A phylotranscriptome study using silica gel-dried leaf tissues produces an updated robust phylogeny of Ranunculaceae

Running title: RNA-seq using silica gel-dried tissues

Jian He[1†], Rudan Lyu[1†], Yike Luo[1†], Jiamin Xiao[1†], Lei Xie[1]*, Jun Wen[2]*, Wenhe Li[1], Linying Pei[4], Jin Cheng[3]

[1] School of Ecology and Nature Conservation, Beijing Forestry University, Beijing, 100083 PR China

[2]Department of Botany, National Museum of Natural History, MRC 166, Smithsonian Institution, Washington, DC 20013-7012, USA

[3] Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, College of Biological Sciences and Technology, Beijing Forestry University, Beijing, 100093 PR China

[4]Beijing Engineering Technology Research Center for Garden Plants, Beijing Forestry University Forest Science Co. Ltd., Beijing, 100083, PR China

[†]These authors contributed equally to this work.

Correspondence: Lei Xie, email: xielei@bjfu.edu.cn; Jun Wen, email: wenj@si.edu.

Abstract

The utility of transcriptome data in plant phylogenetics has gained popularity in recent years. However, because RNA degrades much more easily than DNA, the logistics of obtaining fresh tissues has become a major limiting factor for widely applying this method. Here, we used Ranunculaceae to test whether silica-dried plant tissues could be used for RNA extraction and subsequent phylogenomic studies. We sequenced 27 transcriptomes, 21 from silica gel-dried (SD-samples) and six from liquid nitrogen-preserved (LN-samples) leaf tissues, and downloaded 27 additional transcriptomes from GenBank. Our results showed that although the LN-samples produced slightly better reads than the SD-samples, there were no significant differences in RNA quality and quantity, assembled contig lengths and numbers, and BUSCO comparisons between two treatments. Using this data, we conducted phylogenomic analyses, including concatenated- and coalescent-based phylogenetic reconstruction, molecular dating, coalescent simulation, phylogenetic network estimation, and whole genome duplication (WGD) inference. The resulting phylogeny was consistent with previous studies with higher resolution and statistical support. The 11 core Ranunculaceae tribes grouped into two chromosome type clades (T- and R-types), with high support. Discordance among gene trees is likely due to hybridization and introgression, ancient genetic polymorphism and incomplete lineage sorting. Our results strongly support one ancient hybridization event within the R-type clade and three WGD events in Ranunculales. Evolution of the three Ranunculaceae chromosome types is likely not directly related to WGD events. By clearly resolving the Ranunculaceae phylogeny, we demonstrated that SD-samples can be used for RNA-seq and phylotranscriptomic studies of angiosperms.

Keywords

chromosomal type, phylotranscriptomics, Ranunculaceae, RNA-seq, silica-dried leaf tissue

1 INTRODUCTION

With the recent advances in high-throughput sequencing and analytical methods, phylogenetic reconstruction using genome-wide sequence data has become widely used in plant evolutionary studies (Johnson et al., 2012; Yu et al., 2018). However, whole-genome sequencing of densely sampled phylogenetic analyses has remained impractical and unnecessary due to high costs and computational limitations. Hence researchers have developed reduced-representation methods (e.g., genome skimming, restriction site-associated DNA sequencing or RAD-seq, target enrichment sequencing such as Hyb-seq, and transcriptome sequencing or RNA-seq) as practical tools for phylogenetic studies (Zimmer & Wen, 2015; McKain et al., 2018).

Genome skimming is one of the most widely applied partitioning strategies for phylogenetic inferences, and especially efficient for obtaining complete plastid genome sequences of plants (Dodsworth, 2015; Liu et al., 2018; Zhai et al., 2019; He et al, 2019; Liu et al., 2020; Wang et al, 2020). However, it is often of limited utility to obtain enough single copy nuclear genes for phylogenetic analyses, especially for non-model plant taxa possessing both huge genome sizes and no whole-genome reference (McKain et al., 2018; but see Liu et al., 2021). Hyb-seq is another widely used genome partitioning method that can target low-copy nuclear genes. Like genome skimming, Hyb-seq can use almost all kinds of tissue samples (e.g., silica gel-preserved, flash-frozen, fresh, or even old herbarium materials) (Yu et al., 2018; McKain et al., 2018; Reichelt et al., 2021; Wang et al., 2021). However, Hyb-seq requires a complex laboratory protocol involving bait capture. Furthermore, this method often results in a high proportion of missing data, and also cannot be used to detect ancient whole genome duplication (WGD) based on paralogous genes.

In recent years, using RNA-seq to reconstruct phylogenetic relationships (phylotranscriptomics) and gene family evolution has gained popularity because of its relatively low cost and improved analytical pipelines (Wen et al., 2013, 2015; Wickett et al., 2014; Landis et al., 2017; Zeng et al., 2017; One Thousand Plant Transcriptomes Initiative, 2019; Cheon et al., 2020; Alejo-Jacuinde et al., 2020). Using that method, researchers can often assemble thousands of genes (especially single-copy nuclear genes) from plant taxa under study and use them for both inferring phylogenetic relationships and gene family evolution (Yang and Smith, 2014; Yang et al., 2015; Xiang et al., 2017). Compared to Hyb-seq, RNA-seq uses relatively simple experimental protocols to generate more complete nuclear data, and paralogous genes from RNA-seq data can also be used to infer WGD (McKain et al., 2018).

While improvements in sequencing and extraction protocols have made RNA-seq much easier in plants (Romero et al., 2014; Yang et al., 2017), its application in phylogenetic study still remains challenging. Because RNA is more unstable than DNA, RNA-seq requires more stringent material preservation techniques. Previous studies have used fresh or liquid-nitrogen flash frozen plant tissues or fresh tissue quickly soaked in RNA stabilization solution, and then subsequently preserved in an ultra-low temperature (-80 °C) freezer (Yu et al., 2018; McKain et al., 2018; Dodsworth et al., 2019). However, because phylotranscriptomic studies often focus on non-model plant taxa, they usually require extensive field work and broad taxon sampling schemes. The logistics of using liquid nitrogen tanks in the field or expensive RNA stabilization solution to preserve collected plant tissues at multiple locations, not to mention quick access to an ultra-low temperature freezer for subsequent laboratory work, greatly limits the practicality of using RNA-seq for phylogenetic studies (Zimmer & Wen, 2015; Yang et al., 2017).

Traditional DNA-based molecular phylogenetics, DNA barcoding, as well as genome skimming and RAD-seq methods, often use silica gel-dried leaf tissues (Narzary et al., 2015; Yu et al., 2018). Such sampling method is cheaper and amenable to collecting and transporting large number of samples. Much emphasis has been placed on how different plant tissues, preservation methods, and RNA extraction protocols may impact the quantity and quality of extracted RNA (Johnson et al., 2012; Romero et al., 2014; Yang et al., 2017), but no empirical study has explored silica gel-dried plant tissues for RNA-seq. This may be due to the reasons that even though total RNA may be extracted from silica gel-dried plant tissues, it is not possible to quantitatively measure gene expression using this kind of samples for evo-devo studies. However, a phylotranscriptomic study needs to obtain large nuclear data sets for pertinent plant taxa for analysis, and it does not need to quantitatively measure gene expression. Therefore, if silica gel-dried and moderately frozen (-20 ℃) plant tissues can be sequenced via RNA-seq, it may greatly expand the utility of RNA-seq in phylogenetic studies.

As a member of the early-diverging eudicots, the family Ranunculaceae consists of more than 50 genera and 2,000 species, and it has been widely used as a model for angiosperm evolutionary studies (Tamura, 1995; Emadzade et al., 2010; Johansson and

Jansen 1993; Ro et al. 1997; Wang et al., 2016; Zhai et al., 2019). Systematists have also long focused on the family (Wiegand 1895; Smith 1926; Langlet 1927, 1932; Wodehouse 1936; Gregory 1941; Tamura, 1995; Ro et al., 1997; Zhai et al., 2019), and it has been used as a model for evolutionary studies addressing questions on floral development, genome duplication, and key innovations (Kramer et al., 2003, Kramer, 2009; Rasmussen et al., 2009; Kramer and Hodges 2010; Galimba et al., 2012; Zhang et al., 2013; Wang et al., 2015; Damerval and Becker, 2017; Aköz and Nordborg 2019; Shi and Chen 2020).

Traditionally, systematic studies of the Ranunculaceae emphasized morphological characters (especially floral traits and fruit types) (Prantl, 1887). Because of its complex evolutionary history, the infrafamilial classifications of Ranunculaceae have remained highly controversial (Prantl, 1887; Hutchinson, 1923; Takhtajan, 1980; Cronquist, 1981; Tamura, 1995). Since the studies by Langlet (1927, 1932), cytological characters, including chromosome size and base number, have become important for delimiting subfamilies and tribes of the family (Janchen, 1949; Tamura, 1995). Since the 1990s, many molecular phylogenetic studies of the Ranunculaceae based on Sanger sequencing (Johansson and Jansen, 1993; Hoot, 1995; Johansson, 1995; Ro et al., 1997; Hoot et al., 2008; Wang et al., 2009, 2016; Cossard et al., 2016) have helped clarify many systematic issues, including the delimitation of the family and its phylogenetic framework. Yet detailed tribal relationships within the family have remained controversial and conflicts among different gene trees were common in the family. Recently, phylogenomic studies using complete plastid genome data have greatly improved the resolution and statistical support of many previously unresolved clades within the family (Zhai et al., 2019; He et al., 2019). However, because plastid genome sequences are uniparentally inherited, they may not truly reflect phylogenetic relationships among taxa (Greiner et al., 2015). Therefore, the tribal relationships, especially within the core Ranunculaceae (defined below), still need further clarifications using multiple single-copy nuclear markers.

Using molecular data, Wang et al. (2009, plastid and nrDNA sequences) and Zhai et al. (2019, complete plastid genome sequences) resolved 14 tribes within the Ranunculaceae. Except for three early diverged tribes (Glaucideae, Hydrastideae, and Coptideae) that have *Coptis* type (C-type) chromosomes, the remaining 11 tribes formed the core Ranunculaceae clade. Within the core Ranunculaceae clade, there are two major chromosome types: the *Ranunculus* type bearing large chromosomes with a base number of 8 (R-chromosome group, or R-type) and the *Thalictrum* type having short, small chromosomes with a base number of 7 or 9 (T-chromosome group, T-type) (Gregory, 1941). The genome sizes between these two chromosome types vary 180-fold, with the smallest genome size (1C value: 240.10 MB) in *Thalictrum rhychocarpum* (T-type chromosome) and the largest genome size in *Hepatica nobilis* (R-type chromosome, 1C value: 43,708 MB) (https://cvalues.science.kew.org/). The C-type chromosome genome sizes, from 1,150 MB (*Coptis*, Liu et al., 2021) to 1,310 MB (*Hydrastis*, https://cvalues.science.kew.org/), lie between those of the T- and R-type chromosome taxa, but closer to the T-type.

Genome size variations across eukaryotic taxa often result from macro-mutations, such as WGDs and repetitive element proliferation (Marburger et al., 2018), and WGD provides raw materials for evolutionary novelty in seed plants and has been widely reported in angiosperms (Ren et al., 2018; Cai et al., 2019; One Thousand Plant Transcriptomes Initiative, 2019; Park et al., 2020). Although several recent studies have discussed WGD in some Ranunculaceae taxa (Park et al., 2020; Xie et al., 2020; Liu et al., 2021), their sampling was limited. Until now, the lack of a well-sampled phylogenetic framework of Ranunculaceae impedes our ability to understand the evolutionary processes of key characters, such as floral and fruit characters. Moreover, the exploration of genome-wide gene tree discordance is essential for understanding the underlying evolutionary processes of Ranunculaceae. We also need a robust phylogeny within which to investigate how the three chromosome types evolved and whether ancient WGD events were important in that evolution.

Here, we extracted RNA from both liquid-nitrogen frozen and silica gel-dried leaf tissues and conducted a phylotranscriptomic study of Ranunculaceae. We compared the quality and quantity of extracted RNA between liquid-nitrogen and silica-preserved samples, assessed whether silica-dried plant tissues can be used for phylogenomic study, used multiple single-copy nuclear datasets to reconstruct the Ranunculaceae phylogenetic framework, analyzed the discordance among the gene trees (nuclear as well as cyto-nuclear discordance), and explored divergence times and WGD events in Ranunculaceae and its closely related taxa.

2 MATERIALS AND METHODS

2.1 Plant Material

Because previous phylotranscriptomic studies mainly used transcriptome data from vegetative tissues (Johnson et al., 2012; Zeng et al., 2017; One Thousand Plant Transcriptomes Initiative, 2019), we likewise collected 27 leaf samples from 22 genera of Ranunculales (20 Ranunculaceaae and two Circaeasteraceae genera) for our study. Among them, 21 leaf samples were dried with silica gel (SD-sample) and the other six were liquid nitrogen-preserved leaf tissues (LN-sample). According to other studies (Johnson et al., 2012; McKain et al., 2018), young developing leaf tissues of the 21 SD-samples were dried in blue Tel-Tale silica gel (Sinopharm Chemical Reagent Co. Ltd., Shanghai, size of particles: 2-8 mm), which are widely used in preserving DNA for plant molecular phylogenetic studies. Once completely dried, the leaf tissues were deposited in a -20 °C freezer. The six LN-samples (also young developing leaf tissues) were quickly deposited in liquid nitrogen and then ground for RNA extraction. Voucher specimens were deposited in the Herbarium of Beijing Forestry University (Table 1).

Table 1 Information about samples used in this study

Family	Species	Plant material preservation	Locality of newly sequenced samples	SRA[†] number and voucher	Dates of collecting/library construction	Reference
Ranunculaceae	*Adonis sibirica*	silica-dried	Altay, Xinjiang, China	J. He 20180525-01 (BJFC)	25 May 2018/ 29 July 2019	This study
Ranunculaceae	*Anemoclema glaucifolium*	silica-dried	Cangshan, Yunnan, China	L. Xie 20190715-04 (BJFC)	15 July 2019/ 9 October 2019	This study
Ranunculaceae	*Asteropyrum peltatum*	silica-dried	Emei, Sichuan, China	L. Xie 20180411-1 (BJFC)	11 April 2018/ 29 July 2019	This study
Ranunculaceae	*Beesia calthifolia*	silica-dried	Zhouqu, Gansu, China	J. He 2019060803 (BJFC)	08 June 2019/ 29 July 2019	This study
Ranunculaceae	*Callianthemum alatavicum*	silica-dried	Urumqi, Xinjiang, China	Z. Z. Yang 0527 (BJFC)	27 May 2018/ 29 July 2019	This study
Ranunculaceae	*Callianthemum taipaicum*	silica-dried	Taibai, Shaanxi, China	J. He 2019061506 (BJFC)	15 June 2019/ 29 July 2019	This study
Ranunculaceae	*Caltha scaposa*	silica-dried	Cangshan, Yunnan, China	L. Xie 20190715-02 (BJFC)	15 July 2019/ 21 August 2019	This study
Ranunculaceae	*Cimicifuga mairei*	silica-dried	Kangding, Sichuan, China	L. Xie 2018081003 (BJFC)	10 August 2018/ 29 July 2019	This study
Ranunculaceae	*Clematis leschenaultiana2*	silica-dried	Tiane, Guangxi, China	L. Xie 2017112107 (BJFC)	21 November 2017/ 16 January 2020	This study
Ranunculaceae	*Clematis montana*	silica-dried	Xiaozhongdian, Yunnan, China	L. Xie 2019071601 (BJFC)	16 July 2019/ 16 January 2020	This study
Ranunculaceae	*Clematis tangutica*	silica-dried	Ma'erkang, Sichuan, China	L. Xie 2018070414 (BJFC)	04 July 2018/ 16 January 2020	This study
Ranunculaceae	*Delphinium anthriscifolium*	silica-dried	Gulin, Sichuan, China	L. Xie 20180406-8 (BJFC)	06 April 2018/ 29 July 2019	This study
Ranunculaceae	*Dichocarpum dalzielii*	silica-dried	Emei, Sichuan, China	L. Xie 20180411-2 (BJFC)	11 April 2018/ 29 July 2019	This study
Ranunculaceae	*Halerpestes cymbalaria*	silica-dried	Erdu, Beijing, China	L. Xie 2018052112 (BJFC)	21 May 2018/ 29 July 2019	This study
Ranunculaceae	*Hepatica henryi*	silica-dried	Emei, Sichuan, China	L. Xie 20180412-2 (BJFC)	12 April 2018/ 29 July 2019	This study
Ranunculaceae	*Leptopyrum fumarioides*	silica-dried	Weichang, Hebei, China	W.H. Li, 202005001 (BJFC)	10 May 2020/ 27 May 2020	This study
Ranunculaceae	*Oxygraphis glacialis*	silica-dried	Ganzi, Sichuan, China	L. Xie 2018081205 (BJFC)	12 August 2018/ 29 July 2019	This study
Ranunculaceae	*Oxygraphis glacialis*	silica-dried	Urumqi, Xinjiang, China	Z. Z. Yang 0527 (BJFC)	27 May 2018/ 29 July 2019	This study
Ranunculaceae	*Paraquilegia microphylla*	silica-dried	Lhasa, Tibet, China	S. J. Xu 2018060707 (BJFC)	07 June 2018/ 29 July 2019	This study
Ranunculaceae	*Souliea vaginata*	silica-dried	Zhouqu, Gansu, China	J. He 2019060815 (BJFC)	08 June 2019/ 29 July 2019	This study
Ranunculaceae	*Trollius ranunculoides*	silica-dried	Ganzi, Sichuan, China	L. Xie 20190721-11 (BJFC)	21 July 2019/ 21 August 2019	This study
Circaeasteraceae	*Circaeaster agrestis*	liquid nitrogen	Zhouqu, Gansu, China	J.He zq03 (BJFC)	8 June 2019/ 29 June 2019	This study
Circaeasteraceae	*Kingdonia uniflora*	liquid nitrogen	Zhouqu, Gansu, China	J.He zq04 (BJFC)	8 June 2019/ 29 June 2019	This study
Ranunculaceae	*Aquilegia viridiflora*	liquid nitrogen	Jiufen, Beijing, China	J. He RNAHJ-002 (BJFC)	4 April 2019/ 11 April 2019	This study
Ranunculaceae	*Clematis leschenaultiana*	liquid nitrogen	Tiane, Guangxi, China	L. Xie 2017112107 (BJFC)	10 November 2019/15 November 2019	This study

Ranunculaceae	*Helleborus thibetanus*	liquid nitrogen	Cult. in Beijing Forestry University	J. He C2018001 (BJFC)	20 March 2019/ 25 March 2019	This study
Ranunculaceae	*Pulsatilla chinensis*	liquid nitrogen	Jiufen, Beijing, China	J. He RNAHJ-003 (BJFC)	4 April 2019/ 11 April 2019	This study
Amborellaceae	*Amborella trichopoda*†	Traditional		ERR364329		Matasci et al., 2014
Poaceae	*Avena fatua*†	Traditional		SRR8926019		Chabikwa et al., 2020
Proteaceae	*Macadamia integrifolia*†	Traditional		SRR8926019		Chabikwa et al., 2020
Trochodendraceae	*Trochodendron aralioides*†	Traditional		ERR2040149		Matasci et al., 2014
Eupteleaceae	*Euptelea pleiosperma*†	Traditional		ERR2040156		Matasci et al., 2014
Papaveraceae	*Chelidonium majus*†	Traditional		ERR2040181		Matasci et al., 2014
Lardizabalaceae	*Akebia trifoliata*†	Traditional		ERR2040157		Matasci et al., 2014
Menispermaceae	*Menispermum dauricum*†	Traditional		SRR8298315		Dong et al., 2019
Berberidaceae	*Berberis thunbergii*†	Traditional		SRR7441386		NA
Ranunculaceae	*Aconitella aconiti*†	Traditional		SRR2154260		NA
Ranunculaceae	*Aconitum japonicum*†	Traditional		DRR062390		NA
Ranunculaceae	*Adonis amurensis*†	Traditional		SRR8567627		Zhou et al., 2019
Ranunculaceae	*Anemone flaccida*†	Traditional		SRR3233423		Zhan et al., 2016
Ranunculaceae	*Anemone hupehensis*†	Traditional		ERR2040184		Matasci et al., 2014
Ranunculaceae	*Aquilegia formosa*†	Traditional		SRR7830420		Filiault et al., 2018
Ranunculaceae	*Coptis chinensis*†	Traditional		SRR5947132		He et al., 2018
Ranunculaceae	*Coptis teeta*†	Traditional		SRR6705074		He et al., 2018
Ranunculaceae	*Eranthis hyemalis*†	Traditional		SRR7271011		Meesapyodsuk et al., 2018
Ranunculaceae	*Hydrastis canadensis*†	Traditional		SRR341992		NA
Ranunculaceae	*Nigella sativa*†	Traditional		SRR341997		Hagel et al., 2015
Ranunculaceae	*Pulsatilla patens*†	Traditional		SRR10230555		NA
Ranunculaceae	*Ranunculus brotherusii*†	Traditional		SRR1822558		Zhao et al., 2016
Ranunculaceae	*Ranunculus bungei*†	Traditional		SRR1822529		Chen et al., 2015
Ranunculaceae	*Ranunculus sceleratus*†	Traditional		SRR3291759		Zhao et al., 2016
Ranunculaceae	*Thalictrum hernandezii*†	Traditional		SRR6869419		Morales-Briones et al., 2019
Ranunculaceae	*Thalictrum thalictroides*†	Traditional		ERR2040185		Matasci et al., 2014
Ranunculaceae	*Xanthorhiza simplicissima*†	Traditional		SRR765879		Hagel et al., 2015

†Downloaded from NCBI Sequence Read Archive

Next, we downloaded 27 published transcriptomes from GenBank—18 from the Ranunculaceae, six from other Ranunculales families, and three from distantly related angiosperms that served as outgroups. Our sampling covered 13 of all 14 Ranunculaceae tribes (Wang et al., 2009, 2016; Zhai et al., 2019), only the first diverged clade, Glaucideae, was not available for the study. All 27 downloaded transcriptomes were obtained from either freshly collected, liquid-nitrogen-preserved, or RNA stabilization solution-treated leaf tissues. We refer to them as the traditionally treated samples, T-samples.

