Original Articles

Acupuncture and Herbal Medicine

# Chloroplast genome characterization and divergence time estimation of *Persicaria capitata*

Xianfa Zeng[1], Chang Liu[1,*], Xiaoying Yang[1], Xinyue Wang[1], Jiangli Luo[1], Ying Zhou[1], Liyan Zhang[1], Zengjie Gao[1], Xiang Pu[1,*]

[1]College of Pharmacy, Guizhou University of Traditional Chinese Medicine, Guiyang, China

**Abstract**

**Objective:** With *Persicaria capitata* as test materials, we compared and analyzed the chloroplast (cp) genome characteristics as well as their phylogenetic relationships and evolutionary history with related species of *Persicaria nepalensis*, *Persicaria japonica*, *Persicaria chinensis*, *Persicaria filiformis*, *Persicaria perfoliata*, *Persicaria pubescens*, *Persicaria hnydropiper.*

**Methods:** The Illumina HiSeq high-throughput sequencing platform was used for the first time for *P. capitata* cp genome sequencing. NOVOPlasty and CpGAVAS2 were used for assembly and annotation, and Codon W, DnaSP, and MISA were used to conduct a series of comparative genomic analyses between the plant and seven species of the same genus. A phylogenetic tree was constructed using the maximum likelihood (ML) and neighbor-joining (NJ) methods, and divergence time was estimated using BEAST.

**Results:** The total length of *P. capitata* cp genome was 158,821 bp, with a guanine and cytosine (GC) content of 38.0%, exhibiting a typical circular tetrad structure. The genome contains 127 annotated genes, including 82 protein-coding and 45 tRNA-encoding genes. The cp genome harbors simple sequence repeat (SSR) loci primarily composed of A/T. The conserved species structure of this genus is reinforced by the expansion and contraction of the inverted repeat (IR) region. The non-coding regions of the cp genomes exhibited significant differences among the genera. Six different mutation hotspots (*psbK-psbI*, *atpI-rps2*, *petN-psbD*, *atpB-rbcL*, *cemA-petA*, *ndhI-ndhA-ycf1*) were screened from the non-coding regions of genes with high nucleotide variability (pI). These hotspots were expected to define the phylogenetic species of *Persicaria*. Furthermore, phylogenetic analysis of Polygonaceae plants showed that *P. capitata* was more closely related to *P. chinensis* than *P. nepalensis*. Analysis of divergence time indicated that Polygonaceae originated in the Late Cretaceous (~180 Ma) and began to differentiate during the Middle Miocene. *Persicaria* differentiated ~66.44 million years ago, during the Miocene.

**Conclusions:** Our findings will serve as a scientific basis for further research on species identification and evolution, population genetics, and phylogenetic analysis of *P. capitata*. Further, we provide valuable information for understanding the origin and evolution of *Persicaria* in Polygonaceae and estimating the differentiation time of *Persicaria* and its population.

**Keywords:** Chloroplast genomes, Comparative analysis, Divergence time, *Persicaria capitata*, Phylogenetic relationship

**Graphical abstract:** http://links.lww.com/AHM/A140.

## Introduction

*Persicaria capitata* (Buch.-Ham. ex D. Don) H. Gross, also named "Sijihong" "Shimangcao" "Shuixiuqiu," is commonly used medicine in China. It is one of the characteristic Miao medicines of Guizhou Province, used to cure many diseases such as clearing away heat, detoxification, pain relief, diuresis, etc[1]. *P. capitata* is widely distributed in China and abroad, mainly in northern India, Nepal, Sikkim, Bhutan, Myanmar, and Vietnam. It is produced in Guizhou, Yunnan, Sichuan, Tibet, Guangxi, and other provinces[2]. Specifically, *P. capitata* is one of the priority developments of the "Six major Miao medicines" and the "Seven major Chinese medicine industry chains" in Guizhou[3]. *P. capitata* is a Traditional Chinese medicine (TCM) that has long been used by the Miao people in China to treat various urological disorders[4]. It has considerable antibacterial and anti-inflammatory activities and little toxicity[5].

*P. capitata* has red flowers, round and purple leaves, and a slightly bitter and astringent taste. Zhong Miao harvested it for medicinal use in Guizhou (Tongzhi) and Pingyuan (Zhi) during the Qing dynasty. It was originally known by its herbal name, "Suodongxue"[6]. According to the Second Volume of 1963 Guangxi TCM, *P. capitata* is known as the "shimangcao." It is bitter and pungent in flavor, flat in nature, non-toxic, and capable of detoxifying and anti-inflammatory. The 1977 "Chinese Dictionary of Herbal Medicine" states that *P. capitata* has a bitter, pungent, and

*Corresponding author. Chang Liu, E-mail: 19liuchang@163.com; Xiang Pu, E-mail: 314745595@qq.com.

Received 4 March 2024 / Accepted 19 December 2024

**How to cite this article:** Zeng XF, Liu C, Yang X, Wang XY, Luo JL, Zhou Y, Zhang LY, Gao ZJ, Pu X. Chloroplast genome characterization and divergence time estimation of Persicaria capitata. Acupunct Herb Med 2024;5(1):89–102. DOI: 10.1097/HM9.0000000000000133

cool flavor. It clears away heat, promotes diuresis, blood circulation, and detoxification, and is mainly used for treating diseases such as heat stranguria, bloody stranguria, dysentery, and traumatic injury. *P. capitata* was chronicled in "Materia Medica of GuiZhou" showing channel tropism of the lung, kidney, and bladder meridian. It can also be used to treat tract infection, rheumatism, and bruises caused by falls. The Chinese Herbal Medicines of Honghe states that it was used to treat tuberculosis and hemoptysis. Shel gong shel phreng also records that it treats lung disease and benefits the small intestine. Simultaneously, many documents and works, such as "Chinese Herbal Medicine of Yunnan" "Yi Medicine Plant Medicine (sequel)" "Color Atlas of Common Chinese Herbal Medicine in China," have recorded its medication[7].

We have been studying *P. capitata* for over 20 years, and published several studies, such as "Quality evaluation study of *P. capitata* based on the combination-effect relationship" "Phase IV clinical study of the Miao medicine Relinqing granule after its marketing" and "Study on the mechanism of regulating PI3K/Akt/mTOR pathway to treat pneumonia by Miao medicine *P. capitata* based on the theory of syndrome differentiation for treatment of traditional Chinese medicine." We have obtained many remarkable research results about *P. capitata*, such as those in "Fingerprint data and pharmacodynamic activity indexes of different extracts were established for *P. capitata*"[8–9], "The contents of gallic acid, quercitrin, protocatechuic acid, and quercetin in *P. capitata* were determined"[10], "The curative effect and mechanism of Miao medicine *P. capitata* in treating pneumonia by regulating PI3K/Akt/mTOR signal pathway were discussed." Notably, the in-depth systematic research and industrialization development of key technologies for the promotion and application of Miao medicine *P. capitata* won the first prize in the Guizhou Provincial Scientific and Technological Progress. Furthermore, we found that *P. capitata* has a curative effect on urinary tract and lung infections, an inhibitory effect on *Helicobacter pylori, Staphylococcus aureus, Escherichia coli*, and other bacteria, and a broad-spectrum antibacterial effect[11–13].

*P. capitata* has been documented and researched for the treatment of various diseases, especially infectious diseases, since ancient times. However, there are no reports on the precise identification of the source of the drug, the relationship between relatives, and the evolutionary history of *Persicaria* plants, leading to adulteration of plants of the same genus and closely related species often occurring in the market of the drug. This makes it difficult to ensure the quality and therapeutic efficacy of *P. capitata* and its related products[14]. Molecular identification has the advantages of accuracy, speed, and low cost, and has become a common means and research hotspot. Chloroplasts (cps) are important organelles in plant cells, and their genomes, with high sequence conservation and low base substitution rates, are powerful tools for studying species identification and phylogenetic development[15].

Therefore, we assembled and annotated the cp genomes of *P. capitata* and *P. nepalensis* and the final genome sequences were submitted to the official NCBI website. Meanwhile, we downloaded some of the published information on the cp genomes of Polygonaceae and utilized bioinformatics techniques to conduct comparative analyses of the cp genomes of *P. capitata* and eight species of plants in the genus to identify the sequence fragments and to analyze the phylogeny and evolutionary history of plants in the genus. This study provides data to support further in-depth studies of the *Persicaria* cp genome, provides new ideas for species identification, and establishes a foundation for studying the phylogeny and evolutionary history of *Persicaria* plants.

## Materials and methods

### *Plant material, DNA extraction, and sequencing*

Fresh and healthy leaves of *P. capitata* and *P. nepalensis* were harvested from the germplasm resource nursery of Guizhou University of Traditional Chinese Medicine in Guiyang, Guizhou, China (26°23′N, 106°37′E) and confirmed as *P. capitata* and *P. nepalensis* by Prof. Wei Shenghua. The leaves were washed numerous times with sterile water, dried, and subsequently stored at −80°C. DNA was extracted from the leaves of *P. capitata* and *P. nepalensis* using the Beijing Tiangen Biotech Plant DNA Extraction Kit (Tiangen Biotech Co., Beijing, China). The extracted DNA was subjected to agarose gel electrophoresis on a 0.8% gel to check for degradation and impurities. Subsequently, the DNA concentration was determined. DNA samples that passed quality control were used to create a double-ended sequencing library with an insert size of 350 bp, following the standard Illumina DNA library construction procedure. Quality control pass sequencing of the high-throughput sequencing libraries was performed using the Illumina Novaseq6000 high-throughput sequencing platform adopting the PE150 (Pair-End 150) sequencing strategy.

