Research Article
Genomics and Bioinformatics

The complete chloroplast genome of *Papaver setigerum* and comparative analyses in Papaveraceae

Luxian Liu[1*], Yingxue Du[1*], Cheng Shen[2], Rui Li[3], Joongku Lee[4] and Pan Li[2] iD

[1]*Henan University, School of Life Sciences, Key Laboratory of Plant Stress Biology, Kaifeng, China.*
[2]*Zhejiang University, College of Life Sciences, The Key Laboratory of Conservation Biology for Endangered Wildlife of the Ministry of Education, Hangzhou, China.*
[3]*Food inspection and Testing Institute of Henan Province, Physical and Chemical Laboratory, Zhengzhou, China.*
[4]*Chungnam National University, Department of Environment and Forest Resources, Daejeon, South Korea.*

Abstract

Papaver setigerum is an annual herb that is closely related to the opium poppy, *P. somniferum*. Genetic resources for *P. setigerum* are scarce. In the present study, we assembled the complete chloroplast (cp) genome of *P. setigerum* based on genome skimming data, and we conducted comparative cp genome analyses to study the evolutionary pattern in Papaveraceae. The cp genome of *P. setigerum* is 152,862 bp in length with a typical quadripartite structure. Comparative analyses revealed no gene rearrangement in the Papaveraceae family, although differences were evident in genome size, gene losses, as well as inverted repeats (IR) region expansion and contraction. The *rps*15 gene has been lost from the genomes of *Meconopsis racemosa*, *Coreanomecon hylomeconoides*, *P. orientale*, *P. somniferum*, and *P. setigerum*, and the *ycf*15 gene is found only in *C. hylomeconoides*. Moreover, 13 cpDNA markers, including *psb*A-*trn*H, *rps*16-*trn*Q, *trn*S-*trn*G, *trn*C-*pet*N, *trn*E-*trn*T, *trn*L-*trn*F, *trn*F-*ndh*J, *pet*A-*psb*J, *ndh*F-*rpl*32, *rpl*32-*trn*L, *ccs*A-*ndh*D, *ndh*E-*ndh*G, and *rps*15-*ycf*1, were identified with relatively high levels of variation within *Papaver*, which will be useful for species identification in this genus. Among those markers, *psb*A-*trn*H is the best one to distinguish *P. somniferum* and *P. setigerum*.

Keywords: *Papaver setigerum*, chloroplast genomes, chloroplast hotspot, species identification.

Introduction

Papaver setigerum DC., an annual herb of the poppy family (Kalis, 1979), occurs in the Mediterranean region, especially in southwestern Europe (Portugal, Spain, France, Italy, Greece) and North Africa (Pignatti, 1982). This plant is closely related to and sometimes treated as a variety or subspecies of opium poppy (*P. somniferum* L.) due to its similarity in flower-shape, color, fruit, and production of small amounts of morphine alkaloids (La Valva *et al.*, 1985; Osalou *et al.*, 2013). Of the 110 species of the genus *Papaver*, only *P. somniferum* and *P. setigerum* are controlled species in most countries (Choe *et al.*, 2012). However, the cytological evidence shows that *P. somniferum* is diploid (2n = 22), while *P. setigerum* is tetraploid (2n = 44) (Fulton, 1944; Choe *et al.*, 2012), indicating that *P. setigerum* is not likely the wild ancestral species of the cultivated *P. somniferum*

(Farmilo *et al.*, 1953). For *Papaver* species, inter-specific identification based only on morphological characteristics is difficult because of the similarities in appearance mentioned above (Osalou *et al.*, 2013). Phytochemical methods (Zhang and Cheng, 2009; Osalou *et al.*, 2013) and various molecular markers (Fan *et al.*, 1987; Hosokawa *et al.*, 2004; Choe *et al.*, 2012; Zhang *et al.*, 2015) have been used to identify *Papaver* species in previous studies. However, current studies involving *P. setigerum* have mostly focused on its chemical composition, largely ignoring its genetic background.

Chloroplasts (cp), the photosynthetic organelles of most green plants, are known to be derived from cyanobacteria through endosymbiosis and co-evolution (Dagan *et al.*, 2012; Asaf *et al.*, 2017). In most angiosperms, cp genomes have a typically circular and quadripartite structure. The genome size is usually from 115 to 165 kb in length, consisting of two regions of inverted repeats (IRs), separated by a large single-copy (LSC) region and a small single-copy (SSC) region (Wicke *et al.*, 2011). Compared with nuclear and mitochondrial genomes, the cp genome is more conserved, not only in gene content and organization, but also in genome structure (Raubeson and Jansen, 2005). Due to its

Send correspondence to Pan Li. Zhejiang University, College of Life Sciences, The Key Laboratory of Conservation Biology for Endangered Wildlife of the Ministry of Education, Hangzhou 310058, Zhejiang, China. E-mail: panli_zju@126.com.
*These authors contributed equally to this work

relatively conserved gene content and simple structure, small size, uniparental inheritance, and the fact that it is non-recombinant, the cp genome has been used as an ideal model for phylogenetic reconstruction (Liu *et al.*, 2017), evolutionary and comparative genomic studies (Liu *et al.*, 2018b), species identification (Thomson *et al.*, 2010; Greiner *et al.*, 2015) and markers development (Liu *et al.*, 2018a). Currently, the rapid development and improvement of next-generation sequencing technology have made the assembly of the cp genome cheaper and more efficient compared with traditional sequencing (Alkan *et al.*, 2011). In addition, the releases of many assembly processes or pipelines, such as SOAPdenovo2 (Luo *et al.*, 2012), CLC Genomics Workbench (CLC Inc., Rarhus, Denmark) and GetOrganelle (Jin *et al.*, 2018), have made cp genome reconstruction easier and much more effective.

