



# Complete chloroplast genome sequence determination of *Rheum* species and comparative chloroplast genomics for the members of Rumiceae

Tao Zhou<sup>1,2</sup> · Honghong Zhu<sup>1</sup> · Jian Wang<sup>1</sup> · Yucan Xu<sup>1</sup> · Fusheng Xu<sup>1</sup> · Xumei Wang<sup>1,2</sup>

Received: 21 October 2019 / Accepted: 13 March 2020 / Published online: 27 March 2020  
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

## Abstract

**Key message** Complete plastomes of *Rheum* species facilitated to clarify the phylogeny of Polygonaceae, and comparative chloroplast genomics contributed to develop genetic markers for the authentication of *Rheum* species.

**Abstract** *Rheum* (Polygonaceae) is widely distributed throughout the temperate and subtropical areas of Asian interior. *Rheum* species are usually perennial herbs, and half of them are endemic to China with important medicinal properties. On account of similar morphological characteristics, species delimitation of *Rheum* still remains unclear. Chloroplast genomes of eight *Rheum* species, *Rumex crispus* and *Oxyria digyna* were characterized. Based on the comparison of genome structure of these species and the two published *Rheum* species, it is shown that plastome sequences of these species are relatively conserved with the same gene order, and three Sect. *Palmata* species remarkably showed high sequence similarities. Some hotspots could be used to discriminate the *Rheum* species, and 17 plastid genes were subject to positive selection. The phylogenetic analyses indicated that all the Polygonaceae species were clustered in the same group and showed that *Rheum* species, except for *Rheum wittrockii*, formed a monophyletic group with high maximum parsimony/maximum likelihood bootstrap support values and Bayesian posterior probabilities. The molecular dating based on plastomes indicated that the divergences within Polygonaceae species were dated to the Upper Cretaceous period [73.86–77.99 million years ago (Ma)]. The divergence of Sect. *Palmata* species was estimated to have occurred around 1.60 Ma, indicating that its diversification was affected by the repeated climatic fluctuation in the Quaternary.

**Keywords** *Rheum* · Chloroplast genome · Positive selection · Molecular markers · Phylogeny

---

Communicated by Rachel Wells.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00299-020-02532-0>) contains supplementary material, which is available to authorized users.

---

✉ Xumei Wang  
wangxumei@mail.xjtu.edu.cn

<sup>1</sup> School of Pharmacy, Xi'an Jiaotong University, Xi'an 710061, China

<sup>2</sup> Key Laboratory of Qiyao Resources and Anti-Tumor Activities, Shaanxi Administration of Traditional Chinese Medicine, Xi'an Jiaotong University, Xi'an 710061, China

## Introduction

The genus *Rheum* (Polygonaceae) is widely distributed throughout the temperate and subtropical areas of Asian interior and contains about 60 species (Losina-Losinskaya 1936), 38 of which are found in China, especially in Qinghai–Tibetan Plateau (QTP) (Bao and Grabovskaya-Borodina 2003). *Rheum* species are usually perennial herbs and prefer growing in the mountainous areas at high altitudes ranging from 2000 to 4000 m. Species of this genus evolved diversified morphological traits in response to selection pressures under the harsh distributional environment (Sun et al. 2012). The congeneric relationships of *Rheum* based on the morphological characters were disputed, and the variations in pollen ornamentation of this genus are inconsistent with the morphological classification (Wang et al. 2005; Yang et al. 2001). Although phylogeny of some *Rheum* species was inferred based on chloroplast (cp) DNA fragments, the

species delimitation of this genus still remains challenging due to the paucity of sufficient genetic markers (Sun et al. 2012; Wang et al. 2005). Therefore, more specific genetic markers are needed to infer the phylogeny of *Rheum* species and discriminate them from other related species.

Eight sections have been established and acknowledged under *Rheum* according to the morphological traits (Bao and Grabovskaya-Borodina 2003). Half of *Rheum* species are endemic to China, and many *Rheum* species possess important medicinal properties. For example, three species of Sect. *Palmata*, namely *Rheum palmatum* Linn., *Rheum officinale* Baill. and *Rheum tanguticum* (Maxim. ex Regel) Maxim. ex Balf., are the source plants of rhubarb which is as known as “lord or king of herbs” in China and has been used as an important herb in traditional Chinese medicine for more than 2000 years with the functions including cooling blood, detoxification, removal of blood stasis, removing dampness, abating jaundice, etc. (Chinese Pharmacopoeia Committee 2015). Our previous studies have revealed that the above-mentioned three species could be treated as one species based on genetic and morphological data, and the plasticity of morphological traits is frequently observed among these species due to influences from environmental conditions (Wang et al. 2014, 2018a). Other relatives, such as *Rheum franzenbachii* Munt., *Rheum hotaoense* C. Y. Cheng and T. C. Kao, *Rheum wittrockii* C. E. Lundstr., *Rheum racemiferum* Maxim., *Rheum pumilum* Maxim., *Rheum acuminatum* J. D. Hook. and Thomson and *Rheum przewalskyi* Los. were closely related to Sect. *Palmata*, and *Rheum franzenbachii*, *Rheum hotaoense* and *Rheum wittrockii* were usually regarded as adulterants of rhubarb (Xiao 1981). Previous study indicated that *Rheum* species were originated from a common radiation in the QTP (Sun et al. 2012), which resulted in the complex intraspecific relationships of these species. Therefore, accurate species delimitation of *Rheum* may not only help to resolve the phylogenetic relationships, but to facilitate reliable genetic authentication of medicinal herb in this genus.

Chloroplast is an essential organelle in plant cell that plays a significant role in the process of photosynthesis and carbon fixation. Chloroplast is presumed to originate from cyanobacteria according to endosymbiosis theory (Raven and Allen 2003) and has its own circular genome which usually encodes 110–130 unique genes (Palmer 1985). The cp genome is uniparentally inherited and generally has a quadripartite structure consisting of two copies of inverted repeat (IR) regions divided by one small single-copy (SSC) region and one large single-copy (LSC) region (Bendich 2004). The cp genome is highly conserved compared to nuclear and mitochondrial genomes in terms of gene structure and composition (Asaf et al. 2017). Therefore, a large amount of cpDNA markers were selected for the utility of phylogeny or DNA barcoding. However, commonly used cpDNA markers

may have the limited resolution in species discrimination or phylogenetic analyses for closely related taxa (Dong et al. 2017). The complete cp genomes which contain more hotspot regions with single nucleotide polymorphisms and insertion/deletions (InDels) have been proven to be more informative than cpDNA fragments in inferring evolutionary relationships at different taxonomic levels (Caron et al. 2000; Cho et al. 2015; Wang et al. 2018b; Yang et al. 2016; Yao et al. 2019; Zhou et al. 2018). With the advent of next generation sequencing in recent years, it has become comparatively easy to sequence the complete cp genome of non-model taxa and infer phylogenetic relationships based on whole cp genomes (Guo et al. 2017; Ruhsam et al. 2015; Saarela et al. 2018).

In this study, we characterized the complete cp genomes of eight *Rheum* species, one *Rumex* and one *Oxyria* species. We compared the plastome differences of these species and inferred the phylogeny of Polygonaceae based on the available cp genomes. Our results will be useful for marker development, species discrimination, and the inference of phylogenetic relationships in the genus *Rheum*.

## Materials and methods

### Plant materials and DNA extraction

Ten samples, including eight *Rheum* species, *Rumex crispus* L. and *Oxyria digyna* (Linn.) Hill, were collected from Shaanxi, Gansu, Qinghai, Sichuan, Tibet, Yunnan and Shanxi provinces, China (Table S1). Young leaves were put into silica gel for DNA extraction. Voucher specimens were deposited at the herbarium of Xi'an Jiaotong University, Xi'an, China. Total genomic DNA was isolated using a cetyl trimethyl ammonium bromide protocol (Doyle 1987), and the quantity and quality of the extracted DNA was determined by both agarose gel electrophoresis and a NanoDrop 2000 Spectrophotometer.