### *Cp genome assembly and annotation*

Trimmomatic was used to filter the raw data[16]. The cp genome was assembled using NOVOPlasty molding[17]. Subsequently, gap closure was performed to resolve

**Table 1**

**Fossil information for the BEAST analyses**

| | *Muehlenbeckia* | *Polygonum* | *Calligonum* | *Persicaria* | *Armeria* |
|---|---|---|---|---|---|
| Expo/log | Exponential | Exponential | Exponential | Exponential | Exponential |
| | Mean: 44 | Mean: 11.6 | Mean: 5.3 | Mean: 65.8 | Mean: 11.6 |
| | Offset: 39.6 | Offset: 5.3 | Offset: 2.6 | Offset: 55.8 | Offset: 5.3 |
| Uniform | Lower: 39.6 | Lower: 5.3 | Lower: 2.6 | Lower: 55.8 | Lower: 5.3 |
| | Upper: 125.0 | Upper: 125.0 | Upper: 125.0 | Upper: 125.0 | Upper: 125.0 |

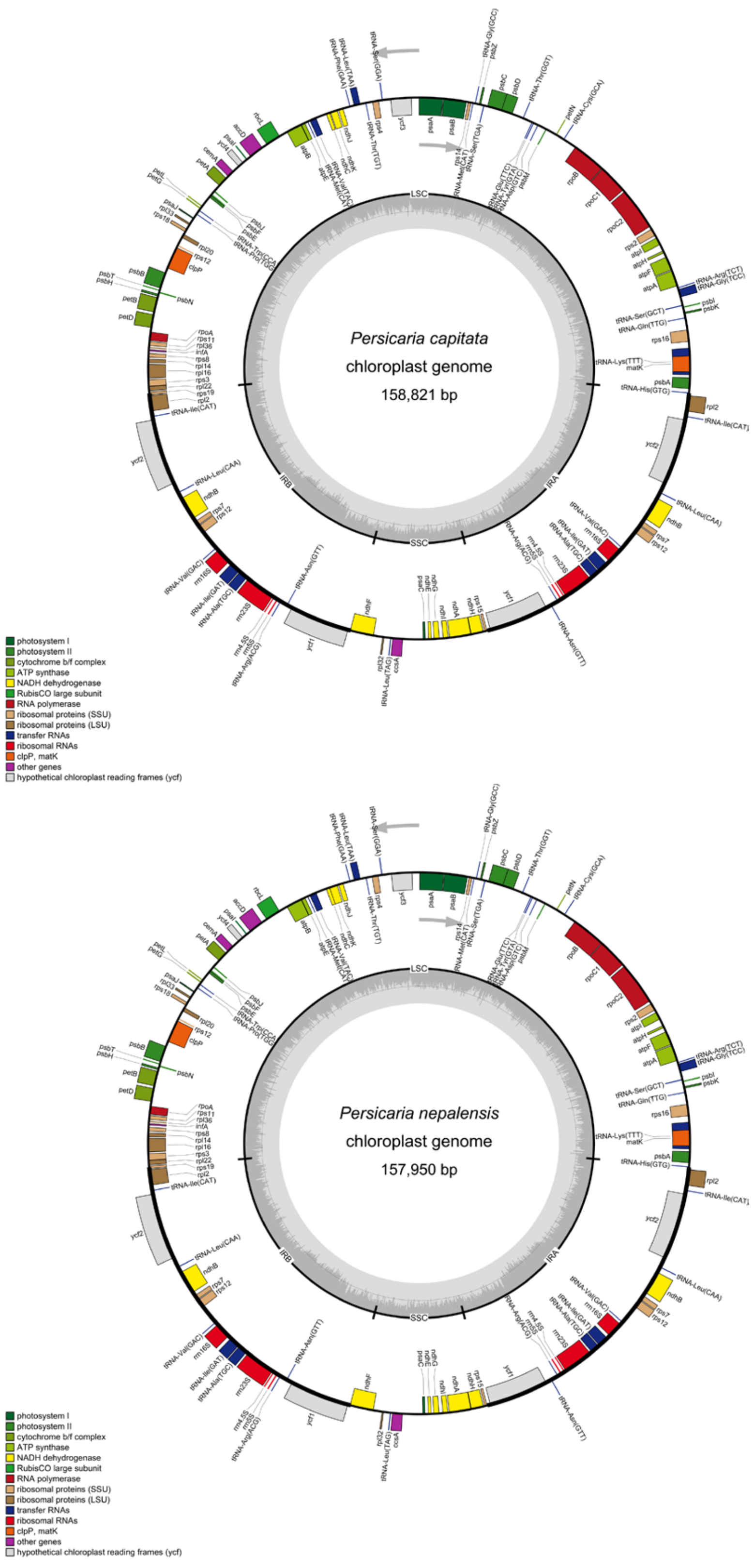


**Figure 1.** Cp genome map of the *Persicaria capitata* and *Persicaria nepalensis*. Length of the cp genome is shown in the circle. Genes located inside the circle are transcribed in the clockwise direction, while those outside are transcribed counter-clockwise. The inner circle's shading denotes varying GC and AT content. AT: Adenine and thymine; Cp: Chloroplast; GC: Guanine and cytosine.

the internal gaps. The genome was manually corrected using CpGAVAS2, resulting in a complete cp genome sequence[18]. Whole cp genome maps were generated using the OGDRAW tool[19]. The annotated genome sequence has been deposited in GenBank (accession numbers OR148685 and OR148686).

### *Analysis of codon usage and repeat sequences*

CodonW was used to examine synonymous codons and relative synonymous codon usage (RSCU) of both *Persicaria* species[20]. Simple sequence repeats (SSRs) were analyzed using MISA software, whereby the parameters were set to 1, 2, 3, 4, 5, and 6 for SSRs and 9, 4, 3, 3, 3, 3, and 3 for nucleotides. The distance between two SSRs was required to be no less than 100 bp[21]. This analysis was repeated using REPuter[22]. These included forward (F), reverse (R), complementary (C), and palindromic repeats (P). The minimum repeat size was set to 30 bp with 90% sequence identity. The abbreviations for the technical terms are explained when they were initially used.

### Comparative genome analysis and sequence variations

The boundary information of the inverted repeats (IR), large single copy (LSC), and small single copy (SSC) areas in the four regions was visualized using IRScope in the cp genomes. The on-line genome analysis program Mvista was employed to compare all eight cp genomes, with *P. capitata* serving as a reference, in the Shuffle-LAGAN mode. MAFFT was used to compare the entire cp genome sequences of the eight *Persicaria* species, followed by sliding window analysis using DNAsp v.6.10[23] with a window length of 600 bp and a step length of 200 bp.

### Phylogenetic and divergence time analysis

The two *Persicaria* cp genomes and the complete cp genomes of 43 species, including *Persicaria*, *Polygonum*, *Bistorta*, *Calligonum*, *Muehlenbeckia*, *Rheum*, *Fagopyrum*, *Oxyria*, *Pleuropterus*, and *Rumex* were obtained from the NCBI for Biotechnology Information database to create a phylogenetic and divergence time tree [S1 Table, http://links.lww.com/AHM/A146]. *Limonium sinense* and *Limonium bicolor* were used as the outgroups. MAFFT (https://mafft.cbrc.jp/alignment/

**Table 2**

**Complete cp genome features of *Persicaria capitata and Persicaria nepalensis***

| | *P. capitata* | | *P. nepalensis* | |
|---|---|---|---|---|
| | Length (bp) | GC content (%) | Length (bp) | GC content (%) |
| Total length (bp) | 158,821 | 38.00 | 157,950 | 37.90 |
| LSC length (bp) | 84,059 | 36.13 | 83,439 | 36.05 |
| IR length (bp) | 30,931 | 41.42 | 30,837 | 41.46 |
| SSC length (bp) | 12,900 | 33.30 | 12,837 | 33.05 |

cp: Chloroplast; GC: Guanine and cytosine; IR: Inverted repeat; LSC: Large single copy; SSC: Small single copy.

**Table 3**

**Annotated genes and their classification in the cp genomes of *Persicaria capitata***

| Category | Gene group | Gene name |
|---|---|---|
| Self-replication | Large subunit of ribosome | *rpl2***(x2), *rpl14*, *rpl16*, *rpl20*, *rpl22*, *rpl32*, *rpl33*, *rpl36* |
| | Small subunit of ribosome | *rps2*, *rps3*, *rps4*, *rps7*(x2), *rps8*, *rps11*, *rps12*(x2), *rps14*, *rps15*, *rps16*, *rps18*, *rps19* |
| | RNA polymerase | *rpoA*, *rpoB*, *rpoC1**, *rpoC2* |
| | rRNA | *rrn4.5S*(x2), *rrn5S*(x2), *rrn16S*(x2), *rrn23S*(x2) |
| | tRNA | *tRNA-Val*(x3), *tRNA-Tyr*, *tRNA-Trp*, *tRNA-Thr*(x2), *tRNA-Ser*(x3), *tRNA-Pro*, *tRNA-Phe*, *tRNA-Met*(x2), *tRNA-Lys*, *tRNA-Leu*(x4), *tRNA-Ile*(x4), *tRNA-His*, *tRNA-Gly*(x2), *tRNA-Glu*, *tRNA-Gln*, *tRNA-Cys*, *tRNA-Asp*, *tRNA-Asn*(x2), *tRNA-Arg*(x3), *tRNA-Ala*(x2) |
| Photosynthesis | Photosystem I | *psaA*, *psaB*, *psaC*, *psaI*, *psaJ* |
| | Photosystem II | *psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbM*, *psbN*, *psbT*, *psbZ* |
| | NADH dehydrogenase | *ndhA**, *ndhB***(x2), *ndhC*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK* |
| | Cytochrome b/f complex | *petA*, *petB**, *petD**, *petG*, *petL*, *petN* |
| | ATP synthase | *atpA*, *atpB*, *atpE*, *atpF*, *atpH*, *atpI* |
| | Large subunit of rubisco | *rbcL* |
| Other genes | Translational initiation factor | *infA* |
| | Maturase | *matK* |
| | Protease | *clpP** |
| | Envelope membrane protein | *cemA* |
| | Subunit of Acetyl-carboxylase | *accD* |
| | C-type cytochrome synthesis | *ccsA* |
| | Open reading frames | *ycf1*(x2), *ycf2*(x2), *ycf3**, *ycf4* |

cp: Chloroplast.
Gene *, genes with one intron; Gene **, genes with two introns; Genes (x2) and (x3), number of copies of multi-copy genes. rpl. Large ribosomal protein (rps) The ribosomal protein subunit rpo. RNA polymerase; psa. Photosynthetic apparatus (ndh) NADH dehydrogenase, peptides polypeptide; rbc. Ribulose bisphosphate Carboxylase Mats The mature enzyme, clp. The caseinolytic protease, cem. cp envelope membrane protein Acc. Acetyl-CoA carboxylases CoAs; ccs. The TypeC cytochrome synthesis gene, ycf. Hypothetical cp openreading frame.