2.2 RNA extraction and sequencing, and cleaning raw reads

We sent our collected leaf tissues to Biomarker Technologies Corporation (http://www.biomarker.com.cn) for RNA extraction, library construction, and next generation sequencing as follows. For each SD- and LN-sample, 100 mg leaf tissues were ground and then total RNA was extracted using a conventional TRIzol-based RNA extraction kit (TRIzon, CoWin Biosciences, Jiangsu, PR China) following the manufacturer's protocol (see Suplementary Materials). Then, the extracted RNA samples were run on standard RNase-free 1.0% agarose gels and measured using a NanoDrop 2000 spectrophotometer (ThermoFisher Scientific, Waltham, Massachusetts, USA) at 230, 260, and 280 nm to assess the purity and intactness of the samples. Subsequently, the OD260/280 and OD260/230 ratios were calculated. RNA integrity numbers (RIN) and rRNA ratios (28S/18S) were detected by a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California, USA). Next, the cDNA library was generated from the total RNA extracts by using reverse transcription PCR. The libraries were sequenced on a NovaSeq 6000 platform (Illumina, San Diego, California, USA), ultimately yielding about 6 Gb of raw data per sample. To obtain high-quality clean data, the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/download.html) was used to remove adaptors and low-quality reads from the raw reads. All raw reads were deposited in the NCBI Sequence Read Archive (SRA, BioProject: PRJNA739247).

The 27 downloaded transcriptomes retrieved from the NCBI SRA database were unzipped using the sratoolkit v.2.9.2 (https://ncbi.github.io/sra-tools/). Low-quality bases were removed using Trimmomatic v.0.38 (Bolger et al., 2014).

2.3 Transcriptome assembly, gene family clustering, and inferring orthology

The 54 clean transcriptome data were *de novo* assembled using Trinity v.2.5.1 (Grabherr et al., 2011) with default parameters, and the longest isoforms of related contigs were selected using the Trinity helper script (get_longest_isoform_seq_per_trinity_gene.pl). To reduce redundancy, we ran the raw assemblies through CD-HIT v.4.6.2 (Fu et al., 2012) with parameters -c 0.95 and -n 8 (Feng et al., 2019) and then assessed each assembly's completeness by comparing them against single-copy orthologs in the BUSCO database (Simão et al., 2015). Coding regions of each putative transcript were predicted using TransDecoder v.5.0 (https://github.com/TransDecoder/TransDecoder/releases).

To assign the *de novo* assembled sequences into single copy orthologous gene (SCOG) families, we used an integrated method proposed by Cai et al. (2019) that rigorously takes sequence similarity and species phylogeny into account. To begin, we first constructed transcriptome homology scans using Proteinortho v.6.0.10 (Lechner et al., 2011) in the Diamond mode (Buchfink et al., 2015) and then searched the resulting clusters to identify gene families containing at least 38 in-group species (> 70%), which then obtained 3,634 candidate homologous clusters.

To reduce errors that sometimes occur in this similarity-based homology search because of deep paralogs, mis-assembly, or frame shifts (Yang & Smith, 2014), we ran multiple scans to increase statistical and inferential powers. Since the SCOGs could have been clustered not only from nuclear genomes but also from organellar genomes, we deleted all the organellar SCOGs by using a BLAST analysis against a *Clematis* chloroplast genome (NC039579) and a *Nelumbo* mitochondrial genome (NC030753).

Considering that a sequence clustered to an erroneous ortholog often shows an unexpectedly long branch in the gene tree (Mai and Mirarab, 2018), we used Treeshrink v.1.3.9 (Mai and Mirarab, 2018) to scrutinize each gene tree, filtering out sequences with long branches, and deleting them from the alignments.

Each gene tree was generated by Treeshrink v.1.3.9 in the following four steps: (1) align the peptide sequences of each putative SCOG with MAFFT v.7.475 (Katoh and Standley, 2013); (2) convert the alignments into the corresponding nucleotide alignments; (3) delete each nucleotide site with more than 20% missing data by using a python script (https://github.com/Jhe1004/DelMissingSite); and (4) reconstruct gene trees using a maximum likelihood (ML) method implemented in RAxML v.8.2.12 (Stamatakis, 2014) and the GTR+G model with 100 bootstrap replicates. Sequences that generated long branches in gene trees were removed from the alignments (Mai and Mirarab, 2018) and then, using a python script (https://github.com/Jhe1004/ConcatenateAlignment), the remaining alignments were concatenated into a super-alignment for further analysis.

Finally, to improve the accuracy of gene trees, we reconstructed the Ranunculaceae phylogenetic framework using three different strategies. (1) Limit the minimum length of each SCOG alignment (Yang et al., 2015). (2) Limit each SCOG tree's minimum average bootstrap values (Molloy and Warnow, 2018). (3) Limit the allowed maximum of missing taxa for each SCOG alignment (Yang et al., 2015). For each strategy, we set two different screening thresholds (described below).

We obtained eight data sets for phylogenetic analysis: (1) 3,611 concatenated SCOGs (SCOG-concatenated), (2) 3,611 separate SCOGs (SCOG-coalescent), (3) SCOGs each with at least 1,000 bp (SCOG-1000bp), (4) SCOGs each with at least 2,000 bp (SCOG-2000bp), (5) SCOGs each with at least 50% mean bootstrap values (SCOG-50BS), (6) SCOGs each with at least 80% mean bootstrap values (SCOG-80BS), (7) SCOGs each with no more than 20% missing taxa (SCOG-20MS), and (8) SCOGs each with no more than 10% missing taxa (SCOG-10MS).

2.4 Phylogenetic reconstruction

For phylogenomic analysis, we used both coalescence- and concatenation-based methods for each data set. For the coalescent-based method, the Accurate Species Tree Algorithm v.4.4.4 (ASTRAL; Zhang et al., 2018) was applied. Single-gene ML trees for seven of the data sets (not the SCOG-concatenated data set) were reconstructed in RAxML v.8.2.12 on the CIPRES Science Gateway v.3.3 (Miller et al., 2010) using the GTR+G model. All those trees were used as input in ASTRAL for species tree inference. For the concatenation method, all the SCOG-concatenated data sets were calculated for using ML method implemented in RAxML v.8.2.12 with the GTR+G model and bootstrap percentages computed after 200 replicates. In order to test cyto-nuclear discordance, the plastome sequence data were obtained from previously published studies (Table S1) and the plastome phylogeny was inferred using the ML method as described above.

2.5 Molecular dating

The molecular clock estimation used the SCOG-concatenated data set and constrained with the species tree topology. We used three widely accepted fossil calibrations in our dating analysis. (1) The Ranunculales stem group (the tree root) was constrained at a minimum age of 113 Mya (range = 113–125 Mya), which was based on the earliest (early Cretaceous) mega-fossil (flower) record for the order, *Teixeiria lusitanica* (von Balthazar et al., 2005; Magallón et al., 2015). (2) *Prototinomiscium vangerowii* gave the stem age of the Menispermaceae (Knobloch and Mai, 1986; Kadereit et al., 2019) at a minimum of 91 Mya (range = 91–99 Mya). (3) The stem of the *Actaea* clade was constrained at a minimum age of 56 Mya (range = 56–67 Mya), based on the fossil record of *Paleoactea nagelii* (Pigg and DeVore, 2005; Wang et al., 2016).

Divergence times were estimated using MCMCTree implemented in the PAML v.4.9a package (Yang, 2007), and it was run in two steps. First, the topology was constrained by the ASTRAL species tree; then the ML estimates of the branch lengths, the gradient vector, and the Hessian matrix were generated using the codeml model from the PAML v.4.9a package (Yang, 2007). Then, those variables and data (Hessian matrix) were fed into MCMCTree control files, together with the GTR substitution model. An autocorrelated relaxed clock model (Rannala and Yang, 2007) was used to estimate the divergence time posterior distribution and the MCMC chain ran for 100,000,000 generations, with tree sampling every 1,000 generations after a burn-in of 25,000,000 iterations. We used FigTree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree) to generate the chronogram.

2.6 Visualizing and assessing conflicts among the gene trees

We used Phyparts v.0.0.1 (Smith et al., 2015) to explore conflicts among the nuclear gene trees. To reduce computational time, we mapped gene trees from the SCOG-1000bp data set onto the species tree generated by the SCOG-coalescent data set, and the numbers of discordant and concordant nodes of the gene trees compared to those of the species tree were then counted and summarized. To further visualize conflicts, we built cloud tree plots using the python package Toytree v.2.0.5 (Eaton, 2020). Because cloud tree plots cannot accommodate missing taxa among gene trees, we used the python package ETE3 v.3.1.2 (Huerta-Cepas et al., 2016) to filter and prune the gene trees to include all Ranunculaceae samples and only one outgroup (*Berberis thunbergii*). Ultrametric gene trees were time-calibrated with TreePL v.1.0 (Smith and O'Meara 2012) and, using the python package DendroPy v.4.5.2 (Sukumaran and Holder 2010), each gene tree's nodes were calibrated by age estimates extracted from the MCMCTree dating results described above. Ultimately, we obtained 92 ultrametric gene trees for cloud tree plots.

2.7 Coalescence simulations

Discordance among gene trees is likely caused by incomplete lineage sorting (ILS), hybridization, or other sources of error and noise (Sang & Zhong, 2000; Roos et al., 2011; Morales-Briones et al., 2021). Among them, ILS has attracted great attention in previous studies (Edwards 2009), and several inference methods have been developed to accommodate ILS as the source of discordance (Edwards et al. 2016). To evaluate if ILS alone could explain the gene tree discordance in this study, we performed a coalescent simulating analysis following published studies (Wang et al., 2018; Yang et al., 2020; Morales-Briones et al., 2021). If the coalescent model fit the empirical gene trees well, the simulated gene trees from that model would be consistent with the empirical gene trees and ILS most likely explained the gene tree incongruence well. We used the function “sim.coaltree.sp” implemented in the R package Phybase v.1.5 (Liu and Yu, 2010) to simulate 10,000 gene trees under the multispecies coalescent (MSC) model. After the input ASTRAL ultrametric species tree was modified using a Perl script (Wang et al., 2018), we calculated the distribution of tree-to-tree distances (Robinson and Foulds, 1981) between the ASTRAL species tree and each empirical gene tree (SCOG-1000bp data set) using the Python package DendroPy v.4.5.2 (Sukumaran and Holder 2010). We then compared the distribution of distances between the ASTRAL species tree and the simulated gene trees. Because that calculation requires both trees to have the same set of taxa, we only kept 13 tribal (*sensu* Wang et al., 2016) representatives of the Ranunculaceae for this analysis.

We also carried out a coalescent simulation (Rose et al., 2021) to evaluate if ILS alone could explain cyto-nuclear discordance in Ranunculaceae. If ILS explained the cyto-nuclear discordance, a number of simulated gene trees would be consistent with the plastid genome tree. So, using the R package Phybase v.1.5, we simulated 10,000 gene trees from our ASTRAL species tree under the MSC model. Then we used Phyparts v.0.0.1 to explore conflicts among the simulated gene trees and the plastid genome tree. The proportion of gene tree concordance compared to the plastid genome tree was then counted and summarized.

2.8 Species network analysis

We used a Julia package, PhyloNetworks (Solís-Lemus et al., 2017), that models ILS and gene flow using a maximum pseudolikelihood method (Yu and Nakhleh, 2015) to investigate high probability hybridization events and to calculate γ, the vector that reflects the proportion of genes inherited from each parent in a hybrid lineage. Due to computational restrictions, the package cannot run large numbers of samples, so we kept the 13 tribal representatives of the Ranunculaceae for this analysis. From the input gene subtrees extracted from the SCOG-1000bp data set, we obtained 614 gene subtrees for the 13 taxa and, in PhyloNetworks, performed SNaQ (Species Networks applying Quartets) analyses that allowed for zero to five maximum hybridization events with ten independent runs. The best number of hybridization events was selected by plotting the pseudolikelihood scores. A sharp score increase is expected until it reaches the best number, after which the score increases slowly. Finally, we used 100 bootstrap replicates to perform hybrid branch support on the best network.

2.9 Detecting and dating WGD in the Ranunculaceae

In this study, WGD events were detected and located using two complimentary methods: a tree-based method to infer WGD and the absolute dating method to verify that result. For the tree-based method, input gene family trees were constructed from all 54 samples by using Proteinortho v.6.0.10 (Lechner et al., 2011) with a cut-off of $E = 1 \times 10^{-5}$. For each gene family, ML trees were estimated by RAxML v.8.2.12 under the GTR+G model and using a customized Python script (https://github.com/Jhe1004/reroot). The RAxML analysis output file (bestTree file) was re-rooted and transformed into an input file for the next step. Tree-based WGD inference was conducted using PhyParts v.0.0.1, which mapped the gene family trees to the species tree (based on the SCOG-coalescent data set using ASTRAL). After that, we recorded the nodes that contained duplication events when the descendants of the node had multiple gene copies for at least two terminal samples (Smith et al., 2015).

For the dating analyses, we followed methods of Vanneste et al. (2014). We used Proteinortho v.6.0.10 (Lechner et al., 2011) to obtain homologous gene families (including both orthologous and paralogous gene families) among the Ranunculaceae samples, ultimately finding 12,332 gene families that contain at least two paralogous copies. Then, those gene families were aligned using MAFFT v.7.475 and ML trees were reconstructed by RAxML v.8.2.12 under the GTR+G model. Next, we used TreePL (Smith & O’Meara, 2012) to date the divergence times of each gene family. For this analysis, we constrained both the root and the crown ages of the clade (Berberidaceae, Ranunculaceae) by using the time estimations found in our MCMCTree dating analysis. Then, we filtered the gene trees to ensure that Ranunculaceae and

Berberidaceae were sister groups. After all the gene families had been calculated, divergence time estimates of gene duplicates were extracted and WGD curves were plotted using Gaussian mixture models in the Python package sklearn v.0.24.1 (Pedregosa et al., 2011).

3 RESULTS

3.1 RNA-seq quality from silica gel -dried and traditionally treated tissues

The mean values of some RNA quality and quantity reference indices of LN-samples were higher than those of SD-samples (e.g., RIN, 28S/18S, and OD260/280), but were lower for other indices (e.g., OD260/230, total RNA, and RNA concentration) (Figure 1 a-b, Table S2). However, one-way analysis of variance (ANOVA) showed that none of those differences was significant at $P < 0.05$.

We then counted and compared the obtained reads, bases, Q20, Q30, and GC content of each treatment. Except for GC content, the other reference indices of LN-samples were significantly better than those of SD-samples (Figure 1 c-d, Table S2). We further assembled the transcripts for all the samples and determined the numbers and average lengths of the contigs (NC and ALC, respectively), the N50s, and the completeness of the assemblies and compared them to BUSCO (Figure 1e-h, Table S3). The results showed that, except for the mean value of NC, which is significantly lower in SD-samples than in LN-samples, all the reference indices were not significantly different between the two treatments.

We next pooled the transcriptome data of our six LN-samples with that of the 27 downloaded T-samples (collectively called T+LN-samples) and compared their combined NC, ALC, and N50 and the completeness of their assemblies, compared to BUSCO, to those indices of the SD-samples (Figure 1i-l, Table S4). The results demonstrated that SD-samples had significantly higher N50 than the T+LN-samples had (Figure 1k, $P < 0.05$, Supplementary Table S4), but there were no statistically significant differences between the NC, ALC, and BUSCO comparison indices (Figure 1i-l, Table S4). Finally, of the 3,611 SCOGs we had obtained for all the 54 samples, *Ranunculus sceleratus* (T-sample) had the most (3,453) and *Avena fatua* (T-sample) had the least (539) (Table S4). Also, the average number of SCOGs was higher in SD-samples (3,252) than in LN-samples (3,085) and significantly greater than in T-samples (2,939) (Figure 1h and l, Figure 2, Tables S2 and S3).

