scientific reports

Check for updates

OPEN

Comparative and phylogenetic analyses of the chloroplast genomes of species of Paeoniaceae

Liwei Wu[1,2], Liping Nie[1,2], Qing Wang[1,2], Zhichao Xu[1,2], Yu Wang[1], Chunnian He[1], Jingyuan Song[1,2] & Hui Yao[1,2]✉

Plants belonging to family Paeoniaceae are not only economically important ornamental plants but also medicinal plants used as an important source of traditional Chinese medicine. Owing to the complex network evolution and polyploidy evolution of this family, its systematics and taxonomy are controversial and require a detailed investigation. In this study, three complete chloroplast genomes of sect. *Paeonia*, one of the sections of *Paeonia*, were sequenced and then analysed together with 16 other published chloroplast genomes of Paeoniaceae species. The total lengths of the chloroplast genomes of these species were 152,153–154,405 bp. A total of 82–87 protein-coding genes, 31–40 tRNA genes and 8 rRNA genes were annotated. Bioinformatics analysis revealed 61–74 simple sequence repeats (SSRs) in the chloroplast genomes, most of which have A/T base preference. Codon usage analysis showed that A/U-ending codons were more positive than C/G-ending codons, and a slight bias in codon usage was observed in these species. A comparative analysis of these 19 species of Paeoniaceae was then conducted. Fourteen highly variable regions were selected for species relationship study. Phylogenetic analysis revealed that the species of sect. *Paeonia* gathered in one branch and then divided into different small branches. *P. lactiflora*, *P. anomala*, *P. anomala* subsp. *veitchii* and *P. mairei* clustered together. *P. intermedia* was related to *P. obovata* and *P. obovata* subsp. *willmottiae*. *P. emodi* was the sister to all other species in the sect. *Paeonia*.

Paeonia is a single genus in family Paeoniaceae, which is derived from family Ranunculaceae. Since the Swedish taxonomist Carl von Linne (1735) established the genus *Paeonia*, the widely used classification systems back then, such as the Engler system and the Hooker system, placed *Paeonia* in family Ranunculaceae and remained classified under this family for over 200 years[1]. At the beginning of the twentieth century, Worsdell discovered that the stamens of *Paeonia* develop centrifugally unlike the other genera in Ranunculaceae. Consequently, *Paeonia* was finally separated from this family and classified under Paeoniaceae, but it was still placed in Order Ranunculales[2]. Numerous extensive studies on *Paeonia* covering plant morphology, anatomy, palynology, embryology, phytochemistry, chromosome number determination and karyotyping, plant reproductive genetics, and phytogeography have led to the unanimous opinion supporting the establishment of Paeoniaceae[3]. In 1946, Stern divided *Paeonia* into three sections, namely, sect. *Moutan*, sect. *Paeonia* and sect. *Onaepia*[4]. Species of sect. *Paeonia* and sect. *Onaepia* are herbaceous, whereas species of sect. *Moutan* are subshrubs[5]. Sect. *Onaepia* has only two species, which are distributed in western North America. Sect. *Paeonia*, which is the most diverse section, has over 20 species that are widely distributed in temperate climate areas of Eurasia. Seven species and two subspecies of sect. *Paeonia* are distributed in China. Eight species and six subspecies of sect. *Moutan*, which are mainly distributed in the southwest and northwest of China, are endemic to the country[6]. Paeoniaceae plants are economically important ornamental plants known for their attractive flowers. Moreover, these plants have high medicinal value. Monoterpene glycosides, flavonoids, tannins, stilbene, triterpenoids and other compounds have been found in species of Paeoniaceae[7–16]. These compounds have antioxidation, antitumor and antipathogenic properties and play a role in immune system regulation, cardiovascular system protection, central nervous system protection and optic nerve protection[17–23].

Section *Paeonia* is the only section with chromosome ploidy changes. Species of sect. *Moutan* and sect. *Onaepia* are all diploid, whereas species of sect. *Paeonia* have diploid and tetraploid chromosomes[24]. The existence of

[1]Key Laboratory of Bioactive Substances and Resources Utilization of Chinese Herbal Medicine, Ministry of Education, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100193, China. [2]Engineering Research Center of Chinese Medicine Resources, Ministry of Education, Beijing 100193, China. ✉email: scauyaoh@sina.com

different ploidy in sect. *Paeonia* makes its phylogenetic relationship very complicated. The tetraploid species in this section were initially thought to be homologous tetraploids[4,25]. However, morphological, cytogenetic and molecular phylogenetic studies showed that this section has allotetraploid species, and some tetraploid groups originated from interspecific hybridisation of two different subgroups[26–35]. Although evidence exists that tetraploid groups are mostly of heterogenetic origin, the origin and classification of sect. *Paeonia* are controversial because of the consistent karyotype, similar morphology and overlapping geographical distributions of species belonging to this section[36]. In classical taxonomy, the use of phenotypic traits alone to infer phylogenetic relationships between taxa with different genotypes is replete with problems[33]. These taxonomic problems caused by different understandings of morphological variations can be resolved using molecular markers independent of morphological features. Early researchers focused on the study of DNA fragment-labelling techniques or phylogenetic analysis based on nuclear or chloroplast DNA fragments[37–43]. Based on the results of ITS and *matK* phylogenetic analysis, Sang et al.[33] constructed the reticular evolution model diagram of sect. *Paeonia*. However, phylogenetic analysis showed that the results of ITS and *matK* only provided partial information on the origin of allotetraploid groups[36]. Owing to chromosomal ploidy, complex network evolution and polyploidy evolution[26], limited nuclear or chloroplast DNA fragments cannot provide sufficient phylogenetic information to effectively solve the interspecies relationships of sect. *Paeonia*. Research on the genetic diversity of sect. *Paeonia* is relatively slow, and related studies at the molecular level are not comprehensive[44]. The relationships among species of sect. *Moutan* have such problems because of their complex evolution and phylogeny[32,33,45–49].

The chloroplast genome is an organ independent of the nuclear genome. The chloroplast genome can be maternally inherited, has highly conserved gene content and order and has a slow molecular evolution and a low recombination rate, making it an ideal material for species authentication and phylogenetic studies[50–53]. Most of the chloroplast genomes of angiosperms have a circular tetrad structure, which consists of two inverted repeats (IRs), a large single copy (LSC) and a small single copy (SSC)[54]. The chloroplast genome has been applied to phylogenetic analysis and species identification of multiple plants[55–60]. Therefore, we can use the chloroplast genome to analyse the relationship among species of Paeoniaceae. In this study, the complete chloroplast genomes of three species of sect. *Paeonia* were sequenced and analysed together with other Paeoniaceae species. Comparative and phylogenetic analyses were then performed on the chloroplast genomes of 19 species of Paeoniaceae, including 8, 10 and 1 chloroplast genomes of sect. *Paeonia*, sect. *Moutan* and sect. *Onaepia*, respectively.

Results and discussion

Statistics and genetic composition of 19 Paeoniaceae chloroplast genomes. The chloroplast genomes of 19 Paeoniaceae species were all classical tetrad structures containing an LSC, an SSC and a pair of IRs (Fig. 1). The total lengths of the chloroplast genomes were 152,153 (*P. ostii*)–154,405 bp (*P. delavayi*). Total GC contents ranged from 38.32% (*P. ostia* and *P. rockii*) to 38.55% (*P. brownii*). The lengths of the IR, LSC and SSC regions were 24,729–26,049, 84,241–86,316, and 16,679–17,059 bp, respectively. The GC contents of the four regions were not balanced. The IR regions had the highest GC content (42.98–43.16%), followed by the LSC (36.63–36.83%) and the SSC regions (32.57–33.02%) (Supplementary Table S1).

The structure and gene composition of the chloroplast genomes of Paeoniaceae species can be divided into 14 categories (Fig. 1). The proteins produced by different combinations of domains are different in nature, and the identification of protein domains is particularly important for analysing protein functions. The protein functional domains of protein-coding genes in the Paeoniaceae species are listed in Fig. 2. A total of 82–87 protein-coding genes, 31–40 tRNA genes and 8 rRNA genes were annotated in the Paeoniaceae species. In the three chloroplast genomes obtained in this study, seven protein-coding genes (*rpl2*, *rpl23*, *ycf2*, *ycf15*, *ndhB*, *rps7* and *rps12*), seven tRNAs (*trnI-CAU*, *trnL-CAA*, *trnV-GAC*, *trnI-GAU*, *trnA-UGC*, *trnR-ACG* and *trnN-GUU*) and four rRNAs (*rrn16*, *rrn23*, *rrn4.5* and *rrn5*) were located in the IR regions. Introns play an important role in the regulation of gene expression, and it can enhance the expression of exogenous genes at specific loci of plants and produce ideal agronomic traits[61]. Among the protein-coding genes, 18 genes contained introns, of which 3 genes (*clpP*, *rps12* and *ycf3*) contained two introns, whereas the remaining 15 genes contained only one intron (Supplementary Table S2). *rps12* gene is a trans-splicing gene with a 5′ end in the LSC region and a 3′ end in the IR region, similar to that of many other plants[62–64].

Analysis of codon usage bias of the chloroplast genomes. The choice of synonymous codons for amino acids encoded by an organism's genes is not completely random, and there is codon usage bias[65]. Codon usage bias not only plays an important regulatory role in gene expression level but also helps to improve the accuracy and efficiency of translation[66,67]. In addition to being affected by selection and mutation, codon usage is affected by tRNA abundance, base composition, gene position on chromosomes, gene length and expression level, amino acid hydrophobicity and mRNA secondary structure[68–74]. Paeoniaceae species can be divided into groups A and B according to the codon usage of chloroplast genomes. Group B included *P. lactiflora*, *P. obovata*, *P. rockii* and *P. rockii* subsp. *taibaishanica*, whereas group A included the remaining 15 species.