In the present study, one *P. setigerum* individual was selected for genome skimming, and the complete chloroplast genome sequence was assembled and reported. We also compared the cp genomes among representatives of Papaveraceae and detected highly divergent regions of the cp genomes within the genus *Papaver*.

Material and Methods

Plant material, DNA extraction, and sequencing

We extracted whole-genomic DNA from silica-dried leaf tissue of one cultivated *P. setigerum* plant collected in Taizhou (Zhejiang, China), using modified CTAB reagent Plant DNAzol (Invitrogen, Shanghai, China) according to the manufacturer's protocol. High quality DNA was sheared to yield fragments with length less than or equal to 800 bp. The quality of fragmentation was checked on an Agilent Bioanalyzer 2100 (Agilent Technologies). The 500 bp short-insert length paired-end library was prepared and sequenced by Beijing Genomics Institute (BGI, Wuhan, China). The library was run in one lane of an Illumina HiSeq X10 and obtained reads with length of 150 bp.

Chloroplast genome assembly and annotation

The raw reads were first screened for Phred score < 30 to remove low-quality sequences. In order to ensure the accuracy of cp genome assembly, we employed two different methods to assemble the cp genome. In the first method, all the remaining reads were assembled into contigs implemented in the CLC genome workbench (CLC Inc., Rarhus, Denmark). The parameters set in CLC were as follows: 200 bp for minimum contig length, 3 for deletion and insertion costs, bubble size of 98, 0.9 for length fraction and similarity fraction, and 2 for mismatch cost. Then, the principal contigs representing the cp genome were separated from the total contigs using a BLAST (NCBI BLAST V2.2.31) search, with the cp genome of *P. somniferum* set as the reference. The representative cp contigs were oriented and ordered on the basis of the reference cp genome, and the complete chloroplast genome of *P. setigerum* was reconstructed by

connecting overlapping terminal sequences. In the second method, the cp genome of *P. setigerum* was de novo assembled using the GetOrganelle pipeline (Jin *et al.*, 2018), with SPAdes 3.10.1 as assembler (Bankevich *et al.*, 2012).

Geneious R11 (https://www.geneious.com) was used to annotate the cp genome of *P. setigerum*, and putative starts, stops, and intron positions were identified on the basis of comparisons with homologous genes of the *P. somniferum* cp genome. The tRNA genes were verified with tRNAscan-SE v1.21 (Schattner *et al.*, 2005) with the default setting. We drew the circular chloroplast genome map of *P. setigerum* using the OrganellarGenomeDRAW program (OGDRAW, Lohse *et al.*, 2013).

Comparative chloroplast genomic analyses

In order to study the sequence variation within Papaveraceae, we downloaded multiple publicly available cp genomes of the family from GenBank to compare the overall similarities, using *Leontice incerta* (Berberidaceae, MH940295) as the reference, according to the results of Kim and Kim (2016). The GenBank accession numbers for the Papaveraceae species are as follows (Table S1): *P. orientale* (NC_037832), *P. rhoeas* (NC_037831), *P. somniferum* (NC_029434), *Meconopsis racemosa* (NC_039625), *Coreanomecon hylomeconoides* (NC_031446), and *Macleaya microcarpa* (NC_039623). The sequence identities of the seven Papaveraceae cp genomes were implemented in the mVISTA program with LAGAN mode (Frazer *et al.*, 2004). The cp DNA rearrangement analyses of seven Papaveraceae cp genomes were based on Mauve Alignment (Darling *et al.*, 2004).

Molecular markers development for *Papaver*

In order to screen variable characters within *Papaver*, multiple alignments of the four *Papaver* species cp genomes were carried out using MAFFT version 7.017 (Katoh and Standley, 2013). The nucleotide diversity (*Pi*) was determined by calculating the total number of mutations (*Eta*) and average number of nucleotide differences (*K*) using DnaSP v5.0 (Librado and Rozas, 2009).

Phylogenetic inferences

The phylogenetic relationships of Papaveraceae were inferred using the whole chloroplast genome sequences of seven species; two species from Ranunculaceae (*Ranunculus macranthus*) and Berberidaceae (*Leontice incerta*) were chosen as the outgroups, according to the results of Kim and Kim (2016). The phylogeny inferences were conducted using Bayesian inference (BI) and maximum likelihood (ML) methods. ML analysis was performed with RAxML-HPC v8.1.11 on the CIPRES cluster (Miller *et al.*, 2010) with GTR + I + G set as the best-fit nucleotide substitution model. BI analysis was implemented in MrBayes v3.2.3 using the same substitution model mentioned above (Ronquist and Huelsenbeck, 2003).

Results

P. setigerum cp genome assembly, organization and gene content

The complete cp genomes of *P. setigerum* assembled from two different assembly strategies were identical. However, using GetOrganelle to assemble the cp genome of *P. setigerum* was much faster and more effective than using CLC genome workbench (< 1h vs. > 6h). The cp genome size was 152,862 bp, and had a typical quadripartite structure that was similar to the majority of land plant cp genomes, consisting of an 83,022 bp large single copy region (LSC), a 17,944 bp small single copy region (SSC) and two 25,948 bp inverted repeats. The *P. setigerum* cp genome contains 113 unique genes, including 79 protein-coding genes, 30 tRNA genes and four ribosomal RNA genes (Figure 1 and Table 1). Eight protein-coding, seven tRNA, and four rRNA genes are duplicated and located in the IR regions. Among the 113 genes, nine protein-coding genes and six tRNA genes contain one intron; three protein-coding genes (*clp*P, *ycf*3 and *rps*12) contain two introns. We submitted the cp genome of *P. setigerum* to GenBank with the accession number MK820043.