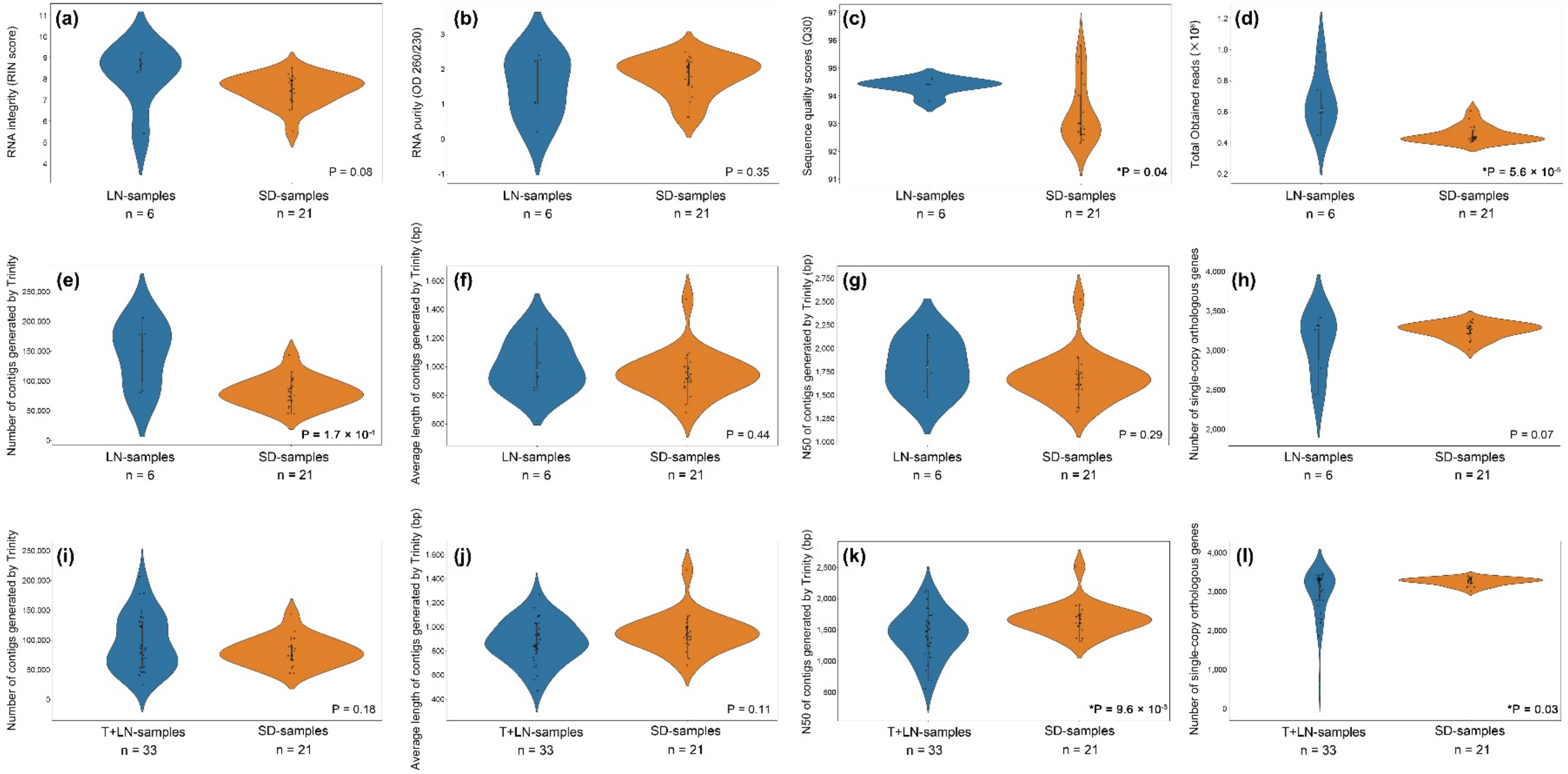


Figure 1 Quality and quantity assessments of RNA extracted from Ranunculales leaf tissues preserved in one of three different ways. Comparisons are of (a-d) genomic RNA, (e-g and i-k) de-novo sequences assembled by Trinity v.2.1.5 (Grabherr et al., 2011), and (h and l the numbers of single-copy orthologous genes compared to BUSCO (Simão et al., 2015). The LN+T samples category includes the 27 samples downloaded from GenBank (see Table 1) plus the liquid nitrogen-preserved samples (n = 6). One-way ANOVA, * $P < 0.05$. OD: optical density; Q30: Phred quality score 30 threshold. For detailed sampling information see Table 1

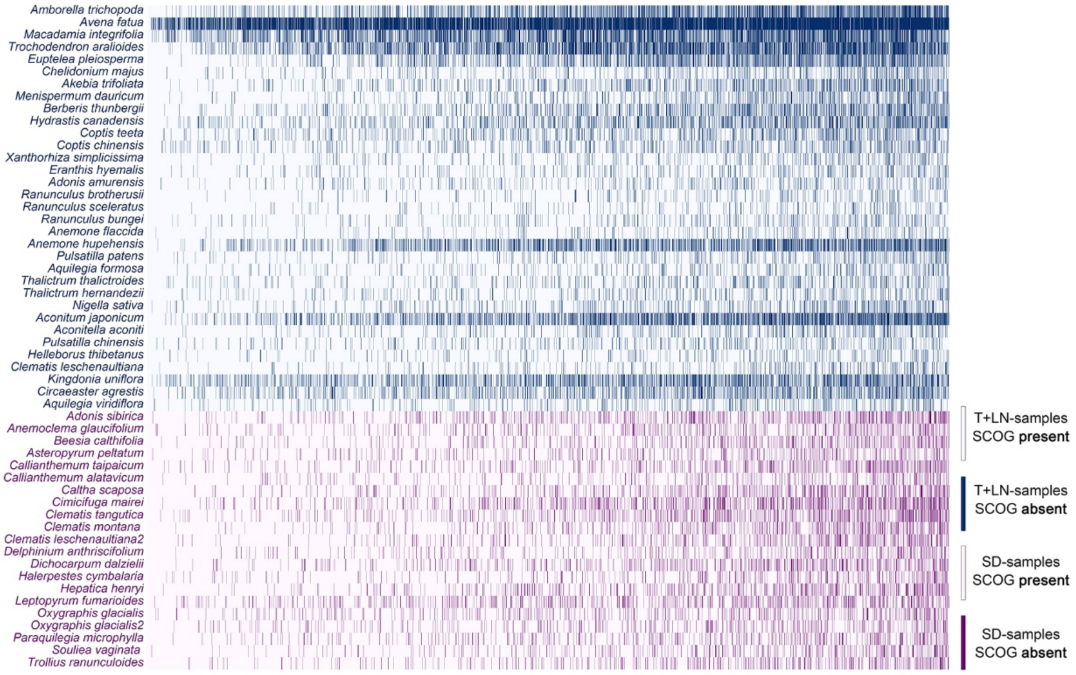


Figure 2 Heatmap of gene recovery efficiency. Each row represents one sample and each column represents one gene (3,611 genes total). Shading indicates a gene's absence. The blue zone shows traditional treatment samples (i.e., tissues preserved in liquid nitrogen or in RNA stabilization solution) (see Table 1). The purple zone shows the 21 silica-dried samples from the current study

3.2 Phylogenetic inference

The ML analysis of the 3,611 SCOG-concatenated data set produced a highly resolved Ranunculaceae phylogeny in which all clades, except Caltheae + Nigelleae and Delphininieae + Cimicifugeae, were supported by 100 bootstrap values (Figure 3a). The 3,611 SCOG-coalescent-based analysis yielded a phylogeny very similar to the concatenated one, with high resolution and local posterior probability support (LPP, Figure 3b). However, both methods failed to confidently resolve *Nigella*'s position within the R-type clade. The other six more rigorously filtered data sets rendered phylogenetic frameworks that were completely congruent with the two 3,611 SCOG datasets (Figure S1). We then used the topology of the 3,611 SCOG-coalescent-based phylogeny (species tree) to discuss Ranunculaceae phylogeny and to conduct further analyses as the topology constraint.

The coalescent-based phylogeny (Figure 3b) resolved the 13 Ranunculaceae tribes we sampled (the first diverged Glaucideae was not available for this study), with the result consistent with previous studies (Wang et al., 2009, 2016; Zhai et al., 2019). Of those tribes, the Hydrastideae lineage diverged first within Ranunculaceae, followed by Coptideae (LPP = 1). The remaining eleven tribes (the core Ranunculaceae clade) formed two well supported major clades: T-type and R-type. The T-type clade included Isopyreae, in which all the genera have T-type chromosomes. The R-type clade included the remaining ten tribes whose genera have R-type chromosomes. There are two sub-clades within the R-type clade. One includes Helleboreae, Callianthemeae, Ranunculeae, and Anemoneae (LPP = 1) and the other contains the remaining six tribes. In the latter subclade, Asteropyreae and Adonideae are sister groups (LPP = 0.99), Caltheae is sister to Nigelleae, and the Caltheae – Nigelleae clade is then sister to the Delphinieae -Cimicifugeae clade.

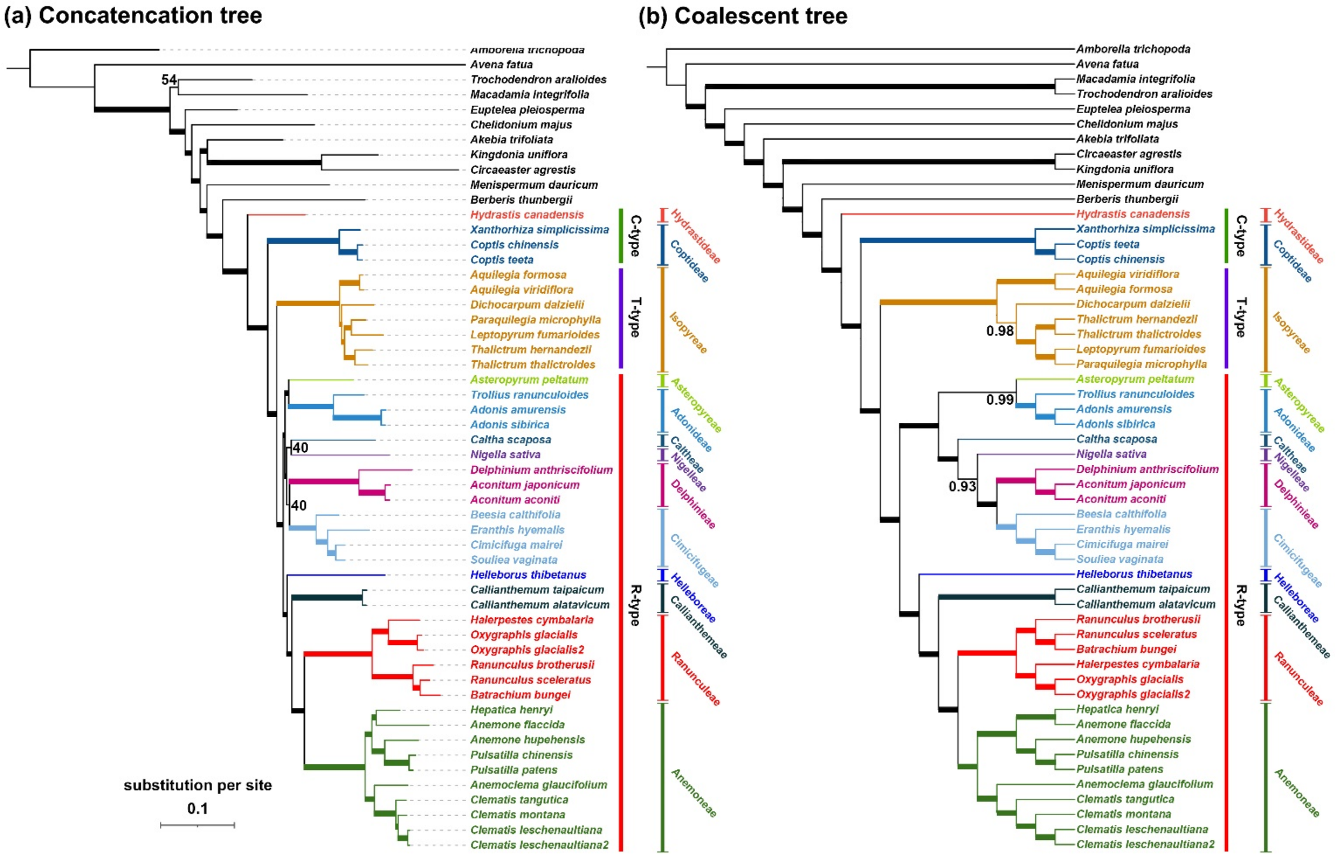

(a) Concatencation tree
(b) Coalescent tree
Amborella trichopoda
Avena fatua
Trochodendron aralioides
Macadamia integrifolia
Euptelea pleiosperma
Chelidonium majus
Akebia trifoliata
Kingdonia uniflora
Circaeaster agrestis
Menispermum dauricum
Berberis thunbergii
Hydrastis canadensis
Xanthorhiza simplicissima
Coptis chinensis
Coptis teeta
Aquilegia formosa
Aquilegia viridiflora
Dichocarpum dalzielii
Paraquilegia microphylla
Leptopyrum fumarioides
Thalictrum hernandezii
Thalictrum thalictroides
Asteropyrum peltatum
Trollius ranunculoides
Adonis amurensis
Adonis sibirica
Caltha scaposa
Nigella sativa
Delphinium anthriscifolium
Aconitum japonicum
Aconitum aconiti
Beesia calthifolia
Eranthis hyemalis
Cimicifuga mairei
Souliea vaginata
Helleborus thibetanus
Callianthemum taipaicum
Callianthemum alatavicum
Halerpestes cymbalaria
Oxygraphis glacialis
Oxygraphis glacialis2
Ranunculus brotherusii
Ranunculus sceleratus
Batrachium bungei
Hepatica henryi
Anemone flaccida
Anemone hupehensis
Pulsatilla chinensis
Pulsatilla patens
Anemoclema glaucifolium
Clematis tangutica
Clematis montana
Clematis leschenaultiana
Clematis leschenaultiana2
54
40
40
0.98
0.99
0.93
substitution per site
0.1
C-type
T-type
R-type
Hydrastideae
Coptideae
Isopyreae
Asteropyreae
Adonideae
Caltheae
Nigelleae
Delphinieae
Cimicifugeae
Helleboreae
Callianthemeae
Ranunculeae
Anemoneae

Figure 3 Phylogenetic relationships of Ranunculaceae. (a) The phylogeny inferred by the concatenated 3,611 single-copy co-orthologous nuclear gene (SCOG) data was produced using the maximum likelihood (ML) method. Bootstrap percentages are indicated on the branches and bold branches showed ML bootstrap values of 100. (b) The coalescence-based species tree topology was inferred by ASTRAL (Zhang, et al., 2018) from 3,611 nuclear genes. Numbers at branches are local posterior probabilities (ASTRAL-pp), but bold branches mark local posterior probabilities equal to 1.00. C-, T-, and R-types are chromosome groups and well-supported clades. The species in black are outgroups

3.3 Divergence time estimation for the Ranunculaceae

Dating results showed that the Ranunculaceae and Berberidaceae diverged in the Cretaceous, about 90.1 Mya with 95% highest posterior density interval (HPD) of 81.2–97.8 Mya (Figure S2). The crown age of the Ranunculaceae was estimated to be 80.2 Mya (95% HPD: 70.8–88.8 Mya) (Figure 4, Figure S2) and then the Coptideae clade diverged at about 71.7 Mya (95% HPD: 62.6–80.7 Mya). The T- and R-type clades diverged close to the Cretaceous-Paleogene (K-Pg) boundary at 66.7 Mya (95% HPD: 58.1–75.9 Mya) and the ten R-type tribes diversified in the early to middle Paleogene, from about 63 to 50 Mya.

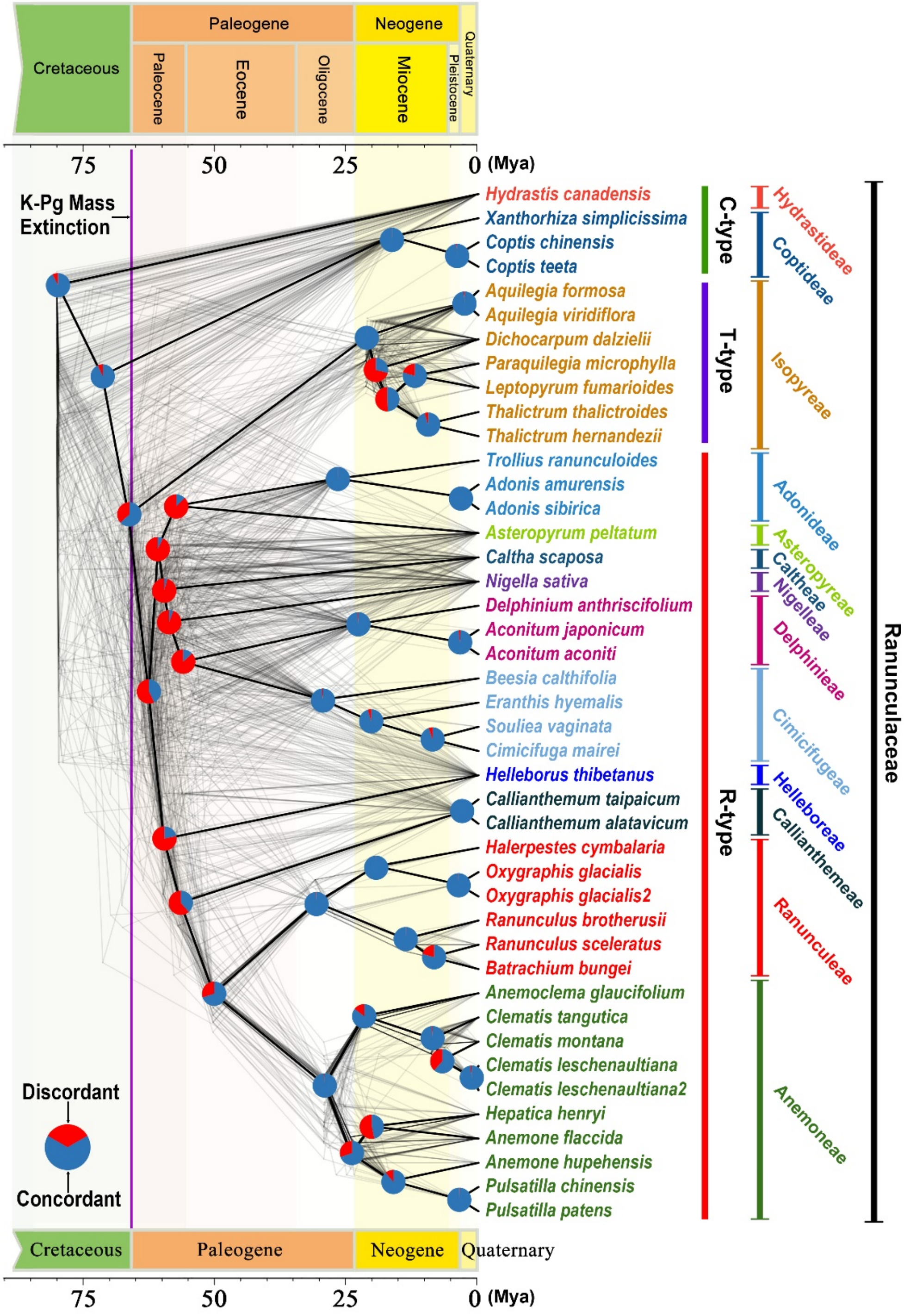

Cretaceous
Paleogene
Neogene
Quaternary
Paleocene
Eocene
Oligocene
Miocene
Pleistocene
75
50
25
0 (Mya)
K-Pg Mass Extinction
Hydrastis canadensis
Xanthorhiza simplicissima
Coptis chinensis
Coptis teeta
Aquilegia formosa
Aquilegia viridiflora
Dichocarpum dalzielii
Paraquilegia microphylla
Leptopyrum fumarioides
Thalictrum thalictroides
Thalictrum hernandezii
Trollius ranunculoides
Adonis amurensis
Adonis sibirica
Asteropyrum peltatum
Caltha scaposa
Nigella sativa
Delphinium anthriscifolium
Aconitum japonicum
Aconitum aconiti
Beesia calthifolia
Eranthis hyemalis
Souliea vaginata
Cimicifuga mairei
Helleborus thibetanus
Callianthemum taipaicum
Callianthemum alatavicum
Halerpestes cymbalaria
Oxygraphis glacialis
Oxygraphis glacialis2
Ranunculus brotherusii
Ranunculus sceleratus
Batrachium bungei
Anemoclema glaucifolium
Clematis tangutica
Clematis montana
Clematis leschenaultiana
Clematis leschenaultiana2
Hepatica henryi
Anemone flaccida
Anemone hupehensis
Pulsatilla chinensis
Pulsatilla patens
C-type
T-type
R-type
Hydrastideae
Coptideae
Isopyreae
Adonideae
Asteropyreae
Caltheae
Nigelleae
Delphinieae
Cimicifugeae
Helleboreae
Callianthemeae
Ranunculeae
Anemoneae
Ranunculaceae
Discordant
Concordant
Cretaceous
Paleogene
Neogene
Quaternary
75
50
25
0 (Mya)

Figure 4 Coalescence-based species tree (Figure 3b) topology showing discordance among genes. The ASTRAL (Zhang et al., 2018) Ranunculaceae tree has all nodes resolved in most of the species trees (in heavy black lines). The gray-colored trees were sampled from 94 single-copy co-orthologous nuclear genes (from the 3,611 gene data set in which all the taxa were present in a single gene tree) and constructed with RAxML v.8.2.12. Pie charts on the nodes show proportions of concordant and discordant tree topologies compared to the species tree. C-, T-, and R-types are chromosome groups and well-supported clade

3.4 Conflict analyses

The conflicts between nuclear gene trees, as assessed by Phyparts v.0.0.1 showed a high proportion of concordant gene trees with the species tree, thus confirming the monophyly of the Ranunculaceae tribes that contain more than one genus (Figure 4). For example, Isopyreae (T-type clade) was supported by 97% (1,969/2,025) concordant, informative nuclear gene trees; Adonideae was supported by 1,609 of 1,705 (97%) informative gene trees; Ranunculeae was supported by 1,973 of 1,994 (99%) informative gene trees; and Anemoneae was supported by 2,037 of 2,068 (99%) informative gene trees. However, tribal relationships (nodes in Paleocene and early Eocene) were characterized by high levels of gene tree discordance (Figure 4).