server/) was used to align the cp genomes of all the species. Manual corrections were then performed. Phylogenetic trees were constructed using the neighbor-joining (NJ) method in MEGAX software[24], with a bootstrap repeat value of 1,000. Phylogenetic trees were constructed using the maximum likelihood (ML) method with the IQ-TREE 2.0.5 software (http://www.iqtree.org/). The conformation tree parameters were configured as -m MFP-B 1000 -alert 1000 and TVM + F + I + G4 was selected as the optimal conformation tree. Chiplot software (https://www.chiplot.online/) was used to embellish the developmental trees. To estimate the Polygonaceae *Persicaria* divergence time, BEAST v1.8.0[25] was employed. Fossive information was obtained from the records of Tanja Schuster et al.[26] regarding the historical data of *Persicaria*, *Muehlenbeckia*, *Polygonum*, *Calligonum*, and *Plumbaginaceae pollen* (Table 1). The MAFFT comparison results were loaded into Geneious software to generate the Nex file, which was subsequently imported into BEAUTi v1.8.0. Here, GTR + Gamma was chosen as the nucleotide replacement model (shape = 0.241) while employing the Relaxed Clock Exponential (RCE) model. The *a priori* value of the tree was then computed using The Yule Model, and the analysis was run for 650,0000 generations, with sampling performed at intervals of 1,000 generations. Discarding the last 650,0000 generations (10%) as burn-in is necessary. The stability of the results was evaluated using Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer/). Figtree v1.4 was used to view the tree file, and Chiplot (https://www.chiplot.online/) was used for developmental tree beautification.

## Results

### *Cp genome features of the two Persicaria species*

The complete cp genomes were 158,821- and 157,950-bp long in *P. capitata* and *P. nepalensis*, respectively. The *Persicaria* cp genomes had a typical quadripartite


**Figure 2.** Analysis of codon preference between (A) *Persicaria capitata* and (B) *Persicaria nepalensis*.

structure similar to that of most angiosperms. This structure comprised LSC, a pair of IR, and an SSC region that are 84,059-, 30,931-, and 12,900-bp long, respectively [S2, S3 Tables, http://links.lww.com/AHM/A146]. The overall guanine and cytosine (GC) content of both plants was similar, ranging from 37.9% to 38%. The GC content in the IR regions (41.42%–41.46%) surpassed that in the LSC and SSC regions (36.05%–36.13% and 33.3%–33.05%) (Figure 1 and Table 2).

The cp genomes of *P. capitata* and *P. nepalensis* contain 127 genes. It comprises 82 protein-coding genes, 37 tRNA genes, and eight rRNA genes. These genes were divided into three categories based on their function: self-replication-related genes, photosynthesis-related genes, and other genes. Furthermore, one gene (*ycf1*) has been identified as a pseudogene. Among these genes, 10 protein-coding genes (*ycf1*, *ycf2*, *ycf15*, *rpl2*, *rps12*, *rps19*, *rps7*, *rpl23*, *ndhB*, *ndhF*), six genes encoding transfer RNA (*tRNA-Val*, *tRNA-Arg*, *tRNA-Leu*, *tRNA-Ile*, *tRNA-Asn*, *tRNA-Ala*), and four genes encoding ribosomal RNA (*rrn5S*, *rrn4.5S*, *rrn16S*, *rrn23S*) appeared in both the IR regions. Moreover, out of the 127 genes, only 11 genes contained introns. Among these 11 genes, nine genes (*petB*, *petD*, *rps16*, *ycf3*, *ndhA*, *clpP*, *rpoC1*, *atpF*, *and rpl16*) had 1 intron, whereas two genes (*rpl2* and *ndhB*) possessed two introns (Table 3 and S4 Table, http://links.lww.com/AHM/A146).

### Codon usage analysis

Analysis of codon usage is crucial for evaluating the evolution of the cp genome. RSCU values were calculated for the cp genomes of *P. capitata* and *P. nepalensis* based on their protein-coding sequences. Figure 2 illustrates the coding content of all protein-coding genes in the cp genomes of the two species, including 20 amino acids and stop codons. Leucine (Leu, L) had the highest number of codons among all amino acids. Based on the RSCU value analysis of its codons, it was found that *P. capitata* had 32 codons with RSCU >1, of which five had G/C endings and 27 had A/U endings. Similarly, *P. nepalensis* had 31 codons with RSCU >1, of which five had G/C endings and 26 had A/U endings. These results suggest a preference for A/U-ending codons over G/C-ending codons in the cp genomes of the two species [S5 Table, http://links.lww.com/AHM/A146].

### Repeat sequence analysis

Structures longer than 30 bp are referred to as long repeats and can be subdivided into four types: forward,

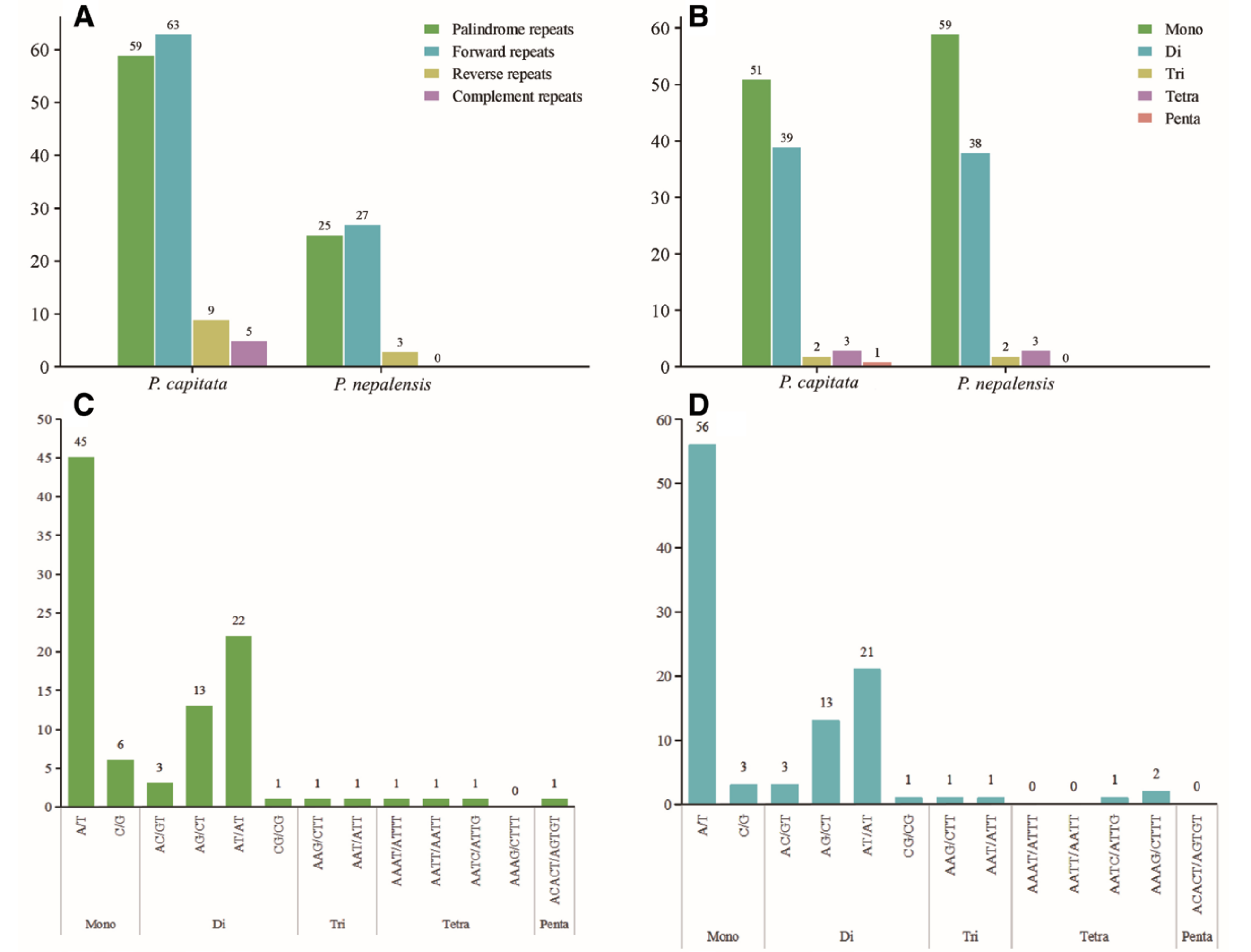


**Figure 3.** Repeat sequences in the cp genomes of *Persicaria capitata* and *Persicaria nepalensis*. (A) Numbers of different long repeat types. Frequencies of different SSR types in two *Persicaria* species: (B) *P. capitata* (C) and *P. nepalensis* (D). Cp: Chloroplast; SSR: Simple sequence repeat.

palindromic, reverse, and complementary. Analysis of long repetitive sequences revealed that *P. capitata* and *P. nepalensis* contain 136 and 55 long repetitive sequences, respectively, including forward (F), palindromic (P), reverse (R), and complementary repeats (C). Most of the repetitive sequences found in the cp genomes of both species were between 30 and 39 bp in length, and no complementary repeats were found in *P. nepalensis* (Figure 3A; S6 Table, http://links.lww.com/AHM/A146). Upon analysis of SSRs, both species exhibited 96 and 102 SSR sites, respectively, characterized by unidentifiable hexanucleotide repeat patterns (Figure 3B–D; S7 Table, http://links.lww.com/AHM/A146).