### Plastome sequencing, assembly, annotation and validation

The DNA library with an insert size of 270 bp was prepared according to the description by Zhou et al. (2018) and then sequenced on an Illumina HiSeq X Ten platform. The raw reads were filtered to obtain high-quality reads by removing adapters, low-quality sequences (reads with unknown bases “N”), and reads with more than 50% low-quality bases (quality value  $\leq 10$ ) with the NGS QC Toolkit v2.3.3 (Patel and Jain 2012). The clean reads were firstly aligned to the plastome sequences of *Rheum palmatum* (NCBI accession NC\_027728), *Rheum wittrockii* (NCBI accession NC\_035950) and *Rumex acetosa* Linn. (NCBI

accession KC817303) to obtain chloroplast-like sequences using Bowtie v2.2.6 with default parameters (Langmead and Salzberg 2012). The matched paired-end clean reads were initially de novo assembled by SPAdes v3.6.0 (Bankovich et al. 2012). The derived longest contig was selected as seed sequence for further assembly using NOVOPlasty v2.6.2 (Dierckxsens et al. 2017). To validate the accuracy of the assembled plastome, all the clean reads were mapped against the unannotated cp genome in Geneious v10.1 with bowtie2 algorithm (Biomatters Ltd., Auckland, New Zealand). Some ambiguous regions with low coverage were confirmed by PCR-based Sanger sequencing using the primers designed for gap-flanking regions (Table S2). PCR amplifications were performed in a reaction volume of 25  $\mu$ L with 12.5  $\mu$ L 2 $\times$  Taq PCR Master Mix, 0.4  $\mu$ M of each primer, 2  $\mu$ L template DNA and 10.1  $\mu$ L ddH<sub>2</sub>O. All amplifications were performed in SimpliAmp™ Thermal Cycler (Applied Biosystems, Carlsbad, CA, USA) as follow: denaturation at 94 °C for 5 min, followed by 35 cycles of 94 °C for 50 s, at specific annealing temperature for 40 s, 72 °C for 90 s and a final extension at 72 °C for 7 min. PCR products were visualized on a 1.5% agarose gel, and the DNA fragments were sequenced by Sangon Biotech (Shanghai, China). The cp genome was aligned to its reverse complement to determine inverted repeat regions, and the boundaries of the inverted repeats and single-copy regions were also verified by Sanger sequencing (Table S3, Fig. S1). Annotation was primarily conducted using automatic annotator DOGMA (Wyman et al. 2004). The draft annotated results were subsequently inspected and adjusted manually according to the annotation information of reference species in Geneious v10.1. Gene boundaries were manually checked to preserve reading frame and start/stop codons. The circular map of each plastid genome was drawn using the online program OrganellarGenome DRAW (Lohse et al. 2013). The complete plastomes have been submitted to Genbank with accession numbers: MN564922–MN564931.

### Genome comparison, repeat structure analysis and markers identification

Ten newly sequenced plastomes and two previously published plastomes of *Rheum palmatum* and *Rheum wittrockii* were compared and visualized using mVISTA software (Frazer et al. 2004) in a Shuffle-LAGAN mode with the annotation of *Rheum palmatum* as a reference. The alignment of 12 cp genomes was retrieved by MAFFT v7.402 (Katoh and Standley 2013). Sliding window analysis was conducted to analyze DNA polymorphism and nucleotide diversity ( $Pi$ ) with 200 bp step size and 600 bp window length using DnaSP v6.0 (Rozas et al. 2017). Some of the highly divergent non-coding regions were selected to design primers and validated based on PCR-based Sanger

sequencing (same as the previous PCR protocols) after the comparison of these plastomes. IR expansion/contraction and gene distribution at the IR/SC borders of these plastomes were also compared. Tandem repeat sequences were identified by the Tandem Repeats Finder v4.09 (Benson 1999) with the following parameters: 2 for the alignment parameter match and 7 for mismatch and InDels. REPuter program was utilized to identify dispersed and palindromic repeats with a minimum repeat size of 30 bp and sequence identity of no less than 90% (hamming distance equal to 3) (Kurtz et al. 2001). Simple sequence repeats (SSRs) were searched using the MISA Perl script (Thiel et al. 2003) with the following minimum number of repeats: ten for mono-, five for di-, four for tri-, and three for tetra-, penta-, and hexanucleotide SSRs.

### Estimation of substitution and selective pressure analysis

To detect selective pressures on the plastid genes, non-synonymous (dN) and synonymous (dS) substitution rates as well as the dN/dS ratios of each protein-coding gene were calculated using the yn00 program in PAML v4.9 (Yang 2007). The site-specific model was further selected to detect signatures of natural selection using EasyCodeML (Gao et al. 2019) based on the CODEML algorithm (Yang 2007). Six substitution models were used to test each plastid gene, namely M0, M1a, M2a, M3, M7, and M8 models. The site-specific model assumed that  $\omega$  (dN/dS) ratio is the same across branches of the phylogeny but varying among different sites (Gao et al. 2019). Log likelihood values of every model were compared against neutral model by means of a likelihood-ratio test to detect statistical significance.

### Phylogenetic inference

Phylogenetic analysis was conducted based on 28 taxa (Table S4), including 25 Polygonaceae species, and three Plumbaginaceae species (*Plumbago auriculata*, *Cerato stigma willmottianum* and *Limonium tenellum*) were set as outgroups. Ten cp genomes were obtained from the present study, while others were retrieved from NCBI GenBank (accession numbers were summarized in Table S4). All the cp genome sequences were aligned using MAFFT v7.402 with default parameters (Katoh and Standley 2013). The ambiguous and most variable sites in the multiple alignments were removed using Gblocks v0.91b (Talavera and Castresana 2007). To obtain robust phylogenetic relationships, three methods, including maximum parsimony (MP), maximum likelihood (ML) and Bayesian inference (BI), were used to construct phylogenetic trees. MP analysis was performed in PAUP v4.0b10 with 1000 bootstrap replicates (Swofford 1998), and addition-sequence was set as 1000

replications for heuristic search. The ML analysis was conducted using IQ-TREE (Nguyen et al. 2015) with the best best-fit model selected by ModelFinder (Kalyanamoorthy et al. 2017) (Table S5), and the bootstrap replicates were 1000. MP/ML bootstrap support (BS) values were shown at each node. BI analysis was conducted using MrBayes v3.2.6 (Ronquist et al. 2012) with the nucleotide substitution model inferred from Modeltest v3.7 (Table S5). The Markov Chain Monte Carlo (MCMC) algorithm was run for one million generations and sampled every 100 generations with one cold chain and three incrementally heated chains. The first 25% of trees were discarded as burn-in, and the remaining trees were used to build a majority-rule consensus tree with posterior probability values for each node. Likewise, the above-mentioned three phylogenetic-inference methods were used to infer the phylogenetic tree from the 58 shared cp genes (Table S6) with the same settings.

### The estimation of divergence time

Divergence time was estimated using BEAST v2.4.5 under uncorrelated lognormal relaxed clock and birth–death model of speciation (Bouckaert et al. 2014). For BEAST analysis, all the InDel sites in the multiple alignments were removed using Gblocks v0.91b (Talavera and Castresana 2007). GTR + G + I model of nucleotide substitution selected by Modeltest v3.7 was used to construct tree. Due to lack of specific fossil records for *Rheum* species, we used the fossils from other Polygonaceae species and the estimated divergence time in previous studies as the calibration points (Manchester and O’Leary 2010; Yao et al. 2019). All calibration priors were treated as a normal distribution with a mean and a wide SD. The Polygonaceae crown group was constrained to a mean age of 75.7 million years ago (Ma) and SD of 1.07 Ma (Yao et al. 2019). The Rumiceae, Polygonaceae and Fagopyreae clustered in the group which was constrained to a mean age of 66 Ma and SD of 0.5 Ma (Manchester and O’Leary 2010). The MCMC simulation was run for  $2.0 \times 10^{10}$  generations with 10% discarded as burn-in. Tracer v1.7.1 was used for checking the convergence of the chains through sufficient effective sample sizes (ESS) value (> 200) (Rambaut et al. 2018).

## Results

### Plastome features and genome composition

A total of 7,087,845–8,586,533 paired-end reads were retrieved with a sequence length of 150 bp. After filtering and mapping, 255,572–1,088,699 clean reads were successfully mapped to the reference plastome (Table 1). Average sequencing depth ranged from about  $224 \times$  (*Rheum*

**Table 1** Comparison of chloroplast genomic characteristics in 10 *Rheum* species, *Rumex crispus* and *Oxyria digyna*, including aspects of genome size, G-C content, gene number and sequence reads (raw and mapped)

Species	Section	Total Length (bp)	G-C (%)	LSC length (bp)	SSC length (bp)	IR length (bp)	Number of genes	CDS	tRNAs	rRNAs	Raw reads	Mapped reads
<i>Rheum palmatum</i>	<i>Palmata</i>	161,541	37.3	86,518	13,111	30,956	131 (17)	84 (5)	37 (7)	8 (4)	-	-
<i>Rheum tanguticum</i>	<i>Palmata</i>	161,492	37.3	86,480	13,106	30,953	131 (17)	84 (5)	37 (7)	8 (4)	8,313,325	284,003
<i>Rheum officinale</i>	<i>Palmata</i>	161,534	37.3	86,508	13,114	30,956	131 (17)	84 (5)	37 (7)	8 (4)	8,372,685	255,572
<i>Rheum acuminatum</i>	<i>Acuminata</i>	161,306	37.4	86,145	13,169	30,996	131 (17)	84 (5)	37 (7)	8 (4)	8,586,533	470,636
<i>Rheum pumilum</i>	<i>Deseritcola</i>	161,749	37.3	86,997	12,812	30,970	131 (17)	84 (5)	37 (7)	8 (4)	7,578,994	568,781
<i>Rheum racemiferum</i>	<i>Deseritcola</i>	161,688	37.4	87,152	12,822	30,854	131 (17)	84 (5)	37 (7)	8 (4)	8,456,188	907,905
<i>Rheum hotoaense</i>	<i>Rheum</i>	162,048	37.4	86,935	13,155	30,979	131 (17)	84 (5)	37 (7)	8 (4)	8,352,786	1,088,699
<i>Rheum wittrockii</i>	<i>Rheum</i>	159,051	37.6	84,751	12,999	30,651	131 (17)	84 (5)	37 (7)	8 (4)	-	-
<i>Rheum franzenbachii</i>	<i>Rheum</i>	161,688	37.4	86,946	12,784	30,979	131 (17)	84 (5)	37 (7)	8 (4)	8,522,394	811,550
<i>Rheum przewalskyi</i>	<i>Rheum</i>	161,740	37.3	86,542	13,152	31,023	131 (17)	84 (5)	37 (7)	8 (4)	7,126,080	241,140
<i>Rumex crispus</i>	-	161,292	37.4	87,213	13,011	30,534	131 (17)	84 (5)	37 (7)	8 (4)	7,087,845	1,075,989
<i>Oxyria digyna</i>	-	160,698	37.5	85,749	13,172	30,888	131 (17)	84 (5)	37 (7)	8 (4)	8,522,394	516,734

