

Research Article

Phylogenomic and comparative analyses of *Rheum*
(Polygonaceae, Polygonoideae)

Hua-Jie Zhang^{1,2†} , Xu Zhang^{1,2,3†} , Jacob B. Landis^{4,5} , Yan-Xia Sun^{1,2} , Jiao Sun^{1,2,3} , Tian-Hui Kuang⁶ ,
Li-Juan Li^{1,2,3} , Bashir B. Tihamy^{1,2,3} , Tao Deng⁶ , Hang Sun^{6*} , and Heng-Chang Wang^{1,2*} 

¹CAS Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China

²Center of Conservation Biology, Core Botanical Gardens, Chinese Academy of Sciences, Wuhan 430074, China

³University of Chinese Academy of Sciences, Beijing 100049, China

⁴School of Integrative Plant Science, Section of Plant Biology and the L.H. Bailey Hortorium, Cornell University, Ithaca, NY 14850, USA

⁵BTI Computational Biology Center, Boyce Thompson Institute, Ithaca, NY 14853, USA

⁶Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China

[†]These authors contributed equally to this study.

*Authors for correspondence. Hang Sun. E-mail: hsun@mail.kib.ac.cn; Heng-Chang Wang. E-mail: hcwang@wbgcas.cn

Received 7 September 2021; Accepted 4 November 2021; Article first published online 06 November 2021

Abstract Species of *Rheum* have high medicinal value, with the center of diversity in the Qinghai–Tibet Plateau (QTP) and adjacent regions. However, phylogenetic relationships of *Rheum* are still unclear due to fragment markers providing insufficient informative loci. Here, we sequenced and annotated plastomes of nine *Rheum* species, and compared the genome structure among the novel nine species along with three published species. Comparative analyses revealed that plastomes of *Rheum* share a relatively conserved structure. Five highly divergent regions (*accD*, *ccsA*, *matK*, *ndhF*, and *ndhH*) can be used as valuable molecular markers for further species delimitation and population genetic studies. Twenty-two accessions representing 17 species were used for phylogenetic analysis, which generated a robust phylogenetic tree and revealed two major clades within *Rheum*. Phylogenetic results showed that glasshouse structures and cushions of *Rheum* are results of parallel evolution during adaptation to similar environments. Inconsistent tree topology between concatenated and coalescent methods was detected, implying that incomplete lineage sorting and hybridization may have occurred in the evolutionary history of *Rheum*. Divergence time estimation based on two fossil calibrations and three secondary calibrations revealed a Miocene to middle Oligocene origin of *Rheum*. Our study provides valuable genomic resources for the medicinally important genus *Rheum*, while gaining helpful insights into its systematics and evolution.

Key words: incomplete lineage sorting and hybridization, molecular marker, phylogenomic, plastomes, *Rheum*.

1 Introduction

Rheum (Polygonoideae, Polygonaceae), comprising approximately 60 species, has a wide distribution range throughout temperate and subtropical areas of the Asian interior (Chin & Youngken, 1947). The Qinghai–Tibetan Plateau (QTP) and adjacent regions are considered to be the center of origin and diversity, with ca. 40 species endemic to this region (Bao & Grabovskaya-Borodina, 2003). Nine sections with high morphological differentiation were proposed by Losina-Losinskaya (1936), and eight sections were acknowledged in the present Flora of China (Bao & Grabovskaya-Borodina, 2003). Species of Sect. *Palmata* A. Los including *R. officinale*, *R. palmatum*, and *R. tanguticum* are used as medicinal plants in China. The rhubarb of these three species,

known as the “lord” or “king of herbs,” has been used for over 2000 years due to its purgative function, and thus is widely cultivated in the southwest region of China (Chinese Pharmacopoeia Committee, 2015). The molecular phylogeny of *Rheum* was first constructed based on the chloroplast DNA *trnL-F* region with 26 species, which revealed that *Rheum* may have experienced a rapid radiation combined with introgressive hybridization and reticulate evolution (Wang et al., 2005). Sun et al. (2012) used eight chloroplast DNA fragments of 34 species to reconstruct the phylogenetic relationships. Although several multi-species clades of *Rheum* were well-supported, relationships of more than half of the species remained unclear (Sun et al., 2012). Moreover, Wan et al. (2014) revealed that *Rheum* underwent a rapid radiation

and extensive hybridization using ITS and cpDNA. These fragment molecular markers cannot fully resolve the phylogeny of the genus due to the lack of phylogenetically informative sites (Wang et al., 2005; Sun et al., 2012). Zhou et al. (2020) first sequenced and annotated eight plastomes of *Rheum* and suggested complete plastomes can improve phylogenetic resolution of *Rheum* and provide high branch support. For phylogenetic reconstruction, more species of *Rheum* possessing highly diversified morphological features need to be collected, especially the species distributed in QTP and adjacent regions, allowing for the exploration into the origins and evolution of *Rheum*.

The chloroplast is proposed to originate from cyanobacteria according to the endosymbiosis theory (Raven & Allen, 2003), often possessing a circular genome consisting of 110–130 unique genes (Palmer, 1985). The chloroplast genome is usually maternally inherited in most angiosperms and consists of two copies of inverted repeat (IR) regions divided by one small single-copy region (SSC) and one large single copy region (LSC) (Bendich, 2004). With advancements in second generation sequencing technology, chloroplast genomes have been frequently used to address phylogenetic relationships among closely related species (Ma et al., 2020; Mu et al., 2020; Ye et al., 2020). The chloroplast genome contains sequence variation that can offer sufficient phylogenetic information to resolve closely related species, including species that have experienced complex evolution histories. Additionally, the use of chloroplast genomes has been shown to be a useful tool for resolving phylogenies of *Calligonum* and *Fagopyrum*, which also belong to Polygonoaceae (Song et al., 2020; Yang et al., 2021).

In this study, we sequenced and annotated nine plastomes of *Rheum* and compared the plastome structure within the genus. Twenty-two accessions representing 17 species of *Rheum* were sampled for phylogenetic analyses. Divergence times of *Rheum* were estimated to investigate the origins and diversification of the genus. Our study not only provides valuable molecular markers for identification of *Rheum*, but also gains insights into the phylogeny, origination, and evolution of *Rheum*.

2 Material and Methods

2.1 Taxon sampling and sequencing

Samples from nine *Rheum* species were collected from Sichuan, Qinghai, Yunnan, Shanxi, and Ningxia provinces (Table S1). Additionally, one species of *Muehlenbeckia* (*Muehlenbeckia complexa*) was collected in the Wuhan Botanical Garden. Fresh leaves were collected in silica gel for DNA extraction. Voucher specimens were deposited at the herbarium of the Wuhan Botanical Garden (Table S1). Total genomic DNA was isolated using a modified CTAB method (Doyle & Doyle, 1987). DNA degradation and contamination were monitored on 1% agarose gels, while DNA purity was checked using a NanoPhotometer® spectrophotometer (Implen, Calabasas, CA, USA). A total of 1.5 µg of DNA per sample was used for sequencing. Libraries were generated using the Truseq Nano DNA HT Sample preparation kit (Illumina, San Diego, CA, USA). Samples were fragmented by sonication to a size of 350 bp, then DNA

fragments were end polished, A-tailed, and ligated with the full-length adapter for Illumina sequencing with further polymerase chain reaction (PCR) amplification. PCR products were purified (AMPure XP system, Beckman Coulter, Brea, CA, USA) and libraries were analyzed for size distribution using an Agilent2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and quantified using real-time PCR. Completed libraries were sequenced using an Illumina HiSeq platform with 150 bp paired-end reads. The target for each sample was a minimum of 2GB of raw sequencing data.