Genome comparison of Papaveraceae

The chloroplast genomes of the seven Papaveraceae species were relatively conservative, and the IR region is more conserved compared to the LSC and SSC regions (Figure 2). No rearrangements, such as translocations or inversions, occurred in gene organization after verification in this family (Figure 3). However, differences existed in genome size, gene losses, and IR expansion and contraction.

In terms of the cp genome size observed among the representative Papaveraceae species, the four *Papaver* species were the smallest and had similar genome sizes ranging from 152,799 bp to 152,931 bp (Figure 4). Of the other species, *Macleaya microcarpa* (161,124 bp) exhibited the largest cp genome, while *Meconopsis racemosa* (153,763 bp) had the smallest one.

The *rps*15 gene has been lost from the genomes of *M. racemosa*, *C. hylomeconoides*, *P. orientale*, *P. somniferum*, and *P. setigerum*, although it is present in *P. rhoeas* and the reference genome. In addition, the *ycf*15 gene occurred only
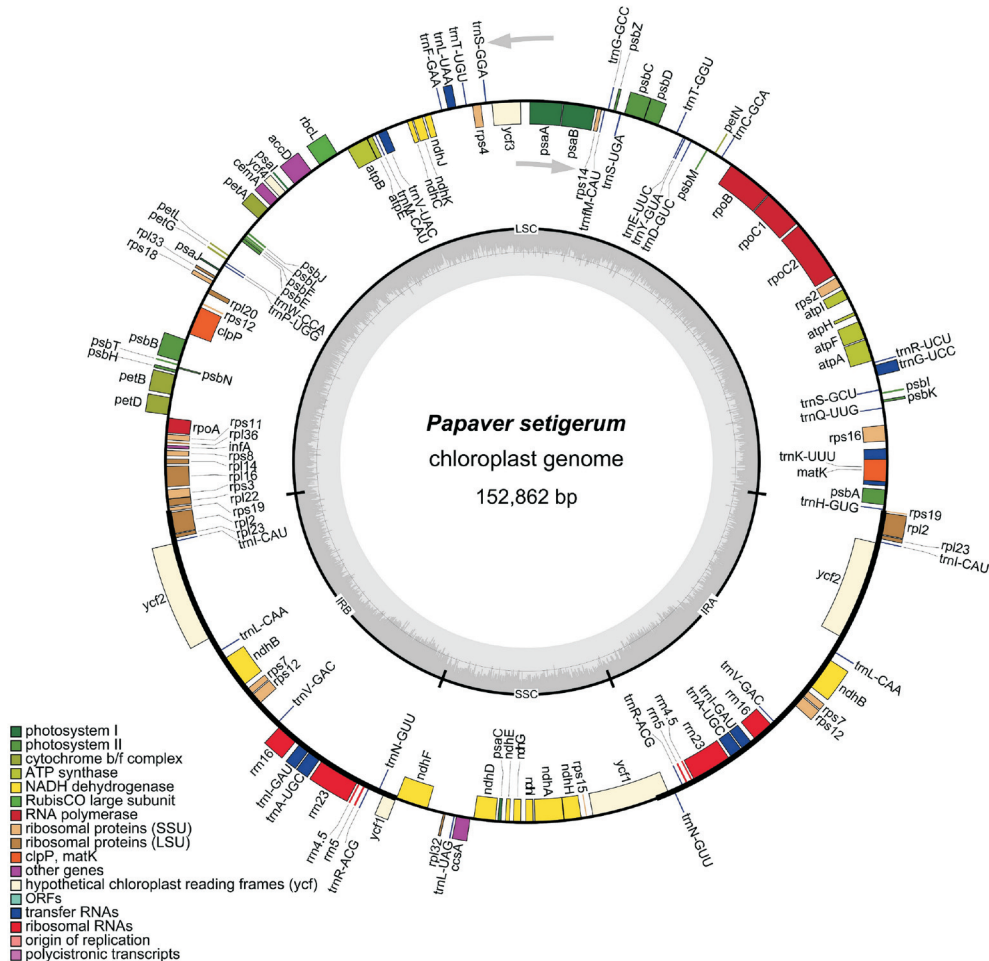

Figure 1 - Chloroplast genome map of *Papaver setigerum*. Genes inside the circle are transcribed clockwise, genes outside are transcribed counter-clockwise. The light gray inner circle corresponds to the AT content, the dark gray to the GC content. Genes belonging to different functional groups are shown in different colors.

Table 1 - Genes contained in *P. setigerum* chloroplast genome (113 genes in total).

