 512; ICA = 0.98), and full QS support (1/-/1) (Fig. 2a;
 471 Supplementary Figs. S15-S22).



472 **Figure 2.** A represented tree-like phylogenetic backbone of the subfamily
 473 Campanuloideae inferred from SCN genes, emphasizing the major clades in the tribe

474 Campanuleae. **a)** Species tree of the tribe Campanuleae in the framework of the
475 subfamily Campanuloideae inferred from ASTRAL-III of the nuclear Monophyletic
476 Outgroup (MO) orthologs. Summarized phylogenetic supports of the focal nine nodes
477 from four trees based on the nuclear MO dataset were presented above the branch.
478 From left to right (labeled in black above branch), the SH-aLRT support and Ultrafast
479 Bootstrap (UFBoot) estimated from IQ-TREE2 (details referring to Supplementary Fig.
480 S8); the bootstrap support (BS) values from RAxML analysis (details referring to
481 Supplementary Fig. S7); Bayesian posterior probability values (pp) from MrBayes
482 (details referring to Supplementary Fig. S9); the local posterior probability (LPP) from
483 ASTRAL-III (details referring to Supplementary Fig. S10) (e.g., 100/100; 100; 1; 0.97);
484 asterisks (*) indicated full support (100/100; 100; 1; 1). Values for Quartet
485 Concordance/ Quartet Differential/ Quartet Informativeness estimated from Quartet
486 Sampling analysis were provided below branches (e.g., 1/-/1, labeled in blue) (details
487 referring to Supplementary Fig. S20). The Pie charts on these nodes illustrated the
488 proportion of gene trees that were concordant with the corresponding clade in the
489 species tree (blue), the proportion that supported the main alternative (green), the
490 proportion that supported the first remaining alternative (red), the proportion considered
491 uninformative (deep gray), and the proportion that are missing in the trees (gray); the
492 value of partial sampling ICA were also presented below (labeled in green) (details
493 referring to Supplementary Fig. S19). Campanuloideae was classified into three tribes,
494 tribe Cyanatheae (orange), tribe Wahlenbergieae (purple), and tribe Campanuleae; tribe

495 Campanuleae was further divided into four clades /subtribes, clade I (subtribe
496 Phyteumatinae with blue background), clade II (subtribe Campanulinae with yellow
497 background), clade III (subtribe Musschiinae with pink background), and clade IV
498 (subtribe Jasioninae with light green background); clade I and II were subdivided into
499 nine subclades (A-I). All species belonging to the genus *Campanula* were highlighted
500 in red, indicating the polyphyly of *Campanula*. **b)** Represented species of four subtribes
501 in tribe Campanuleae, indicating the morphological diversity of flowers. From top to
502 bottom: subtribe Phyteumatinae (*Adenophora potaninii* subsp. *wawreana*), subtribe
503 Campanulinae (*Campanula yunnanensis*), subtribe Musschiinae (*Campanula*
504 *lactiflora*), and subtribe Jasioninae (*Jasione montana*). Photos credit to You-Pai Zeng,
505 Hai-Lei Zheng, Ke Cheng, and Jia-Nong Li (from top to bottom).

506 Additionally, all nuclear trees consistently supported the monophyly of four major
507 clades within the bellflower tribe, such as clade I (blue), clade II (yellow), clade III
508 (pink), and clade IV (light green) (Fig. 2a; Supplementary Figs. S3-S14). All four
509 clades were recovered with maximum support and full QS support (1/-/1), except clade
510 IV which includes five individuals of *Jasione* and one species of *Hesperocodon*
511 *hederaceus*—received relatively lower support in the three ASTRAL-III coalescent
512 trees (LPP = 0.97) and strong QS support (0.54/1/1) (Fig. 2a nodes 6-9; Supplementary
513 Figs. S6, S10, S14, and S20). Monophyly of the clade I, clade II, and clade III were
514 supported by most of the informative trees: clade I with 296 concordant trees (out of
515 370 informative trees; ICA = 0.55; Fig. 2a node 9), clade II with 275 concordant trees

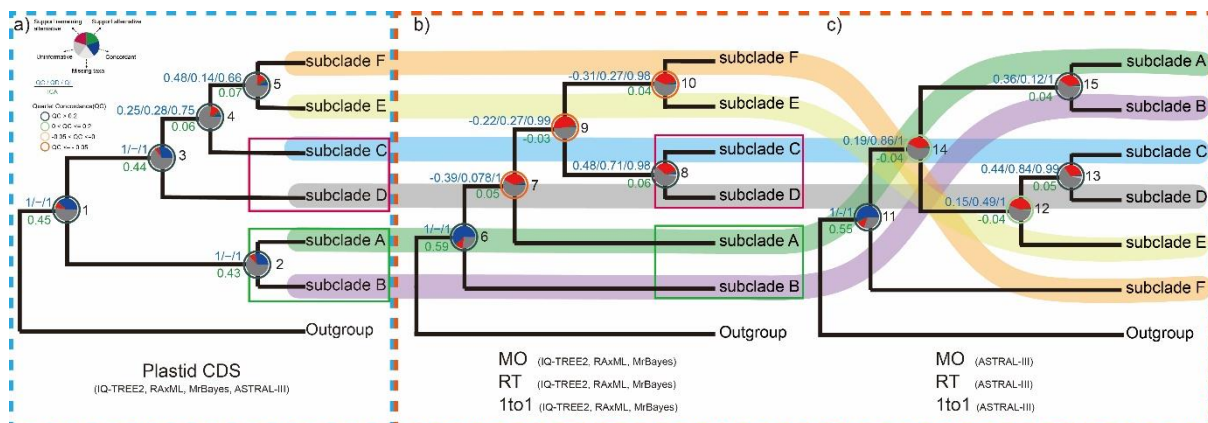
516 (out of 325 informative trees; ICA = 0.63; Fig. 2a node 8), and clade III with 479
517 concordant trees (out of 489 informative trees; Fig. 2a node 7); but clade IV were
518 supported by only 74 concordant trees out of 223 informative trees (ICA = 0.36; Fig. 2a
519 node 6).

520 The sister group consisting of clade I and clade II was recovered with maximum
521 support, 203 concordant trees out of 294 informative trees (ICA = 0.44), and full QS
522 score (1/-/1) (Fig. 2a node 5). Clade III was recovered as sister of clade I + clade II with
523 relatively high support (SH-aLRT/UFBoot = 100/100; BS = 100; pp = 1; LPP = 0.97),
524 only 55 concordant trees (out of 208 informative trees; ICA = 0.16), and counter QS
525 support with a strong majority of quartets supporting an alternative discordant topology
526 (-0.23/0.19/0.98) (Fig. 2a node 4). Finally, clade IV was recovered as the sister to the
527 rest of tribe Campanuleae with maximum support, 265 concordant trees out of 361
528 informative trees (ICA = 0.54), and full QS score (1/-/1) (Fig. 2a node 3).

529 According to the monophyletic groups recovered in the 12 nuclear trees, we
530 subdivided clade I into six subclades (A-F) and clade II into three subclades (G-I) (Fig.
531 2a; Supplementary Figs. S3-S14). The phylogenetic relationships in the 12 nuclear trees
532 were concordant for the three subclades (G-I) in clade II; however, these 12 nuclear
533 trees revealed significant conflicting topologies in clade I.

534 In both the concatenation and coalescent-based tree of plastid CDSs (Fig. 3a;
535 Supplementary Figs. S23-S28), the sister relationship between subclade F and E was
536 recovered by 5 out of 18 informative trees (ICA = 0.07) and strong QS support with

537 discordant skew (0.48/0.14/0.66) (Fig. 3a node 5). Subclade C was recovered as sister
 538 to a combined clade of subclade F and E with weak QS support with discordant skew
 539 (0.25/0.28/0.75), and only four concordant trees (out of 18 informative trees; ICA =
 540 0.06) (Fig. 3a node 4). Subclade D was recovered as sister to a large clade consisting of
 541 subclade C, subclade E, and subclade F, with 22 concordant trees (out of 27 informative
 542 trees; ICA = 0.44) and full QS support (1/-/1) (Fig. 3a node 3). The sister group
 543 composed of subclade A and subclade B was recovered with full QS support (1/-/1) and
 544 21 concordant trees (out of 28 informative trees; ICA = 0.43) (Fig. 3a node 2), and was
 545 placed as sister to the rest of clade I with full QS support (1/-/1) and 29 concordant
 546 trees (out of 36 informative trees; ICA = 0.45) (Fig. 3a node 1).



547 **Figure 3.** Comparative visualization of three conflicting topologies from different
 548 datasets and inference methods for clade I (equivalent to subtribe Phyteumatinae). **a)**
 549 The same topology inferred from IQ-TREE2, RAxML, MrBayes, and ASTRAL-III
 550 based on the plastid CDSs dataset. **b)** The same topology estimated by IQ-TREE2,
 551 RAxML, and MrBayes across all three nuclear datasets (MO, RT, and 1to1). **c)** The
 552 same topology estimated by ASTRAL-III across all three nuclear datasets. Quartet

553 Sampling analysis values for QC/QD/QI are displayed above branches (e.g., 1/-/1, in
554 blue). Pie charts on the nodes, determined by *phyparts*, illustrated the proportion of
555 gene trees that were concordant with the corresponding clade in the species tree (blue),
556 the proportion that supported the main alternative (green), the proportion that supported
557 the first remaining alternative (red), the proportion considered uninformative (deep
558 gray), and the proportion that are missing in the trees (gray). Below the branches,
559 values of the partial sampling ICA are presented (in green). Different color bandings
560 beneath subclades A to F visually highlight the topological discrepancies between them.

561 All the 12 nuclear trees recovered the monophyly of clade I with full QS support
562 (1/-/1), and almost all the informative gene trees being concordant with these nodes
563 (219 out of 262 for node 6, ICA = 0.59; 296 out of 370 for node 11, ICA = 0.55).
564 However, almost all the concerned nodes within the clade I showed conflict with a little
565 percentage of supporting trees, but QS scores of most nodes revealed supported
566 alternative topologies (Fig. 3 nodes 6-15). The sister group consisting of subclade C
567 and subclade D was recovered with strong QS support (i.e., a strong majority of
568 quartets supported the focal nodes, and the low skew in discordant frequencies
569 indicated no alternative history was favored) in both concatenation-based (node 8,
570 0.48/0.71/0.98) and coalescent-based trees (node 13, 0.44/0.84/0.99), moreover, a few
571 informative trees showed concordance on this nodes (ten out of 139 for concatenation-
572 based trees, ICA = 0.06, node 8; nine out of 187 for coalescent-based trees, ICA = 0.05,
573 node 9). The topologies of the rest of clade I showed high level varied among

574 concatenation and coalescent-based trees. In the concatenation-based trees, the sister
575 group composed of subclade E and subclade F was recovered with only nine
576 concordant trees (out of 157, ICA = 0.04, node 10), and counter QS support (-
577 0.31/0.27/0.98) with a skew in discordance suggesting an alternative discordant
578 topology; furthermore, this group was placed as the sister to subclade C + subclade D,
579 with five concordant trees (out of 185, ICA = -0.03, node 9), and counter QS support (-
580 0.22/0.27/0.99). Then subclade A was recovered as sister to the clade consisting of
581 subclades C to F, with 11 concordant trees (out of 153, ICA = 0.05, node 7) and counter
582 QS support. Finally, subclade B was the earliest-divergent lineage within clade I. In the
583 ASTRAL trees based on nuclear, subclade E was placed as the sister to the group
584 consisting of subclade C and subclade D, with only one concordant tree (out of 284 ,
585 ICA = -0.04, node 12), and the QS score also showed weak support with discordant
586 skew (0.15/0.49/1) indicating a possible alternative topology. The group consisting of
587 subclade A and subclade B with was recovered with nine concordant trees (out of 218,
588 ICA = 0.04, node 15), and weak QS support with discordant skew (0.36/0.12/1),
589 moreover, this group was recovered as the sister of the rest subclades within clade I
590 except subclade F, with six concordant trees (out of 230, ICA = -0.04, node 14) and
591 weak QS support (0.19/0.86/1). Finally, subclade F was placed as the earliest-divergent
592 lineage within clade I.

593

Cytonuclear Discordance

594

The results from different datasets (nuclear and plastid CDSs) and various

595

phylogenetic inference methods (concatenation and coalescent-based) consistently

596

indicated similar topologies on the main clade nodes (Fig. 3; Supplementary Figs. S3-

597

S14, S23-S26). However, some conflicts were observed within clade I. In the

598

concatenation and coalescent-based trees inferred from plastid CDSs, clade I was

599

divided into two major monophyly, subclade A and subclade B consisted the first-

600

branching lineage, the other subclades composed another monophyly (Fig. 3a). Similar

601

to the trees of plastid CDSs, the monophyly composed by subclades C to F was

602

recovered in the concatenation-based trees of nuclear too, with two groups consisting

603

respectively of subclade C + subclade D and subclade E + subclade F, but the subclade

604

A and subclade B were placed as the first and the second branching lineages(Fig. 3b).

605

The topology of the ASTRAL trees was vastly different with trees of plastid CDSs,

606

except the monophyly consisting of subclade A and subclade B, subclade C, subclade

607

D, and subclade E consisted a group, the subclade F was placed as the earliest-

608

divergent lineage which was sister to the other with in clade I (Fig. 3c).

609

Reticulate Evolution of the Four Major Clades in the tribe Campanuleae

610

Reticulate evolution events were analyzed in more detail due to the strong cytonuclear

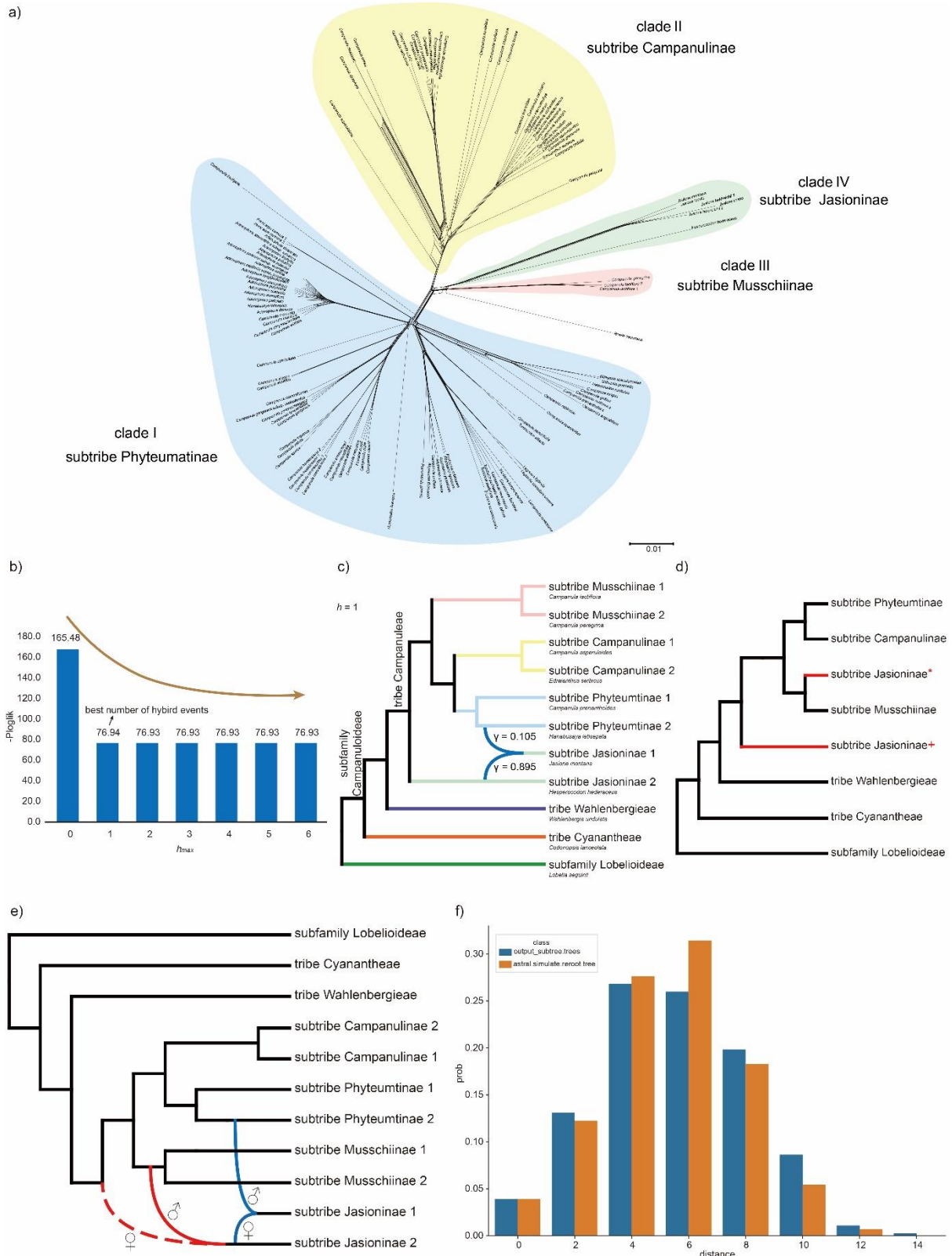
611

discordance among several branches. We built a split network utilizing sequences from

612

the MO dataset and designated *Roella decurrens* as the outgroup. Similar to the nuclear

613 phylogenetic tree, the split network indicated the separation of four clades in the tribe
 614 Campanuleae; the clades were each well differentiated from the rest of the tribe, with
 615 clades III and IV being relatively close together (Fig. 4a). **Figure 4.** Exploring the



616 phylogenetic network of the major clades, including three tribes in the subfamily
617 Campanuloideae and four subtribes in the tribe Campanuleae. **a)** Split network
618 estimated from the nuclear MO dataset, showcasing four distinct subtribes. Clade I,
619 represented in blue, corresponds to the subtribe Phyteumatinae; Clade II, shown in
620 yellow, aligns with the subtribe Campanulinae; Clade III, depicted in pink, associates
621 with the subtribe Musschiinae; and Clade IV, in light green, represents the subtribe
622 Jasioninae. **b)** Bar chart representing the pseudo-loglikelihood scores (-ploglik) across a
623 range of maximum reticulations (from zero to six). The chart highlights the optimal
624 number of hybrid events, with $h_{max} = 1$ being identified as the best number. **c)**
625 Phylogenetic network inferred from SNaQ analysis with $h_{max} = 1$ as the optimal
626 network. On the right, species names are displayed in a reduced font size beneath their
627 corresponding tribe/subtribe names, indicating the species chosen as representatives
628 from various subfamilies, tribes, or subtribes in the analysis. **d)** The simplified most
629 parsimonious multi-labeled trees (MUL-trees) derived from the species tree based on
630 the nuclear MO dataset that includes all taxa. Branches marked in red signify the
631 allopolyploid origin of the subtribe Jasioninae, further highlighted by a red asterisk or a
632 red plus sign from different parents. **e)** Summary of the potential phylogenetic network
633 among the major clades within the subfamily Campanuloideae. Red curves denote the
634 potential allopolyploidization events inferred from the GRAMPA analysis, with the
635 dotted lines indicating extinct ancestors. The symbols ♂ and ♀ on the curving
636 branches signify paternal and maternal parents. Blue curves highlight potential

637 hybridization events based on the SNaQ analysis. **f)** Distribution of tree-to-tree
638 distances between empirical gene trees and the species tree inferred from ASTRAL-III
639 analysis, compared to those from the coalescent simulation.

640 Additionally, it also revealed significant levels of reticulation in the clade,
641 implying intricate relationships among the species within tribe Campanuleae (Fig. 4a).

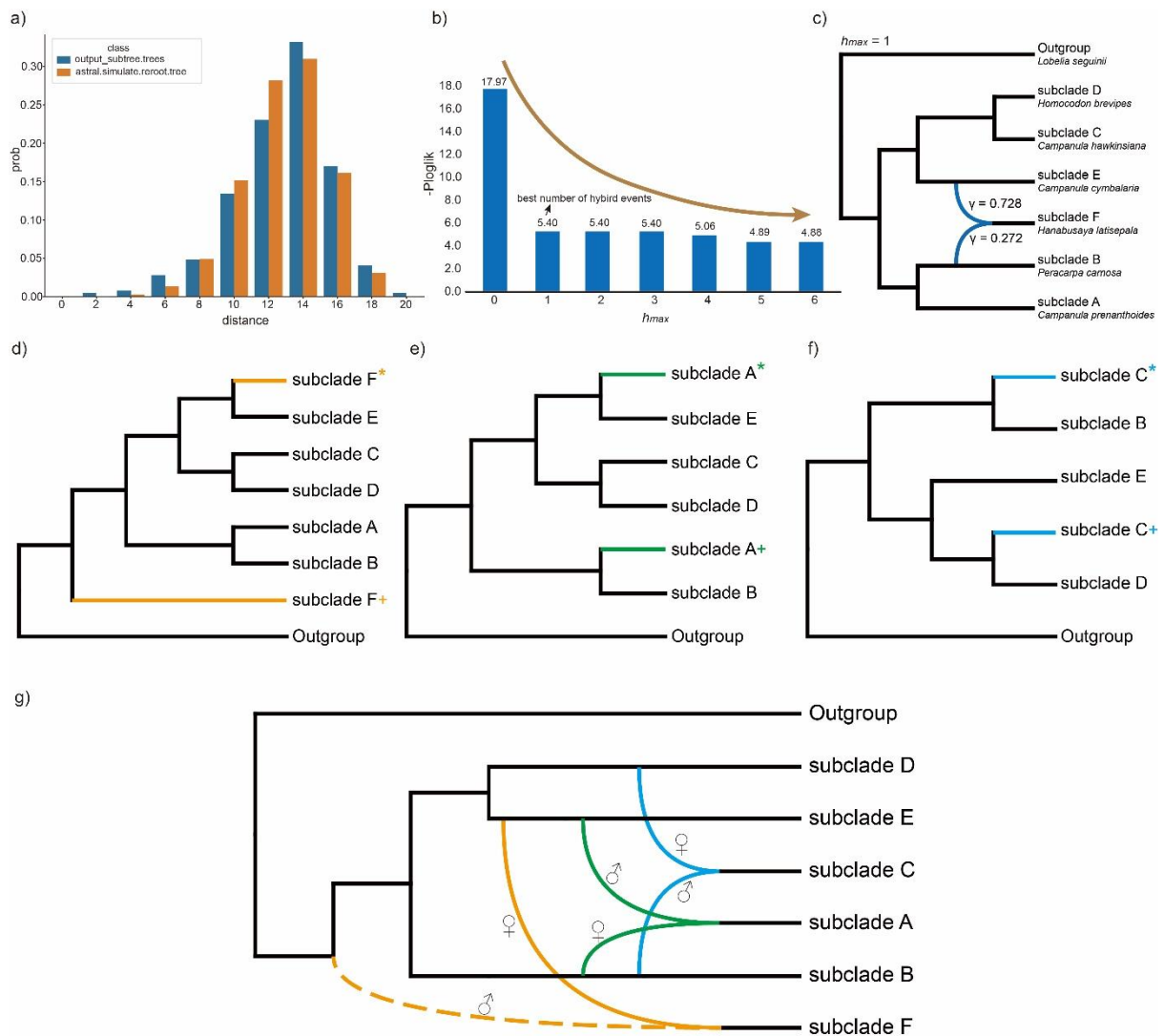
642 The coalescent simulation analysis for tribe Campanuleae suggested that ILS
643 could not fully explain the observed conflict between gene trees and species trees (Fig.
644 4f). We selected specific species as representatives to investigate possible hybrid events
645 within tribe Campanuleae. The plot of pseudo-loglikelihood scores suggested that the
646 optimal number of hybrid events, inferred from the SNaQ network analysis, was one
647 (Fig. 4b; Supplementary Fig. S29). This indicated the hybrid origin of subtribe
648 Jasioninae 1, arising between subtribe Jasioninae 2 ($\gamma = 0.895$) and subtribe
649 Phyteumatinae 2 ($\gamma = 0.105$) (Fig. 4c).

650 GRAMPA was employed to investigate potential allopolyploidy events associated
651 with the origin of the subtribe Jasioninae. The inferred MUL-tree supported that the
652 subtribe Jasioninae was of allopolyploid origin between subtribe Musschiinae and an
653 unsampled or extinct lineage sister to tribe Campanuleae (Fig. 4d; Supplementary Fig.
654 S30).

655 *Reticulate Evolution of the Six Subclades in the Subtribe Phyteumatinae*

656 Given the intensive discordance among the topologies based on different datasets

657 and alternative phylogenetic inference methods, several analyses were employed to test
 658 the causes. We used coalescent simulation analysis to assess the impact of ILS, and the
 659 results indicated that ILS was not the primary factor contributing to the conflict within
 660 clade I (Fig. 5b). Several species were chosen from different subclades as
 661 representatives to identify possible hybrid events within clade I. Using the SNaQ
 662 network analysis, we identified the optimal number of hybrid events ($h_{max} = 1$) with the
 663 best score (Fig. 5c; Supplementary Fig. S31), revealing the hybrid origin of subclade F,
 664 with parental lineages subclade E ($\gamma = 0.728$) and subclade B ($\gamma = 0.272$) (Fig. 5d).



665 **Figure 5.** Phylogenetic network exploration of the six monophyletic groups (subclades)

666 within the subtribe Phyteumatinae. **a)** Coalescent simulation analysis showcasing the
667 distribution of distances between empirical gene trees and the species tree. **b)** Bar chart
668 showcasing pseudo-loglikelihood scores (-ploglik) over a spectrum of maximum
669 reticulations, ranging from zero to six. The chart underscores $h_{max} = 1$ as the optimal
670 count for hybrid events. **c)** Phylogenetic network inferred from SNaQ analysis with h_{max}
671 = 1 as the optimal network. To the right, representative species from various subclades
672 are noted with their names in smaller font sizes, displayed beneath their associated
673 subclades. **d)** The multi-labeled trees (MUL-trees) based on the species tree inferred
674 from nuclear MO dataset, encompassing all taxa within clade I (= subtribe
675 Phyteumatinae). Branches in orange highlight the potential allopolyploid origin of
676 subclade F, further denoted by an orange asterisk or plus sign. **e)** The MUL-trees
677 visualized after excluding subclade F, as depicted in d). Branches colored in green
678 indicate the allopolyploid origin of subclade A, further marked by a green asterisk or
679 plus sign. **f)** The MUL-trees visualized after removing subclades A and F, as referenced
680 in e). Branches in blue highlight the allopolyploid origin of subclade C, complemented
681 by a blue asterisk or plus sign. **g)** Summary of the potential reticulate evolutionary
682 relationships among the six subclades within the subtribe Phyteumatinae. The orange,
683 green, and blue curves depict the allopolyploidy events determined by the GRAMPA
684 analysis, with the dotted line signifying an extinct ancestor. The symbols ♂ and ♀ on
685 the curving branches signify paternal and maternal parents.

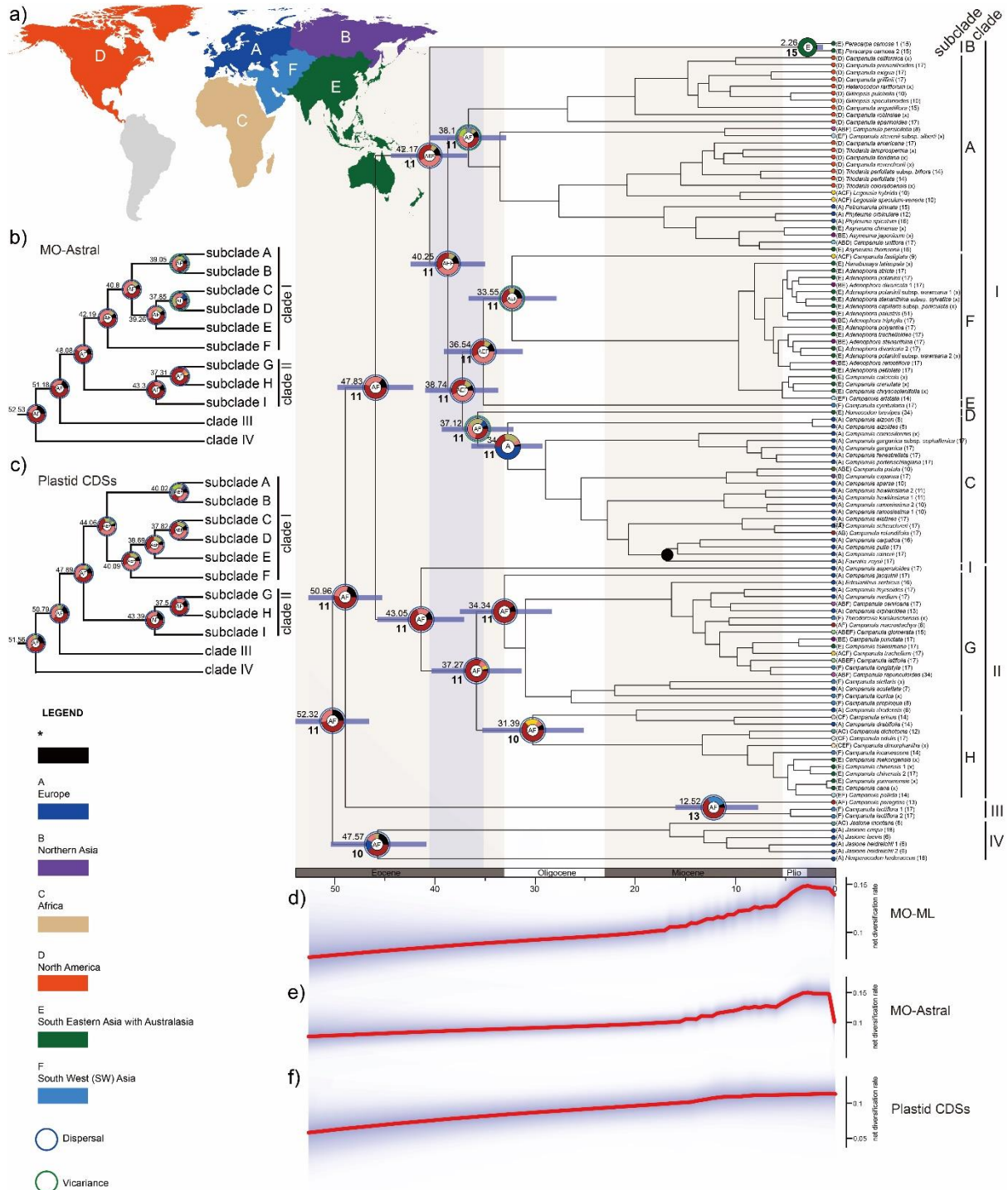
686 The MUL-tree inferred from GRAMPA with the lowest reconciliation score

687 indicated the allopolyploid origin of subclade F, with the subclade E and the common
688 ancestor of clade I as parental lineages (Fig. 5e; Supplementary Fig. S32). After
689 removing subclade F (Fig. 5e orange), the analysis revealed subclade A as an
690 allopolyploid clade, with parental lineages including subclades E and B (Fig. 5f;
691 Supplementary Fig. S33). Finally, upon removing subclade A (Fig. 5f green), the
692 analysis found subclade C to have an allopolyploid origin clade between subclade B
693 and subclade D (Fig. 5g; Supplementary Fig. S34).