LSC large single-copy region, SSC small single-copy region, IR inverted repeat (A or B) regions, CDS protein-coding genes

Numbers in parentheses mean the number of genes duplicated in IRs

A dash line indicates missing data or not applicable



*przewalskyi*) to  $1,007\times$  (*Rheum hotaoense*) (Table 1). With a combination of de novo and reference guided assemblies, plastomes of eight *Rheum* species, *Rumex crispus* and *O. digyna* were obtained. The size of ten newly sequenced cp genomes ranged from 160,698 bp (*O. digyna*) to 162,048 bp (*Rheum hotaoense*) (Table 1). All the plastomes possess the typical quadripartite structure, consisting of a pair of IRs (30,534–31,023 bp) separated by the LSC (86,145–86,997 bp) and SSC regions (12,784–13,172 bp) (Table 1). The overall GC content among the ten newly sequenced plastomes and the two previously published plastomes of *Rheum palmatum* and *Rheum wittrockii* was very similar (37.3–37.6%). Each plastome identically encoded 131 predicted genes, including 79 protein-coding genes (PCGs), 30 tRNA genes and four rRNA genes; 17 genes in the IR regions were found to be duplicated (Fig. S2, Table 1). Across these cp genomes, *rpl23* was found to be a pseudogene with two copies located in the IR regions, respectively. Three PCGs (*ycf3*, *clpP* and *rps12*) harbored two introns, while nine PCGs (*rps16*, *petD*, *atpF*, *rpl16*, *petB*, *rpl2*, *ndhA*, *ndhB*, *rpoC1*) and six tRNA genes (*trnV-UAC*, *trnL-UAA*, *trnG-UCC*, *trnA-UGC*, *trnI-GAU*, *trnK-UUU*) harbored one intron. No introns were detected from the remaining plastid genes.

### Comparative analyses of plastomes

To infer the cp genome divergence among these 12 species, overall sequence alignments were compared using the annotation of *Rheum palmatum* as a reference (Fig. 1). The aligned results showed that cp genome sequences of these species are relatively conserved with the same gene order. Especially, three Sect. *Palmata* species showed highly similar sequences (Fig. 1). No gene relocation or inversion were detected among these plastomes. However, some highly variable hotspots were found in the intergenic regions. The sliding window analysis also indicated that the divergence in intergenic regions is higher than that in genic regions, and the LSC/SSC regions are more variable than the IR regions (Fig. 2). The most divergent non-coding regions among these cp genomes were *rps16-trnQ*(UUG), *trnS*(GCU)-*trnG*(UCC), *trnR*(UCU)-*atpA*, *atpF-atpH*, *psbM-trnD*(GUC), *trnE*(UUC)-*trnT*(GGU), *trnT*(GGU)-*psbD*, *ycf3-trnS*(GGA), *rps4-trnT*(UGU), *trnL*(UAA) intron, *trnT*(UGU)-*trnL*(UAA), *psaJ-rpl33*, *rpl33-rps18*, *rps18-rpl20* and *clpP* intron1. Some of these variable regions were used to design primers (Table S2) for validating the discriminatory powers, which indicated that the above-mentioned variable hotspots could be utilized as potential genetic markers to discriminate *Rheum* species (Fig. S3).

Additionally, we found that the IR/SC boundary regions were still relatively conserved among these species (Fig. 3). Comparison of these 12 plastomes in Rumiceae revealed

a few expansion and contraction of the IRs. Some genes including *rps19*, *ndhF*, *rps15*, *ycf1*, *rpl2* and *trnH*(GUG) were found in the LSC/IR and SSC/IR borders. Of these genes, *rps15* was found to be 64 bp, 49 bp, 80 bp away from the SSC/IRa border in *Rheum wittrockii*, *O. digyna* and *Rumex crispus*, respectively (Fig. 3), but it was similarly distributed in the remaining species. Except *rps15*, the remaining five genes showed similar distribution status at SC/IR borders in 12 plastomes. Especially, the above-mentioned six genes located in SC/IR borders of three Sect. *Palmata* species were never shifted.

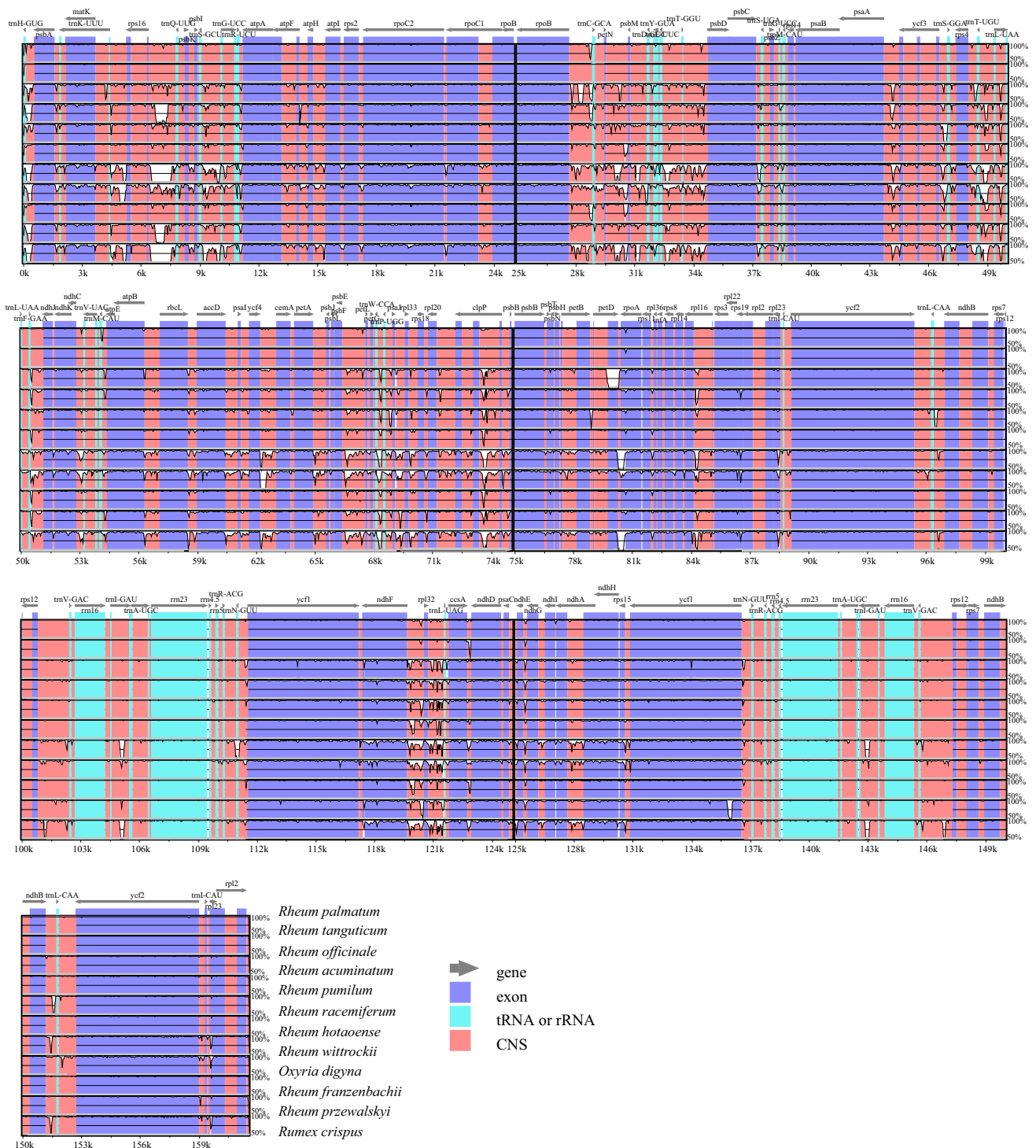
### Identification of SSRs and repeat sequences

A total of 809 SSRs were identified across the cp genomes of 12 Rumiceae species by MISA analysis. The number of SSRs per plastome/species ranged from 53 (*Rheum przewalskyi*) to 77 (*Rheum palmatum*). The majority of SSRs were mononucleotide repeats (423), followed by dinucleotide (178), trinucleotide (111), tetranucleotide (84), pentanucleotide (12) and hexanucleotide (1) repeats (Fig. 4). Mono nucleotide repeat motif (A/T), dinucleotide repeat motif (AT/TA), trinucleotide repeat motif (AAT/ATT) and tetranucleotide repeat motif (mostly AAAT/ATTT or rarely ACAG/CTGT) were present in each plastome (Fig. 4). Most of SSRs were located in non-coding regions, but a few of them were distributed in genic regions (Table S7). Over half of the SSRs were found in the LSC region, whereas low proportions of SSRs were found in the SSC or IR regions.