## IR contraction and expansion

A comparison of the IR/LSC and IR/SSC boundaries among the cp genomes of the eight *Persicaria* species is shown in Figure 4. *Persicaria perfoliata* had the longest cp genome (160,730 bp) while *P. nepalensis* had the shortest (157,950 bp) among the eight species. In all eight *Persicaria* species, the IRb-LSC border was positioned in the *rps19* gene's coding or intergenic region with three genes, *rpl22, rps19*, and *rpl2*, present at the border. For the IRb-SSC boundary, the coding region of the *ndhF* was located in all eight species. Similarly, for the IRa-SSC boundary, the intergenic regions of *ycf1* and *rps15* housed all eight species. The IRa-LSC boundary was identified solely in the *rpl2* and *rps19* intergenic regions of *P. perfoliata*, whereas the other seven species were situated in the *trnH* intergenic region.

## Comparative genomic analysis

To investigate genomic divergence, the percentage of sequence identity was calculated for the eight species of *Persicaria* using the mVISTA program, with *P. capitata* as the reference. There was high similarity among the eight species and the variability in the IR regions was less than that in the LSC and SSC regions. Furthermore, the cp genomes were more highly variable in their noncoding regions than in their coding regions, which is consistent with the pattern found in most angiosperms. The cp genome contains many variable nucleotides that can be used to identify closely related species or genera for valuable DNA barcoding. Furthermore, the coding regions with large variations in the ten cp genomes were *matK*, *rps16*, *atpF*, *rpoC1*, *ycf3*, *clpP*, *petB*, *petD*, *ndhA,* and *rpl16*, whereas other genes were highly conserved (Figure 5). The number of variant loci in the intergenic regions was significantly higher than that in the gene regions. To clarify the variation in higher regions, we calculated the nucleotide diversity values (pi) using DNAsp v.6.10 software. Six divergent loci (*psb*K-*psb*I, *atp*I-*rps*2, *pet*N-*psb*D, *ndh*I-*ndh*A-*ycf*1, *atp*B-*rbc*L, and *cem*A-*pet*A.) had a *P* value >0.01, with *psb*K-*psb*I, *atp*I-*rps*2, *pet*N-*psb*D, *atp*B-*rbc*L, and *cem*A-*pet*A located in the LSC region, whereas the other loci were located in the SSC region, and none were detected in the IR region (Figure 6). These results confirmed that the LSC and SSC regions exhibited greater variability than the IR regions. The cp genes of eight *Persicaria* species were examined using the Mauve software. Multiple genome comparisons revealed two local collinear blocks (LCB) shared

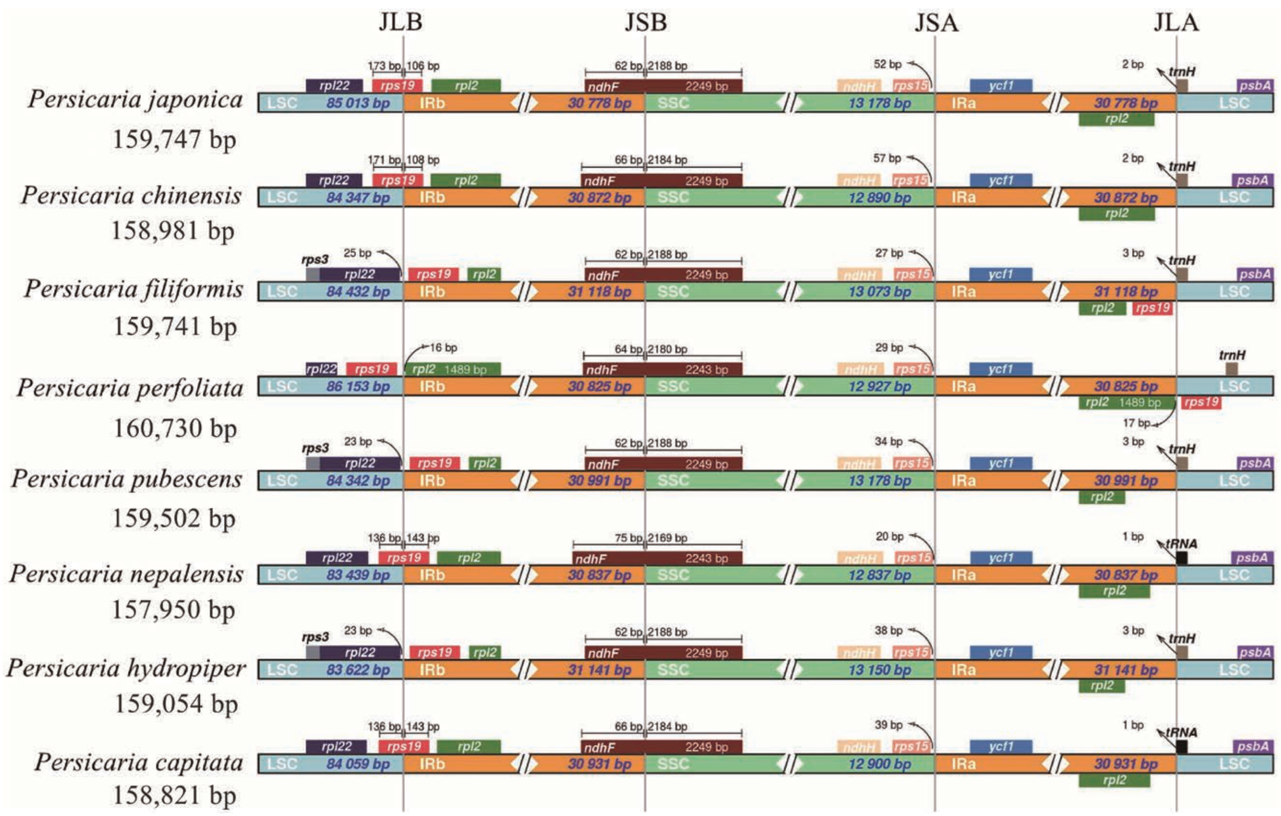


Figure 4. Comparison of IR boundaries in the cp genomes of eight *Persicaria* species. cp: Chloroplast; IR: Inverted repeat; JLA: Junction between LSC and IRa; JLB: Junction between LSC and IRb; JSA: Junction between SSC and IRa; JSB: Junction between SSC and IRb; LSC: Large single copy; SSC: Small single copy.

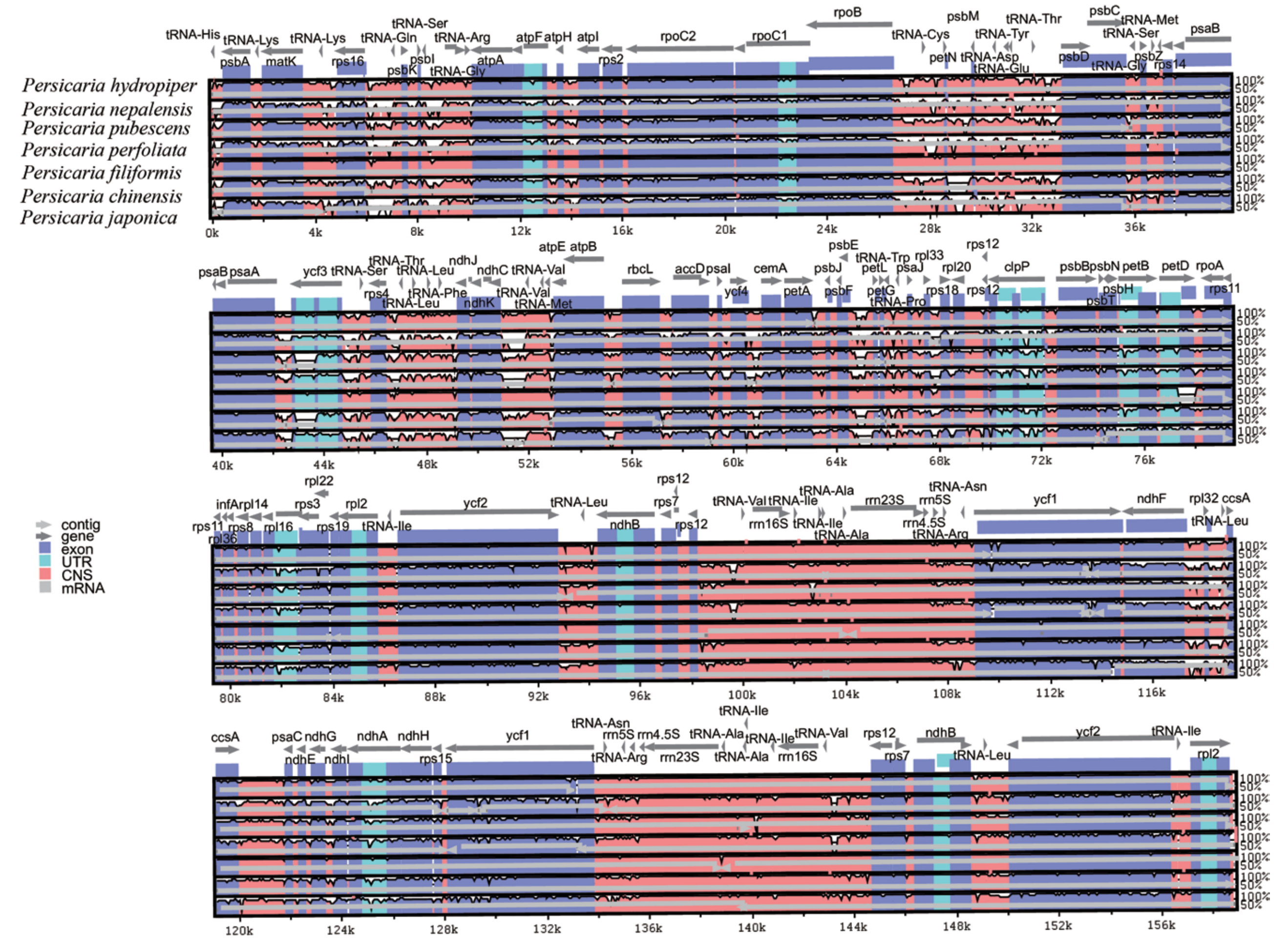


**Figure 5.** Global alignment in the cp genomes of the eight *Persicaria* species. The top arrows show the transcription direction; blue indicates the exons of protein-coding genes (exon); red shows the CNS; The X-axis represents the positions in the cp genome, while the Y-axis represents percentage identity within 50% to 100%. CNS: Conserved non-coding sequence; cp: Chloroplast.