694 *Dating Analyses, Ancestral Area, and Chromosome Number Reconstructions*

695 We constrained nodes with a fossil calibration point to estimate divergence times
696 and diversification dynamics based on three different phylogenies (ML tree and
697 ASTRAL tree inferred from the MO dataset; ML tree inferred from the plastid CDSs)
698 within the subfamily Campanuloideae. The result based on all three phylogenetics
699 indicated that the tribe Campanuleae originated in Europe-South West Asia during the
700 early Eocene, diverging at *c.* 52.32 million years ago (Mya) (95% highest posterior
701 density (HPD): 56.19-48.5 Mya) (Fig. 6a; Supplementary Figs. S35-S37).

702 **Figure 6.** Divergence time estimation and geographical range evolution of the
 703 bellflower tribe Campanuleae. a) Dated chronogram of the tribe Campanuleae inferred
 704 from PAML based on the nuclear MO dataset. Focal nodes feature estimated



705 divergence times and ancestral geographical ranges, with the purple shading
706 highlighting the divergence time range for the six subclades within clade I (equivalent
707 to subtribe Phyteumatinae). The inset map in the upper left outlines the six distribution
708 areas used for geographical analysis: (A) Europe, (B) Northern Asia, (C) Africa, (D)
709 North America, (E) Southern East Asia with Australasia, and (F) South West (SW)
710 Asia. A black circle denotes the fossilized seed constraint. The gametophytic
711 chromosome count was displayed in brackets next to the relevant species name, and the
712 boldface numbers represent the reconstructed ancestral chromosome counts. **b)** The
713 ASTRAL-estimated species tree of the nuclear MO dataset with estimated divergence
714 time and geographical range marked on focal nodes. **c)** The plastid CDSs-inferred
715 topology with estimated divergence time and geographical range marked on focal
716 nodes. **d)** Net diversification rate (species/my) of Campanuleae estimated from the
717 Maximum Likelihood (ML) phylogeny using the nuclear MO dataset. The red line
718 represents the median, while the shaded blue area demarcates the 95% credible
719 intervals for the rate. **e)** The net diversification rate estimated from the species tree of
720 ASTRAL-III based on the nuclear MO dataset. **f)** The net diversification rate from the
721 plastid phylogeny.

722 Furthermore, the majority of clades underwent divergence during the Eocene to
723 the Oligocene period. Clade I had its origins in Europe-South Eastern Asia with
724 Australasia-South West Asia by the middle Eocene, and diverged at *c.* 42.17 Mya (95%
725 HPD: 46.22-38.22 Mya) based on the ML-trees inferred from the MO dataset and

726 plastid CDSs (Fig. 6a-c; Supplementary Figs. S35-S43). In contrast, the result based on
727 the ASTRAL-tree inferred from the MO dataset indicated Europe-South West Asia as
728 the origin of clade I (Fig. 6b). Within clade I, six subclades underwent divergence
729 during the Eocene by around *c.* 42.17-36.54 Mya based on the MO dataset (Fig. 6a light
730 purple shadow, 6b), but a little earlier based on the plastid CDSs which indicated the
731 divergence by *c.* 44.06-37.82 Mya (Fig. 6c). In addition, clade II diverged around *c.*
732 43.05 Mya (95% HPD: 47.62-38.55 Mya), and clade IV followed suit at *c.* 47.57 Mya
733 (95% HPD: 52.5-42.48 Mya). In contrast, clade III experienced diversification in the
734 Miocene, occurring approximately *c.* 12.52 Mya (95% HPD: 16.46-7.9 Mya) (Fig. 6a).
735 The diversification rates, as inferred from both the MO dataset and the plastid CDSs,
736 exhibited an increasing trend from the Eocene to the present. Moreover, the rates
737 inferred from the MO dataset showed sudden increases in the late Miocene, which
738 indicated rapid speciation (Fig. 6d-f).

739 The model CONST_RATE_DEMI_EST, as determined by ChromEvol, was
740 identified as the best fit for chromosome evolution (AIC = 503.2). Using this model, we
741 deduced that the most probable ancestral haploid chromosome number ($n = 11$) was
742 shared by the subfamily Campanuloideae and the tribe Campanuleae. This count was
743 consistently observed in the MRCA of the tribes Cyanantheae, Wahlenbergieae, and
744 Campanuleae. Within the tribe Campanuleae, the MRCA of both clades I and II
745 maintained the ancestral haploid chromosome number 11. In contrast, the MRCA of
746 clade III presented an ancestral chromosome number of 13, and clade IV had a count of

747 10. For most subclades within clades I and II, the ancestral number remained 11, with
748 the exception of subclade B, which exhibited a chromosome number of 15 (Fig. 6a;
749 Supplementary Fig. S44).

750

751 DISCUSSION

752 By integrating multiple genomic data sources and various phylogenetic inference
753 methods, we produced a robust phylogenetic backbone for the bellflower tribe
754 Campanuleae within the subfamily Campanuloideae. Our analyses of deep reticulate
755 evolution suggest that hybridization and allopolyploidization have played a pivotal role
756 in the diversification of the Campanuleae tribe, especially in the early diversification of
757 the subtribe Phytentinae. We demonstrated that fully untangling deep reticulation is
758 optimally accomplished by leveraging data from multiple genomic sources, especially
759 when inferring the phylogenetic backbone of non-model plant lineages. Additionally,
760 thorough taxon sampling is also critically important to fully dissect deep reticulation.
761 Moreover, it is essential to note that conflicting topologies derived from separate
762 genomic datasets and inference methods can significantly influence downstream
763 inferences, such as dating analyses, ancestral area determination, and diversification
764 rate estimates.

765 *Exploring the Reticulate Phylogenetic Backbone of the Bellflower Tribe Campanuleae*

766 In traditional phylogenetic models, evolutionary relationships are portrayed as
767 bifurcating branches, suggesting that species or lineages diverge and proceed to evolve
768 independently. However, this strictly branching framework often fails to capture the
769 complexity of evolutionary dynamics as they occur in nature (Jin et al. 2023). In many
770 instances, evolutionary processes are better represented as being network-like, where
771 species or lineages may engage in various forms of hybridization or experience
772 intricate, web-like interactions (Stull et al. 2023). Such network-like patterns can arise
773 from various complex processes involving multiple and ongoing events, such as ILS,
774 introgression, hybridization, and/or allopolyploidization.

775 In this study, we undertook a comprehensive approach integrating extensive taxon
776 sampling with diverse genomic data types, including DGS, Hyb-Seq, RNA-Seq, and
777 WGS. Utilizing state-of-the-art automated tree-based orthology inference methods
778 (Yang and Smith 2014; Morales-Briones et al. 2022), we carefully estimated paralogs
779 and generated three alternative nuclear data matrices (1to1, MO, and RT), along with a
780 complete dataset of 79 plastid CDSs for phylogenomic analysis. Employing various
781 phylogenomic inference methods, such as concatenated and coalescent-based methods,
782 we produced 16 phylogenetic trees—12 nuclear and four plastid. These trees confirmed
783 the classification of three tribes as previously outlined by Hong and Wang (2015) and
784 recovered four strongly supported major clades—namely Clade I, II, III, and IV—in the
785 tribe Campanuleae (Fig. 2; Supplementary Figs. S3-S14, S23-S26). Nevertheless, our

786 findings also highlighted the ambiguous phylogenetic position of Clade IV (or the
787 subtribe Jasioninae), which appears to be either a sister group to the other three clades
788 in Campanuleae (Fig. 2; Supplementary Figs. S3-S14, S23, S25 and S26) or related to a
789 combined clade (Clade I + II, as shown in Supplementary Fig. S24). We hypothesized
790 that the evolutionary relationships among these four major clades in the Campanuleae
791 tribe may not be strictly bifurcating but could exhibit a more network-like structure.
792 This hypothesis was corroborated by our SplitsTree network analysis (Fig. 4a).

793 Tree incongruence has been a prevalent and challenging issue in molecular
794 phylogenetic studies (Wendel and Doyle 1998; Som 2015; Kapli et al. 2020; Steenwyk
795 et al. 2023). Such incongruence can arise from both biological and non-biological
796 factors. Non-biological factors primarily encompass limited taxon and/or gene
797 sampling, the use of paralogous genes, and errors in sequence alignment (Steenwyk et
798 al. 2023). In our study, these factors seem insufficient to explain the observed
799 incongruence. Our extensive taxon sampling and use of genomic-level data largely rule
800 out stochastic uncertainties from insufficient taxon and/or gene sampling. Furthermore,
801 we used three strategies of orthology inference to reduce the effect of paralogs, and it is
802 unlikely the observed topological inconsistencies among individual gene trees can be
803 attributed solely to paralogy. In addition, sequence alignments have undergone
804 meticulous visual examination and are mostly well-aligned. Tools like trimAl help trim
805 low-quality alignments, and TreeShrink filters out samples with aberrantly long
806 branches. This suggests that our sequence alignment is unlikely a major factor in tree

807 incongruence. Biologically, tree incongruences can result from ILS,
808 hybridization/introgression, or horizontal gene transfer (HGT; Kapli et al. 2020). While
809 HGT is more common in bacteria and archaea, it is less so in eukaryotes (Kurland et al.
810 2003; Keeling and Palmer 2008; Boto 2010), making it an unlikely cause of the tree
811 incongruence in our case. On the other hand, both ILS and hybridization/introgression
812 are established contributors to tree incongruence across the tree of life (Degnan and
813 Rosenberg 2009; Mallet et al. 2016). Thus, it is plausible that these factors, individually
814 or in combination, are behind the observed tree incongruence in Campanuleae.

815 Distinguishing between ILS and hybridization/introgression based solely on
816 patterns of tree incongruence is challenging. Both processes can produce similar
817 incongruence patterns, complicating the identification of the specific cause of
818 incongruence in phylogenetic studies (Degnan and Rosenberg 2009; Degnan 2018). To
819 further dissect the potential events contributing to this network-like evolution, we
820 performed step-by-step analyses to distinguish the possible events in this net-like
821 evolution. Coalescent simulation analyses indicated that ILS alone could not account
822 for the observed reticulate patterns (Fig. 4f). Subsequent PhyloNetwork analyses using
823 the SNaQ algorithm and polyploidy assessments with the GRAMPA reconciliation
824 algorithm were carried out to explore the potential roles of hybridization and
825 polyploidy.

826 Given the strongly supported gene tree and species tree discordance for the
827 phylogenetic position of clade III and IV in different datasets and inference methods

828 (Fig. 1), the phylogenetic position of subtribe Jasioninae remains a subject of debate.
829 As we delved into the diversification of the three tribes and four subtribes, we
830 pinpointed an allopolyploidy event (Fig. 4b,c; Supplementary Fig. S31) and a distinct
831 hybridization event (Fig. 4d; Supplementary Fig. S30). The GRAMPA analysis
832 unveiled the allopolyploid origin of subtribe Jasioninae, resulting from the MRCA of
833 the tribe Campanuleae (acting as the maternal lineage) and the MRCA of the subtribe
834 Musschiinae (serving as the paternal lineage, Fig. 4d,e). Our analyses of dating and
835 ancestral area reconstruction consistently identified the likely occurrence of this
836 allopolyploidization event in Europe and SW Asia (around the Tethys Sea) during the
837 early Eocene, as supported by both nuclear (52.32 Mya from the ASTRAL species tree
838 and 52.53 Mya from the ML tree) and plastid (51.56 Mya) topologies (Supplementary
839 Figs. S35-S43). Further, our PhyloNetwork results confirmed the hybrid origin of
840 *Jasione montana* between the MRCA of *Hanabusaya latisepala* (paternal lineage) and
841 the MRCA of *Hesperocodon hederaceus* (maternal lineage). Nevertheless, the dating
842 analyses of the stem clade of *Jasione montana* showed disparities between nuclear
843 (17.08 Mya estimated from the ML tree and 16.53 Mya from the ASTRAL species tree)
844 and plastid data (2.08 Mya). Considering these strongly supported but discordant time
845 estimates, we hypothesize that the MRCA of *Jasione montana* might have captured the
846 chloroplast genome from a particular *Jasione* species, an event possibly occurred in
847 Europe during the early Quaternary.

848 Within the subtribe Phyteumatinae, we conducted thorough taxon sampling

849 encompassing all 15 clades, as confirmed by Xu and Hong (2021). Significant
850 nuclear/plastid gene trees and cytonuclear discordance were detected among the 16
851 generated trees (Fig. 2; Supplementary Figs. S3-S14, S23-S26). We classified the 15
852 clades of subtribe Phyteumatinae into six monophyletic groups, labeled as subclades A-
853 F (Figs. 3, 5). We utilized an iterative strategy to estimate the potential polyploidy
854 origin for each subclade. The first GRAMPA reconciliation identified an allopolyploid
855 origin of subclade F, arising between the MRCA of subclade E and the MRCA of
856 subtribe Phyteumatinae (Fig. 5d). The close relationship of subclade F with subclade E
857 in the plastid topologies (Fig. 3a) suggests that the MRCA of subclade E likely served
858 as the maternal parent, while the MRCA of subtribe Phyteumatinae may have acted as
859 the paternal parent. This allopolyploidization is partially consistent with the hybrid
860 origin proposed by the SNaQ analysis, designating subclade E ($\gamma = 0.728$) as the
861 maternal parent and subclade B ($\gamma = 0.272$) as the paternal parent (Fig. 5b,c).
862 Interestingly, these two methodologies inferred different paternal origins. This
863 discrepancy, marked by an uneven inheritance probability ($\gamma = 0.728$ vs. 0.272), could
864 clarify the conflicting topologies of subclade F in both the nuclear concatenated-based
865 analyses and species tree (Fig. 3b,c). As a result, we concluded that the origin of the
866 subclade F could be traced to multiple, undistinguishable hybrid swarms involving the
867 MRCA of subclade E and the MRCA of subtribe Phyteumatinae, all of which were
868 accompanied by a whole genome duplication (WGD) event. After excluding subclades
869 F and A, the origins of subclades A and C were identified as allopolyploid, arising from

870 the combination of subclades E and B, and B and D, respectively (Fig. 5e,f). Our dating
871 analysis, based on three distinct topologies depicted in Figure 6abc, supports the rapid
872 radiation hypothesis for the six subclades (A to F) highlighted in the purple-shaded area
873 of Figure 6a. This radiation possibly took place during the Middle to Late Eocene, with
874 date ranges of 42.17-36.54 Mya from the nuclear ML topology, 42.19-37.85 Mya from
875 the nuclear species tree, and 44.06-37.82 Mya from the plastid ML topology. Notably,
876 discrepancies were observed between our nuclear and plastid topologies regarding the
877 ancestral areas of the subtribe Phyteumatinae. The nuclear topologies suggest Europe &
878 SW Asia (the Ancient Mediterranean region) as the ancestral areas, while the plastid
879 topology points to a broader region encompassing Europe, SW Asia, and Southern East
880 Asia and Australasia—essentially spanning much of Southern Eurasia. Given that
881 plastid data may have a narrower genetic diversity (due to uniparental inheritance), they
882 might not capture the full range of ancestral distributions compared to nuclear data. We
883 concluded that the possible ancestral areas of the subtribe Phyteumatinae included the
884 Ancient Mediterranean region. The European and Southwest Asian regions experienced
885 pivotal transitions in both climate and tectonics. This era marked a departure from the
886 extremely warm in the early Eocene to a slightly cooler, yet still warm global
887 environment. It was also a time of significant tectonic events; the collision between the
888 African, Eurasian, and Anatolian tectonic plates continued during this period, resulting
889 in the formation of the Alpine mountain chain and uplift of mountain ranges like the
890 Taurus and the Zagros. Additionally, the closure of the Tethys Sea also had significant

891 impacts on ocean circulation patterns and climate. The extreme climate and tectonic
892 changes may have promoted the rapid diversification of the six major clades in subtribe
893 *Phyteumatinae*.

894 The notable cytonuclear inconsistency in dating analyses, ancestral area
895 reconstruction, net diversification rates analyses (Fig. 6) underscores a critical lesson:
896 relying solely on a singular genomic data source might not furnish a comprehensive or
897 precise depiction of biogeographic evolutions for potential reticulate lineages (Dong et
898 al. 2022; Liu et al. 2022). This disparity underscores the need for integrative and multi-
899 faceted approaches in phylogenomic research, ensuring a more holistic understanding
900 of evolutionary histories and events. Utilizing multiple data sources can bridge
901 knowledge gaps, validate findings, and offer richer insights into the complexities of
902 lineage diversification and adaptation.

903 *Multi-source Genomic Data for Phylogenomic Analyses of Non-model Plants*

904 Phylogenomics has revolutionized our understanding of the evolutionary
905 relationships among organisms by incorporating hundreds to thousands of nuclear
906 genes in advanced phylogenetic inference methods. However, applying phylogenomics
907 to non-model plants presents several challenges. Non-model plants display vast
908 diversity, making it difficult to obtain representative samples that encompass the entire
909 phylogenetic and geographical diversity of a particular group. Researchers from
910 specific regions may have an advantage in collecting taxon samples from their areas,

911 but conducting comprehensive species sampling, especially for cosmopolitan lineages,
912 becomes a challenge for a single laboratory. As a result, previous phylogenomic studies
913 have mainly focused on resolving the phylogenetic backbone of major lineages at the
914 family or even higher level, such as Xiang et al. (2017), Zhao et al. (2021), Huang et al.
915 (2022), Hu et al. (2023), and Zhang et al. (2023), due to the ease of conducting
916 comprehensive taxon sampling covering most major clades. However, incomplete
917 taxon sampling has substantially hindered shallow-level phylogenomic studies of non-
918 model plants, except for some ecologically and economically important lineages, such
919 as apples (Liu et al. 2022) and *Rhododendron* (Xia et al. 2022). Moreover, limited
920 sampling can lead to biased inferences and inaccurate estimation of evolutionary
921 relationships (Heath et al. 2008; Nabhan and Sarkar 2011; Young et al. 2020). Limited
922 or biased sampling can be particularly problematic when reticulation is prevalent
923 because recent hybridization can obscure accurate inference of more ancient
924 reticulation events.

925 To address these challenges, Guo et al. (2021) proposed several criteria for an
926 optimal sequencing strategy for non-model plants, which recommended generation of
927 hundreds to thousands of SCNs from a large number of samples. As Next-Generation
928 Sequencing (NGS) technologies have developed, researchers in plant taxonomy,
929 evolutionary biology, and horticulture have sequenced various types of genomic-level
930 data from different species within their respective countries. Most of these raw data
931 have been deposited in public data repositories, such as Sequence Read Archive (SRA)

932 in the National Center for Biotechnology Information (NCBI), National Genomics Data
933 Center (NGDC), and European Nucleotide Archive (ENA). Since a single sequencing
934 technology does not entirely fulfill the requirements for non-model plant
935 phylogenomics, our objective is to propose an alternative method by integrating all
936 available genomic data and employing a feasible bioinformatic pipeline. Liu et al.
937 (2021, 2022) introduced a practical and innovative approach for assembling SCN genes
938 and plastomes, as well as conducting phylogenomic discordance analyses in non-model
939 plants. The researchers identified three primary data types that proved to be particularly
940 suitable for their methodology: DGS, WGS, and RNA-Seq data. Each data type offered
941 unique advantages in contributing to a comprehensive understanding of the genomic
942 landscape and evolutionary history of non-model plants. DGS data, which were
943 generated with approximately 10× coverage, and WGS data, with higher coverage at
944 around 20×, were both derived from whole genome DNA libraries. These two
945 approaches complemented each other, with DGS providing a cost-effective means of
946 obtaining broader genomic coverage and WGS offering higher sequencing depth for
947 more accurate variant calling and detection of rare genetic variations. The combination
948 of both DGS and WGS data allowed for a robust analysis of genomic diversity and
949 evolutionary patterns within the studied plant lineages. In addition to DGS and WGS,
950 the researchers utilized RNA-Seq data, which represented the complete set of RNA
951 transcripts of expressing coding regions. Moreover, the focus on nuclear protein-coding
952 genes, particularly nuclear SCN genes, in current phylogenomic studies offered an

953 effective strategy to resolve the phylogenetic relationships of non-model plants. The
954 conserved nature of nuclear SCN gene ensures their suitability for inferring
955 evolutionary relationships across a wide range of plant species. Leveraging these genes
956 provided a stable and well-supported phylogenetic backbone for the studied lineages,
957 allowing researchers to confidently infer the evolutionary history of non-model plants.

958 Collectively, the integration of DGS, WGS, and RNA-Seq data, along with the
959 emphasis on nuclear SCN genes, has opened up new avenues of research in the
960 phylogenomics of non-model plants. This integration of multi-source genomic data
961 approach has empowered researchers to explore the rich genomic diversity and
962 evolutionary patterns of diverse plant lineages more comprehensively and accurately.
963 By leveraging the various types of available data and employing innovative
964 bioinformatic pipelines, researchers can now gain deeper insights into the evolutionary
965 relationships, historical biogeography, and functional implications of genetic variations
966 in non-model plants, ultimately advancing our understanding of plant evolution and
967 diversification. Continued advancements in sequencing technologies and analytical
968 methods are expected to further enhance the potential of phylogenomics, opening up
969 exciting possibilities for unraveling the mysteries of plant biodiversity and adaptation.

970 *An updated infra-tribal classification of the tribe Campanuleae*

971 **Campanulaceae** Juss., Gen. Pl. [Jussieu] 163 (1789), nom. cons. Type: *Campanula* L.

- 972 Subfamily **Campanuloideae** Burnett, *Outlines Bot.* 942, 1094, 1110 (1835). Type:
- 973 *Campanula* L.
- 974 Tribe **Campanuleae** Dumort., *Fl. Belg.* 58. 1827. Type: *Campanula* L.
- 975 = tribe Jasionae Dumort., *Fl. Belg.* 59. 1827. Type: *Jasione* L.
- 976 = tribe Phyteumateae Dumort., *Fl. Belg.* 59. 1827, ‘Phyteumeae’. Type: *Phyteuma* L.
- 977 = tribe Peracarpeae Fed., *Fl. URSS* 24: 471. 1957. Type: *Peracarpa* Hook.f. &
- 978 Thomson.
- 979 = tribe Michauxieae Fed., *Fl. URSS* 24: 472. 1957. Type: *Michauxia* L’Hér.
- 980 = tribe Edraiantheae Fed., *Fl. URSS* 24: 475. 1957. Type: *Edraianthus* A.DC.
- 981 = tribe Annaeae Kolak., *Bot. Zhurn. (Moscow & Leningrad)* 72(12): 1575. 1987,
- 982 ‘Annaea’. Type: *Annaea* Feer.
- 983 = tribe Azorineae Kolak., *Bot. Zhurn. (Moscow & Leningrad)* 72(12): 1575. 1987.
- 984 Type: *Azorina* Feer.
- 985 = tribe Echinocodoneae Kolak., *Bot. Zhurn. (Moscow & Leningrad)* 72(12): 1575.
- 986 1987, nom. illeg. Type: *Echinocodon* Kolak., *Soobshch. Akad. Nauk Gruz. SSR*,
- 987 121(2): 387. 1986, hom. illeg. non D.Y. Hong, *Acta Phytotax. Sin.*, 22(3): 183.
- 988 1984.
- 989 = tribe Gadellieae Kolak., *Bot. Zhurn. (Moscow & Leningrad)* 72(12): 1576. 1987.
- 990 Type: *Gadellia* Schulk.
- 991 = tribe Muehlbergelleae Kolak., *Bot. Zhurn. (Moscow & Leningrad)* 72(12): 1575.
- 992 1987. Type: *Muehlbergella* Feer.

- 993 = tribe *Musschieae* Kolak., Bot. Zhurn. (Moscow & Leningrad) 72(12): 1575. 1987.
- 994 Type: *Musschia* Dumort.
- 995 = tribe *Mzymteleae* Kolak., Bot. Zhurn. (Moscow & Leningrad) 72(12): 1578. 1987.
- 996 Type: *Mzymtella* Kolak.
- 997 = tribe *Neocodoneae* Kolak., Bot. Zhurn. (Moscow & Leningrad) 72(12): 1577. 1987.
- 998 Type: *Neocodon* Kolak. & Serdyuk.
- 999 = tribe *Sachokieleae* Kolak., Bot. Zhurn. (Moscow & Leningrad) 72(12): 1578. 1987.
- 1000 Type: *Sachokiella* Kolak.
- 1001 = tribe *Sergieae* Kolak., Bot. Zhurn. (Moscow & Leningrad) 72(12): 1577. 1987. Type:
- 1002 *Sergia* Fed.
- 1003 = tribe *Theodorovieae* Kolak., Bot. Zhurn. (Moscow & Leningrad) 72(12): 1575. 1987.
- 1004 Type: *Theodorovia* Kolak.
- 1005 = tribe *Echinococonieae* Kolak., Bot. Zhurn. (Moscow & Leningrad) 79(1): 114. 1994.
- 1006 Type: *Echinocodonia* Kolak.
- 1007 = tribe *Pseudocampanuleae* Kolak., Bot. Zhurn. (Moscow & Leningrad) 79(1): 115.
- 1008 1994. Type: *Pseudocampanula* Kolak.
- 1009 I. subtribe **Campanulinae** R.Schönland (clade II) in H.G.A. Engler & K.A.E. Prantl,
- 1010 Nat. Pflanzenfam. IV, 5(48). 1889. Type: *Campanula* L.
- 1011 Included genera: *Azorina*, *Campanula*, *Edraianthus*, *Michauxia*, *Muehlbergella*,
- 1012 *Sachokiella*, *Theodorovia*, *Trachelium*, and *Zeugandra*.

- 1013 II. subtribe **Phyteumatinae** Caruel (clade I), *Epit. Fl. Europ.* 2: 248. 1894,
1014 ‘Phyteumeae’. Type: *Phyteuma* L.
1015 Included genera: *Adenophora*, *Astrocodon*, *Asyneuma*, *Brachycodonia*,
1016 *Cryptocodon*, *Cylindrocarpa*, *Decaprisma*, *Eastwoodiella*, *Favratia*, *Githopsis*,
1017 *Hanabusaya*, *Hayekia*, *Heterocodon*, *Homocodon*, *Legousia*, *Loreia*, *Melanocalyx*,
1018 *Palustricodon*, *Peracarpa*, *Petromarula*, *Physoplexis*, *Phyteuma*, *Poolea*, *Protocodon*,
1019 *Ravenella*, *Rotantheta*, *Sergia*, *Smithiastrum*, and *Triodanis*.
1020 III. subtribe **Musschiinae** B.B.Liu (clade III), **stat. nov.** basionym: tribe Musschieae
1021 Kolak. *Bot. Zhurn. (Moscow & Leningrad)* 72: 1575. 1987. Type: *Musschia* L.
1022 Included genera: *Echinocodonia*, *Gadellia*, and *Musschia*.
1023 IV. subtribe **Jasioninae** Endl. (clade IV), *Gen. Pl. [Endlicher]*. 514. 1838, ‘Jasioneae’.
1024 Type: *Jasione* L.
1025 Included genera: *Feeria*, *Hesperocodon*, and *Jasione*.

1026 CONCLUSIONS

1027 We focused on the bellflower tribe Campanuleae, a non-model plant lineage
1028 known for its extensive history of hybridization and introgression. We presented a
1029 comprehensive and versatile framework for deciphering the network-like phylogenetic
1030 relationships within such lineages, relying on various genomic data sources and
1031 analytical methods. Our results unveiled compelling evidence supporting the role of

1032 allopolyploidization and hybridization in promoting the early diversification of the
1033 Campanuleae tribe. Notably, the rapid radiation of six major subclades in the subtribe
1034 Phyteumatinae may have been driven by multiple and continuous allopolyploidization
1035 events, taking place in the ancient Mediterranean region during the Middle to Late
1036 Eocene epochs. Furthermore, our study emphasizes a significant challenge in
1037 evolutionary biology research: conflicting tree topologies derived from different
1038 genomic datasets and phylogenetic inference methods can lead to substantial variation
1039 in our downstream estimates of the timing of evolutionary events, ancestral geographic
1040 origins, and diversification rates.

1041 SUPPLEMENTARY MATERIAL

1042 Data available from the Dryad Digital Repository:

1043 <https://doi.org/10.5061/dryad.hqbkzkh1nz>.

1044 FUNDING

1045 This work was supported by the National Natural Science Foundation of China
1046 (grant number 32270216 to B.B.L., 32000163 to B.B.L.) and the Youth Innovation
1047 Promotion Association CAS (grant number 2023086 to B.B.L.).

1048 ACKNOWLEDGEMENTS

1049 All the phylogenomic analyses have been run on the PhyloAI supercomputer,
1050 owned by B.B.L., Institute of Botany, Chinese Academy of Sciences.

1051

AUTHOR CONTRIBUTIONS

1052

B.B.L. designed the project and supervised the study. B.B.L., Z.T.J., S.Y.X., and

1053

C.X. wrote the draft manuscript. Z.T.J., S.Y.X., and Y.Z. carried out the phylogenomic

1054

analyses. C.X. performed the deep genome skimming sequencing. B.L. provided

1055

suggestions for taxonomic classification. G.N.L., H.W., X.H.L., R.G.J.H., D.K.M.,

1056

S.H.J., L.Z., C.R., and D.Y.H. provided suggestions for structuring the paper. All the

1057

authors contributed to the writing and interpreting of the results and approved the final

1058

manuscript.

1059

1060

REFERENCES

- 1061 Andrews S. 2018. FastQC: A quality control tool for high throughput sequence data.
1062 Available from: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- 1063 Baker W.J., Bailey P., Barber V., Barker A., Bellot S., Bishop D., Botigué L.R.,
1064 Brewer G., Carruthers T., Clarkson J.J., Cook J., Cowan R.S., Dodsworth S.,
1065 Epitawalage N., Françoso E., Gallego B., Johnson M.G., Kim J.T., Leempoel K.,
1066 Maurin O., McGinnie C., Pokorny L., Roy S., Stone M., Toledo E., Wickett N.J.,
1067 Zuntini A.R., Eiserhardt W.L., Kersey P.J., Leitch I.J., Forest F. 2022. A
1068 comprehensive phylogenomic platform for exploring the angiosperm tree of life.
1069 *Syst. Biol.* 71:301–319.
- 1070 Boissier P.E. 1875. *Flora Orientalis* Vol. 3. Genève and Basel: H. Georg.
- 1071 Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina
1072 sequence data. *Bioinformatics* 30:2114–2120.
- 1073 Borowiec M.L. 2016. AMAS: a fast tool for alignment manipulation and computing of
1074 summary statistics. *PeerJ*. 4:e1660.
- 1075 Borowiec M.L. 2019. Spruceup: fast and flexible identification, visualization, and
1076 removal of outliers from large multiple sequence alignments. *J. Open Source*
1077 *Softw.* 4:1635.
- 1078 Boto L. 2010. Horizontal gene transfer in evolution: facts and challenges. *Proc. Biol.*
1079 *Sci.* 277:819–827.