The analysis of the 12 plastomes using Reputer recognized 523 repeats. The repeat number of each species ranged from 34 (*O. digyna*) to 49 (*Rheum racemiferum* and *Rheum przewalskyi*). Forward repeats are the most abundant types (260), followed by palindromic (235) and reverse (22) repeats (Table S8). The reverse repeats are rare in these 12 cp genomes, and only 22 reverse repeats were found. By searching the tandem repeats with Tandem Repeats Finder, the results indicated that the tandem repeat number of 12 plastomes ranged from 3 (*O. digyna*) to 25 (*Rumex crispus*) with a total number of 116 (Table S9). We found that most repeats were located in the intergenic or intron regions, and only a few repeats were distributed in protein-coding regions (*ycf1*, *ycf2* and *psaA*, *psaB*, *rpl14*) (Table S8, S9).

### Estimation of substitution and selective pressure analysis

The dS values of 76 PCGs between 12 paired Rumiceae species ranged from 0 to 1.4488 with an average value of 0.0361, and dN values ranged from 0 to 3.0185 with an average value of 0.0161 (Table S10). A total of 3709 paired dN/dS values were recovered, most of which were less than 1, indicating that the majority of cp genes

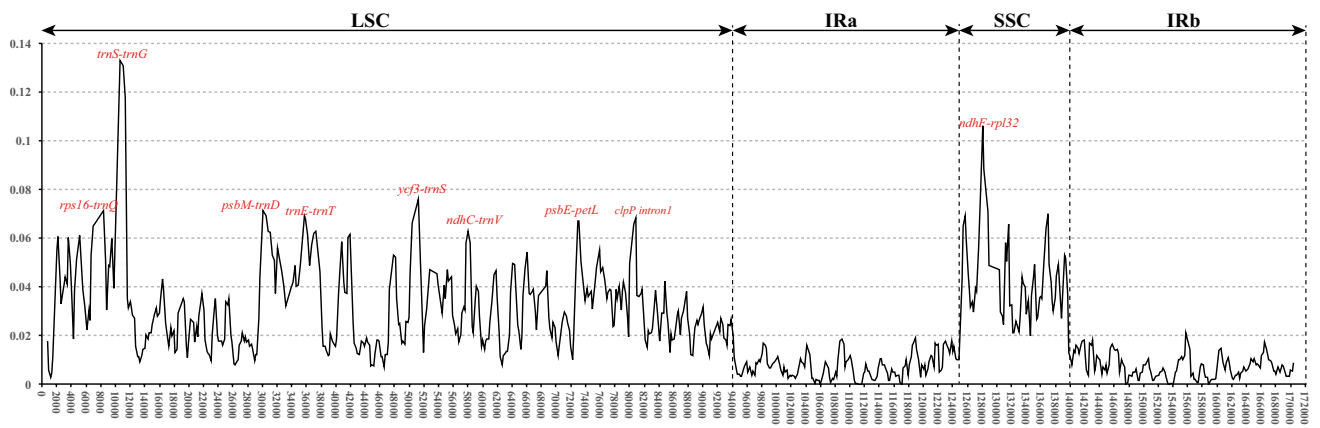


**Fig. 1** Percentages of identity comparing 12 plastomes of 10 *Rheum* species, *Rumex crispus* and *Oxyria digyna* with the *R. palmatum* as the reference (mVISTA). The y axis represents the percent identity

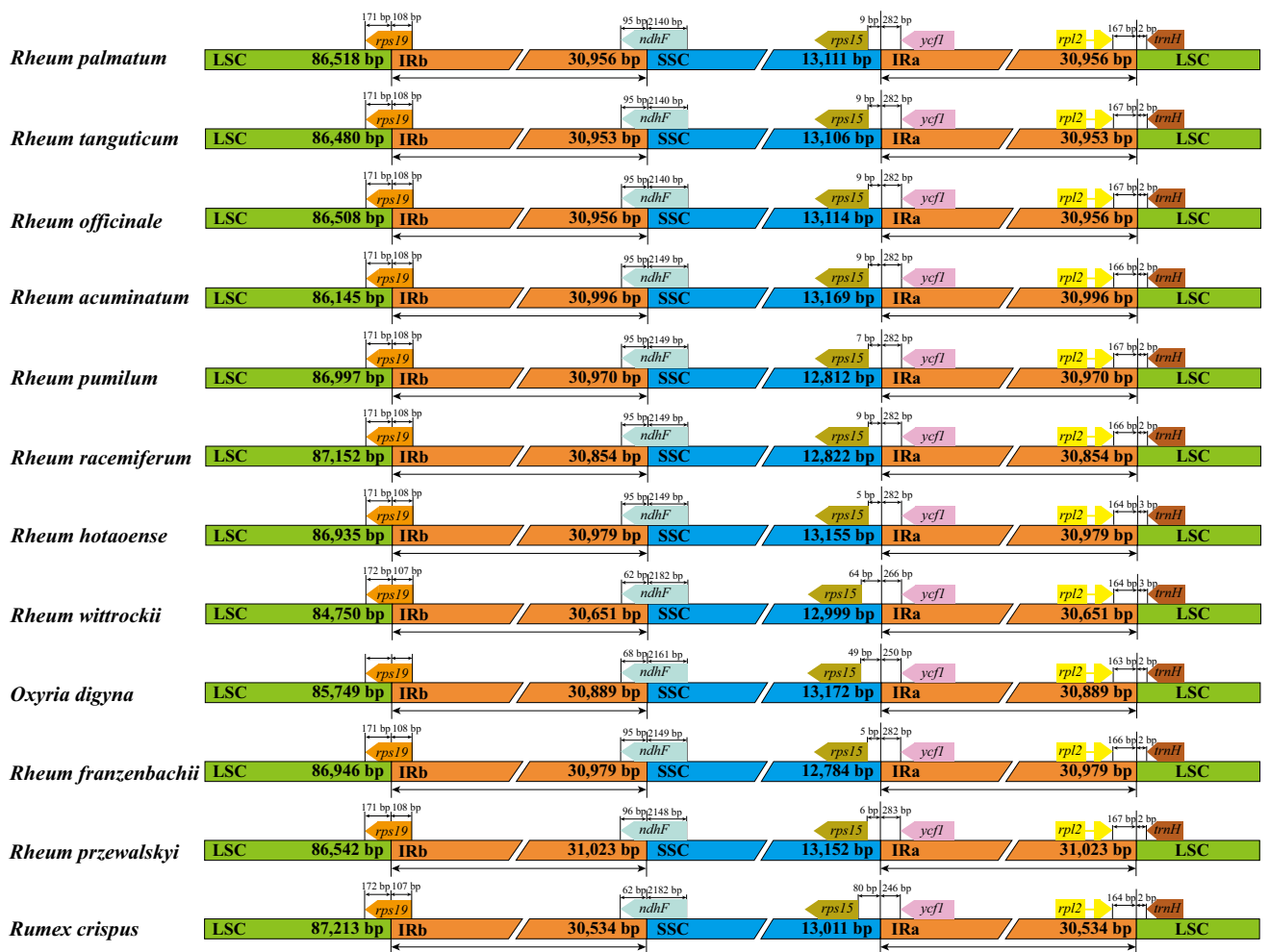
within 50–100%. Genome regions are color-coded as protein-coding (purple), rRNA or tRNA coding genes (blue), and non-coding sequences (pink)

were under purifying selection. Twelve cp genes including *matK*, *ndhB*, *ndhD*, *ndhF*, *ndhG*, *psaJ*, *psbL*, *rpoA*, *cemA*, *rpl32*, *ycf1* and *ycf2* were detected with dN/dS values > 1, indicating that these genes had undergone

positive selection. Based on the site-specific model of CODEML, nine genes (*atpE*, *ndhD*, *psaJ*, *psbC*, *rpl16*, *rpoC2*, *rps3*, *ycf1* and *ycf2*) with positively selected sites