2.2 Plastome assembly and annotation

Raw sequencing reads were filtered with Trimmomatic v0.39 (Bolger et al., 2014) to remove adapters and low quality reads. Clean reads were *de novo* assembled using NOVOPlasty v4.3.1 (Dierckxsens et al., 2017) with the plastome sequence of *Rheum palmatum* (NCBI accession NC027728) as the reference. Assembled genomes were annotated with GeSeq (<https://chlorobox.mpimp-golm.mpg.de/geseq.html>). Start/stop codons and intron/exon boundaries were manually checked and modified in Geneious v9.0.2 (Kearse et al., 2012). tRNAs were annotated with tRNAscan-SE (Lowe & Eddy, 1997) as implemented in Geseq.

2.3 Comparative plastome analysis

To detect the expansion or contraction of the IR among *Rheum* species, we chose *R. officinale* (MN 564922) as the reference to visualize the borders of the LSC, SSC, and IR of 12 species (10 species of *Rheum*, one species of *Oxyria*, and one species of *Rumex*) using Irscope (Amiryousefi et al., 2018). The 10 selected *Rheum* species represent major clades based on the results of the phylogenetic analyses (see Section 3). The percentage of sequence identity was also plotted with the 12 selected species using mVISTA with Shuffle-LAGAN mode (Frazer et al., 2004).

We used the MISA-web application (Beier et al., 2017; <https://webblast.ipk-gatersleben.de/misa/>) to detect simple sequence repeats (SSRs) of the nine newly sequenced *Rheum* species. A minimum number of repeat units were set as: 10 for mononucleotide, five for dinucleotide, four for trinucleotide, and three for tetranucleotide, pentanucleotide, and hexanucleotide. Tandem repeat sequences were detected by Tandem Repeat Finder (<https://tandem.bu.edu/trf/trf.html>) (Benson, 1999) with default parameters: two for the alignment parameter match, seven for mismatch and indels, and 50 for minimum alignment score to report repeat. Dispersed and palindromic repeat were detected using REPuter with following parameter: three for hamming distance, 90 for maximum computed repeats, and 30 for minimal repeat size.

2.4 Estimation of substitution and positive selection on plastid genes

To detect genes with high divergence, 75 shared protein-coding genes (PCGs) among *Rheum* and closely related outgroups were extracted and then aligned with MAFFT v7.313 as implemented in PhyloSuite v1.2.2 (Zhang et al., 2020). We detected the degree of sequence divergence by calculating nucleotide diversity (π) of each PCG with DNAsp v6 (Rozas et al., 2017). To detect selective pressures on the plastid genes, non-synonymous substitution rates (dN),

synonymous substitution rates (dS), and dN/dS ratios (omega) of each PCG were calculated using the branch-site model using codeml in PAML (Yang, 1997). The site-specific model assumes that omega is the same across branches of the tree but can vary among different sites. We choose the maximum-likelihood (ML) tree from the phylogenetic analyses as the constraint topology.

2.5 Phylogenetic analyses

To explore the phylogenetic relationships among *Rheum*, 28 accessions were collected including 19 previously published plastomes (Table S1) plus the nine newly sequenced plastomes in this study. We selected two *Fagopyrum* species *F. leptopodum* (MW 017634) and *F. luojishanense* (NC 037706) as the outgroups. We extracted shared 75 PCGs, aligned them with the codon-aware program MACSE v2.03 (Ranwez et al., 2018), and then concatenated sequences into a supermatrix with PhyloSuite v1.2.2 (Zhang et al., 2020). We calculated the parsimony informative sites of the supermatrix with MEGAX (Kumar et al., 2018). We conducted both ML and Bayesian inference (BI) analyses. For the ML analysis, we used raxmlHPC-PTHREADS-AVX (Stamatakis, 2014) with the general time reversible model for nucleotide substitution with the gamma model of rate heterogeneity (GTR + G), and conducted 1000 rapid bootstrap replicates. Bayesian analyses were conducted with MrBayes v3.2.7 (Huelsenbeck & Ronquist, 2001). The best-fit model was calculated with jModeltest2 (Darriba et al., 2012) under the Bayesian information criterion (BIC). Two runs with four Markov chains were applied with 2 000 000 generations from a random starting tree with sampling every 500 generations. The initial 25% of sampled trees were discarded as burn-in, with the remaining trees used to construct a majority-rule consensus tree and calculate the posterior probability. Partitioned BI analyses were also conducted with the best-fit models for each subset calculated with PartitionFinder v2.1.1 (Lanfear et al., 2017). For the coalescent analysis, each PCG was used to construct a ML tree using raxmlHPC-PTHREADS-AVX (Stamatakis, 2014) under the GTR + G model with 500 bootstrap replicates. All generated gene trees were used to estimate the species tree with ASTRAL v5.6.3 (Zhang et al., 2018). The final phylogenetic topologies were viewed in FigTree v1.3.1 (Rambaut, 2009). In addition, we used the SplitsTree v4.13.1 (Huson & Bryant, 2006) to infer neighbor-net networks based on uncorrected p-distances calculated from the concatenated supermatrix.

2.6 Estimation of divergence time

For the estimation of divergence time, more genera were added to the sequence matrix, including *Muehlenbeckia*, *Afrobrunnichia*, *Antigonon* and *Symmeria*. We selected three species of Plumbaginaceae (*Ceratostigma willmottianum* NC041261, *Plumbago auriculata* NC041245, and *Limonium tenellum* NC041279) as the outgroups (Table S1). Divergence time was estimated using BEAST v1.10.4 (Drummond & Rambaut, 2007). For the BEAST analysis, we used the GTR + G as the substitution model, the relaxed molecular clock model and Yule process as the tree prior. According to previous studies (Manchester & O'Leary, 2010; Schuster et al., 2013; Yao et al., 2019), two fossil calibrations and three secondary calibrations were used. The crown age of

Polygonaceae was set to an age range of 72.1–66 Ma with a lognormal calibration prior distribution (Manchester & O'Leary, 2010). The crown age of *Muehlenbeckia* was set to an age range of 22–19.0 Ma with a lognormal calibration prior (Pole, 1992; Schuster et al., 2013). The three secondary calibrations included the crown age of *Afrobrunnichia* and *Antigonon* set to a mean age 67 Ma with a SD of 3 Ma, the crown age of Plumbaginaceae set to a mean age of 60 Ma with a SD of 3 Ma, and the crown age of all species set to a mean age of 91.8 Ma with a SD of 0.5 Ma (Yao et al., 2019); each of the three calibration priors were set as a normal distribution. The MCMC was run for 1×10^9 generations, sampling every 10 000 generations until a sufficient effective sample size (ESS) > 200 for all relevant parameters were evaluated in Tracer v1.7 (Rambaut et al., 2018). The first 25% trees were discarded as burn-in and the maximum clade credibility tree was summarized with TreeAnnotator v1.10.4 (Suchard et al., 2018) with mean heights reported for node heights.

3 Results

3.1 Genome assembly and plastome features

Illumina sequencing generated a total of 14 893 932–21 253 448 pair-end clean reads for each species (Table 1). The mean sequencing coverage of the observed plastome data ranged from 524x (*R. delavayi*) to 1282x (*R. rhabarbarum*). All nine plastomes displayed a typical quadripartite structure, consisting of a LSC (86 507–87 578 bp), SSC (12 717–12 842 bp), and two IRs (IRa and IRb; 30 855–31 007 bp). The length of the complete plastomes ranged from 161 131 bp (*R. officinale*) to 162 374 bp (*R. delavayi*) with a similar GC content (37.26%–37.41%). The IR regions were highly consistent among the nine *Rheum* species, consisting of 17 genes (6 PCG, 7 tRNA genes). All plastomes of *Rheum* contained 79 PCGs, four rRNAs and 30 tRNA arranged in the same gene order (Table 1). In *Rheum*, 18 genes possess introns, among which 12 possess only one intron, including eight PCGs (*atpF*, *ndhA*, *petB*, *petD*, *rpl16*, *rpoC1*, *rps12*, *rps16*), while four tRNA (*trnG-UCC*, *trnK-UUU*, *trnL-UAA*, and *trnV-UAC*), four PCGs (*clpP*, *ndhB*, *rpl2*, and *ycf3*), and two tRNA (*trnA-UGC* and *trnI-GAU*) possess two introns.