- 1080 Brown J.W., Walker J.F., Smith S.A. 2017. Phyx: phylogenetic tools for unix.
1081 *Bioinformatics* 33:1886–1888.
- 1082 Capella-Gutiérrez S., Silla-Martínez J.M., Gabaldón T. 2009. trimAl: a tool for
1083 automated alignment trimming in large-scale phylogenetic analyses.
1084 *Bioinformatics* 25:1972–1973.
- 1085 Cellinese N., Smith S.A., Edwards E.J., Kim S.T., Haberle R.C., Avramakis M.,
1086 Donoghue M.J. 2009. Historical biogeography of the endemic Campanulaceae of
1087 Crete. *J. Biogeogr.* 36:1253–1269.
- 1088 Chamala S., García N., Godden G.T., Krishnakumar V., Jordon-Thaden I.E., De Smet
1089 R., Barbazuk W.B., Soltis D.E., Soltis P.S. 2015. MarkerMiner 1.0: a new
1090 application for phylogenetic marker development using angiosperm
1091 transcriptomes. *Appl. Plant Sci.* 3:1400115.
- 1092 Crowl A.A., Mavrodiev E., Mansion G., Haberle R., Pistarino A., Kamari G., Phitos
1093 D., Borsch T., Cellinese N. 2014. Phylogeny of Campanuloideae
1094 (Campanulaceae) with emphasis on the utility of nuclear pentatricopeptide repeat
1095 (PPR) genes. *Plos One.* 9:e94199.
- 1096 Crowl A.A., Miles N.W., Visger C.J., Hansen K., Ayers T., Haberle R., Cellinese N.
1097 2016. A global perspective on Campanulaceae: biogeographic, genomic, and
1098 floral evolution. *Am. J. Bot.* 103:233–245.

- 1099 Crowl A.A., Myers C., Cellinese N. 2017. Embracing discordance: phylogenomic
1100 analyses provide evidence for allopolyploidy leading to cryptic diversity in a
1101 mediterranean *Campanula* (Campanulaceae) clade. *Evolution* 71:913–922.
- 1102 Dagan T., Martin W. 2009. Getting a better picture of microbial evolution en route to a
1103 network of genomes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364:2187–2196.
- 1104 Damboldt J. 1976. Materials for a flora of Turkey XXXII: Campanulaceae. *Edinb. J.*
1105 *Bot.* 35:39–52.
- 1106 Damboldt J. 1978. *Campanula* L. In: Davis P.H., Edmondson J.R., Mill R.R., Tan K.,
1107 editors. *Flora of Turkey and east Aegean islands Vol. 6*. Edinburgh: Edinburgh
1108 Univ. Press. p. 2–64.
- 1109 Darwin C.R. 1859. *On the origin of species by means of natural selection, or the*
1110 *preservation of favoured races in the struggle for life*. London: John Murray.
- 1111 De Candolle A.P., De Candolle A., Buek H.W. 1830. *Prodromus systematis naturalis*
1112 *regni vegetabilis, sive enuemeratio contracta ordinum generum specierumque*
1113 *plantarum huc usque cognitarum, juxta methodi naturalis normas digesta*. Parisii:
1114 *Sumptibus Sociorum Treuttel et Würtz*.
- 1115 Degnan J.H. 2018. Modeling hybridization under the network multispecies coalescent.
1116 *Syst. Biol.* 67:786–799.
- 1117 Degnan J.H., Rosenberg N.A. 2009. Gene tree discordance, phylogenetic inference and
1118 the multispecies coalescent. *Trends Ecol. Evol.* 24:332–340.

- 1119 Dong W.P., Li E.Z., Liu Y.L., Xu C., Wang Y.S., Liu K.J., Cui X.Y., Sun J.H., Suo
1120 Z.L., Zhang Z.X., Wen J., Zhou S.L. 2022. Phylogenomic approaches untangle
1121 early divergences and complex diversifications of the olive plant family. *BMC*
1122 *Biol.* 20:92.
- 1123 Eddie W.M.M., Shulkina T., Gaskin J., Haberle R.C., Jansen R.K. 2003. Phylogeny of
1124 *Campanulaceae* s. str. inferred from its sequences of nuclear ribosomal DNA.
1125 *Ann. Mo. Bot. Gard.* 90:554–575.
- 1126 Fedorov A. 1957. *Campanula* L. In: Komarov K., editor. *Flora of the U.R.S.S.* Vol. 24.
1127 Moscow: URSS Academy of Sciences. p. 126–501.
- 1128 Gitzendanner M.A., Soltis P.S., Yi T.S., Li D.Z., Soltis D.E. 2018. Plastome
1129 phylogenetics: 30 years of inferences into plant evolution. *Adv. Bot. Res.* 85:293–
1130 313.
- 1131 Glick L., Itay M. 2014. Chromevol: assessing the pattern of chromosome number
1132 evolution and the inference of polyploidy along a phylogeny. *Mol. Biol. Evol.*
1133 31:1914–1922.
- 1134 Gontier N. 2015. Reticulate evolution everywhere. In: Gontier N., editor. *Reticulate*
1135 *evolution, symbiogenesis, lateral gene transfer, hybridization and infectious*
1136 *heredit.* Springer: Cham, Switzerland. p. 1–40.
- 1137 Gregg W.C.T., Ather S.H., Hahn M.W. 2017. Gene-tree reconciliation with mul-trees
1138 to resolve polyploidy events. *Syst. Biol.* 66:1007–1018.

- 1139 Guo X., Fang D.M., Sahu S.K., Yang S., Guang X.M., Folk R., Smith S.A.,
1140 Chanderbali A.S., Chen S.S., Liu M., Yang T., Zhang S.Z., Liu X., Xu X., Soltis
1141 P.S., Soltis D.E., Liu H. 2021. Chloranthus genome provides insights into the
1142 early diversification of angiosperms. *Nat. Commun.* 12:6930.
- 1143 He J., Lyu R., Luo Y.K., Xiao J.M., Xie L., Wen J., Li W.H., Pei L.Y., Cheng J. 2022.
1144 A phylotranscriptome study using silica gel-dried leaf tissues produces an updated
1145 robust phylogeny of Ranunculaceae. *Mol. Phylogenet. Evol.* 174:107545.
- 1146 Heath T.A, Hedtke S.M., Hillis D.M. 2008. Taxon sampling and accuracy of
1147 phylogenetic analyses. *J. Syst. Evol.* 46:239–257.
- 1148 Hong D.Y., Wang Q. 2015. A new taxonomic system of the Campanulaceae s.s. *J.*
1149 *Syst. Evol.* 53:203–209.
- 1150 Hu H.Y., Sun P.C., Yang Y.Z., Ma J.X., Liu J.Q. 2023. Genome-scale angiosperm
1151 phylogenies based on nuclear, plastome, and mitochondrial datasets. *J. Integr.*
1152 *Plant Biol.* 65:1479–1489.
- 1153 Huang W.C., Zhang L., Columbus J.T., Hu Y., Zhao Y.Y., Tang L., Guo Z.H., Chen
1154 W.L., McKain M., Bartlett M., Huang C.-H., Li D.Z., Ge S., Ma H. 2022. A well-
1155 supported nuclear phylogeny of Poaceae and implications for the evolution of C₄
1156 photosynthesis. *Mol. Plant* 15:755–777.
- 1157 Huson D.H., Bryant D. 2006. Application of phylogenetic networks in evolutionary
1158 studies. *Mol. Biol. Evol.* 23:254–267.

1159 Jarvis E.D., Mirarab S., Aberer A.J., Li B., Houde P., Li C., Ho S.Y.W., Faircloth B.C.,
1160 Nabholz B., Howard J.T., Suh A., Weber C.C., da Fonseca R.R., Li J.W., Zhang
1161 F., Li H., Zhou L., Narula N., Liu L., Ganapathy G., Boussau B., Bayzid M.S.,
1162 Zavidovych V., Subramanian S., Gabaldón T., Capella-Gutiérrez S., Huerta-Cepas
1163 J., Rekepalli B., Munch K., Schierup M., Lindow B., Warren W.C., Ray D., Green
1164 R.E., Bruford M.W., Zhan X.J., Dixon A., Li S.B., Li N., Huang Y.H., Derryberry
1165 E.P., Bertelsen M.F., Sheldon F.H., Brumfield R.T., Mello C.V., Lovell P.V.,
1166 Wirthlin M., Schneider M.P.C., Prosdocimi F., Samaniego J.A., Vargas Velazquez
1167 A.M., Alfaro-Núñez A., Campos P.F., Petersen B., Sicheritz-Ponten T., Pas A.,
1168 Bailey T., Scofield P., Bunce M., Lambert D.M., Zhou Q., Perelman P., Driskell
1169 A.C., Shapiro B., Xiong Z.J., Zeng Y.L., Liu S.P., Li Z.Y., Liu B.H., Wu K., Xiao
1170 J., Xiong Y.Q., Zheng Q.M., Zhang Y., Yang H.M., Wang J., Smeds L., Rheindt
1171 F.E., Braun M., Fjeldsa J., Orlando L., Barker F.K., Jønsson K.A., Johnson W.,
1172 Koepfli K.-P., O'Brien S., Haussler D., Ryder O.A., Rahbek C., Willerslev E.,
1173 Graves G.R., Glenn T.C., McCormack J., Burt D., Ellegren H., Alström P.,
1174 Edwards S.V., Stamatakis A., Mindell D.P., Cracraft J., Braun E.L., Warnow T.,
1175 Jun W., Gilbert M.T.P., Zhang G.J. 2014. Whole-genome analyses resolve early
1176 branches in the tree of life of modern birds. *Science* 346:1320–1331.
1177 Jin Z.T., Richard G.J.H., Ma D.K., Wang H., Liu G.N., Ren C., Ge B.J., Fan Q., Jin
1178 S.H., Xu C., Wu J., Liu B.B. 2023. Nightmare or delight: taxonomic

- 1179 circumscription meets reticulate evolution in the phylogenomic era. *Mol.*
1180 *Phylogenet. Evol.* 189:107914.
- 1181 Johnson M.G., Gardner E.M., Liu Y., Medina R., Goffinet B., Shaw A.J., Zerega
1182 N.J.C., Wickett N.J. 2016. HybPiper: Extracting coding sequence and introns for
1183 phylogenetics from high-throughput sequencing reads using target enrichment.
1184 *Appl. Plant Sci.* 4:1600016.
- 1185 Jones K.E., Korotkova N., Petersen J., Henning T., Borsch T., Kilian N. 2017.
1186 Dynamic diversification history with rate upshifts in Holarctic bell-flowers
1187 (*Campanula* and allies). *Cladistics* 33:637–666.
- 1188 Kapli P., Yang Z.H., Telford M.J. 2020. Phylogenetic tree building in the genomic age.
1189 *Nat. Rev. Genet.* 21:428–444.
- 1190 Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S.,
1191 Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P.,
1192 Drummond A. 2012. Geneious basic: an integrated and extendable desktop
1193 software platform for the organization and analysis of sequence data.
1194 *Bioinformatics* 28:1647–1649.
- 1195 Keeling P.J., Palmer J.D. 2008. Horizontal gene transfer in eukaryotic evolution. *Nat.*
1196 *Rev. Genet.* 9:605–618.
- 1197 Kurland C.G., Canback B., Berg O.G. 2003. Horizontal gene transfer: a critical view.
1198 *Proc. Natl. Acad. Sci. U. S. A.* 100:9658–9662.

- 1199 Lammers T.G. 2007a. World checklist and bibliography of Campanulaceae. Kew,
1200 Richmond: Royal Botanic Gardens.
- 1201 Lammers T.G. 2007b. Campanulaceae. In: Kadereit J.W., Jeffrey C., editors. The
1202 families and genera of vascular plants. Vol. VIII. Flowering plants. Eudicots.
1203 Asterales. Springer Berlin: Heidelberg.
- 1204 Łańcucka-Środoniowa M. 1979. Macroscopic plant remains from the freshwater
1205 miocene of the Nowy Sącz Basin (West Carpathians, Poland). *Acta Palaeobot.*
1206 20:3–117.
- 1207 Lanfear R., Frandsen P.B., Wright A.M., Senfeld T., Calcott B. 2016. PartitionFinder
1208 2: new methods for selecting partitioned models of evolution for molecular and
1209 morphological phylogenetic analyses. *Mol. Biol. Evol.* 34:772–773.
- 1210 Leebens-Mack J.H., Barker M.S., Carpenter E.J., Deyholos M.K., Gitzendanner M.A.,
1211 Graham S.W., Grosse I., Li Z., Melkonian M., Mirarab S., Porsch M., Quint M.,
1212 Rensing S.A., Soltis D.E., Soltis P.S., Stevenson D.W., Ullrich K.K., Wickett
1213 N.J., DeGironimo L., Edger P.P., Jordon-Thaden I.E., Joya S., Liu T., Melkonian
1214 B., Miles N.W., Pokorny L., Quigley C., Thomas P., Villarreal J.C., Augustin
1215 M.M., Barrett M.D., Baucom R.S., Beerling D.J., Benstein R.M., Biffin E.,
1216 Brockington S.F., Burge D.O., Burris J.N., Burris K.P., Burtet-Sarramegna V.,
1217 Caicedo A.L., Cannon S.B., Çebi Z., Chang Y., Chater C., Cheeseman J.M., Chen
1218 T., Clarke N.D., Clayton H., Covshoff S., Crandall-Stotler B.J., Cross H.,
1219 dePamphilis C.W., Der J.P., Determann R., Dickson R.C., Di Stilio V.S., Ellis S.,

1220 Fast E., Feja N., Field K.J., Filatov D.A., Finnegan P.M., Floyd S.K., Fogliani B.,
1221 García N., Gâteblé G., Godden G.T., Goh F., Greiner S., Harkess A., Heaney
1222 J.M., Helliwell K.E., Heyduk K., Hibberd J.M., Hodel R.G.J., Hollingsworth P.M.,
1223 Johnson M.T.J., Jost R., Joyce B., Kapralov M.V., Kazamia E., Kellogg E.A.,
1224 Koch M.A., Von Konrat M., Könyves K., Kutchan T.M., Lam V., Larsson A.,
1225 Leitch A.R., Lentz R., Li F.-W., Lowe A.J., Ludwig M., Manos P.S., Mavrodiev
1226 E., McCormick M.K., McKain M., McLellan T., McNeal J.R., Miller R.E., Nelson
1227 M.N., Peng Y., Ralph P., Real D., Riggins C.W., Ruhsam M., Sage R.F., Sakai
1228 A.K., Scascitella M., Schilling E.E., Schlösser E.-M., Sederoff H., Servick S.,
1229 Sessa E.B., Shaw A.J., Shaw S.W., Sigel E.M., Skema C., Smith A.G., Smithson
1230 A., Stewart C.N., Stinchcombe J.R., Szövényi P., Tate J.A., Tiebel H., Trapnell
1231 D., Villegente M., Wang C.-N., Weller S.G., Wenzel M., Weststrand S.,
1232 Westwood J.H., Whigham D.F., Wu S.X., Wulff A.S., Yang Y., Zhu D., Zhuang
1233 C.L., Zuidof J., Chase M.W., Pires J.C., Rothfels C.J., Yu J., Chen C., Chen L.,
1234 Cheng S.F., Li J.J., Li R., Li X., Lu H.R., Ou Y.X., Sun X., Tan X.M., Tang J.B.,
1235 Tian Z.J., Wang F., Wang J., Wei X.F., Xu X., Yan Z.X., Yang F., Zhong X.N.,
1236 Zhou F.Y., Zhu Y., Zhang Y., Ayyampalayam S., Barkman T.J., Nguyen N.-p.,
1237 Matasci N., Nelson D.R., Sayyari E., Wafula E.K., Walls R.L., Warnow T., An
1238 H., Arrigo N., Baniaga A.E., Galuska S., Jorgensen S.A., Kidder T.I., Kong H.H.,
1239 Lu-Irving P., Marx H.E., Qi X.S., Reardon C.R., Sutherland B.L., Tiley G.P.,
1240 Welles S.R., Yu R.P., Zhan S., Gramzow L., Theißen G., Wong G.K.-S. 2019 One

- 1241 thousand plant transcriptomes and the phylogenomics of green plants. *Nature*
1242 574:679–685.
- 1243 Li C.J., Wang R.N., Li D.Z. 2020. Comparative analysis of plastid genomes within the
1244 Campanulaceae and phylogenetic implications. *Plos One*. 15:e0233167.
- 1245 Li H.T., Yi T.S., Gao L.M., Ma P.F., Zhang T., Yang J.B., Gitzendanner M.A., Fritsch
1246 P.W., Cai J., Luo Y., Wang H., van der Bank M., Zhang S.D., Wang Q.F., Wang
1247 J., Zhang Z.R., Fu C.N., Yang J., Hollingsworth P.M., Chase M.W., Soltis D.E.,
1248 Soltis P.S., Li D.Z. 2019. Origin of angiosperms and the puzzle of the Jurassic
1249 gap. *Nat. Plants*. 5:461–470.
- 1250 Li J.L., Wang S., Yu J., Wang L., Zhou S.L. 2013. A modified CTAB protocol for
1251 plant DNA extraction. *Chin. Bull. Bot.* 48:72–78.
- 1252 Liu B.B., Ma Z.Y., Ren C., Hodel R.G.J., Sun M., Liu X.Q., Liu G.N., Hong D.Y.,
1253 Zimmer E.A., Wen J. 2021. Capturing single-copy nuclear genes, organellar
1254 genomes, and nuclear ribosomal DNA from deep genome skimming data for plant
1255 phylogenetics: A case study in Vitaceae. *J. Syst. Evol.* 59:1124–1138.
- 1256 Liu B.B., Ren C., Kwak M., Hodel R.G.J., Xu C., He J., Zhou W.B., Huang C.H., Ma
1257 H., Qian G.Z., Hong D.Y., Wen J. 2022. Phylogenomic conflict analyses in the
1258 apple genus *Malus* s.l. reveal widespread hybridization and allopolyploidy driving
1259 diversification, with insights into the complex biogeographic history in the
1260 Northern Hemisphere. *J. Integr. Plant Biol.* 64:1020–1043.

- 1261 Liu L., Yu L.L. 2010. Phybase: an R package for species tree analysis. *Bioinformatics*
1262 26:962–963.
- 1263 Mai U., Mirarab S. 2018. TreeShrink: fast and accurate detection of outlier long
1264 branches in collections of phylogenetic trees. *BMC Genomics* 19:272.
- 1265 Mallet J., Besansky N., Hahn M.W. 2016. How reticulated are species? *Bioessays*
1266 38:140–149.
- 1267 Mansion G., Parolly G., Crowl A.A., Mavrodiev E., Cellinese N., Oganessian M.,
1268 Fraunhofer K., Kamari G., Phitos D., Haberle R., Akaydin G., Ikinici N., Raus T.,
1269 Borsch T. 2012. How to handle speciose clades? mass taxon-sampling as a
1270 strategy towards illuminating the natural history of *Campanula*
1271 (*Campanuloideae*). *PLoS One*. 7:e50076.
- 1272 Matzke N.J. 2018. BioGeoBEARS: BioGeography with Bayesian (and likelihood)
1273 evolutionary analysis with R scripts. version 1.1.1. Available from:
1274 <http://dx.doi.org/10.5281/zenodo.1478250>
- 1275 Mindell D.P. 2013. The tree of life: metaphor, model, and heuristic device. *Syst. Biol.*
1276 62:479–489.
- 1277 Minh B.Q., Schmidt H.A., Chernomor O., Schrempf D., Woodhams M.D., Von
1278 Haeseler A., Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for
1279 phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37:1530–1534.
- 1280 Morales-Briones D.F., Kadereit G., Tefarikis D.T., Moore M.J., Smith S.A.,
1281 Brockington S.F., Timoneda A., Yim W.C., Cushman J.C., Yang Y. 2021.

- 1282 Disentangling sources of gene tree discordance in phylogenomic data sets: testing
1283 ancient hybridizations in *Amaranthaceae* s.l. *Syst. Biol.* 70:219–235.
- 1284 Morales-Briones D.F., Gehrke B., Huang C.-H., Liston A., Ma H., Marx H.E., Tank
1285 D.C., Yang Y. 2022. Analysis of paralogs in target enrichment data pinpoints
1286 multiple ancient polyploidy events in *Alchemilla* s.l. (*Rosaceae*). *Syst. Biol.*
1287 71:190–207.
- 1288 Nabhan A.R., Sarkar I.N. 2011. The impact of taxon sampling on phylogenetic
1289 inference: a review of two decades of controversy. *Brief Bioinform.* 13:122–134.
- 1290 Nakamura T., Yamada K.D., Tomii K., Katoh K. 2018. Parallelization of MAFFT for
1291 large-scale multiple sequence alignments. *Bioinformatics* 34:2490–2492.
- 1292 Nemcok M., Pospisil L., Lexa J., Donelick R.A. 1998. Tertiary subduction and slab
1293 break-off model of the Carpathian-Pannonian region. *Tectonophysics* 295:307–
1294 340.
- 1295 Olesen J.M., Alarcon M., Ehlers B.K., Aldasoro J.J., Roquet C. 2012. Pollination,
1296 biogeography and phylogeny of oceanic island bellflowers (*Campanulaceae*).
1297 *Perspect. Plant Ecol. Evol. Syst.* 14:169–182.
- 1298 Oszast J., Stuchlik L. 1977. The Neogene vegetation of the Podhale (West Carpathians,
1299 Poland). *Acta Palaeobot.* 18:45–86.
- 1300 Pease J.B., Brown J.W., Walker J.F., Hinchliff C.E., Smith S.A. 2018. Quartet
1301 sampling distinguishes lack of support from conflicting support in the green plant
1302 tree of life. *Am. J. Bot.* 105:385–403.

- 1303 Rabosky D.L., Donnellan S.C., Grundler M., Lovette I.J. 2014a. Analysis and
1304 visualization of complex macroevolutionary dynamics: an example from
1305 australian scincid lizards. *Syst. Biol.* 63:610–627.
- 1306 Rabosky D.L., Grundler M., Anderson C., Title P., Shi J.J., Brown J.W., Huang H.,
1307 Larson J.G. 2014b. BAMMtools: an R package for the analysis of evolutionary
1308 dynamics on phylogenetic trees. *Methods Ecol. Evol.* 5:701–707.
- 1309 Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Posterior
1310 summarization in bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* 67:901–
1311 904.
- 1312 Ronquist F., Teslenko M., van der Mark P., Ayres D.L., Darling A., Höhna S., Larget
1313 B., Liu L., Suchard M.A., Huelsenbeck J.P. 2012. MrBayes 3.2: efficient bayesian
1314 phylogenetic inference and model choice across a large model space. *Syst. Biol.*
1315 61:539–542.
- 1316 Rothfels C.J. 2021. Polyploid phylogenetics. *New Phytol.* 230:66–72.
- 1317 Salichos L., Stamatakis A., Rokas A. 2014. Novel information theory-based measures
1318 for quantifying incongruence among phylogenetic trees. *Mol. Biol. Evol.*
1319 31:1261–1271.
- 1320 Smith S.A., Moore M.J., Brown J.W., Yang Y. 2015. Analysis of phylogenomic
1321 datasets reveals conflict, concordance, and gene duplications with examples from
1322 animals and plants. *BMC Evol. Biol.* 15:150.

- 1323 Solís-Lemus C., Ané C. 2016. Inferring phylogenetic networks with maximum
1324 pseudolikelihood under incomplete lineage sorting. *PLoS Genet.* 12:e1005896.
- 1325 Solís-Lemus C., Bastide P., Ané C. 2017. PhyloNetworks: a package for phylogenetic
1326 networks. *Mol. Biol. Evol.* 34:3292–3298.
- 1327 Som A. 2015. Causes, consequences and solutions of phylogenetic incongruence.
1328 *Brief. Bioinform.* 16:536–548.
- 1329 Stamatakis A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic
1330 analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
- 1331 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-
1332 analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
- 1333 Steenwyk J.L., Li Y.N., Zhou X.F., Shen X.X., Rokas A. 2023. Incongruence in the
1334 phylogenomics era. *Nat. Rev. Genet.* Online 10.1038/s41576-023-00620-x.
- 1335 Stull G.W., Qu X.J., Parins-Fukuchi C., Yang Y.Y., Yang J.B., Yang Z.Y., Hu Y., Ma
1336 H., Soltis P.S., Soltis D.E., Li D.Z., Smith S.A., Yi T.S. 2021. Gene duplications
1337 and phylogenomic conflict underlie major pulses of phenotypic evolution in
1338 gymnosperms. *Nat. Plants* 7:1015–1025.
- 1339 Stull G.W., Pham K.K., Soltis P.S., Soltis D.E. 2023. Deep reticulation: the long legacy
1340 of hybridization in vascular plant evolution. *Plant J.* 114:743–766.
- 1341 Sukumaran J., Holder M.T. 2010. DendroPy: a python library for phylogenetic
1342 computing. *Bioinformatics* 26:1569–1571.

- 1343 Thomas G.W.S., Ather S.H., Hahn M.W. 2017. Gene-tree reconciliation with MUL-
1344 trees to resolve polyploidy events. *Syst. Biol.* 66:1007–1018.
- 1345 Upham N.S., Esselstyn J.A., Jetz W. 2019. Inferring the mammal tree: species-level
1346 sets of phylogenies for questions in ecology, evolution, and conservation. *PLoS*
1347 *Biol.* 17:e3000494.
- 1348 Weitemier K., Straub S.C.K., Cronn R.C., Fishbein M., Schmickl R., McDonnell A.,
1349 Liston A. 2014. Hyb-Seq: combining target enrichment and genome skimming for
1350 plant phylogenomics. *Appl. Plant Sci.* 2:1400042.
- 1351 Wen D.J., Yu Y., Zhu J.F., Nakhleh L. 2018. Inferring phylogenetic networks using
1352 phyloNet. *Syst. Biol.* 67:735–740.
- 1353 Wendel J.F., Doyle J.J. 1998. Phylogenetic incongruence: window into genome history
1354 and molecular evolution. In: Soltis D.E., Soltis P.S., Doyle J.J. editors. *Molecular*
1355 *systematics of plants II: DNA sequencing*. Dordrecht: Kluwer Academic
1356 Publishing. p. 265–296.
- 1357 Xia X.M., Yang M.Q., Li C.L., Huang S.X., Jin W.T., Shen T.T., Wang F., Li X.H.,
1358 Yoichi W., Zhang L.H., Zheng Y.R., Wang X.Q. 2022. Spatiotemporal evolution
1359 of the global species diversity of *Rhododendron*. *Mol. Biol. Evol.* 39:314.
- 1360 Xiang Y.Z., Huang C.-H., Hu Y., Wen J., Li S.S., Yi T.S., Chen H.Y., Xiang J., Ma H.,
1361 2017. Evolution of rosaceae fruit types based on nuclear phylogeny in the context
1362 of geological times and genome duplication. *Mol. Biol. Evol.* 34:262–281.

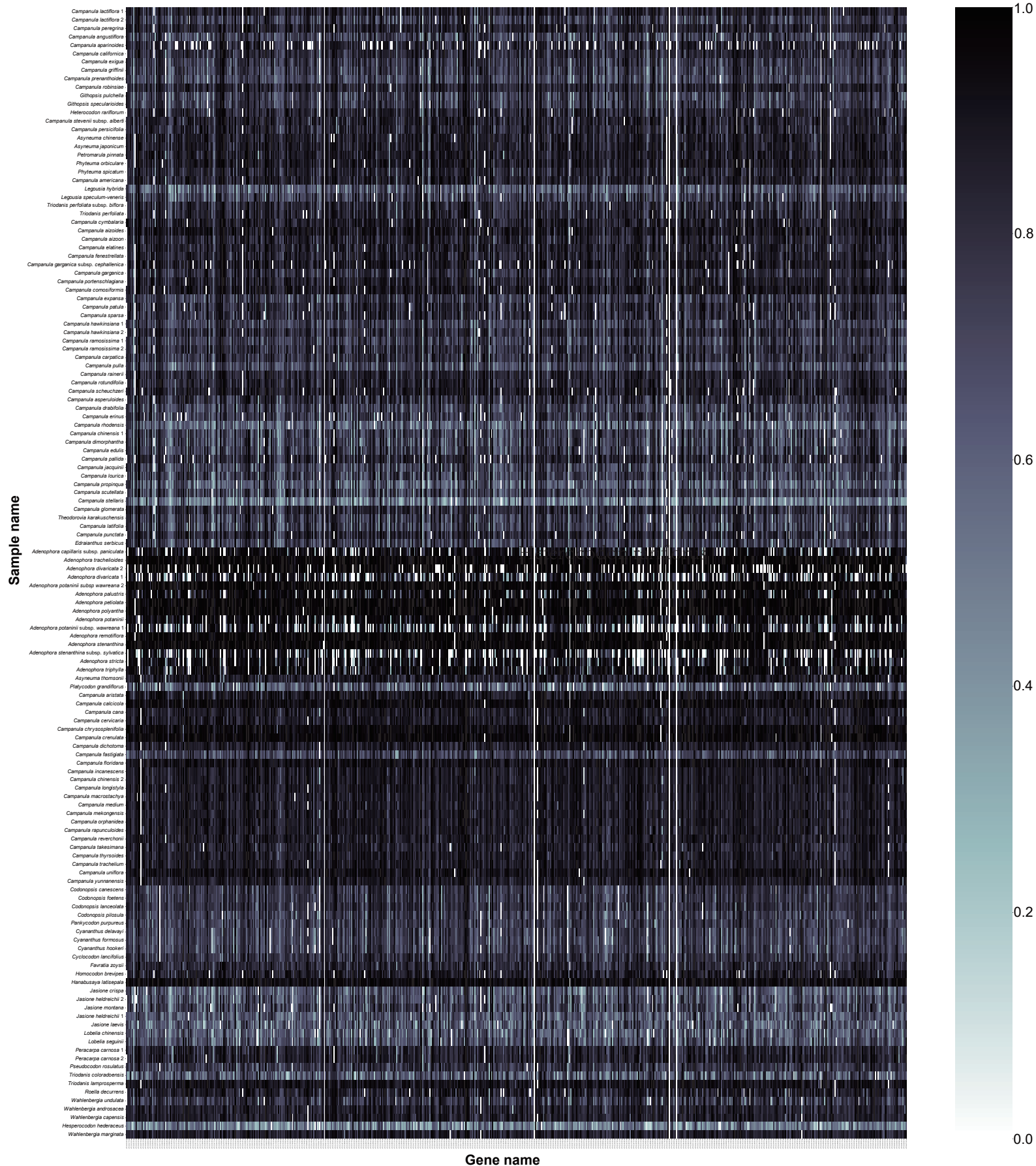
- 1363 Xu C., Hong D.Y. 2021. Phylogenetic analyses confirm polyphyly of the genus
1364 *Campanula* (Campanulaceae s. str.), leading to a proposal for generic reappraisal.
1365 J. Syst. Evol. 59:475–489.
- 1366 Yang Y., Smith S.A. 2014. Orthology inference in nonmodel organisms using
1367 transcriptomes and low-coverage genomes: improving accuracy and matrix
1368 occupancy for phylogenomics. Mol. Biol. Evol. 31:3081–3092.
- 1369 Yoo K.O., Crowl A.A., Kim K.A., Cheon K.S., Cellinese N. 2018. Origins of east asian
1370 Campanuloideae (Campanulaceae) diversity. Mol. Phylogenet. Evol. 127:468–
1371 474.
- 1372 Young A.D., Gillung J.P. 2020. Phylogenomics — principles, opportunities and pitfalls
1373 of big-data phylogenetics. Syst. Entomol. 45:225–247.
- 1374 Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: polynomial time
1375 species tree reconstruction from partially resolved gene trees. BMC bioinformatics
1376 19:153.
- 1377 Zhang G.J., Hu Y., Huang M.Z., Huang W.C., Liu D.K., Zhang D.Y., Hu H.H.,
1378 Downing J.L., Liu Z.J., Ma H. 2023. Comprehensive phylogenetic analyses of
1379 Orchidaceae using nuclear genes and evolutionary insights into epiphytism. J.
1380 Integr. Plant Biol. 65:1204–1225.
- 1381 Zhao Y.Y., Zhang R., Jiang K.W., Qi J., Hu Y., Guo J., Zhu R.B., Zhang T.K., Egan
1382 A.N., Yi T.S., Huang C.-H., Ma H. 2021. Nuclear phylotranscriptomics and
1383 phylogenomics support numerous polyploidization events and hypotheses for the

