



# Organization, Phylogenetic Marker Exploitation, and Gene Evolution in the Plastome of *Thalictrum* (Ranunculaceae)

## OPEN ACCESS

### Edited by:

Jiayu Zhou,

Nankai University, China

### Reviewed by:

Xiaoguo Xiang,

Nanchang University, China

Weishu Fan,

Kunming Institute of Botany (CAS),  
China

### \*Correspondence:

Zhi-Qiang Wu

wuzhiqiang@caas.cn

Wen-Chuang He

hewenchuang@caas.cn

Ying-Xue Yang

yxyxue32@163.com

<sup>†</sup>These authors have contributed  
equally to this work and share last  
authorship

### Specialty section:

This article was submitted to  
Plant Bioinformatics,  
a section of the journal  
Frontiers in Plant Science

Received: 16 March 2022

Accepted: 11 April 2022

Published: 20 May 2022

### Citation:

Xiang K-L, Mao W, Peng H-W,  
Erst AS, Yang Y-X, He W-C and  
Wu Z-Q (2022) Organization,  
Phylogenetic Marker Exploitation, and  
Gene Evolution in the Plastome of  
*Thalictrum* (Ranunculaceae).  
*Front. Plant Sci.* 13:897843.  
doi: 10.3389/fpls.2022.897843

Kun-Li Xiang<sup>1,2</sup>, Wei Mao<sup>3</sup>, Huan-Wen Peng<sup>2,4</sup>, Andrey S. Erst<sup>5,6</sup>, Ying-Xue Yang<sup>1\*†</sup>,  
Wen-Chuang He<sup>1\*†</sup> and Zhi-Qiang Wu<sup>1,7\*†</sup>

<sup>1</sup>Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China, <sup>2</sup>State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing, China, <sup>3</sup>College of Ecology and Environment, Hainan University, Haikou, China, <sup>4</sup>College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China, <sup>5</sup>Central Siberian Botanical Garden, Russian Academy of Sciences, Novosibirsk, Russia, <sup>6</sup>Laboratory Herbarium (TK), Tomsk State University, Tomsk, Russia, <sup>7</sup>Kunpeng Institute of Modern Agriculture at Foshan, Chinese Academy of Agricultural Sciences, Foshan, China

*Thalictrum* is a phylogenetically and economically important genus in the family Ranunculaceae, but is also regarded as one of the most challengingly difficult in plants for resolving the taxonomical and phylogenetical relationships of constituent taxa within this genus. Here, we sequenced the complete plastid genomes of two *Thalictrum* species using Illumina sequencing technology *via de novo* assembly. The two *Thalictrum* plastomes exhibited circular and typical quadripartite structure that was rather conserved in overall structure and the synteny of gene order. By updating the previously reported plastome annotation of other nine *Thalictrum* species, we found that the expansion or contraction of the inverted repeat region affect the boundary of the single-copy regions in *Thalictrum* plastome. We identified eight highly variable noncoding regions—*infA-rps8*, *ccsA-ndhD*, *trnS<sup>UGA</sup>-psbZ*, *trnH<sup>GUG</sup>-psbA*, *rpl16-rps3*, *ndhG-ndhI*, *ndhD-psaC*, and *ndhJ-ndhK*—that can be further used for molecular identification, phylogenetic, and phylogeographic in different species. Selective pressure and codon usage bias of all the plastid coding genes were also analyzed for the 11 species. Phylogenetic relationships showed *Thalictrum* is monophyly and divided into two major clades based on 11 *Thalictrum* plastomes. The availability of these plastomes offers valuable genetic information for accurate identification of species and taxonomy, phylogenetic resolution, and evolutionary studies of *Thalictrum*, and should assist with exploration and utilization of *Thalictrum* plants.

**Keywords:** *Thalictrum*, plastid genome, genome structure, molecular markers, phylogeny

## INTRODUCTION

*Thalictrum* L., comprising ca. 200 species, is a phylogenetically and economically important genus in the family Ranunculaceae (Tamura, 1995) and is worldwide with main distribution in northern temperate regions. *Thalictrum* plants are rich in benzylisoquinoline-derived alkaloids; at least 250 such compounds have been isolated from 60 species, and most of them show strong biological activities (Zhu and Xiao, 1991). *Thalictrum* plants are used in folk medicine for the treatment of many kinds of diseases by various ethnic groups of China, which has a long history (Wang and Xiao, 1979; Zhu and Xiao, 1989; Wu et al., 1998; Wang et al., 2001). In some place, roots of *Thalictrum* were used as substitutes for *Rhizoma coptidis* to treat enteritis and dysentery (Wu et al., 1998). Furthermore, bearing luxuriant foliage, extended branches, and attractive flowers, *Thalictrum* species have previously been mainly applied as perennial garden plants. At present, the horticultural values of *Thalictrum* plants, such as *Thalictrum delavayi*, *Thalictrum reniforme*, and *Thalictrum grandiflorum* have been widely paid attention with great commercial prospects (Wang and Xiao, 1979).

*Thalictrum* is taxonomically and phylogenetically regarded as one of the most challengingly difficult taxa in plants. Traditionally, *Thalictrum* was classified into 14 sections based on morphological traits such as leaf, flower, and fruit characteristics (Tamura, 1995). Molecular phylogenetic analyses have consistently suggested only that *Thalictrum* is a monophyletic group containing two major clades, based on the nuclear ribosomal internal transcribed spacer (ITS) region (ITS1, ITS2, and 5.8S) and the chloroplast DNA (cpDNA) *rpl16* intron (Soza et al., 2012). Then, a revised phylogeny yielded better resolution based on nuclear ribosomal ITS region, external transcribed spacer (ETS) region, and the cpDNA 3'*trnV*-*ndhC* (*trnV*-*ndhC*) intergenic region (Soza et al., 2013). Nonetheless, none of the sections traditionally assigned to the genus (Tamura, 1995) are monophyletic (Soza et al., 2012, 2013). Moreover, numerous species and varieties in *Thalictrum* are poorly defined owing to insufficient field studies and lack of consistent characteristics for diagnostic methods in the literature (Wang et al., 2001). Therefore, further exploiting more stable genetic variations and effective molecular markers in *Thalictrum* species is greatly important for conservation and utilization of the plants from this genus.

The popularity of the ITS region for infrageneric studies within angiosperms is well-known (Baldwin et al., 1995; Hughes et al., 2006; Mort et al., 2007). Levels of ITS sequence divergence within *Thalictrum* are relatively high (Soza et al., 2012, 2013). However, *Thalictrum* exhibits an enormous range of ploidy, from  $2n=2x=14$  to  $2n=24x=168$  (Löve, 1982; Tamura, 1995), with very small chromosomes known as the T-type in Ranunculaceae (Langlet, 1927). In *Thalictrum*, the ITS region is often presented as more than one copy (Soza et al., 2012, 2013). Owing to their haploidy, maternal inheritance, and high conservation in gene content and genome structure, the plastomes have been popular in researches on evolutionary relationships at almost any taxonomic level in plants. Although sequence divergence among the interspecific cpDNAs is generally less than ITS (Hughes et al., 2006; Mort et al., 2007), it is necessary to utilize cpDNA regions that exhibit relatively high rates of

substitution in *Thalictrum*. With the advent of high-throughput sequencing technologies, it is now more practical and inexpensive to obtain plastome sequences and to upgrade cp-based phylogenetics to phylogenomics.

In the present study, we sequenced the complete plastid genomes of two *Thalictrum* species by using the next-generation sequencing platform and performed the first comprehensive analysis of *Thalictrum* plastomes by combining these data with previously reported plastomes of other nine species (Park et al., 2015; He et al., 2019, 2021b; Morales-Briones et al., 2019). Our study aims were as follows: (1) to investigate global structural patterns of the 11 *Thalictrum* plastomes; (2) to identify the most variable regions of these plastomes as prospective DNA barcodes for future species identification; (3) to choose more effective molecular markers *via* reconstruction of phylogenetic relationships among the 11 *Thalictrum* species using various makers; and (4) to test for the presence of adaptive evolution in all genes located in the two single-copy regions, and one of the two inverted-repeat (IR) regions by analyses of selective pressure and codon usage bias. The results will provide abundant information for further species identification, phylogenetic, and phylogeographic studies on *Thalictrum*, and will assist in exploration and utilization of *Thalictrum* plants.

## MATERIALS AND METHODS

### Sample Preparation, Sequencing, Assembly, and Annotation

The sequenced two *Thalictrum* species (*Thalictrum minus* var. *hypoleucum* and *Thalictrum simplex*) are growing in the Beijing Botanical Garden, Beijing, China. Genomic DNA was extracted from fresh leaves and purified using the Tiangen Isolation/Extraction/Purification Kit [Tiangen Biotech (Beijing) Co., Ltd.]. Short insert of 300–500 bp libraries were prepared for sequencing on the Illumina HiSeq X-Ten platform.