The relative synonymous codon usage (RSCU) of the chloroplast genomes of Paeoniaceae species was calculated on the basis of all protein-coding genes (Supplementary Table S3). Results showed that the chloroplast genomes of Paeoniaceae species contained 64 types of codons encoding 20 amino acids. In group A, of all amino acid codons, leucine had the highest number of codons, whereas cysteine had the lowest number of codons. Thirty-one codons were found with an RSCU of > 1, of which 29 were A/U-ending codons; 33 codons were found with an RSCU of ≤ 1, of which 30 were G/C-ending codons. The highest RSCU value was recorded for UUA and

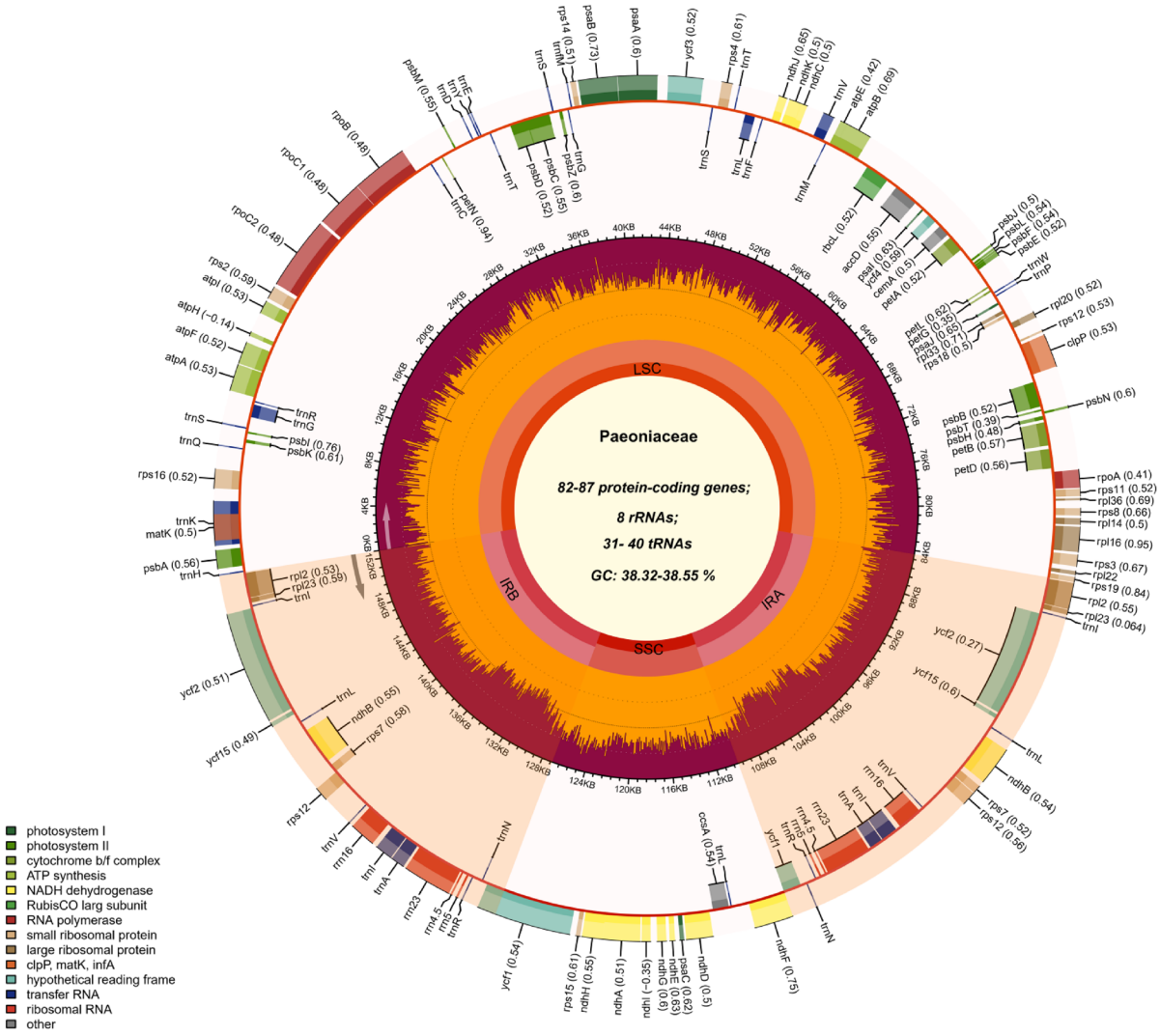

Figure 1. Chloroplast genome map of Paeoniaceae species, using *P. intermedia* as the template. The gradient GC content of the genome was plotted in the second circle with zero level based on the outer circle. The gene names and their codon usage bias were labeled on the outermost layer. The gene specific GC content was depicted with the proportion of shaded areas. Represented with arrows, the transcription directions for the inner and outer genes were listed clockwise and anticlockwise, respectively.

the lowest for UAC, which encode leucine and tyrosine, respectively. In group B, serine had the highest number of codons, and methionine had the lowest number of codons. Furthermore, 29 (*P. obovata*) and 30 (*P. lactiflora*, *P. rockii* and *P. rockii* subsp. *taibaishanica*) codons were found with an RSCU of > 1, of which 26 (*P. lactiflora* and *P. obovata*) and 27 (*P. rockii* and *P. rockii* subsp. *taibaishanica*) were A/U-ending codons. Moreover, 34 (*P. lactiflora*, *P. rockii* and *P. rockii* subsp. *taibaishanica*) and 35 (*P. obovata*) codons were found with an RSCU of ≤ 1, of which 29 (*P. obovata*) and 32 (*P. lactiflora*, *P. rockii* and *P. rockii* subsp. *taibaishanica*) were G/C-ending codons. The highest RSCU value was recorded for AGA and the lowest for CUG, which encode arginine and leucine, respectively. To conclude, A/U-ending codons were more positive than G/C-ending codons (Fig. 3).

GC refers to the total content of all codons G and C, and GC3s pertains to the frequency of G and C bases in the third codon base of synonymous codons encoding the same amino acid. GC reflects the strength of directional mutation pressure, and GC3s is closely related to codon bias[75]. The main difference in synonymous codons is reflected in the third base, and a change in this base of codons usually does not cause changes in encoded amino acids. Therefore, the selection pressure on the third base of the codon is less selective. GC3s is used as an important basis for analysing codon usage pattern[76]. GC and GC3s in the codons of these 19 chloroplast genomes were all less than 0.5, indicating that the chloroplast genomes of Paeoniaceae species tended to use A/T bases and A/T-ending codons. Codon adaptation index values and effective number of codon values indicated a slight bias in codon usage in the Paeoniaceae species. Frequency of optimal codons was relatively low. In addition, the hydrophobicity of the protein (i.e., Gravy) and the aromatic protein (i.e., Aromo) had little effect on codon usage

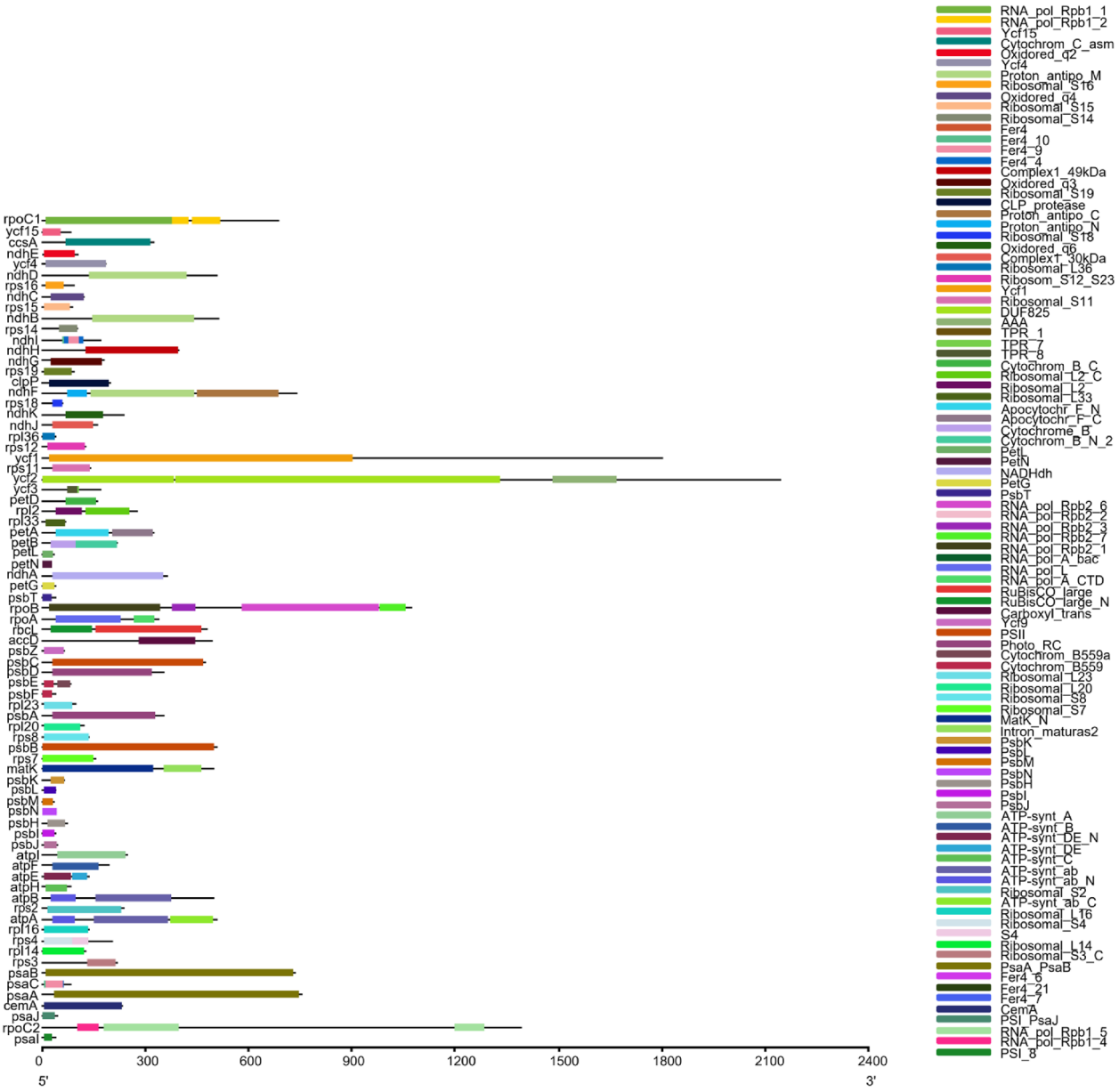

Figure 2. Protein functional domains of protein-coding genes in Paeoniaceae species.

bias. Compared with those in group A, the species in group B had higher GC and GC3s contents and slighter codon usage bias, and Gravy and Aromo had a greater influence on codon usage bias (Table 1).

Long repeat sequences and SSRs. Long repeats play an important role in genome rearrangement and are often used to study phylogenetic relationships between species; moreover, they promote intermolecular recombination in the chloroplast genomes of plants to produce diversity[77]. Long repeat sequences include forward, palindrome, reverse and complement. For all repeat types, repeat length is ≥ 30 bp and sequence similarity is ≥ 90%. In Paeoniaceae species, our results revealed 36–63 long repeats, most of which were forward (17–29) and palindrome (18–31) repeats. Complement repeats were the least distributed and found only in *P. anomala*, *P. anomala* subsp. *veitchii*, *P. lactiflora*, *P. mairei* and *P. rockii*. In addition, the length of these repeats was mainly within the range of 30–39 bp. Repeats with a length of ≥ 70 bp only existed in *P. brownie*, *P. ostia* and *P. rockii* subsp. *taibaishanica* (Fig. 4) (Supplementary Table S4).