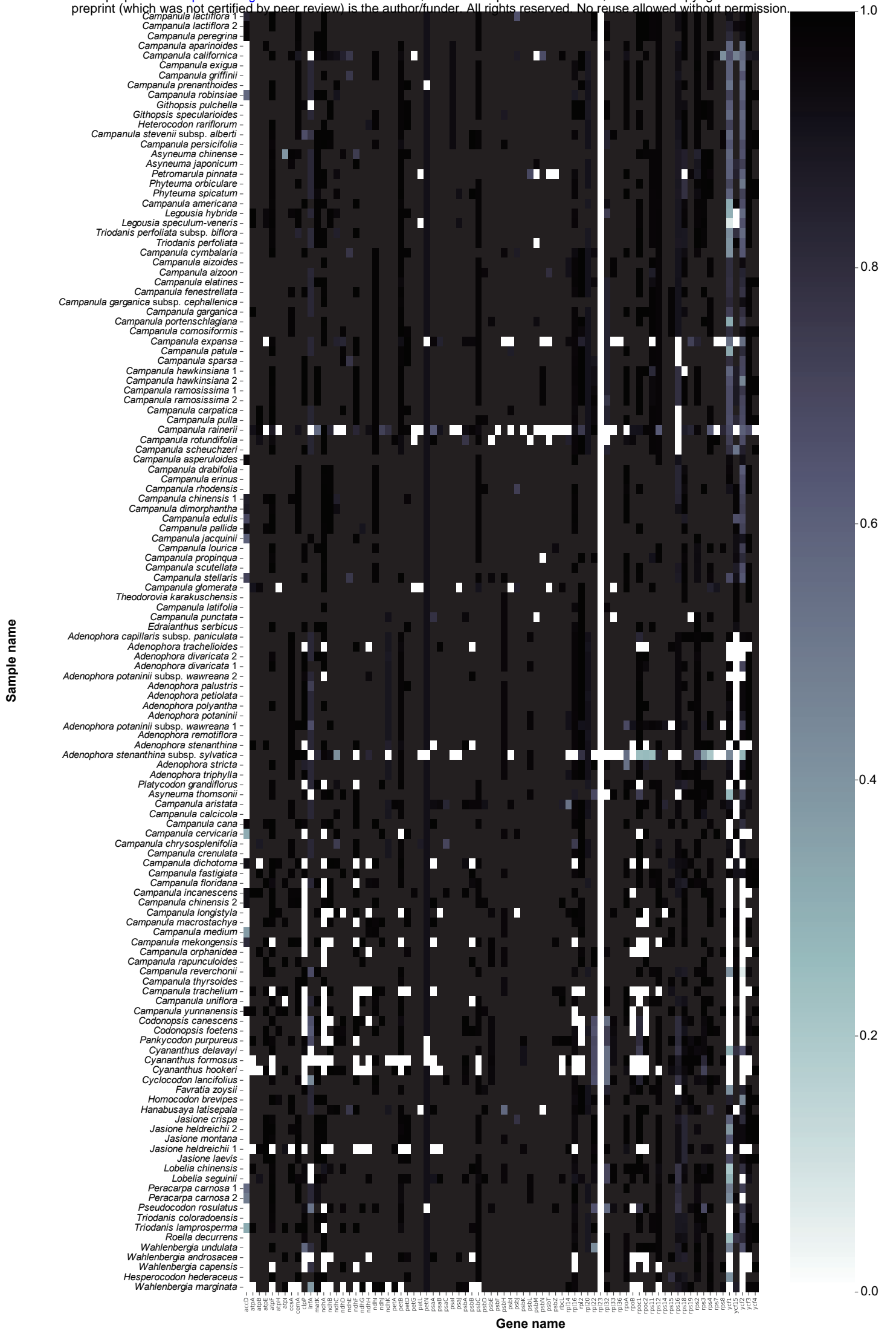
1384 evolution of rhizobial nitrogen-fixing symbiosis in Fabaceae. *Mol. Plant* 14:748–

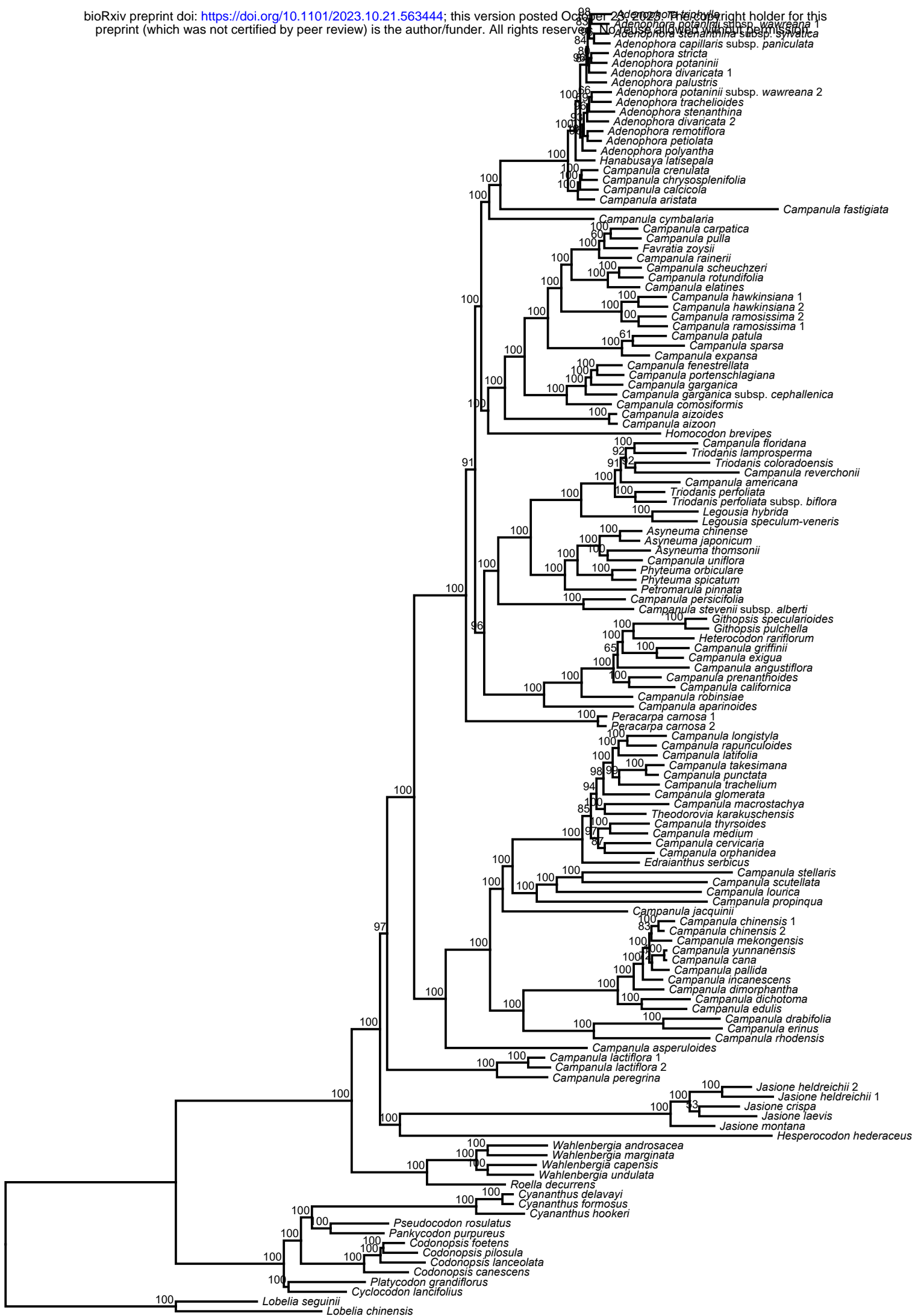
1385 773.

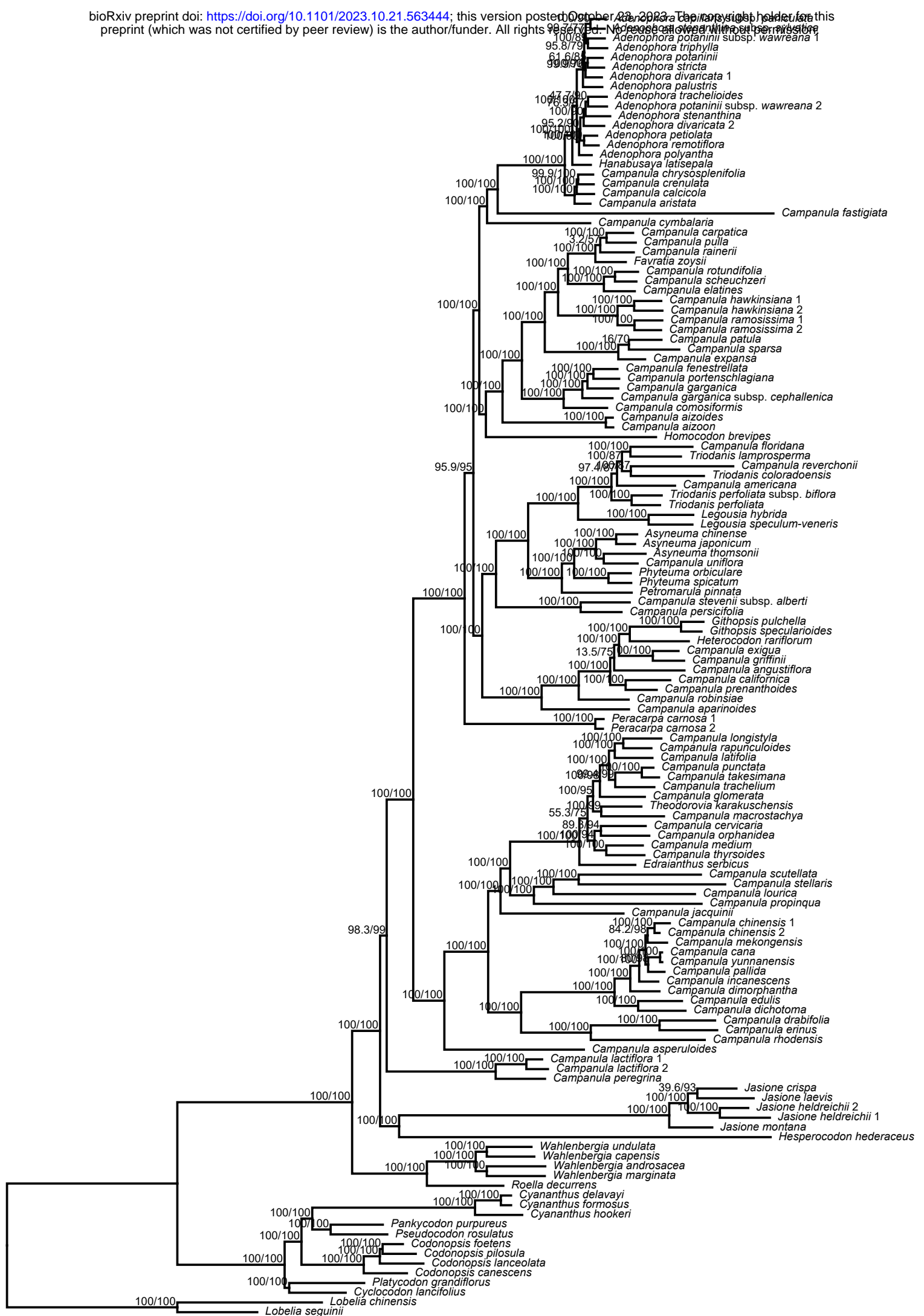
1386

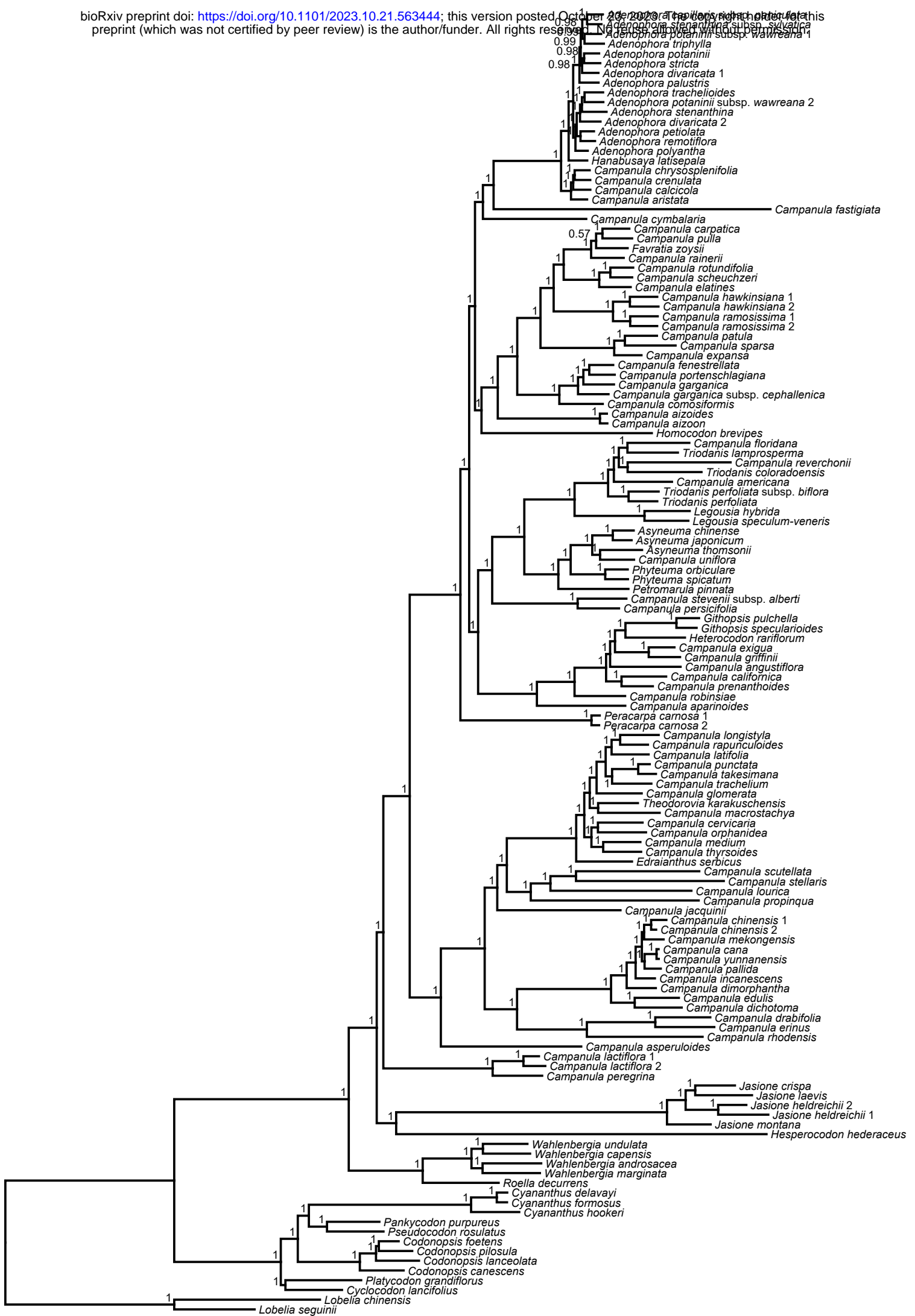
Percentage length recovery for each gene, relative to mean of targetfile references