Before assembly of the short reads, plastome original reads were extracted by mapping all short reads to the nine plastomes as reference with BWA (Li and Durbin, 2009) and SAMtools (Danecek et al., 2021). Then the two plastomes were *de novo* assembled with SPAdes v3.15.2 (Bankevich et al., 2012) as described in He et al. (2021a). Highly accurate annotation of organelle genomes was performed by using the Organellar Genome GeSeq tool (Tillich et al., 2017) with subsequent manual correction. Three chloroplast genomes from *Thalictrum coreanum* (GenBank accession No. NC\_026103), *Thalictrum minus* (NC\_041544), and *Thalictrum thalictroides* (NC\_039433) were used as reference sequences. The circular plastomes were visualized by using OGDRAW v1.3.1 (Greiner et al., 2019), with subsequent manual editing. We also updated the annotation of plastomes for the other 11 species in this study.

### Detection and Annotation for Plastid Genomic Variations

Multiple sequence alignments of whole plastome sequences from the 11 *Thalictrum* species that have the representatives

of the two major clades of this genus in previous studies (Soza et al., 2012, 2013), as well as *Paraquilegia anemonoides* and *Leptopyrum fumarioides* in Thalictreae as outgroups were implemented using MAFFT v7 (Kato and Toh, 2010) with standard parameters, and further adjusted manually in Geneious v8.0.4 (Kearse et al., 2012). For comparison, the gene order and structure of the 13 plastomes were compared by using IRscope.<sup>1</sup>

To identifying hypervariable regions, the sequence alignment of *Thalictrum* plastomes without outgroups was subjected to a sliding window analysis in DNAsp v6.12.03 (Rozas et al., 2017) to evaluate nucleotide diversity ( $\pi$ ) of all genes, genes without introns, and intergenic spacer (IGS) regions. Functional annotations for the nucleotide variations were conducted by using snpEff v5.1 (Cingolani, 2012).

## Phylogenetic Analysis

Phylogenetic analyses of *Thalictrum* were performed with maximum likelihood (ML) method in RAxML v8.2.11 (Stamatakis, 2014) with 1,000 replicates under GTRGAMMA model. The analyses were carried out based on the following nine data sets, including the complete plastid DNA sequences, concatenation of 115 IGS regions, concatenation of 114 gene sequences, and six genes and/or their introns and spacers (*rpl16* intron, *ndhC-trnV<sup>UAC</sup>*, *ndhA* intron, *trnL<sup>UAA</sup>-trnF<sup>GAA</sup>*, *rpl32-trnL<sup>UAG</sup>*, and *rbcL*) that have been employed in previous studies on *Thalictrum* (Soza et al., 2012, 2013; Wang et al., 2019).

## Selective Pressure Analysis

Selective pressures were detected throughout the phylogenetic tree of *Thalictrum* for each plastid gene. Nonsynonymous ( $d_N$ ) and synonymous ( $d_S$ ) substitution rates of each plastid gene were assessed by using the CODEML program in PAML v4.9 (Yang, 2007). We tested different hypotheses *via* branch models, H0: the one-ratio model (m0), assumes the same  $d_N/d_S$  ratio ( $\omega$  ratio) for all branches in the phylogeny, HA: the free-ratio model (m1) that assumes an independent  $\omega$  ratio for each branch. Likelihood ratio tests were used to test each model's fit. The double log-likelihood difference between the two models (2 $\Delta$ L) was compared to a chi-square distribution with N-1 degrees of freedom, where N is the number of branches in the phylogeny (Whelan and Goldman, 1999).

## Codon Usage Analysis

The program DNAsp v6.12.03 (Rozas et al., 2017) was used to examine the synonymous codon usage of 79 protein-coding genes in the plastome of *Thalictrum* and to calculate several related parameters such as the effective number codons (ENC), codon bias index (CBI), and relative synonymous codon usage (RSCU). The ENC and CBI are often used to evaluate codon bias at the level of an individual gene (Frank, 1990). RSCU is the observed codon frequency divided by the expected frequency. An RSCU value close to 1.0 indicates that the deviation is not significant (Sharp et al., 1986). Amino acid

(AA) frequency was calculated as the percentage of codons encoding the same amino acid divided by the total codons.

## RESULTS

### Genome Features

The 11 plastomes of the *Thalictrum* species ranged in size from 154,924 bp (*T. thalictroides*) to 156,258 bp (*T. minus* var. *hypoleucum*). All these plastomes displayed the typical quadripartite structure of nearly all land plants, consisting of a pair of inverted repeats (IRs, 26,273–26,521 bp) separated by a single-copy (LSC) region (84,733–85,700 bp) and a small single-copy (SSC) region (17,479–17,655 bp; **Table 1**). The average GC content was ~38.39%, which is almost identical with each other among the 11 complete *Thalictrum* plastomes. In the IR region, the GC content (43.22%) was found to be much higher than that in the LSC (36.62%) and SSC regions (32.45%). Although overall genomic structure including gene number and gene order were well-conserved (**Figure 1**), the 11 *Thalictrum* plastomes exhibited obvious differences in the IR-SC boundary regions (**Figure 2**). The gene *ycf1* spanned the SSC-IR<sub>B</sub> region while a pseudogene fragment  $\psi$ *ycf1* was located at the IR<sub>A</sub> region with a length range of 1,144–1,152 bp. The gene *rps19* spanned the LSC-IR<sub>A</sub> region and a pseudogene fragment  $\psi$ *rps19* (100–122 bp) was located in the IR<sub>B</sub> region of all *Thalictrum* species except *T. thalictroides*. At the junction of IR<sub>A</sub> and SSC regions in most species, the distance between  $\psi$ *ycf1* and *ndhF* ranged from 0 to 752 bp, except for that of *Thalictrum foeniculaceum* with an overlap region of 39 bp between  $\psi$ *ycf1* and *ndhF*. At the junction of IR<sub>B</sub> and LSC regions, the distances between  $\psi$ *rps19* and *trnH* ranged from 42 to 81 bp.

All the 11 plastomes each identically encoded 131 predicted functional genes and three pseudo genes, of which seven protein-coding genes, seven tRNA genes, four rRNA genes, and two pseudo genes were duplicated in the IR regions (**Figure 1**). Two introns were detected in each of two protein-coding genes (*clpP* and *ycf3*) while a single intron was detected in each of 11 protein-coding genes (*atpF*, *ndhA*, *ndhB*, *petB*, *petD*, *rpl2*, *rpl16*, *rpoC1*, *rps12*, *rps16*, and *ycf15*) and six tRNA genes (*trnA<sup>UGC</sup>*, *trnG<sup>UCC</sup>*, *trnI<sup>GAU</sup>*, *trnK<sup>UUU</sup>*, *trnL<sup>UAA</sup>*, and *trnV<sup>UAC</sup>*; **Supplementary Table S1**). Among 79 protein-coding genes, 75 contained standard AUG as the initiation codon, while three genes (*ndhD*, *rps19*, and *ycf15*) contained GUG instead, and the *rpl2* started with ACG.

### Polymorphic Variation and Hypervariable Regions

Nucleotide variations among the complete plastid genomes of the 11 *Thalictrum* species were identified to elucidate the level of sequence divergence (**Figure 3**). The aligned matrix of the 11 *Thalictrum* plastomes (159,334 bp) contained 2,957 single-nucleotide polymorphisms (SNPs) and 1,016 insertion-deletions (indels). The vast majority of SNPs from coding genes were functionally silent (synonymous), while 594 SNPs (43.8%) and six SNPs (0.4%), from altogether 79 coding genes, were missense

<sup>1</sup><https://irscope.shinyapps.io/irapp/>

**TABLE 1** | Summary of characteristics of plastome sequences used in the study.

Species	GenBank numbers	Total genome size (GC content)	LSC size (GC content)	IR size (GC content)	SSC size (GC content)	No. total gene (unique gene)	No. protein-coding gene (unique gene)	No. tRNA gene (unique gene)	No. rRNA gene (unique gene)	No. pseudo gene
<i>Thalictrum aquilegifolium</i> L.	MZ442608	156,253 (38.35%)	85,695 (36.55%)	26,480 (43.23%)	17,598 (32.41%)	134 (114)	86 (79)	37 (30)	8 (4)	3
<i>Thalictrum baicalense</i> Turcz. ex Ledeb.	MW133265	155,859 (38.39%)	85,258 (36.63%)	26,482 (43.22%)	17,637 (32.41%)	134 (114)	86 (79)	37 (30)	8 (4)	3
<i>Thalictrum coreanum</i> H. Lévl.	NC_026103	155,088 (38.44%)	84,733 (36.68%)	26,403 (43.25%)	17,549 (32.49%)	134 (114)	86 (79)	37 (30)	8 (4)	3
<i>Thalictrum foeniculaceum</i> Bunge	NC_053570	155,923 (38.34%)	85,323 (36.57%)	26,486 (43.21%)	17,628 (32.30%)	134 (114)	86 (79)	37 (30)	8 (4)	3
<i>Thalictrum foliolosum</i> DC.	MZ196217	155,764 (38.46%)	85,086 (36.71%)	26,521 (43.22%)	17,636 (32.58%)	134 (114)	86 (79)	37 (30)	8 (4)	3
<i>Thalictrum minus</i> var. <i>hypoleucum</i> (Siebold & Zucc.) Miq.	OM501079	156,258 (38.35%)	85,700 (36.55%)	26,480 (43.23%)	17,598 (32.41%)	134 (114)	86 (79)	37 (30)	8 (4)	3
<i>Thalictrum petaloideum</i> L.	MK253449	155,876 (38.42%)	85,326 (36.64%)	26,480 (43.23%)	17,590 (32.55%)	134 (114)	86 (79)	37 (30)	8 (4)	3
<i>Thalictrum simplex</i> L.	OM501080	156,211 (38.36%)	85,662 (36.56%)	26,481 (43.22%)	17,587 (32.46%)	134 (114)	86 (79)	37 (30)	8 (4)	3
<i>Thalictrum tenue</i> Franch.	MK253448	156,103 (38.37%)	85,507 (36.59%)	26,504 (43.2%)	17,588 (32.43%)	134 (114)	86 (79)	37 (30)	8 (4)	3
<i>Thalictrum thalictroides</i> (L.) A. J. Eames & B. Boivin	NC_039433	154,924 (38.43%)	84,899 (36.66%)	26,273 (43.26%)	17,479 (32.5%)	133 (114)	86 (79)	37 (30)	8 (4)	2
<i>Thalictrum viscosum</i> W. T. Wang & S. H. Wang	MZ442609	155,984 (38.38%)	85,339 (36.63%)	26,495 (43.2%)	17,655 (32.36%)	134 (114)	86 (79)	37 (30)	8 (4)	3
<i>Leptopyrum fumarioides</i> (L.) Rchb.	NC_041542	157,448 (38.41%)	84,907 (36.41%)	27,821 (43.34%)	16,899 (32.19%)	133 (113)	86 (79)	37 (30)	8 (4)	2
<i>Paraquilegia microphylla</i> (Royle) J. R. Drumm. & Hutch.	NC_041479	164,383 (38.87%)	84,925 (36.62%)	30,979 (43.77%)	17,500 (32.42%)	134 (114)	86 (79)	37 (30)	8 (4)	3