Category	Group of gene	Name of gene			
Self-replication	Ribosomal RNA genes	*rrn*4.5[a]	*rrn*5[a]	*rrn*16[a]	*rrn*23[a]
	Transfer RNA gene	*trnA*-UGC[a]*	*trnC*-GCA	*trnD*-GUC	*trnE*-UUC
		trnF-GAA	*trnfM*-CAU	*trnG*-GCC*	*trnG*-UCC
		trnH-GUG	*trnI*-CAU[a]	*trnI*-GAU[a]*	*trnK*-UUU*
		trnL-CAA[a]	*trnL*-UAA*	*trnL*-UAG	*trnM*-CAU
		trnN-GUU[a]	*trnP*-UGG	*trnQ*-UUG	*trnR*-ACG[a]
		trnR-UCU	*trnS*-GCU	*trnS*-GGA	*trnS*-UGA
		trnT-GGU	*trnT*-UGU	*trnV*-GAC[a]	*trnV*-UAC*
		trnW-CCA	*trnY*-GUA		
	Small subunit of ribosome	*rps*2	*rps*3	*rps*4	*rps*7[a]
		*rps*8	*rps*11	*rps*12[a,b]**	*rps*14
		*rps*15	*rps*16*	*rps*18	*rps*19[a]
	Large subunit of ribosome	*rpl*2[a]*	*rpl*14	*rpl*16*	*rpl*20
		*rpl*22	*rpl*23[a]	*rpl*32	*rpl*33
		*rpl*36			
	RNA polymerase subunits	*rpo*A	*rpo*B	*rpo*C1*	*rpo*C2
Photosynthesis	Subunits of photosystem I	*psa*A	*psa*B	*psa*C	*psa*I
		*psa*J	*ycf*3**		
	Subunits of photosystem II	*psb*A	*psb*B	*psb*C	*psb*D
		*psb*E	*psb*F	*psb*H	*psb*I
		*psb*J	*psb*K	*psb*L	*psb*M
		*psb*N	*psb*T	*psb*Z	
	Subunits of cytochrome	*pet*A	*pet*B*	*pet*D*	*pet*G
		*pet*L	*pet*N		
	Subunits of ATP synthase	*atp*A	*atp*B	*atp*E	*atp*F*
		*atp*H	*atp*I		
	Large subunit of Rubisco	*rbc*L			
	Subunits of NADH	*ndh*A*	*ndh*B[a]*	*ndh*C	*ndh*D
	Dehydrogenase	*ndh*E	*ndh*F	*ndh*G	*ndh*H
		*ndh*I	*ndh*J	*ndh*K	
Other genes	Translational initiation factor	*inf*A			
	Maturase	*mat*K			
	Envelope membrane protein	*cem*A			
	Subunit of acetyl-CoA	*acc*D			
	C-type cytochrome synthesis gene	*ccs*A			
	Protease	*clp*P**			
Unknown function	Conserved open reading frames	*ycf*1[a](part)	*ycf*2[a]	*ycf*4	

[a] Two gene copies in IRs; [b] gene divided into two independent transcription units; one and two asterisks indicate one- and two-intron containing genes, respectively.

in *C. hylomeconoides* compared to the other analyzed cp genomes.