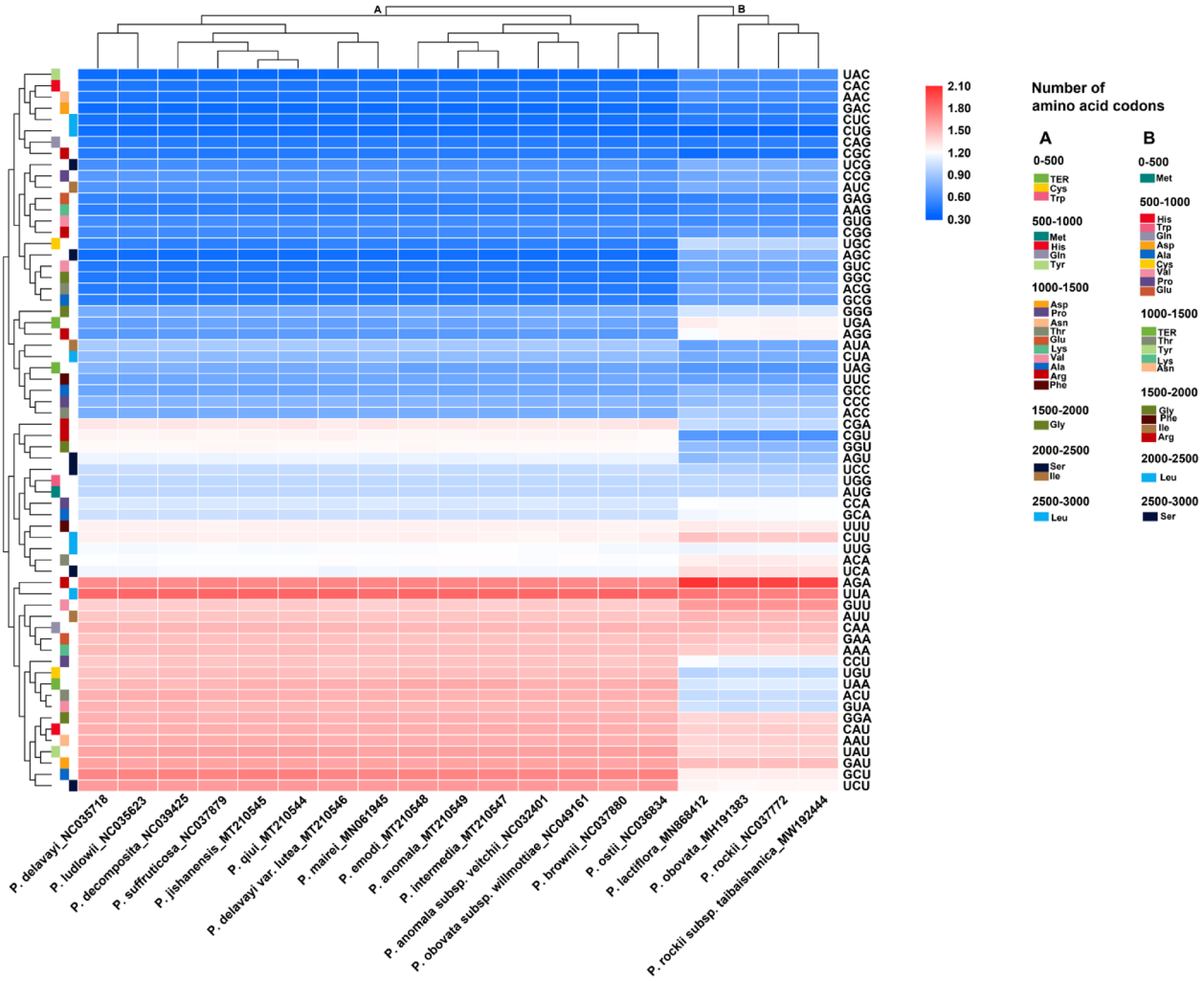

Figure 3. Heat map of relative synonymous codon usage (RSCU) values among Paeoniaceae species.

SSRs, also known as microsatellite sequences, are a kind of tandem repeat sequences consisting of 1–6 repeating nucleotide units that are widely distributed throughout chloroplast genomes[78]. Owing to their high polymorphism, SSRs are increasingly used as molecular markers, in species identification and in studying population genetics and phylogenetic relationships[79–81]. A total of 61–74 SSRs were identified in the chloroplast genomes of the Paeoniaceae species. In addition, the base composition of the repeating motifs from mononucleotide repeats to trinucleotide repeats had a certain base preference, mainly the repeating motifs rich in A–T. In these SSRs, mononucleotide repeats were the largest in number, which were found 39–49 times in these chloroplast genomes. A/T repeats (91.7–100%) were the most common mononucleotide repeats, whereas the majority of dinucleotide repeat sequences comprised of AT/AT repeats (88.9–92.9%), and all of trinucleotide repeats were AAT/ATT, except for *P. delavayi*. These results were consistent with A-T enrichment in complete chloroplast genomes[82]. Moreover, compared with polyC and polyG, polyA and polyT occupy a relatively high proportion in the SSRs of many plants[83]. ACG/CGT, AAAAG/CTTTT and AATAT/ATATT were found to be unique in *P. delavayi*, *P. anomala* and *P. jishanensis*, respectively (Fig. 4) (Supplementary Table S5).

Comparative analysis of chloroplast genomes of Paeoniaceae. In this study, the complete chloroplast genomes of 19 species of Paeoniaceae were compared using mVISTA[84] with the *P. intermedia* genome as the reference genome (Fig. 5). Overall, the comparative genomic analysis revealed that the 19 Paeoniaceae chloroplast genomes were relatively conserved. Intergenic spacers and intron regions showed more variations than protein-coding regions. Most protein-coding regions had a very high degree of conservation (most had > 90% similarity), and rRNA genes (*rrn4.5*, *rrn5*, *rrn16* and *rrn23*) were highly conserved with almost no variation. Variations in the SSC and LSC regions were considerably greater than those in the IR regions, similar to studies in other plants[85–89].

Species	T3s	C3s	A3s	G3s	GC3s	GC	CAI	ENc	Fop	Gravy	Aromo	L_sym	L_aa
P. intermedia	0.4582	0.1788	0.4213	0.1914	0.283	0.386	0.166	50.92	0.354	− 0.089477	0.111029	24,652	25,714
P. emodi	0.4585	0.1795	0.4207	0.1909	0.284	0.387	0.166	50.91	0.354	− 0.088465	0.111016	24,558	25,618
P. anomala	0.4581	0.1791	0.4214	0.191	0.283	0.386	0.166	50.9	0.354	− 0.089123	0.110835	24,651	25,714
P. anomala subsp. *veitchii*	0.4583	0.1783	0.4216	0.1915	0.283	0.386	0.166	50.9	0.354	− 0.088383	0.110857	24,698	25,763
P. mairei	0.4574	0.1796	0.421	0.1907	0.284	0.387	0.166	50.96	0.354	− 0.083818	0.110868	24,936	26,004
P. obovata subsp. *willmottiae*	0.4583	0.179	0.4206	0.1916	0.284	0.387	0.166	50.95	0.354	− 0.088775	0.111241	24,655	25,719
P. jishanensis	0.4595	0.1782	0.4223	0.1897	0.282	0.386	0.166	50.8	0.354	− 0.08939	0.111236	24,645	25,711
P. decomposita	0.4593	0.1784	0.4219	0.1895	0.282	0.386	0.166	50.81	0.354	− 0.081256	0.111128	24,722	25,790
P. qiui	0.4596	0.1783	0.4223	0.1897	0.282	0.386	0.166	50.79	0.354	− 0.090311	0.111159	24,644	25,711
P. ostii	0.4596	0.18	0.4214	0.1892	0.283	0.386	0.167	50.87	0.354	− 0.086684	0.1115	24,642	25,713
P. suffruticosa	0.4586	0.1789	0.422	0.1905	0.283	0.386	0.166	50.86	0.354	− 0.088544	0.111268	24,493	25,551
P. ludlowii	0.4569	0.1807	0.4208	0.1904	0.284	0.387	0.166	51.05	0.355	− 0.084929	0.110735	24,942	26,017
P. delavayi var. *lutea*	0.4572	0.1805	0.4214	0.1899	0.284	0.387	0.167	50.99	0.355	− 0.089308	0.110296	24,160	25,214
P. delavayi	0.4571	0.1802	0.4212	0.1906	0.284	0.387	0.166	51.03	0.354	− 0.082886	0.110866	24,763	25,833
P. brownii	0.4566	0.1813	0.4209	0.191	0.285	0.387	0.166	51.03	0.354	− 0.089699	0.110977	24,541	25,600
P. lactiflora	0.4117	0.2366	0.3993	0.2224	0.351	0.395	0.16	53.75	0.363	− 0.308232	0.146809	23,536	24,440
P. obovata	0.413	0.2353	0.3955	0.2294	0.354	0.395	0.158	54.1	0.359	− 0.320514	0.147034	23,524	24,457
P. rockii	0.4139	0.2347	0.3959	0.2289	0.353	0.394	0.158	54	0.359	− 0.321144	0.147478	23,538	24,485
P. rockii subsp. *taibaishanica*	0.4138	0.2348	0.3962	0.2283	0.352	0.394	0.158	54.01	0.359	− 0.319911	0.147729	23,538	24,484

Table 1. Codon usage of the Paeoniaceae species. *T3s/C3s/A3s/G3s/GC3s* the thymine/cytosine/adenine/guanine/GC content at synonymous third codon position, *GC* the total GC content, *CAI* codon adaptation index, *ENc* effective number of codons, *Fop* Frequency of optimal codons, *Gravy* the influence of protein hydrophobicity on codon usage bias, *Aromo* the influence of aromatic protein on codon usage bias, *L_sym* number of synonymous codons, *L_aa* total number of synonymous and non-synonymous codons.

A co-linear analysis of the 19 Paeoniaceae chloroplast genomes was conducted with the *P. intermedia* genome as the reference genome. Results showed that the entire genome sequence was a homologous region with no big indels. The 19 chloroplast genomes connected with a line, indicating that the chloroplast genomes of these species were relatively conserved, and no rearrangement occurred in gene organisation (Fig. 6).

Mutational hotspots of shared genes and intergenic spacers of the chloroplast genomes of the 19 Paeoniaceae species were identified by DnaSP[90]. The intergenic spacers had more polymorphisms (average Pi = 0.00955) than the gene regions (average Pi = 0.00393). Moreover, the largest nucleic acid variation was observed in the SSC regions (average Pi in intergenic spacers = 0.01107; average Pi in gene regions = 0.00568), followed by that in the LSC regions (average Pi in intergenic spacers = 0.01045; average Pi in gene regions = 0.00408) and that in the IR regions (average Pi in intergenic spacers = 0.00391; average Pi in gene regions = 0.00124). These results were consistent with those of mVISTA analysis. Seven protein-coding genes (*rps18*, *ndhF*, *rps3*, *rpl16*, *psbH*, *rps16* and *matK*) positioned at the single copy regions exhibited high Pi values (> 0.008) (Fig. 7A). By comparison, seven intergenic spacers (*petG-trnW-CCA*, *petA-psbJ*, *petL-petG*, *psbK-psbI*, *accD-psaI*, *ndhE-ndhG* and *rpl14-rpl16*) showed high diversity values (> 0.015) (Fig. 7B).