The detection of the IR/SC boundary implied conservation among the examined species, with a few expansions and contractions detected (Fig. 1). High similarity in the complete plastomes was detected within the ten *Rheum* species showing high conservation within the genus (Fig. 2). High divergence regions can be seen in the LSC with non-coding sequences (Fig. 2).

3.2 SSRs and repeat sequences

A total of 565 SSRs were identified among the newly sequenced *Rheum* species using MISA. The number of SSRs ranged from 55 (*R. pumilum* and *R. uninerve*) to 77 (*R. officinale*). The number of mononucleotide repeats is the largest (289), then the dinucleotide (129), trinucleotide (76), tetranucleotide (62), pentanucleotide (7), and hexanucleotide (2) repeats (Fig. S1c). The mononucleotide repeat motif (A/T), dinucleotide repeat motif (AT/TA), trinucleotide

Table 1 Assembly information of newly sequenced plastomes of nine *Rheum* species

Species	<i>R. rhomboideum</i>	<i>R. spiciforme</i>	<i>R. delavayi</i>	<i>R. pumilum</i>	<i>R. uninerve</i>	<i>R. alexandrae</i>	<i>R. rhabarbarum</i>	<i>R. officinale</i>	<i>R. racemiferum</i>
Total reads	148 93 932	21 253 448	17 148 308	17 892 900	15 018 740	1 684 678	16 115 476	16 643 358	17 493 796
Assembled reads	755 164	874 124	536 534	755 940	213 880	1 110 486	1 294 264	421 148	1 156 252
Average coverage	731	867	524	736	212	1081	1282	419	1132
Plastome size (bp)	161 833	161 614	162 374	162 208	161 927	161 698	161 655	161 131	161 915
LSC length (bp)	87 151	86 924	87 578	87 449	87 297	87 011	86 936	86 507	87 408
IR length (bp)	30 981	30 933	30 996	31 007	30 931	30 961	30 979	30 956	30 855
SSC length (bp)	12 720	12 842	12 804	12 745	12 768	12 765	12 761	12 712	12 797
GC content (%)	37.38	37.41	37.26	37.27	37.41	37.32	37.37	37.33	37.41
Number of genes	131	131	131	131	131	131	131	131	131
tRNAs	37	37	37	37	37	37	37	37	37
rRNAs	8	8	8	8	8	8	8	8	8

repeat motif (AAG/CTT and AAT/ATT), and tetranucleotide repeat motif (AAAT/ATTT and AACT/AGTT) can be detected in each plastome (Fig. S1d).

A total of 386 repeats were detected by Reputer, with the repeat number ranging from 30 (*R. spiciforme*) to 52 (*R. pumilum*). Palindromic repeats were the most abundant types with a total of 190. A total of 129 forward repeats, four complement repeats and two reverse repeats were found (Fig. S1a). A total of 386 tandem repeats were detected by Tandem Repeat Finder, ranging from 38 (*R. spiciforme*) to 52 (*R. pumilum*) (Fig. S1b).

3.3 Nucleotide diversity and selective pressure analysis

We identified divergent genes with high nucleotide diversity (π), with the top seven highly divergent genes (π ranging from 0.02518 to 0.06562) being: *accD*, *ccsA*, *matK*, *ndhF*, *ndhH*, *psbM*, and *rpl32* (Fig. 3A; Table S2). Among the top seven genes, two genes (*psbM* and *rpl32*) are under 500 bp in length (174–414 bp) with a low number of parsimony informative sites (5 and 17) (Table S2). The remaining five genes (*accD*, *ccsA*, *matK*, *ndhF*, *ndhH*) can be utilized as effective molecular markers for phylogenetic and population analyses. The dN value of the 75 shared PCGs ranged from 0 (*psbI* and *rps2*) to 0.1372 (*ccsA*), while the dS value ranged from 0.0448 (*ndhC*) to 0.7435 (*ndhH*) (Fig. 3B; Table S2). The dS of *psaI* and *psaB* is near 0 (0.0001 and 0.0002) which may result in a large ratio of dN/dS. The low dS value may result from the highly similar and short sequence of *psaI* and *psaB*. The ratio of the other 73 PCGs ranged from 0 (*psbI* and *rps2*) to 0.69111 (*ycf1*) (Fig. 3B; Table S2). The results imply that all the genes are under purifying selection, with no signals of positive selection detected.

3.4 Phylogenetic relationship

For the phylogenetic analysis, the alignment of 75 PCGs extracted from the 28 accessions was 69 123 bp in length, containing 4359 parsimony-informative sites. The ML, BI, and partitioned BI analyses generated identical tree topologies (Fig. 4A). Twenty-two individuals representing 17 species of *Rheum* formed a monophyletic clade with strong support (BS/PP = 100/1.00). Within *Rheum*, two main clades were resolved (BS/PP = 100/1.00). In clade A, *R. racemiferum* was the earliest diverging clade (BS/PP = 91/1.00), followed by *R. alexandrae* (BS/PP = 100/1.00). *Rheum hotaoense* formed a sister relationship with *R. rhabarbarum*, which was then sister to *R. franzenbachii* (BS/PP = 100/1.00). In the *R. palmatum* complex, *R. officinale* and *R. tanguticum* did not form monophyletic species (Fig. 4A). In clade B, two small clusters were formed (BS/PP = 97/1.00). In B1, *R. rhomboideum* and *R. spiciforme* formed a sister clade and then were sister to *R. nobile*, which then clustered with *R. acuminatum* (BS/PP = 100/1.00) (Fig. 4A). In cluster B2, *R. delavayi* formed a sister clade with *R. likiangense*, and *R. przewalskyi* formed a sister clade with *R. pumilum* (BS/PP = 100/1.00). The above two clades clustered as sister clades, with *R. uninerve* sister to the rest (BS/PP = 100/1.00) (Fig. 4A). The presented phylogeny covers six of the eight previously proposed sections, with our results showing four sections being polyphyletic (Fig. 4A). The coalescent analysis generated an inconsistent tree topology compared to the concatenated analyses (Fig. 4B). Clade B1 formed a sister relationship with