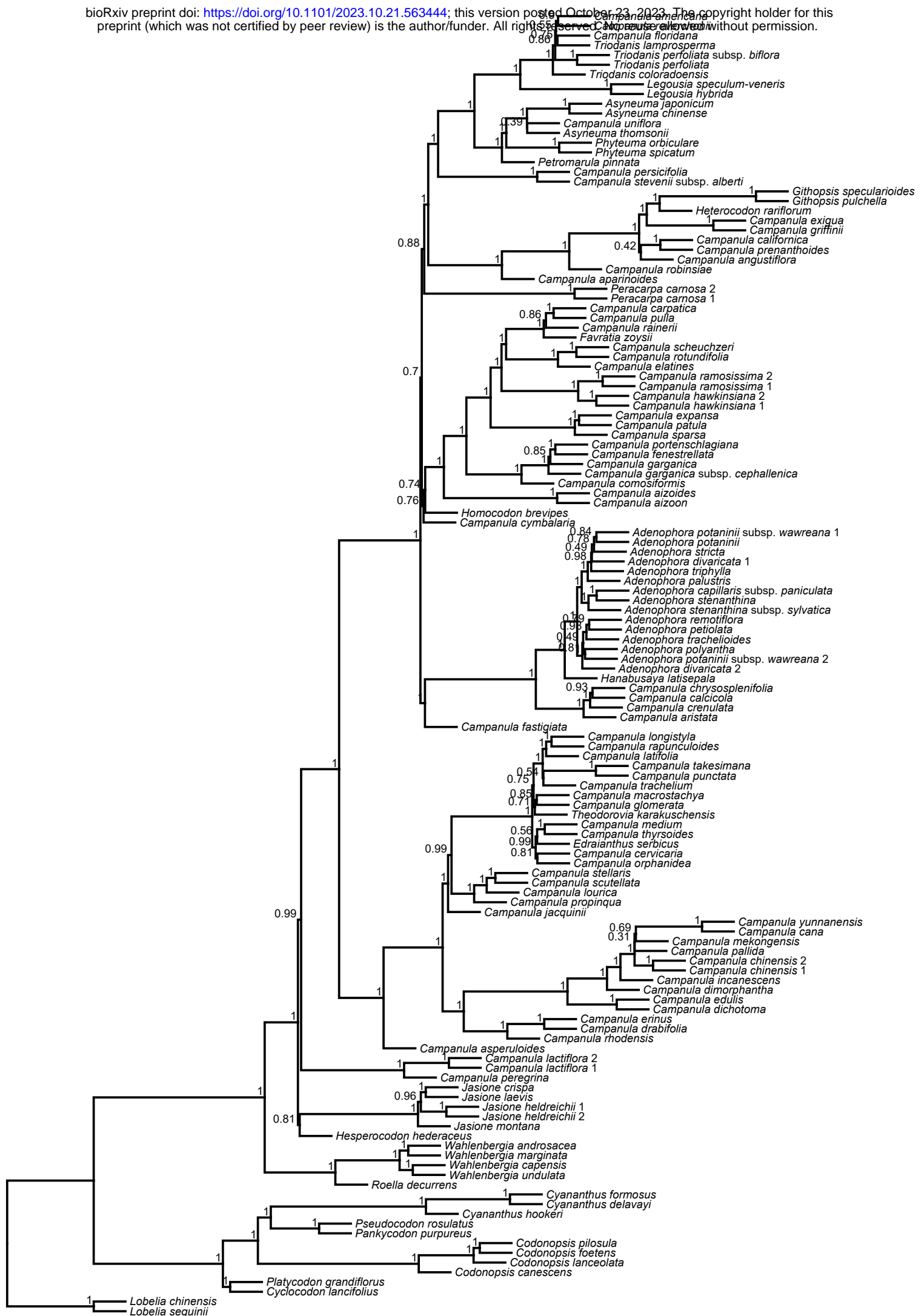


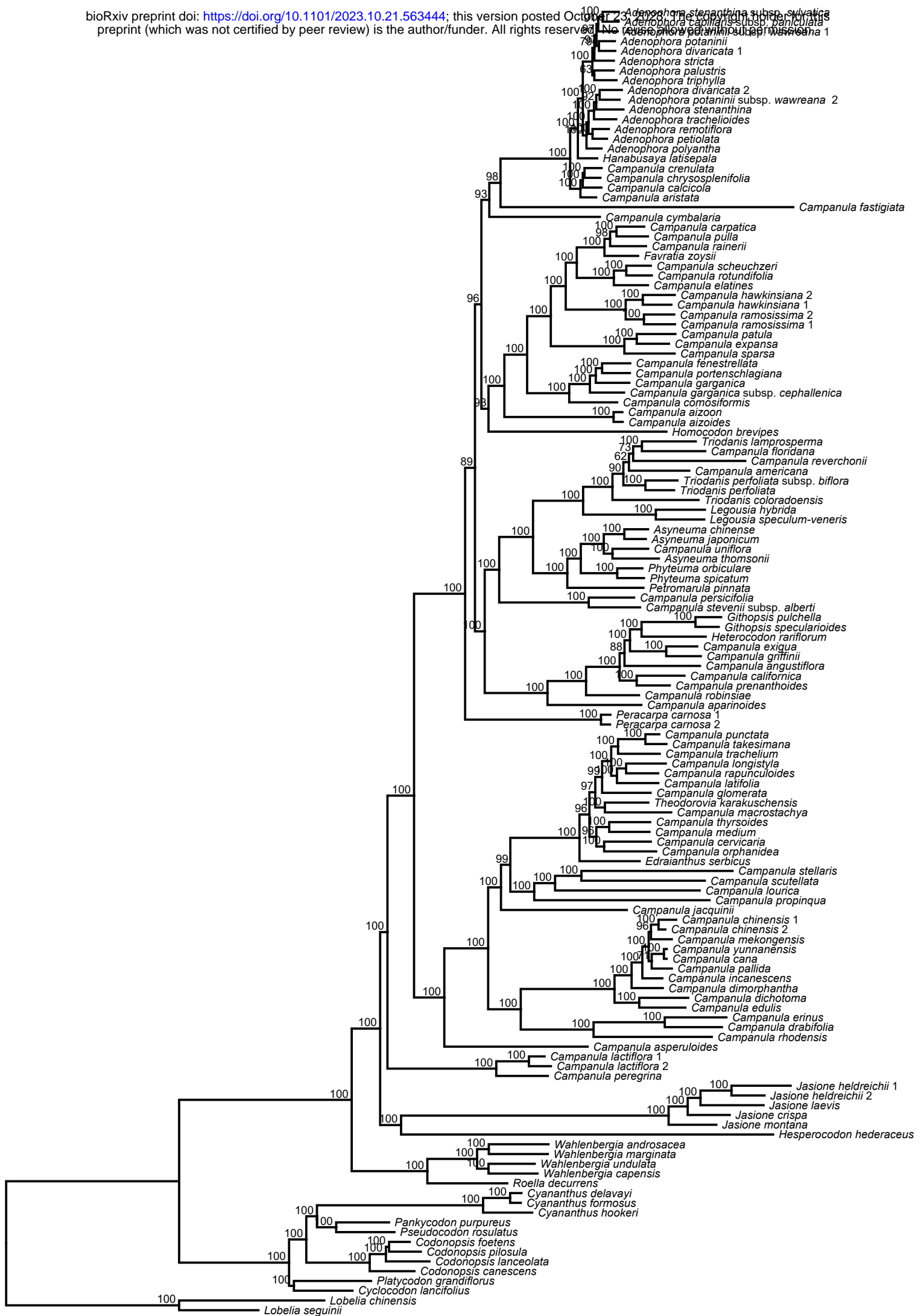


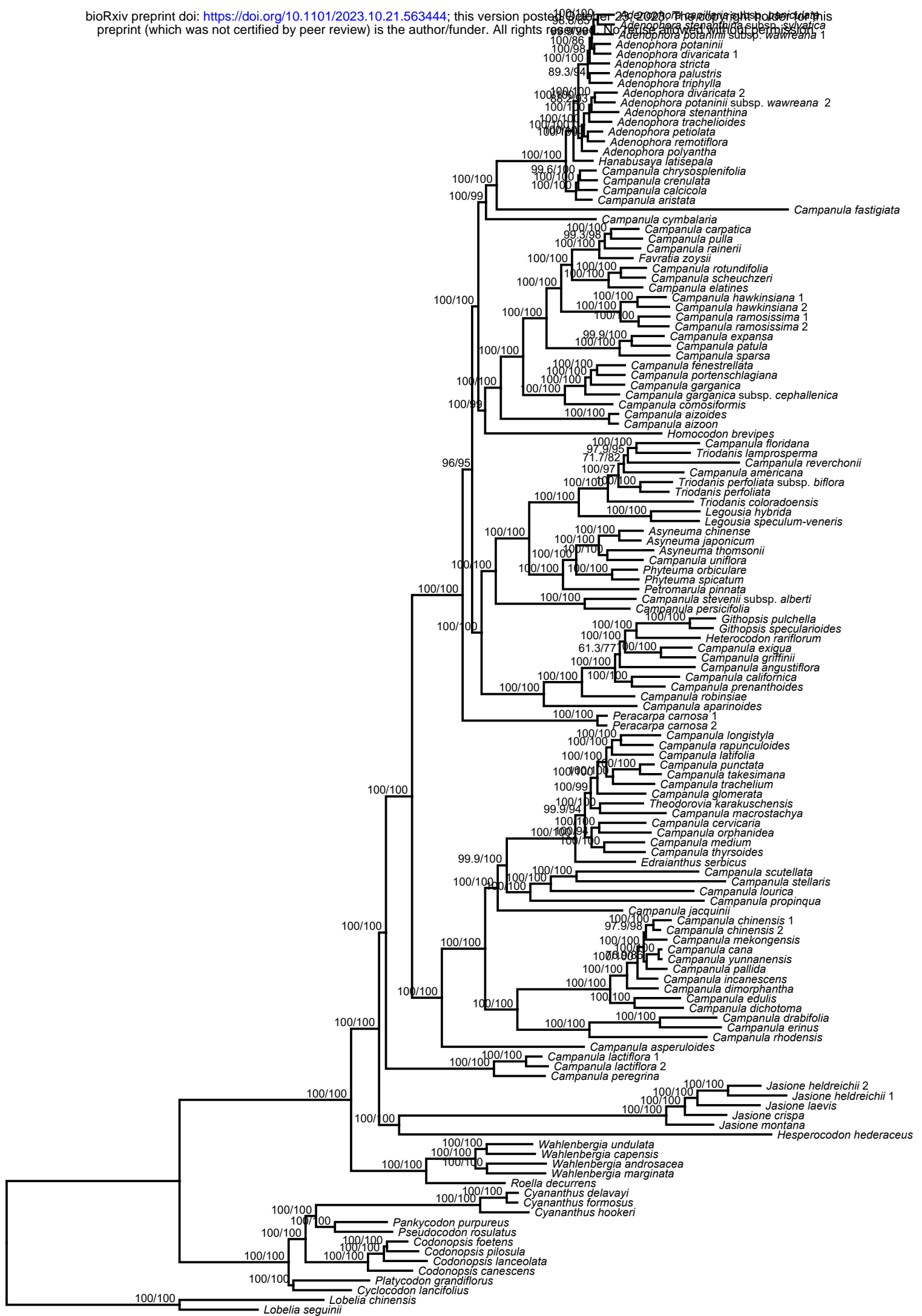


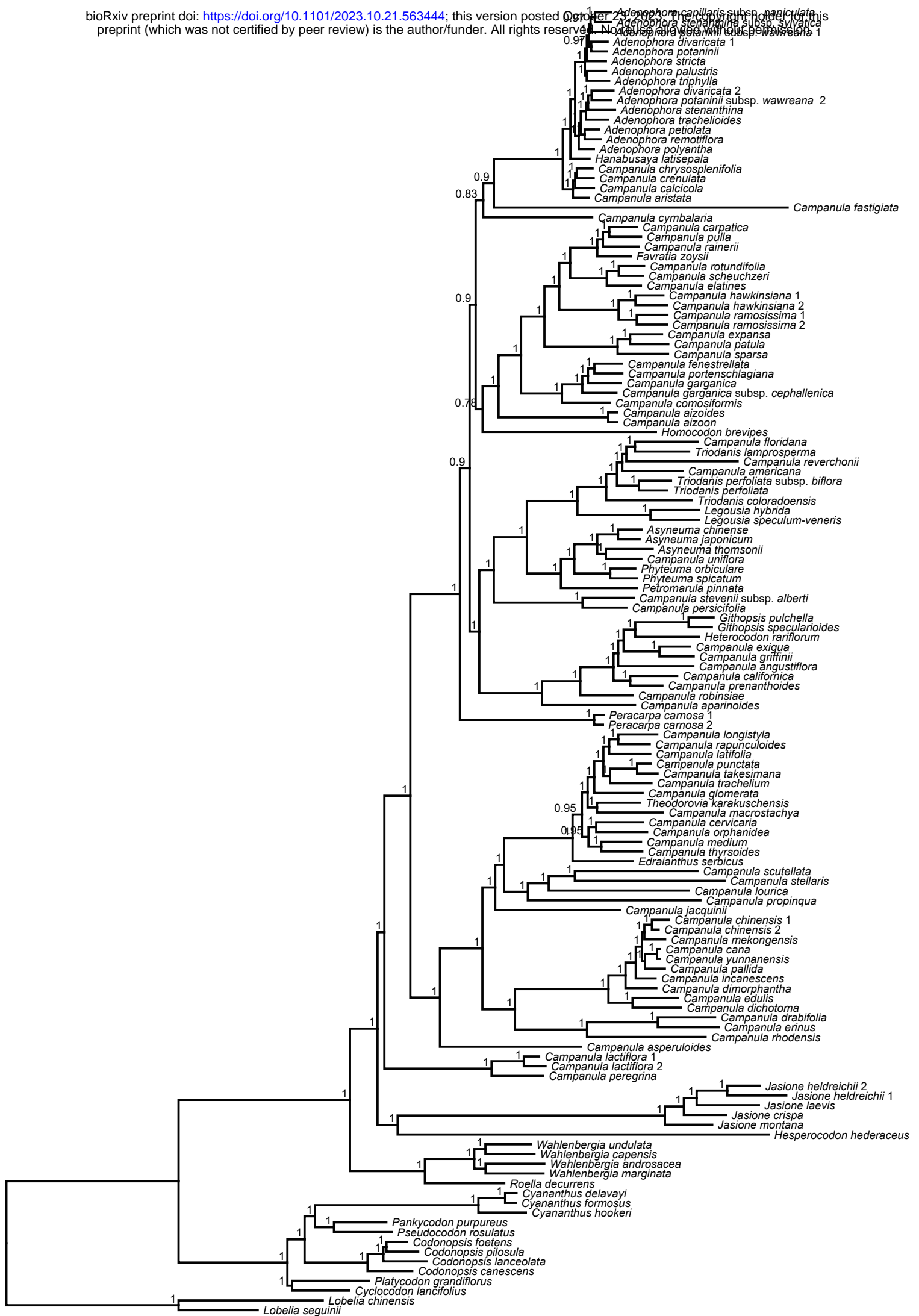


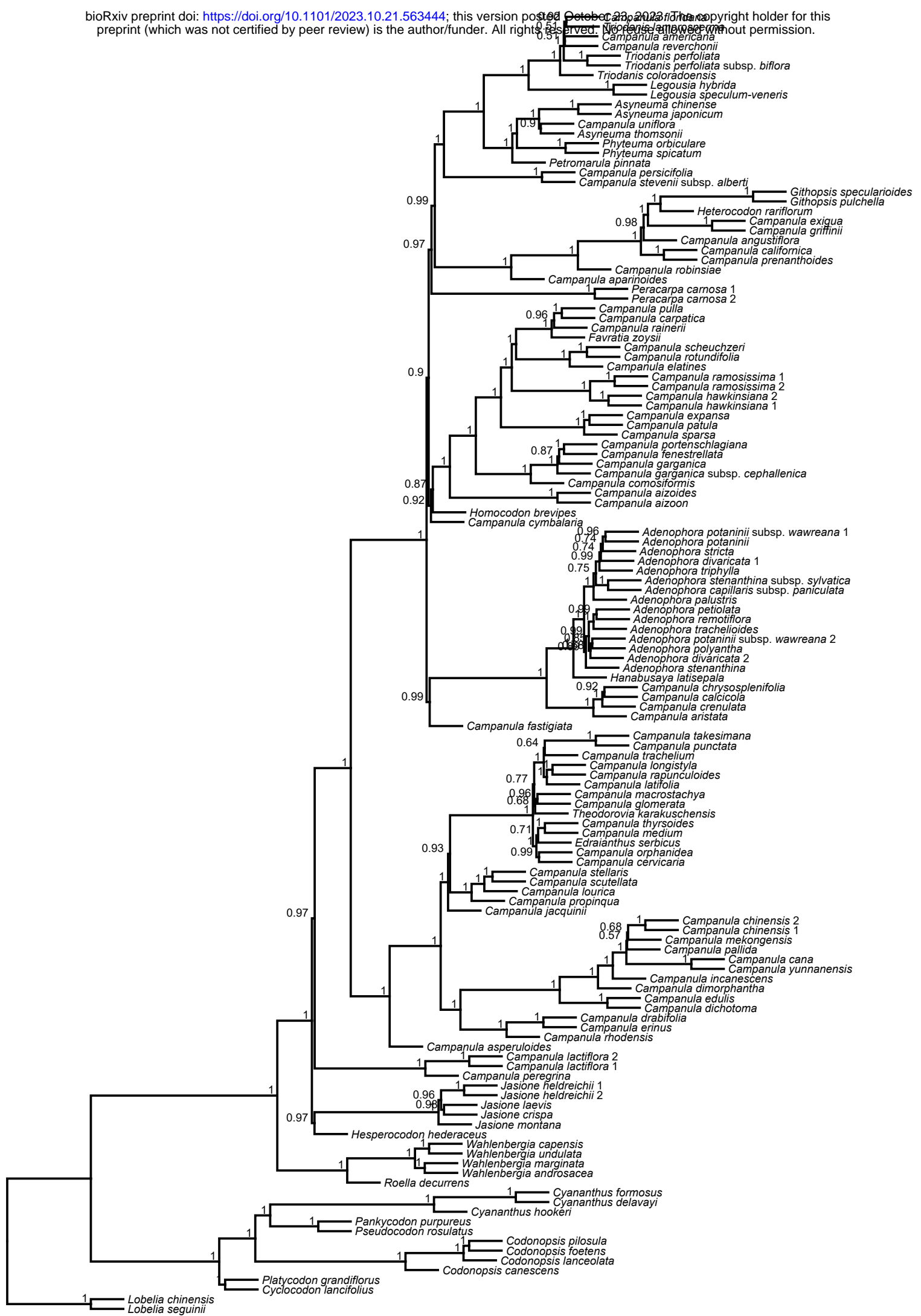


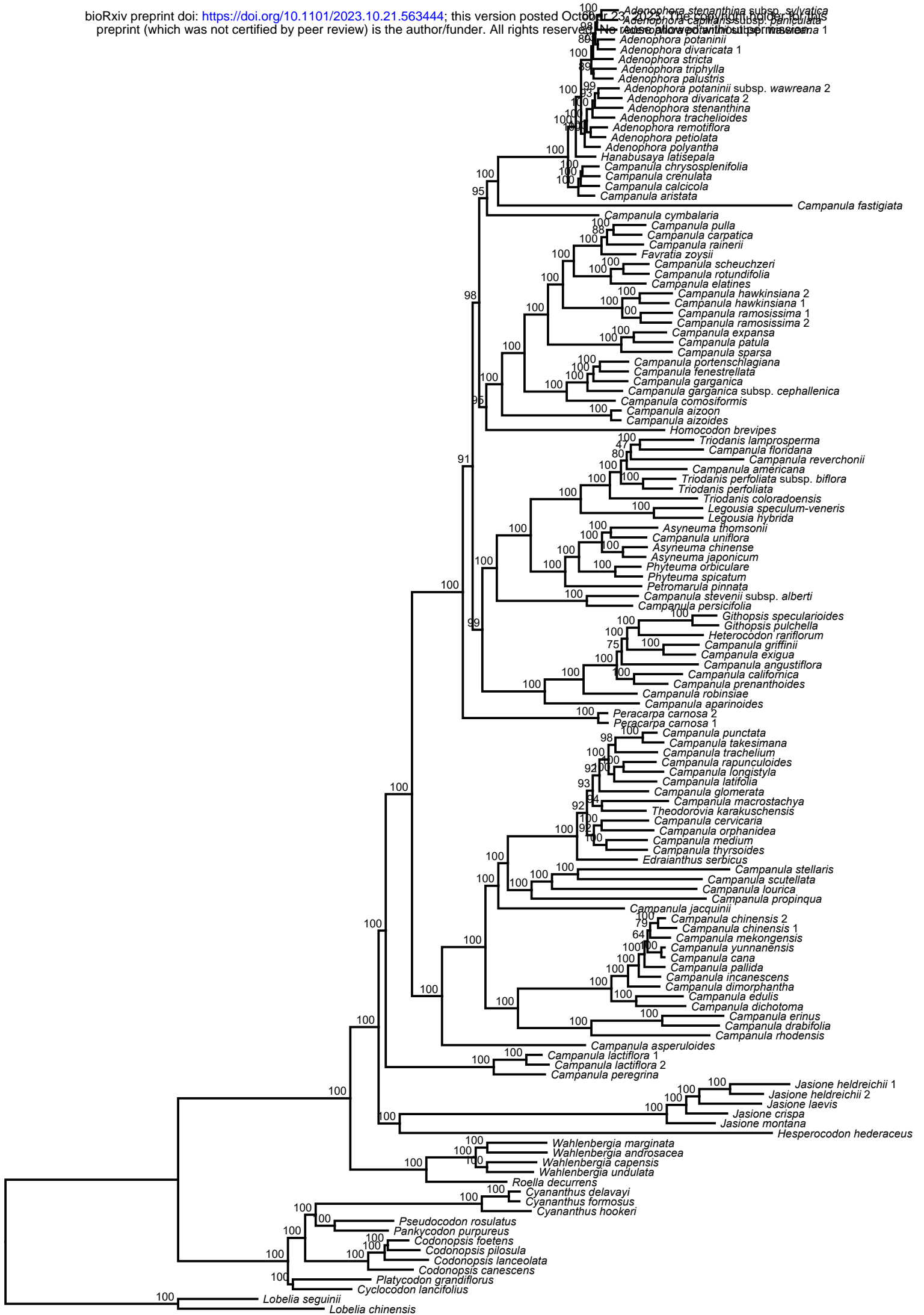


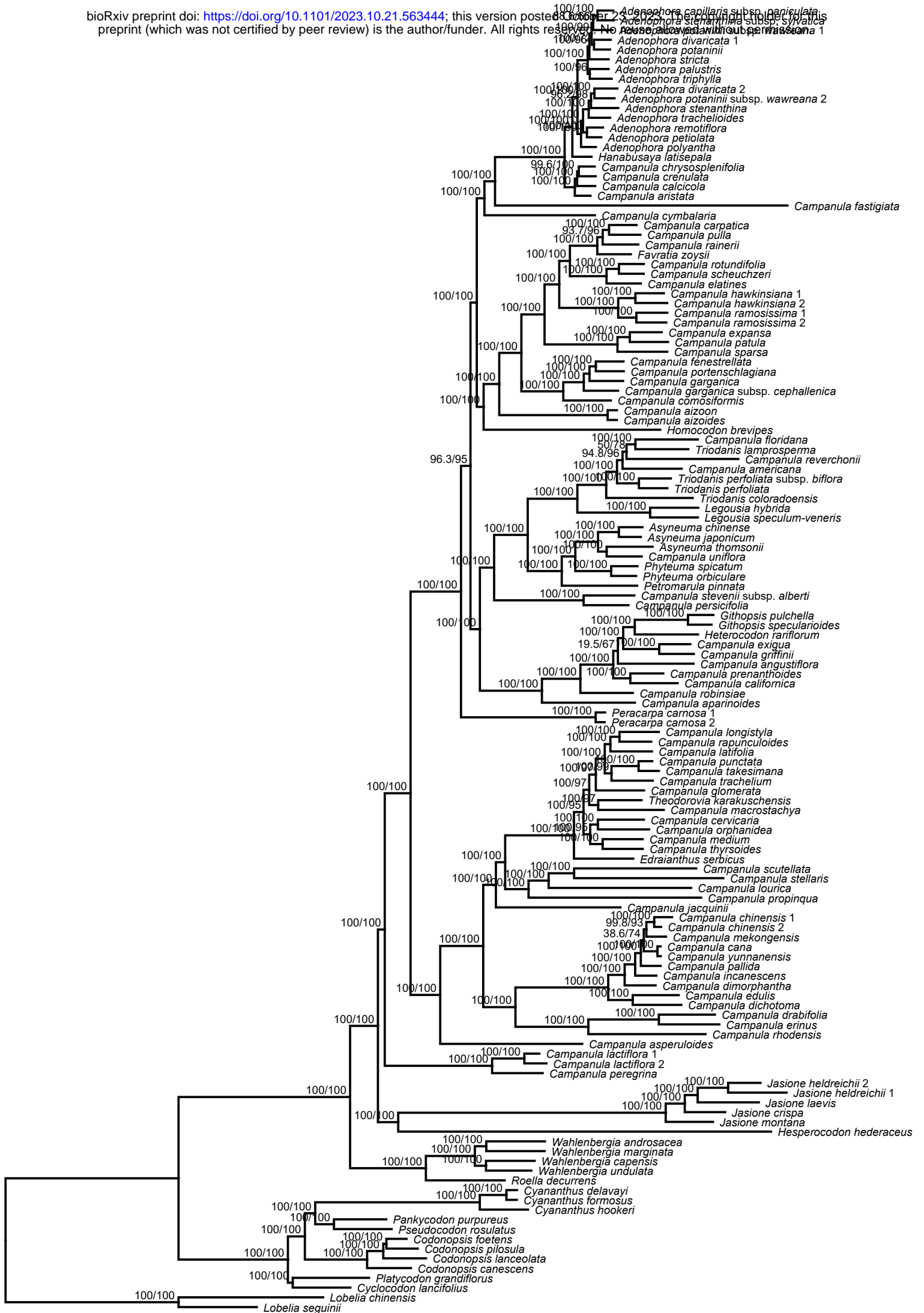


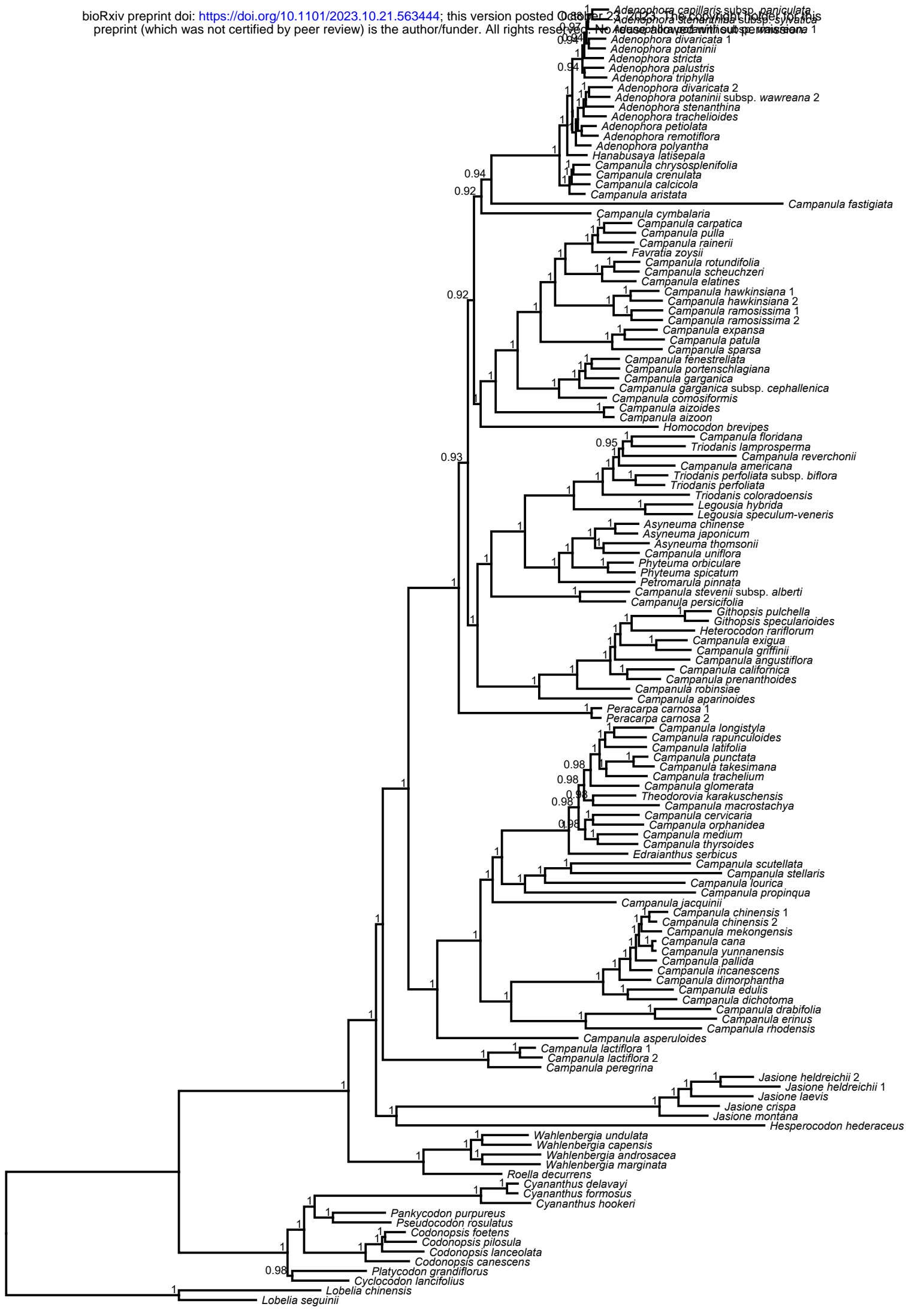


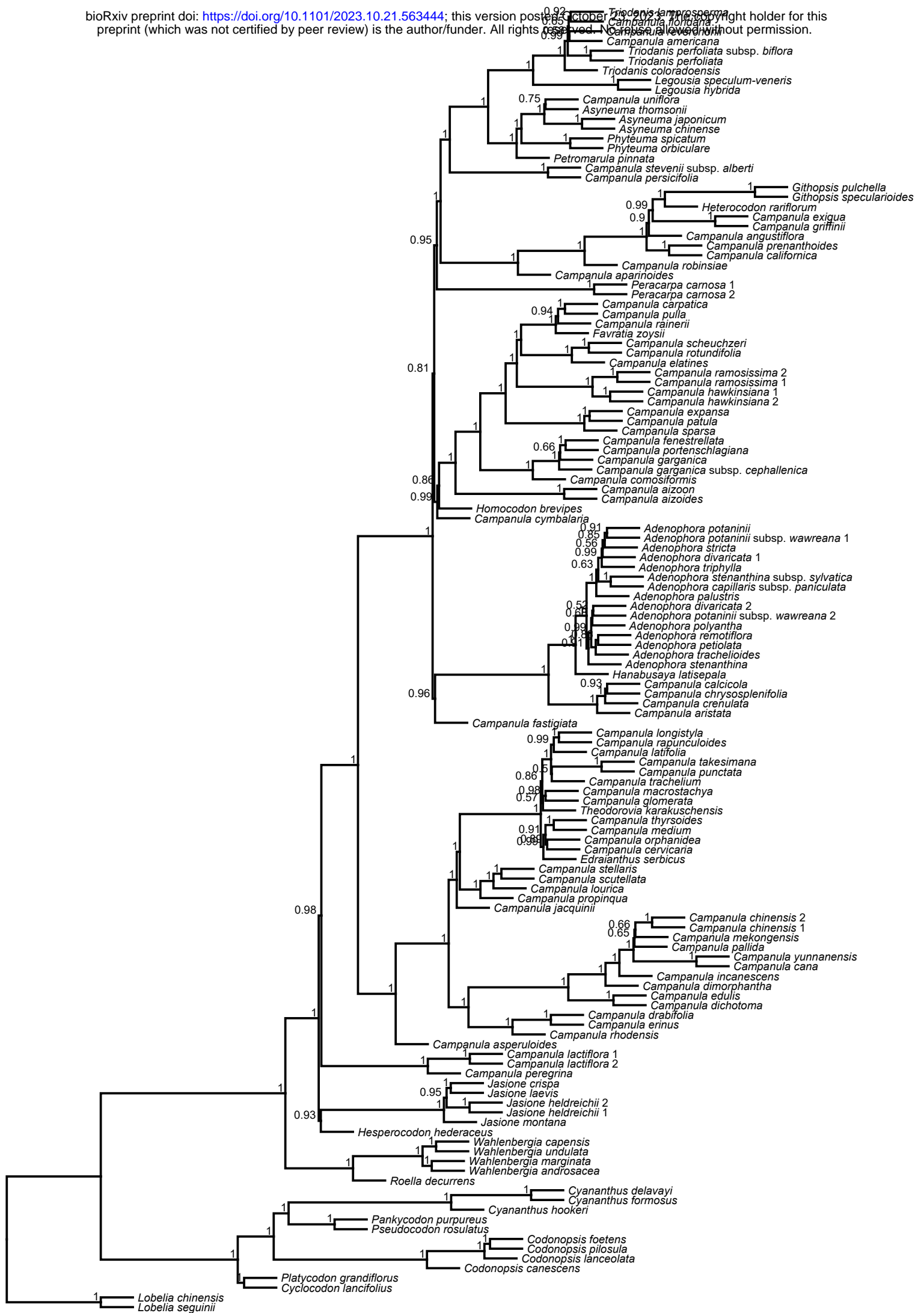


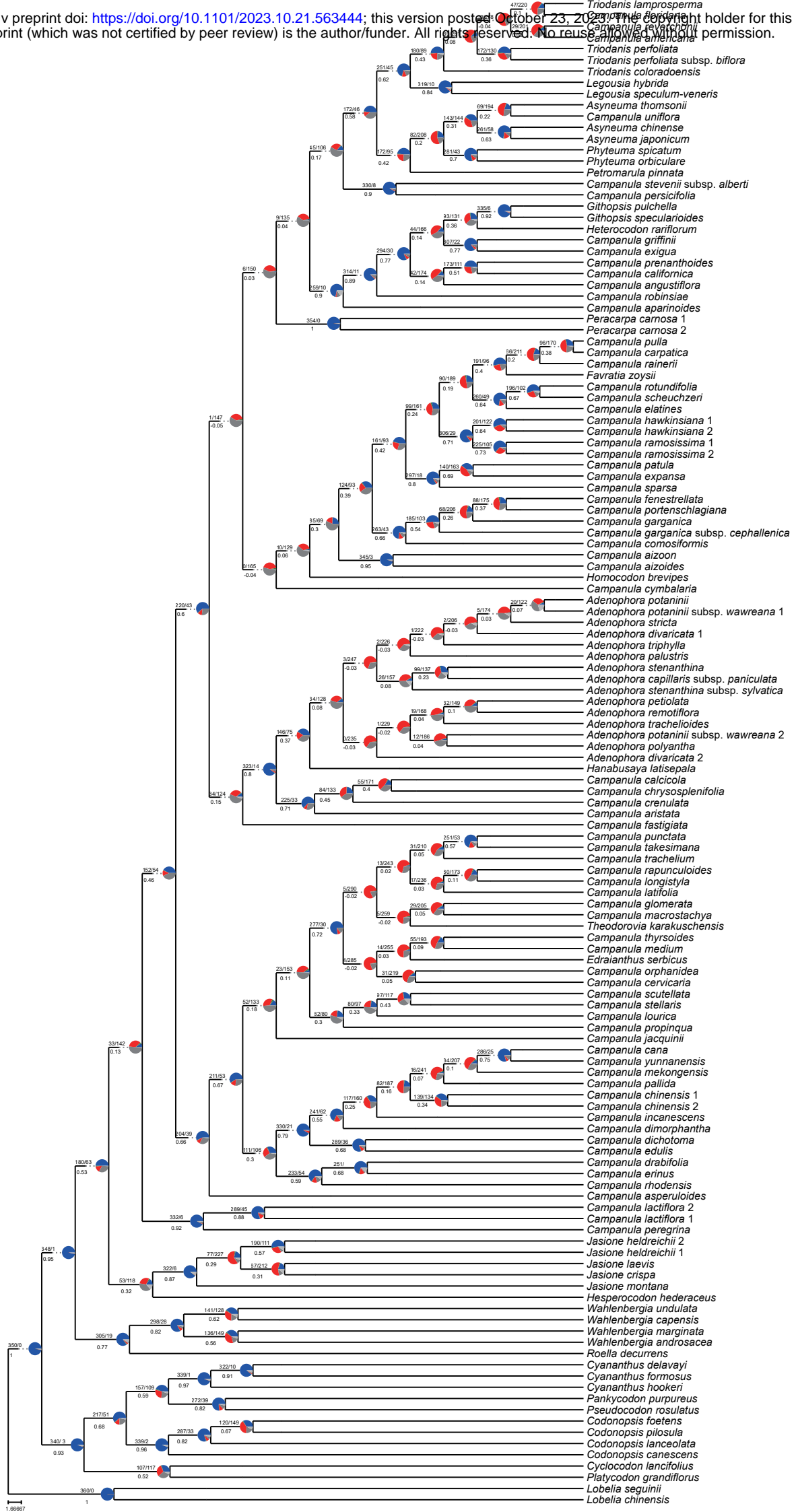






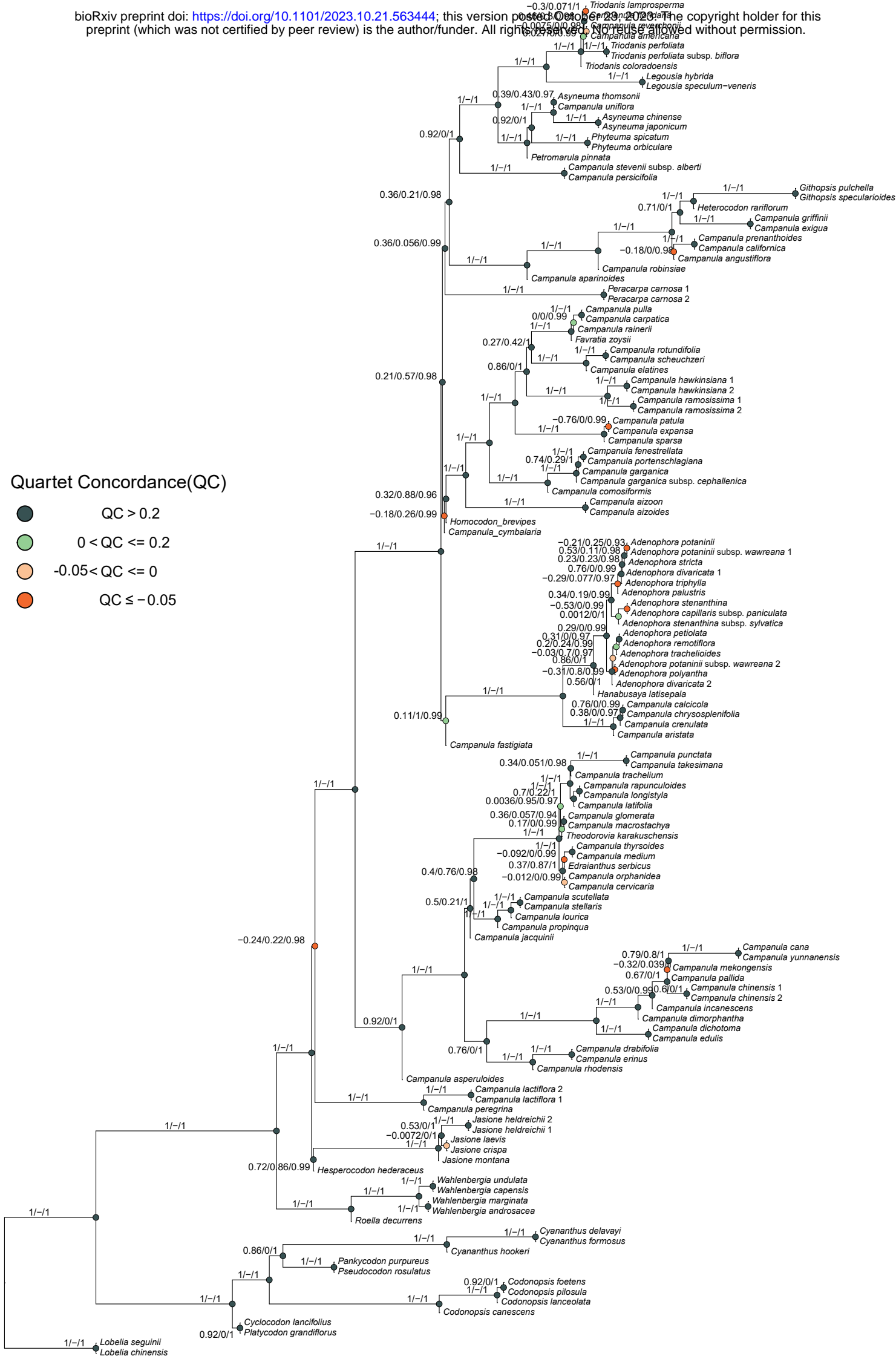


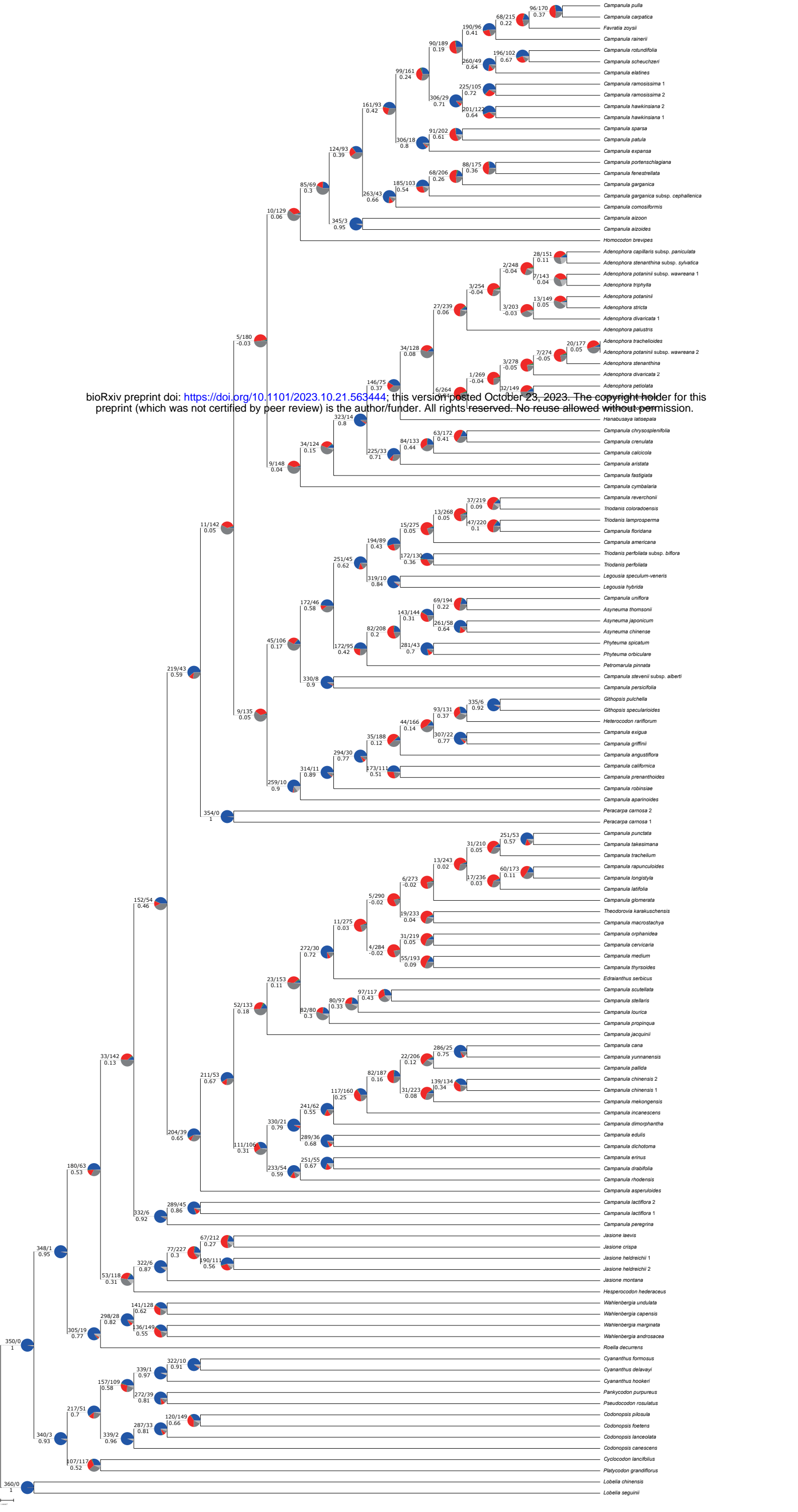




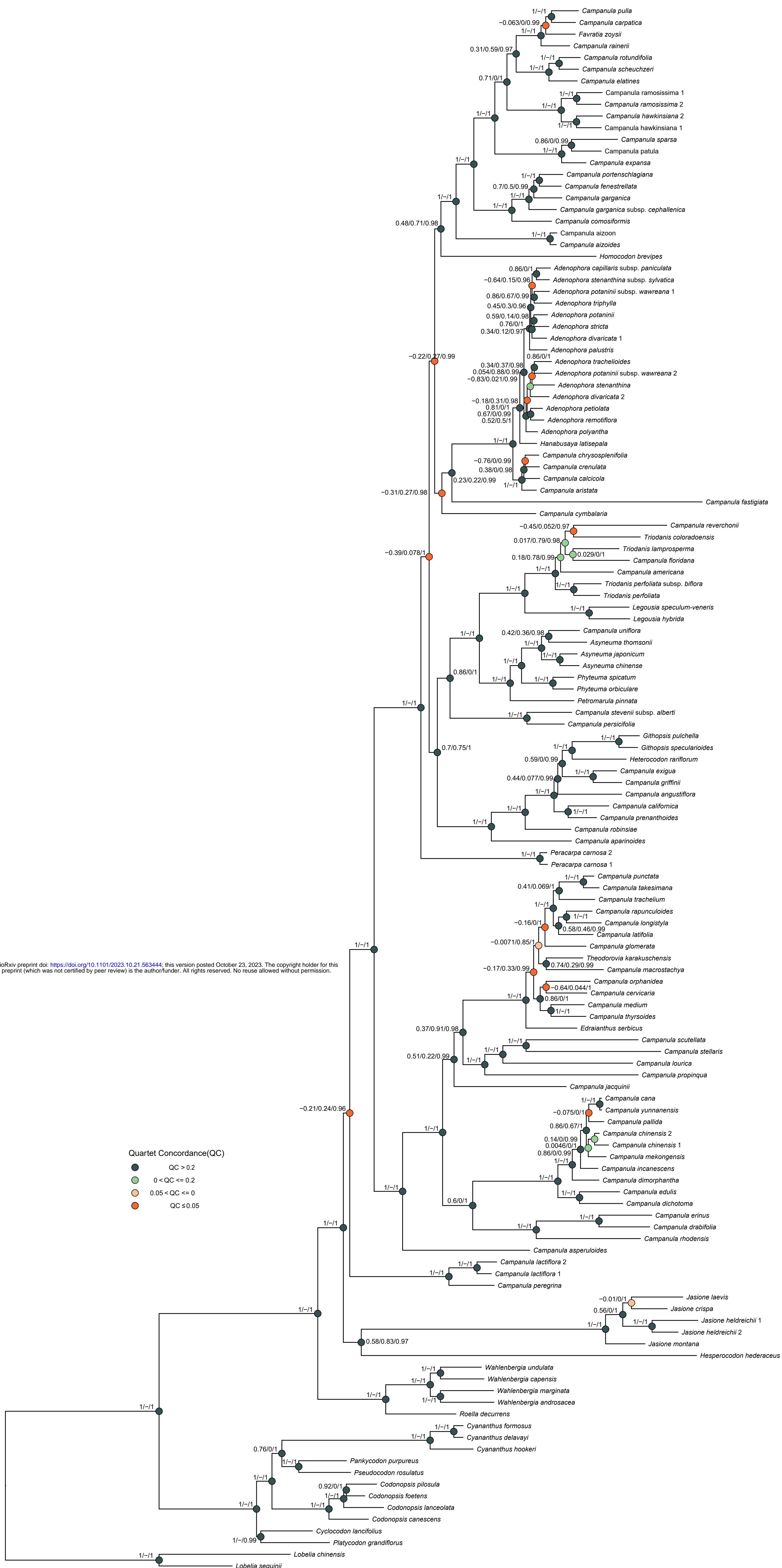
Quartet Concordance(QC)