We also compared the Ranunculaceae nuclear and plastid phylogenies (Figure 5) and found that the plastid phylogeny, having low support values at several nodes, was not better resolved than the nuclear phylogeny. Otherwise, the two phylogenies were largely consistent. The only supported cyto-nuclear discordance is the position of Adonineae (R-type). In the nuclear phylogeny, Adonieae is well-resolved in the R-type clade, but in the plastid phylogeny it is sister to Isopyreae in the T-type clade.

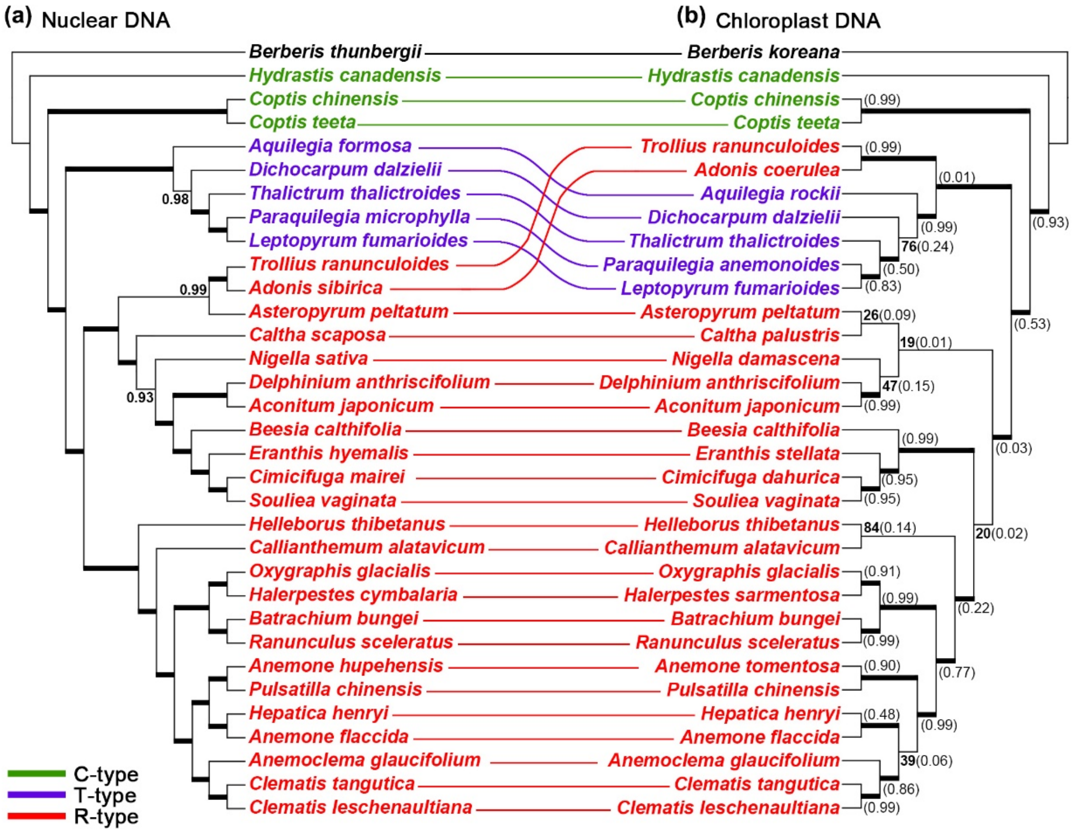

(a) Nuclear DNA
(b) Chloroplast DNA
Berberis thunbergii
Hydrastis canadensis
Coptis chinensis
Coptis teeta
Aquilegia formosa
Dichocarpum dalzielii
Thalictrum thalictroides
Paraquilegia microphylla
Leptopyrum fumarioides
Trollius ranunculoides
Adonis sibirica
Asteropyrum peltatum
Caltha scaposa
Nigella sativa
Delphinium anthriscifolium
Aconitum japonicum
Beesia calthifolia
Eranthis hyemalis
Cimicifuga mairei
Souliea vaginata
Helleborus thibetanus
Callianthemum alatavicum
Oxygraphis glacialis
Halerpestes cymbalaria
Batrachium bungei
Ranunculus sceleratus
Anemone hupehensis
Pulsatilla chinensis
Hepatica henryi
Anemone flaccida
Anemoclema glaucifolium
Clematis tangutica
Clematis leschenaultiana
0.98
0.99
0.93
Berberis koreana
Hydrastis canadensis
Coptis chinensis
Coptis teeta
Trollius ranunculoides
Adonis coerulea
Aquilegia rockii
Dichocarpum dalzielii
Thalictrum thalictroides
Paraquilegia anemonoides
Leptopyrum fumarioides
Asteropyrum peltatum
Caltha palustris
Nigella damascena
Delphinium anthriscifolium
Aconitum japonicum
Beesia calthifolia
Eranthis stellata
Cimicifuga dahurica
Souliea vaginata
Helleborus thibetanus
Callianthemum alatavicum
Oxygraphis glacialis
Halerpestes sarmentosa
Batrachium bungei
Ranunculus sceleratus
Anemone tomentosa
Pulsatilla chinensis
Hepatica henryi
Anemone flaccida
Anemoclema glaucifolium
Clematis tangutica
Clematis leschenaultiana
(0.99)
(0.99)
(0.01)
(0.93)
(0.99)
76(0.24)
(0.50)
(0.83)
26(0.09)
(0.53)
19(0.01)
47(0.15)
(0.99)
(0.99)
(0.03)
(0.95)
(0.95)
84(0.14)
20(0.02)
(0.91)
(0.99)
(0.22)
(0.99)
(0.90)
(0.77)
(0.48)
(0.99)
39(0.06)
(0.86)
(0.99)
C-type
T-type
R-type

Figure 5 Comparison of the Ranunculaceae nuclear and plastid phylogenies. Left: Coalescence-based species tree topology inferred by ASTRAL (Zhang et al., 2018) from 3,611 nuclear genes (see Figure 3b). Numbers at branches are local posterior probabilities (ASTRAL-pp) but local posterior probabilities equal to 1.00 are marked with bold branches. Right: The Ranunculaceae phylogeny inferred from the concatenated coding regions of the complete plastome sequences using the maximum likelihood (ML) method. Bootstrap percentages are indicated on the branches and bold branch indicates ML bootstrap values of 100%. Numbers in brackets show the contribution of incomplete lineage sorting to the conflicts between the nuclear and plastid gene trees based on the multispecies coalescent model implemented in Phybase v.1.5 (Liu and Yu, 2010). Plastome sequence data are from previous studies (see Table S1)

3.5 Coalescent simulations and phylogenetic network estimating

To explain the cyto-nuclear discordance, we compared conflicts among the Phybase-simulated gene trees with the plastid genome phylogeny. The results showed that because of a very low proportion (0.01) of concordant simulated trees, the well-supported contradiction of the Adonineae between the nuclear and plastid trees could not be explained by ILS (Figure 5).

Nuclear gene discordance analysis (Figure 6a) showed much overlap between the empirical and simulated distance distributions, suggesting that ILS alone can explain most of the observed gene tree heterogeneity (Maureira-Butler et al., 2008). Meanwhile, during SNaQ analyses in PhyloNetworks, a plot of pseudo-loglikelihood scores (Figure 6c) suggested that the best network model was one with a maximum of one hybridization event (-ploglik = 246.7). We detected the most likely reticulation event to be between the *Caltha* and the ((*Callianthemum*, (*Ranunculus*, *Anemone*)), *Helleborus*) clades ($\gamma = 0.368$) with bootstrap value (BS) of 44.1 (Figure 6b). Also, when more than one hybridization event was allowed, we detected two additional reticulations in the Ranunculaceae species trees (Figure 6c), but the γ values of those two were minor (0.006–0.09).

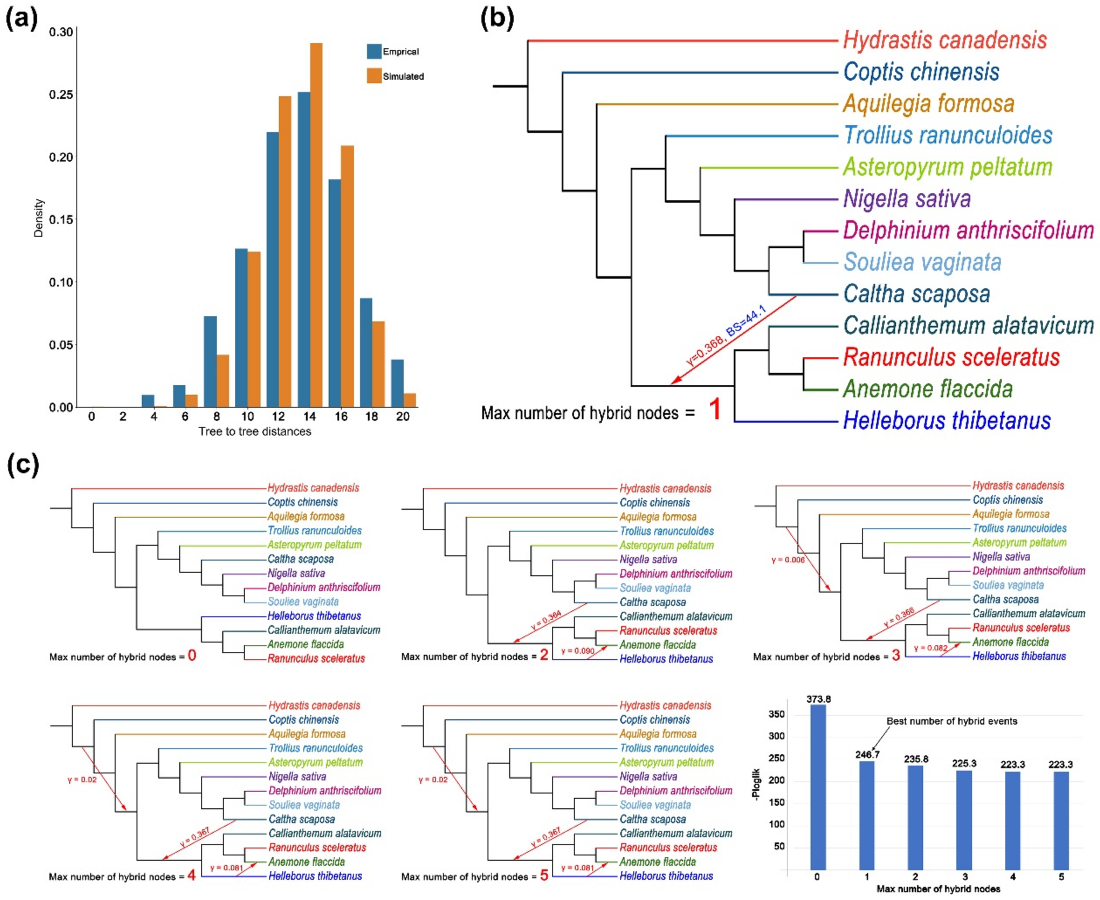


Figure 6 Testing the contributions of incomplete lineage sorting (ILS) and hybridization to the conflicts between the Ranunculaceae gene trees and the species tree inferred by ASTRAL (Zhang et al., 2018). (a) Coalescent simulations of tree-to-tree distance distributions between the empirical gene trees and the ASTRAL species tree (Empirical) and those from the coalescent simulation (Simulated). (b) Introgression model of unstable Ranunculaceae clades estimated using Phylonetworks. The results display maximum pseudolikelihood trees with a maximum zero to five allowed reticulations, and (c) the best model had a maximum allowed hybrid node of one according to the pseudo-loglikelihood scores (-ploglik) shown in the bar chart in (b). Red arrows mark possible reticulation events, γ is the gene contribution, and BS is the bootstrap support for the placement of minor hybrid branches

3.6 WGD detection

The results of Phyparts analysis showed a markedly increased number of gene duplications in several clades (Figure 7a), including the Ranunculales clade (R clade, 92 gene trees with duplication), the Ranunculaceae clade without Hydrastideae (RH clade, 86 gene trees with duplication), and the Delphinieae clade (DEL clade, 86 gene trees with duplication). We also found several clades with many gene trees with duplication (Figure 7a): the clade with all the Ranunculales families except Eupteleaceae (RE clade, 72 gene trees with duplication), the core Ranunculaceae clade (CR clade, 73 gene trees with duplication), and Anemoneae (ANE clade, 64 gene trees with duplication).

Our absolute dating analysis (Figure 7b) found an obvious peak at around 80 Mya for each RH clade sample except *Hydrastis canadensis*, thus suggesting an ancient WGD event for this clade. We detected another pronounced peak at about 60 Mya for the DEL clade samples, suggesting an ancient WGD event for Delphinieae. We also detected a peak at about 110 Mya in some Ranunculaceae samples (e.g., *Hydrastis*, *Aquilegia*, *Callianthemum*, and *Helleborus*) and in the outgroups, and that may represent an ancient WGD event for Ranunculales. However, a few Ranunculaceae samples (e.g., *Aconitum* and *Caltha*), had no such peak. It is most likely because that later duplication may have relieved the selection pressure on gene copies and caused previously maintained copies to disappear (Ren et al., 2018). Those three WGD events corresponded with the three clades possessing the highest number of gene trees with duplication, as inferred by Phyparts analysis.

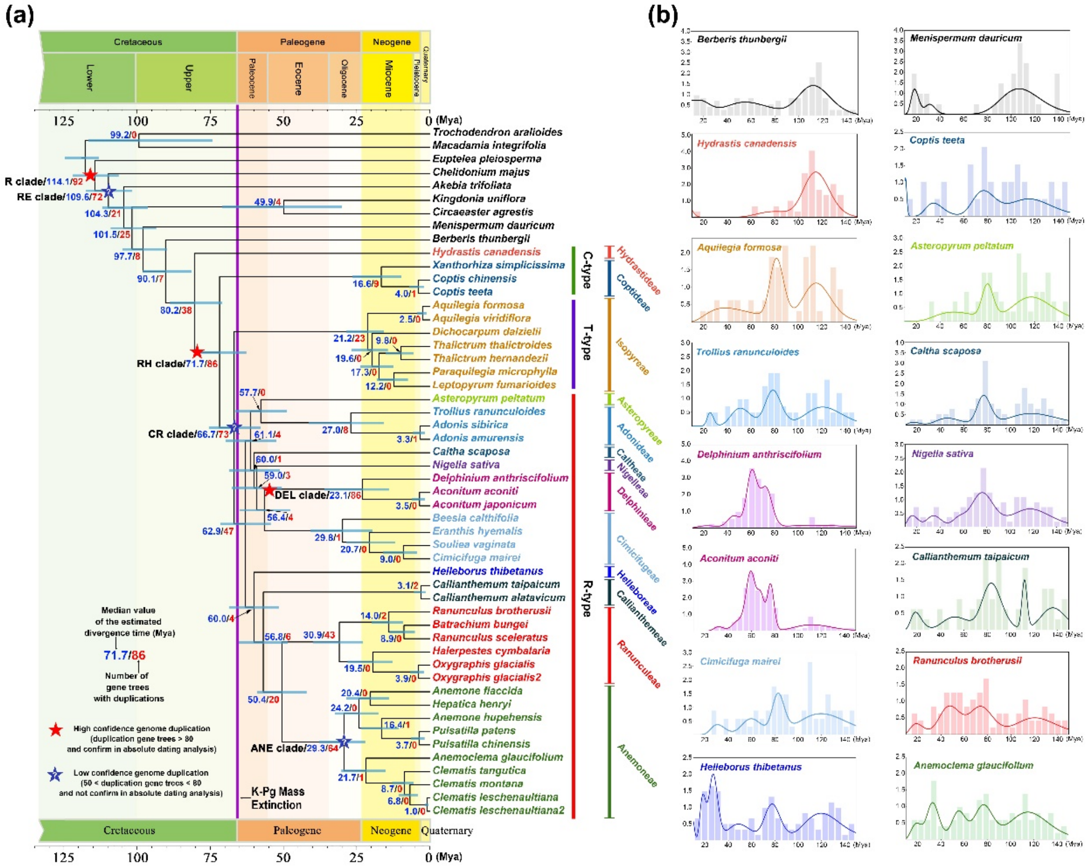


Figure 7 Genome duplication analysis under the absolute dating context. (a) Chronogram showing estimated divergence times according to MCMCTree and based on the Ranunculaceae coalescent species tree using 3,611 single-copy orthologous genes. The lineages have different colors and the taxa in a given lineage are grouped in one color. The outgroups are in black. Divergence times and numbers of gene trees with duplications (estimated by Phyparts) are marked under (or above) the branches. Starred clades are R: all Ranunculales, RE: R without Eupteleaceae, RH: R without Hydrastideae, CR: core Ranunculaceae, DEL: CR without Delphinieae, and ANE: CR without Anemoneae. (b) Absolute age distributions of genome duplications among major Ranunculales clades

4 DISCUSSION

4.1 On using silica gel-dried leaf tissues in phylotranscriptomic analyses

For transcriptome partition, previous studies suggested that the contents and concentrations of total RNA, RIN, and rRNA ratio are the most important indices for assessing the success of RNA extraction (Jordon-Thaden et al., 2015; Yang et al., 2017). For example, an extraction was considered successful if the total RNA content was ≥ 30 μg, the RIN was higher than 5, and the rRNA ratio (26S/18S) was greater than 1 (Jordon-Thaden et al., 2015). Yang et al. (2017) considered an RIN higher than 6 and a concentration higher than 20 ng/μL acceptable. However, those indices may not be comparable among different studies because the acceptable RNA quality and quantity are greatly dependent on the next-generation sequencing platform used (Johnson et al., 2012). Nevertheless, our RNA extraction results showed that all the RIN values from the SD-samples were higher than six (average RIN = 7.5, Table S2). Although, most of those SD-samples did not yield ≥ 30 μg total RNA, the concentration of the extracted RNA was higher than 131.3 ng/μL (Table S2). Also, we found no significant differences in those RNA indices between the SD and LN treatments (Table S2).