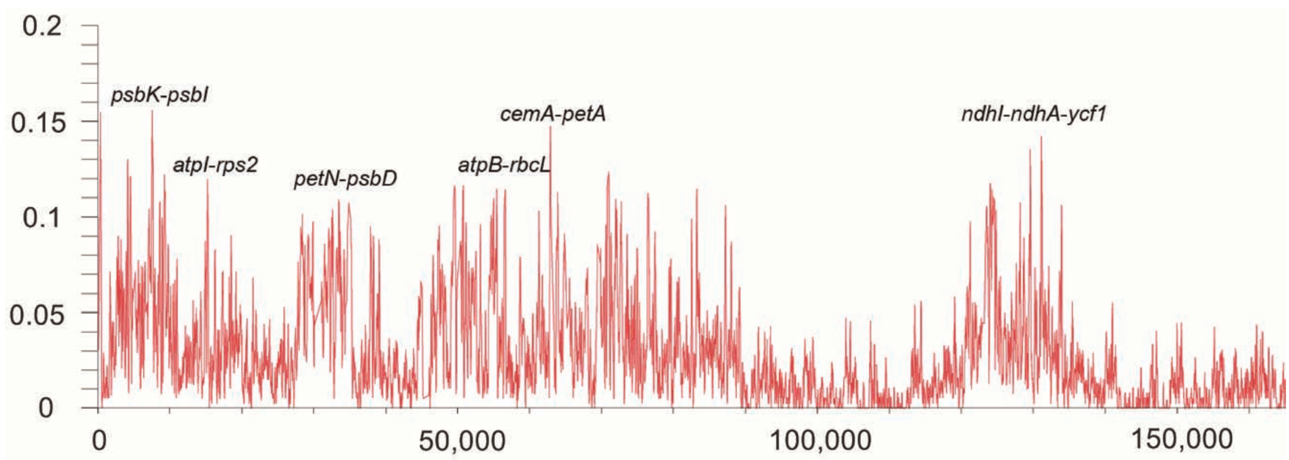


**Figure 6.** Nucleotide polymorphism analysis of the cp genomes of eight species of *Persicaria*. The location of the mutation site was along the X-axis, and the nucleotide diversity (Pi) value in each window was along the Y-axis. cp: Chloroplast.

among the cp genomes of the eight species (Figure 7). The genomes of the eight species displayed a high degree of similarity. Eight species within the genus exhibited no rearrangements or inversions in their cp genomes. However, common mutations in the identified regions (0–90,000; 120,000–135,000) were characterized by a high degree of gene sequence variation in the aligned cp genomes.

### *Phylogenetic analysis*

To determine the systematic position of *Persicaria*, the phylogenetic tree was constructed based on Polygonaceae cp sequences. The NJ method analysis indicated that there are distinct phylogenetic connections between 43 plants in 10 genera belonging to the Polygonaceae family. Except for *L. sinense* and *L. bicolor* as outgroups, the Polygonaceae branch can be divided into

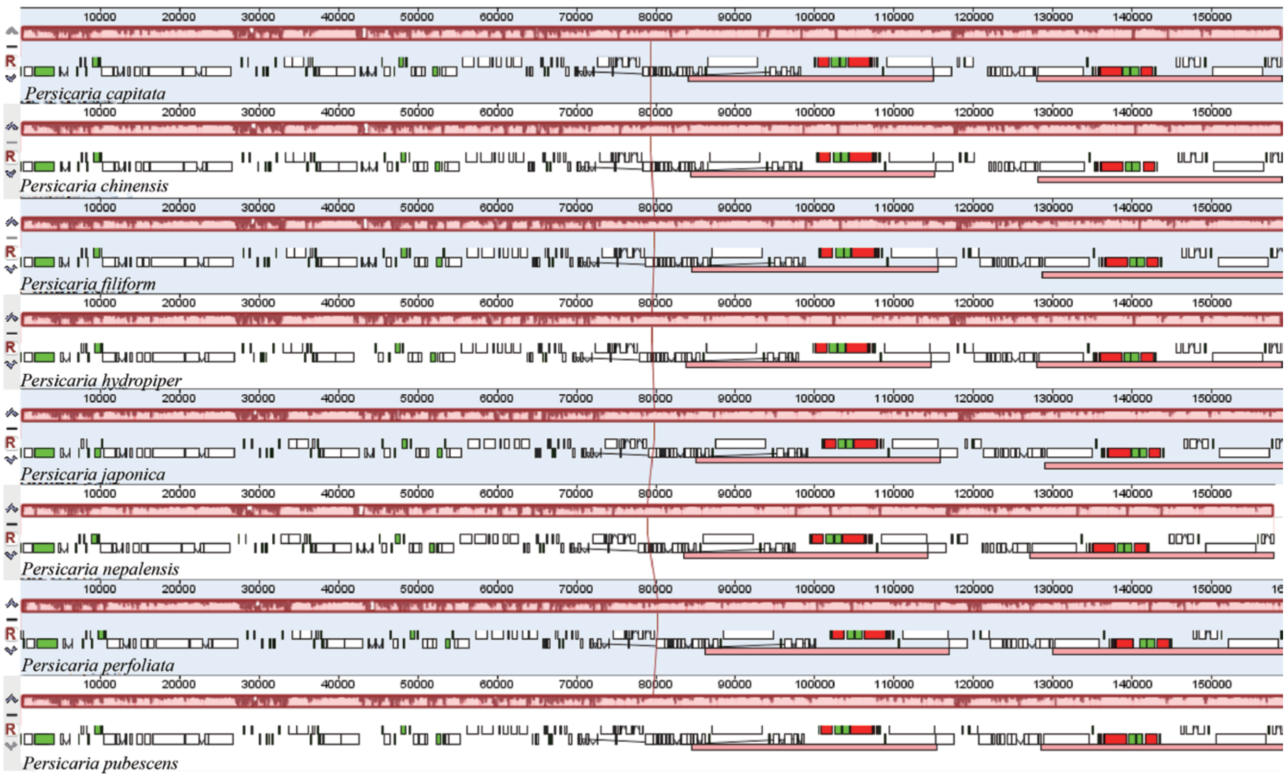


**Figure 7.** Collinearity analysis of cp genomes among eight *Persicaria* species. Within the alignments, local collinear blocks are represented by blocks of the same color connected by lines. cp: Chloroplast.

two major categories: *Persicaria* and *Bistorta* as a single category and the remaining eight genera, *Polygonum, Calligonum, Muehlenbeckia, Rheum, Fagopyrum, Oxyria, Pleuropterus*, and *Rumex*, comprise the second category. The three genera *Rheum, Oxyria*, and *Rumex* were classified into a single branch with 100% support. This implies that these three genera are closely related (Figure 8). *Persicaria, Muehlenbeckia*, and *Pleuropterus* spp. grouped into single branches. *Persicaria* and *Bistorta* were grouped into a single branch with a 100% level of support, indicating a close relationship. Although *P. capitata* and its relative *P. nepalensis* clustered together with 100% support, *P. capitata* and its congener *P. chinensis* formed a separate, but closely related branch with the same level of support. The phylogenetic results of both ML and NJ analyses were essentially identical (Figure 9).

### *Divergence time analysis*

Based on the cp genome sequence data of *P. capitata* and *P. nepalensis*, along with the published cp genome information of 41 species of Polygonaceae and two species of *Plumbaginaceae*, we produced a divergence time tree for the *Polygonaceae* family. This was achieved using *Muehlenbeckia, Polygonum, Armeria, Calligonum*, and *Persicaria* divergence time points as calibration points (Figure 10). Figure 10 illustrates the divergence of *Muehlenbeckia, Polygonum*, and *Fallopia* in the *Polygonaceae* from genera *Calligonum, Persicaria, Fagopyrum, Bistorta, Oxyria, Rumex* during the *Cretaceous* Period, ~180 million years ago. *Rheum* and *Rumex* separated ~86 million years ago, with divergence occurring within the genus Rheum ~18.69 million years ago and within the genus *Rumex* ~13.27 million years ago at a node in the Miocene. *Persicaria* diverged from the more closely related *Bistorta* ~92 million years ago. Approximately 66.44 million years ago, *Persicaria* diverged from *Bistorta* and ~0.55 million years ago, *Bistorta* diverged from *Persicaria*. *P. capitata* and *P. nepalensis* diverged ~7.89 million years ago. The divergence between *P. capitata* and *P. chinensis* occurred in the Miocene epoch specifically ~2.32 million years ago. This event occurred during the Quaternary Period. The timing of the divergence supports the hypothesis that *P. chinensis* and *P. capitata* are more closely related than *P. nepalensis*.