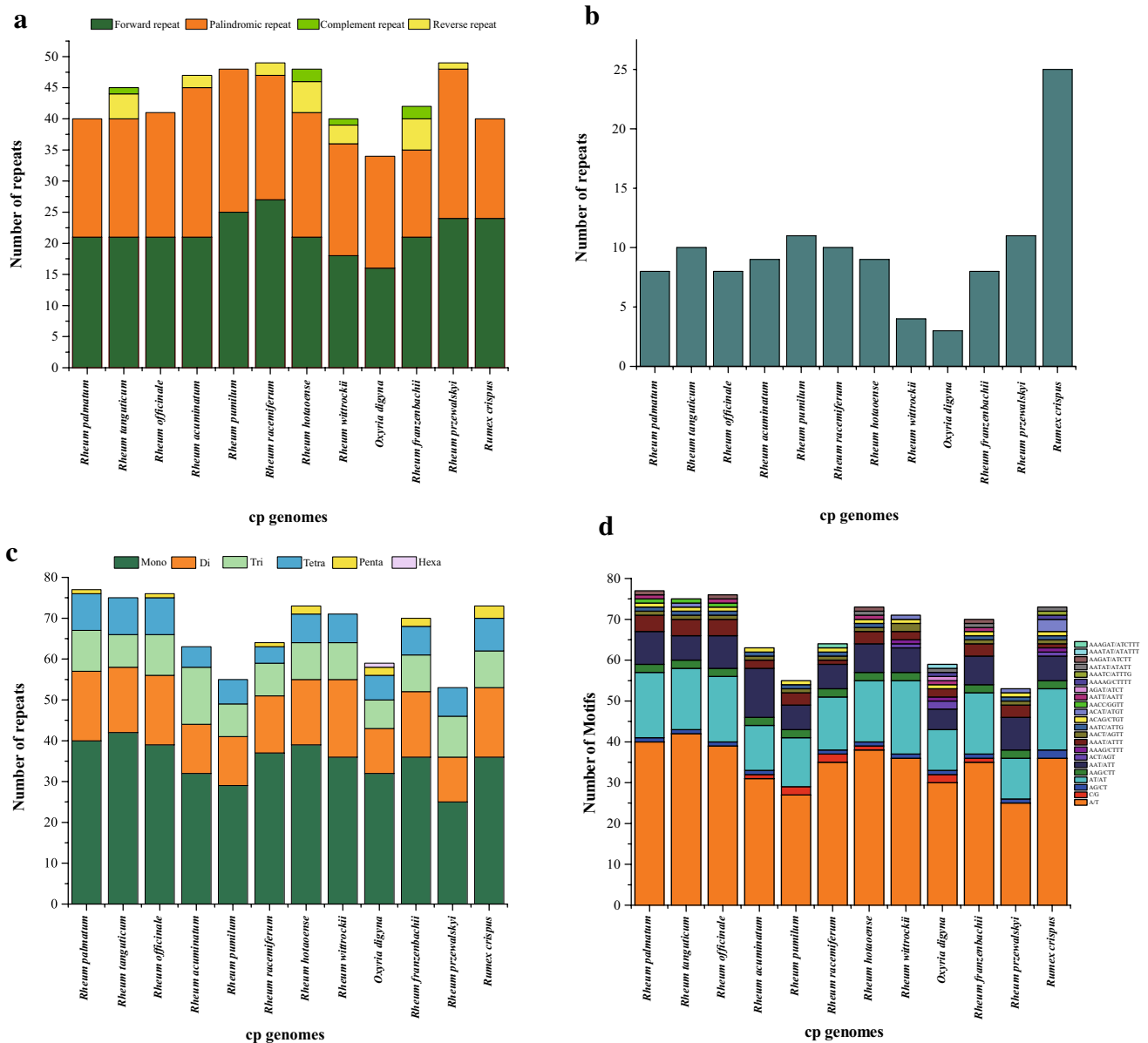


**Fig. 2** Nucleotide diversity (Pi) in the complete cp genomes of 10 *Rheum* species, *Rumex crispus* and *Oxyria digyna*. Sliding window analysis with a window length of 600 bp and a step size of 200 bp



**Fig. 3** The IR/SSC borders in the plastomes of 10 *Rheum* species, *Rumex crispus* and *Oxyria digyna*

were identified (Table S11). By combining the results of



**Fig. 4** Statistics of repeat elements in the plastomes of 10 *Rheum* species, *Rumex crispus* and *Oxyria digyna*. **a** Number of forward, palindromic, complement and reverse repeats. **b** Number of tandem

repeats. **c** Number of different simple sequence repeat (SSR) types detected in 12 plastomes. **d** Frequency of identified SSR motifs in 12 plastomes

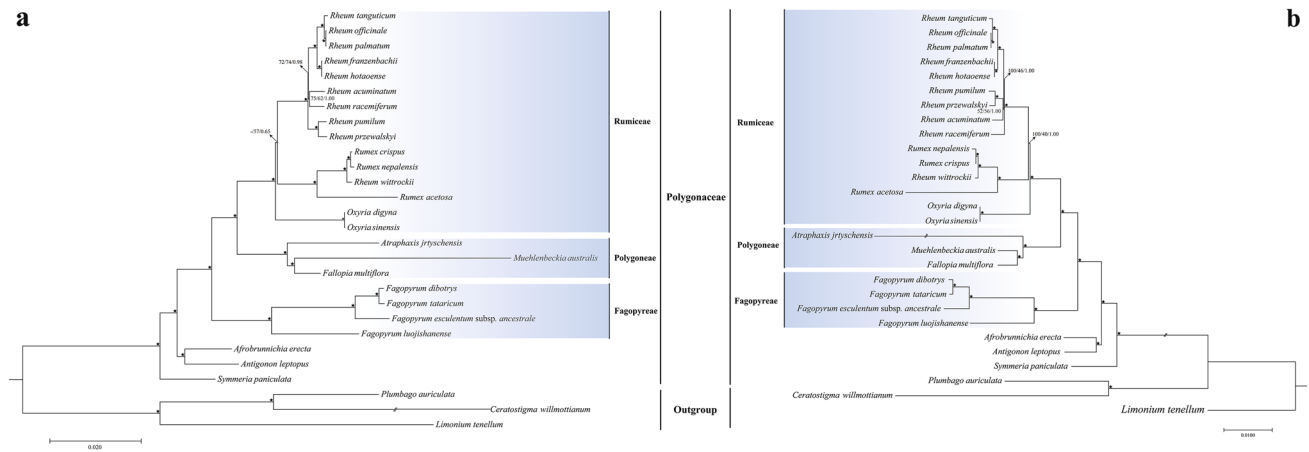
YN and CODEML algorithm, we found that a total of 17 cp genes were under positive selection.

### Phylogenetic inference

Three phylogenetic methods (MP/ML/BI) yielded almost identical topologies with generally high support values. The monophyly of Polygonaceae was strongly supported based on the available cp genome dataset (MP/ML, BS = 100; BI, PP = 1) (Fig. 5a). Species belonging to

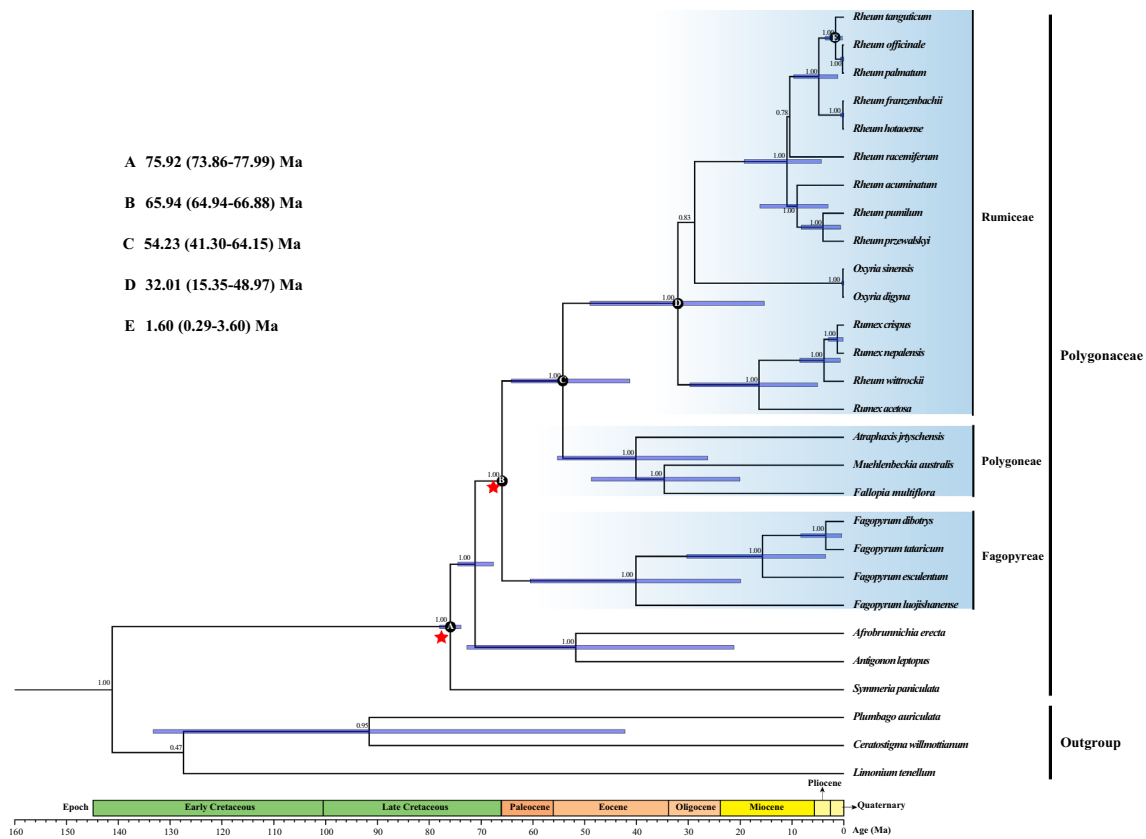
three tribes including Rumiceae, Polygoneae and Fagopyreae were also clustered in the same clade, respectively. Except for *Rheum wittrockii*, which was clustered with *Rumex* species in the same clade, other *Rheum* species formed a monophyletic group with high BS and PP values. This result also confirmed the monophyly of Sect. *Palmata*. The phylogenetic analysis also produced the similar result based on 58 shared CDS regions with three different methods (Fig. 5b).