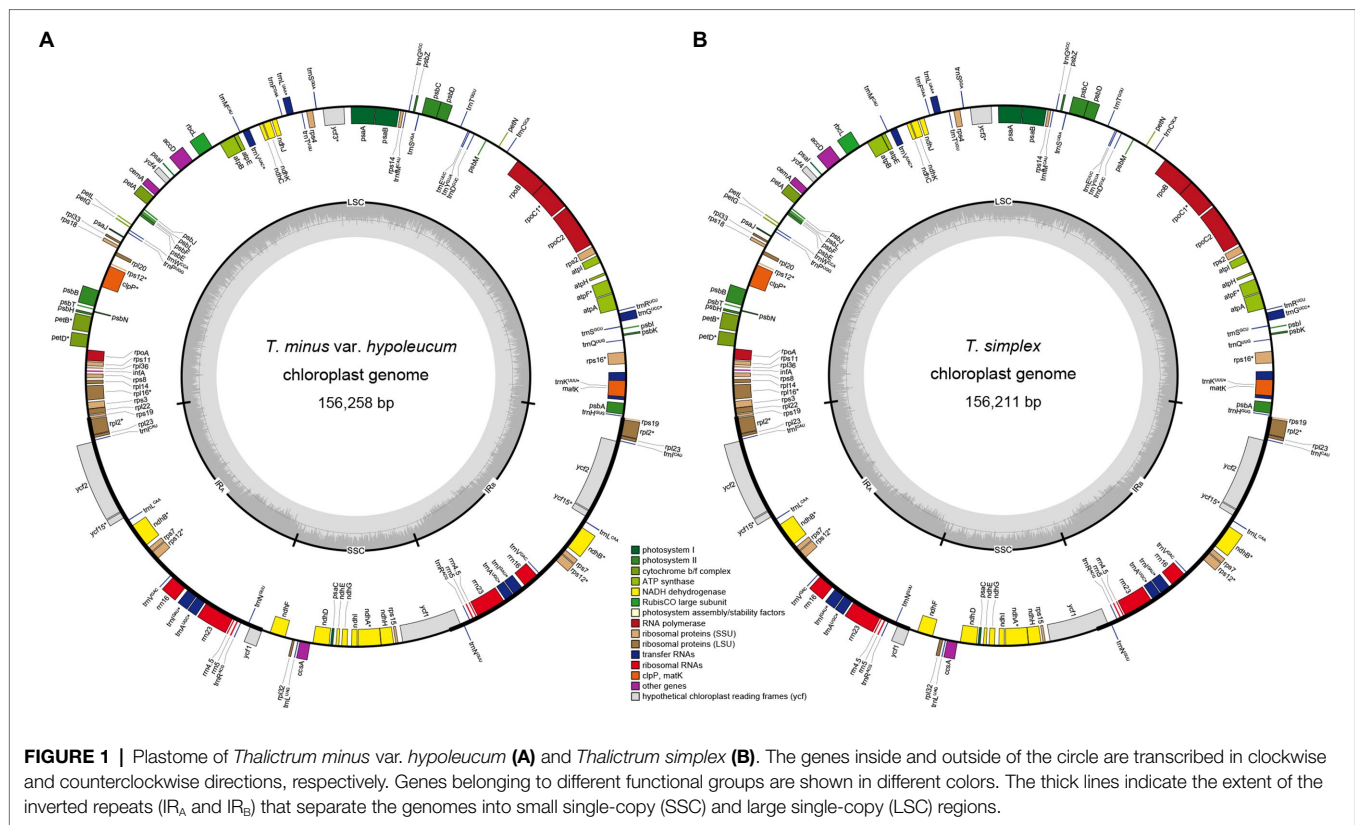
and nonsense variations (**Supplementary Table S2**). A total of 549 simple sequence repeats (SSRs) were identified in the 11 *Thalictrum* plastomes with a range of 39 (*Thalictrum petaloideum*) to 60 (*Thalictrum baicalense*) SSRs were detected in each species (**Supplementary Table S3**), indicating rich polymorphism of the SSRs among plastomes of different species. The SSC regions showed the highest nucleotide diversity ( $\pi=0.01381$ ), followed by the LSC ( $\pi=0.00803$ ) and IR ( $\pi=0.00154$ ) regions. In the 114 unique genes, the nucleotide diversity for each locus ranged from 0 (e.g., *rps7*, *rrn16*, and *trnC<sup>GCA</sup>*) to 0.02608 (*infA*) with an average of 0.00438, whereby 10 regions (i.e., *infA*, *rpl32*, *yef1*, *rpl20*, *ccsA*, *rpl22*, *rpl16*, *rps15*, *rps16*, and *accD*) had remarkably high values ( $\pi>0.0096$ ; **Supplementary Table S1**; **Figure 3A**). For exons in genes, the nucleotide diversity ranged from 0 (e.g., *rps7*, *rrn16*, and *trnA-UGC*) to 0.02608 (*infA*) with an average of 0.00373, while for the 115 IGS regions it ranged from 0 (e.g., *atpE-atpB*, *rpl23-trnI<sup>CAU</sup>*, *rrn16-trnI<sup>GAU</sup>*, and *trnI<sup>GAU</sup>-trnA<sup>UGC</sup>*) to 0.03486 (*rpoC1-rpoB*) with an average of 0.01025, except for the *rpoC1-rpoB*

( $\pi>0.07171$ ) with a targetable sequence of only 5 bp. Additionally, 10 of those regions showed considerably high values ( $\pi>0.0217$ ; i.e., *ndhF-rpl32*, *infA-rps8*, *ccsA-ndhD*, *rpl32-trnL<sup>UAG</sup>*, *trnS<sup>UGA</sup>-psbZ*, *trnH<sup>GUG</sup>-psbA*, *rpl16-rps3*, *ndhG-ndhI*, *ndhD-psaC*, and *ndhJ-ndhK*; see **Supplementary Table S4**; **Figure 3B**).

## Phylogenetic Analysis

Three datasets, the whole complete plastid genome sequences, IGS regions, and gene sequences were constructed to investigate the phylogenetic relationships among the 11 *Thalictrum* species, with *P. anemonoides* and *Leptopyrum fumarioides* as two outgroups. By using ML method, three phylogenetic trees were built based on the three respective datasets, whose topologies were found to be highly concordant between one another (**Figures 4A–C**). The *Thalictrum* was strongly supported as a monophyletic group [bootstrap support (bs)=100%], and contained two major clades that are strongly supported as sister groups: clades I (bs=100%) and II (bs=100%;