## Discussion

*P. capitata* has been developed and marketed as an active pharmaceutical ingredient (API) in many preparations. This research is of great significance as *P. capitata* is one of the "six major Miao medicines" and "seven major Chinese medicine industry chains" developed in Guizhou Province. TCM has gained increasing popularity worldwide for the treatment and prevention of diseases[27]. The *P. capitata* herb-based medicinal product, Relinqing Granue, has been approved by the Chinese National Medical Products Administration for the treatment of tract infections[28]. Researchers have investigated its pharmacological activity[6], chemical composition[29], quality standards[30], and extraction processes[31]. However, studies on the cp genome of *P. capitata* have not yet been reported. Cps are the leading sites of photosynthesis in plants and are essential for plant growth and development. Their genome sequences

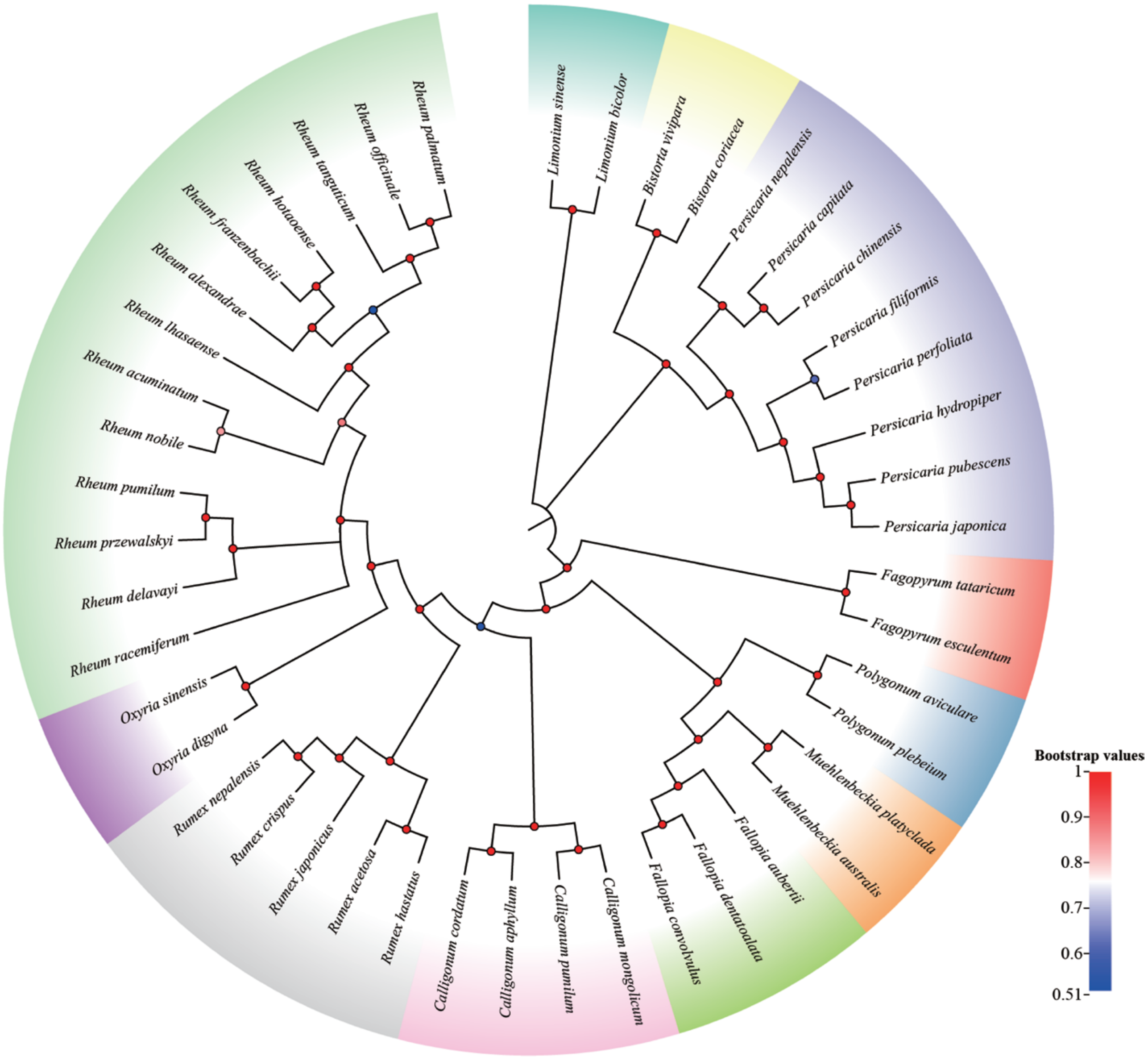


**Figure 8.** Neighbor-joining tree reconstructed based on the complete cp genome sequences of 45 species. *Limonium sinense* and *Limonium bicolor* were used as the outgroups. Colors were used to distinguish between the various genera. cp: Chloroplast.

are highly conserved, and their genome structure resolution and variation analysis can better understand their phylogenetic processes.

In the present study, the cp genomes of *P. capitata* and *P. nepalensis* were sequenced and analyzed. The results showed that the cp genomes of both were structured as a double-stranded circular tetrad structure, mainly consisting of four parts, IRa, IRb, LSC, and SSC, with a genome length of 158,821 bp and 157,950 bp, 127 genes, 20 different types of *tRNA*-encoded genes, and four *rRNAs*. The cp genome structural characteristics, gene length range, and number of genes are consistent with existing studies[17–18], reflecting a stable cp genome structure and a slow overall evolutionary rate[32–33]. The lengths of the LSC, SSC, and IR regions of the cp genome are 80 to 90, 16 to 27, and 20 to 28 kb, respectively[34]. Furthermore, *P. capitata* and *P. nepalensis* had LSC lengths of 84.059 and 83.439 kb, SSC lengths of 12.9 and 12.837 kb, and IR lengths of 30.931 and 30.837 kb, respectively. Shorter SSC regions reduce the chance of gene variation in the cp genomes, and the coding region with the most gene variation is the SSC region. Additionally, longer IR regions play an essential role in maintaining cp genome stability as they make the cp genome less susceptible to structural rearrangements[35–37]. SSR are widely used as molecular markers in population genetics[38], genealogical geography[39], and other related studies because of their abundance, high polymorphism, high information content, independence from external environmental factors, and uniparental inheritance patterns. Our findings revealed that the SSRs in the cp whole gene sequences of *P. capitata* and *P. nepalensis* were dominated by single-nucleotide repeats (53.12%–55.66% of the total), and single-nucleotide and dinucleotide SSR were mainly A/T and AT/TA, respectively, and the repeat units of trinucleotide SSRs were also mainly composed of A and T bases combinations, indicating that SSRs in the cp genome sequence are mainly composed of poly-A or

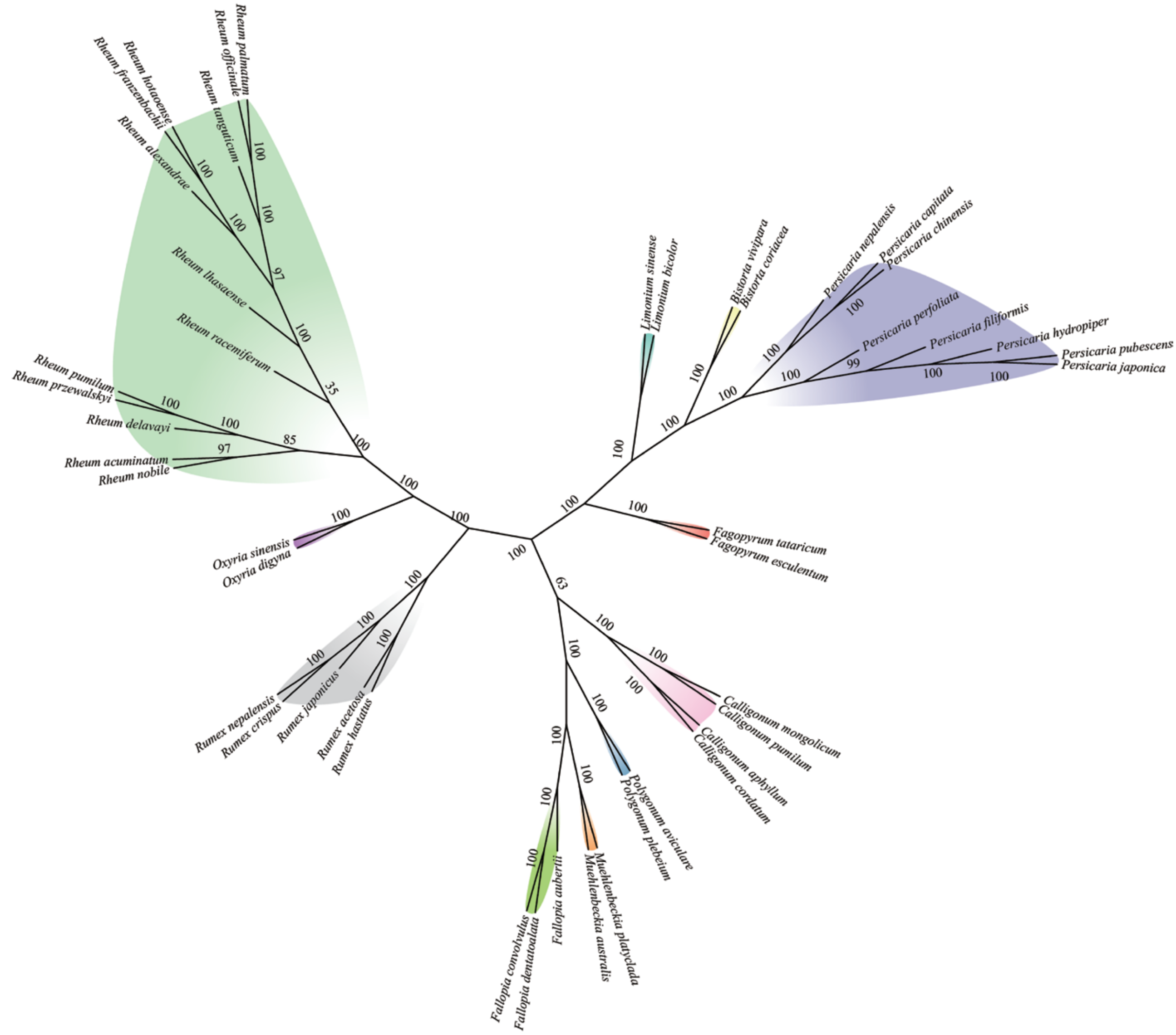


**Figure 9.** Maximum likelihood phylogenetic tree reconstructed based on the complete cp genome sequences of 45 species. *Limonium sinense* and *Limonium bicolor* were used as the outgroups. The numbers represent bootstrap values (%). Cp: Chloroplast.

poly-T, but less C or G tandem repeats, which is consistent with other scholars' studies[40], and these SSRs can be used as candidate molecular markers for molecular genetics-related studies in *Persicaria*.