**Fig. 5** Phylogenetic tree of 28 taxa using maximum likelihood, maximum parsimony and Bayesian inference based on dataset of the 58 shared genes and complete plastome sequence. **a** The phylogenetic tree inferred from complete plastome sequence dataset. **b** The phylo-

genetic tree inferred from the dataset of 58 shared genes. ML topology was shown with ML bootstrap value/MP bootstrap value /Bayesian posterior probability at each node. The pentastar indicates that the support rate of branch is 100/100/1.0



**Fig. 6** BEAST-derived chronograms of Polygonaceae based on the plastome sequences with two calibration points (red pentastar) derived from previous study. Blue bars indicate the 95% highest pos-

terior density (HPD) credibility intervals for node ages (Ma). Posterior probabilities are sequentially labelled above nodes. Mean divergence dates for major nodes (**a–e**) are labeled

## Estimation of divergence times

Using three Plumbaginaceae species as outgroups and the phylogeny generated here, divergence times of the major clades within Polygonaceae were estimated based on the dataset comprised of 28 plastome sequences (Fig. 6). Our analyses based on the two combined MCMC runs produced sufficient effective sample sizes (> 200) for all relevant parameters and suggested adequate sampling of the posterior distribution. The results suggest that the splits in crown group Polygonaceae were dated to the Upper Cretaceous (especially in the Campanian period, 73.86–77.99 Ma). The divergence between Rumiceae and Polygonaceae was estimated as *c.* 54.23 Ma (41.30–64.15 Ma). The divergences within Rumiceae clade were dated to Oligocene (32.01 Ma, 15.35–48.97 Ma). The divergence time of Sect. *Palmata* clade was estimated to have occurred as *c.* 1.60 Ma.

## Discussion

The ten newly obtained cp genomes in the present study not only offered an opportunity to comprehensively compare cp genome sequences of Rumiceae, but also provided sufficient genetic resources for species discrimination and phylogenetic utility in *Rheum*. The cp genomes display the typical quadripartite structure with a LSC and a SSC region which are separated by two IRs. All the analyzed *Rheum* species exhibit 131 predicted genes and share the same gene content and gene synteny. Major structural changes such as gene loss and genome rearrangement have been reported in several angiosperm lineages (Cai et al. 2008; Weng et al. 2013; Wicke et al. 2016), but such changes were not found in the plastomes of *Rheum*. Especially, the overall cp genome structure of Sect. *Palmata* species was extremely conserved, which, to some extent, confirmed the close relationships of these three species.

The sizes of plastid genomes vary considerably in different angiosperms (63–242 kb) and even show high discrepancy in the same family (46–190 kb in Orobanchaceae) (Frailey et al. 2018; Gruenstaeudl et al. 2017; Wicke et al. 2013). On the contrary, *Rheum* species show minor difference in plastome sizes, from 159,051 bp in *Rheum wittrockii* to 162,048 bp in *Rheum hotaoense*. It was reported that the size increases of plastomes are usually caused by the expansions of the IR regions (Weng et al. 2013; He et al. 2017). Of compared *Rheum* plastomes, *Rheum wittrockii* had the shortest length of IRs, while *Rheum hotaoense* had the longest. It is similarly presumed that the differences in plastome size recorded in the present study are mainly the result of extractions or contractions of the IR regions. Genes located in SC/IR borders of *Rheum* plastomes were identical, and only minor length differences were detected between these genes

and SSC/IR border. As *Rheum* species with close interspecific relationships mainly derived from a rapid radiation, we deduced that the highly conserved nature of cp genome resulted in the similar gene distributions at SC/IR junctions.

Although cp genome is highly conserved, some variable hotspots that include InDels have been detected (Aldrich et al. 1988). These hotspots in plastomes may provide several highly variable cpDNA markers. Species of Sect. *Palmata* are important in traditional Chinese medicines; therefore, the accurate species discrimination of *Rheum* is vital for the utilization of rhubarb. However, morphological similarities have further hampered the authentication of rhubarb species. Thus the molecular identification based on DNA markers will provide a more effective way to overcome these problems. The plastome has a conserved sequence length from 110 to 160 kb, which far exceeds the length of commonly used molecular markers and can provide more variation to distinguish the closely related species (Li et al. 2015; Nguyen et al. 2017). Therefore, some mutation hotspot regions could be tested as *Rheum* specific DNA markers and used to discriminate the rhubarb from its adulterants. These regions might also provide sufficient genetic variations for resolving the phylogenetic relationships of *Rheum* species.

Repeat elements detected in plastomes have been proven to be correlated with rearrangement, sequence divergence and recombination (Asano et al. 2004; Timme et al. 2007; Weng et al. 2013). The numbers and distributions of tandem, dispersed (forward and reverse), and palindromic repeats were surveyed in this study. Our results indicated that the repeat distribution status is similar in cp genomes of *Rheum*. Interestingly, most repeats were found in the intergenic or intron regions, and a few repeats were distributed in same gene regions (*ycf1*, *ycf2*) or gene with similar functions (*psaA*, *psaB*). Chloroplast simple sequence repeats (cpSSRs) are usually variable within the same species and have been proven to be an important molecular marker for species discrimination and population genetics at lower taxonomic levels (Provan et al. 2001; Ruhsam et al. 2015; Xue et al. 2012). The SSRs distributed in *Rheum* plastomes were similar, with the mononucleotide (A/T) being the most abundant repeat type. Poly (A)/(T) SSRs are pervasively found in plant cp genomes (Wang et al. 2018b; Yang et al. 2016; Zhou et al. 2019). Most cpSSRs were found in non-coding regions, but only a few SSRs were detected in coding regions. CpSSRs located in non-coding regions usually show high intraspecific variation in repeat numbers (Eguiluz et al. 2017). Therefore, the cpSSRs will provide more valuable genetic resources for the species identification and population genetics of *Rheum*.

Previous studies have revealed signatures of natural (purifying or positive) selection in some cp gene regions (e.g. *psbA*, *matK*, *rbcL*) encoding proteins involved in photosynthesis (Carbonell-Caballero et al. 2015; Ye et al. 2018). We

found 17 protein-coding genes (*atpE*, *matK*, *ndhB*, *ndhD*, *ndhF*, *ndhG*, *psaJ*, *psbL*, *psbC*, *rpoA*, *rpoC2*, *cemA*, *rpl16*, *rpl32*, *rps3*, *ycf1* and *ycf2*) being under the positive selection. Of these, *matK* with sufficient variant sites has been used as a standard DNA barcode for species discrimination, and it was highly divergent in Caryophyllales (Cuénoud et al. 2002). Eleven *ndh* genes in plant cp genome encode NAD(P)H dehydrogenase (NDH) complex which is essential for photosystem I cyclic electron transport and chlororespiration (Kofer et al. 1998). As NDH monomer is sensitive to high light intensity, we inferred that the *ndh* genes might have changed greatly to generate new functions for the stress resistance, and previous studies also showed the similar results (Peng et al. 2011; Wang et al. 2018b; Yang et al. 2016). *Psa* and *psb* genes are primary members of photosystem which may evolve rapidly in some Rumiceae species. Plastid genes, including *rpoA* and *rpoC2*, encoding proteins involved in transcription and post-transcriptional modification have been found to evolve under positive selection (Piot et al. 2018). *CemA* gene is related to the synthesis of PPR7 protein and may have coevolved with nuclear genes (Jalal et al. 2015). Therefore, we presumed that *cemA* gene may have a fast evolution rate in Rumiceae species, and some previous studies on other green plants also obtained the similar result (Xu et al. 2015; Zhou et al. 2016). It has been suggested that *rpl* and *rps* encode ribosomal proteins that have more divergent sequences than proteins related to photosynthesis (Xu et al. 2015). *Ycf1* and *ycf2* are two of the largest genes encoding a putative membrane protein (Cuénoud et al. 2002; Kikuchi et al. 2013) and have rapidly evolved in several species (Cho et al. 2015; Park et al. 2018; Wang et al. 2018b; Yang et al. 2016; Zhou et al. 2019).

The plastome has been proven to be the most important genetic resource for inferring the phylogeny of green plant and disentangling phylogenetic relationships of species that have experienced rapid radiations (Davis et al. 2014; Li et al. 2019; Ma et al. 2014; Yao et al. 2019). *Rheum* species were originated from a common radiation in the QTP (Sun et al. 2012) and shared the similar morphological traits. Phylogenetic relationships of *Rheum* species were inferred based on the available plastomes using MP, ML and Bayesian methods. Most *Rheum* species were clustered in the same clade with a high BS value, which was compatible with the previous phylogeny inferred from cpDNA fragments (Sun et al. 2012). However, *Rheum wittrockii* was clustered with *Rumex* species. Unexpectedly, the overall genomic structure of *Rheum wittrockii* was much more different from the congeneric *Rheum* species. Therefore, high sequence divergence of cp genome resulted in a discordant phylogenetic position. The phylogenetic result also showed that *Rheum palmatum*, *Rheum officinale* and *Rheum tanguticum* were constantly clustered in the same clade with a high resolution value, and the plastome sequences of these three species also showed

the high similarity. Previous molecular and morphological data indicated that these three species should be treated as one species complex (Wang et al. 2018a). Therefore, our study further confirmed the close relationships of species belonging to Sect. *Palmata*. In addition, based on the phylogenetic trees, species from three tribes (Rumiceae, Polygonaceae and Fagopyreae) separately formed monophyletic groups with high support values, indicating that cp genomes are suitable for resolving the phylogeny of Polygonaceae. However, limited taxon sampling may provide insufficient phylogenetic information which may lead to a discrepant tree topology (Eguiluz et al. 2017; Leebens-Mack et al. 2005). Therefore, sufficient samples should be recovered to obtain a more reliable inference of the phylogenetic relationships of Polygonaceae.