In addition, we compared the exact IR border positions and their adjacent genes between the seven Papaveraceae cp genomes and the reference genome (Figure 4). The results showed that the *ycf*1 gene spanned the SSC/IR$_A$ region and the pseudogene fragment of $^{\Psi}$*ycf*1 varied from 912 bp to 1379 bp. The *ndh*F gene shares some nucleotides (25 bp) with the *ycf*1 pseudogene in *Meconopsis racemosa* but is separated from $^{\Psi}$*ycf*1 by spacers in the other analyzed species. The *trn*H-GUG gene was located in the LSC region of all genomes, but varied from 5 bp to 117 bp apart from the IR$_A$/LSC junctions. In addition, the *rps*19 pseudogene appeared in all the representative Papaveraceae species due to the *rps*19 gene extending to the IR region.
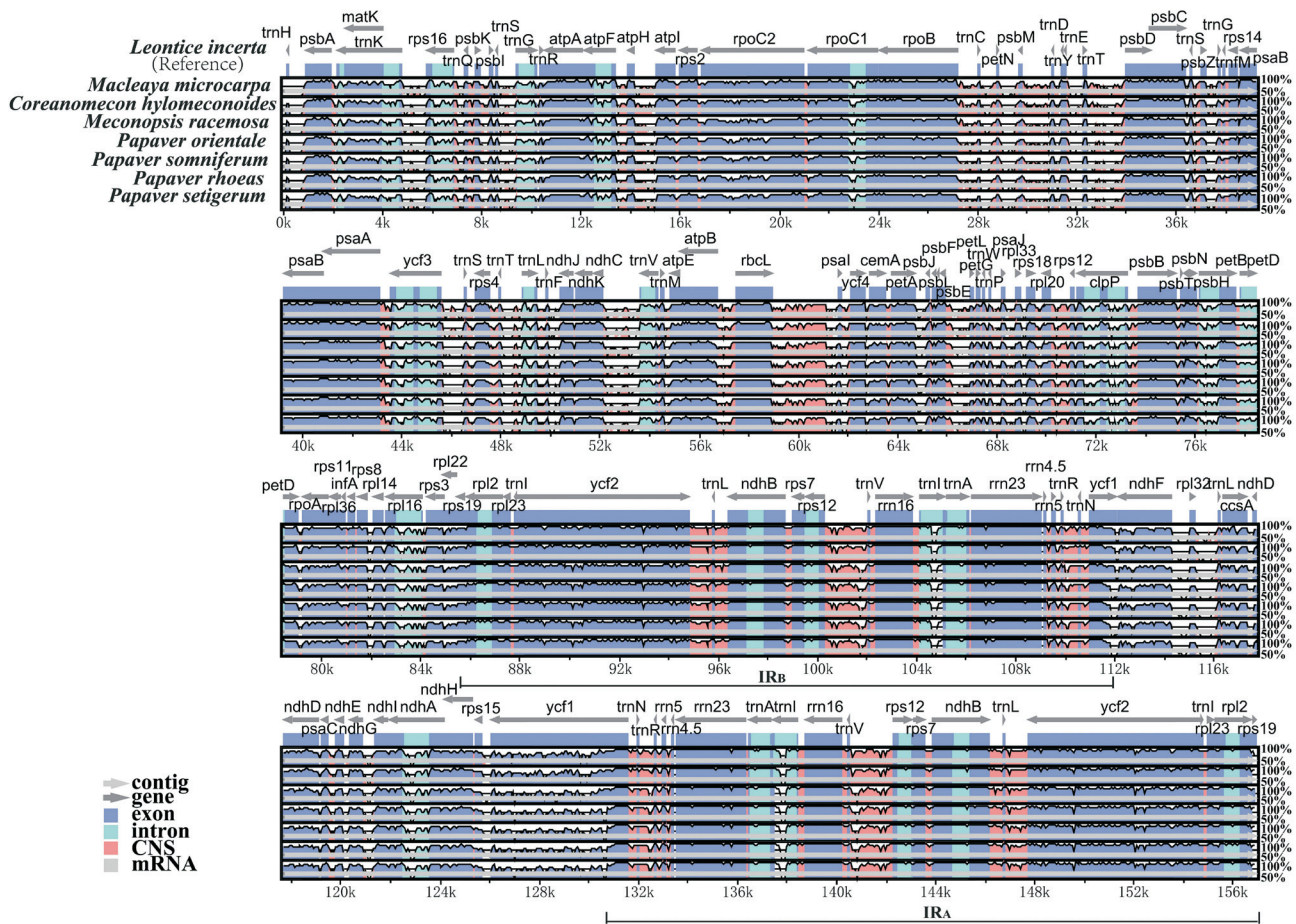
Figure 2 - Visualization of alignment of the seven Papaveraceae chloroplast genome sequences, with *Leontice incerta* as the reference. The horizontal axis indicates the coordinates within the chloroplast genome. The vertical scale indicates the percentage of identity, ranging from 50 to 100%. Genome regions are color coded as protein coding, intron, mRNA, and conserved non-coding sequences (CNS).
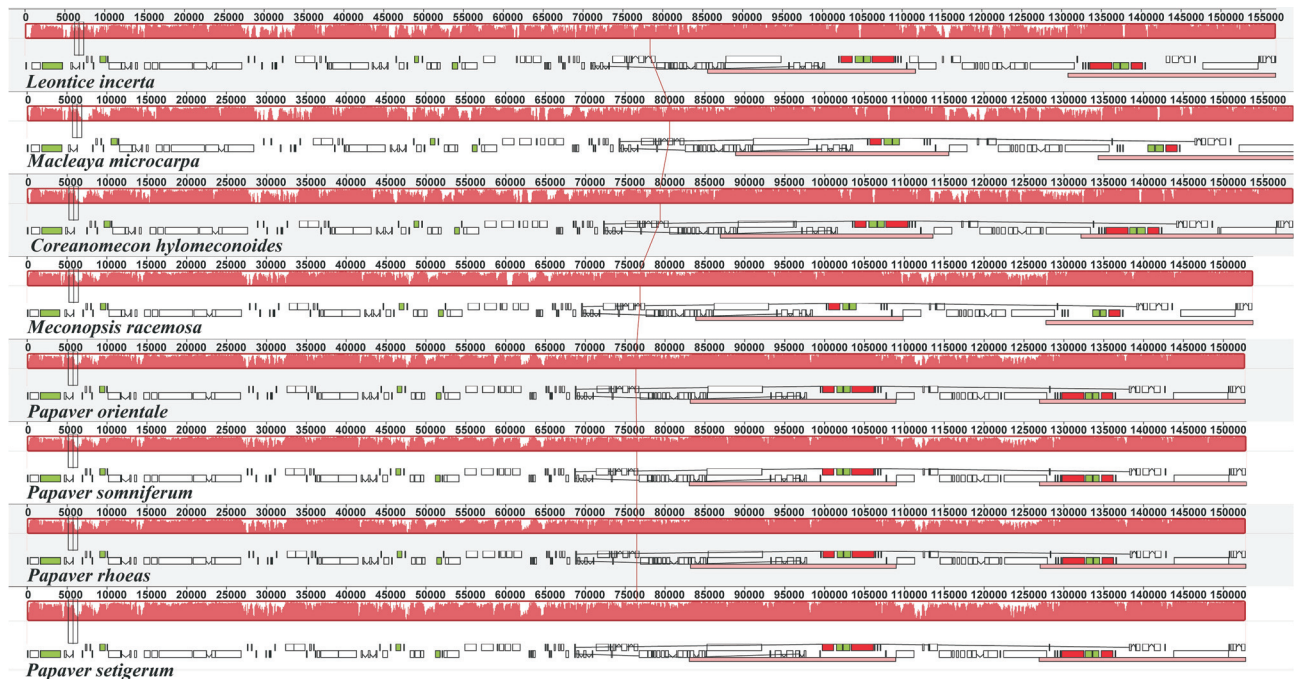