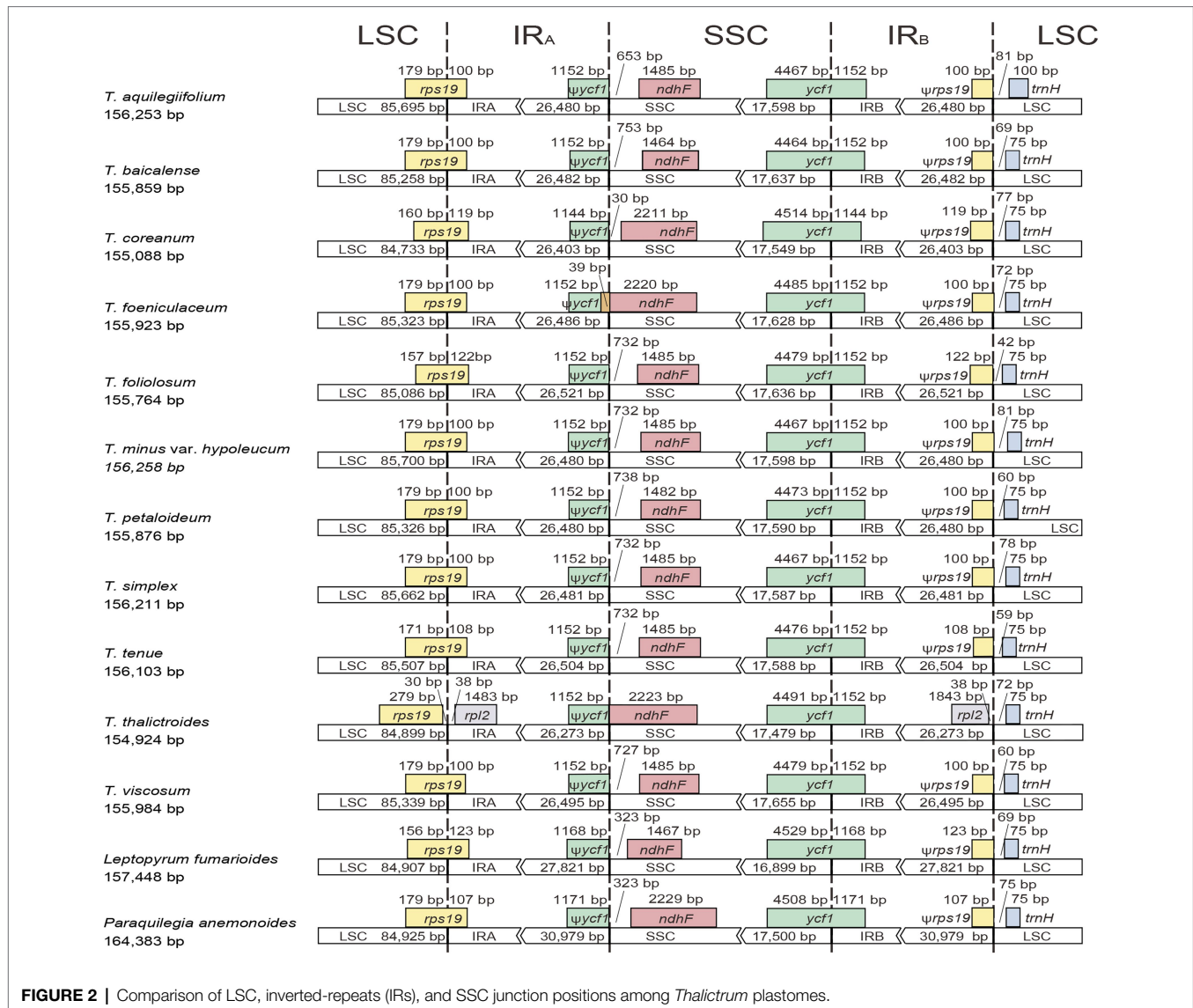
**Figures 4A–C.** The resolution of previously used six molecular fragments was also evaluated for *Thalictrum* species. Five genes and/or their introns and spacers yielded similar results except for the *rpl32-trnL*<sup>UAG</sup> (**Figures 4D–I**). However, different supporting values were observed from the nodes based on different sequence dataset. For example, two nodes in clades II derived from the dataset of gene sequences both showed weaker supports (bs=54% and bs=68%; **Figure 4C**) than those derived from complete plastid genome sequences (bs=100% and bs=95%; **Figure 4A**) and IGS regions (bs=100% and bs=95%; **Figure 4B**). Additionally, the *rpl16* intron had the strongest support within clades II (**Figure 4D**), while *rbcl* had the weakest support in them (**Figure 4I**). These results indicated a much stronger resolving power of complete plastid genome sequences as well as IGS and intron regions as compared to the exon regions, which may serve as a reliable source of phylogenetic information in *Thalictrum*.

## Selective Pressure and Codon Usage Analysis

Selective pressure analysis was conducted for CDS of all the 79 plastid protein-coding genes. A total of 66 genes are fit of m1 model in which *atpF* showed the highest  $\omega$  ratio (1.13) except for *rpl23* ( $\omega=999$ ), while other 13 genes (*psbL*, *psaC*, *rps12*, *rps19*, *petB*, *psbN*, *psbF*, *psaJ*, *psbE*, *rpl36*, *psbZ*, *petN*, and *rps7* are fit of m0 model; **Table 2**). Among the 66 genes, most (50/66) were located in LSC region following by IR (7/66)

and SSC (9/66) regions. The values of  $\omega$  are significantly different ( $p < 0.05$ ) between *Thalictrum* species for *ndhG* (SSC), *petA* (LSC), and *rpl22* (LSC) gene based on likelihood ratio tests, within some species have positive selection (e.g., *ndhG* in *T. coreanum*, *T. foeniculaceum*, *Thalictrum foliolosum*, and *T. thalictroides*; *petA* in *T. minus* var. *hypoleucum*). No genes in IR regions were detected significantly different between different species. However, 12 genes (LSC: *atpF*, *rpl33*, *rpl20*, *rps16*, *rps18*, *petG*, *rpl2*, *petL*, *psbJ*, *psbM*; IR: *rpl23*; SSC: *rps15*) were subject to positive selection in most species (median of  $\omega > 1$ ; see **Table 2**; **Supplementary Tables S6, S7**), although their values of  $\omega$  are not significantly different between different species.

We further analyzed the codon usage bias of the 79 protein coding genes in the plastomes of the 11 *Thalictrum* species. Most codons (55/64) were found to be used without bias or with only a slight bias ( $0.5 \leq \text{RSCU} \leq 1.5$ ) in the protein-coding genes (**Supplementary Table S8**). The effective number of codons (ENC) and codon bias index (CBI) of all the 79 genes varied within a wide range, e.g., from 25.02 to 61.00 and from 0.28 to 0.85, respectively, with a median value of 49.0 and 0.50, respectively (**Figure 5**; **Supplementary Figure S1**; **Supplementary Table S8**). The data indicated that these genes were probably expressed in different levels due to their different usage frequencies of the rare and optimal codons, although they are all highly conserved in the plastomes. Most genes in SSC region (80.0%) showed relatively strong bias in the codon usage



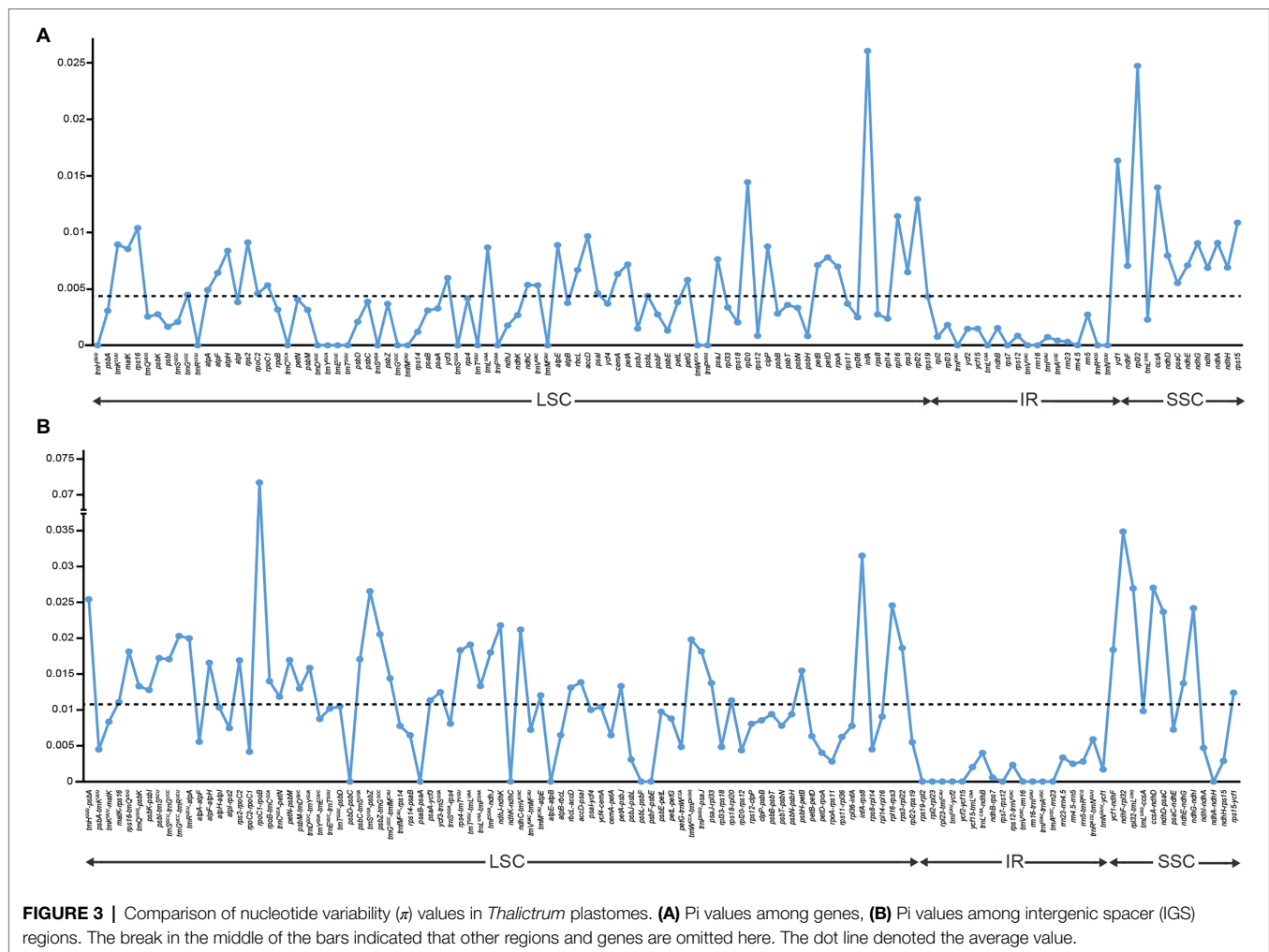
( $ENC \leq ENC_{\text{median}} = 49.0$  or  $CBI \geq CBI_{\text{median}} = 0.5$ ), while 67.2% of genes in LSC region and 50.0% of genes in IR region performed relatively strong codon usage bias. Notably, almost all genes under positively selective pressures in more than half species performed relatively strong bias in the codon usage ( $ENC \leq ENC_{\text{median}} = 49.0$  or  $CBI \geq CBI_{\text{median}} = 0.5$ ), e.g., *atpF* featured a relatively strong codon usage bias with a low ENC of 41.61. This finding suggested that those important genes with higher expression levels may played important roles in the evolution and divergence of *Thalictrum* plastomes.