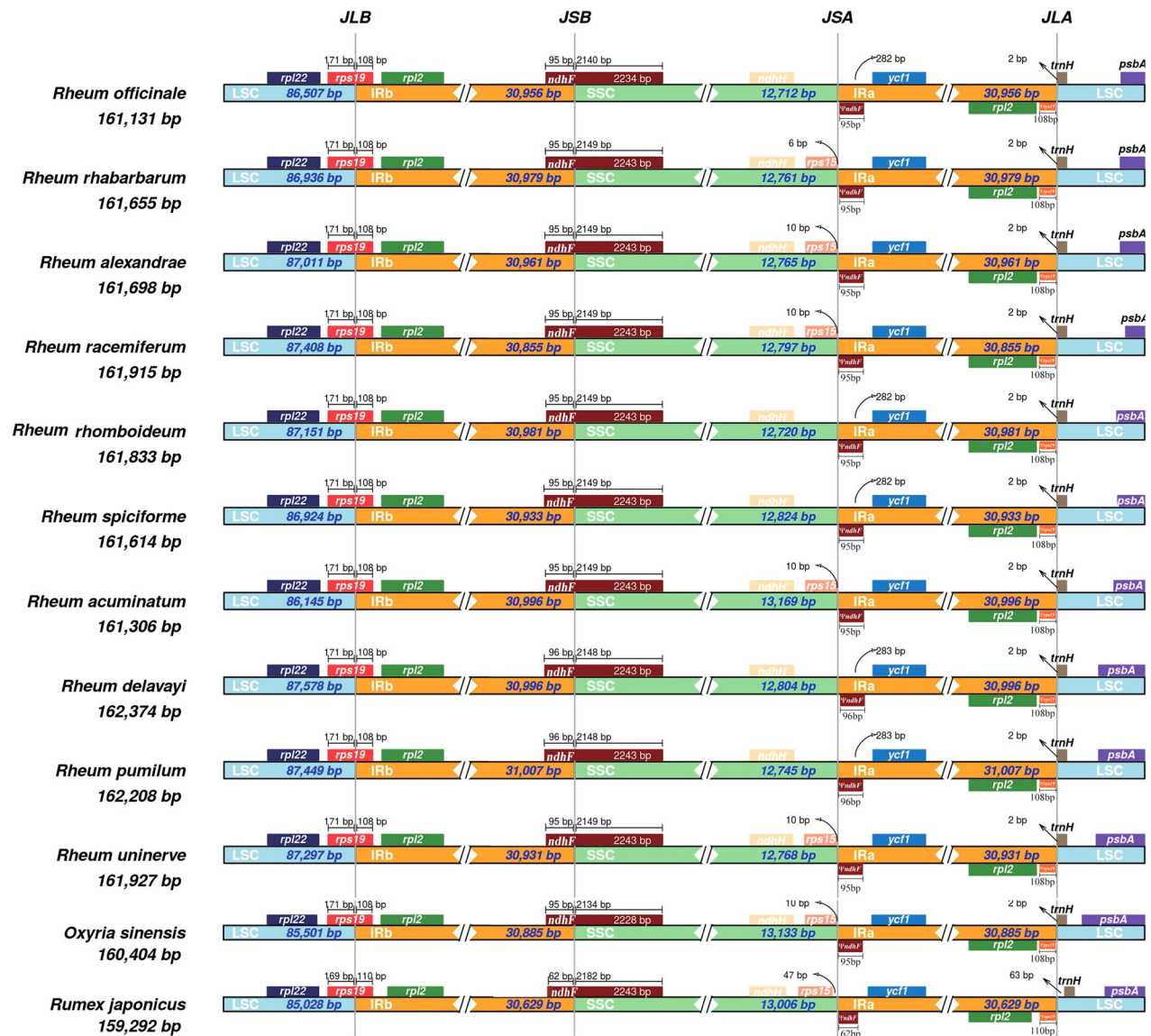


Fig. 1. The IR/SSC borders in the plastomes of 10 *Rheum* species, 1 *Rumex* species, and 1 *Oxyria* species. JLB, JSB, JSA, and JLA denote the junction sites of LSC/IRb, SSC/IRb, SSC/IRa and LSC/IRa, respectively.

clade A, which was then sister to the rest of clade B (Fig. 4B). However, species relationships in the small clades were almost identical with the concatenated analyses results. The network analysis revealed reticulate evolution may have occurred within *Rheum* (Fig. S2).

3.5 Estimation of divergence times

For the BEAST analysis, 36 accessions representing 31 species were included, with 74 shared PCGs extracted and aligned for a total length of 74 058 bp. The BEAST analysis revealed that the split between *Rheum* and *Oxyria* was dated to ca. 27.73 Ma (95% HPD 16.86–39.73 Ma) (Fig. 5). The divergence of the whole genus was estimated to ca. 18.67 Ma (95% HPD 9.95–28.37 Ma), and the two major clades within *Rheum* diversified almost at the same time (ca. 16.32–16.60 Ma) (Fig. 5). Most species originated after the middle Miocene, suggesting a rapid recent diversification of *Rheum*.

4 Discussion

In this study, we newly assembled and analyzed nine complete plastomes of *Rheum*. All *Rheum* species share a similar gene content and structure, possessing a typical quadripartite structure with two IRs separated by the LSC and SSC region. Generally, the length of the chloroplast genome is primarily affected by the expansion and contraction of IR regions (Weng et al., 2014; He et al., 2017). However, we see no obvious expansion or contraction in *Rheum*. Moreover, structural changes such as gene loss and rearrangement were not detected either suggesting that the plastomes of *Rheum* are highly conserved. A high degree of plastome conservation has also been detected in other genera of Polygonaceae such as *Calligonum* and *Fagopyrum* (Song et al., 2020; Yang et al., 2021). Selective pressure analysis implies that all PCGs of the plastomes are under

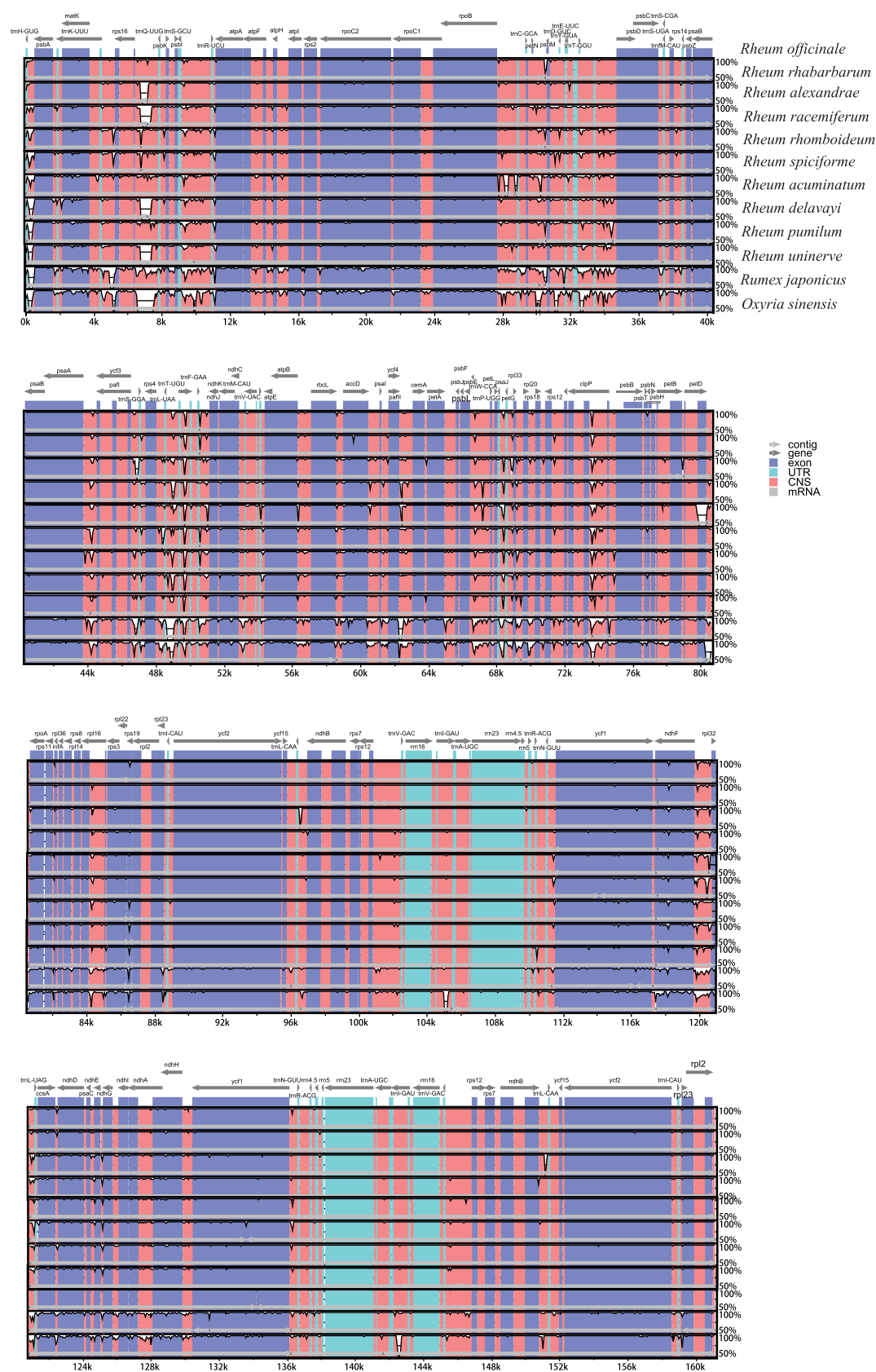
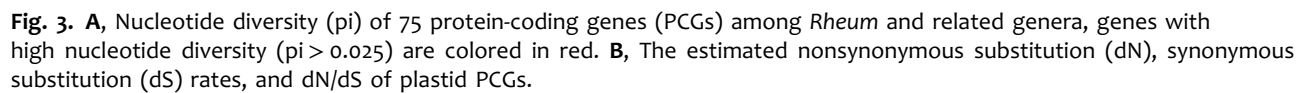