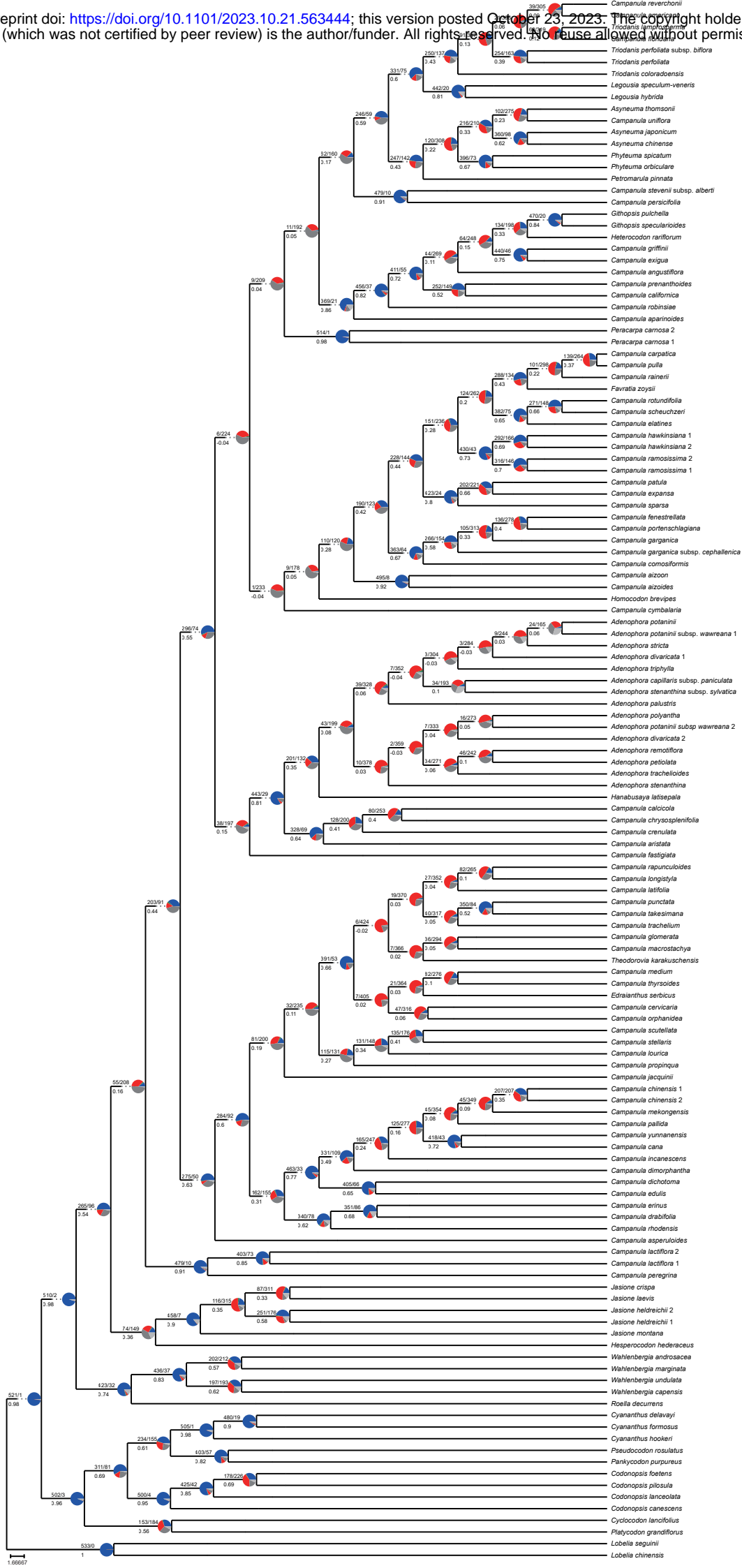
- QC > 0.2
- 0 < QC ≤ 0.2
- -0.05 < QC ≤ 0
- QC ≤ -0.05



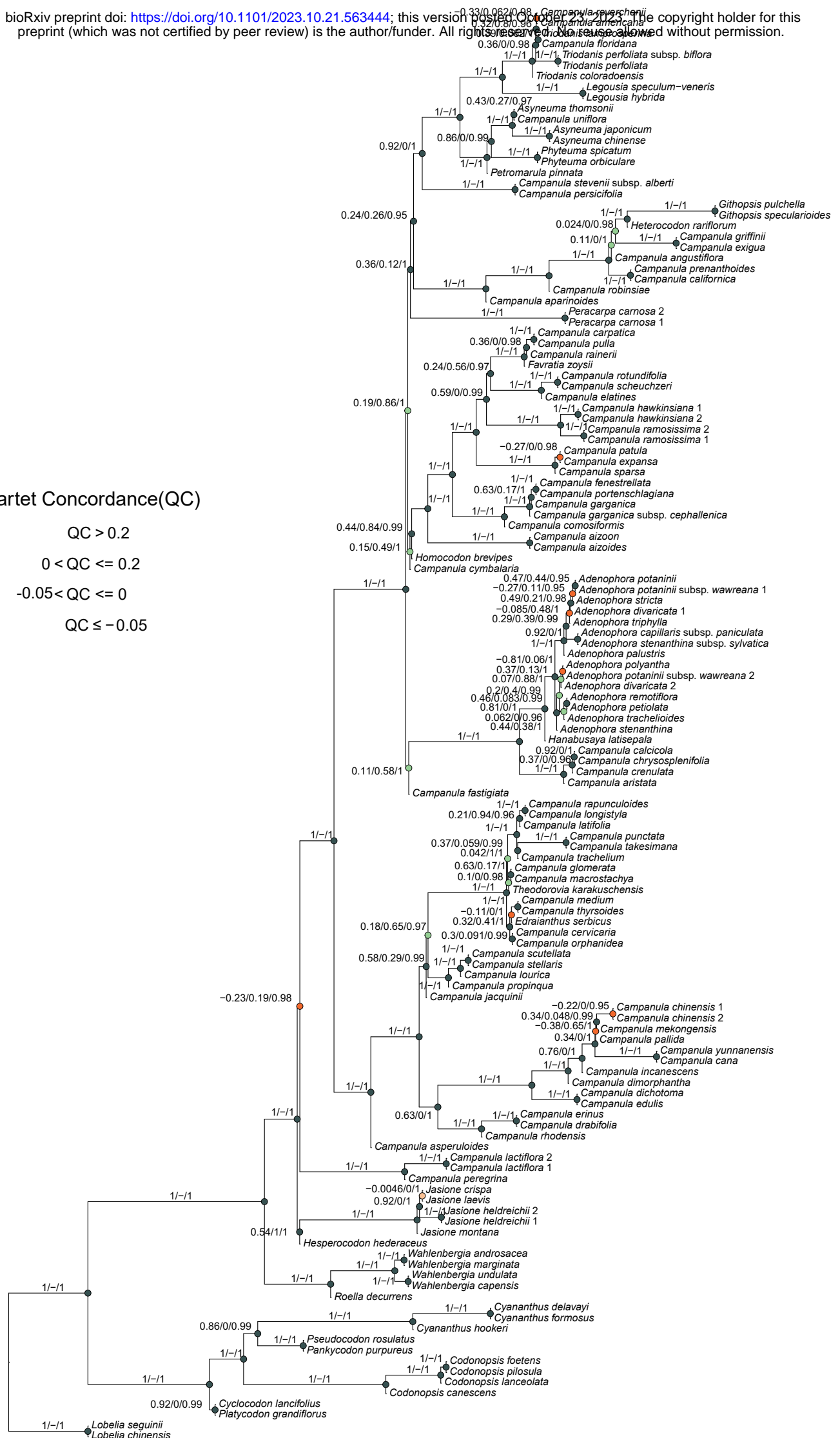
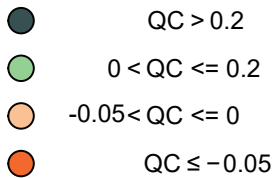


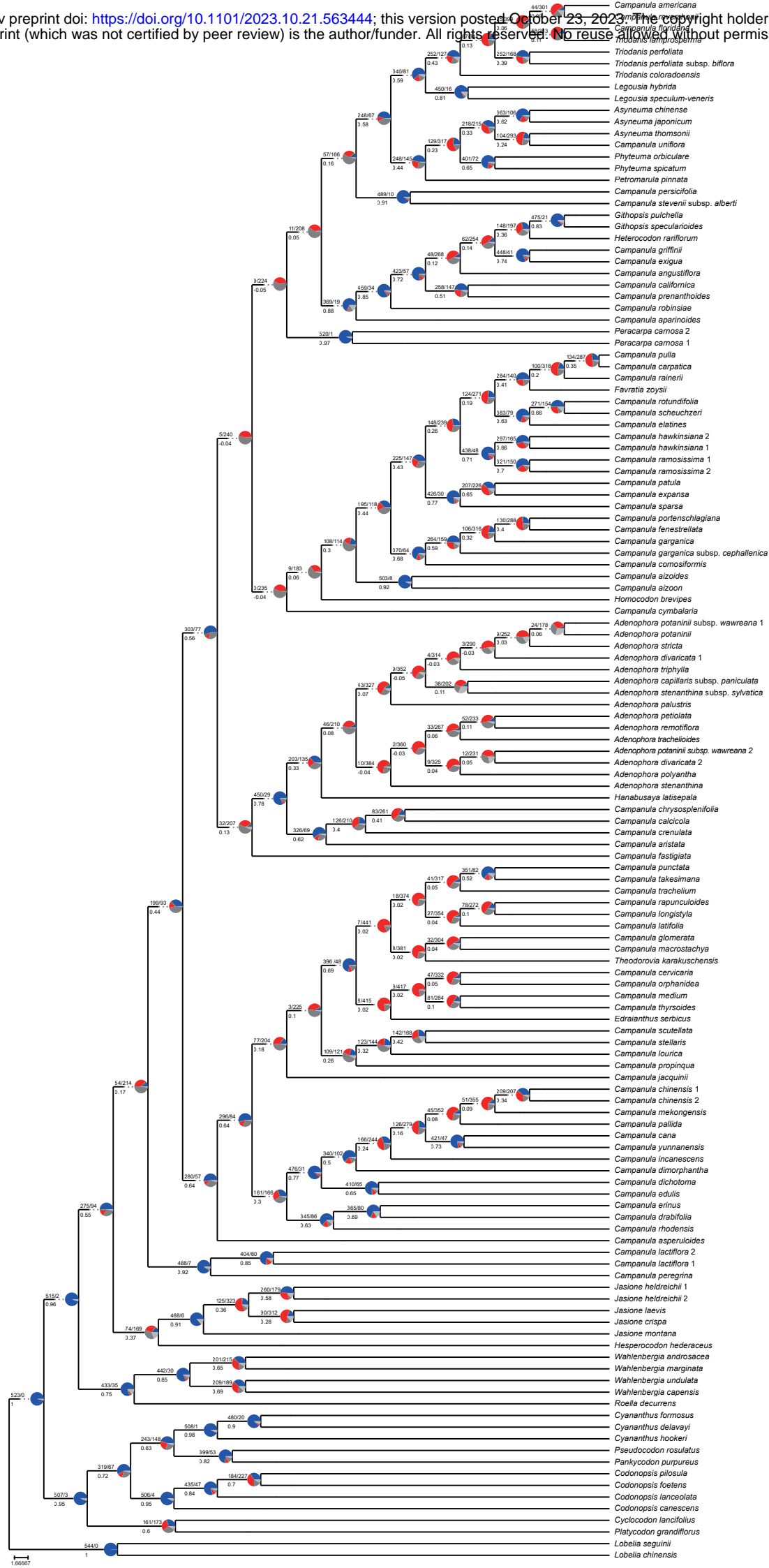
bioRxiv preprint doi: <https://doi.org/10.1101/2023.10.21.563444>; this version posted October 23, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



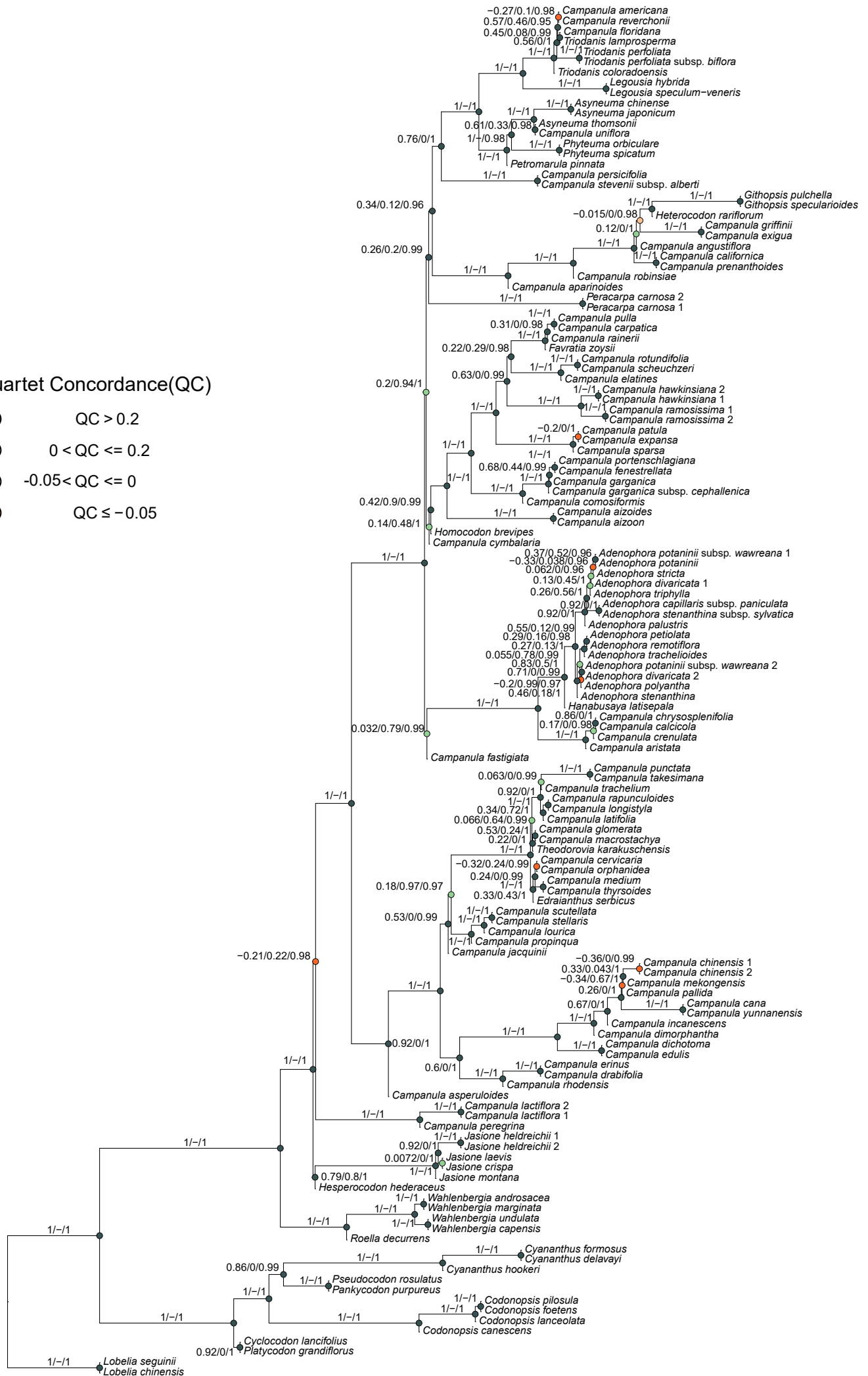
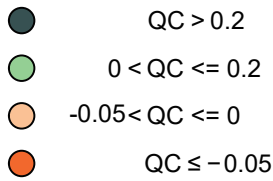


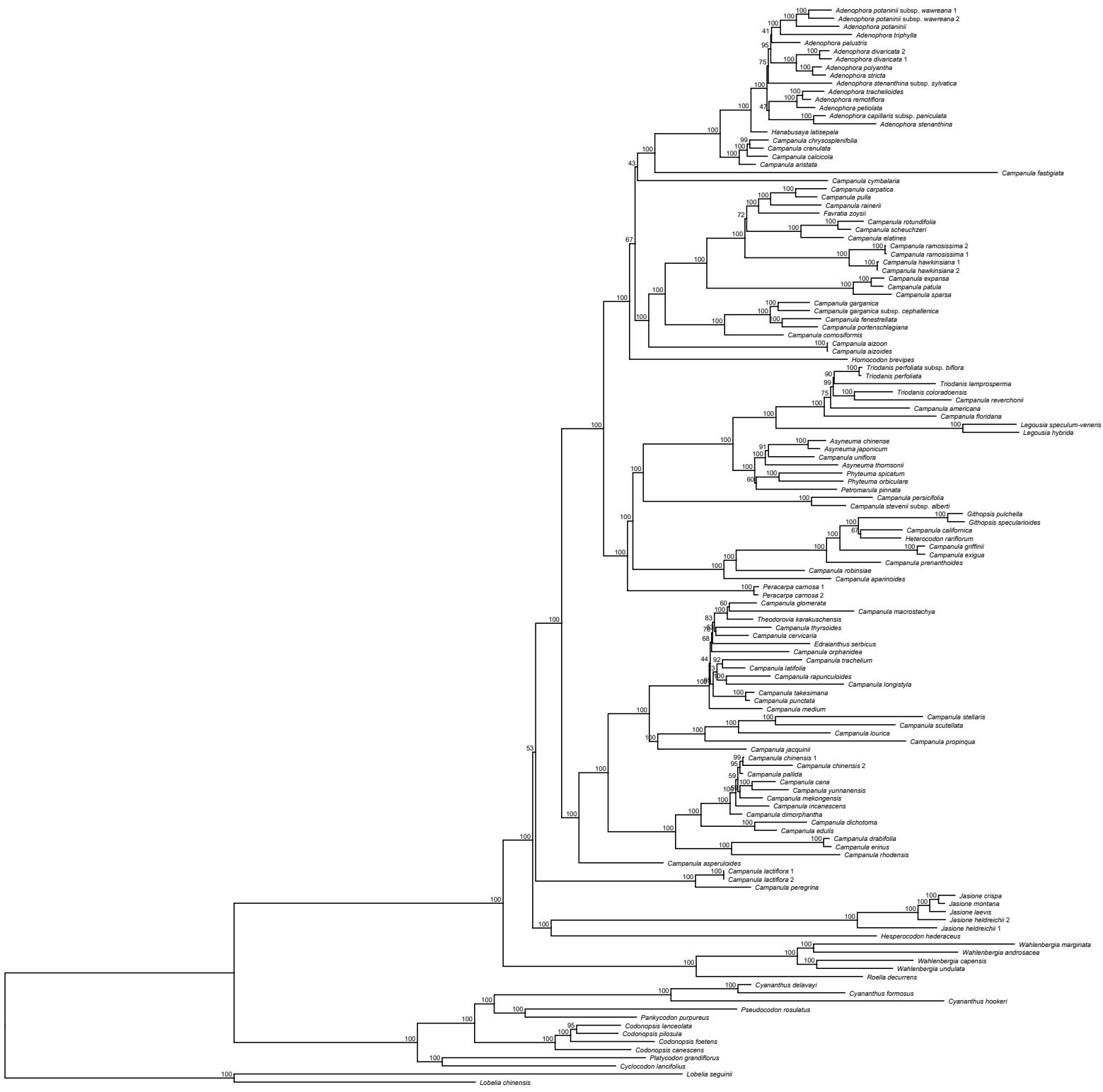
Quartet Concordance(QC)

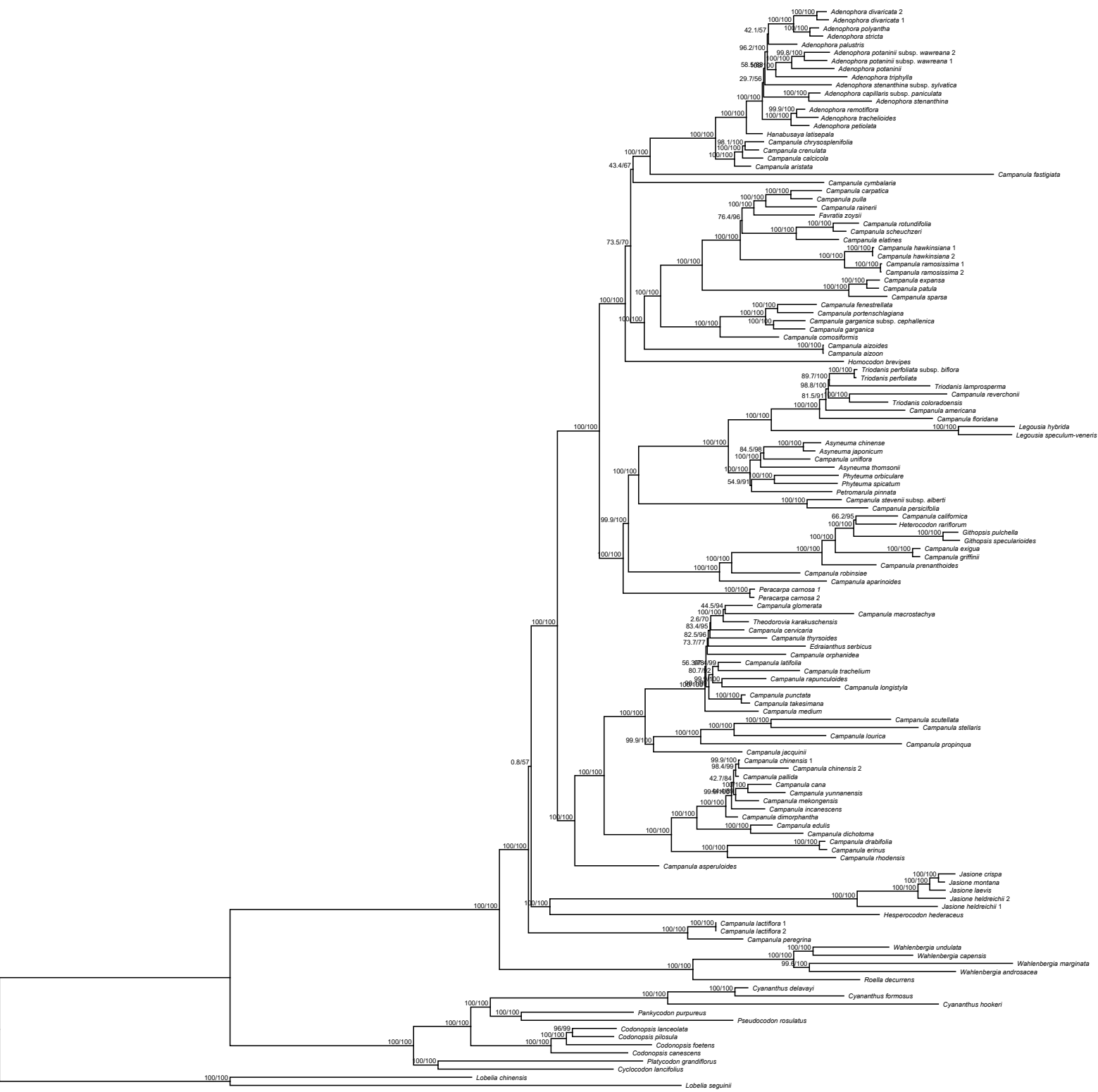


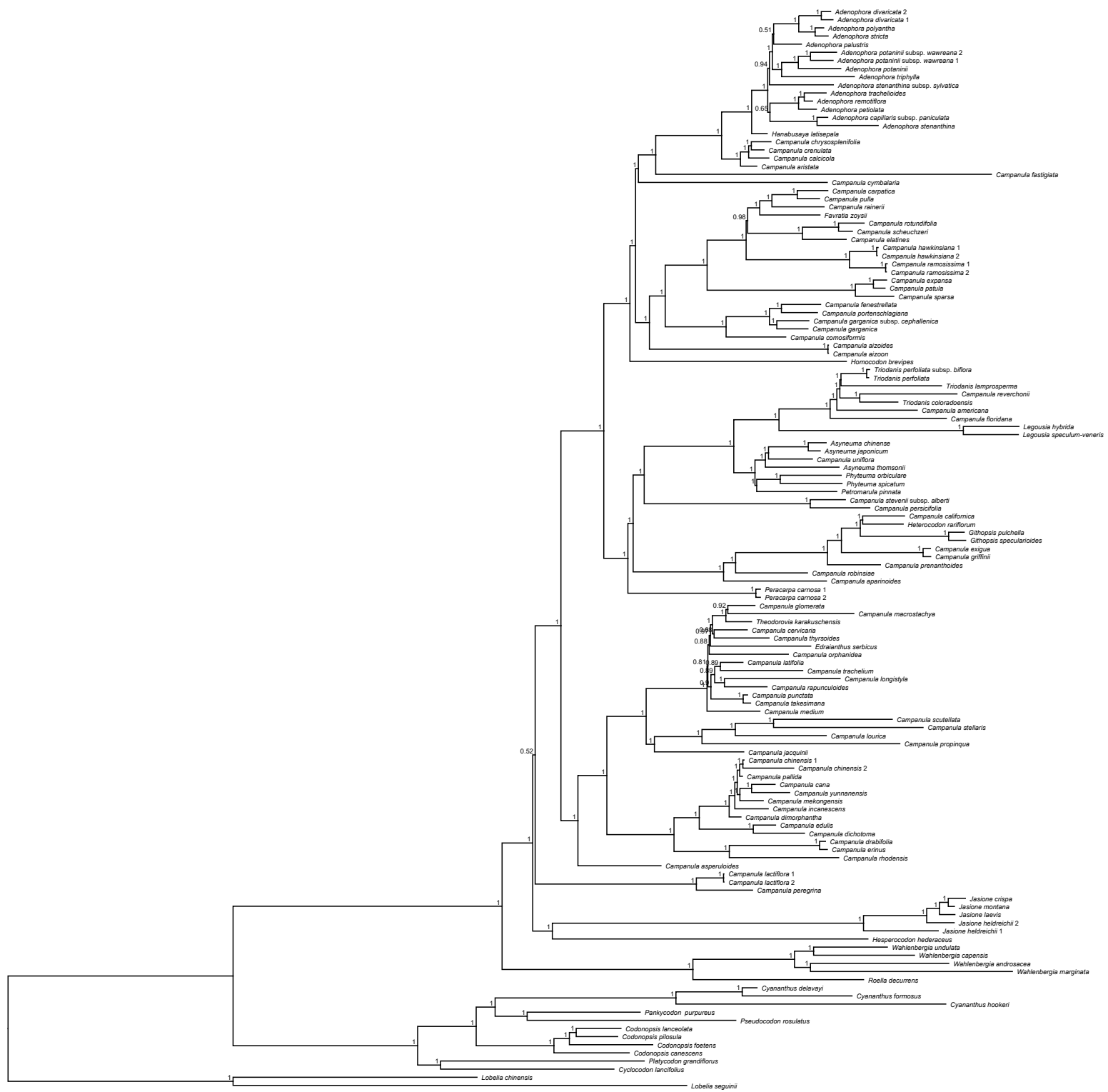


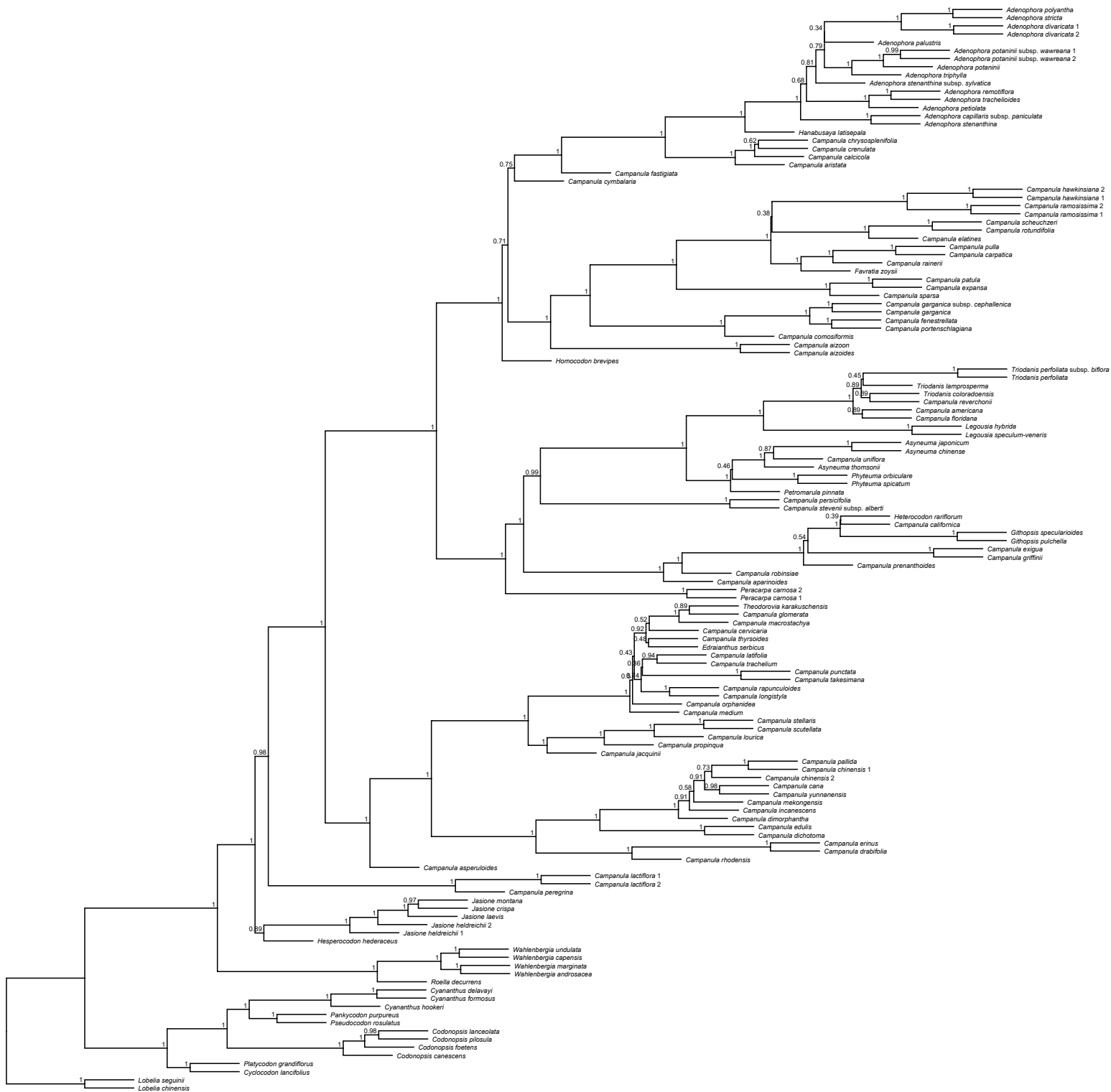
Quartet Concordance(QC)