## DISCUSSION

### Plastome Characteristics of *Thalictrum*

In the present study, complete plastome sequences were firstly assembled for *T. minus* var. *hypoleucum* and *T. simplex* in the

*Thalictrum* genus, with a total length of 156,211 and 156,258 bp, respectively (Table 1). The two plastomes are also highly similar in overall structure and gene order when compared to the majority of previously published plastomes of other nine species in *Thalictrum* (Park et al., 2015; He et al., 2019, 2021b; Morales-Briones et al., 2019). However, there was obvious variation in the IR-SC boundary regions among the 11 *Thalictrum* plastomes (Figure 2). The variations in IR-SC boundary regions in the 11 *Thalictrum* plastomes led to their length variation of the four regions and whole genome sequences. The expansion and contraction of the IR-SC boundary regions was considered as a primarily mechanism causing the length variation of angiosperm plastomes (Kim and Lee, 2004). In general, such expansions or contractions of the IRs into or out of adjacent single-copy regions are frequently observed in angiosperm plastomes (e.g., Yang et al., 2016; Zhang et al., 2016; Ye et al., 2018).



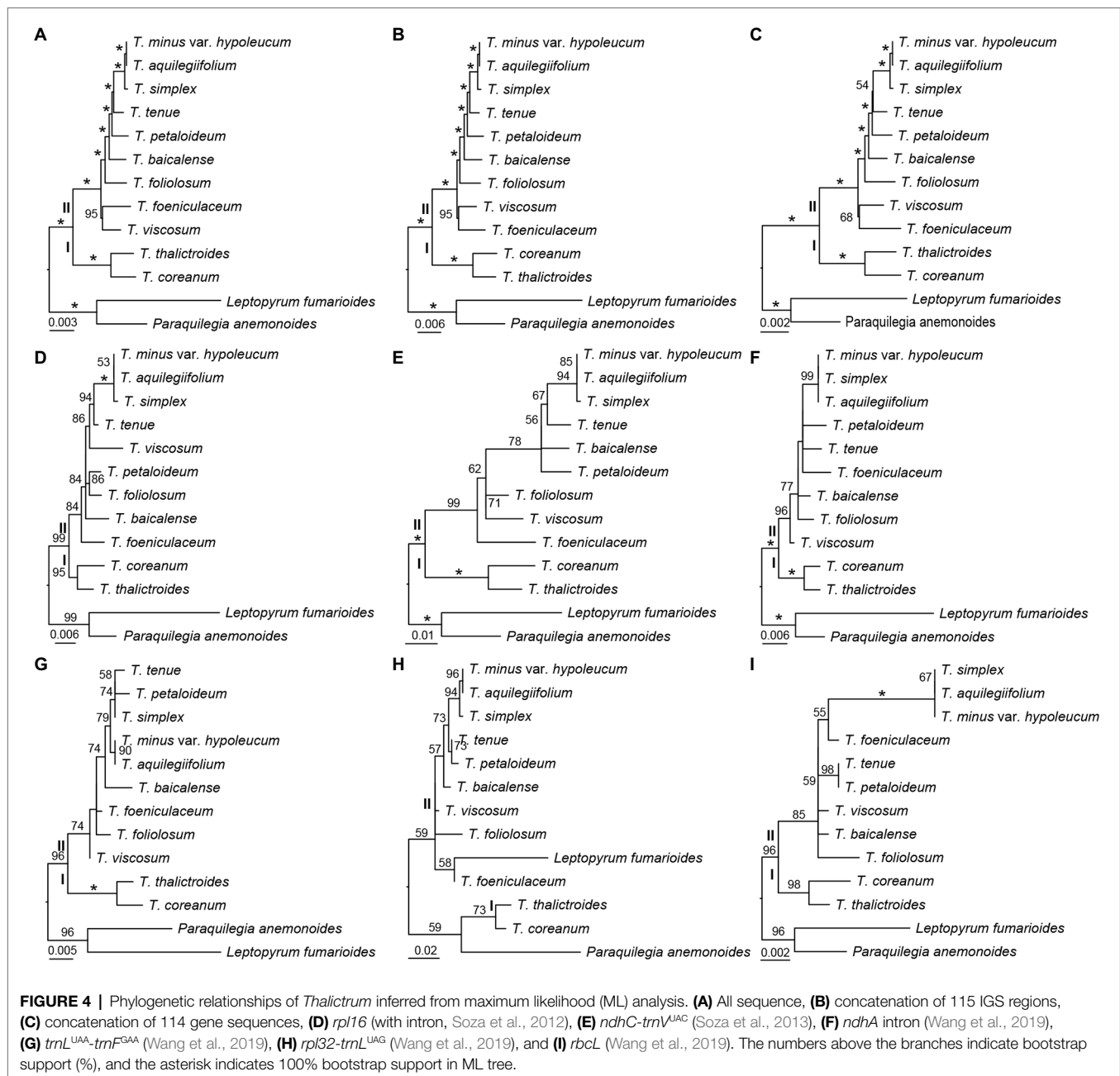
Nonetheless, there are particular genes, especially *ycf1*, *rps19*, *ndhF*, *ycf15*, and  $\psi rpl32$ , which deserve closer scrutiny. For instance, in various members of *Thalictrum*, *ycf1* is duplicated, with a shorter copy ( $\psi ycf1$ , 1,144–1,152 bp) and a larger copy (*ycf1*, 5,616–5,658 bp) located at the SSC-IR<sub>A</sub> and SSC-IR<sub>B</sub> boundaries, respectively (Figure 2). Similarly, the *rps19* is present as two copies including  $\psi rps19$  (100–122 bp) and *rps19* (279 bp) at the SSC-IR<sub>B</sub> and SSC-IR<sub>A</sub> boundaries respectively except in *T. thalictroides* (Figure 2). Both shorter copies apparently resulted from incomplete duplication. Similar pseudogenizations of *ycf1* and locations of  $\psi ycf1$  copies are known from other plants (Yang et al., 2013; Szczecińska and Sawicki, 2015; Ye et al., 2018), and two copies of *rps19* have been found in Podophylloideae (Berberidaceae; Ye et al., 2018). As for the *ndhF*, the coding sequence was unexpectedly terminated by a stop-codon-gained event caused by nucleotide variation of a poly-A region in eight *Thalictrum* species except for *T. coreanum*, *T. foeniculaceum*, and *T. thalictroides*. For the *ycf15*, an intact copy and an interrupted gene have been found in other plants, with lengths of c. 150–300 bp (Raubeson et al., 2007; Shi et al., 2013). By contrast, an interrupted *ycf15* gene has

been annotated in the sequenced chloroplast genomes in *Thalictrum* species. Additionally,  $\psi rpl32$  is incomplete because the *rpl32* gene was found to be transferred to the nucleus in the ancestor of the subfamily Thalictroideae (Park et al., 2015).

Regarding the initiation codon, *ndhD*, *rps19*, and *ycf15* used GUG, while *rpl2* used ACG in *Thalictrum*. The ACG codon may be restored to the canonical start codon (AUG) by RNA editing (Hoch et al., 1991; Takenaka et al., 2013), whereas GUG has been detected in in other plastomes (Kuroda et al., 2007; Gao et al., 2009; Zhang et al., 2016).

### Noncoding Regions as a Source of Phylogenetic Information in *Thalictrum*

Given that the nuclear-genome coded ITS region is often presented as more than one copy in *Thalictrum*, sequences of cpDNA intergenic spacers have been employed to uncover intraspecific variability in *Thalictrum* (Soza et al., 2012, 2013). The IRs usually showed lower sequence divergence than the SC regions in most of higher plants and possibly due to copy correction between IR sequences by gene



conversion (Khakhlova and Bock, 2006; Zhang et al., 2016). In the present study, the whole genome and IGS regions manifested higher sequence divergence than genes did, and genes with introns showed higher sequence divergence than genes without introns in *Thalictrum* species (Figure 3). In general, the non-coding regions (introns and spacers) had higher variability proportions than coding regions, which was also true for most higher plants (Shaw et al., 2014; Zhang et al., 2016).