Fig. 2. Percentage of identity comparing 12 plastomes of 10 *Rheum* species, 1 *Oxyria* species, and 1 *Rumex* species, with *R. officinale* as the reference (mVISTA). The y-axis represents the percent identity spanning 50–100%. Genome regions are color-coded as protein-coding region (purple), rRNA or tRNA coding genes (cyan), non-coding sequences (red).



informative sites revealed five genes (*accD*, *ccsA*, *matK*, *ndhF*, and *ndhH*) that can be used as genetic markers for species delimitation and phylogenetic studies in the future.

Three species of Sect. *Palmeta* including *R. officinale*, *R. palmatum*, and *R. tanguticum* are treated as important traditional medicine in China. Previous studies proposed that these three species should be treated as one species due to genetic similarity, although morphological features such as the degree of leaf blade dissection and the shape of the lobes are different among these species (Wang et al., 2014, 2018). In the phylogenetic analyses presented here, three accessions of *R. officinale* and two accessions of *R. tanguticum* did not form respective monophyletic units. However, all six accessions of these three species did form a monophyletic group in both ML and MrBayes analyses (Fig. 4A). While our research supports previous studies that these three species can be treated as one species (Wang et al., 2014, 2018), further investigations based on the nuclear genome is still needed to illustrate the divergence history among populations of these species.

Discordant phylogenetic results were detected between concatenated and coalescent approaches in our analyses. Doyle (2021) proposed that the plastome should be treated as a single estimate of the underlying species phylogeny. The tree topology generated by the concatenated data showed strong support values and is overall similar to previous studies (Zhou et al., 2020), with the results improving our understanding of the backbone of *Rheum*. Meanwhile, coalescent analyses of the plastome loci can show other events such as incomplete lineage sorting (ILS) and hybridization (Gonçalves et al., 2019, 2020; Walker et al., 2019).

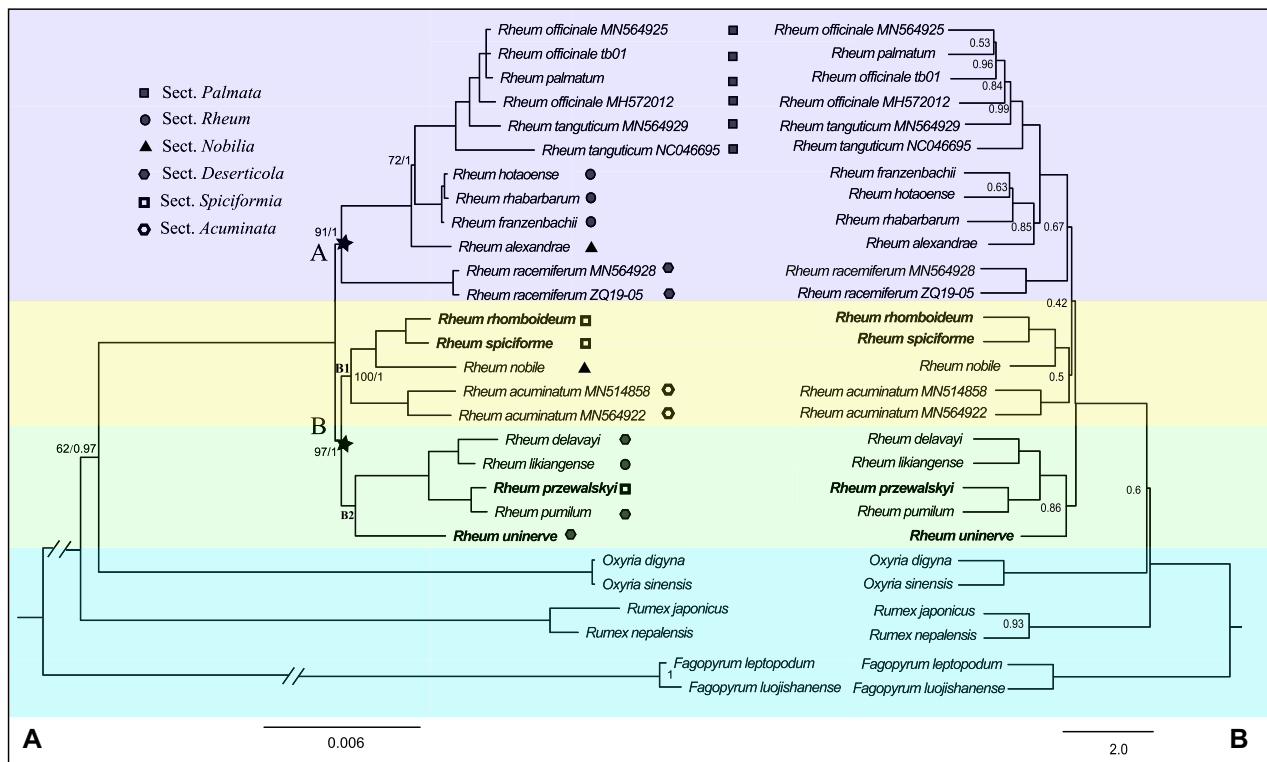


Fig. 4. **A**, Phylogenetic tree of 31 taxa using maximum likelihood (ML) and Bayesian inference (BI) based on 75 shared genes. ML bootstrap values (BS) and posterior probabilities (PP) are shown at nodes. Branches with no values listed have 100% BS and PP of 1.0. **B**, Coalescent analyses of 75 shared genes. Branches with no values listed have 1.0 local posterior probability. The same species from different accessions were distinguished with genbank ID or voucher information after the name of species. Different colors represent major clades in our study. Two star symbols represent the two major clades in *Rheum*, which correspond to clade A and B. B1 and B2 are two smaller clades in clade B; species of B1 correspond to species with yellow background, species of B2 correspond to species with green background.