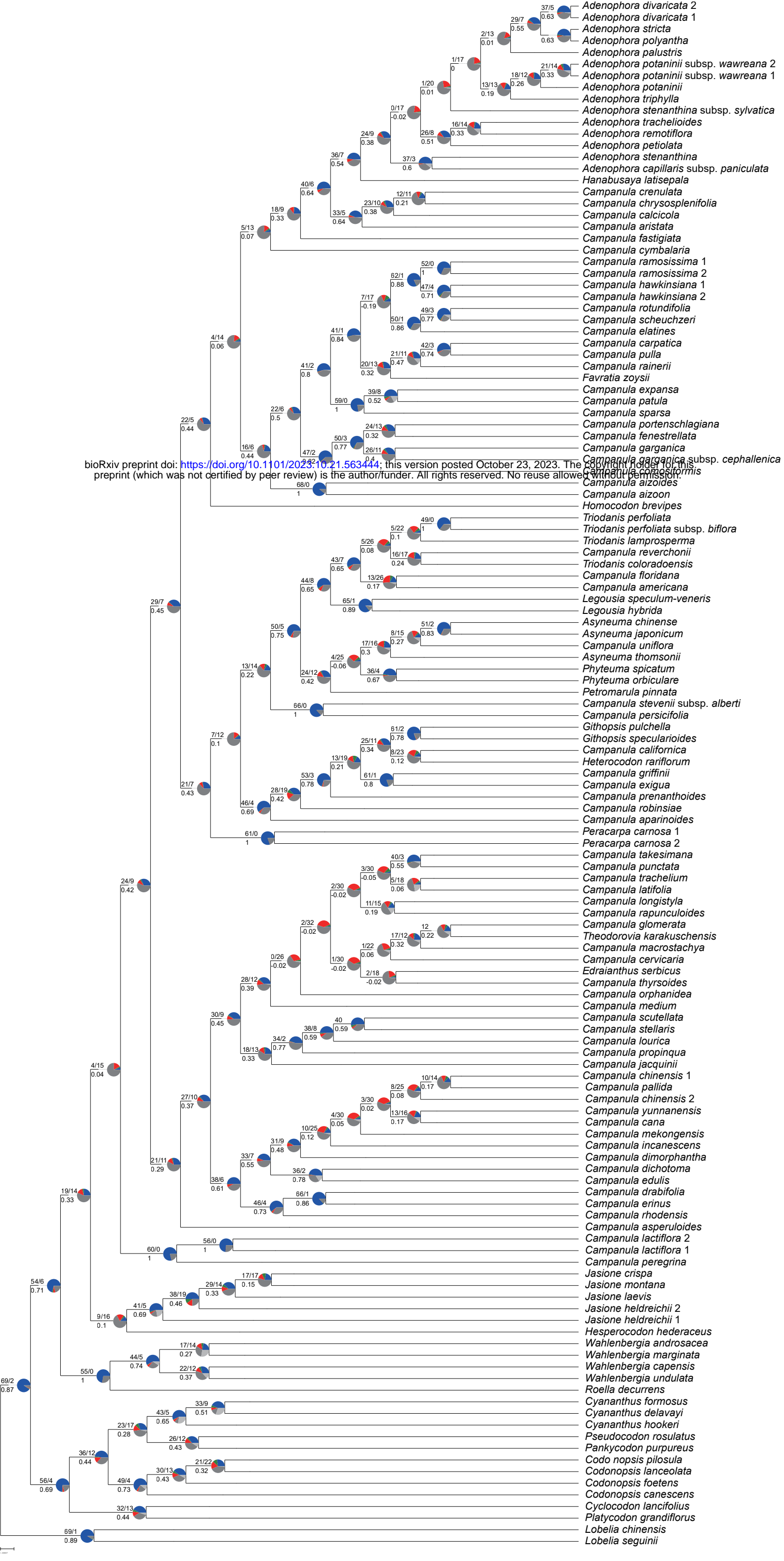






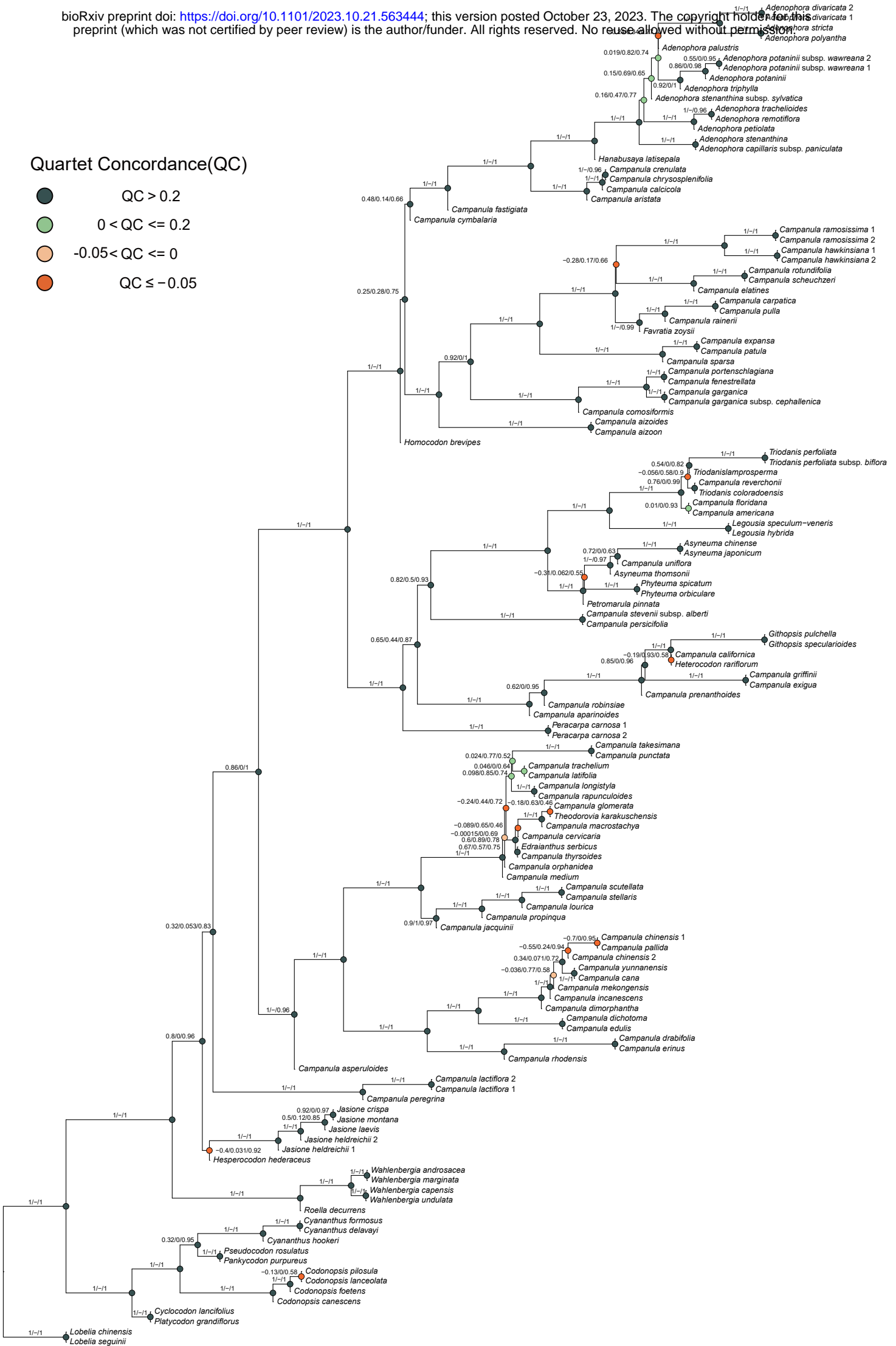


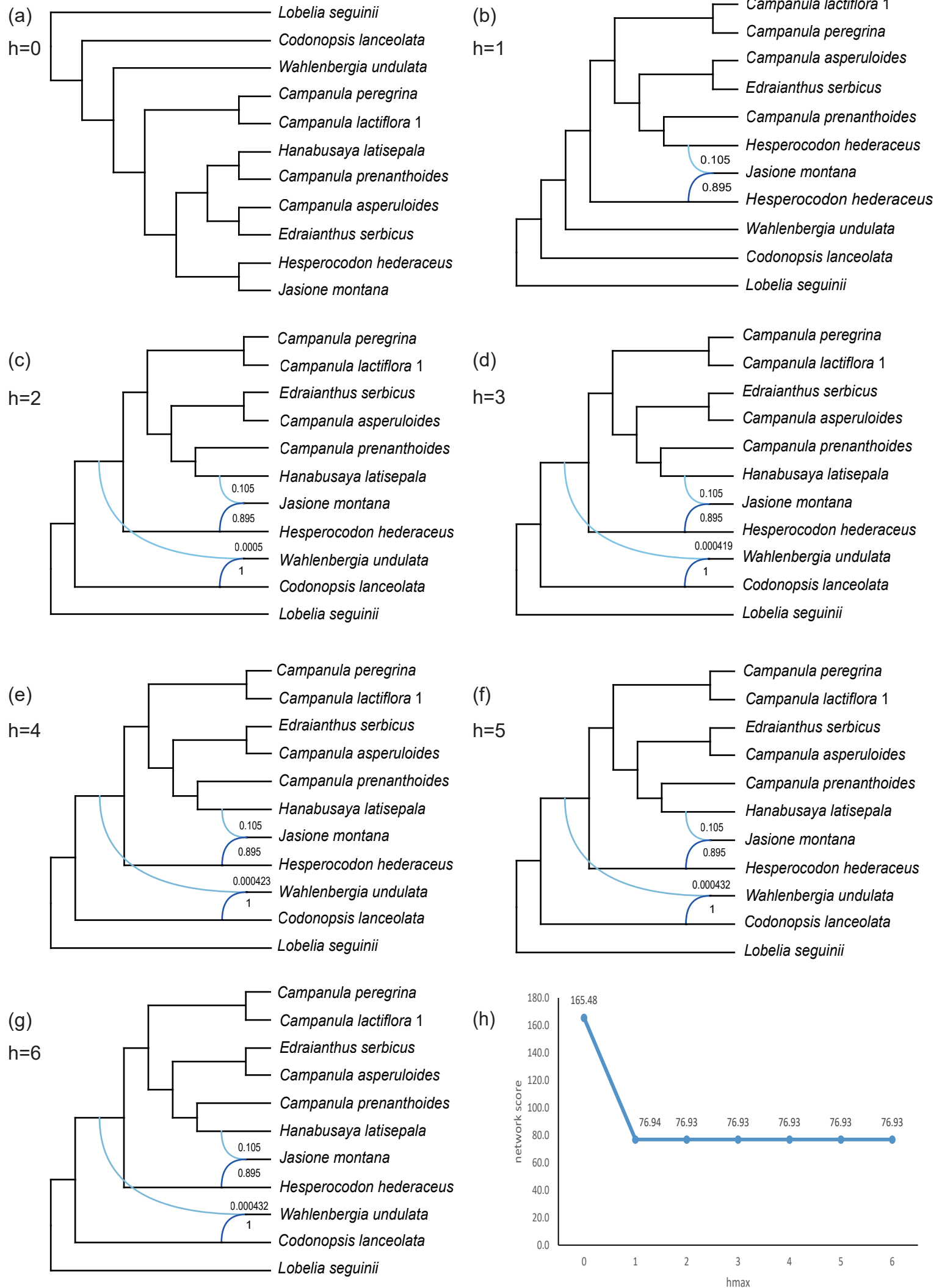




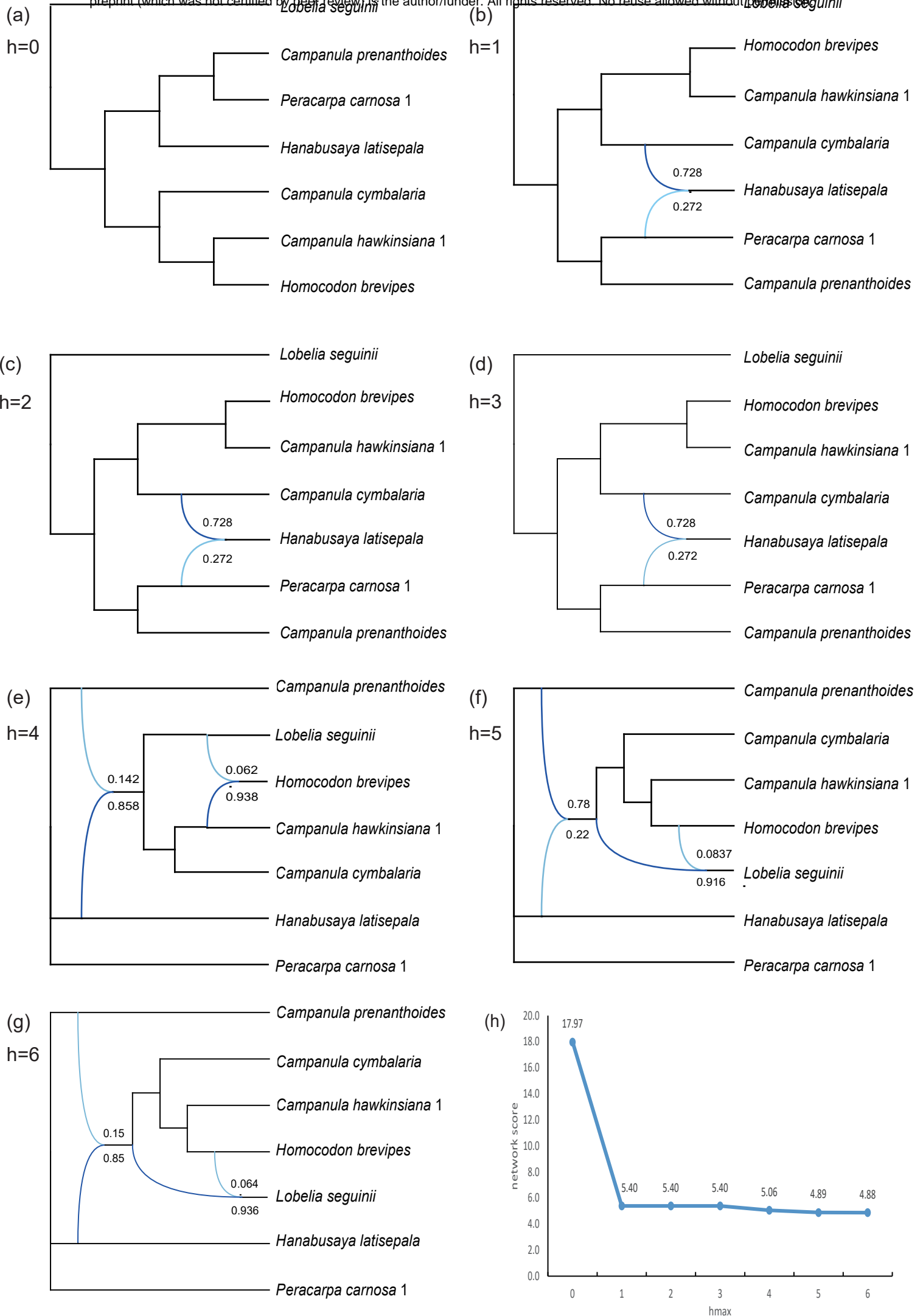
Quartet Concordance(QC)

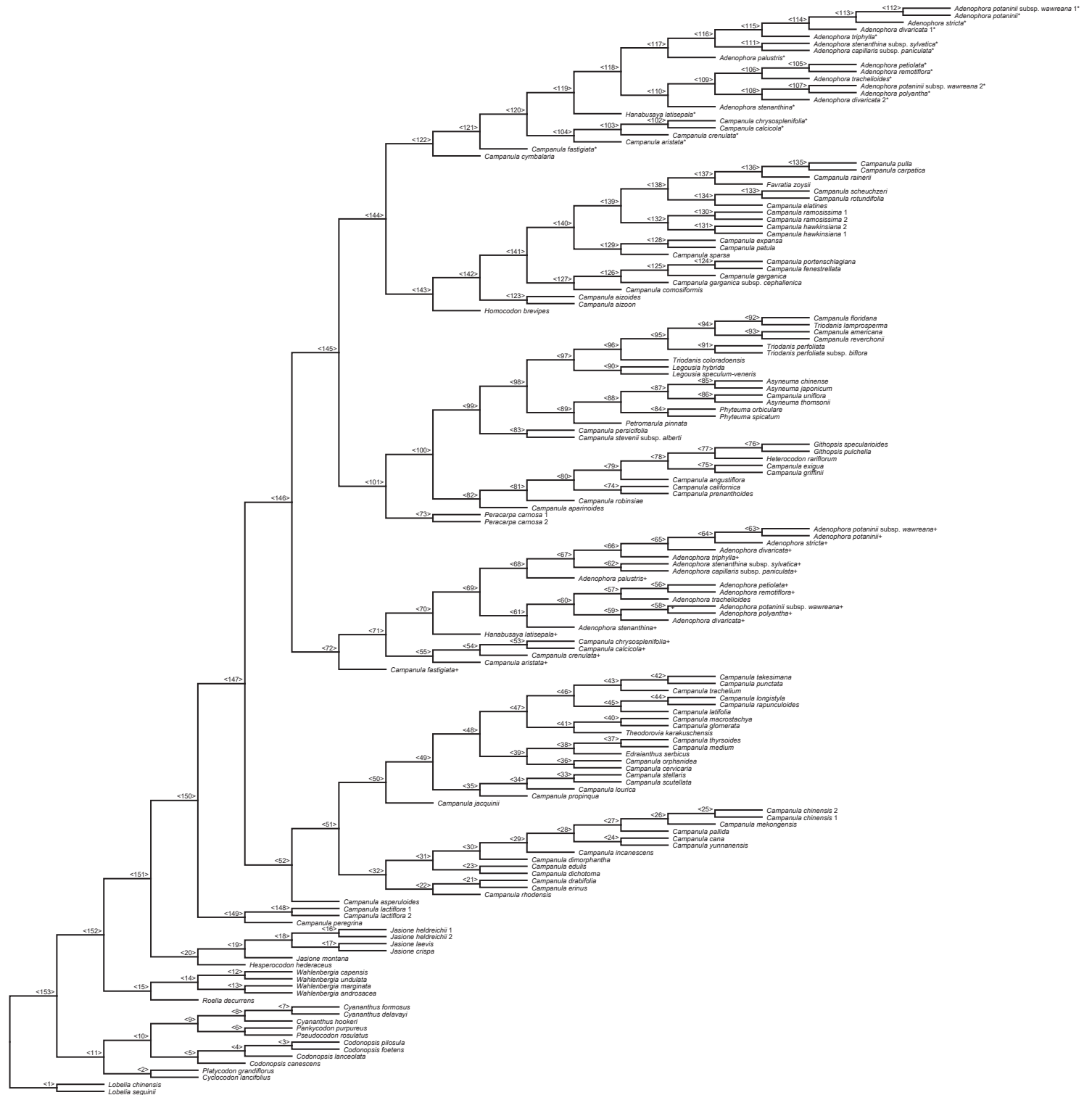
- QC > 0.2
- 0 < QC ≤ 0.2
- -0.05 < QC ≤ 0
- QC ≤ -0.05