In some studies, eight noncoding regions (*ndhF-rpl32*, *rpl32-trnL<sup>UAG</sup>*, *ndhC-trnV-UAC*, *rps16-trnQ<sup>UUG</sup>*, *psbE-petL*, *trnI<sup>GGU</sup>-psbD*, *petA-psbJ*, and *rpl16* intron) have been

identified as the best possible choices for low-level phylogenetic studies on angiosperms (Shaw et al., 2014). Among these regions, *ndhF-rpl32*, *rpl32-trnL<sup>UAG</sup>*, and the *rpl16* intron were also identified as highly divergent loci among *Thalictrum* species in the present study. Nonetheless, two IGS regions related to *rpl32* are not suitable as molecular markers in *Thalictrum* because the *rpl32* gene is often transferred to the nucleus (Park et al., 2015). Aside from these loci, we also observed high nucleotide diversity in *infA-rps8*, *ccsA-ndhD*, *trnS<sup>UGA</sup>-psbZ*, *trnH<sup>GUG</sup>-psbA*, *rpl16-rps3*, *ndhG-ndhI*, and *ndhD-psaC* regions. Additionally, an intron of *rps16* also showed highly variable here, similarly to



**TABLE 2** | Summary of models H0 and HA analyzed in the study.  $d_N$ ,  $d_S$ , and  $\omega$  are presented as medians.

Gene	H0: m0				HA: m1				2*(HA-H0)	p-value
	$d_N$	$d_S$	$\omega$	lnL	$d_N$	$d_S$	$\omega$	lnL		
<i>accD</i>	0.0005	0.0019	0.2489	-2437.30	0.000832	0.00428	0.193579	-2433.08	8.44	0.75
<i>atpA</i>	0.0001	0.0021	0.0705	-2272.75	0	0.002509	0.0001	-2266.30	12.90	0.38
<i>atpB</i>	0.0002	0.0022	0.0699	-2221.92	0	0.002701	0.0001	-2213.35	17.14	0.14
<i>atpE</i>	0	0	0.1804	-642.17	0.000001	0.000005	0.0372454	-635.33	13.69	0.32
<i>atpF</i>	0	0	1.1265	-866.78	0.000002	0.000002	124.242	-862.26	9.02	0.70
<i>atpH</i>	0	0	0.0001	-366.55	0.000002	0.000005	0.0001	-366.55	0.000258	1.00
<i>atpI</i>	0.0004	0.0043	0.0857	-1137.90	0.000001	0.004595	0.0001	-1130.12	15.55	0.21
<i>ccsA</i>	0.0022	0.0071	0.3153	-1836.83	0.002604	0.010913	0.169018	-1831.85	9.97	0.62
<i>cemA</i>	0.001	0.0034	0.2884	-1096.08	0.000001	0.000006	0.0001	-1086.23	19.71	0.07
<i>clpP</i>	0	0	0.0694	-935.05	0.000001	0.000004	0.0001	-929.55	11.01	0.53
<i>infA</i>	0.0096	0.0123	0.7783	-269.79	0.013353	0.000013	74.8743	-264.81	9.97	0.62
<i>matK</i>	0.0027	0.0058	0.4631	-2572.82	0.002523	0.006385	0.911524	-2562.74	20.15	0.06
<i>ndhA</i>	0.0008	0.0053	0.1467	-1710.61	0.000002	0.003808	0.329949	-1704.63	11.97	0.45
<i>ndhB</i>	0	0	0.1178	-2133.67	0	0.000006	0.0001	-2129.17	8.99	0.70
<i>ndhC</i>	0	0	0.0916	-566.08	0	0.000005	0.0001	-562.34	7.48	0.82
<i>ndhD</i>	0.0011	0.0108	0.0998	-2601.20	0.00089	0.011321	0.0627638	-2593.10	16.20	0.18
<i>ndhE</i>	0	0	0.0235	-456.81	0	0.000006	0.0001	-454.86	3.92	0.98
<i>ndhF</i>	0.0009	0.0071	0.1292	-2463.23	0.001765	0.007192	0.147131	-2455.95	14.55	0.27
<i>ndhG</i>	0.001	0.005	0.1954	-909.26	0.000003	0.000074	0.0001	-896.00	26.52	0.01
<i>ndhH</i>	0.0005	0.006	0.0866	-1923.38	0.001108	0.003891	0.0698539	-1919.64	7.48	0.82
<i>ndhI</i>	0.0007	0.0061	0.113	-908.57	0.000001	0.008613	0.0001	-902.89	11.35	0.50
<i>ndhJ</i>	0	0	0.3505	-688.53	0	0.000005	0.0001	-686.35	4.38	0.98
<i>ndhK</i>	0	0	0.3155	-1048.40	0	0.000006	0.0001	-1041.06	14.69	0.26
<i>petA</i>	0.0004	0.003	0.1455	-1568.16	0.000002	0.004392	0.0001	-1555.79	24.74	0.02
<i>petD</i>	0	0	0.0307	-741.85	0.000001	0.000005	0.0102435	-739.71	4.29	0.98
<i>petG</i>	0	0	0.8509	-156.93	0.000002	0	999	-155.55	2.77	1.00
<i>petL</i>	0	0	0.1008	-136.46	0.000002	0	313.224	-135.59	1.74	1.00
<i>psaA</i>	0	0.0018	0.0084	-3309.32	0	0.001795	0.0001	-3306.01	6.62	0.88
<i>psaB</i>	0.0002	0.0031	0.0672	-3293.87	0	0.003743	0.0001	-3286.29	15.16	0.23
<i>psaI</i>	0	0	0.8811	-177.10	0	0.000005	0.106004	-174.86	4.48	0.97
<i>psbA</i>	0.0001	0.0037	0.0157	-1554.95	0	0.003857	0.0001	-1552.80	4.31	0.98
<i>psbB</i>	0.0002	0.0021	0.0878	-2235.59	0	0.002695	0.0001	-2229.89	11.39	0.50
<i>psbC</i>	0.0001	0.0024	0.0259	-2138.08	0	0.002509	0.0001	-2131.91	12.33	0.42
<i>psbD</i>	0	0	0.075	-1533.73	0	0.000005	0.0001	-1529.39	8.68	0.73
<i>psbH</i>	0	0	0.12	-321.61	0.000002	0.000006	0.0978274	-319.64	3.93	0.98
<i>psbI</i>	0	0	0.0001	-152.19	0	0.000003	0.0001	-152.19	0.005686	1
<i>psbJ</i>	0	0	0.1454	-163.03	0.000002	0	113.399	-162.51	1.05	1.00
<i>psbK</i>	0	0	0.3253	-268.48	0	0.000005	0.0001	-265.68	5.59	0.94
<i>psbM</i>	0	0	0.0001	-126.66	0.000002	0	212.117	-126.66	0.000884	1.00
<i>psbT</i>	0	0	0.245	-142.20	0	0.000004	0.0001	-141.67	1.05	1.00
<i>rbcL</i>	0.0004	0.0018	0.1906	-2281.62	0	0.002963	0.0001	-2274.83	13.59	0.33
<i>rpl14</i>	0	0	0.037	-529.92	0.000001	0.000006	0.0001	-526.77	6.30	0.90
<i>rpl16</i>	0	0	0.1546	-631.29	0.000001	0.000005	0.0001	-624.98	12.62	0.40
<i>rpl2</i>	0	0	0.183	-1139.62	0.000002	0	31.3706	-1138.57	2.10	1.00
<i>rpl20</i>	0	0	0.4022	-628.11	0.000002	0.000004	9.55435	-623.99	8.24	0.77
<i>rpl22</i>	0	0	0.4957	-931.12	0	0.000007	0.0001	-919.97	22.31	0.03
<i>rpl23</i>	0	0	999	-396.07	0.000002	0.000001	55.6975	-396.07	0.00	1.00
<i>rpl33</i>	0	0	0.1364	-307.30	0.000002	0	491.843	-303.04	8.51	0.74
<i>rpoA</i>	0.001	0.006	0.1692	-1751.50	0.000001	0.010121	0.0001	-1742.58	17.85	0.12
<i>rpoB</i>	0.0003	0.0029	0.1019	-4816.32	0.000418	0.001825	0.0461232	-4807.91	16.82	0.16
<i>rpoC1</i>	0.0007	0.004	0.1796	-3224.37	0.000638	0.002151	0.0743791	-3216.52	15.71	0.20
<i>rpoC2</i>	0.0012	0.0042	0.3139	-6585.10	0.001256	0.004513	0.215304	-6576.06	18.07	0.11
<i>rps11</i>	0	0	0.0797	-634.65	0	0.000005	0.0001	-631.31	6.68	0.88
<i>rps14</i>	0	0	0.1102	-420.73	0	0.000005	0.0001	-418.47	4.51	0.97
<i>rps15</i>	0	0	0.6425	-434.79	0.000002	0	76.3582	-430.76	8.08	0.78
<i>rps16</i>	0	0	0.4244	-380.25	0.000002	0	15.7794	-377.71	5.07	0.96
<i>rps18</i>	0	0	0.0918	-411.02	0.000002	0	192.024	-408.76	4.52	0.97
<i>rps2</i>	0	0	0.4331	-1203.16	0.000002	0.000006	0.134231	-1198.26	9.80	0.63
<i>rps3</i>	0.0005	0.005	0.1081	-1071.79	0.000001	0.005949	0.0001	-1064.71	14.16	0.29
<i>rps4</i>	0.0007	0.0051	0.1334	-944.70	0.000001	0.000006	0.0001	-937.89	13.61	0.33

(Continued)



Our phylogenetic trees based on whole complete plastid genome sequences, 116 IGS regions, and 114 gene sequences revealed that *Thalictrum* contains two major clades that is consistent with previous studies (Figures 4A–C; Soza et al., 2012, 2013; Morales-Briones et al., 2019; Wang et al., 2019). However, the relationships along the backbone of the clades are not well-supported in their studies. None of the sections traditionally circumscribed for this genus (Tamura, 1995) is monophyletic. It is necessary to apply more samplings and find more efficient molecular markers for *Thalictrum*.