Based on the available plastome dataset, we estimated the divergence time of these Polygonaceae species. The ages of the major Polygonaceae splits are in agreement with the previously published result (Yao et al. 2019). Most *Rheum* species mainly diverged in the middle Miocene to Pliocene. It has been proven that the uplift of the eastern edge of the QTP during the Miocene and Late Pliocene facilitated the radiations of species (Sun et al. 2011; Wen et al. 2014). Therefore, the divergence and diversification of *Rheum* species might have been affected by such important tectonic events, which was consistent with the previous phylogenetic study (Sun et al. 2012). According to our results, the earliest divergence of Sect. *Palmata* group was dated to Pleistocene. It can be inferred that the repeated climatic fluctuation in the Quaternary promote the divergence of these three species, which is compatible with the previous studies based on plastid and nuclear markers (Sun et al. 2012; Wang et al. 2018a).

## Conclusions

The complete plastomes of eight *Rheum* species, one *Rumex* and one *Oxyria* species were sequenced and compared. All the plastomes showed identical structure with the same gene order, and no gene relocation or inversion were detected among these plastomes. However, some highly variable hotspots were found in the intergenic regions which provided candidate genetic markers for species authentication and phylogeny for *Rheum*. Besides, based on comparative cp genome analyses, repeat elements were detected from the plastomes, and the abundant SSRs could be used to the species discrimination and population genetics of *Rheum*. We found that most plastid genes have undergone purifying selection, and 17 genes were subjected to positive selection. The phylogenetic analyses further confirmed the monophyly of Polygonaceae based on the available cp genomes, and also indicated that plastomes could facilitate to reconstruct the phylogeny of Polygonaceae. According to the molecular

dating based on plastome sequences, the divergence time of three tribes of Polygonaceae was estimated. We confirmed that the diversification of Sect. *Palmata* was caused by fluctuant climate in the Quaternary.

**Acknowledgements** This study was supported by the National Natural Science Foundation of China (Nos. 31770364 and 81903739) and the China Postdoctoral Science Foundation (No. 2018M643680).

**Author contribution statement** XW and TZ conceived and designed the experiments. TZ, HZ, JW, YX, and FX performed the experiments and analyzed the data. TZ and XW wrote the paper. All authors read and approved the final manuscript.

## Compliance with ethical standards

**Conflict of interest** All authors did not have any conflict of interest.

**Ethical approval** This article does not contain any studies with human participants or animals performed by any of the authors.

## References

- Aldrich J, Cherney BW, Merlin E (1988) The role of insertions/deletions in the evolution of the intergenic region between *psbA* and *trnH* in the chloroplast genome. *Curr Genet* 14:137–146. <https://doi.org/10.1007/bf00569337>
- Asaf S et al (2017) Comparative analysis of complete plastid genomes from wild soybean (*Glycine soja*) and nine other *Glycine* species. *PLoS ONE* 12:e0182281. <https://doi.org/10.1371/journal.pone.0182281>
- Asano T, Tsudzuki T, Takahashi S, Shimada H, Kadowaki K (2004) Complete nucleotide sequence of the sugarcane (*Saccharum officinarum*) chloroplast genome: a comparative analysis of four monocot chloroplast genomes. *DNA Res* 11:93–99. <https://doi.org/10.1093/dnares/11.2.93>
- Bankevich A et al (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>
- Bao B, Grabovskaya-Borodina A (2003) *Rheum*, vol 5. Flora of China. Science Press, Beijing
- Bendich AJ (2004) Circular chloroplast chromosomes: the grand illusion. *Plant Cell* 16:1661–1666. <https://doi.org/10.1105/tpc.160771>
- Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27:573. <https://doi.org/10.1093/nar/27.2.573>
- Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol*. <https://doi.org/10.1371/journal.pcbi.1003537>
- Cai Z et al (2008) Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. *J Mol Evol* 67:696–704. <https://doi.org/10.1007/s00239-008-9180-7>
- Carbonell-Caballero J, Alonso R, Ibañez V, Terol J, Talon M, Dopazo J (2015) A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus *Citrus*. *Mol Biol Evol* 32:2015–2035. <https://doi.org/10.1093/molbev/msv082>
- Caron H, Dumas S, Marque G, Messier C, Bandou E, Petit RJ, Kremer A (2000) Spatial and temporal distribution of chloroplast DNA polymorphism in a tropical tree species. *Mol Ecol* 9:1089–1098. <https://doi.org/10.1046/j.1365-294x.2000.00970.x>
- Cho K-S, Yun B-K, Yoon Y-H, Hong S-Y, Mekapogu M, Kim K-H, Yang T-J (2015) Complete chloroplast genome sequence of tartary buckwheat (*Fagopyrum tataricum*) and comparative analysis with common buckwheat (*F. esculentum*). *PLoS ONE* 10:e0125332. <https://doi.org/10.1371/journal.pone.0125332>
- Chinese Pharmacopoeia Committee CP (2015) Pharmacopoeia of the People's Republic of China. China Medical Science Press, Beijing
- Cuénoud P, Savolainen V, Chatrou LW, Powell M, Grayer RJ, Chase MW (2002) Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbcl*, *atpB*, and *matK* DNA sequences. *Am J Bot* 89:132–144. <https://doi.org/10.3732/ajb.89.1.132>
- Davis CC, Xi Z, Mathews S (2014) Plastid phylogenomics and green plant phylogeny: almost full circle but not quite there. *BMC Biol* 12:11. <https://doi.org/10.1186/1741-7007-12-11>
- Dierckxsens N, Mardulyn P, Smits G (2017) NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res* 45:e18–e18. <https://doi.org/10.1093/nar/gkw955>
- Dong W et al (2017) Phylogenetic resolution in *Juglans* based on complete chloroplast genomes and nuclear DNA sequences. *Front Plant Sci* 8:1148. <https://doi.org/10.3389/fpls.2017.01148>
- Doyle JJ (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19:11–15
- Eguiluz M, Rodrigues NF, Guzman F, Yuyama P, Margis R (2017) The chloroplast genome sequence from *Eugenia uniflora*, a Myrtaceae from Neotropics. *Plant Syst Evol*. <https://doi.org/10.1007/s00606-017-1431-x>
- Frailely DC, Chaluvadi SR, Vaughn JN, Coatney CG, Bennetzen JL (2018) Gene loss and genome rearrangement in the plastids of five Hemiparasites in the family Orobanchaceae. *BMC Plant Biol* 18:30. <https://doi.org/10.1186/s12870-018-1249-x>
- Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comparative genomics. *Nucleic Acids Res* 32:W273–W279. <https://doi.org/10.1093/nar/gkh458>
- Gao F, Chen C, Arab DA, Du Z, He Y, Ho SYW (2019) EasyCodeML: a visual tool for analysis of selection using CodeML. *Ecol Evol* 9:3891–3898. <https://doi.org/10.1002/ece3.5015>
- Gruenstaedl M, Nauheimer L, Borsch T (2017) Plastid genome structure and phylogenomics of Nymphaeales: conserved gene order and new insights into relationships. *Plant Syst Evol*. <https://doi.org/10.1007/s00606-017-1436-5>
- Guo X et al (2017) Plastome phylogeny and early diversification of Brassicaceae. *BMC Genomics* 18:176. <https://doi.org/10.1186/s12864-017-3555-3>
- He L, Qian J, Li X, Sun Z, Xu X, Chen S (2017) Complete chloroplast genome of medicinal plant *Lonicera japonica*: genome rearrangement, intron gain and loss, and implications for phylogenetic studies. *Molecules* 22:249. <https://doi.org/10.3390/molecules22020249>
- Jalal A, Schwarz C, Schmitz-Linneweber C, Vallon O, Nickelsen J, Bohne A-V (2015) A small multifunctional pentatricopeptide repeat protein in the chloroplast of *Chlamydomonas reinhardtii*. *Mol Plant* 8:412–426. <https://doi.org/10.1016/j.molp.2014.11.019>
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587. <https://doi.org/10.1038/nmeth.4285>
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>
- Kikuchi S et al (2013) Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science* 339:571–574. <https://doi.org/10.1126/science.1229262>