Phylogenetic analysis of Paeoniaceae. Chloroplast genomes play an important role in phylogenetic studies[91,92]. In the current study, the complete chloroplast genome sequences of 19 Paeoniaceae species and 32 Ranunculaceae species were used to construct a phylogenetic tree. *Stephania tetrandra* served as the outgroup (Fig. 8). Results showed that all nodes in the phylogenetic tree had high bootstrap values. Paeoniaceae species clustered in one branch, whereas Ranunculaceae species were clearly distinguished from Paeoniaceae species, supporting the argument that Paeoniaceae is a family independent from Ranunculaceae. Species of subsect. *Vaginatae* and subsect. *Delavayanae* of sect. *Moutan* clustered in different branches, and the species relationship in sect. *Moutan* was consistent with that reported by a previous study[93]. With regard to species of sect. *Paeonia*, *P. lactiflora*, *P. anomala*, *P. anomala* subsp. *veitchii* and *P. mairei* clustered together. Pan[36] found that *P. sterniana* is closely related to them. Xia[43] analysed the genetic relationship of sect. *Paeonia* and found that *P. lactiflora* is closely related to *P. anomala* and *P. anomala* subsp. *veitchii*. Zhang et al.[94] found that *P. anomala*, *P. anomala* subsp. *veitchii* and *P. mairei* are closely related and far from *P. obovata* according to the results of ML-based phylogenetic analysis using complete chloroplast genomes. The results of the current study were consistent with those of the aforementioned studies. *P. obovata*, *P. obovata* subsp. *willmottiae* and *P. intermedia* clustered together, and *P. emodi* was the sister to all other species in the sect. *Paeonia*. This branching pattern was consistent with that of the phylogenetic tree constructed by Zhou et al. by using chloroplast markers[95]. Furthermore,

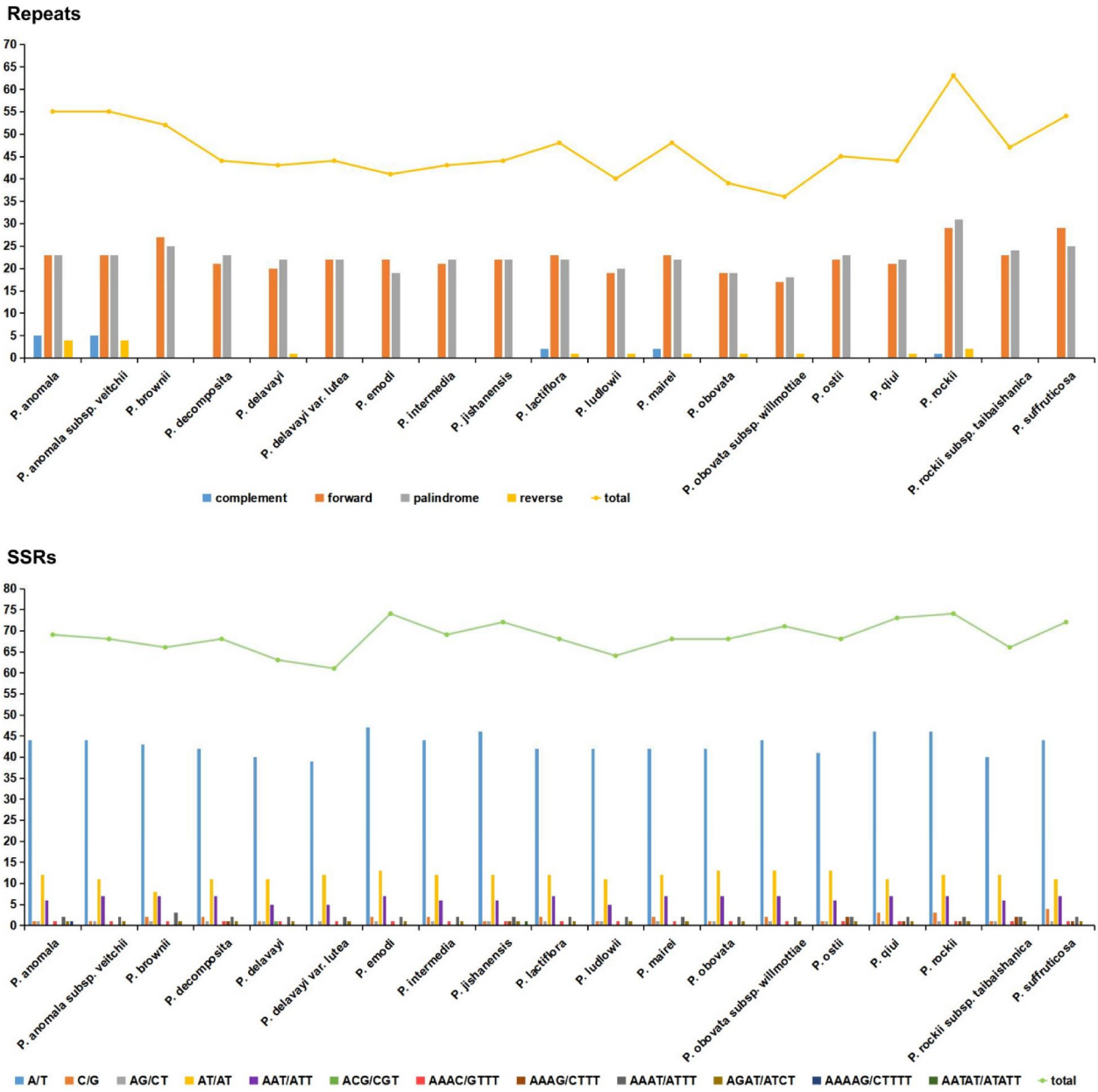

Figure 4. Types and amounts of repeats and SSRs in Paeoniaceae chloroplast genomes.

the relationship between sect. *Moutan* and sect. *Onaepia* was close. The result of phylogenetic tree based on SNPs (single-nucleotide polymorphisms) showed that the species relationship in Paeoniaceae was consistent with that based on the complete chloroplast genome sequences (Fig. 9). However, the relationship between sect. *Moutan* and sect. *Paeonia* was close in the tree based on SNPs, which was more coincident with the geographical distribution.

Among the 14 highly variable regions, *matK* appeared suitable for phylogenetic analysis of the species of sect. *Paeonia*, which was consistent with that based on complete chloroplast genomes. The other 13 highly variable regions were found to be unsuitable for the identification and phylogenetic analysis of Paeoniaceae species (Supplementary Fig. S1) mainly because of inadequate variations provided by a limited number of DNA loci, which was insufficient to distinguish these conservative species[96]. A previous study also demonstrated that complete chloroplast genome sequences have a higher resolution than highly variable regions and can be used to identify related species[97], consistent with the current study.

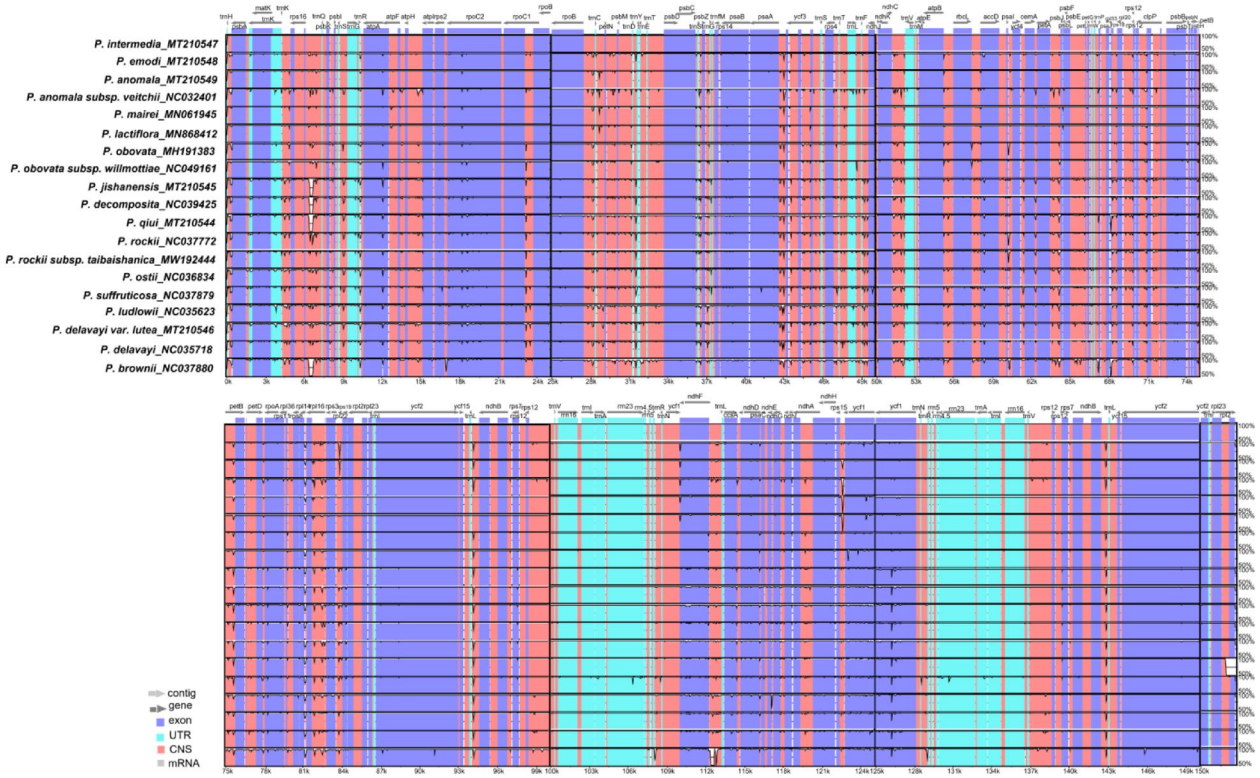

Figure 5. Global alignment of chloroplast genomes of 19 Paeoniaceae species. The x-axis represents the coordinates in the chloroplast genome. The y-axis indicates the average percent identity of sequence similarity in the aligned regions, ranging between 50 and 100%. Genome regions are color coded as protein coding, rRNA coding, tRNA coding, or conserved noncoding sequences (CNS).

Materials and methods

DNA sources. Fresh leaves of *P. intermedia*, *P. emodi* and *P. anomala* were collected from a garden in Xiaohongcun, Luanchuan County, Henan Province, China, which were transplanted from Tacheng in Xinjiang Autonomous Region (*P. intermedia* and *P. anomala*) and Shannan in Xizang Autonomous Region (*P. emodi*). These three species were identified by Prof. Peigen Xiao and Prof. Chunnian He from the Institute of Medicinal Plant Development (IMPLAD), Chinese Academy of Medical Sciences and Peking Union Medical College. Voucher specimens were deposited in the herbarium at IMPLAD, and the ID numbers were Y19075 (*P. intermedia*), Y19076 (*P. emodi*), and Y19078 (*P. anomala*). The collected fresh leaves were stored in a − 80 ℃ refrigerator until use.

DNA extraction, sequencing, assembly and annotation. Total DNA was extracted using a DNease plant mini kit (Qiagen, Germany). Total DNA concentration was detected using a microspectrophotometer (Nanodrop 2000, USA), and total DNA quality was detected via 1% agarose gel electrophoresis. The DNA was then used to generate libraries with an average insert size of 500 bp and sequenced using Illumina Hiseq X in accordance with standard protocols. Paired-end sequencing was performed to obtain 150 bp sequences at both ends of each read. The NGS data was stated in Supplementary Table S6. Low-quality regions in the original data were trimmed by Trimmomatic software[98], and mapped back using bwa to get the sequencing depth. The average genome coverage depth for *P. intermedia*, *P. emodi*, and *P. anomala* was 937×, 1037×, and 1113×, respectively. The Basic Local Alignment Search Tool database was constructed from the chloroplast genome sequences published on the National Centre for Biological Information. Clean reads were then compared with this database, and mapped reads were extracted according to coverage and similarity. The extracted reads were spliced into several contigs by using SOAPdenovo 2[99] and NOVOPlasty[100]. The contigs were connected to complete chloroplast genome sequences by using the SSPACE software[101], and gaps were filled using the GapCloser software[99]. The sequences were initially annotated by using the CPGAVAS software[102] and the GeSeq[103] and corrected manually. tRNAs were annotated using the tRNAscan-SE software[104]. Genes, introns and the boundaries of coding regions were compared with reference sequences.