Polygonaceae plants belong to the Angiospermae family, and the sequences of some highly variable regions in the cp genome sequences of angiosperms can often be used as molecular markers for species identification and related studies, such as phylogenetic relationship analysis[41]. Through a comparative analysis of the whole genome sequence of the cps of *P. capitata* and *P. nepalensis* and six other species of *Persicaria*, it was found that the cp genomes of *Persicaria* were also highly conserved, and the gene order was the same, with no gene deletions, mutations, inversions, or translocations. Polygonaceae have been studied in *P. multiflorus*, *Fagopyrum dibotrys*, *Fagopyrum tataricum*, and other studies are consistent[42]. A phylogenetic tree constructed from the full-length cp genomes of 19 Polygonaceae species by Hu et al.[43] showed that *Rheum*, *Rumex*, and *Oxyria* clustered into one large branch with 100% support. The monophyly of *Rheum*, *Rumex*, and *Oxyria* also received 100% support, which is the same conclusion as the present study. A study based on cp genomic data showed that *P. capitata* and *P. nepalensis* belonged to *Persicaria* of Polygonaceae and are more closely related to *Bistorta*.

Green plants originated 1.6 to 1 billion years ago, and the Earth's climatic environment was relatively smooth during this geological period[44]. However, three ice ages suddenly disrupted the billion-year-long stable environment on Earth in the Neogene, namely the *Sturtian* (~717–662 Ma), *Marinoan* (~639–635 Ma), and *Gaskiers* (~580 Ma) ice ages[44–46]. Beker speculated that the Gaskiers ice age might have been the main driver of land plant evolution[44]. The results of the present study indicate that Polygonaceae vulgare diverged ~180 million years ago during the Cretaceous, but it is impossible to prove whether it first appeared. The Tertiary Period is the oldest epoch of the Cenozoic (65–2.6 million years ago), these include the Early Tertiary (65–23.3 million years ago) and the Late Tertiary (23.3–1.64 million years

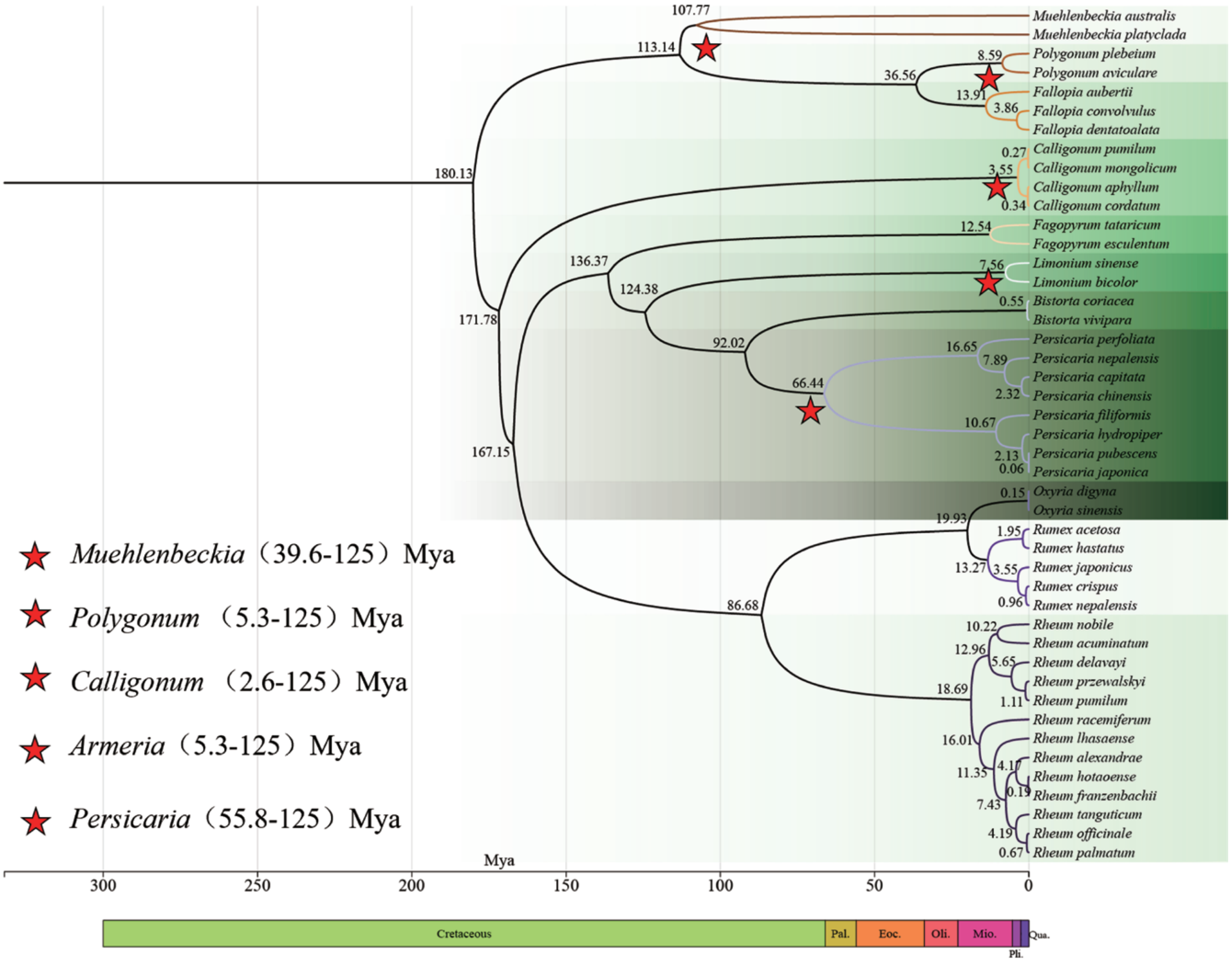


**Figure 10.** Divergence time of Polygonaceae based on cp genes. Cp: Chloroplast.

ago), in which angiosperms, insects, birds, and mammals flourished the most. The divergence of various medicinal plants within the Polygonaceae *Rheum*, such as *Rheum officinale*, *Rheum palmatum,* and *Rheum tanguticum*, took approximately 4.19 million years, *Oxyria* and *Rumex* separated in ~19.93 million years, *Bistorta* took ~0.55 million years, *Calligonum* 3.55 million years, *Fallopia* 13.91 million years, *Polygonum* 8.59 million years, and *P. capitata*, *P. nepalensis*, *Persicaria chinensis*, *Persicaria japonica,* and *Persicaria filiformis* all diverged in the Tertiary Neogene. During the Tertiary period, the final disappearance of the ancient Mediterranean Sea, the final formation of the Asian continent, the rise of the Tibetan Plateau, the formation of modern mountain systems such as the Alps, the Himalayas, the Rocky Mountains and the Andes, the appearance or disappearance of the Turgai Strait and the Isthmus of Panama, and the gradual development of climatic zonation were among the significant events[47] that may be closely related to the differentiation of angiosperms and the differentiation of plants of the *Polygonaceae*. Due to the deficiency of *Polygonaceae's* fossils and cp genome information, the evolutionary history of *Polygonaceae* could not be fully explored in this study, and the evolutionary survival of *Polygonaceae* under extreme environments such as seafloor spreading, ancient land disintegration, and climate change cannot be clarified.

## Conclusion

This study determined the complete cp genomes and revealed the basic structure, conservation, and variability of the sequences of *P. capitata* and *P. nepalensis* in *Persicaria* family of Polygonaceae. The IR regions were more conserved than the LSC and SSC regions, whereas the non-coding regions were more mutable than the gene-coding regions. SSRs and hotspot regions can be used to develop molecular markers for population genetics and phylogenetic studies. Phylogenetic relationships established using complete cp genome information can determine the relationships between and within genera; however, the IR region is not suitable for identification or phylogenetic analyses. Constructing a divergence tree based on the available fossil information on Polygonaceae can offer an initial understanding of the timing of divergence between various genera, as well as between different species within those genera in Polygonaceae. This will provide a specific theoretical foundation for further evolutionary studies on Polygonaceae plants. The survey results offer significant insights for identifying Polygonaceae species, determining evolutionary relationships, understanding divergence times, and cultivating genetic resources.

## Conflict of interest statement

The authors declare no conflict of interest.

## Funding

This study was supported by the National Natural Science Foundation of China (82060913).