Figure 3 - MAUVE alignment of seven Papaveraceae chloroplast genomes. *Leontice incerta* is shown at the top as the reference. Within each of the alignments, local collinear blocks are represented by blocks of the same color connected by lines.
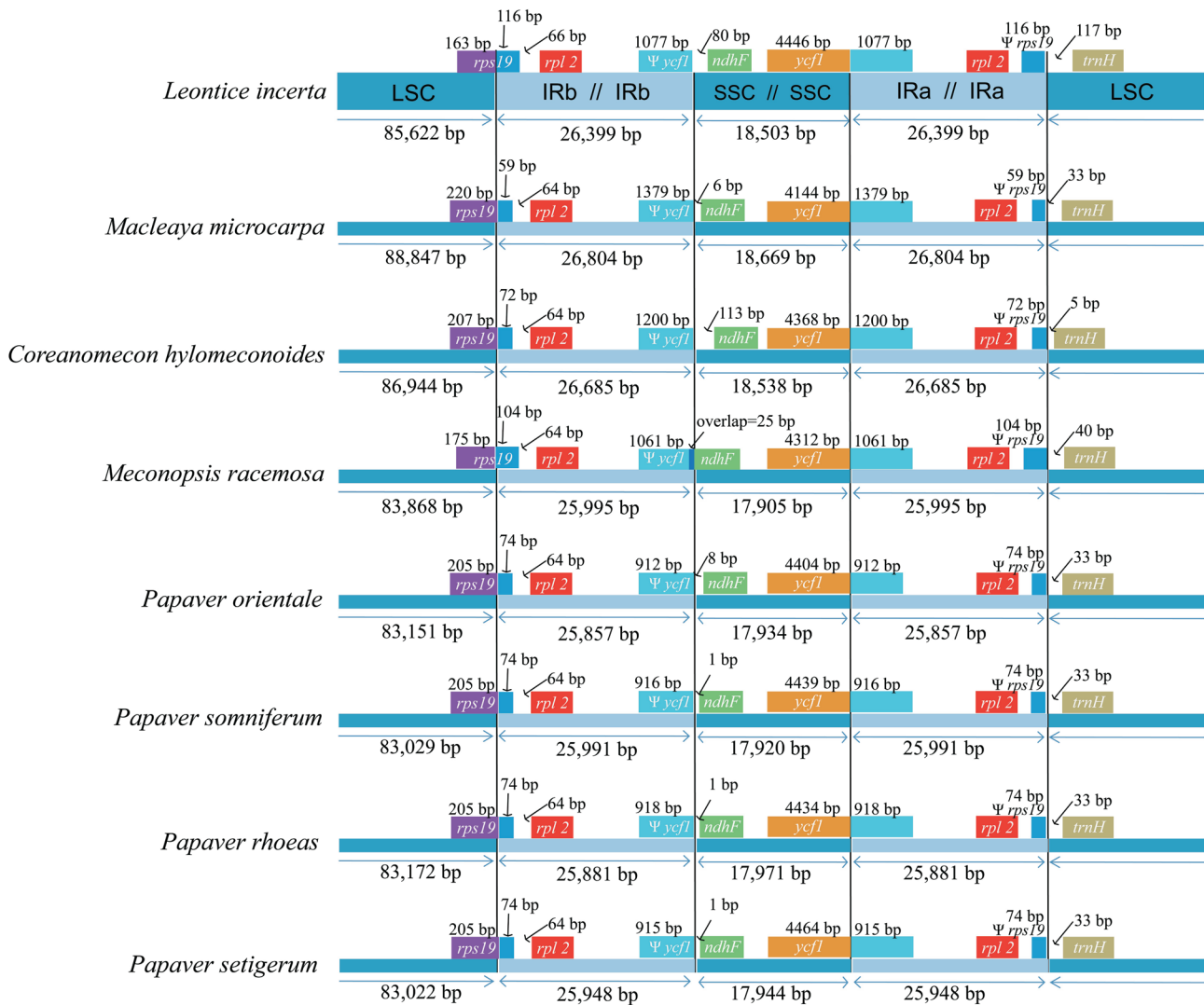
Figure 4 - Comparison of the borders of large single-copy (LSC), small single-copy (SSC), and inverted repeat (IR) regions among the seven Papaveraceae chloroplast genomes, with the *Leontice incerta* cp genome shown at the top as the reference.

Molecular markers development for *Papaver*

In order to explore the divergence hotspot regions in *Papaver*, we divided the genome alignment of the four *Papaver* species into non-coding regions, coding genes, and intron regions. We eventually identified 125 loci (53 coding genes, 55 inter-genic spacers, and 17 intron regions) within *Papaver* having more than 200 bp in length (Figure 5). Of these 125 regions, nucleotide variability (Pi) values ranged from 0.0003 (*rrn*16) to 0.0474 (*psb*A-*trn*H). Thirteen of these variable loci ($Pi > 0.02$), including *psb*A-*trn*H, *rps*16-*trn*Q, *trn*S-*trn*G, *trn*C-*pet*N, *trn*E-*trn*T, *trn*L-*trn*F, *trn*F-*ndh*J, *pet*A-*psb*J, *ndh*F-*rpl*32, *rpl*32-*trn*L, *ccs*A-*ndh*D, *ndh*E-*ndh*G, and *rps*15-*ycf*1, showed high levels of intra-generic variation. Within the 13 regions, seven varied between *P. setigerum* and *P. somniferum* (Table S2), and these can be candidate markers to identify these two species. Among those markers, *psb*A-*trn*H is the best one to distin-

guish *P. somniferum* and *P. setigerum*, which are different by seven site mutations.

Phylogenetic inferences

The tree topologies from both ML and Bayesian analyses were consistent with each other (Figure 6). All but one node within Papaveraceae have full surport (maximum likelihood bootstrap, MLBS = 100%, Bayesian inference posterior probabilities, BIPP = 1). The four *Papaver* species formed one clade with full support and is sister to *Meconopsis racemosa*. The remaining two species, *Macleaya microcarpa* and *Coreanomecon hylomeconoides*, formed another clade.