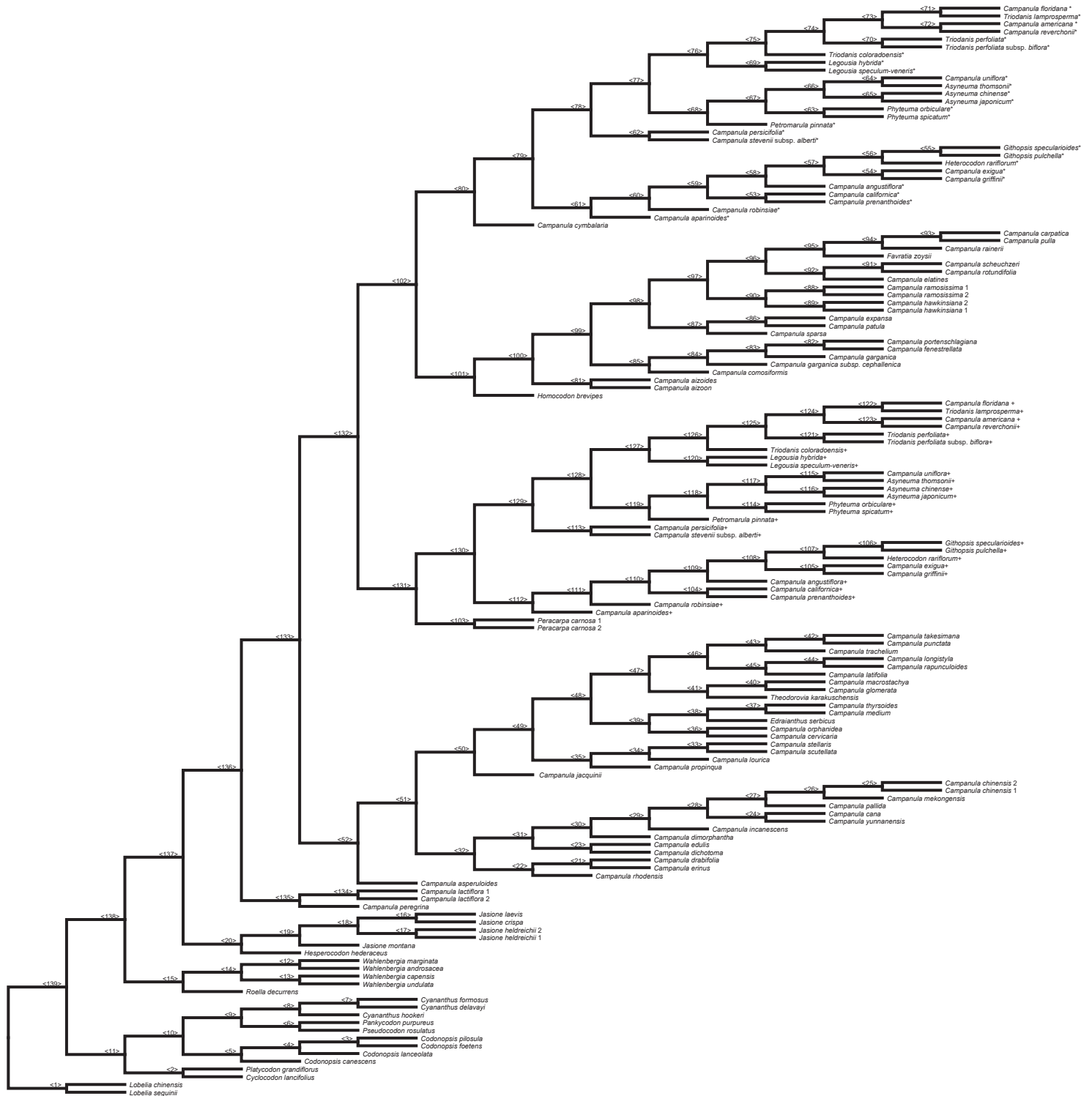


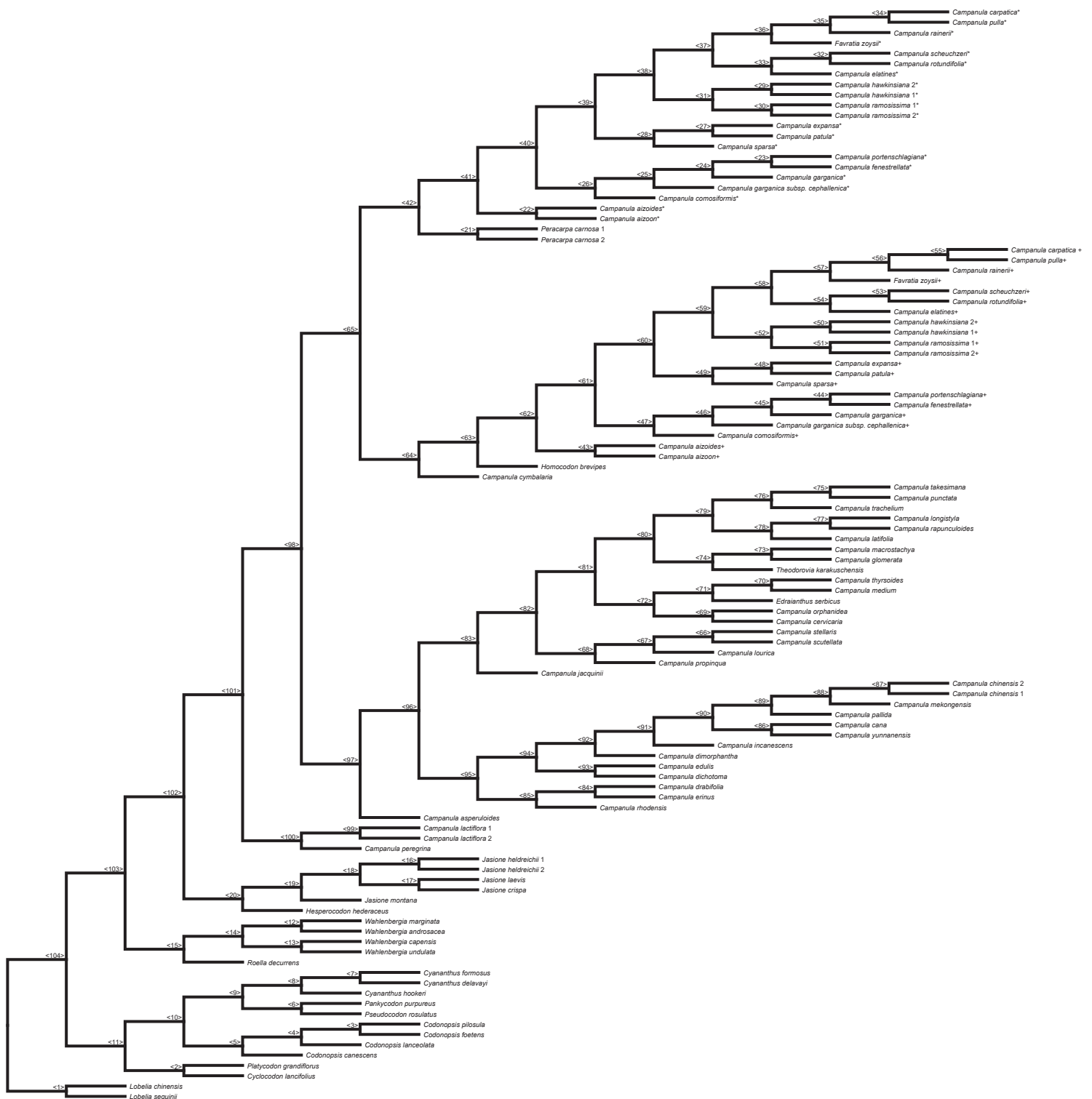


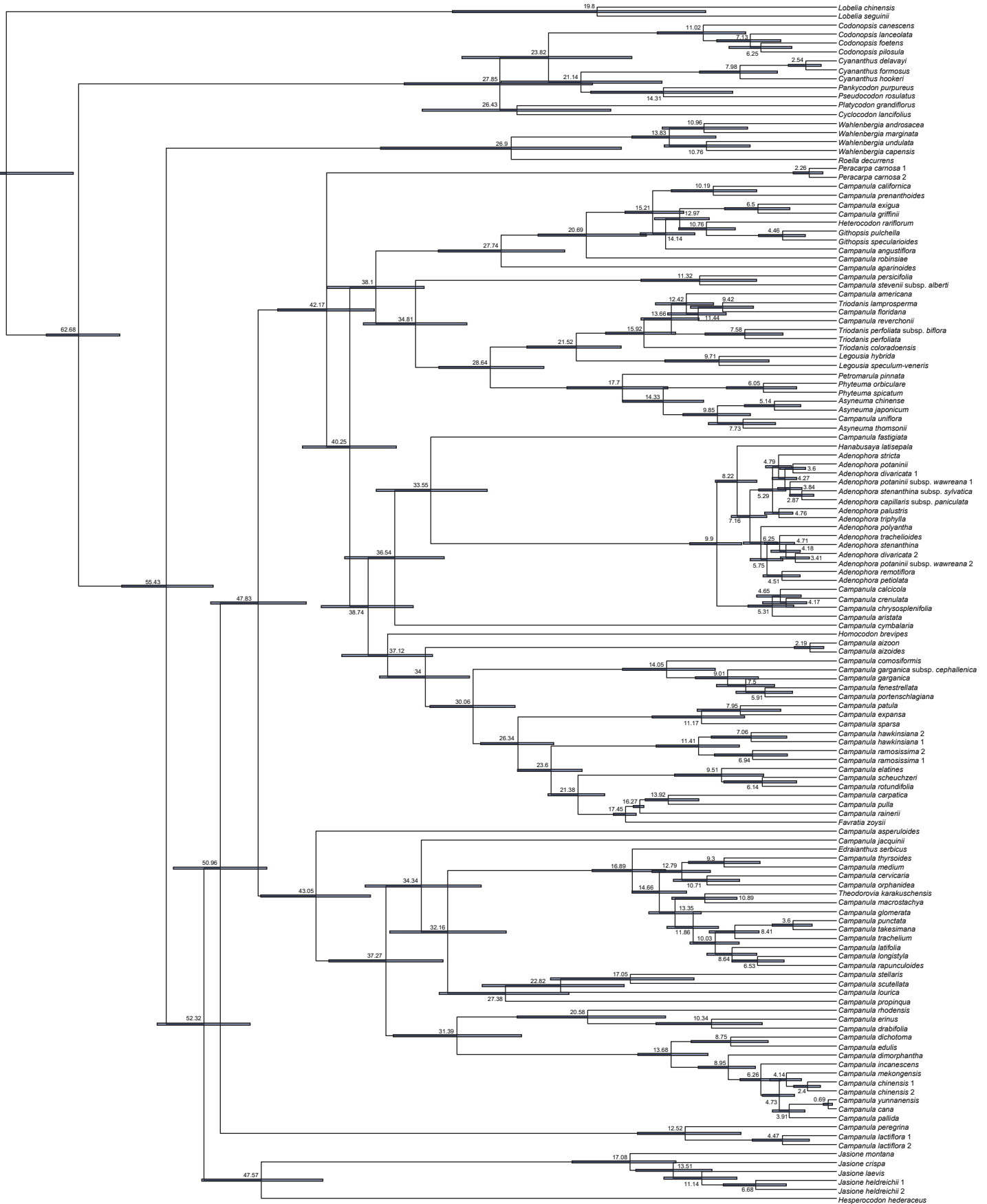


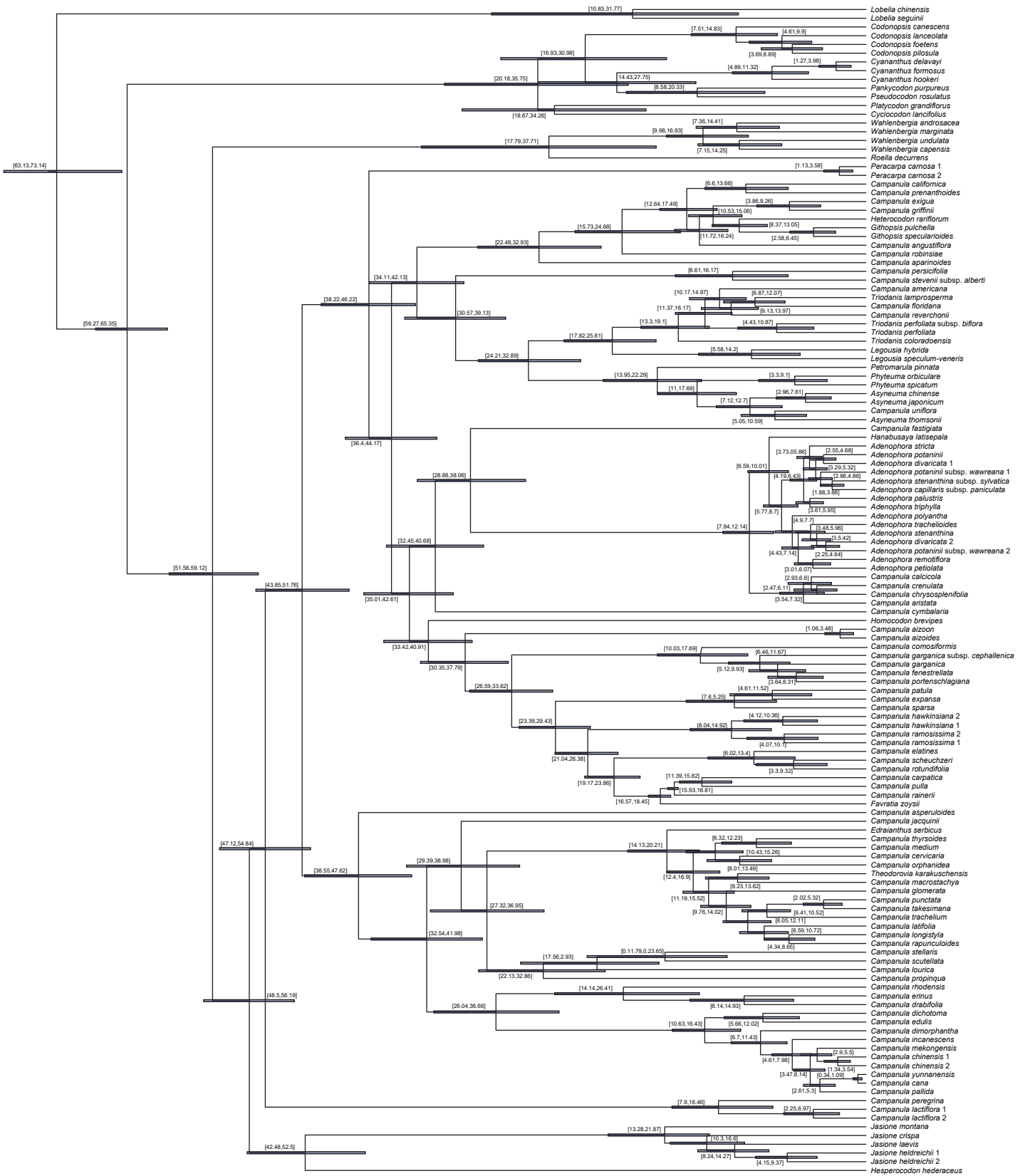




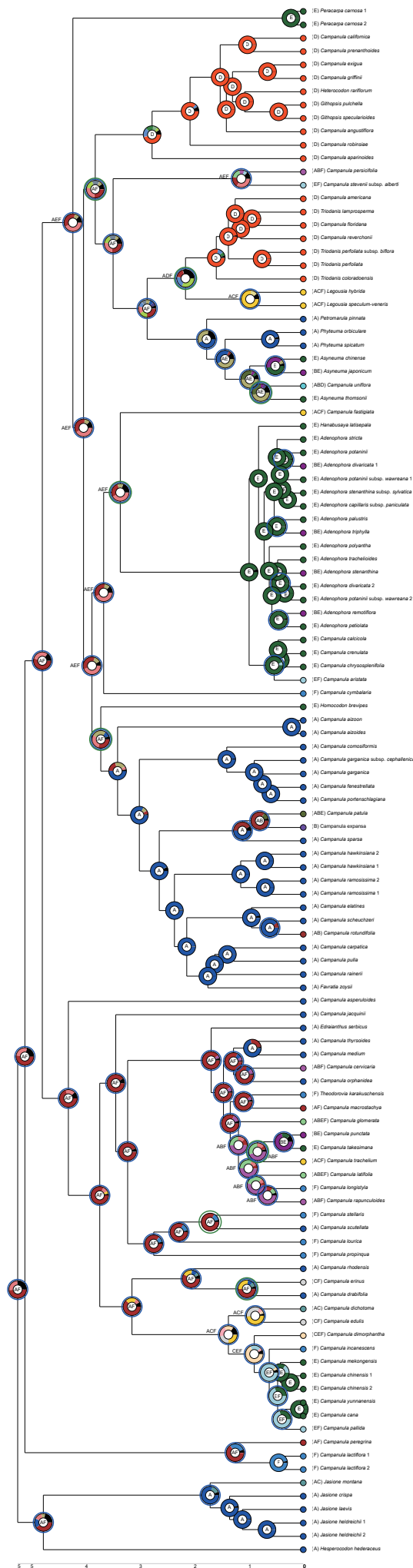
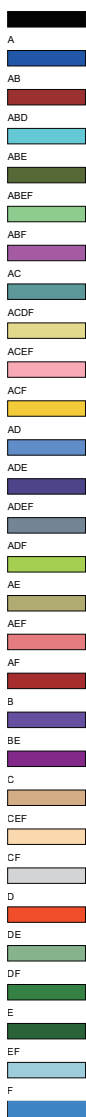


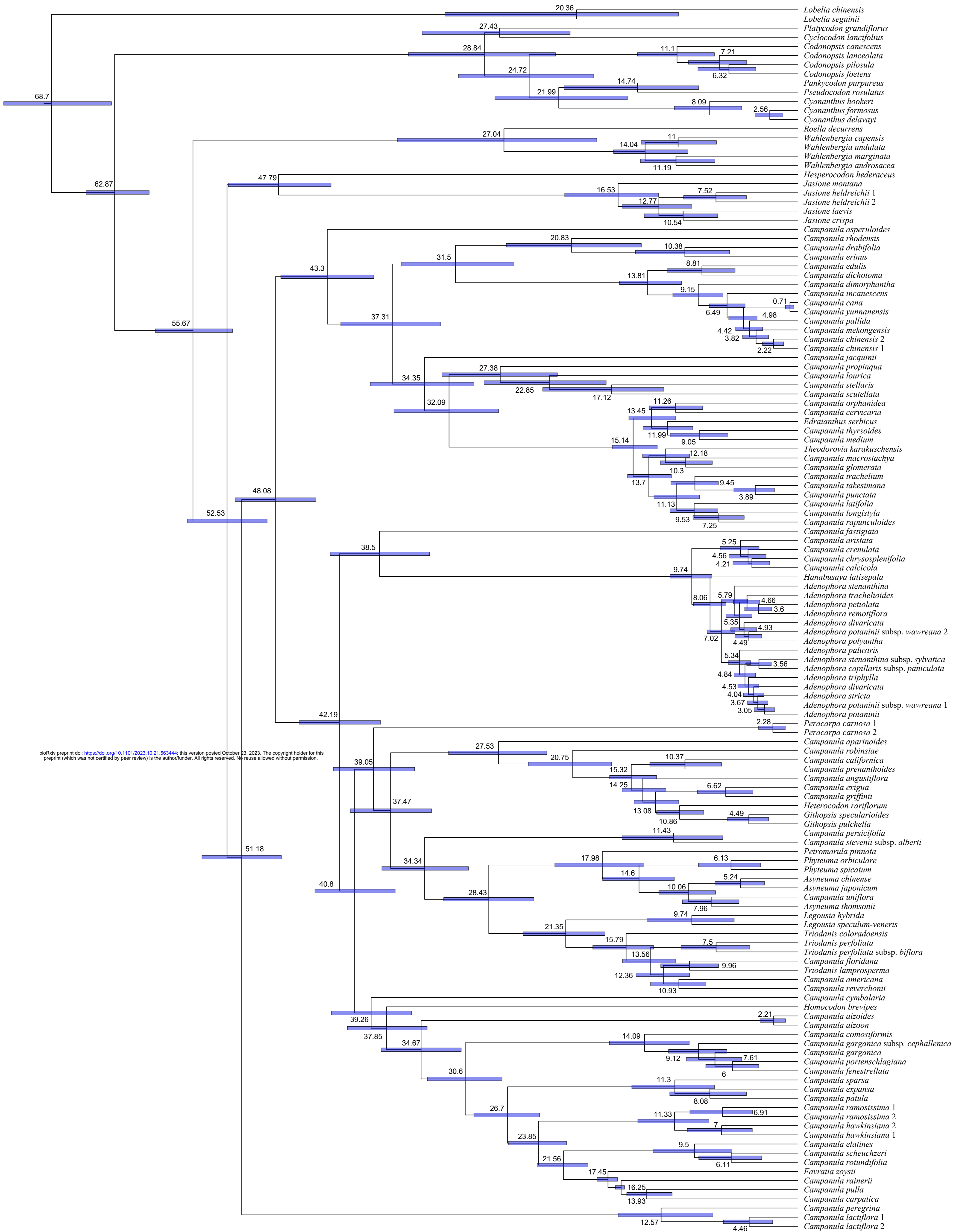




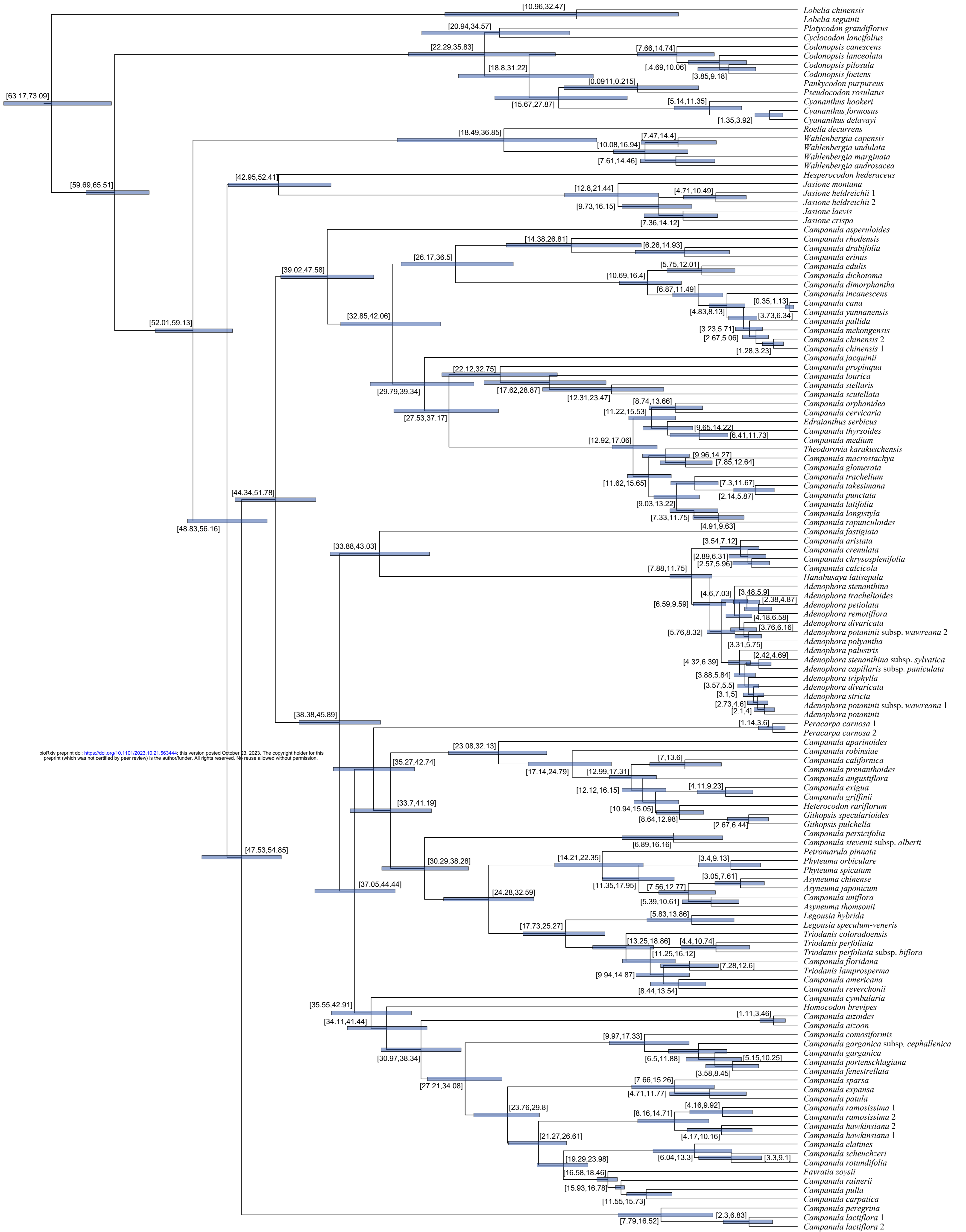


LEGEND



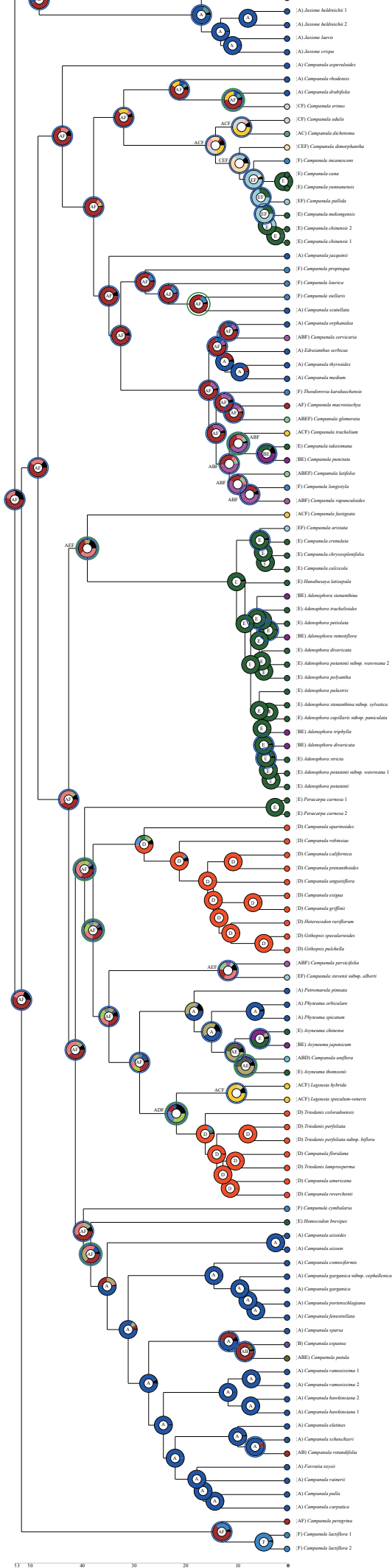
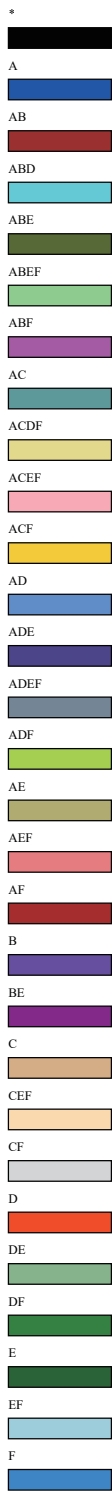


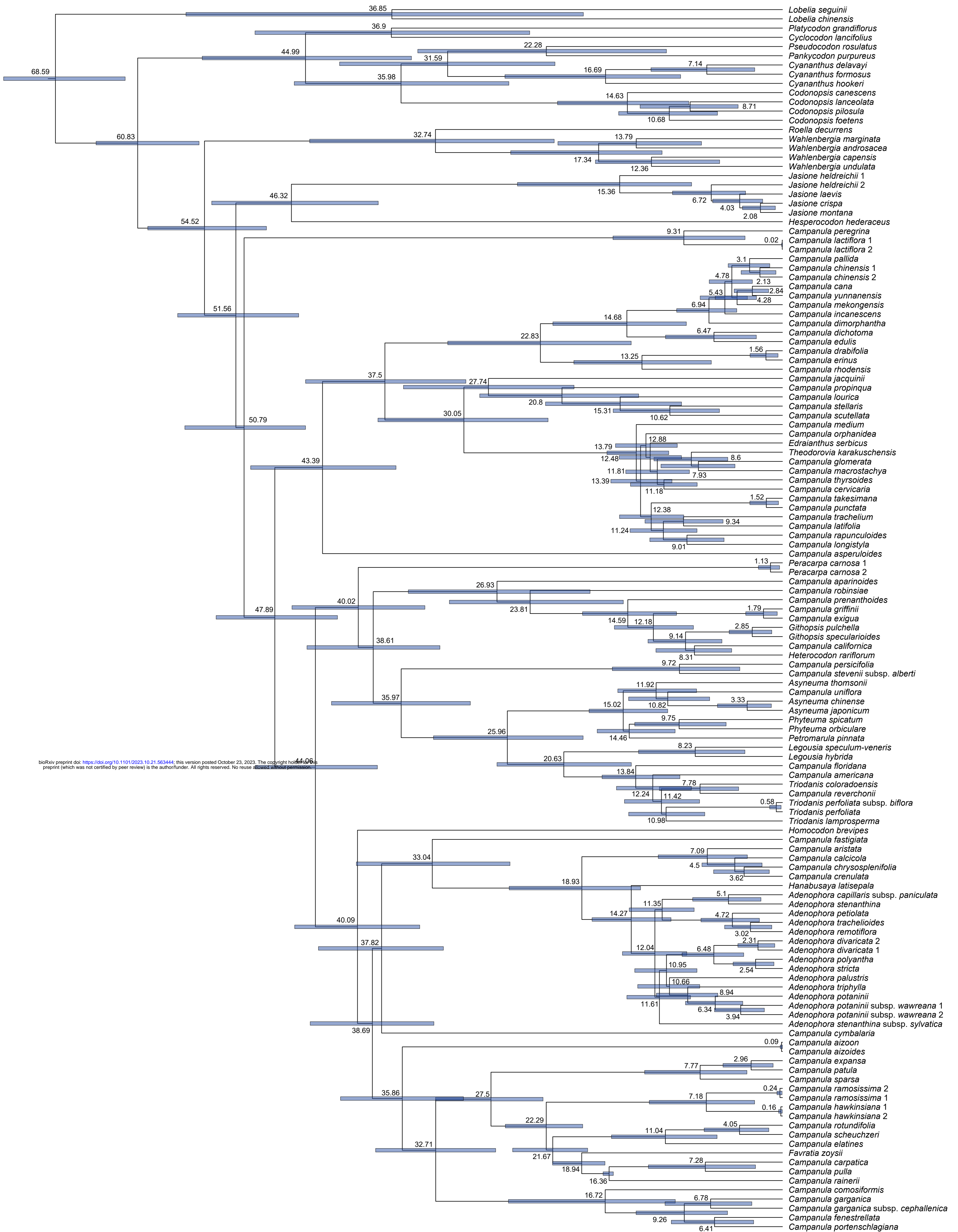
bioRxiv preprint doi: <https://doi.org/10.1101/2023.10.21.563444>; this version posted October 23, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



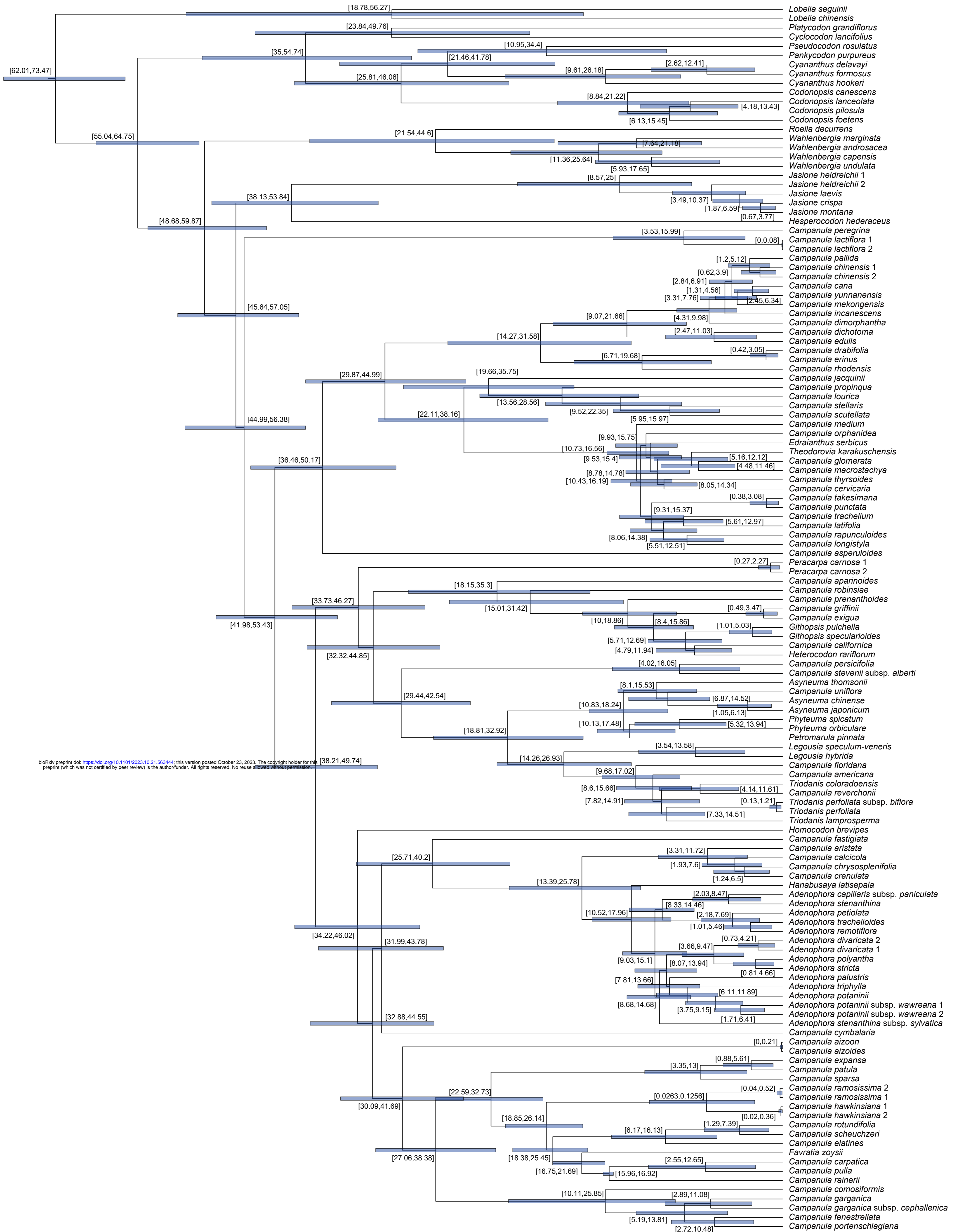
bioRxiv preprint doi: <https://doi.org/10.1101/2023.10.21.563444>; this version posted October 23, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

LEGEND



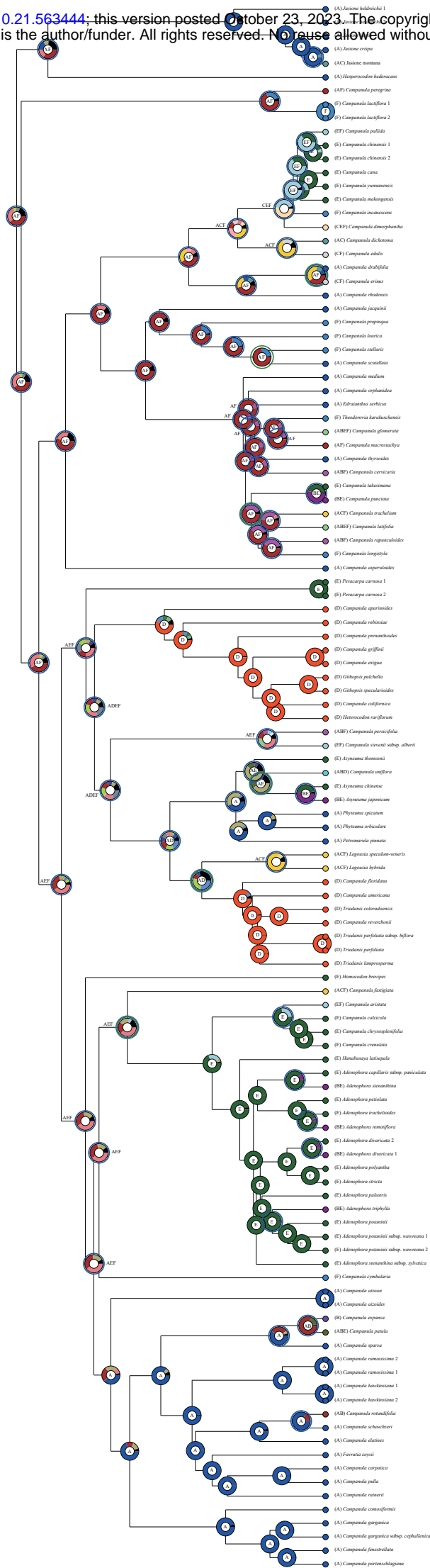


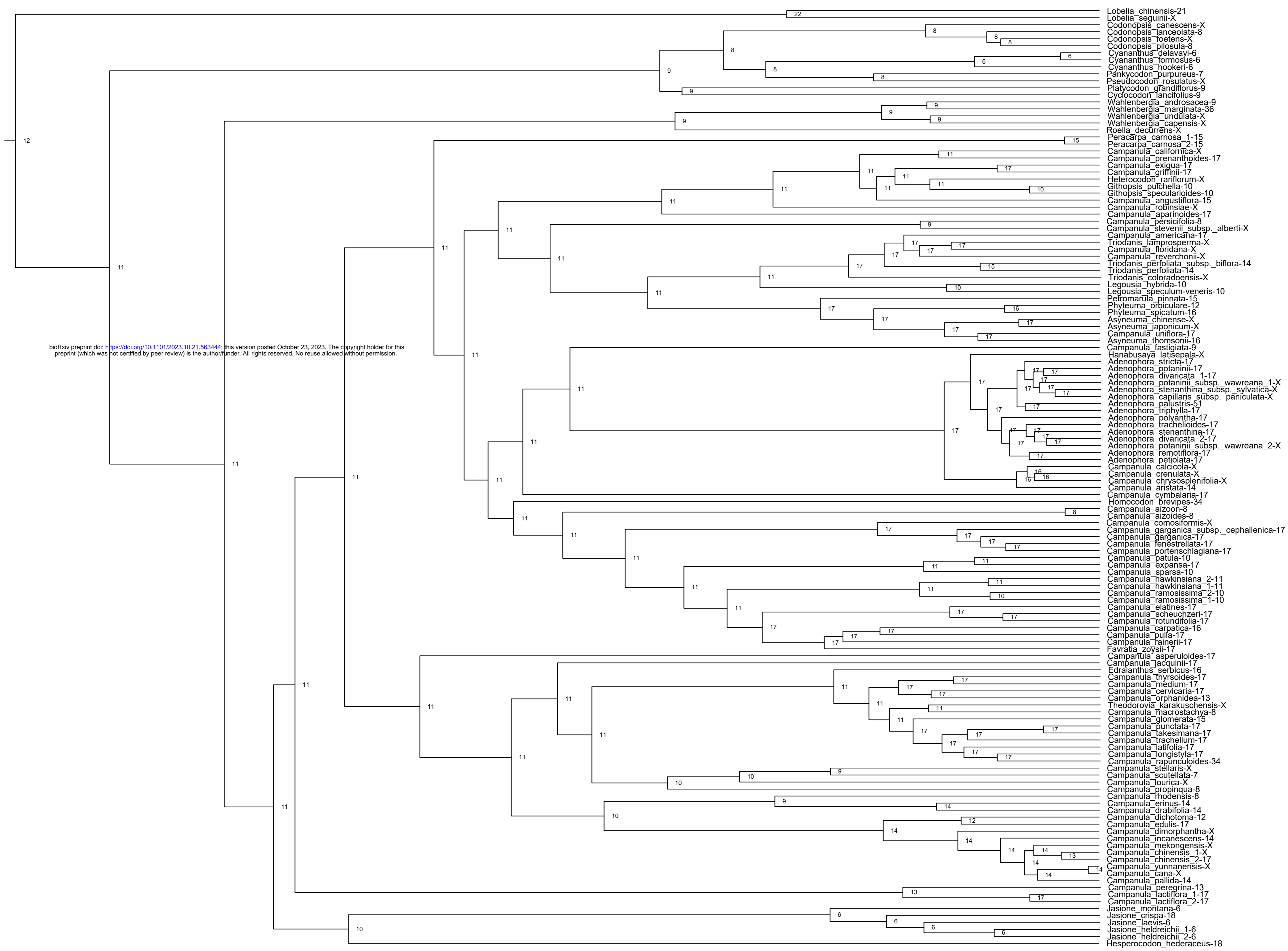
bioRxiv preprint doi: <https://doi.org/10.1101/2023.10.21.563444>; this version posted October 23, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



bioRxiv preprint doi: <https://doi.org/10.1101/2023.10.21.563444>; this version posted October 23, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

LEGEND





bioRxiv preprint doi: <https://doi.org/10.1101/2023.10.21.563444>; this version posted October 23, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.