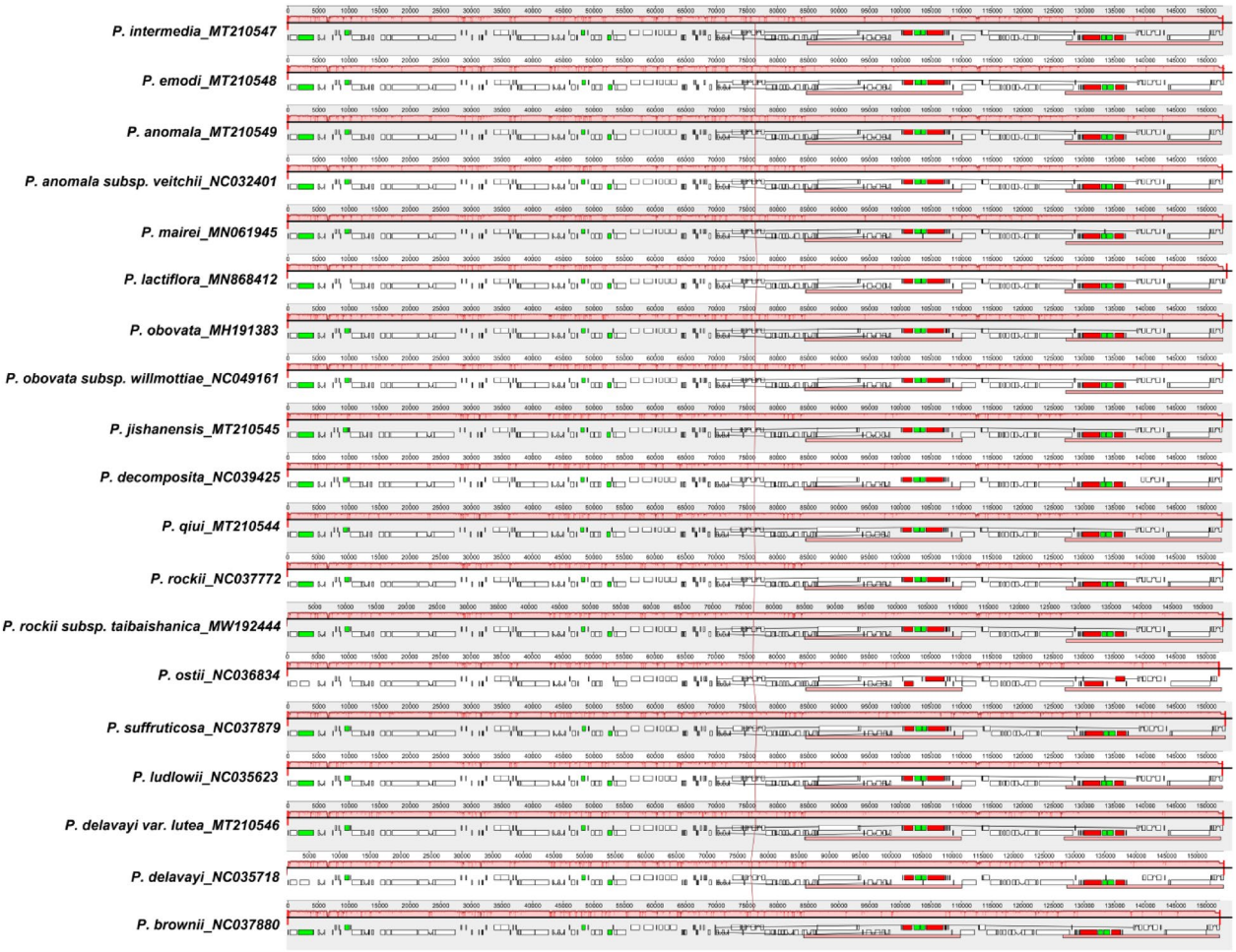

Figure 6. Co-linear analysis of 19 Paeoniaceae species chloroplast genomes. Local collinear blocks are represented by blocks of the same color connected by lines.

Structural, comparative and phylogenetic analyses. Chloroplast genome maps were generated using Chloroplot[105] and then manually corrected. Protein functional domains of protein-coding genes were searched by Pfam[106]. The CodonW software[107] was adopted to analyse the usage of codon. SSRs and long repeat sequences were detected using the MISA[108] and REPuter[109], respectively. The chloroplast genomes were compared by using the mVISTA software[84] to detect variations within the Paeoniaceae species. Chloroplast genome sequence homology and collinearity were analysed using the Mauve software[110]. The nucleotide diversity values (Pi) of chloroplast genomes of Paeoniaceae species were computed using DnaSP v5.10[90], and 14 highly variable regions were selected. All chloroplast genome sequences were aligned by MAFFT software[111]. Tree models were selected and phylogenetic trees were constructed by using the IQTREE software[112]. Phylogenetic trees were constructed based on the complete chloroplast genomes, SNPs, and 14 highly variable regions by ML methods. The SNPs were from the chloroplast genome sequences of Paeoniaceae species after alignment and removal of all indels[113]. ML analysis was conducted with a bootstrap of 1000 repetitions based on the TVM + F + R4 (complete chloroplast genomes), TVMe + ASC + R2 (SNPs), K3Pu + F (*accD-psaI*, *matK*, *rps3* and *rps16*), TPM2u + F (*ndhE-ndhG*), K3Pu + F + I (*ndhF*, *psbK-psbI* and *rpl16*), TIM + F + G4 (*petA-psbJ*), K2P + I (*petG-trnW-CCA*), and F81 + F (*petL-petG*, *psbH*, *rpl14-rpl16* and *rps18*) models.

All the experiment has been done in the accordance with relevant institutional, national, and international guidelines and legislation.

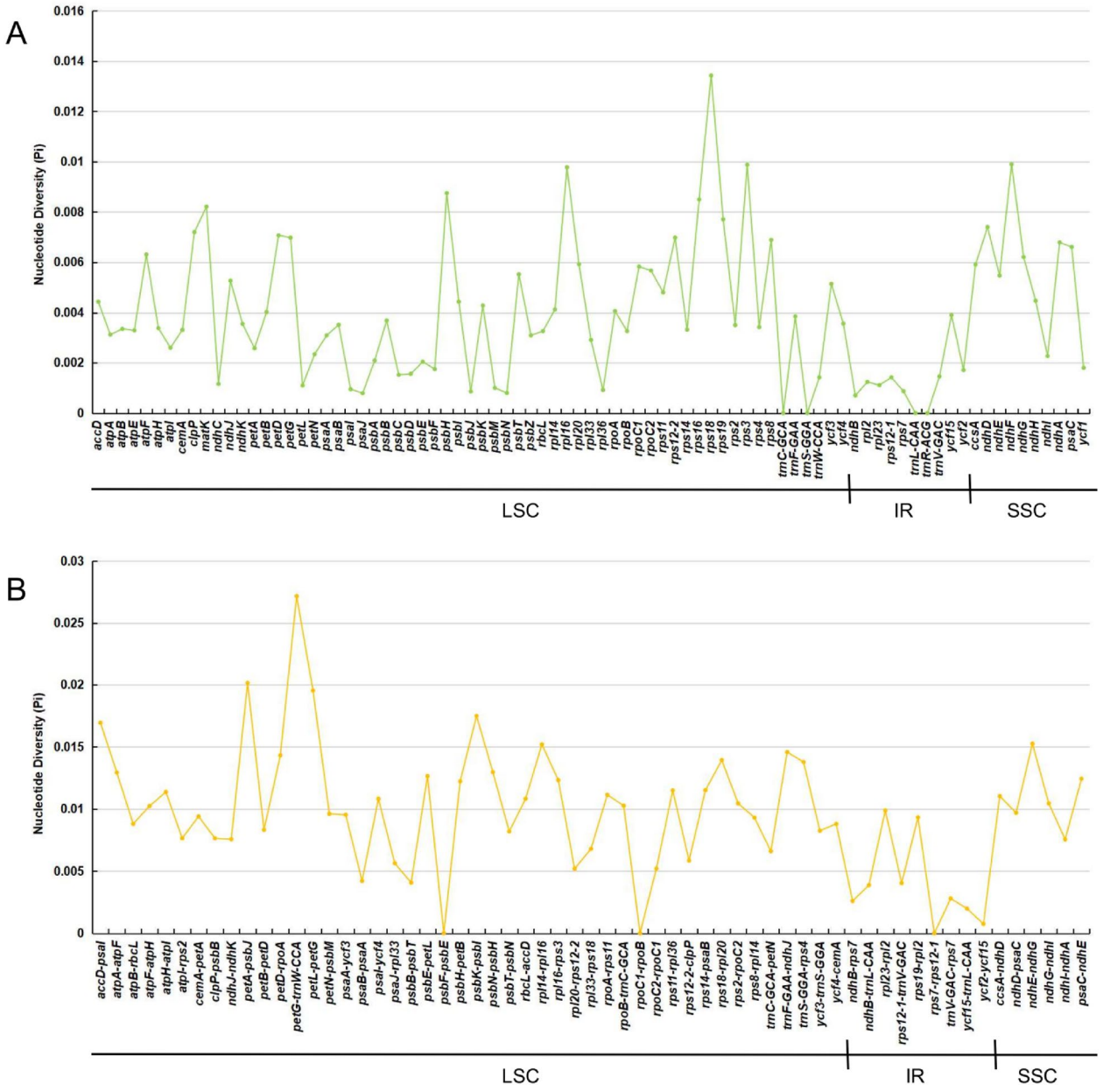

Figure 7. Nucleotide diversity (Pi) of shared various regions in 19 Paeoniaceae species chloroplast genomes. (**A**) Pi values in the genes regions. (**B**) Pi values in the intergenic spacers regions.