Our results also showed that RNA does not decay as easily as previously thought, at least not in our Ranunculales samples. All 21 SD-samples yielded usable total RNA and among them, the SD-sample of *Clematis leschenaultiana* was stored in the -20 °C freezer for more than two years (Table 1, Table S2). Although SD-samples generated relatively lower quality transcriptome reads than the LN- and T-samples, and may not be useful for evo-devo studies, they are of sufficient quality for phylotranscriptomic analysis. Indeed, the assembled SCOGs from our SD-samples produced good or even better results than did those from the LN-samples and the T-samples used by the One Thousand Plant Transcriptomes Initiative (2019). If silica gel dried sampled can be used for RNA-seq in most angiosperm taxa, the utility of transcriptome data could become much more feasible for phylogenetic analyses andmay greatly improve the efficiency of data generation and reduce the cost. This SD plant tissue preservation strategy tested here provides novel insight into the feasibility of methods to easily obtain RNA datafor researchers in plant phylogenomics and may stimulate the exploration into broader groups of plants using SD preservation for RNA-seq phylogenetics.

4.2 Phylogenetic framework of the Ranunculaceae

Using Sanger sequencing data, molecular phylogenetic studies have not satisfactorily resolved the relationships of the 11 tribes that comprise the core Ranunculaceae (Johansson and Jansen, 1993; Hoot, 1995; Johansson, 1995; Ro et al., 1997; Hoot et al., 2008; Wang et al., 2009, 2016). Two chromosome types (R-type and T-type) are recognized in the core Rasnunculaceae (Tamura, 1995; Wang et al., 2009, 2016; Zhai et al., 2019).

Based on our 3,611 nuclear SCOGs (as well as the other rigorously selected data sets), we showed that the T- and R-type clades are robustly resolved as sister to each other within the core Ranunculaceae. According to our nuclear phylogeny, the T- and R-type chromosome groups seemed to have diverged from the C-type chromosome group. Our results are consistent with those of other published Ranunculaceae phylogenetic studies (Hoot, 1995; Hoot et al., 2008), and specifically support the Ranunculaceae phylogenetic classification proposed by Wang et al. (2009).

However, complete plastid genome data resolved different relationships within the core Ranunculaceae clade (Zhai et al., 2019; He et al., 2019). The sister relationship of Adonideae (R-type chromosome group) and Isopyreae (T-type chromosome group) were resolved and supported by the plastome sequences, while the other nine R-type tribes formed another clade and was sister to the Adonideae-Isopyreae clade (Zhai et al., 2019, Figure 5). This result indicated that T-type chromosomes may have derived from R-type chromosomes (Zhai et al., 2019). Strongly supported cyto-nuclear discordance may be the result of ILS or hybridization (Roos et al., 2011), but our coalescent simulating analysis suggests that ILS can not explain the cyto-nuclear discordance in Ranunculaceae. The different positions of Adonideae in the nuclear versus the plastid phylogenies may have been caused by ancient hybridization and/or subsequent introgression events, which are known to have happened in many angiosperm lineages (Liu et al., 2020; Hodel et al., 2021; Wang et al., 2021).

Within the R-type clade, the position of Cimicifugeae created another cyto-nuclear discordance. In our species tree (Figure 3b), Cimicifugeae was clustered with strong support as the sister group of Delphinieae. However, in the plastid phylogeny (Figure 5 and Zhai et al., 2019) Cimicifugeae was sister to a large R-type sub-clade including Helleboreae, Callianthemeae, Ranunculeae, and Anemoneae, but it had low support values. This discordance was likely caused by insufficient information in the plastid genome data.

Using complete plastid genome data, Zhai et al. (2019) estimated two waves of radiation in the Ranunculaceae after the Cretaceous-Paleogene (K-P) boundary. Our molecular dating analysis yielded similar divergence time estimates and consistently revealed two waves of radiation after the K-P boundary (Figure 4). Almost all the ten tribes of the R-type clade evolved in the first wave of radiation during the Paleocene and early Eocene eras. Furthermore, we detected widespread nuclear gene discordance during that diversification, thus showing complicated evolutionary processes at work during that period. Our coalescent simulating results showed that the conflict among nuclear trees was mainly due to ancestral genetic polymorphism and ILS, and supported one hybridization event between Caltheae and the (Callianthemeae-[Ranunculeae-Anemoneae]-Helleboreae) clades (Figure 6). Our results have hence provided important insights into why previous phylogenetic studies, based on either limited numbers of genes or uniparentally transmitted markers, could not satisfactorily resolve the tribal relationships of Ranunculaceae.

4.3 Whole genome duplications

Our WGD analyses support three WGD events within the Ranunculales (Figure 7). The first event occurred in the Ranunculales ancestor at around 110 Mya, as previous studies had also suggested (Pabón-Mora et al., 2013; One Thousand Plant Transcriptomes Initiatives, 2019; Pei et al., 2021). After that, we detected another WGD event around 80 Mya after the divergence of Hydrastideae, but before the divergence of Coptideae and the core Ranunculaceae clade. Previous studies also reported an ancient WGD event either after the divergence of Berberidaceae and Ranunculaceae (Xie et al., 2020) or shared by the Ranunculaceae (Liu et al., 2021). However, those studies did not include samples of the Ranunculaceae tribes that diverged the earliest, the Glaucideae and Hydrastideae. So, those reported WGDs are likely the second WGD event that we detected before the divergence of the Coptideae and the core Ranunculaceae clade. A third WGD event was detected in the stem lineage of Delphinieae by Park et al. (2020) who, using plastid genome and transcriptome data, found that plastid genes *rpl*32 and *rps*16 were transferred to the nucleus in Delphinieae and duplicated in *Aconitum* samples. They assumed that this duplication was associated with an ancient WGD event in *Aconitum*. However, the WGD hypothesis based on the two duplicated genes needs to be tested with more robust data. Our results suggest that both *Aconitum* and *Delphinium* may have shared an ancient WGD event. Both the CR and R-type clades had large numbers of gene trees with gene duplication (73 for the CR clade and 47 for the R-type clade, Figure 7a), but the dating analysis did not support the WGD events in these two clades. The high gene duplication detected in those clades may have been caused by small scale or single gene duplications or other mechanisms.

4.4 Conclusions

In this study, we tested whether silica gel-dried plant tissues could be used for RNA extraction and phylotranscriptomic studies. Our results showed that the RNA extracted from silica gel-dried leaf tissue samples of members of the Ranunculaceae and the close relatives performed well in subsequent phylogenetic analyses. Based on the transcriptome data we generated, we successfully resolved long-standing phylogenetic controversies in the Ranunculaceae, such as core Ranunculaceae tribal relationships, and cyto-nuclear discordance of the T- and R-type clades. We also obtained a resolved

Ranunculaceae phylogeny that had better statistical support than in previous studies. The 11 tribes of the core Ranunculaceae grouped into two clades corresponding to the chromosome types (the T- and R-type clades), with high support. Two waves of radiation, previously inferred based on plastome data (Zhai et al., 2019), were also supported by our nuclear SCOG data. The first wave, which represented the deep diversification of all 11 tribes of the core Ranunculaceae, underwent extremely complicated evolutionary processes, such as gene duplication, hybridization, introgression, and incomplete lineage sorting. However, WGD was important in the evolution of Ranunculaceae, but not a direct factor in chromosome type evolution in the family. Our results indicated that using either a limited number of genes or plastid genome data does not satisfactorily resolve Ranunculaceae tribal relationships. Our results provide novel insights into RNA-seq methods that may greatly increase the utility of RNA sequencing in phylogenomics, as well as into the phylogenetics and evolution of the widely contested Ranunculaceae family.

ACKNOWLEDGEMENTS

This study was supported by the National Natural Science Foundation of China [grant number 31670207] and the Beijing Natural Science Foundation [grant number 5182016].

REFERENCES

Aköz G., & Nordborg M. (2019). The *Aquilegia* genome reveals a hybrid origin of core eudicots. *Genome Biology*, 20:256. https://doi.org/10.1186/s13059-019-1888-8

Alejo-Jacuinde G., Gonzalez-Morales S. I., Oropeza-Aburto A., Simpson J., & Herrera-Estrella L. (2020). Comparative transcriptome analysis suggests convergent evolution of desiccation tolerance in *Selaginella* species. *BMC Plant Biology*, 20:468. https://doi.org/10.1186/s12870-020-02638-3

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

Buchfink B., Xie C., & Huson D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12:59–60. https://doi.org/10.1038/nmeth.3176

Cai L., Xi Z., Amorim A. M., Sugumaran M., Rest J. S., Liu L., & Davis C. C. (2019). Widespread ancient whole-genome duplications in Malpighiales coincide with Eocene global climatic upheaval. *New Phytologist*, 221:565–576. https://doi.org/10.1111/nph.15357

Chabikwa T. G., Barbier F. F., Tanurdzic M., & Beveridge C. A. (2020). *De novo* transcriptome assembly and annotation for gene discovery in avocado, macadamia and mango. *Scientific Data*, 7:9. https://doi.org/10.1038/s41597-019-0350-9

Chen L. Y., Zhao S. Y., Wang Q. F., & Moody M. L. (2015). Transcriptome sequencing of three *Ranunculus* species (Ranunculaceae) reveals candidate genes in adaptation from terrestrial to aquatic habitats. *Scientific Reports*, 5:10098. https://doi.org/10.1038/srep10098

Cheon S., Zhang J., & Park C. (2020). Is phylotranscriptomics as reliable as phylogenomics? *Molecular Biology and Evolution*, 37:3672–3683. https://doi.org/10.1093/molbev/msaa181

Cossard G., Sannier J., Sauquet H., Damerval C., Ronse de Craene L., Jabbour F., & Nadot S. (2016). Subfamilial and tribal relationships of Ranunculaceae: evidence from eight molecular markers. *Plant Systematics and Evolution*, 302:419–431. https://doi.org/10.1007/s00606-015-1270-6

Cronquist A. (1981). *An integrated system of classification of flowering plants*. Columbia University Press.

Damerval C., & Becker A. (2017). Genetics of flower development in Ranunculales—a new, basal eudicot model order for studying flower evolution. *New Phytologist*, 216:361–366. https://doi.org/10.1111/nph.14401

Dodsworth S. (2015). Genome skimming for next-generation biodiversity analysis. *Trends in Plant Science*, 20:525–527. https://doi.org/10.1016/j.tplants.2015.06.012

Dodsworth S., Pokorny L., Johnson M. G., Kim J. T., Maurin O., Wickett N. J., Forest F., & Baker W. J. (2019). Hyb-Seq for flowering plant systematics. *Trends in Plant Science*, 24:887–891. https://doi.org/10.1016/j.tplants.2019.07.011

Dong Y. B., Chen S. C., Cheng S. F., Zhou W. B., Ma Q., Chen Z. D., Fu C. X., Liu X., Zhao Y. P., Soltis P. S., Wong G. K. S., Soltis D. E., & Xiang Q. Y. (2019). Natural selection and repeated patterns of molecular evolution following allopatric divergence. *eLife*, 8:e45199. https://doi.org/10.7554/eLife.45199

Eaton D. A. R. (2020). Toytree: A minimalist tree visualization and manipulation library for Python. *Methods in Ecology and Evolution*, 11:187–191. https://doi.org/10.1111/2041-210X.13313

Edwards S. V. (2009). Is a new and general theory of molecular systematics emerging? *Evolution*, 63(1):1–19. https://doi.org/10.1111/j.1558-5646.2008.00549.x

Edwards S. V., Xi Z., Janke A., Faircloth B. C., McCormack J. E., Glenn T. C., Zhong B., Wu S., Lemmon E. M., Lemmon A. R., & Leaché A. D. (2016). Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. *Molecular Phylogenetics and Evolution*, 94:447–462. https://doi.org/10.1016/j.ympev.2015.10.027

Emadzade K., Lehnebach C., Lockhart P., & Hörandl E. (2010). A molecular phylogeny, morphology and classification of genera of Ranunculeae (Ranunculaceae). *Taxon*, 59:809–828. https://doi.org/10.1002/tax.593011

Feng S., Ru D., Sun Y., Mao K., Milne R., & Liu J. (2019). Trans-lineage polymorphism and nonbifurcating diversification of the genus *Picea*. *New Phytologist*, 222:576–587. https://doi.org/10.1111/nph.15590

Filiault D. L., Ballerini E. S., Mandáková T., Aköz G., Derieg N. J., Schmutz J., Jenkins J., Grimwood J., Shu S., Hayes R. D., Hellsten U., Barry K., Yan J., Mihaltcheva S., Karafiátová M., Nizhynska V., Kramer E. M., Lysak M. A., Hodges S. A., & Nordborg M. (2018). The *Aquilegia* genome provides insight into adaptive radiation and reveals an extraordinarily polymorphic chromosome with a unique history. *eLife*, 7:e36426. https://doi.org/10.7554/eLife.36426

Fu L., Niu B., Zhu Z., Wu S., & Li W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28:3150–3152. https://doi.org/10.1093/bioinformatics/bts565

Galimba K. D., Tolkin T. R., Sullivan A. M., Melzer R., Theißen G., & Di Stilio V. S. (2012). Loss of deeply conserved C-class floral homeotic gene function and C- and E-class protein interaction in a double-flowered ranunculid mutant. *Proceedings of the National Academy of Sciences*, 109:E2267–E2275. https://doi.org/10.1073/pnas.1203686109

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

Gregory W. C. (1941). Phylogenetic and cytological studies in the Ranunculaceae Juss. *Transactions of the American Philosophical Society*, 31:443–521. https://doi.org/10.2307/1005611

Greiner S., & Sobanski J., Bock R. (2015). Why are most organelle genomes transmitted maternally? *Bioessays*, 37(1):80–94. https://doi.org/10.1002/bies.201400110

Hagel J. M., Morris J. S., Lee E. J., Desgagné-Penix I., Bross C. D., Chang L., Chen X., Farrow S. C., Zhang Y., Soh J., Sensen C. W., & Facchini P. J. (2015). Transcriptome analysis of 20 taxonomically related benzylisoquinoline alkaloid-producing plants. *BMC Plant Biology*, 15:227. https://doi.org/10.1186/s12870-015-0596-0

He J., Yao M., Lyu R. D., Lin L. L., Liu H. J., Pei L. Y., Yan S. X., Xie L., & Cheng J. (2019). Structural variation of the complete chloroplast genome and plastid phylogenomics of the genus *Asteropyrum* (Ranunculaceae). *Scientific Reports*, 9: 15285. https://doi.org/10.1038/s41598-019-51601-2

He S. M., Liang Y. L., Cong K., Chen G., Zhao X., Zhao Q. M., Zhang J. J., Wang X., Dong Y., Yang J. L., Zhang G. H., Qian Z. L., Fan W., & Yang S. C. (2018). Identification and characterization of genes involved in benzylisoquinoline alkaloid biosynthesis in *Coptis* species. *Frontiers in Plant Science*, 9:731. https://doi.org/10.3389/fpls.2018.00731

Hodel R. G. J., Zimmer, E., Wen, J. (2021). A phylogenomic approach resolves the backbone of Prunus (Rosaceae) and identifies signals of hybridization and allopolyploidy. *Molecular Phylogenetics and Evolution* 160: 107118.

Hoot S. B. (1995). Phylogeny of the Ranunculaceae based on preliminary *atp*B, *rbc*L and 18S nuclear ribosomal DNA sequence data. In U. Jensen, J. Kadereit (Eds.), *Systematics and Evolution of the Ranunculiflorae* (pp. 241–251). Springer.

Hoot S. B., Kramer J., & Arroyo M. T. K. (2008). Phylogenetic position of the South American dioecious genus *Hamadryas* and related Ranunculeae (Ranunculaceae). *International Journal of Plant Sciences*, 169:3. https://doi.org/10.1086/526460

Huerta-Cepas J., Serra F., & Bork P. (2016). ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Molecular Biology and Evolution*, 33:1635–1638. https://doi.org/10.1093/molbev/msw046

Hutchinson J. (1923). Contributions towards a phylogenetic classification of flowering plants. *Bulletin of Miscellaneous Information (Royal Botanic Gardens, Kew)*, 1923:65–89. https://doi.org/10.2307/4118622

Janchen E. (1949). *Die Systematische Gliederung der Ranunculaceen und Berberidaceen*. Springer.

Johansson J. T. (1995). A revised chloroplast DNA phylogeny of the Ranunculaceae. In U. Jensen, J. Kadereit (Eds.), *Systematics and Evolution of the Ranunculiflorae* (pp. 253–261). Springer.

Johansson J. T., & Jansen R. K. (1993). Chloroplast DNA variation and phylogeny of the Ranunculaceae. *Plant Systematics and Evolution*, 187:29–49. https://doi.org/10.1007/BF00994090

Johnson M. T., Carpenter E. J., Tian Z., Bruskiewich R., Burris J. N., Carrigan C. T., Chase M. W., Clarke N. D., Covshoff S., dePamphilis C. W., Edger P. P., Goh F., Graham S., Greiner S., Hibberd J. M., Jordon-Thaden I., Kutchan T. M., Leebens-Mack J., Melkonian M., Miles N., Myburg H., Patterson J., Pires J. C., Ralph P., Rolf M., Sage R. F., Soltis D. E., Soltis P. S., Stevenson D., Stewart Jr. C. N., Surek B., Thomsen C. J. M., Villarreal J. C., Wu X., Zhang Y., Deyholos M. K., & Wong G. K. S. (2012). Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. *PLoS*

ONE, 7:e50226. https://doi.org/10.1371/journal.pone.0050226

Jordon-Thaden I. E., Chanderbali A. S., Gitzendanner M. A., & Soltis D. E. (2015). Modified CTAB and TRIzol protocols improve RNA extraction from chemically complex Embryophyta. *Applications in Plant Sciences*, 3:1400105. https://doi.org/10.3732/apps.1400105

Kadereit J. W., Lauterbach M., Kandziora M., Spillmann J., & Nyffeler R. (2019). Dual colonization of European high-altitude areas from Asia by *Callianthemum* (Ranunculaceae). *Plant Systematics and Evolution*, 305:431–443. https://doi.org/10.1007/s00606-019-01583-5

Katoh K., & Standley D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30:772–780. https://doi.org/10.1093/molbev/mst010

Knobloch E., & Mai D. H. (1986). *Monographie der Früchte und Samen in der Kreide von Mitteleuropa*. Alexander Doweld.

Kramer E. M. (2009). *Aquilegia*: A new model for plant development, ecology, and evolution. *Annual Review of Plant Biology*, 60:261–277. https://doi.org/10.1146/annurev.arplant.043008.092051

Kramer E. M., Di Stilio V. S., & Schlüter P. M. (2003). Complex patterns of gene duplication in the *APETALA3* and *PISTILLATA* lineages of the Ranunculaceae. *International Journal of Plant Sciences*, 164:1–11. https://doi.org/10.1086/344694

Kramer E. M., & Hodges S. A. (2010). *Aquilegia* as a model system for the evolution and ecology of petals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365:477–490. https://doi.org/10.1098/rstb.2009.0230

Landis J. B., Soltis D. E., & Soltis P. S. (2017). Comparative transcriptomic analysis of the evolution and development of flower size in *Saltugilia* (Polemoniaceae). *BMC Genomics*, 18:475. https://doi.org/10.1186/s12864-017-3868-2

Langlet O. (1927). Beitra¨ge zur Zytologie der Ranunculaceen. *Svensk botanisk tidskrift*, 21:1–17.