- Kofer W, Koop HU, Wanner G, Steinmüller K (1998) Mutagenesis of the genes encoding subunits A, C, H, I, J and K of the plastid NAD(P)H-plastoquinone-oxidoreductase in tobacco by polyethylene glycol-mediated plastome transformation. *Mol Gen Genet* 258:166–173. <https://doi.org/10.1007/s004380050719>
- Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res* 29:4633–4642. <https://doi.org/10.1093/nar/29.22.4633>
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>
- Leebens-Mack J et al (2005) Identifying the basal angiosperm node in chloroplast genome phylogenies: sampling one's way out of the felsenstein zone. *Mol Biol Evol* 22:1948–1963. <https://doi.org/10.1093/molbev/msi191>
- Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S (2015) Plant DNA barcoding: from gene to genome. *Biol Rev* 90:157–166. <https://doi.org/10.1111/brv.12104>
- Li H-T et al (2019) Origin of angiosperms and the puzzle of the Jurassic gap. *Nat Plants* 5:461–470. <https://doi.org/10.1038/s41477-019-0421-0>
- Lohse M, Drechsel O, Kahlau S, Bock R (2013) OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res* 41:W575–W581. <https://doi.org/10.1093/nar/gkt289>
- Losina-Losinskaya A (1936) The genus *Rheum* and its species. *Acta Instituti Botanici Academiae Scientiarum Unionis Rerum Publicarum Sovieticarum Socialisticarum, Ser* 1:5–141
- Ma P-F, Zhang Y-X, Zeng C-X, Guo Z-H, Li D-Z (2014) Chloroplast phylogenomic analyses resolve deep-level relationships of an intractable bamboo tribe Arundinarieae (Poaceae). *Systematic Biol* 63:933–950. <https://doi.org/10.1093/sysbio/syu054>
- Manchester SR, O'Leary EL (2010) Phylogenetic distribution and identification of fin-winged fruits. *Bot Rev* 76:1–82. <https://doi.org/10.1007/s12229-010-9041-0>
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274. <https://doi.org/10.1093/molbev/msu300>
- Nguyen VB, Park H-S, Lee S-C, Lee J, Park JY, Yang T-J (2017) Authentication markers for five major *Panax* species developed via comparative analysis of complete chloroplast genome sequences. *J Agric Food Chem* 65:6298–6306. <https://doi.org/10.1021/acs.jafc.7b00925>
- Palmer JD (1985) Comparative organization of chloroplast genomes. *Annu Rev Genet* 19:325–354. <https://doi.org/10.1146/annurev.ev.ge.19.120185.001545>
- Park I, Yang S, Kim WJ, Noh P, Lee HO, Moon BC (2018) The complete chloroplast genomes of six *Ipomoea* species and indel marker development for the discrimination of authentic *Pharbitidis* Semen (Seeds of *I. nil* or *I. purpurea*). *Front Plant Sci* 9:965. <https://doi.org/10.3389/fpls.2018.00965>
- Patel RK, Jain M (2012) NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS ONE* 7:e30619. <https://doi.org/10.1371/journal.pone.0030619>
- Peng L, Yamamoto H, Shikanai T (2011) Structure and biogenesis of the chloroplast NAD(P)H dehydrogenase complex. *BBA-bioenergetics* 1807:945–953. <https://doi.org/10.1016/j.bbabi.2010.10.015>
- Piot A, Hackel J, Christin P-A, Besnard G (2018) One-third of the plastid genes evolved under positive selection in PACMAD grasses. *Planta* 247:255–266. <https://doi.org/10.1007/s00425-017-2781-x>
- Provan J, Powell W, Hollingsworth PM (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends Ecol Evol* 16:142–147. [https://doi.org/10.1016/S0169-5347\(00\)02097-8](https://doi.org/10.1016/S0169-5347(00)02097-8)
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior summarization in bayesian phylogenetics using Tracer 1.7. *Syst Biol* 67:901–904. <https://doi.org/10.1093/sysbio/syy032>
- Raven JA, Allen JF (2003) Genomics and chloroplast evolution: what did cyanobacteria do for plants? *Genome Biol* 4:209. <https://doi.org/10.1186/gb-2003-4-3-209>
- Ronquist F et al (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61:539–542. <https://doi.org/10.1093/sysbio/sys029>
- 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. *Mol Biol Evol* 34:3299–3302. <https://doi.org/10.1093/molbev/msx248>
- Ruhsam M et al (2015) Does complete plastid genome sequencing improve species discrimination and phylogenetic resolution in *Araucaria*? *Mol Ecol Resour* 15:1067–1078. <https://doi.org/10.1111/1755-0998.12375>
- Saarela JM et al (2018) A 250 plastome phylogeny of the grass family (Poaceae): topological support under different data partitions. *PeerJ* 6:e4299. <https://doi.org/10.7717/peerj.4299>
- Sun B-N et al (2011) Reconstructing Neogene vegetation and climates to infer tectonic uplift in western Yunnan, China. *Palaeogeogr Palaeoclimatol Palaeoecol* 304:328–336. <https://doi.org/10.1016/j.palaeo.2010.09.023>
- Sun Y, Wang A, Wan D, Wang Q, Liu J (2012) Rapid radiation of *Rheum* (Polygonaceae) and parallel evolution of morphological traits. *Mol Phylogenet Evol* 63:150–158. <https://doi.org/10.1016/j.ympev.2012.01.002>
- Swofford DL (1998) Commands used in the PAUP Block in PAUP 4.0: phylogenetic analysis using parsimony. Smithsonian Institution, Washington, DC, pp 132–135. [https://doi.org/10.1007/978-1-4020-6754-9\\_12413](https://doi.org/10.1007/978-1-4020-6754-9_12413)
- Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 56:564–577. <https://doi.org/10.1080/10635150701472164>
- Thiel T, Michalek W, Varshney R, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet* 106:411–422. <https://doi.org/10.1007/s00122-002-1031-0>
- Timme RE, Kuehl JV, Boore JL, Jansen RK (2007) A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: identification of divergent regions and categorization of shared repeats. *Am J Bot* 94:302–312. <https://doi.org/10.3732/ajb.94.3.302>
- Wang A, Liu J, Yang M (2005) Molecular phylogeny, recent radiation and evolution of gross morphology of the rhubarb genus *Rheum* (Polygonaceae) Inferred from chloroplast DNA *trnL-F* sequences. *Ann Bot* 96:489–498. <https://doi.org/10.1093/aob/mci201>
- Wang X-M, Hou X-Q, Zhang Y-Q, Li Y (2014) Morphological variation in leaf dissection of *Rheum palmatum* complex (Polygonaceae). *PLoS ONE* 9:e110760. <https://doi.org/10.1371/journal.pone.0110760>
- Wang X et al (2018a) Genetic and chemical differentiation characterizes top-geoherb and non-top-geoherb areas in the TCM herb rhubarb. *Sci Rep* 8:9424. <https://doi.org/10.1038/s41598-018-27510-1>
- Wang X, Zhou T, Bai G, Zhao Y (2018b) Complete chloroplast genome sequence of *Fagopyrum dibotrys*: genome features, comparative analysis and phylogenetic relationships. *Sci Rep* 8:12379. <https://doi.org/10.1038/s41598-018-30398-6>
- Wen J, Zhang J-Q, Nie Z-L, Zhong Y, Sun H (2014) Evolutionary diversifications of plants on the Qinghai-Tibetan Plateau. *Front Genet* 5:4. <https://doi.org/10.3389/fgene.2014.00004>



- Weng M-L, Blazier JC, Govindu M, Jansen RK (2013) Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats and nucleotide substitution rates. *Mol Biol Evol* 31:645–659. <https://doi.org/10.1093/molbev/mst257>
- Wicke S et al (2013) Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. *Plant Cell* 25:3711–3725. <https://doi.org/10.1105/tpc.113.113373>
- Wicke S, Müller KF, dePamphilis CW, Quandt D, Bellot S, Schneeweiss GM (2016) Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. *Proc Natl Acad Sci* 113:9045–9050. <https://doi.org/10.1073/pnas.1607576113>
- Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20:3252–3255. <https://doi.org/10.1093/bioinformatics/bth352>
- Xiao PG (1981) The resource utilization of medicinal plant of *Rheum*. *China J Chin Mat Med* 6:11–13
- Xue J, Wang S, Zhou S-L (2012) Polymorphic chloroplast microsatellite loci in *Nelumbo* (Nelumbonaceae). *Am J Bot* 99:e240–e244. <https://doi.org/10.3732/ajb.1100547>
- Xu J-H, Liu Q, Hu W, Wang T, Xue Q, Messing J (2015) Dynamics of chloroplast genomes in green plants. *Genomics* 106:221–231. <https://doi.org/10.1016/j.ygeno.2015.07.004>
- Yang M, Zhang D, Zheng J, Liu J (2001) Pollen morphology and its systematic and ecological significance in *Rheum* (Polygonaceae) from China. *Nord J Bot* 21:411–418. <https://doi.org/10.1111/j.1756-1051.2001.tb00789.x>
- Yang Y, Zhou T, Duan D, Yang J, Feng L, Zhao G (2016) Comparative analysis of the complete chloroplast genomes of five *Quercus* species. *Front Plant Sci* 7:959. <https://doi.org/10.3389/fpls.2016.00959>
- Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591. <https://doi.org/10.1093/molbev/msm088>
- Yao G et al (2019) Plastid phylogenomic insights into the evolution of Caryophyllales. *Mol Phylogenet Evol* 134:74–86. <https://doi.org/10.1016/j.ympev.2018.12.023>
- Ye W-Q, Yap Z-Y, Li P, Comes HP, Qiu Y-X (2018) Plastome organization, genome-based phylogeny and evolution of plastid genes in Podophylloideae (Berberidaceae). *Mol Phylogenet Evol* 127:978–987. <https://doi.org/10.1016/j.ympev.2018.07.001>
- Zhou T et al (2016) Comparative transcriptome and chloroplast genome analyses of two related *Dipteronia* Species. *Front Plant Sci* 7:1512. <https://doi.org/10.3389/fpls.2016.01512>
- Zhou T, Wang J, Jia Y, Li W, Xu F, Wang X (2018) Comparative chloroplast genome analyses of species in *Gentiana* section *Cruciata* (Gentianaceae) and the development of authentication markers. *Int J Mol Sci* 19:1962. <https://doi.org/10.3390/ijms19071962>
- Zhou T et al (2019) The complete chloroplast genome of *Euphrasia regelii*, pseudogenization of *ndh* genes and the phylogenetic relationships within Orobanchaceae. *Front Genet* 10:444. <https://doi.org/10.3389/fgene.2019.00444>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.