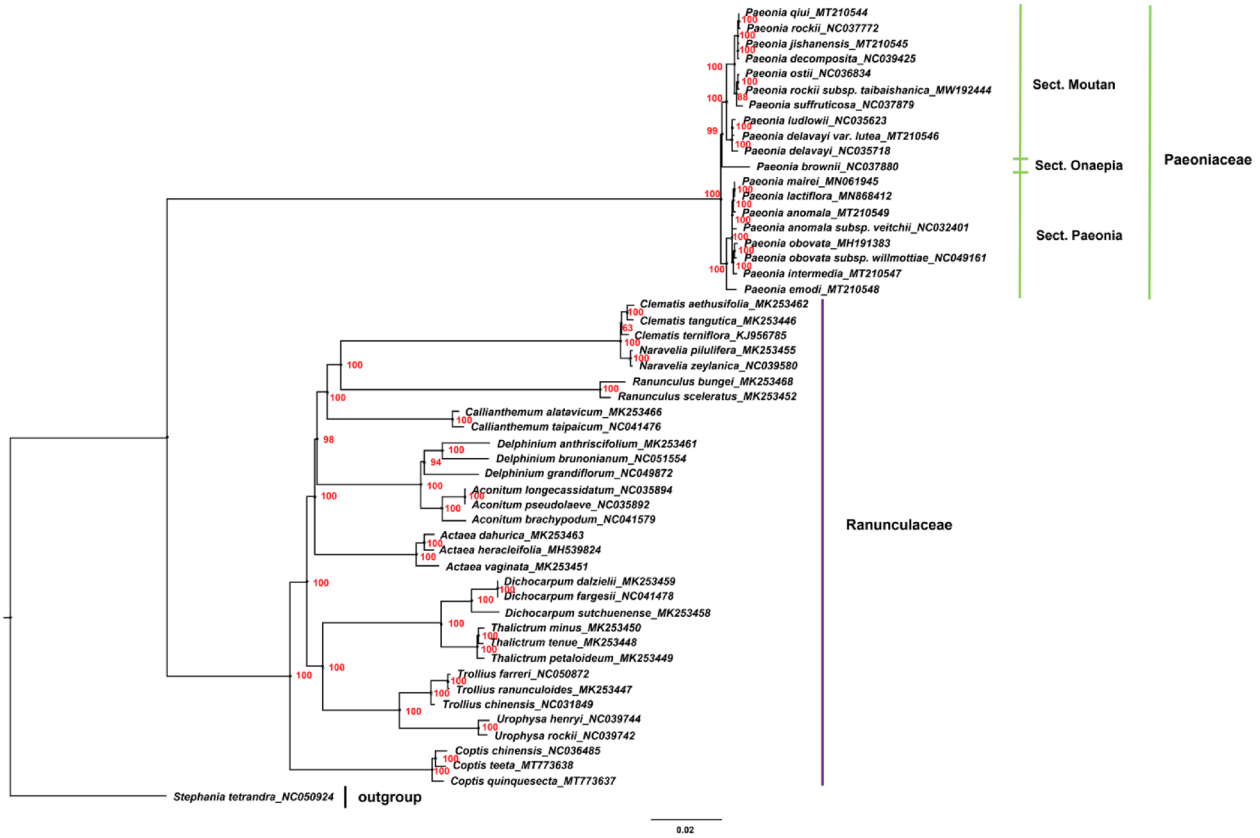

Figure 8. Phylogenetic tree constructed using Maximum Likelihood (ML) method based on the complete chloroplast genome sequences of 19 Paeoniaceae species and 32 Ranunculaceae species. Red numbers at nodes are values for bootstrap support.

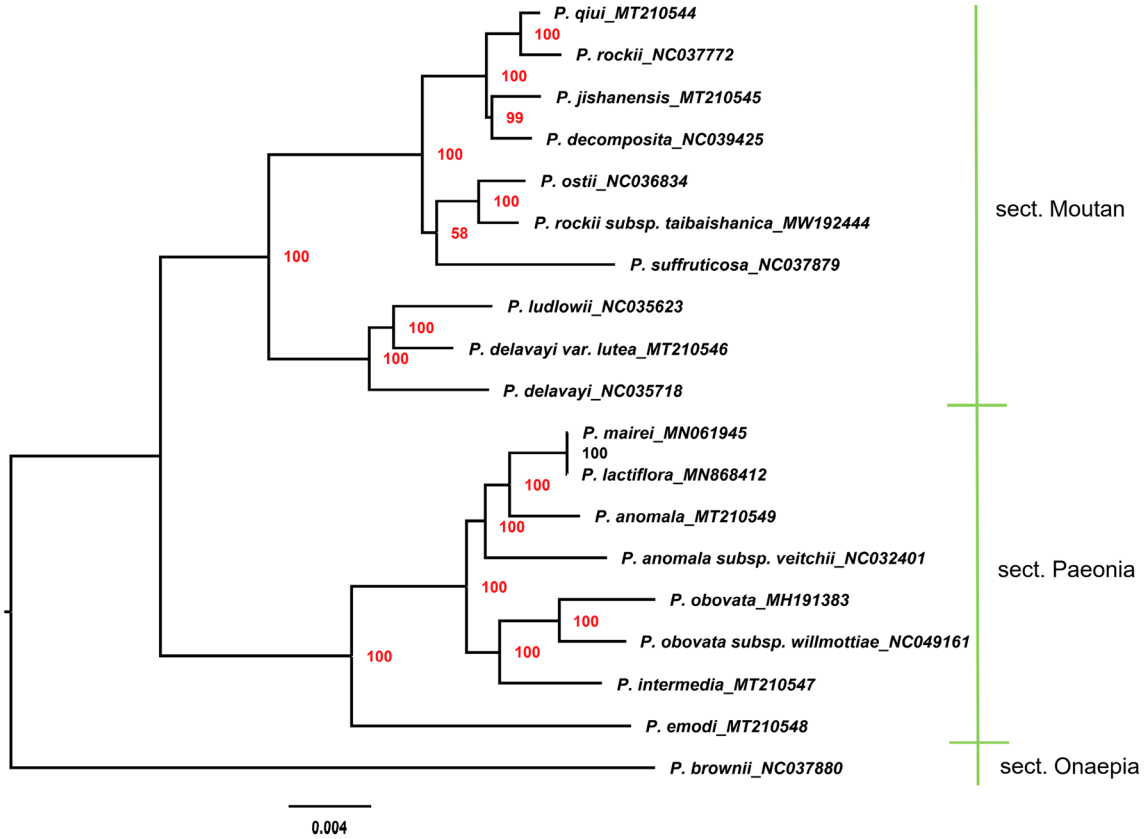

Figure 9. Phylogenetic tree constructed using Maximum Likelihood (ML) method based on SNPs of 19 Paeoniaceae species. Red numbers at nodes are values for bootstrap support.

Specimen collection statement. The collection of fresh leaves obtained the permission of the owner.

Data availability

The assembled chloroplast genomes of *P. intermedia*, *P. emodi* and *P. anomala* were deposited in GenBank with the accession numbers MT210547, MT210548 and MT210549. The sequences are available on NCBI now: https://www.ncbi.nlm.nih.gov/nuccore/MT210547, https://www.ncbi.nlm.nih.gov/nuccore/MT210548, https://www.ncbi.nlm.nih.gov/nuccore/MT210549.

Received: 22 February 2021; Accepted: 6 July 2021
Published online: 19 July 2021

References

1. Guo, X. F. Advances in classification of Chinese herbaceous peony. *J. Beijing For. Univ.* **24**, 99–102 (2002).
2. Wordell, W. C. A study of the vascular system in certain orders of the Ranales. *Ann. Bot.* **22**, 603 (1908).
3. Peng, Z. & Jiang, S. Distinctive characters and origin of Paeoniaceae. *J. Anhui Agric. Univ.* **27**, 209–213. https://doi.org/10.3969/j.issn.1672-352X.2000.03.001 (2000).
4. Stern, F. C. *A Study of the Genus Paeonia*. (The Royal Horticultural Society, 1946).
5. Fang, Q. B. Classification, distribution and medicinal use of Paeonia section Paeonia in China. *Res. Pract. Chin. Med.* **18**, 28–30. https://doi.org/10.3969/j.issn.1673-6427.2004.03.009 (2004).
6. Hong, D. Y., Pan, K. Y. & Turland, N. J. *Flora of China*, vol. 6. (Science Press & St.Louis: Missouri Botanical Garden Press, 2001).
7. Kamiya, K., Yoshioka, K., Saiki, Y., Ikuta, A. & Satake, T. Triterpenoids and flavonoids from *Paeonia lactiflora*. *Phytochemistry* **44**, 141–144. https://doi.org/10.1016/S0031-9422(96)00518-3 (1997).
8. Hsu, F. L., Lai, C. W. & Cheng, J. T. Antihyperglycemic effects of paeoniflorin and 8-debenzoylpaeoniflorin, glucosides from the root of *Paeonia lactiflora*. *Planta Med.* **63**, 323–325. https://doi.org/10.1055/s-2006-957692 (1997).
9. Tanaka, T., Fukumori, M., Ochi, T. & Kouno, I. Paeonianins A–E, new dimeric and monomeric ellagitannins from the fruits of *Paeonia lactiflora*. *J. Nat. Prod.* **66**, 759–763. https://doi.org/10.1021/np020608g (2003).
10. Kim, H. J. *et al.* Cytotoxic and antimutagenic stilbenes from seeds of *Paeonia lactiflora*. *Arch. Pharm. Res.* **25**, 293–299. https://doi.org/10.1007/bf02976629 (2002).
11. Takechi, M. & Tanaka, Y. Antiviral substances from the root of Paeonia species. *Planta Med.* **45**, 252–253. https://doi.org/10.1055/s-2007-971387 (1982).
12. Ikuta, A., Kamiya, K., Satake, T. & Saiki, Y. Triterpenoids from callus tissue cultures of Paeonia species. *Phytochemistry* **38**, 1203–1207. https://doi.org/10.1016/0031-9422(94)00445-Y (1995).
13. Lee, S. C., Kwon, Y. S., Son, K. H., Kim, H. P. & Heo, M. Y. Antioxidative constituents from *Paeonia lactiflora*. *Arch. Pharm. Res.* **28**, 775–783. https://doi.org/10.1007/bf02977342 (2005).