Langlet O. (1932). Über chromosomenverhaltnisse und systematik der Ranunculaceae. *Svensk botanisk tidskrift*, 26:381–400.

Lechner M., Findeiß S., Steiner L., Marz M., Stadler P. F., & Prohaska S. J. (2011). Proteinortho: detection of (co-) orthologs in large-scale analysis. BMC *Bioinformatics*, 12:1–9. https://doi.org/10.1186/1471-2105-12-124

Liu B. B., Campbell C. S., Hong D. Y., & Wen J. (2020). Phylogenetic relationships and chloroplast capture in the *Amelanchier-Malacomeles-Peraphyllum* clade (Maleae, Rosaceae): evidence from chloroplast genome and nuclear ribosomal DNA data using genome skimming. *Molecular Phylogenetics and Evolution*, 147:106784. https://doi.org/10.1016/j.ympev.2020.106784

Liu B. B., Ma Z. Y., Ren C., Hodel R. G. J., Sun M., Liu X. Q., Liu G. N., Hong D. Y., Zimmer E. A., & Wen J. (2021). Capturing single-copy nuclear genes, organellar genomes, and nuclear ribosomal DNA from deep genome skimming data for plant phylogenetics: A case study in Vitaceae. *BioRxiv*, 2021:02.25.432805. https://doi.org/10.1101/2021.02.25.432805

Liu H., He J., Ding C., Lyu R., Pei L., Cheng J., & Xie L. (2018). Comparative analysis of complete chloroplast genomes of *Anemoclema*, *Anemone*, *Pulsatilla*, and *Hepatica* revealing structural variations among genera in tribe Anemoneae (Ranunculaceae). *Frontiers in Plant Science*, 9:1097. https://doi.org/10.3389/fpls.2018.01097

Liu L., Yu L. (2010). Phybase: an R package for species tree analysis. *Bioinformatics*, 26:962–963. https://doi.org/10.1093/bioinformatics/btq062

Liu Y., Wang B., Shu S., Li Z., Song C., Liu D., Niu Y., Liu J., Zhang J., Liu H., Hu Z., Huang B., Liu X., Liu W., Jiang L., Alami M. M., Zhou Y., Ma Y., He X, Yang Y., Zhang T., Hu H., Barker M. S., Chen S., Wang X., & Nie J. (2021). Analysis of the *Coptis chinensis* genome reveals the diversification of protoberberine-type alkaloids. *Nature Communications*, 12:1–13. https://doi.org/10.1038/s41467-021-23611-0

Magallón S., Gómez-Acevedo S., Sánchez-Reyes L. L., & Hernández-Hernández T. (2015). A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist*, 207:437–453. https://doi.org/10.1111/nph.13264

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

Marburger S., Alexandrou M. A., Taggart J. B., Creer S., Carvalho G., Oliveira C., & Taylor M. I. (2018). Whole genome duplication and transposable element proliferation drive genome expansion in Corydoradinae catfishes. *Proceedings of the Royal Society B: Biological Sciences*, 285:20172732. https://doi.org/10.1098/rspb.2017.2732

Matasci N., Hung L. H., Yan Z., Carpenter E. J., Wickett N. J., Mirarab S., Nguyen N., Warnow T., Ayyampalayam S., Barker M., Burleigh J. G., Gitzendanner M. A., Wafula E., Der J. P., dePamphilis C. W., Roure B., Philippe H., Ruhfel B. R., Miles N. W., Graham S. W., Mathews S., Surek B., Melkonian M., Soltis D. E., Soltis P. S., Rothfels C., Pokorny L., Shaw J. A., DeGironimo L., Stevenson D. W., Villarreal J. C., Chen T., Kutchan T. M., Rolf M., Baucom R. S., Deyholos M. K., Samudrala R., Tian Z., Wu X., Sun X., Zhang Y., Wang J., Leebens-Mack J., & Wong G. K. S. (2014). Data access for the 1,000 Plants (1KP) project. *GigaScience*, 3:2047–217X. https://doi.org/10.1186/2047-217X-3-17

Maureira-Butler I. J., Pfeil B. E., Muangprom A., Osborn T. C., & Doyle J. J. (2008). The reticulate history of *Medicago* (Fabaceae). *Systematic Biology*, 57:466–482. https://doi.org/10.1080/10635150802172168

McKain M. R., Johnson M. G., Uribe-Convers S., Eaton D., & Yang, Y. (2018). Practical considerations for plant phylogenomics. *Applications in Plant Sciences*, 6:e1038. https://doi.org/10.1002/aps3.1038

Meesapyodsuk D., Ye S., Chen Y., Chen Y., Chapman R. G., & Qiu X. (2018). An engineered oilseed crop produces oil enriched in two very long chain polyunsaturated fatty acids with potential health-promoting properties. *Metabolic Engineering*, 49:192–200. https://doi.org/10.1016/j.ymben.2018.08.009

Miller M. A., Pfeiffer W., & Schwartz T. (2012). The CIPRES science gateway:

enabling high-impact science for phylogenetics researchers with limited resources. *InProceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the extreme to the campus and beyond*, 2012:1–8. https://doi.org/10.1145/2335755.2335836

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

Morales-Briones D. F., Arias T., Di Stilio V. S., & Tank D. C. (2019). Chloroplast primers for clade-wide phylogenetic studies of *Thalictrum*. *Applications in Plant Sciences*, 7:e11294. https://doi.org/10.1002/aps3.11294

Morales-Briones D. F., Kadereit G., Tefarikis D. T., Moore M. J., Smith S. A., Brockington S. F., Timoneda A., Yim W. C., Cushman J. C., & Yang Y. (2021). Disentangling sources of gene tree discordance in phylogenomic data sets: Testing ancient hybridizations in Amaranthaceae s.l. *Systematic Biology*, 70:219–235. https://doi.org/10.1093/sysbio/syaa066

Narzary D., Verma S., Mahar K. S., & Rana T. S. (2015). A rapid and effective method for isolation of genomic DNA from small amount of silica-dried leaf tissues. *National Academy Science Letters*, 38:441–444. https://doi.org/10.1007/s40009-015-0357-5

One Thousand Plant Transcriptomes Initiative. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature*, 574:679–685. https://doi.org/10.1038/s41586-019-1693-2

Pabón-Mora N., Hidalgo O., Gleissberg S., & Litt A. (2013). Assessing duplication and loss of APETALA1/FRUITFULL homologs in Ranunculales. *Frontiers in Plant Science*, 4:358. https://doi.org/10.3389/fpls.2013.00358

Park S., An B., & Park S. (2020). Recurrent gene duplication in the angiosperm tribe Delphinieae (Ranunculaceae) inferred from intracellular gene transfer events and heteroplasmic mutations in the plastid *mat*K gene. *Scientific Reports*, 10:1–11. https://doi.org/10.1038/s41598-020-59547-6

Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., & Vanderplas J. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12:2825–2830.

Pei L., Wang B., Ye J., Hu X., Fu L., Li K., Ni Z., Wang Z., Wei Y., Shi L., Zhang Y., Bai X., Jiang M., Wang S., Ma C., Li S., Liu K., Li W., & Cong B. (2021). Genome and transcriptome of *Papaver somniferum* Chinese landrace CHM indicates that massive genome expansion contributes to high benzylisoquinoline alkaloid biosynthesis. *Horticulture Research*, 8:1–13. https://doi.org/10.1038/s41438-020-00435-5

Pigg K. B., & DeVore M. L. (2005). *Paleoactaea* gen. nov. (Ranunculaceae) fruits from the Paleogene of North Dakota and the London Clay. *American Journal of Botany*, 92:1650–1659. https://doi.org/10.3732/ajb.92.10.1650

Prantl K. (1887). Beiträge zur morphologie und systematik der Ranunculaceen. *Botanischer Jahrbücher für Systematik*, 9:225–273.

Rannala B., & Yang Z. (2007). Inferring speciation times under an episodic molecular clock. *Systematic Biology*, 56:453–466. https://doi.org/10.1080/10635150701420643

Rasmussen D. A., Kramer E. M., & Zimmer E. A. (2009). One size fits all? Molecular evidence for a commonly inherited petal identity program in Ranunculales. *American Journal of Botany*, 96:96–109. https://doi.org/10.3732/ajb.0800038

Reichelt N., Wen J., Pätzold C., & Appelhans M. S. (2021). Target enrichment improves phylogenetic resolution in the genus *Zanthoxylum* (Rutaceae) and indicates both incomplete lineage sorting and hybridization events. *Annals of Botany*, https://doi.org/10.1093/aob/mcab092

Ren R., Wang H., Guo C., Zhang N., Zeng L., Chen Y., Ma H., & Qi J. (2018). Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms. *Molecular Plant*, 11:414–428. 10.1016/j.molp.2018.01.002

Ro K. E., Keener C. S., & McPheron B. A. (1997). Molecular phylogenetic study of the Ranunculaceae: Utility of the nuclear 26S ribosomal DNA in inferring intrafamilial relationships. *Molecular Phylogenetics and Evolution*, 8:117–127. https://doi.org/10.1006/mpev.1997.0413

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

Romero I. G., Pai A. A., Tung J., & Gilad Y. (2014). RNA-seq: impact of RNA degradation on transcript quantification. *BMC Biology*, 12:1–13. https://doi.org/10.1186/1741-7007-12-42

Roos C., Zinner D., Kubatko L. S., Schwarz C., Yang M., Meyer D., Nash S. D., Xiang J., Batzer M. A., Brameier M., Leendertz F. H., Ziegler T., Perwitasari-Farajallah D., Nadler T., Walter L., & Osterholz M. (2011). Nuclear versus mitochondrial DNA: evidence for hybridization in colobine monkeys. *BMC Evolutionary Biology*, 11:1–13. https://doi.org/10.1186/1471-2148-11-77

Rose J. P., Toledo C. A. P., Lemmon E. M., Lemmon A. R., & Sytsma K. J. (2021). Out of sight, out of mind: widespread nuclear and plastid-nuclear discordance in the flowering plant genus *Polemonium* (Polemoniaceae) suggests widespread historical gene flow despite limited nuclear signal. *Systematic Biology*, 70:162–180. https://doi.org/10.1093/sysbio/syaa049

Sang T., & Zhong Y. (2000). Testing hybridization hypotheses based on incongruent gene trees. *Systematic Biology*, 49:422–434. https://doi.org/10.1080/10635159950127321

Shi T., & Chen J. (2020). A reappraisal of the phylogenetic placement of the *Aquilegia* whole-genome duplication. *Genome Biology*, 21:295. https://doi.org/10.1186/s13059-020-02212-y

Simão F. A., Waterhouse R. M., Ioannidis P., Kriventseva E. V., & Zdobnov E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Smith G. H. (1926). Ascular anatomy of Ranalian flowers I. Ranunculaceae. *Botanical Gazette*, 82:1–29. https://doi.org/10.1086/333631

Smith S. A., Moore M. J., Brown J. W., & Yang Y. (2015). Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evolutionary Biology*, 15:1–15. https://doi.org/10.1186/s12862-015-0423-0

Smith S. A., & O'Meara B. C. (2012). treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics*, 28:2689–2690. https://doi.org/10.1093/bioinformatics/bts492

Solís-Lemus C., Bastide P., & Ané C. (2017). PhyloNetworks: a package for phylogenetic networks. *Molecular Biology and Evolution*, 34:3292–3298. https://doi.org/10.1093/molbev/msx235

Stamatakis A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Sukumaran J., & Holder M. T. (2010). DendroPy: a Python library for phylogenetic computing. *Bioinformatics*, 26:1569–1571. https://doi.org/10.1093/bioinformatics/btq228

Takhtajan A. L. (1980). Outline of the classification of flowering plants (Magnoliophyta). *The Botanical Review*, 46:225–359. https://doi.org/10.1007/BF02861558

Tamura M. (1995). *Archiclematis & Clematis.* In P. Hiepko (Ed.), *Die Natürlichen Pflanzen- Familien* (Vol. 17, pp. 366–387). Duncker & Humblot.

Vanneste K., Baele G., Maere S., & Van de Peer Y. (2014). Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous–Paleogene boundary. *Genome Research*, 24:1334–1347. http://www.genome.org/cgi/doi/10.1101/gr.168997.113

von Balthazar, M., Pedersen, K. R., & Friis, E. M. (2005). *Teixeiria lusitanica*, a new fossil flower from the Early Cretaceous of Portugal with affinities to Ranunculales. *Plant Systematics and Evolution*, 255:55–75. https://doi.org/10.1007/s00606-005-0347-z

Wang H.X., Morales-Briones D.F., Moore M.J., Wen J., Wang H.F. (2021). A phylogenomic perspective on gene tree conflict and character evolution in Caprifoliaceae using target enrichment data, with Zabelioideae recognized as a new subfamily. *Journal of Systematics and Evolution* https://doi.org/10.1111/jse.12745

Wang K., Lenstra J. A., Liu L., Hu Q., Ma T., Qiu Q., & Liu J. (2018). Incomplete lineage sorting rather than hybridization explains the inconsistent phylogeny of the wisent. *Communications Biology*, 1:169. https://doi.org/10.1038/s42003-018-0176-6

Wang P., Liao H., Zhang W., Yu X., Zhang R., Shan H., Duan X., Yao X., & Kong H. (2015). Flexibility in the structure of spiral flowers and its underlying mechanisms. *Nature Plants*, 2:15188. https://doi.org/10.1038/nplants.2015.188

Wang W., Lin L., Xiang X. G., Ortiz R. D. C., Liu Y., Xiang K. L., Yu S. X., Xing Y. W.,

& Chen Z. D. (2016). The rise of angiosperm-dominated herbaceous floras: insights from Ranunculaceae. *Scientific Reports*, 6:27259. https://doi.org/10.1038/srep27259

Wang W., Lu A. M., Ren Y., Endress M. E., & Chen Z. D. (2009). Phylogeny and classification of Ranunculales: evidence from four molecular loci and morphological data. *Perspectives in Plant Ecology, Evolution and Systematics*, 11:81–110. https://doi.org/10.1016/j.ppees.2009.01.001

Wang Y. B., Liu B. B., Nie Z. L., Chen H. F., Chen F. J., Figlar R. B., & Wen J. (2020). Major clades and a revised classification of *Magnolia* and Magnoliaceae based on whole plastid genome sequences via genome skimming. *Journal of Systematics and Evolution*, 58:673–695. https://doi.org/10.1111/jse.12588

Wen J., Egan A. N., Dikow R. B., & Zimmer E. A. (2015). Utility of transcriptome sequencing for phylogenetic inference and character evolution. In E. Hörandl & M. S. Appelhans (Eds.), *Next-generation sequencing in plant systematics* (pp. 51–91). International Association for Plant Taxonomy (IAPT). https://doi.org/10.14630/000003

Wen J., Xiong Z., Nie Z. L., Mao L., Zhu Y., Kan X. Z., Ickert-Bond S. M., Gerrath J., Zimmer E. A., & Fang X. D. (2013). Transcriptome sequences resolve deep relationships of the grape family. *PLoS ONE*, 8:e74394. https://doi.org/10.1371/journal.pone.0074394

Wickett N. J., Mirarab S., Nguyen N., Warnow T., Carpenter E., Matasci N., Ayyampalayam S., Barker M. S., Burleigh J. G., Gitzendanner M. A., Ruhfel B. R., Wafula E., Der J. P., Graham S. W., Mathews S., Melkonian M., Soltis D. E., Soltis P. S., Miles N. W., Rothfels C. J., Pokorny L., Shaw A. J., DeGironimo L., Stevenson D. W., Surek B., Villarreal J. C., Roure B., Philippe H., dePamphilis C. W., Chen T., Deyholos M. K., Baucom R. S., Kutchan T. M., Augustin M. M., Wang J., Zhang Y., Tian Z., Yan Z., Wu X., Sun X., Wong G. K. S., & Leebens-Mack J. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences*, 111:E4859–E4868. https://doi.org/10.1073/pnas.1323926111

Wiegand K. M. (1895). The structure of the fruit in the order Ranunculaceae. *Proceedings of the American Microscopical Society*, 16:69–100. https://doi.org/10.2307/3220722

Wodehouse R. P. (1936). Pollen grains in the identification and classification of Plants. VII. The Ranunculaceae. *Bulletin of the Torrey Botanical Club*, 63:495–514. https://doi.org/10.2307/2480930

Xiang Y., Huang C. H., Hu Y., Wen J., Li S., Yi T., Chen H., Xiang J., & Ma H. (2017). Evolution of Rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication. *Molecular Biology and Evolution*, 34:262–281. https://doi.org/10.1093/molbev/msw242

Xie J., Zhao H., Li K., Zhang R., Jiang Y., Wang M., Guo X., Yu B., Kong H., Jiao Y., & Xu G. (2020). A chromosome-scale reference genome of *Aquilegia oxysepala* var. *kansuensis*. *Horticulture Research*, 7:1–13. https://doi.org/10.1038/s41438-020-0328-y

Yang Y., Moore M. J., Brockington S. F., Soltis D. E., & Wong G. K. (2015). Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. *Molecular Biology and Evolution*, 32:2001–2014. https://doi.org/10.1093/molbev/msv081

Yang Y., Moore M. J., Brockington S. F., Timoneda A., Feng T., Marx H. E., Walker J. F., & Smith S. A. (2017). An efficient field and laboratory workflow for plant phylotranscriptomic projects. *Applications in Plant Sciences*, 5:1600128. https://doi.org/10.3732/apps.1600128

Yang Y., & Smith S. A. (2014). Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: Improving accuracy and matrix occupancy for phylogenomics. *Molecular Biology and Evolution*, 31:3081–3092. https://doi.org/10.1093/molbev/msu245

Yang Y., Sun P., Lv L., Wang D., Ru D., Li Y., Ma T., Zhang L., Shen X., Meng F., & Jiao B. (2020). Prickly waterlily and rigid hornwort genomes shed light on early angiosperm evolution. *Nature Plants*, 6:215–222. https://doi.org/10.1038/s41477-020-0594-6

Yang Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24:1586–1591. https://doi.org/10.1093/molbev/msm088

Yu X., Yang D., Guo C., & Gao L. (2018). Plant phylogenomics based on genome-partitioning strategies: Progress and prospects. *Plant Diversity*, 40:158–164. https://doi.org/10.1016/j.pld.2018.06.005

Yu Y., & Nakhleh L. (2015). A maximum pseudo-likelihood approach for phylogenetic networks. *BMC Genomics*, 16:1–10. https://doi.org/10.1186/1471-2164-16-S10-S10

Zeng L., Zhang N., Zhang Q., Endress P. K., Huang J., & Ma H. (2017). Resolution of deep eudicot phylogeny and their temporal diversification using nuclear genes from transcriptomic and genomic datasets. *New Phytologist*, 214:1338–1354. https://doi.org/10.1111/nph.14503

Zhai W., Duan X., Zhang R., Guo C., Li L., Xu G., Shan H., Kong H., & Ren Y. (2019). Chloroplast genomic data provide new and robust insights into the phylogeny and evolution of the Ranunculaceae. *Molecular Phylogenetics and Evolution*, 135:12–21. https://doi.org/10.1016/j.ympev.2019.02.024

Zhan C. S., Li X. H., Zhao Z. Y., Yang T. W., Wang X. K., Luo B. B., Zhang Q. Y., Hu Y. R., & Hu X. B. (2016). Comprehensive analysis of the triterpenoid saponins biosynthetic pathway in *Anemone flaccida* by transcriptome and proteome profiling. *Frontiers in Plant Science*, 7:1094. https://doi.org/10.3389/fpls.2016.01094

Zhang C., Rabiee M., Sayyari E., & Mirarab S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, 19:15–30. https://doi.org/10.1186/s12859-018-2129-y

Zhang R., Guo C., Zhang W., Wang P., Li L., Duan X., Du Q., Zhao L., Shan H., & Hodges S. A. (2013). Disruption of the petal identity gene APETALA3-3 is highly correlated with loss of petals within the buttercup family (Ranunculaceae). *Proceedings of the National Academy of Sciences*, 110:5074–5079.

https://doi.org/10.1073/pnas.1219690110

Zhao S. Y., Chen L. Y., Wei Y. L., Wang Q. F., & Moody M. L. (2016). RNA-seq of *Ranunculus sceleratus* and identification of orthologous genes among four *Ranunculus* species. *Frontiers in Plant Science*, 7:732. https://doi.org/10.3389/fpls.2016.00732

Zhou A., Sun H. W., Dai S. Y., Feng S., Zhang J. Z., Gong S. F., & Wang J. G. (2019). Identification of transcription factors involved in the regulation of flowering in *Adonis amurensis* through combined RNA-seq transcriptomics and iTRAQ proteomics. *Genes*, 10:305. https://doi.org/10.3390/genes10040305

Zimmer E. A, & Wen J. (2015). Using nuclear gene data for plant phylogenetics: Progress and prospects II. Next-gen approaches. *Journal of Systematics and Evolution*, 53:371–379. https://doi.org/10.1111/jse.12174

DATA AVAILABILITY STATEMENT

Raw sequence reads are deposited in the NCBI SRA BioProject ID PRJNA739247; Tree, alignment and code are deposited in the Zenodo (https://doi.org/10.5281/zenodo.5144194).