Our phylogenetic trees indicated that 116 IGS regions had stronger support than 114 gene sequences (Figures 4B,C). Additionally, the *rpl16* intron—that was used by Soza et al. (2012) with high sequence divergence in the studies—showed also strong support in clades II here (Figure 4D). While the coding regions of *rbcL* employed by Wang et al. (2019) showed lower supports within clades II in our analysis (Figure 4I). The non-coding regions (introns and spacers) are more variable molecular markers. For the ML tree of *rpl32-trnL<sup>UAG</sup>* used by Wang et al. (2019), the outgroups are embedded in *Thalictrum* probably because the matrix of *rpl32-trnL<sup>UAG</sup>* contained lots of indels (Figure 4H). The *rpl32* gene is often transfers to the nucleus (Park et al., 2015) that make the *ndhF-rpl32*, *rpl32-trnL<sup>UAG</sup>*, and *rpl32* regions not reliable to be markers for phylogeny in *Thalictrum*.

## Positive Selection in Different Genes

It is believed that selection is the most probable components of the evolutionary forces acting on most highly expressed genes, although all genes are basically subjected to a certain degree of natural selection (Gouy and Gautier, 1982; Sueoka, 1999; Sharp et al., 2010). And the degeneracy of genetic code leads to the expression of variation contained in a gene through its manifestation in protein, which varied among different species (Edelman and Gally, 2001; Wan et al., 2004; Chakraborty et al., 2020). In the present study, we observed different codon usage frequency on different genes under positive pressure. For example, 12 plastid genes (*atpF*, *rpl33*, *rps15*, *rpl20*, *rps16*, *rps18*, *petG*, *rpl2*, *petL*, *psbJ*, *psbM*, and *rpl23*) were observed under positive selective pressure in most of the 11 *Thalictrum* species among which 11 showed relatively higher CBI values (>0.5) suggesting high expression level *in vivo*; while three plastid genes that are relative with NADH oxidoreductase (*ndhG*), cytochrome b6/f complex (*petA*), and ribosomal proteins (*rpl22*) were observed under significantly strong positive selective pressure ( $p < 0.05$  based on likelihood ratio tests) in only 1–4 *Thalictrum* species, showing relatively lower CBI (<0.5). The former and latter genes performed different codon usage bias suggesting different expression levels due to different usage frequency of the rare and optimal codons, which could further affected the functional patterns of those genes during their evolution process. Additionally, it also indicated potential functional divergence among plastid genomes of different *Thalictrum* species, according to abundant differences observed

between selective pressures and usage codon frequencies for different plastid genes in these species.

## CONCLUSION

This is the first report to describe a comprehensive landscape of plastomic variations among *Thalictrum* species on the basis of 11 complete plastomes. Comparison between these plastomes uncovered not only high similarities in overall structure, gene order, and content but also some structural variations caused by the expansion or contraction of the IR regions into or out of adjacent single-copy regions. DNA sequence divergence across 11 *Thalictrum* plastomes revealed that *infA-rps8*, *ccsA-ndhD*, *trn<sup>SUGA</sup>-psbZ*, *trnH<sup>GUG</sup>-psbA*, *rpl16-rps3*, *ndhG-ndhI*, *ndhD-psaC*, and *ndhJ-ndhK* are among the fastest-evolving loci and are promising molecular markers. Therefore, these highly variable loci should be valuable for future phylogenetic and phylogeographic studies on *Thalictrum*. Our phylogenomic analyses based on whole complete plastid genome sequences, 116 IGS regions and 114 gene sequences were all supported the monophyly of *Thalictrum* and two major clades within this genus. Furthermore, among 79 plastome-derived protein-coding genes (CDSs), 15 genes were identified as fast evolving genes, which were all proved to be under positive selection but showed different bias in their codon usage frequencies. Overall, our results demonstrate the ability of plastid phylogenomics to improve phylogenetic resolution, and will expand the understanding of plastid gene evolution in *Thalictrum*.

## DATA AVAILABILITY STATEMENT

All raw sequencing reads generated in the study have been deposited in NCBI under the BioProject accession PRJNA817687. The complete sequences and annotations of plastomes have also been deposited at GenBank under the accessions OM501079 and OM501080. The updated annotations of plastomes for the other 11 species in this study have been deposited to the Figshare online database (<https://doi.org/10.6084/m9.figshare.19108097.v1>).

## AUTHOR CONTRIBUTIONS

W-CH and Z-QW conceived the research. H-WP carried out taxon sampling and generated all the data. K-LX and W-CH performed the data analyses. K-LX, W-CH, and WM wrote the manuscript with help from Z-QW. Y-XY revised the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the China Postdoctoral Science Foundation (grant number 2021M703540), the Training of

Excellent Science and Technology Innovation talents in Shenzhen-Basic Research on Outstanding Youth (grant number RCYX20200714114538196), the National Natural Science Foundation of China (grant number 32011530072), the Initial fund of Shenzhen Agricultural Genome Research Institute, Chinese Academy of Agricultural Sciences (grant number SJXW19073), and the Russian Science Foundation [grant number 19-74-10082 (preparation of material)], within state assignments for CSBG SB RAS [grant number AAAA-A21-121011290024-5 (study of herbarium collections)].

## REFERENCES

Baldwin, B. G., Sanderson, M. J., Porter, J. M., Wojciechowski, M. F., Campbell, C. S., and Donoghue, M. J. (1995). The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Ann. Missouri Bot. Gard.* 82, 247–277. doi: 10.2307/2399880

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Chakraborty, S., Yengkhom, S., and Uddin, A. (2020). Analysis of codon usage bias of chloroplast genes in *Oryza* species. *Planta* 252:67. doi: 10.1007/s00425-020-03470-7

Cingolani, P. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain *w<sup>1118</sup>*; iso-2; iso-3. *Flying* 6, 80–92. doi: 10.4161/fly.19695

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10, 1–4. doi: 10.1093/gigascience/giab008

Edelman, G. M., and Gally, J. A. (2001). Degeneracy and complexity in biological systems. *Proc. Natl. Acad. Sci. U. S. A.* 98, 13763–13768. doi: 10.1073/pnas.231499798

Frank, W. (1990). The 'effective number of codons' used in a gene. *Gene* 87, 23–29. doi: 10.1016/0378-1119(90)90491-9

Gao, L., Yi, X., Yang, Y. X., Su, Y. J., and Wang, T. (2009). Complete chloroplast genome sequence of a tree fern *Alsophila spinulosa*: insights into evolutionary changes in fern chloroplast genomes. *BMC Evol. Biol.* 9:130. doi: 10.1186/1471-2148-9-130

Gouy, M., and Gautier, C. (1982). Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res.* 10, 7055–7074. doi: 10.1093/nar/10.22.7055

Greiner, S., Lehwark, P., and Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 47, W59–W64. doi: 10.1093/nar/gkz238

He, W., Chen, C., Xiang, K., Wang, J., Zheng, P., Tembrock, L. R., et al. (2021a). The history and diversity of rice domestication as resolved from 1464 complete plastid genomes. *Front. Plant Sci.* 12:781793. doi: 10.3389/fpls.2021.781793

He, Y., Wang, R., Gai, X., Lin, P., and Wang, J. (2021b). The complete chloroplast genome of *Thalictrum baicalense* turcz. ex ledeb. *Mitochondrial DNA B Resour.* 6, 437–438. doi: 10.1080/23802359.2020.1870896

He, J., Yao, M., Lyu, R. D., Lin, L. L., Liu, H. J., Pei, L. Y., et al. (2019). Structural variation of the complete chloroplast genome and plastid phylogenomics of the genus *Asteropyrum* (Ranunculaceae). *Sci. Rep.* 9:15285. doi: 10.1038/s41598-019-51601-2

Hoch, B., Maier, R. M., Appel, K., Igloi, G. L., and Kössel, H. (1991). Editing of a chloroplast mRNA by creation of an initiation codon. *Nature* 353, 178–180. doi: 10.1038/353178a0

Huang, H., Shi, C., Liu, Y., Mao, S. Y., and Gao, L. Z. (2014). Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. *BMC Evol. Biol.* 14:151. doi: 10.1186/1471-2148-14-151