In phylogenetic analyses typically inferred from estimated genealogical histories of one or several genes, a gene tree may be topologically inconsistent with its species tree which is known as ILS (Pamilo & Nei, 1988; Maddison, 1997; Nichols, 2001). ILS is mainly the result of genetic polymorphism in the ancestral species and often occurs when the time of divergence between species is short (Pamilo & Nei, 1988). In addition, our network analysis revealed reticulate evolution may have also occurred in *Rheum* species. Overall, the phylogenetic discordance from different methods, along with weakly supported resolution of some relationships, provides new evidence for a complex evolutionary history possibly involving ILS and hybridization. This indicates further investigation using more molecular data such as single copy nuclear genes, as well as a wider species sampling, are necessary in further analyses.

Turbulent climate and tectonic movement are important factors driving species formation and differentiation (Bruch et al., 2006; Fortelius et al., 2006; Johnson et al., 2006). The geological process and climatic changes during the Neogene had tremendous impacts on the evolution of biota in different regions of Eurasia, especially in the QTP (Qin et al., 2013; Yu et al., 2014; Favre et al., 2015). The divergence of *Rheum* was estimated to begin at ca. 18.67 Ma (95% HPD 9.95–28.37 Ma), with most species originating in the Miocene

and the differentiation of most species happened during the middle Miocene to Pliocene (Fig. 5), which is nearly identical with previous studies (Sun et al., 2012; Zhou et al., 2020). The QTP experienced further extension during the Late Miocene and the Pliocene, particularly on the eastern edge, for example, the Hengduan Mountains (Zhisheng et al., 2001; Zhang et al., 2013). Additionally, the uplift of high mountain ranges of the QTP massively altered the air circulation, together with a worldwide cooling since the Middle Miocene climate optimum (Miao et al., 2012), causing the progressive aridification of northwest China (Favre et al., 2015). *Rheum* species are always drought-enduring with rhizomes and stout roots. Previous studies revealed that the QTP and adjacent regions are the center of origin and diversity of *Rheum* (Sun et al., 2012), with high species richness found in the southwest and northwest of China (Bao & Grabovskaya-Borodina, 2003). The uplift of QTP during the Late Miocene and the Pliocene, along with the aridification of northwest China, may have facilitated the radiation of *Rheum* species. Therefore, our results support the hypothesis that geological and climatic fluctuation of the QTP played important roles in driving biological diversification of *Rheum* species.

Eight sections of *Rheum* were proposed by Bao & Grabovskaya-Borodina (2003), with sections defined primarily based on morphological characters such as inflorescence

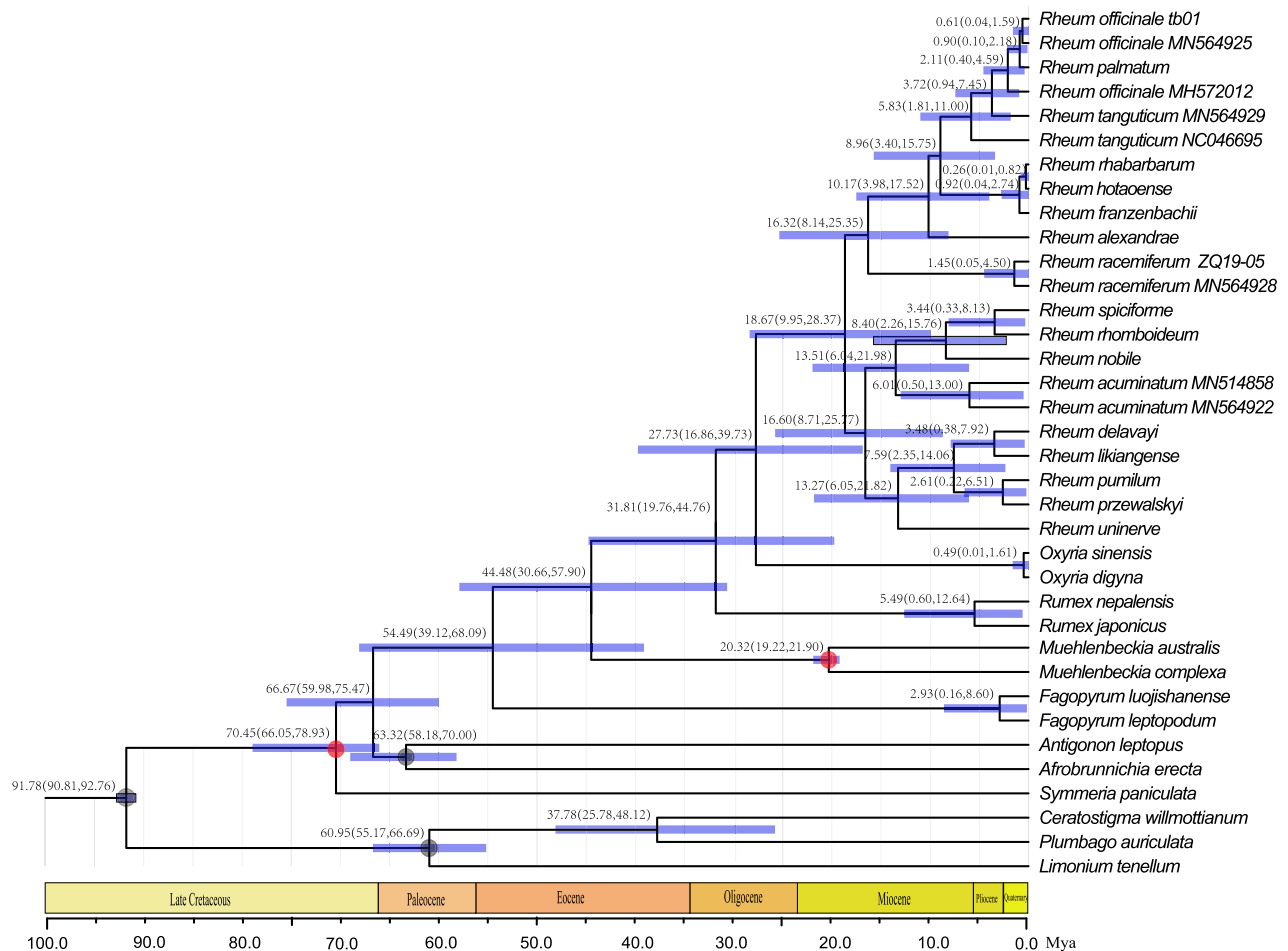


Fig. 5. BEAST-derived chronograms of Polygonaceae based on 74 shared protein-coding genes (PCGs) with 2 fossil calibration (red circles) and 3 second calibration point (gray circles).

styles and shape of leaves. Our study comprised six of the sections of *Rheum* and included the most diverse morphologies of the group. Most sections are not monophyletic, with four section circumscriptions being highly inconsistent with our phylogenomic results. Actually, it is a little difficult to find morphological synapomorphies within sections. However, for some species, phylogenetic placement is supported by morphological characters. For instance, Sect. *Rheum* is characterized with large herbs and triangular-ovate leaves with sinuate. *Rheum likiangense* belongs to Sect. *Rheum*, the leaves of *R. likiangense* are entire or slightly sinuate. *Rheum likiangense* is clustered in clade A, while the other three Sect. *Rheum* species possess obvious sinuate leaves and are clustered in clade B (Fig. 4A). Species of Sect. *Deserticola* always possess short plant heights under 50 cm and four species of Sect. *Deserticola* were included here. *Rheum racemiferum* has a plant height taller than 50–70 cm and is clustered in clade A, while the other three species with low plant height are clustered in clade B (Fig. 4A). In Sect. *Spiciformia*, three species were included and all clustered in clade B, with *R. rhomboideum* and *R. spiciforme* forming a sister relationship. Their panicles are not deliquescently branched (without side branches), while *R. przewalskyi* clustered with another small lineage with panicle deliques-

cently branched (with side branches). Morphological comparisons of species within sections can be validated according to the phylogenetic tree to some extent.