AUTHOR CONTRIBUTIONS

Jian He: Formal analysis, Methodology, Software, Validation, Visualization, Writing original draft. Rudan Lyu: Investigation, Resources, Data curation. Yike Luo: Investigation, Resources. Jiamin Xiao: Investigation, Data curation. Lei Xie: Writing - review & editing, Funding acquisition. Jun Wen: Writing - review & editing, Supervision. Wenhe Li: Investigation, Resources. Linying Pei: Investigation, Resources. Jin Cheng: Project administration, Supervision.

FIGURE LEGENDS

FIGURE 1 Quality and quantity assessments of RNA extracted from Ranunculales leaf tissues preserved in one of three different ways. Comparisons are of (a-d) genomic RNA, (e-g and i-k) de-novo sequences assembled by Trinity v.2.1.5 (Grabherr et al., 2011), and (h and l the numbers of single-copy orthologous genes compared to BUSCO (Simão et al., 2015). The LN+T samples category includes the 27 samples downloaded from GenBank (see Table 1) plus the liquid nitrogen-preserved samples (n = 6). One-way ANOVA, * $P < 0.05$. OD: optical density; Q30: Phred quality score 30 threshold. For detailed sampling information see Table 1

FIGURE 2 Heatmap of gene recovery efficiency. Each row represents one sample and each column represents one gene (3,611 genes total). Shading indicates a gene's absence. The blue zone shows traditional treatment samples (i.e., tissues preserved in liquid nitrogen or in RNA stabilization solution) (see Table 1). The purple zone shows the 21 silica-dried samples from the current study

FIGURE 3 Phylogenetic relationships of Ranunculaceae. (a) The phylogeny inferred by the concatenated 3,611 single-copy co-orthologous nuclear gene (SCOG) data was produced using the maximum likelihood (ML) method. Bootstrap percentages are indicated on the branches and bold branches showed ML bootstrap values of 100. (b) The coalescence-based species tree topology was inferred by ASTRAL (Zhang, et al., 2018) from 3,611 nuclear genes. Numbers at branches are local posterior probabilities (ASTRAL-pp), but bold branches mark local posterior probabilities equal to 1.00. C-, T-, and R-types are chromosome groups and well-supported clades. The species in black are outgroups

FIGURE 4 Coalescence-based species tree (Figure 3b) topology showing discordance among genes. The ASTRAL (Zhang et al., 2018) Ranunculaceae tree has all nodes resolved in most of the species trees (in heavy black lines). The gray-colored trees were sampled from 94 single-copy co-orthologous nuclear genes (from the 3,611 gene data set in which all the taxa were present in a single gene tree) and constructed with RAxML v.8.2.12. Pie charts on the nodes show proportions of concordant and discordant tree topologies compared to the species tree. C-, T-, and R-types are chromosome groups and well-supported clades

FIGURE 5 Comparison of the Ranunculaceae nuclear and plastid phylogenies. Left: Coalescence-based species tree topology inferred by ASTRAL (Zhang et al., 2018) from 3,611 nuclear genes (see Figure 3b). Numbers at branches are local posterior probabilities (ASTRAL-pp) but local posterior probabilities equal to 1.00 are marked with bold branches. Right: The Ranunculaceae phylogeny inferred from the concatenated coding regions of the complete plastome sequences using the maximum likelihood (ML) method. Bootstrap percentages are indicated on the branches and bold branch indicates ML bootstrap values of 100%. Numbers in brackets show the contribution of incomplete lineage sorting to the conflicts between the nuclear and plastid gene trees based on the multispecies coalescent model implemented in Phybase v.1.5 (Liu and Yu, 2010). Plastome sequence data are from previous studies (see Table S1)

FIGURE 6 Testing the contributions of incomplete lineage sorting (ILS) and hybridization to the conflicts between the Ranunculaceae gene trees and the species tree inferred by ASTRAL (Zhang et al., 2018). (a) Coalescent simulations of tree-to-tree distance distributions between the empirical gene trees and the ASTRAL species tree (Empirical) and those from the coalescent simulation (Simulated). (b) Introgression model of unstable Ranunculaceae clades estimated using Phylonetworks. The results display maximum pseudolikelihood trees with a maximum zero to five allowed reticulations, and (c) the best model had a maximum allowed hybrid node of one according to the pseudo-loglikelihood scores (-ploglik) shown in the bar chart in (b). Red arrows mark possible reticulation events, γ is the gene contribution, and BS is the bootstrap support for the placement of minor hybrid branches

FIGURE 7 Genome duplication analysis under the absolute dating context. (a) Chronogram showing estimated divergence times according to MCMCTree and based on the Ranunculaceae coalescent species tree using 3,611 single-copy orthologous genes. The lineages have different colors and the taxa in a given lineage are grouped in one color. The outgroups are in black. Divergence times and numbers of gene trees with duplications (estimated by Phyparts) are marked under (or above) the branches. Starred clades are R: all Ranunculales, RE: R without Eupteleaceae, RH: R without Hydrastideae, CR: core Ranunculaceae, DEL: CR without Delphinieae, and ANE: CR without Anemoneae. (b) Absolute age distributions of genome duplications among major Ranunculales clades

SUPPLEMENTAL INFORMATION:

SUPPLEMENTARY MATERIALS Protocol of RNA extraction used in this study

FIGURE S1 Ranunculaceae phylogenies inferred from nuclear datasets with different selective constrains (see Materials and Methods, 2.3). Phylogenies inferred from concatenated datasets were produced using the maximum likelihood method. Bootstrap percentages are indicated on the branches and bold branch shows ML bootstrap values of 100. Coalescence-based species tree topologies were inferred by ASTRAL (Zhang et al., 2018). Numbers at branches are local posterior probabilities (ASTRAL-pp), but bold branches mark local posterior probabilities equal to 1.00

FIGURE S2 Chronogram of Ranunculaceae inferred from the ASTRAL (Zhang et al., 2018) topology of the 3,611 SCOG coalescent data set using MCMCTree implemented in the PAML v.4.9a package (Yang, 2007). Red stars mark fossil constraints which are explained in the text (Materials and Methods 2.5). Numbers on the branches indicate the median value (95% high posterior density interval) of the estimated divergence time (Mya)

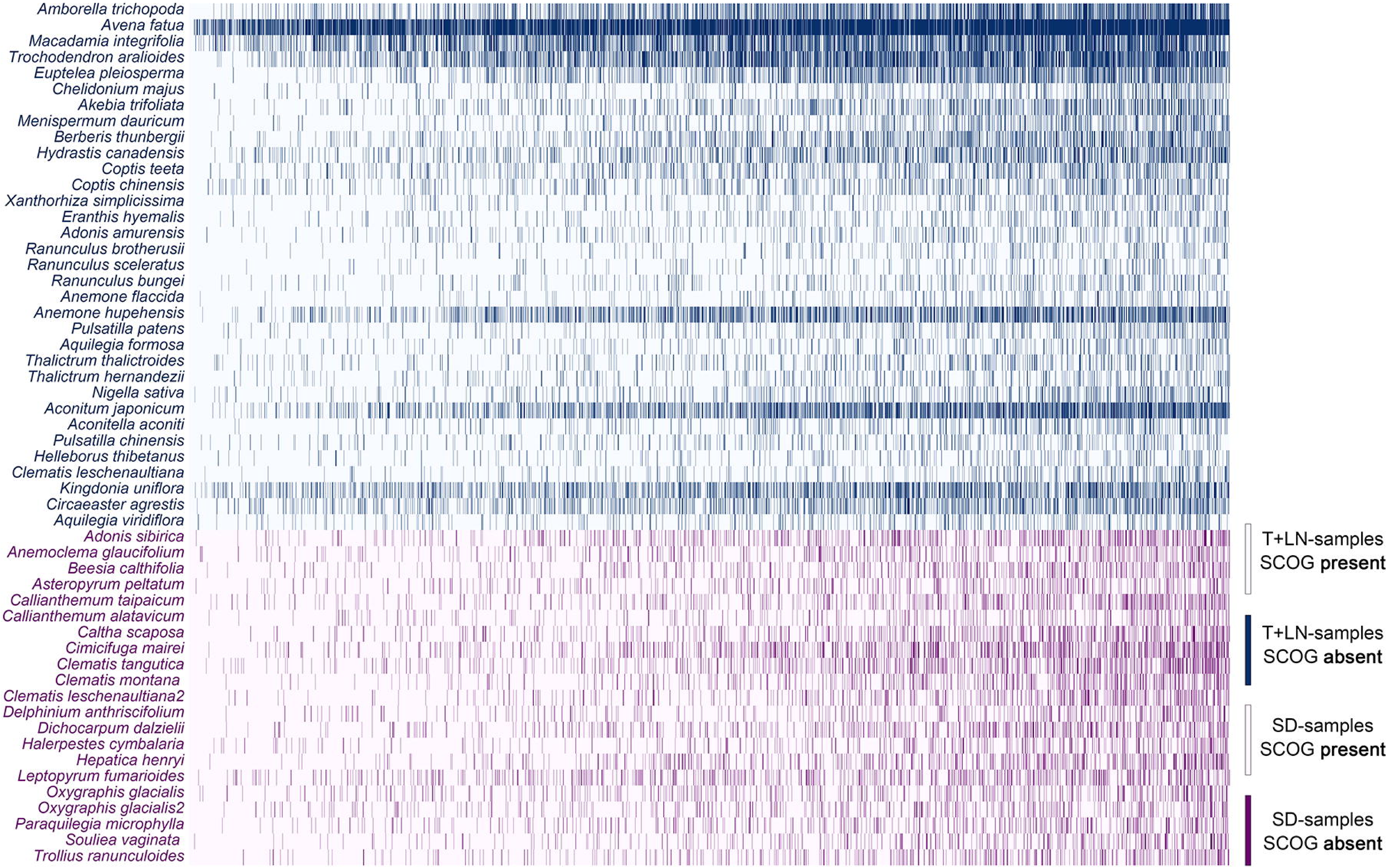
TABLE S1 Species sampling information and references for plastome sequences used in Figure 5

TABLE S2 RNA quality and sequencing information for the newly sequenced samples in this study. One-way ANOVAs for each column are included in worksheets in this Excel workbook

TABLE S3 Information of de novo assembly of newly sequenced samples in this study. One-way ANOVAs for each column are included in worksheets in this Excel workbook

TABLE S4 Information about the de novo assembly of all samples (including newly sequenced and downloaded) in this study. One-way ANOVAs for each column are included in worksheets in this Excel workbook

(a)
RNA integrity (RIN score)
P = 0.08
LN-samples n = 6
SD-samples n = 21
(b)
RNA purity (OD 260/230)
P = 0.35
LN-samples n = 6
SD-samples n = 21
(c)
Sequence quality scores (Q30)
*P = 0.04
LN-samples n = 6
SD-samples n = 21
(d)
Total Obtained reads (×10⁸)
*P = 5.6 × 10⁻⁵
LN-samples n = 6
SD-samples n = 21
(e)
Number of contigs generated by Trinity
P = 1.7 × 10⁻⁴
LN-samples n = 6
SD-samples n = 21
(f)
Average length of contigs generated by Trinity (bp)
P = 0.44
LN-samples n = 6
SD-samples n = 21
(g)
N50 of contigs generated by Trinity (bp)
P = 0.29
LN-samples n = 6
SD-samples n = 21
(h)
Number of single-copy orthologous genes
P = 0.07
LN-samples n = 6
SD-samples n = 21
(i)
Number of contigs generated by Trinity
P = 0.18
T+LN-samples n = 33
SD-samples n = 21
(j)
Average length of contigs generated by Trinity (bp)
P = 0.11
T+LN-samples n = 33
SD-samples n = 21
(k)
N50 of contigs generated by Trinity (bp)
*P = 9.6 × 10⁻³
T+LN-samples n = 33
SD-samples n = 21
(l)
Number of single-copy orthologous genes
*P = 0.03
T+LN-samples n = 33
SD-samples n = 21

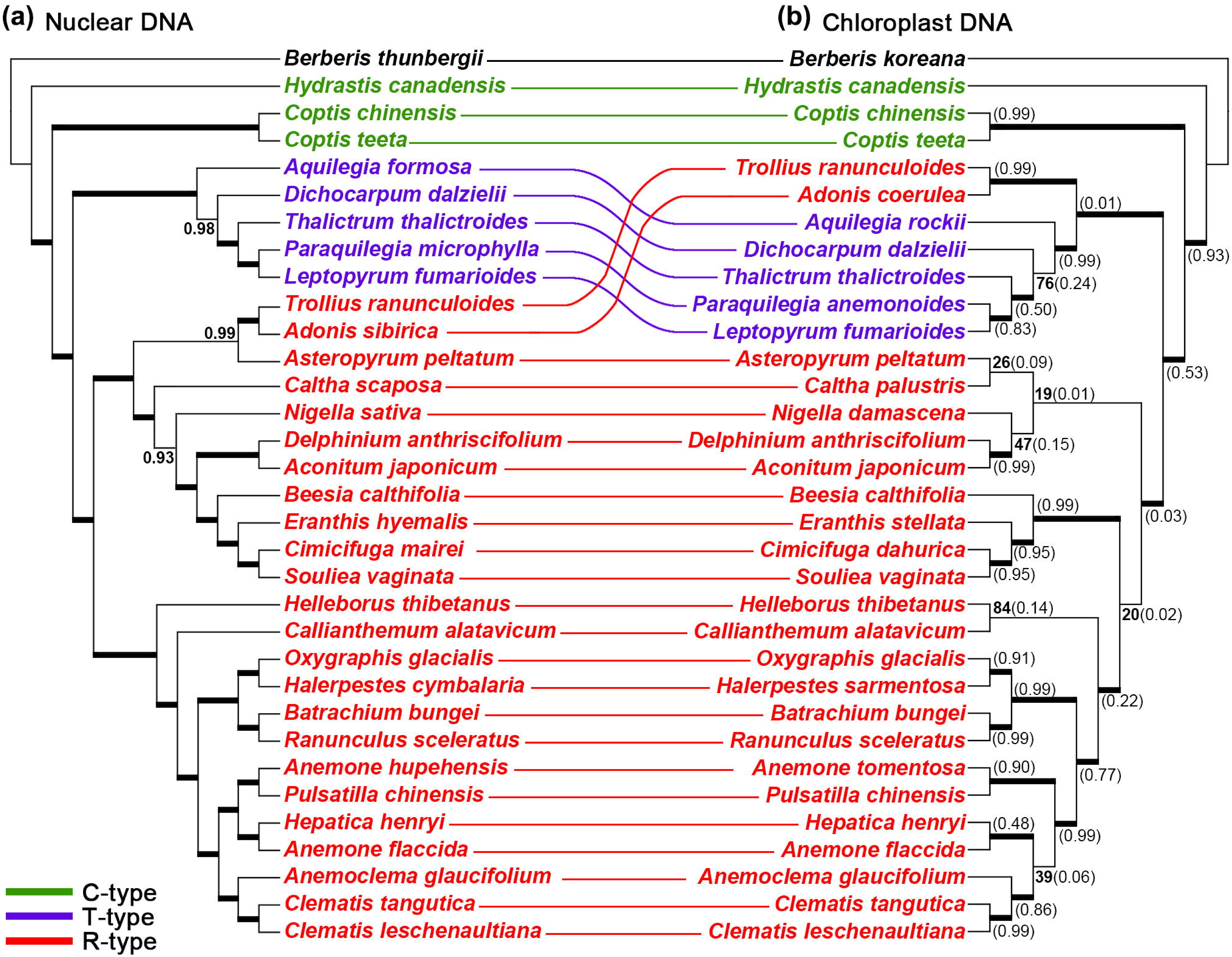

Amborella trichopoda
Avena fatua
Macadamia integrifolia
Trochodendron aralioides
Euptelea pleiosperma
Chelidonium majus
Akebia trifoliata
Menispermum dauricum
Berberis thunbergii
Hydrastis canadensis
Coptis teeta
Coptis chinensis
Xanthorhiza simplicissima
Eranthis hyemalis
Adonis amurensis
Ranunculus brotherusii
Ranunculus sceleratus
Ranunculus bungei
Anemone flaccida
Anemone hupehensis
Pulsatilla patens
Aquilegia formosa
Thalictrum thalictroides
Thalictrum hernandezii
Nigella sativa
Aconitum japonicum
Aconitella aconiti
Pulsatilla chinensis
Helleborus thibetanus
Clematis leschenaultiana
Kingdonia uniflora
Circaeaster agrestis
Aquilegia viridiflora
Adonis sibirica
Anemoclema glaucifolium
Beesia calthifolia
Asteropyrum peltatum
Callianthemum taipaicum
Callianthemum alatavicum
Caltha scaposa
Cimicifuga mairei
Clematis tangutica
Clematis montana
Clematis leschenaultiana2
Delphinium anthriscifolium
Dichocarpum dalzielii
Halerpestes cymbalaria
Hepatica henryi
Leptopyrum fumarioides
Oxygraphis glacialis
Oxygraphis glacialis2
Paraquilegia microphylla
Souliea vaginata
Trollius ranunculoides
T+LN-samples SCOG present
T+LN-samples SCOG absent
SD-samples SCOG present
SD-samples SCOG absent