Discussion

In the last decades, the rapid development of high throughput sequencing technologies have greatly reduced sequencing cost. Considering the large number of copies of
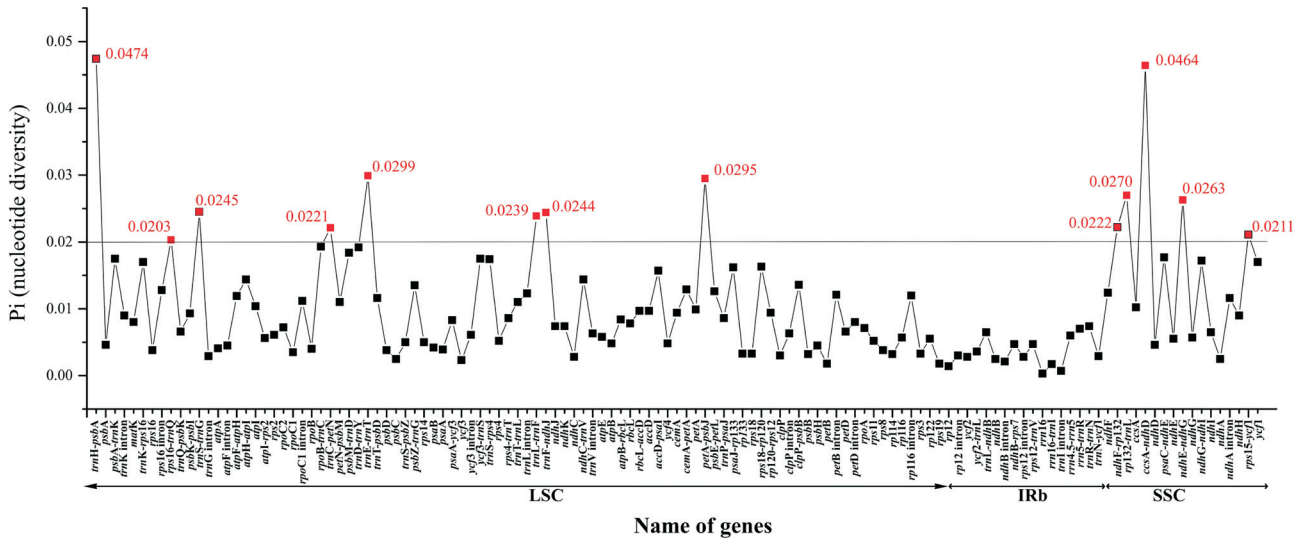
Figure 5 - Comparative analysis of the nucleotide variability (Pi) values among the four *Papaver* species.
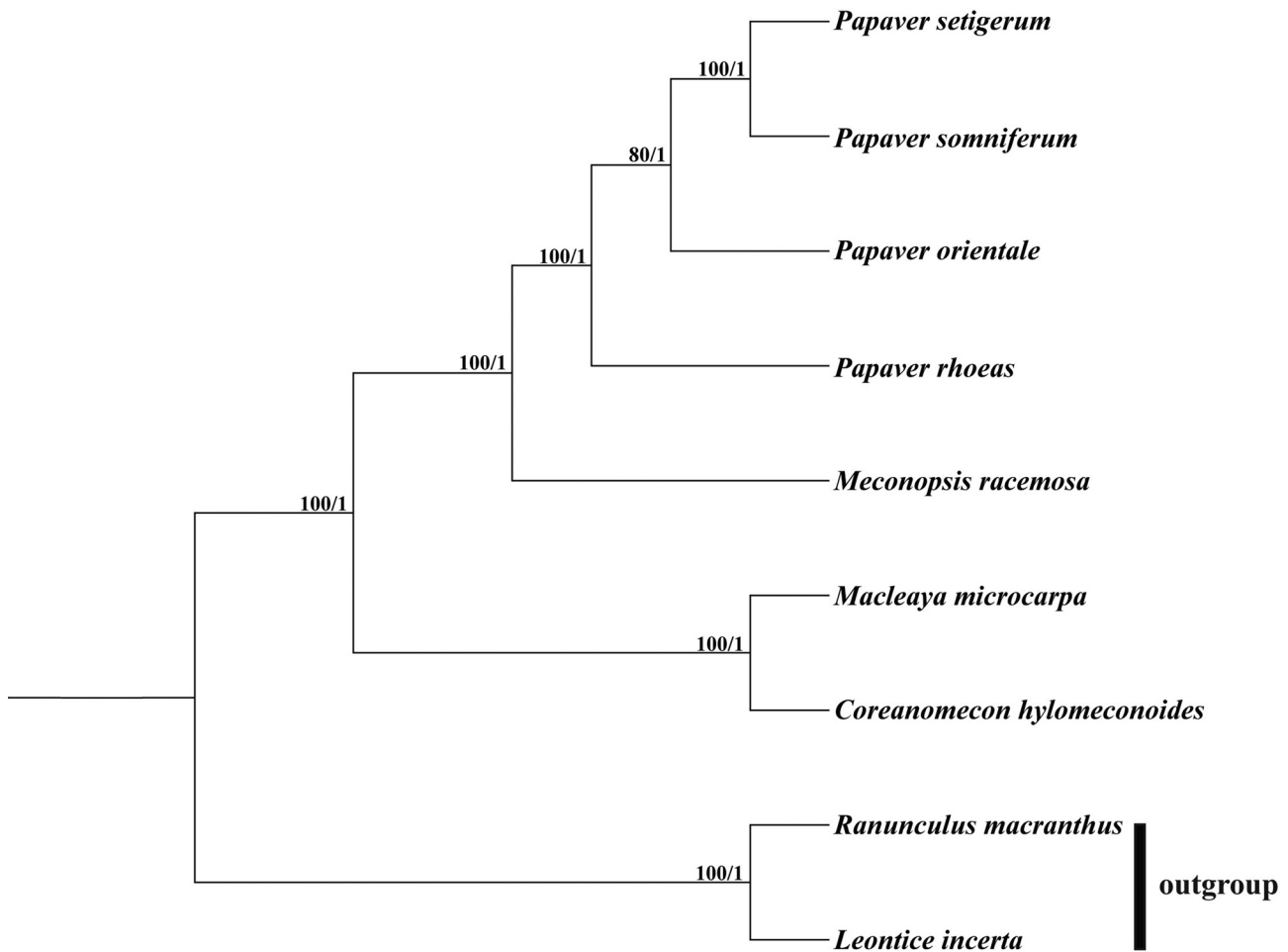

Figure 6 - Phylogenetic tree reconstruction of Papaveraceae using maximum likelihood (ML) based on whole chloroplast genome sequences. Numbers above the branches represent bootstrap values from maximum likelihood analyses and posterior probabilities from Bayesian inference, respectively.

the plastid genome in a single cell, it is easy to get enough reads to reconstruct a complete cp genome from low-coverage, whole-genome sequencing data (Twyford and Ness, 2017), viz. genome skimming data (Straub *et al.*, 2012).