The QTP is the largest and highest plateau in the world and hosts a great amount of biodiversity (Sun et al., 2014, 2017), and is a natural museum of the floristic evolution along with geological evolution (Sun, 2002; Chen et al., 2018). *Rheum nobile* and *R. alexandrae* are known as the “glasshouse” in high mountains, with the upper leaves becoming large translucent or cream-colored bracts that cover the inflorescences. The two species clustered in two clades, implying parallel evolution in different lineages is responsible for the observed phenotypes. The robust phylogenetic trees also provide conditions for exploring the evolution of decumbent/caulescent body-plan traits in the genus. Four species with this decumbent trait within *Rheum* were included: *R. rhomboideum*, *R. spiciforme*, *R. przewalskyi*, and *R. uninerve*. The four species did not form a monophyletic unit, which implies that this decumbent trait of *Rheum* also experienced parallel evolution. Both glasshouse structures and cushions of *Rheum* may have adaptively evolved to cope with hostile alpine environments according to the estimation of divergence times. Natural selection plays an important role in shaping current biodiversity of *Rheum*.

5 Conclusion

Nine complete plastomes of *Rheum* and one species of *Muehlenbeckia* were newly sequenced and annotated. Including previously published plastomes that were downloaded, a total of 18 *Rheum* species were used for phylogenomic analyses. The results indicate that all *Rheum* species possess a conserved plastome structure with the same gene order. We detected five genes with high nucleotide diversity and SSRs that can be used as gene markers for phylogenetic analyses or species delimitation in future analyses. All 75 shared plastid genes are undergoing purifying selection and no genes were subjected to positive selection. The phylogenomic analyses revealed plastomes are useful for resolving the phylogenetic relationships among *Rheum* species. ILS and hybridization likely occurred within *Rheum*, which may be the reasons that previous phylogenetic analyses using DNA fragments could not fully resolve phylogenetic relationships. The estimated divergence times imply the climatic changes and orogeneses during the Miocene of the QTP were the main triggers of *Rheum* origination and diversification. Highly diversified morphological characters of *Rheum* are adaptively evolved to cope with hostile alpine environments.

Acknowledgements

This study was supported by grants from the Second Tibetan Plateau Scientific Expedition and Research (STEP) program (2019QZKK0502), the Key Projects of the Joint Fund of the National Natural Science Foundation of China (U1802232), the Strategic Priority Research Program of Chinese Academy of Sciences (XDA20050203), the Major Program of the National Natural Science Foundation of China (31590823), the Youth Innovation Promotion Association of Chinese Academy of Sciences (2019382), The Ten Thousand Talents Program of Yunnan Province (202005AB160005).

References

- Amiryousefi A, Hyvönen J, Pocai P. 2018. Irscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* 34: 3030–3031.
- Bao BJ, Grabovskaya-Borodina AE. 2003. *Rheum*. In: Wu ZY, Raven PH, Hong DY eds. *Flora of China*. Beijing: Science Press; St. Louis: Missouri Botanical Garden Press. 5: 341–350.
- Beier S, Thiel T, Münch T, Scholz U, Mascher M. 2017. Misa-web: A web server for microsatellite prediction. *Bioinformatics* 33: 2583–2585.
- Bendich AJ. 2004. Circular chloroplast chromosomes: The grand illusion. *The Plant Cell* 16: 1661–1666.
- Benson G. 1999. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Research* 27: 573–580.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for illumina sequence data. *Bioinformatics* 30: 2114–2120.
- Bruch A, Utescher T, Mosbrugger V, Gabrielyan I, Ivanov D. 2006. Late Miocene climate in the circum-Alpine realm—a quantitative analysis of terrestrial palaeofloras. *Palaeogeography, Palaeoclimatology, Palaeoecology* 238: 270–280.
- Chen YS, Deng T, Zhou Z, Sun H. 2018. Is the east Asian flora ancient or not? *National Science Review* 5(6): 142–154.
- Chin TC, Youngken HW. 1947. The cytotoxicity of *Rheum*. *American Journal of Botany* 34: 401–407.
- Chinese Pharmacopoeia Committee. 2015. *Pharmacopoeia of the People's Republic of China*. Beijing: China Medical Science Press.
- Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods* 9: 772.
- Dierckx N, Mardulyn P, Smits G. 2017. Novoplasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research* 45: e18.
- Doyle J, Doyle J. 1987. A rapid procedure for DNA purification from small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.
- Doyle JJ. 2021. Defining coalescent genes: Theory meets practice in organelle phylogenomics. *Systematic Biology* 1–14.
- Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7: 1–8.
- Favre A, Päckert M, Pauls SU, Jähnig SC, Uhl D, Michalak I, Muellner-Riehl AN. 2015. The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. *Biological Reviews* 90: 236–253.
- Fortelius M, Eronen J, Liu LP, Pushkina D, Tesakov A, Vislobokova I, Zhang ZQ. 2006. Late Miocene and Pliocene large land mammals and climatic changes in Eurasia. *Palaeogeography, Palaeoclimatology, Palaeoecology* 238: 219–227.
- Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. 2004. VISTA: Computational tools for comparative genomics. *Nucleic Acids Research* 32: W273–W279.
- Gonçalves DJ, Jansen RK, Ruhlman TA, Mandel JR. 2020. Under the rug: Abandoning persistent misconceptions that obfuscate organelle evolution. *Molecular Phylogenetics and Evolution* 151: 106903.
- Gonçalves DJ, Simpson BB, Ortiz EM, Shimizu GH, Jansen RK. 2019. Incongruence between gene trees and species trees and phylogenetic signal variation in plastid genes. *Molecular Phylogenetics and Evolution* 138: 219–232.
- He L, Qian J, Li XW, Sun ZY, Xu XL, Chen SL. 2017. Complete chloroplast genome of medicinal plant *Lonicera japonica*: Genome rearrangement, intron gain and loss, and implications for phylogenetic studies. *Molecules* 22: 249.
- Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754–755.
- Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23: 254–267.
- Johnson WE, Eizirik E, Pecon-Slatery J, Murphy WJ, Antunes A, Teeling E, O'Brien SJ. 2006. The late Miocene radiation of modern Felidae: A genetic assessment. *Science* 311: 73–77.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647–1649.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* 35: 1547–1549.
- Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2017. PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution* 34: 772–773.