14. Murakami, N. *et al.* New bioactive monoterpene glycosides from Paeoniae Radix. *Chem. Pharm. Bull. (Tokyo)* **44**, 1279–1281. https://doi.org/10.1248/cpb.44.1279 (1996).
15. Stavri, M., Mathew, K. T., Bucar, F. & Gibbons, S. Pangelin, an antimycobacterial coumarin from *Ducrosia anethifolia*. *Planta Med.* **69**, 956–959. https://doi.org/10.1055/s-2003-45109 (2003).
16. Nawaz, H. R., Malik, A., Khan, P. M., Shujaat, S. & Rahman, A. A novel beta-glucuronidase inhibiting triterpenoid from *Paeonia emodi*. *Chem. Pharm. Bull. (Tokyo)* **48**, 1771–1773. https://doi.org/10.1248/cpb.48.1771 (2000).
17. Okubo, T. *et al.* The inhibition of phenylhydroquinone-induced oxidative DNA cleavage by constituents of Moutan Cortex and Paeoniae Radix. *Biol. Pharm. Bull.* **23**, 199–203. https://doi.org/10.1248/bpb.23.199 (2000).
18. Lee, S. M. *et al.* Paeoniae Radix, a Chinese herbal extract, inhibit hepatoma cells growth by inducing apoptosis in a p53 independent pathway. *Life Sci.* **71**, 2267–2277. https://doi.org/10.1016/s0024-3205(02)01962-8 (2002).
19. Li, X., Wang, Y., Wang, K. & Wu, Y. Renal protective effect of Paeoniflorin by inhibition of JAK2/STAT3 signaling pathway in diabetic mice. *Biosci. Trends* **12**, 168–176. https://doi.org/10.5582/bst.2018.01009 (2018).
20. Wang, Y. W. & Wang, Y. J. Pharmacological study and clinical application of total glucosides of peony in autoimmune diseases. *J. Zhejiang Univ. Tradit. Chin. Med.* **31**, 240. https://doi.org/10.16466/j.issn1005-5509.2007.02.060 (2007).
21. Liapina, L. A., Kondashevskaia, M. V., Ziadetdinova, G. A. & Uspenskaia, M. S. Comparative study of anticoagulants obtained from various extracts of *Paeonia anomala*. *Izv. Akad. Nauk. Ser. Biol.* **3**, 345–349. https://doi.org/10.1080/152165400410272 (2000).
22. Zhang, X. J. *et al.* The analgesic effect of paeoniflorin on neonatal maternal separation-induced visceral hyperalgesia in rats. *J. Pain* **9**, 497–505. https://doi.org/10.1016/j.jpain.2007.12.009 (2008).
23. Li, J., Xiong, X. & Liu, Y. Protective effect of paeoniflorin against optic nerve crush. *J. Huazhong Univ. Sci. Technol. Med. Sci.* **27**, 650–652. https://doi.org/10.1007/s11596-007-0607-y (2007).
24. Pan, K. Y. The analysis of distribution pattern in the Paeoniaceae and its formation. *Acta Phytotaxon. Sin.* **33**, 340–349 (1995).
25. Barber, H. N. Evolution in the Genus Paeonia. *Nature* **148**, 227–228. https://doi.org/10.1038/148227b0 (1941).
26. Hong, D. Y. *Peonies of the World: Taxonomy and Phytogeography*. (Royal Botanic Gardens, 2010).
27. Saunders, A. P. & Stebbins, G. L. Cytogenetic studies in Paeonia I. The compatibility of the species and the appearance of the hybrids. *Genetics* **23**, 65–82. https://doi.org/10.1007/BF02982368 (1938).
28. Tzanoudakis, D. Karyotypes of four wild Paeonia species from Greece. *Nord. J. Bot.* **3**, 307–318. https://doi.org/10.1111/j.1756-1051.1983.tb01943.x (1983).
29. Schwarzacher-Robinson, T. Meiosis, SC-formation, and karyotype structure in diploid *Paeonia tenuifolia* and tetraploid *P. officinalis*. *Plant Syst. Evol.* https://doi.org/10.1007/BF00990128 (1986).
30. Punina, E. O. Karyological study of species of the genus Paeonia (Paeoniaceae) from the Caucasus. *Botanicheskii Zhurnal* **72**, 1504–1514 (1987).
31. Hong, D. Y., Wang, X. Q. & Zhang, D. M. *Paeonia saueri* (Paeoniaceae), a new species from the Balkans. *Taxon* **53**, 83–90. https://doi.org/10.2307/4135491 (2004).
32. Sang, T., Crawford, D. J. & Stuessy, T. F. Documentation of reticulate evolution in peonies (Paeonia) using internal transcribed spacer sequences of nuclear ribosomal DNA: Implications for biogeography and concerted evolution. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 6813–6817. https://doi.org/10.1073/pnas.92.15.6813 (1995).
33. Sang, T., Crawford, D. & Stuessy, T. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of Paeonia (Paeoniaceae). *Am. J. Bot.* **84**, 1120–1136. https://doi.org/10.2307/2446155 (1997).
34. Sang, T. & Zhang, D. Reconstructing hybrid speciation using sequences of low copy nuclear genes: Hybrid origins of five Paeonia species based on Adh gene phylogenies. *Syst. Bot.* **24**, 148–163. https://doi.org/10.2307/2419546 (1999).
35. Ferguson, D. & Sang, T. Speciation through homoploid hybridization between allotetraploids in peonies (Paeonia). *Proc. Natl. Acad. Sci. U. S. A.* **98**, 3915–3919. https://doi.org/10.1073/pnas.061288698 (2001).
36. Pan, J. *Variation and Evolution in Paeonia Sect. Paeonia (Paeoniaceae): Evidence from Morphology, Chromosome and Gene Sequences* Ph. D thesis, Institute of Botany, Chinese Academy of Sciences (2006).
37. Hosoki, T. *et al.* Classification of herbaceous peony [*Paeonia lactiflora*] cultivars by random amplified polymorphic DNA (RAPD) analysis. *J. Jpn. Soc. Hortic.* **65**, 843–849 (1997).
38. Wang, X. H. *Analysis of Genetic Diversity and Construction of the Core Collection for Chinese Herbaceous Peony Cultivars Using SRAP Marker* Master thesis, Shandong Agricultural University (2010).
39. Guo, X. F. & Wang, L. Y. Genetic relationship of partial peony germplasm resources with RAPD markers. *Acta Hortic. Sin.* **34**, 1321–1326. https://doi.org/10.3321/j.issn:0513-353x.2007.05.045 (2007).
40. Hu, W. Q., Lu, H., Liu, W., Yuan, J. X. & Zhang, D. M. Paternity identification and genetic structure analysis on the wild population in *Paeonia lactiflora* Pallas (Paeoniaceae). *Acta Hortic. Sin.* **38**, 503–511. https://doi.org/10.16420/j.issn.0513-353x.2011.03.014 (2011).
41. Li, L., Cheng, F. Y. & Zhang, Q. X. Microsatellite markers for the Chinese herbaceous peony *Paeonia lactiflora* (Paeoniaceae). *Am. J. Bot.* **98**, e16-18. https://doi.org/10.3732/ajb.1000410 (2011).
42. Pan, J., Zhang, D. & Sang, T. Molecular phylogenetic evidence for the origin of a diploid hybrid of Paeonia (Paeoniaceae). *Am. J. Bot.* **94**, 400–408. https://doi.org/10.3732/ajb.94.3.400 (2007).
43. Xia, T. *Molecular Phylogenetic Study of Paeoniaceae: Focus on Paeonia Section Paeonia* Ph. D thesis, Institute of Botany, Chinese Academy of Sciences (2008).
44. Yu, X. N., Ji, L. J. & Wang, Q. Research advances in molecular genetic diversity of Paeonia L. *J. Beijing For. Univ.* **34**, 130–136. https://doi.org/10.13332/j.1000-1522.2012.03.017 (2012).
45. Sang, T., Donoghue, M. J. & Zhang, D. Evolution of alcohol dehydrogenase genes in peonies (Paeonia): Phylogenetic relationships of putative nonhybrid species. *Mol. Biol. Evol.* **14**, 994–1007. https://doi.org/10.1093/oxfordjournals.molbev.a025716 (1997).
46. Zhang, J. M., Wang, J. X., Xia, T. & Zhou, S. L. Application of DNA barcoding technique based on phylogenetic analysis in clarifying the species of Paeonia Section Moutan. *Sci. Sin. Vitae* **38**, 1166–1176 (2008).
47. Sun, Y. L. & Hong, S. K. Phylogenetic relationship and evolution analysis of the peony Paeonia species using multi-locus deoxyribonucleic acid (DNA) barcodes. *J. Med. Plants Res.* **6**, 5048–5058. https://doi.org/10.5897/JMPR11.1733 (2012).
48. Zhao, X., Zhou, Z. Q., Lin, Q. B., Pan, K. Y. & Hong, D. Y. Molecular evidence for the interspecific relationships in Paeonia sect. Moutan: PCR-RFLP and sequence analysis of glycerol-3-phosphate acyltransferase (GPAT) gene. *J. Syst. Evol.* **42**, 236–244. https://doi.org/10.1088/1009-0630/6/5/011 (2004).
49. Zou, Y. P., Cai, M. L. & Wang, Z. P. Systematic studies on Paeonia sect. Moutan DC. based on RAPD analysis. *J. Syst. Evol.* **37**, 220–227. https://doi.org/10.1088/0256-307X/16/12/025 (1999).
50. Wolfe, K. H., Li, W. H. & Sharp, P. M. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. U. S. A.* **84**, 9054–9058. https://doi.org/10.1073/pnas.84.24.9054 (1987).
51. Drouin, G., Daoud, H. & Xia, J. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol. Phylogenet. Evol.* **49**, 827–831. https://doi.org/10.1016/j.ympev.2008.09.009 (2008).
52. Smith, D. R. Mutation rates in plastid genomes: They are lower than you might think. *Genome Biol. Evol.* **7**, 1227–1234. https://doi.org/10.1093/gbe/evv069 (2015).