Hughes, C. E., Eastwood, R. J., and Bailey, C. D. (2006). From famine to feast? Selecting nuclear DNA sequence loci for plant species-level phylogeny

## ACKNOWLEDGMENTS

The authors are grateful for Jian-Fei Ye for kind advices on identification of the species used in the present study.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.897843/full#supplementary-material>

reconstruction. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 361, 211–225. doi: 10.1098/rstb.2005.1735

Jansen, R. K., Cai, Z., Raubeson, L. A., Daniell, H., Leebens-Mack, J., Müller, K. F., et al. (2007). Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19369–19374. doi: 10.1073/pnas.0709121104

Katoh, K., and Toh, H. (2010). Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics* 26, 1899–1900. doi: 10.1093/bioinformatics/btq224

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Khakhlova, O., and Bock, R. (2006). Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant J.* 46, 85–94. doi: 10.1111/j.1365-3113X.2006.02673.x

Kim, K. J., and Lee, H. L. (2004). Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res.* 11, 247–261. doi: 10.1093/dnares/11.4.247

Kim, K., Lee, S.-C., Lee, J., Yu, Y., Yang, K., Choi, B.-S., et al. (2015). Complete chloroplast and ribosomal sequences for 30 accessions elucidate evolution of *Oryza* AA genome species. *Sci. Rep.* 5:15655. doi: 10.1038/srep15655

Kuroda, H., Suzuki, H., Kusumegi, T., Hirose, T., Yukawa, Y., and Sugiura, M. (2007). Translation of *psbC* mRNAs starts from the downstream GUG, not the upstream AUG, and requires the extended Shine–Dalgarno sequence in tobacco chloroplasts. *Plant Cell Physiol.* 48, 1374–1378. doi: 10.1093/pcp/pcm097

Langlet, O. F. I. (1927). Beitrage zur zytologie der Ranunculazeen. *Sven. Bot. Tidskr.* 21, 1–17.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H. T., Yi, T. S., Gao, M. L., Ma, P. F., Zhang, T., Yang, J. B., et al. (2019). Origin of angiosperms and the puzzle of the Jurassic gap. *Nat. Plants.* 5, 461–470. doi: 10.1038/s41477-019-0421-0

Löve, Á. (1982). IOPB chromosome number reports LXXIV. *Taxon* 31, 119–128. doi: 10.1002/j.1996-8175.1982.tb02346.x

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

Mort, M. E., Archibald, J. K., Randle, C. P., Levsen, N. D., O'Leary, T. R., Topalov, K., et al. (2007). Inferring phylogeny at low taxonomic levels: utility of rapidly evolving cpDNA and nuclear ITS loci. *Am. J. Bot.* 94, 173–183. doi: 10.3732/ajb.94.2.173

Park, S., Jansen, R. K., and Park, S. (2015). Complete plastome sequence of *Thalictrum coreanum* (Ranunculaceae) and transfer of the *rpl32* gene to the nucleus in the ancestor of the subfamily Thalictrioideae. *BMC Plant Biol.* 15:40. doi: 10.1186/s12870-015-0432-6

Raubeson, L. A., Peery, R., Chumley, T., Dziubek, C., Fourcade, H. M., Boore, J. L., et al. (2007). Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics* 8, 174–200. doi: 10.1186/1471-2164-8-174

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Librado, P. G., Ramos-Onsins, S. E., and Sánchez-Gracia, A. (2017). DnaSP v6: DNA



- sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34, 3299–3302. doi: 10.1093/molbev/msx248
- Sharp, P. M., Emery, L. R., and Zeng, K. (2010). Forces that influence the evolution of codon bias. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 365, 1203–1212. doi: 10.1098/rstb.2009.0305
- Sharp, P. M., Tuohy, T. M., and Mosurski, K. R. (1986). Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res.* 14, 5125–5143. doi: 10.1093/nar/14.13.5125
- Shaw, J., Shafer, H. L., Leonard, O. R., Kovach, M. J., Schorr, M., and Morris, A. B. (2014). Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: the tortoise and the hare IV. *Am. J. Bot.* 101, 1987–2004. doi: 10.3732/ajb.1400398
- Shi, C., Liu, Y., Huang, H., Xia, E. H., Zhang, H. B., and Gao, L. Z. (2013). Contradiction between plastid gene transcription and function due to complex posttranscriptional splicing: an exemplary study of *ycf15* function and evolution in angiosperms. *PLoS One* 8:e59620. doi: 10.1371/journal.pone.0059620
- Soza, V. L., Brunet, J., Liston, A., Smith, P. S., and Di Stilio, V. S. (2012). Phylogenetic insights into the correlates of dioecy in meadow-rues (*Thalictrum*, Ranunculaceae). *Mol. Phylogenet. Evol.* 63, 180–192. doi: 10.1016/j.ympev.2012.01.009
- Soza, V. L., Haworth, K. L., and Di Stilio, V. S. (2013). Timing and consequences of recurrent polyploidy in meadow-rues (*Thalictrum*, Ranunculaceae). *Mol. Biol. Evol.* 30, 1940–1954. doi: 10.1093/molbev/mst101
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Sueoka, N. (1999). Two aspects of DNA base composition: G + C content and translation-coupled deviation from intra-strand rule of A = T and G = C. *J. Mol. Evol.* 49, 49–62. doi: 10.1007/PL00006534
- Szczecińska, M., and Sawicki, J. (2015). Genomic resources of three *Pulsatilla* species reveal evolutionary hotspots, species-specific sites and variable plastid structure in the family Ranunculaceae. *Int. J. Mol. Sci.* 16, 22258–22279. doi: 10.3390/ijms160922258
- Takenaka, M., Zehrmann, A., Verbitskiy, D., Härtel, B., and Brennicke, A. (2013). RNA editing in plants and its evolution. *Annu. Rev. Genet.* 47, 335–352. doi: 10.1146/annurev-genet-111212-133519
- Tamura, M. (1995). “Ranunculaceae,” in *Die Natürlichen Pflanzenfamilien. Vol. 17a*. ed. P. Hiepko (Germany: Duncker & Humblot, Berlin), 223–497.
- Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., et al. (2017). GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45, W6–W11. doi: 10.1093/nar/gkx391
- Wan, X. F., Xu, D., Kleinhofs, A., and Zhou, J. (2004). Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes. *BMC Evol. Biol.* 4:19. doi: 10.1186/1471-2148-4-19
- Wang, T. N., Clifford, M. R., Martínez-Gómez, J., Johnson, J. C., Riffell, J. A., and Di Stilio, V. S. (2019). Scent matters: differential contribution of scent to insect response in flowers with insect vs. wind pollination traits. *Ann. Bot.* 123, 289–301. doi: 10.1093/aob/mcy131
- Wang, W. C., Fu, D. Z., Li, L. Q., Bartholomew, B., Brach, A. R., Dutton, B. E., et al. (2001). “Ranunculaceae,” in *Flora of China. Vol. 6*. eds. Z. Y. Wu, P. H. Raven and D. Y. Hong (Beijing: Science Press), 133–438.
- Wang, W. C., and Xiao, P. G. (1979). “Ranunculaceae,” in *Flora Reipublicae Popularis Sinicae. Vol. 27*. (Beijing: Science Press), 502–592.
- Whelan, S., and Goldman, N. (1999). Distributions of statistics used for the comparison of models of sequence evolution in phylogenetics. *Mol. Biol. Evol.* 16, 1292–1299. doi: 10.1093/oxfordjournals.molbev.a026219
- Wu, Z. Y., Zhou, T. Y., and Xiao, P. G. (1998). *Xin Hua Compendium of Materia Medica*. Shanghai: Science and Technology Press, 1, 133–139.
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088
- Yang, J. B., Yang, S. X., Li, H. T., Yang, J., and Li, D. Z. (2013). Comparative chloroplast genomes of *Camellia* species. *PLoS One* 8:e73053. doi: 10.1371/journal.pone.0073053
- Yang, Y., Zhou, T., Duan, D., Yang, J., Feng, L., and Zhao, G. (2016). Comparative analysis of the complete chloroplast genomes of five *Quercus* species. *Front. Plant Sci.* 7:959. doi: 10.3389/fpls.2016.00959
- Ye, W., Yap, Z., Li, P., Comes, H. P., and Qiu, Y. X. (2018). Plastome organization, genome-based phylogeny and evolution of plastid genes in Podophylloideae (Berberidaceae). *Mol. Phylogenet. Evol.* 127, 978–987. doi: 10.1016/j.ympev.2018.07.001
- Zhang, Y., Du, L., Liu, A., Chen, J., Wu, L., Hu, W., et al. (2016). The complete chloroplast genome sequences of five *Epimedium* species: lights into phylogenetic and taxonomic analyses. *Front. Plant Sci.* 7:306. doi: 10.3389/fpls.2016.00306
- Zhu, M., and Xiao, P. G. (1989). Study on resource utilization of germander (*Thalictrum*). *Chin. Trad. Herb Drugs* 20, 29–31.
- Zhu, M., and Xiao, P. G. (1991). Chemosystematic studies on *Thalictrum* L. in China. *Acta Phytotaxon. Sin.* 29, 358–369.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Xiang, Mao, Peng, Erst, Yang, He and Wu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.