- Losina-Losinskaya A. 1936. The genus *Rheum* and its species. *Acta Instituti Botanici Academiae Scientiarum Unionis Rerum Publicarum Sovieticarum Socialistarum*, Ser 1: 5–141.
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* 25: 955–964.
- Ma YP, Zhao L, Zhang WJ, Zhang YH, Xing X, Duan XX, Hu J, Harris A, Liu PL, Dai SL. 2020. Origins of cultivars of *Chrysanthemum*—evidence from the chloroplast genome and nuclear LFY gene. *Journal of Systematics and Evolution* 58: 925–944.
- Maddison WP. 1997. Gene trees in species trees. *Systematic Biology* 46: 523–536.
- Manchester SR, O'Leary EL. 2010. Phylogenetic distribution and identification of fin-winged fruits. *The Botanical Review* 76: 1–82.
- Maréchal A, Brisson N. 2010. Recombination and the maintenance of plant organelle genome stability. *New Phytologist* 186: 299–317.
- Miao Y, Herrmann M, Wu F, Yan X, Yang S. 2012. What controlled Mid-Late Miocene long-term aridification in Central Asia?—Global cooling or Tibetan Plateau uplift: A review. *Earth-Science Reviews* 112, 155–172.
- Mu XY, Tong L, Sun M, Zhu YX, Wen J, Lin QW, Liu B. 2020. Phylogeny and divergence time estimation of the walnut family (Juglandaceae) based on nuclear RAD-Seq and chloroplast genome data. *Molecular Phylogenetics and Evolution* 147: 106802.
- Nichols R. 2001. Gene trees and species trees are not the same. *Trends in Ecology & Evolution* 16: 358–364.
- Palmer JD. 1985. Comparative organization of chloroplast genomes. *Annual Review of Genetics* 19: 325–354.
- Pamilo P, Nei M. 1988. Relationships between gene trees and species trees. *Molecular Biology and Evolution* 5: 568–583.
- Pole M. 1992. Early Miocene flora of the Manuhirika Group, New Zealand. 2. Conifers. *Journal of the Royal Society of New Zealand* 22: 287–302.
- Qin AL, Wang MM, Cun YZ, Yang FS, Wang SS, Ran JH, Wang XQ. 2013. Phylogeographic evidence for a link of species divergence of *Ephedra* in the Qinghai-Tibetan Plateau and adjacent regions to the Miocene Asian aridification. *PLoS One* 8: e56243.
- Rambaut A. 2009. FigTree v1. 3.1. <http://tree.bio.ed.ac.uk/software/figtree/>. last retrieved on 15.3.2021.
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* 67: 901–904.
- Ranwez V, Douzery EJ, Cambon C, Chantret N, Delsuc F. 2018. MACSE v2: Toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. *Molecular Biology and Evolution* 35: 2582–2584.
- Raven JA, Allen JF. 2003. Genomics and chloroplast evolution: What did cyanobacteria do for plants? *Genome Biology* 4: 1–5.
- Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution* 34: 3299–3302.
- Schuster TM, Setaro SD, Kron KA. 2013. Age estimates for the buckwheat family Polygonaceae based on sequence data calibrated by fossils and with a focus on the Amphipacific *Muehlenbeckia*. *PLoS One* 8: e61261.
- Song F, Li T, Burgess KS, Feng Y, Ge XJ. 2020. Complete plastome sequencing resolves taxonomic relationships among species of *Calligonum* L. (Polygonaceae) in China. *BMC Plant Biology* 20: 1–15.
- Stamatakis A. 2014. RAXML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.
- Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus evolution* 4: vey016.
- Sun H. 2002. Tethys retreat and Himalayas-Hengduan Mountains uplift and their significance on the origin and development of the Sino-Himalayan elements and alpine flora. *Acta Botanica Yunnanica* 24: 273–288.
- Sun H, Niu Y, Chen YS, Song B, Liu CQ, Peng DL, Chen JG, Yang Y. 2014. Survival and reproduction of plant species in the Qinghai-Tibet Plateau. *Journal of Systematics and Evolution* 52: 378–396.
- Sun H, Zhang JW, Deng T, Boufford DE. 2017. Origins and evolution of plant diversity in the Hengduan Mountains, China. *Plant Diversity* 39: 161–166.
- Sun YS, Wang AL, Wan DS, Wang Q, Liu JQ. 2012. Rapid radiation of *Rheum* (Polygonaceae) and parallel evolution of morphological traits. *Molecular Phylogenetics and Evolution* 63: 150–158.
- Walker JF, Walker-Hale N, Vargas OM, Larson DA, Stull GW. 2019. Characterizing gene tree conflict in plastome-inferred phylogenies. *PeerJ* 7: e7747.
- Wan DS, Sun YS, Zhang X, Bai XT, Wang J, Wang AL, Milne R. 2014. Multiple ITS copies reveal extensive hybridization within *Rheum* (Polygonaceae), a genus that has undergone rapid radiation. *PLoS One* 9: e89769.
- Wang AL, Yang MH, Liu JQ. 2005. Molecular phylogeny, recent radiation and evolution of gross morphology of the rhubarb genus *Rheum* (Polygonaceae) inferred from chloroplast DNA trn L-F sequences. *Annals of Botany* 96: 489–498.
- Wang XM, Feng L, Zhou T, Ruhsam M, Huang L, Hou XQ, Sun XJ, Fan K, Huang M, Zhou Y. 2018. Genetic and chemical differentiation characterizes top-geoherb and non-top-geoherb areas in the TCM herb rhubarb. *Scientific Reports* 8: 1–14.
- Wang XM, Hou XQ, Zhang YQ, Li Y. 2014. Morphological variation in leaf dissection of *Rheum palmatum* complex (Polygonaceae). *PLoS One* 9: e110760.
- Weng ML, Blazier JC, Govindu M, Jansen RK. 2014. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Molecular Biology and Evolution* 31: 645–659.
- Yang BB, Li LD, Liu JQ, Zhang LS. 2021. Plastome and phylogenetic relationship of the woody buckwheat *Fagopyrum tibeticum* in the Qinghai-Tibet Plateau. *Plant Diversity* 43: 198–205.
- Yang ZH. 1997. PAML: A program package for phylogenetic analysis by maximum likelihood. *Bioinformatics* 13: 555–556.
- Yao G, Jin JJ, Li HT, Yang JB, Mandala VS, Croley M, Mostow R, Douglas NA, Chase MW, Christenhusz MJ. 2019. Plastid phylogenomic insights into the evolution of Caryophyllales. *Molecular Phylogenetics and Evolution* 134: 74–86.
- Ye JF, Niu YT, Feng YL, Liu B, Hai LS, Wen J, Chen ZD. 2020. Taxonomy and biogeography of *Diapensia* (Diapensiaceae) based on chloroplast genome data. *Journal of Systematics and Evolution* 58: 696–709.
- Yu XQ, Maki M, Drew BT, Paton AJ, Li HW, Zhao JL, Conran JG, Li J. 2014. Phylogeny and historical biogeography of *Isodon* (Lamiaceae): Rapid radiation in south-west China and Miocene overland dispersal into Africa. *Molecular Phylogenetics and Evolution* 77: 183–194.
- Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19: 15–30.

- Zhang D, Gao FL, Jakovlić I, Zou H, Zhang J, Li WX, Wang GT. 2020. Phylosuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Molecular Ecology Resources* 20: 348–355.
- Zhang KX, Wang GC, Xu YD, Luo MS, Ji JL, Xiao GQ, Wang A, Song BW, Liang YP, Jiang SS. 2013. Sedimentary evolution of the Qinghai–Tibet Plateau in Cenozoic and its response to the uplift of the plateau. *Acta Geologica Sinica-English Edition* 87: 555–575.
- Zhisheng A, Kutzbach JE, Prell WL, Porter SC. 2001. Evolution of Asian monsoons and phased uplift of the Himalaya–Tibetan plateau since Late Miocene times. *Nature* 411: 62–66.
- Zhou T, Zhu HH, Wang J, Xu YC, Xu FS, Wang XM. 2020. Complete chloroplast genome sequence determination of *Rheum* species and comparative chloroplast genomics for the members of Rumiceae. *Plant Cell Reports* 39: 811–824.

Supplementary Material

The following supplementary material is available online for this article at <http://onlinelibrary.wiley.com/doi/10.1111/jse.12814/supinfo>:

Fig. S1. Statistics of repeat elements in the plastomes of ten *Rheum* species. **a** Number of forward, palindromic, complement and reverse repeats. **b**, Number of tandem repeats. **c**, Different simple sequence repeat (SSR) types detected in nine species. **d**, Frequency of identified SSR motifs in nine plastomes.

Fig. S2. Network of *Rheum* species.

Table S1. Taxa included in present study. NCBI accession numbers and voucher specimens' information are provided for newly sequenced plastomes. “.” indicates not applicable.

Table S2. Genetic characteristics of 75 protein-coding genes used in analyses, including length, nucleotide diversity (π), parsimony informative sites (PI), nonsynonymous substitution rate (dN), synonymous substitution rate (dS), and dN/dS.