## Author contributions

Zeng XF: Conceptualization (lead), data curation (lead), methodology (equal), resources (lead), software (lead), visualization (lead), and writing the original draft (lead). Liu C: Funding acquisition (lead) and writing, review, and editing (lead). Yang XY: Conceptualization (equal); data curation (equal); writing–original draft (equal); writing–review and editing (equal). Luo JL: Methodology (equal) and resources (equal).Wang XY: Data curation (equal);Zhou Y: Formal analysis (equal); visualization (equal); funding acquisition (equal). Zhang LY: Resources (equal), software (lead);Gao ZJ: Software (lead); Xiang Pu: Funding acquisition (equal); writing–review and editing (equal). All authors contributed to the manuscript revision and approved the final manuscript.

## Ethical approval of studies and informed consent

Not applicable.

## Acknowledgments

None.

## Data availability

The assembled cp genome sequences of *P. capitata* and *P. nepalensis* were uploaded to GenBank under the accession numbers OR148685 and OR148686, respectively.

## References

[1] Sun CS, Liang B, Wang CF. Research progress of Polygonum capitatum. *Mod Chin Med* 2005;7(4):26–28.
[2] Cao F, Tan H, Deng XK. Research progress on chemical constituents and flavonoid content of polygonum capitatum. *J Qiannan Med Coll Nationalities* 2012;25(2):121–125.
[3] Zhao HX, Bai H, Li W, et al. Chemical constituents from Polygonum capitatum. *Nat Prod Res Dev* 2011;23(2):262–266.
[4] Zheng L, Lu Y, Cao X, et al. Evaluation of the impact of Polygonum capitatum, a traditional Chinese herbal medicine, on rat hepatic cytochrome P450 enzymes by using a cocktail of probe drugs. *J Ethnopharmacol* 2014;158 Pt A:276–282.
[5] Liao SG, Zhang LJ, Sun F, et al. Antibacterial and anti-inflammatory effects of extracts and fractions from Polygonum capitatum. *J Ethnopharmacol* 2011;134(3):1006–1009.
[6] Liu YL, Sui Y, Hu CX, et al. Summary of the pharmacological activities of miao medicine Polygonum capitatum. *J Guizhou Univ Tradit Chin Med* 2021;43(1):81–84.
[7] Liang B, Zhang LY, Ran YX. Chinese Hmong medicine of Polygonum capitatum. *Beijing: China Press Tradit Chin Med* 2014;29:3–9.
[8] Xue XY, Liu CX, Zhou Y, et al. Spectrum-effect relationship of anti-inflammatory effect of different part from Polygonum capitatum. *Chin Tradit Herbal Drugs* 2018;49(21):5134–5141.
[9] Pu X, Zhang LY, Yang FW, et al. Treatment of uncomplicated urinary tract infection by relinqing: a systematic review of randomized controlled trials of clinical studies. *Tianjin Med J* 2016;44(8):1048–1052.
[10] Zhang LY, Wang PJ, Tian JH, et al. Meta-analysis of the study of "Miao Medicine" Reilin Qing. *J Guizhou Univ Tradit Chin Med* 2019;41(2):84–87.
[11] Fan ZL, Li LL, Pu X, et al. Research progress on the chemical composition and pharmacological mechanism of action of Polygonum Capitatum. *Asia-Pacific Tradit Med* 2023;19(8):229–233.
[12] Liu YX, Song XY, Kang WY, et al. Antibacterial activity of Polygonum capitatum against multi-resistant Staphylococcus aureus. *Chin Tradit Patent Med* 2014;36(9):1817–1821.
[13] Zhang S, Luo ZX, Mo F, et al. Analysis of antibacterial effect of Polygonum capitatum on Helicobacter pylori. *Chin J Hosp Pharm* 2015;35(2):113–118.
[14] Zhou T, Xie Y, Zhang LY, et al. Study on sequence characterized amplified region (SCAR) markers of Polygonum capitatum. *Chin J Hosp Pharm* 2013;38(16):2577–2580.
[15] Yuan J. Sequence Analysis and Phylogenetic Relationships of Chloroplast Genome of Coptis. Chongqing: Southwest University; 2020:I–II.
[16] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30(15):2114–2120.
[17] Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res* 2017;45(4):e18.
[18] Liu C, Shi L, Zhu Y, et al. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genomics* 2012;13:715.
[19] Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res* 2019;47(W1):W59–W64.
[20] Li ZW, Qiu Q, Lang JQ, et al. Sequence analysis of complete chloroplast genome of Dendrobium heterocarpum and Dendrobium trigonopus. *Chin Tradit Herbal Drugs* 2022;53(16):5159–5169.
[21] Du Q, Wang LQ, Chen ZE, et al. Characterization and phylogenetic analysis of the complete chloroplast genome of Lycopus europaeus. *Acta Pharm Sin* 2022;57(7):2206–2215.
[22] Zhang MY, Zhang YQ, Li YM, et al. Complete plastid genomes of Bupleurum chinense DC. and B. boissieuanum H. Wolff, with comparative and phylogenetic analyses of medicinal Bupleurum species. *Acta Pharm Sin* 2021;56(2):618–629.
[23] Li RJ, Wu LW, Xin TY, et al. Analysis of chloroplast genomes and development of specific DNA barcodes for identifying the original species of Rhei Radix et Rhizoma. *Acta Pharm Sin*. 2022;57(5):1495–1505.
[24] Song Y, Jia MJ, Cao YP, et al. Analysis on chloroplast genomic characteristics of Forsythia suspensa. *Acta Hortic Sin* 2022;49(1):187–199.
[25] Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 2007;7:214.
[26] Schuster TM, Setaro SD, Kron KA. Age estimates for the buckwheat family Polygonaceae based on sequence data calibrated by fossils and with a focus on the amphi-Pacific Muehlenbeckia. *PLoS One* 2013;8(4):e61261.
[27] Han P, Huang Y, Xie Y, et al. Metabolomics reveals immunomodulation as a possible mechanism for the antibiotic effect of Persicaria capitata (Buch.-Ham. ex D. Don) H.Gross. *Metabolomics* 2018;14(7):91.
[28] Zhang CL, Zhang JJ, Zhu QF, et al. Antihyperuricemia and antigouty arthritis effects of Persicaria capitata herba in mice. *Phytomedicine* 2021;93:153765.
[29] Yang JY. Research progress on chemical composition and detection methods of Persicaria capitata in seedling medicine. *Agric Technol Ser* 2019;36(9):46–48.
[30] Zhao YK, Huang DD, Wang X, et al. Quality standard for Polygonum capitatum. *Chin Tradit Patent Med* 2020;42(2):408–415.
[31] Yun CY, Gu MQ, Mu CF, et al. Optimization of the extraction process of Polygonum capitatum and its antimicrobial activity. *Chin Tradit Patent Med* 2018;40(3):725–729.
[32] Tian X, Liu YY, Zhang YM, et al. Comparative and phylogeny analysis of four Veratrum medicinal plants complete chloroplast genomes. *Chin Tradit Herbal Drugs* 2022;53(4):1127–1137.
[33] Fu G, Liu J, Li JQ. Characterization of chloroplast genome structure and phyletic evolution of Elsholtzia densa. *Chin Tradit Herbal Drugs* 2022;53(6):1844–1853.
[34] Serrano M, Wang B, Aryal B, et al. Export of salicylic acid from the chloroplast requires the multidrug and toxin extrusion-like transporter EDS5. *Plant Physiol* 2013;162(4):1815–1821.
[35] Guisinger MM, Kuehl JV, Boore JL, et al. Extreme reconfiguration of plastid genomes in the angiosperm family

Geraniaceae: rearrangements, repeats, and codon usage. *Mol Biol Evol* 2011;28(1):583–600.
[36] Hirao T, Watanabe A, Kurita M, et al. Complete nucleotide sequence of the Cryptomeria japonica D. Don. chloroplast genome and comparative chloroplast genomics: diversified genomic structure of coniferous species. *BMC Plant Biol* 2008;8:70.
[37] Palmer JD, Thompson WF. Rearrangements in the chloroplast genomes of mung bean and pea. *Proc Natl Acad Sci USA* 1981;78(9):5533–5537.
[38] Du Q, Wang B, Wei Z, et al. Genetic diversity and population structure of Chinese White poplar (Populus tomentosa) revealed by SSR markers. *J Hered* 2012;103(6):853–862.
[39] Chmielewski M, Meyza K, Chybicki IJ, et al. Chloroplast microsatellites as a tool for phylogeographic studies: the case of white oaks in Poland. *IForest* 2015;8:765–771.
[40] Yang XY, Liu C, Zeng XF, et al. Sequence structure an phylogenetic analysis of the chloroplast genomes of Alangium chinense (Lour.) Harms and its different subspecies. *Acta Pharm Sinica* 2022;57(10):3229–3239.
[41] Li X, Tan W, Sun J, et al. Comparison of four complete chloroplast genomes of medicinal and ornamental meconopsis species: genome organization and species discrimination. *Sci Rep* 2019;9(1):10567.
[42] Sun MT, Zhang JX, Huang TR, et al. Genome structure and variation of Reynoutria japonica Houtt. chloroplast genome. *Chin J Biotechnol* 2022;38(5):1953–1964.
[43] Hu SW, Ding YN, Bi GY, et al. Characterization and phylogenetic analysis of complete chloroplast genome of medicinal plant Polygonum aviculare. *Chin Tradit Herbal Drugs* 2022;53(9):2776–2785.
[44] Pierrehumbert RT, Abbot DS, Voigt A, et al. Climate of the neoproterozoic. *Annu Rev Earth Planet Sci* 2011;39:417–460.
[45] Prave AR, Condon D, Hoffmann KH, et al. Duration and nature of the end-Cryogenian (Marinoan) glaciation. *Geology* 2016;44(8):G38089.1.
[46] Zhao Chunsheng, Zhang Ziwei, Sun Linlin, et al. Genome sequencing provides potential strategies for drug discovery and synthesis. Acupunct Herb Med. 2023;3(4):244–255.
[47] Xu HD. Stratigraphy and sedimentary features of paleogene sediments of bohai bay basin. *Mar Geol Res* 1981;1(2):10–27.