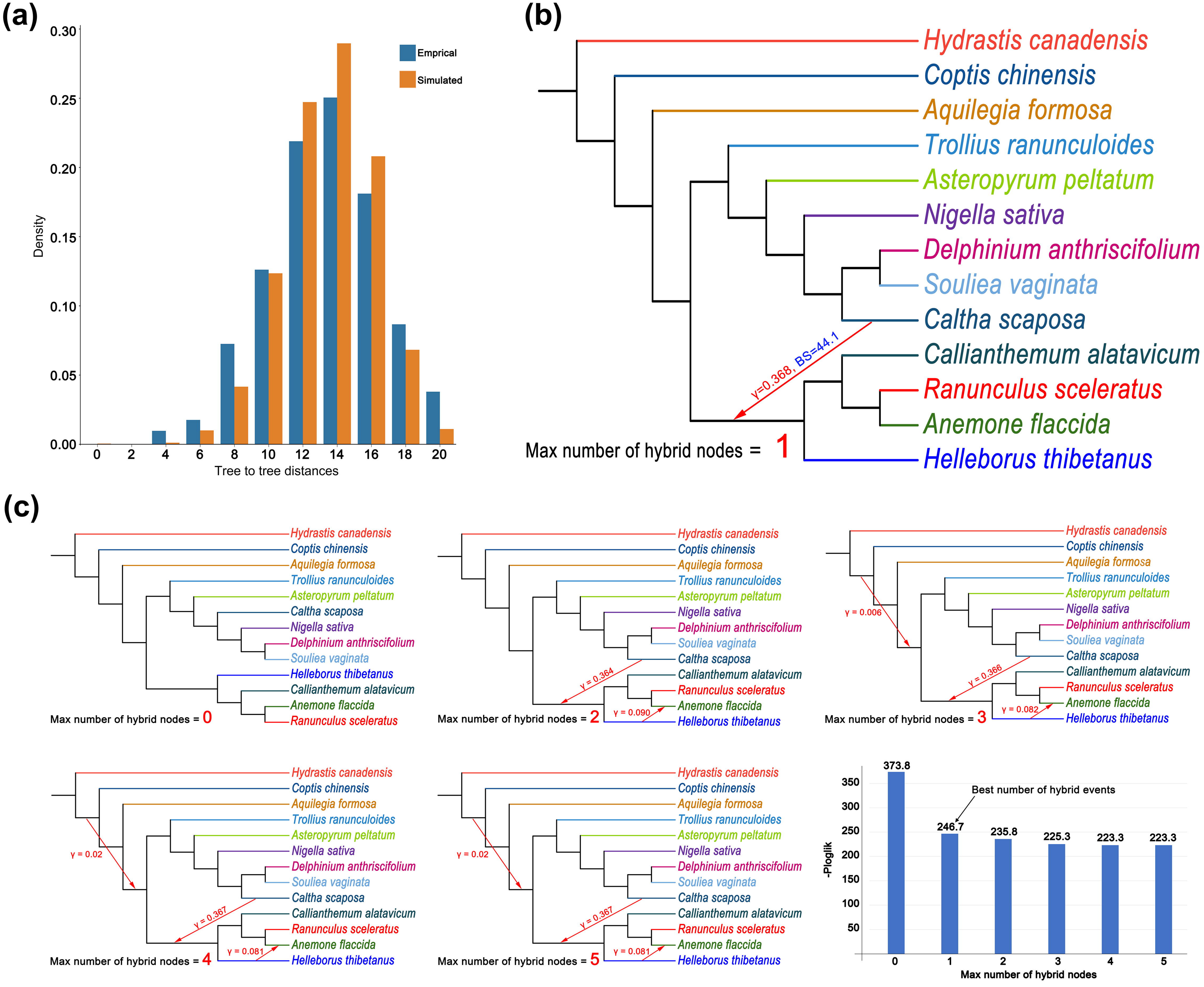
(a) Concatencation tree

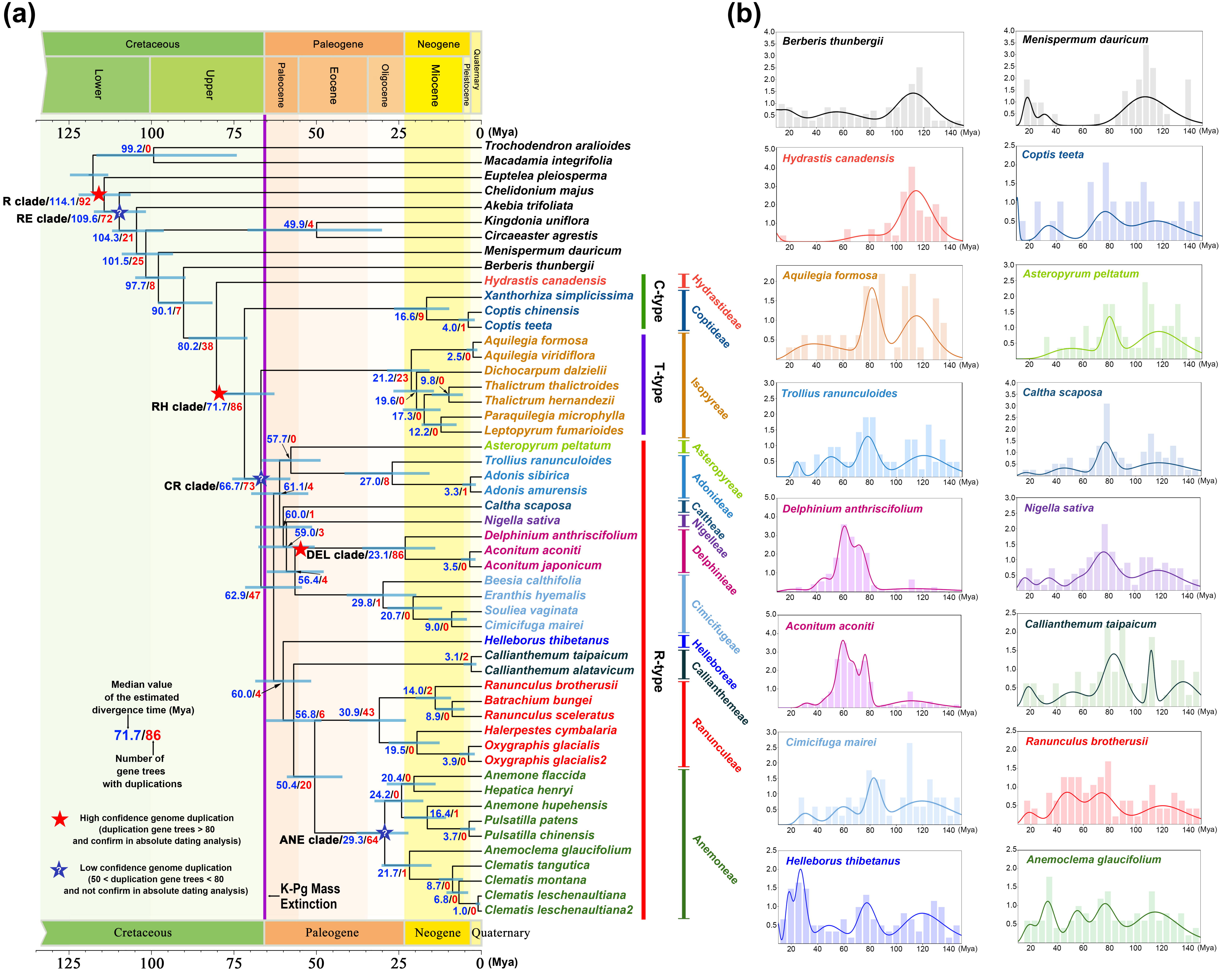
(b) Coalescent tree

Amborella trichopoda
Avena fatua
54
Trochodendron aralioides
Macadamia integrifolia
Euptelea pleiosperma
Chelidonium majus
Akebia trifoliata
Kingdonia uniflora
Circaeaster agrestis
Menispermum dauricum
Berberis thunbergii
Hydrastis canadensis
Xanthorhiza simplicissima
Coptis chinensis
Coptis teeta
Aquilegia formosa
Aquilegia viridiflora
Dichocarpum dalzielii
Paraquilegia microphylla
Leptopyrum fumarioides
Thalictrum hernandezii
Thalictrum thalictroides
Asteropyrum peltatum
Trollius ranunculoides
Adonis amurensis
Adonis sibirica
40
Caltha scaposa
Nigella sativa
Delphinium anthriscifolium
Aconitum japonicum
Aconitum aconiti
40
Beesia calthifolia
Eranthis hyemalis
Cimicifuga mairei
Souliea vaginata
Helleborus thibetanus
Callianthemum taipaicum
Callianthemum alatavicum
Halerpestes cymbalaria
Oxygraphis glacialis
Oxygraphis glacialis2
Ranunculus brotherusii
Ranunculus sceleratus
Batrachium bungei
Hepatica henryi
Anemone flaccida
Anemone hupehensis
Pulsatilla chinensis
Pulsatilla patens
Anemoclema glaucifolium
Clematis tangutica
Clematis montana
Clematis leschenaultiana
Clematis leschenaultiana2
substitution per site
0.1
C-type
T-type
R-type
Hydrastideae
Coptideae
Isopyreae
Asteropyreae
Adonideae
Caltheae
Nigelleae
Delphinieae
Cimicifugeae
Helleboreae
Callianthemeae
Ranunculeae
Anemoneae
Amborella trichopoda
Avena fatua
Macadamia integrifolia
Trochodendron aralioides
Euptelea pleiosperma
Chelidonium majus
Akebia trifoliata
Circaeaster agrestis
Kingdonia uniflora
Menispermum dauricum
Berberis thunbergii
Hydrastis canadensis
Xanthorhiza simplicissima
Coptis teeta
Coptis chinensis
Aquilegia viridiflora
Aquilegia formosa
Dichocarpum dalzielii
0.98
Thalictrum hernandezii
Thalictrum thalictroides
Leptopyrum fumarioides
Paraquilegia microphylla
Asteropyrum peltatum
0.99
Trollius ranunculoides
Adonis amurensis
Adonis sibirica
Caltha scaposa
Nigella sativa
0.93
Delphinium anthriscifolium
Aconitum japonicum
Aconitum aconiti
Beesia calthifolia
Eranthis hyemalis
Cimicifuga mairei
Souliea vaginata
Helleborus thibetanus
Callianthemum taipaicum
Callianthemum alatavicum
Ranunculus brotherusii
Ranunculus sceleratus
Batrachium bungei
Halerpestes cymbalaria
Oxygraphis glacialis
Oxygraphis glacialis2
Hepatica henryi
Anemone flaccida
Anemone hupehensis
Pulsatilla chinensis
Pulsatilla patens
Anemoclema glaucifolium
Clematis tangutica
Clematis montana
Clematis leschenaultiana
Clematis leschenaultiana2
C-type
T-type
R-type
Hydrastideae
Coptideae
Isopyreae
Asteropyreae
Adonideae
Caltheae
Nigelleae
Delphinieae
Cimicifugeae
Helleboreae
Callianthemeae
Ranunculeae
Anemoneae

Cretaceous
Paleogene
Paleocene
Eocene
Oligocene
Neogene
Miocene
Pleistocene
Quaternary
75
50
25
0 (Mya)
K-Pg Mass Extinction
Hydrastis canadensis
Xanthorhiza simplicissima
Coptis chinensis
Coptis teeta
Aquilegia formosa
Aquilegia viridiflora
Dichocarpum dalzielii
Paraquilegia microphylla
Leptopyrum fumarioides
Thalictrum thalictroides
Thalictrum hernandezii
Trollius ranunculoides
Adonis amurensis
Adonis sibirica
Asteropyrum peltatum
Caltha scaposa
Nigella sativa
Delphinium anthriscifolium
Aconitum japonicum
Aconitum aconiti
Beesia calthifolia
Eranthis hyemalis
Souliea vaginata
Cimicifuga mairei
Helleborus thibetanus
Callianthemum taipaicum
Callianthemum alatavicum
Halerpestes cymbalaria
Oxygraphis glacialis
Oxygraphis glacialis2
Ranunculus brotherusii
Ranunculus sceleratus
Batrachium bungei
Anemoclema glaucifolium
Clematis tangutica
Clematis montana
Clematis leschenaultiana
Clematis leschenaultiana2
Hepatica henryi
Anemone flaccida
Anemone hupehensis
Pulsatilla chinensis
Pulsatilla patens
C-type
T-type
R-type
Hydrastideae
Coptideae
Isopyreae
Adonideae
Asteropyreae
Caltheae
Nigelleae
Delphinieae
Cimicifugeae
Helleboreae
Callianthemeae
Ranunculeae
Anemoneae
Ranunculaceae
Discordant
Concordant
Cretaceous
Paleogene
Neogene
Quaternary
75
50
25
0 (Mya)


(a) Nuclear DNA
(b) Chloroplast DNA
Berberis thunbergii
Hydrastis canadensis
Coptis chinensis
Coptis teeta
Aquilegia formosa
Dichocarpum dalzielii
Thalictrum thalictroides
Paraquilegia microphylla
Leptopyrum fumarioides
Trollius ranunculoides
Adonis sibirica
Asteropyrum peltatum
Caltha scaposa
Nigella sativa
Delphinium anthriscifolium
Aconitum japonicum
Beesia calthifolia
Eranthis hyemalis
Cimicifuga mairei
Souliea vaginata
Helleborus thibetanus
Callianthemum alatavicum
Oxygraphis glacialis
Halerpestes cymbalaria
Batrachium bungei
Ranunculus sceleratus
Anemone hupehensis
Pulsatilla chinensis
Hepatica henryi
Anemone flaccida
Anemoclema glaucifolium
Clematis tangutica
Clematis leschenaultiana
0.98
0.99
0.93
Berberis koreana
Hydrastis canadensis
Coptis chinensis
Coptis teeta
Trollius ranunculoides
Adonis coerulea
Aquilegia rockii
Dichocarpum dalzielii
Thalictrum thalictroides
Paraquilegia anemonoides
Leptopyrum fumarioides
Asteropyrum peltatum
Caltha palustris
Nigella damascena
Delphinium anthriscifolium
Aconitum japonicum
Beesia calthifolia
Eranthis stellata
Cimicifuga dahurica
Souliea vaginata
Helleborus thibetanus
Callianthemum alatavicum
Oxygraphis glacialis
Halerpestes sarmentosa
Batrachium bungei
Ranunculus sceleratus
Anemone tomentosa
Pulsatilla chinensis
Hepatica henryi
Anemone flaccida
Anemoclema glaucifolium
Clematis tangutica
Clematis leschenaultiana
(0.99)
(0.99)
(0.01)
(0.93)
(0.99)
76(0.24)
(0.50)
(0.83)
(0.53)
26(0.09)
19(0.01)
47(0.15)
(0.99)
(0.99)
(0.03)
(0.95)
(0.95)
84(0.14)
20(0.02)
(0.91)
(0.99)
(0.22)
(0.99)
(0.90)
(0.77)
(0.48)
(0.99)
39(0.06)
(0.86)
(0.99)
C-type
T-type
R-type


(a)
Emprical
Simulated
Density
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
2
4
6
8
10
12
14
16
18
20
Tree to tree distances
(b)
Hydrastis canadensis
Coptis chinensis
Aquilegia formosa
Trollius ranunculoides
Asteropyrum peltatum
Nigella sativa
Delphinium anthriscifolium
Souliea vaginata
Caltha scaposa
Callianthemum alatavicum
Ranunculus sceleratus
Anemone flaccida
Helleborus thibetanus
γ=0.368, BS=44.1
Max number of hybrid nodes = 1
(c)
Max number of hybrid nodes = 0
Max number of hybrid nodes = 2
γ = 0.364
γ = 0.090
Max number of hybrid nodes = 3
γ = 0.006
γ = 0.366
γ = 0.082
Max number of hybrid nodes = 4
γ = 0.02
γ = 0.367
γ = 0.081
Max number of hybrid nodes = 5
γ = 0.02
γ = 0.367
γ = 0.081
-Ploglik
50
100
150
200
250
300
350
373.8
246.7
235.8
225.3
223.3
223.3
Best number of hybrid events
0
1
2
3
4
5
Max number of hybrid nodes


(a)
Cretaceous
Lower
Upper
Paleogene
Paleocene
Eocene
Oligocene
Neogene
Miocene
Pleistocene
Quaternary
125
100
75
50
25
0 (Mya)
99.2/0
Trochodendron aralioides
Macadamia integrifolia
Euptelea pleiosperma
R clade/114.1/92
Chelidonium majus
RE clade/109.6/72
Akebia trifoliata
49.9/4
Kingdonia uniflora
104.3/21
Circaeaster agrestis
101.5/25
Menispermum dauricum
Berberis thunbergii
97.7/8
Hydrastis canadensis
90.1/7
Xanthorhiza simplicissima
16.6/9
Coptis chinensis
4.0/1
Coptis teeta
80.2/38
Aquilegia formosa
2.5/0
Aquilegia viridiflora
Dichocarpum dalzielii
21.2/23
9.8/0
Thalictrum thalictroides
RH clade/71.7/86
Thalictrum hernandezii
19.6/0
Paraquilegia microphylla
17.3/0
Leptopyrum fumarioides
12.2/0
57.7/0
Asteropyrum peltatum
Trollius ranunculoides
CR clade/66.7/73
27.0/8
Adonis sibirica
61.1/4
3.3/1
Adonis amurensis
Caltha scaposa
60.0/1
Nigella sativa
59.0/3
Delphinium anthriscifolium
DEL clade/23.1/86
Aconitum aconiti
3.5/0
Aconitum japonicum
56.4/4
Beesia calthifolia
62.9/47
Eranthis hyemalis
29.8/1
20.7/0
Souliea vaginata
9.0/0
Cimicifuga mairei
Helleborus thibetanus
3.1/2
Callianthemum taipaicum
Callianthemum alatavicum
Median value of the estimated divergence time (Mya)
60.0/4
Ranunculus brotherusii
14.0/2
Batrachium bungei
56.8/6
30.9/43
8.9/0
Ranunculus sceleratus
71.7/86
Halerpestes cymbalaria
Number of gene trees with duplications
19.5/0
Oxygraphis glacialis
3.9/0
Oxygraphis glacialis2
20.4/0
Anemone flaccida
50.4/20
Hepatica henryi
24.2/0
Anemone hupehensis
★ High confidence genome duplication (duplication gene trees > 80 and confirm in absolute dating analysis)
16.4/1
Pulsatilla patens
ANE clade/29.3/64
3.7/0
Pulsatilla chinensis
Anemoclema glaucifolium
★ Low confidence genome duplication (50 < duplication gene trees < 80 and not confirm in absolute dating analysis)
21.7/1
Clematis tangutica
8.7/0
Clematis montana
K-Pg Mass Extinction
6.8/0
Clematis leschenaultiana
1.0/0
Clematis leschenaultiana2
C-type
T-type
R-type
Hydrastideae
Coptideae
Isopyreae
Asteropyreae
Adonideae
Caltheae
Nigelleae
Delphinieae
Cimicifugeae
Helleboreae
Callianthemeae
Ranunculeae
Anemoneae
Cretaceous
Paleogene
Neogene
Quaternary
125
100
75
50
25
0 (Mya)
(b)
Berberis thunbergii
Menispermum dauricum
Hydrastis canadensis
Coptis teeta
Aquilegia formosa
Asteropyrum peltatum
Trollius ranunculoides
Caltha scaposa
Delphinium anthriscifolium
Nigella sativa
Aconitum aconiti
Callianthemum taipaicum
Cimicifuga mairei
Ranunculus brotherusii
Helleborus thibetanus
Anemoclema glaucifolium
20
40
60
80
100
120
140 (Mya)