53. Chumley, T. W. *et al.* The complete chloroplast genome sequence of Pelargonium × hortorum: Organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Mol. Biol. Evol.* **23**, 2175–2190. https://doi.org/10.1093/molbev/msl089 (2006).
54. Wang, L., Dong, W. P. & Zhou, S. L. Structural mutations and reorganizations in chloroplast genomes of flowering plants. *Acta Bot. Boreali-Occiden. Sin.* **32**, 1282–1288. https://doi.org/10.3969/j.issn.1000-4025.2012.06.031 (2012).
55. Jansen, R. K. *et al.* Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 19369–19374. https://doi.org/10.1073/pnas.0709121104 (2007).
56. Chen, X. *et al.* Identification of ligularia herbs using the complete chloroplast genome as a super-barcode. *Front. Pharmacol.* **9**, 695. https://doi.org/10.3389/fphar.2018.00695 (2018).
57. Park, I. *et al.* Authentication of herbal medicines *Dipsacus asper* and *Phlomoides umbrosa* using DNA barcodes, chloroplast genome, and sequence characterized amplified region (SCAR) marker. *Molecules* https://doi.org/10.3390/molecules23071748 (2018).
58. Zhou, Y., Nie, J., Xiao, L., Hu, Z. & Wang, B. Comparative chloroplast genome analysis of Rhubarb botanical origins and the development of specific identification markers. *Molecules* https://doi.org/10.3390/molecules23112811 (2018).
59. Niu, Z. *et al.* Plastome-wide comparison reveals new SNV resources for the authentication of *Dendrobium huoshanense* and its corresponding medicinal slice (Huoshan Fengdou). *Acta Pharm. Sin. B* **8**, 466–477. https://doi.org/10.1016/j.apsb.2017.12.004 (2018).
60. Wu, M., Li, Q., Hu, Z., Li, X. & Chen, S. The complete *Amomum kravanh* chloroplast genome sequence and phylogenetic analysis of the Commelinids. *Molecules* https://doi.org/10.3390/molecules22111875 (2017).
61. Xu, J. *et al.* The first intron of rice EPSP synthase enhances expression of foreign gene. *Sci. Chin. Ser. C Life Sci.* **46**, 561–569. https://doi.org/10.1360/02yc0120 (2003).
62. Kim, K. J. & Lee, H. L. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res.* **11**, 247–261. https://doi.org/10.1093/dnares/11.4.247 (2004).
63. Li, X. *et al.* Complete chloroplast genome sequence of holoparasite *Cistanche deserticola* (Orobanchaceae) reveals gene loss and horizontal gene transfer from its host *Haloxylon ammodendron* (Chenopodiaceae). *PLoS One* **8**, e58747. https://doi.org/10.1371/journal.pone.0058747 (2013).
64. Samigullin, T. H., Logacheva, M. D., Penin, A. A. & Vallejo-Roman, C. M. Complete plastid genome of the recent holoparasite *Lathraea squamaria* reveals earliest stages of plastome reduction in Orobanchaceae. *PLoS One* **11**, e0150718. https://doi.org/10.1371/journal.pone.0150718 (2016).
65. Wang, Z. & Wang, F. *Biological Genome Evolution and Codon Usage*. (National Defense Industry Press, 2014).
66. Carlini, D. B. & Stephan, W. In vivo introduction of unpreferred synonymous codons into the Drosophila Adh gene results in reduced levels of ADH protein. *Genetics* **163**, 239–243 (2003).
67. Sharp, P. M. & Matassi, G. Codon usage and genome evolution. *Curr. Opin. Genet. Dev.* **4**, 851–860. https://doi.org/10.1016/0959-437x(94)90070-1 (1994).
68. Romero, H., Zavala, A. & Musto, H. Codon usage in *Chlamydia trachomatis* is the result of strand-specific mutational biases and a complex pattern of selective forces. *Nucleic Acids Res.* **28**, 2084–2090. https://doi.org/10.1093/nar/28.10.2084 (2000).
69. Stenico, M., Lloyd, A. T. & Sharp, P. M. Codon usage in *Caenorhabditis elegans*: Delineation of translational selection and mutational biases. *Nucleic Acids Res.* **22**, 2437–2446. https://doi.org/10.1093/nar/22.13.2437 (1994).
70. Olejniczak, M. & Uhlenbeck, O. C. tRNA residues that have coevolved with their anticodon to ensure uniform and accurate codon recognition. *Biochimie* **88**, 943–950. https://doi.org/10.1016/j.biochi.2006.06.005 (2006).
71. Romero, H., Zavala, A., Musto, H. & Bernardi, G. The influence of translational selection on codon usage in fishes from the family Cyprinidae. *Gene* **317**, 141–147. https://doi.org/10.1016/s0378-1119(03)00701-7 (2003).
72. Hiraoka, Y., Kawamata, K., Haraguchi, T. & Chikashige, Y. Codon usage bias is correlated with gene expression levels in the fission yeast *Schizosaccharomyces pombe*. *Genes Cells Devot. Mol. Cell. Mech.* **14**, 499–509. https://doi.org/10.1111/j.1365-2443.2009.01284.x (2009).
73. Knight, R. D., Freeland, S. J. & Landweber, L. F. A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. *Genome Biol.* **2**, Research0010. https://doi.org/10.1186/gb-2001-2-4-research0010 (2001).
74. Gu, W., Zhou, T., Ma, J., Sun, X. & Lu, Z. Folding type specific secondary structure propensities of synonymous codons. *IEEE Trans. Nanobiosci.* **2**, 150–157. https://doi.org/10.1109/tnb.2003.817024 (2003).
75. Zhao, Y., Liu, Z., Yang, P., Cheng, Y. & Yang, Y. Codon bias analysis method and research progress on codon bias in *Camellia sinensis*. *J. Tea Commun.* **43**, 3–7. https://doi.org/10.3969/j.issn.1009-525X.2016.02.001 (2016).
76. Shang, M., Liu, F., Hua, J. & Wang, K. Analysis on codon usage of chloroplast genome of *Gossypium hirsutum*. *Sci. Agric. Sin.* **44**, 245–253. https://doi.org/10.3864/j.issn.0578-1752.2011.02.003 (2011).
77. Park, I., Yang, S., Choi, G., Kim, W. J. & Moon, B. C. The complete chloroplast genome sequences of *Aconitum pseudolaeve* and *Aconitum longecassidatum*, and development of molecular markers for distinguishing species in the Aconitum subgenus Lycoctonum. *Molecules* https://doi.org/10.3390/molecules22112012 (2017).
78. Powell, W., Morgante, M., McDevitt, R., Vendramin, G. G. & Rafalski, J. A. Polymorphic simple sequence repeat regions in chloroplast genomes: Applications to the population genetics of pines. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 7759–7763. https://doi.org/10.1073/pnas.92.17.7759 (1995).
79. Yang, A. H., Zhang, J. J., Yao, X. H. & Huang, H. W. Chloroplast microsatellite markers in *Liriodendron tulipifera* (Magnoliaceae) and cross-species amplification in *L. chinense*. *Am. J. Bot.* **98**, e123-126. https://doi.org/10.3732/ajb.1000532 (2011).
80. Jiao, Y. *et al.* Development of simple sequence repeat (SSR) markers from a genome survey of Chinese bayberry (*Myrica rubra*). *BMC Genom.* **13**, 201. https://doi.org/10.1186/1471-2164-13-201 (2012).
81. Xue, J., Wang, S. & Zhou, S. L. Polymorphic chloroplast microsatellite loci in Nelumbo (Nelumbonaceae). *Am. J. Bot.* **99**, e240-244. https://doi.org/10.3732/ajb.1100547 (2012).
82. Qian, J. *et al.* The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoS One* **8**, e57607. https://doi.org/10.1371/journal.pone.0057607 (2013).
83. Kuang, D. Y. *et al.* Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): Implication for DNA barcoding and population genetics. *Genome* **54**, 663–673. https://doi.org/10.1139/g11-026 (2011).
84. Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M. & Dubchak, I. VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* **32**, W273-279. https://doi.org/10.1093/nar/gkh458 (2004).
85. Lu, R. S., Li, P. & Qiu, Y. X. The complete chloroplast genomes of three Cardiocrinum (Liliaceae) species: Comparative genomic and phylogenetic analyses. *Front. Plant Sci.* **7**, 2054. https://doi.org/10.3389/fpls.2016.02054 (2016).
86. Liu, L. X. *et al.* The complete chloroplast genome of Chinese Bayberry (*Morella rubra*, Myricaceae): Implications for understanding the evolution of fagales. *Front. Plant Sci.* **8**, 968. https://doi.org/10.3389/fpls.2017.00968 (2017).
87. Liu, L. *et al.* Chloroplast genome analyses and genomic resource development for epilithic sister genera Oresitrophe and Mukdenia (Saxifragaceae), using genome skimming data. *BMC Genom.* **19**, 235. https://doi.org/10.1186/s12864-018-4633-x (2018).
88. Menezes, A. P. A. *et al.* Chloroplast genomes of Byrsonima species (Malpighiaceae): Comparative analysis and screening of high divergence sequences. *Sci. Rep.* **8**, 2210. https://doi.org/10.1038/s41598-018-20189-4 (2018).

89. Chaney, L., Mangelson, R., Ramaraj, T., Jellen, E. N. & Maughan, P. J. The complete chloroplast genome sequences for four Amaranthus species (Amaranthaceae). *Appl. Plant Sci.* https://doi.org/10.3732/apps.1600063 (2016).
90. Librado, P. & Rozas, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452. https://doi.org/10.1093/bioinformatics/btp187 (2009).
91. Zhang, Y. J., Ma, P. F. & Li, D. Z. High-throughput sequencing of six bamboo chloroplast genomes: Phylogenetic implications for temperate woody bamboos (Poaceae: Bambusoideae). *PLoS One* **6**, e20596. https://doi.org/10.1371/journal.pone.0020596 (2011).
92. Hu, Y., Woeste, K. E. & Zhao, P. Completion of the chloroplast genomes of five Chinese Juglans and their contribution to chloroplast phylogeny. *Front. Plant Sci.* **7**, 1955. https://doi.org/10.3389/fpls.2016.01955 (2016).
93. Wu, L. *et al.* Comparative and phylogenetic analysis of the complete chloroplast genomes of three Paeonia section Moutan species (Paeoniaceae). *Front. Genet.* **11**, 980. https://doi.org/10.3389/fgene.2020.00980 (2020).
94. Zhang, M. Y. *et al.* Complete chloroplast genome of *Paeonia mairei* H. Lév.: Characterization and phylogeny. *Acta Pharm. Sin.* **55**, 168–176. https://doi.org/10.16438/j.0513-4870.2019-0654 (2020).
95. Zhou, S. *et al.* Out of the Pan-Himalaya: Evolutionary history of the Paeoniaceae revealed by phylogenomics. *J. Syst. Evol.* https://doi.org/10.1111/jse.12688 (2020).
96. Xu, S. *et al.* Evaluation of the DNA barcodes in Dendrobium (Orchidaceae) from mainland Asia. *PLoS One* **10**, e0115168. https://doi.org/10.1371/journal.pone.0115168 (2015).
97. Zhu, S. *et al.* Accurate authentication of *Dendrobium officinale* and its closely related species by comparative analysis of complete plastomes. *Acta Pharm. Sin. B* **8**, 969–980. https://doi.org/10.1016/j.apsb.2018.05.009 (2018).
98. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170 (2014).
99. Luo, R. *et al.* SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18. https://doi.org/10.1186/2047-217x-1-18 (2012).
100. Dierckxsens, N., Mardulyn, P. & Smits, G. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* **45**, e18. https://doi.org/10.1093/nar/gkw955 (2017).
101. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578–579. https://doi.org/10.1093/bioinformatics/btq683 (2011).
102. Liu, C. *et al.* CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genom.* **13**, 715. https://doi.org/10.1186/1471-2164-13-715 (2012).
103. Tillich, M. *et al.* GeSeq-versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **45**, W6-w11. https://doi.org/10.1093/nar/gkx391 (2017).
104. Schattner, P., Brooks, A. N. & Lowe, T. M. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* **33**, W686-689. https://doi.org/10.1093/nar/gki366 (2005).
105. Zheng, S., Poczai, P., Hyvönen, J., Tang, J. & Amiryousefi, A. Chloroplot: An online program for the versatile plotting of organelle genomes. *Front. Genet.* **11**, 576124. https://doi.org/10.3389/fgene.2020.576124 (2020).
106. Mistry, J. *et al.* Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**, D412-d419. https://doi.org/10.1093/nar/gkaa913 (2021).
107. Sharp, P. M. & Li, W. H. The codon Adaptation Index—A measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **15**, 1281–1295. https://doi.org/10.1093/nar/15.3.1281 (1987).
108. Beier, S., Thiel, T., Münch, T., Scholz, U. & Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **33**, 2583–2585. https://doi.org/10.1093/bioinformatics/btx198 (2017).
109. Kurtz, S. *et al.* 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 (2001).
110. Darling, A. C., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **14**, 1394–1403. https://doi.org/10.1101/gr.2289704 (2004).
111. Katoh, K. & Standley, D. M. 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 (2013).
112. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. 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 (2015).
113. Abdullah, *et al.* Chloroplast genome evolution in the *Dracunculus clade* (Aroideae, Araceae). *Genomics* **113**, 183–192. https://doi.org/10.1016/j.ygeno.2020.12.016 (2021).

Acknowledgements

This research was funded by the Major Scientific and Technological Special Project for "Significant New Drugs Creation" (No. 2018ZX09711001-008-007) and Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences (CIFMS) (No. 2016-I2M-3-016).

Author contributions

Conceptualization, H.Y.; data curation, L.W.; formal analysis, L.W., L.N. and Q.W.; funding acquisition, H.Y.; investigation, L.W., L.N. and Q.W.; methodology, L.W., L.N. and Q.W.; project administration, H.Y.; resources, L.N., Y.W. and C.H.; software, Z.X. and J.S.; supervision, H.Y.; validation, H.Y.; visualization, L.W.; writing—original draft, L.W.; writing—review & editing, H.Y. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-94137-0.

Correspondence and requests for materials should be addressed to H.Y.

Reprints and permissions information is available at www.nature.com/reprints.

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