With the publication of many cp genome assembly pipelines (Luo *et al.*, 2012; Jin *et al.*, 2018), cp genome reconstruction by these protocols is more effective than the Sanger method. Since the first complete nucleotide sequence of the cp ge-

nome was generated (*Nicotiana tabacum*; Shinozaki *et al.*, 1986), more than 3000 cp genomes have been submitted to GenBank (Jin *et al.*, 2018). In this study, we tried to assemble the cp genome sequence of *Papaver setigerum* using two different pipelines, the CLC Genomics Workbench (CLC Inc., Rarhus, Denmark) and GetOrganelle (Jin *et al.*, 2018). Cp genome sequences produced by the two pipelines were completely identical in terms of both genome size and base information. However, the GetOrganelle pipeline is faster and more effective in assembling a circular cp genomes of *P. setigerum* or other species (we are preparing to publish a comparative study separately), especially for the low coverage data of the whole genome.

In recent years, comparative studies of cp genomes have been applied to a number of focal species (Young *et al.*, 2011), genera (Greiner *et al.*, 2008; Liu *et al.*, 2018b), or plant families (Daniell *et al.*, 2006; Liu *et al.*, 2017). Comparative analyses of cp genomes are useful for phylogenic inference at higher taxonomic levels (Moore *et al.*, 2010; Li *et al.*, 2017), as well as for understanding the evolution of genome size variations, gene and intron losses, and nucleotide substitutions. In the present study, multiple complete cp genomes of representative Papaveraceae species provide an opportunity to compare the sequence variation within the family. No rearrangement, such as translocations and inversions, occurred in gene organization in this family. However, we identified differences in genome size, gene losses, and IR expansion and contraction. The *rps*15 gene is found in most cp genomes in land plants (Tsuji *et al.*, 2007). However, comparative analysis revealed that this gene was found in *P. rhoeas* and the reference genome *Leonitce incerta*, but was not present in other Papaveraceae species. Previous studies have certified that the *rps*15 loss has also appeared in other families (Tsuji *et al.*, 2007; Krause, 2012). Similarly, the function of the *ycf*15 gene has attracted the attention of previous workers (Raubeson *et al.*, 2007; Shi *et al.*, 2013), and it has apparently been lost in other taxa (Liu *et al.*, 2017; Liu *et al.*, 2018). The *ycf*15 gene, which displays a small open reading frame (ORF), is located immediately downstream of the *ycf*2 gene (Dong *et al.*, 2013). In our study, the *ycf*15 gene occurred only in *Coreanomecon hylomeconoides*, located immediately downstream of the *ycf*2 gene but absent in other analyzed cp genomes. These findings suggest that parallel losses of particular genes have occurred during the evolution of land plant cp genomes.

In the genus *Papaver*, almost all of the species are similar in their flower-shapes (two sepals that fall off as the bud opens and four to six petals), colors, and fruits, complicating species identification based on morphological characteristics alone (Osalou *et al.*, 2013; Zhou *et al.*, 2018). Previous studies have identified *Papaver* species using physicochemical methods, including discrete stationary wavelet transform (Zhang *et al.*, 2009), amplified fragment length polymorphism (Lu *et al.*, 2008), as well as phytochemical methods (Osalou *et al.*, 2013). Hosokawa *et al.* (2004) identified *Papaver* species using the plastid gene *rpl*16 and

*rpl*16-*rpl*14 spacer sequences. Zhang *et al.* (2015) had verified that *trn*L-*trn*F can be considered a novel DNA barcode to identify the *Papaver* genus, and ITS, *mat*K, *psb*A-*trn*H, and *rbc*L can be used as combined barcodes for identification. Zhou *et al.* (2018) screened five hypervariable regions, including *rpo*B-*trn*C, *trn*D-*trn*T, *pet*A-*psb*J, *psb*E-*pet*L, and *ccs*A-*ndh*D, as specific DNA barcodes. In this study, except for the regions mentioned above, we developed nine additional regions (*rps*16-*trn*Q, *trn*S-*trn*G, *trn*C-*pet*N, *trn*E-*trn*T, *trn*F-*ndh*J, *ndh*F-*rpl*32, *rpl*32-*trn*L, *ndh*E-*ndh*G and *rps*15-*ycf*1) with relatively high levels of intrageneric variation, which can be used for identify *Papaver* species in the future. Moreover, *P. setigerum* was formerly treated as a variety or subspecies of *P. somniferum* due to the similar morphological appearance and chemical signature (La Valva *et al.*, 1985; Osalou *et al.*, 2013). However, the cytological evidence rejects this perspective (Fulton, 1944). Besides, there are seven cp regions varied between *P. setigerum* and *P. somniferum* (Table S2). In the phylogenetic tree of the present study, *P. setigerum* is sister to *P. somniferum* with full support within the *Papaver* clade, which cannot be applied for determining the phylogenetic relationship of these two species due to lack of population sampling. Therefore, more samples for each species in subsequent studies will help us to resolve the genetic relationship between *P. setigerum* and *P. somniferum*.

Conclusion

In the present study, we assembled the complete chloroplast genome sequence of *Papaver setigerum* based on genome skimming data. The chloroplast genome of *P. setigerum* had a typical quadripartite structure with similar size and organization to other sequenced angiosperms. The evolutionary pattern of cp genomes in Papaveraceae was also detected utilizing seven representative species. Moreover, we screened additional cp hotspots regions for the genus *Papaver*, which will contribute to identification of species in this genus. The inter-genic region *psb*A-*trn*H is the best marker to distinguish *P. somniferum* and *P. setigerum*.

Acknowledgments

Conflict of Interest

The authors declare that they have no potential conflict of interest.

Author Contributions

     PL and JL designed the study. PL conducted the sampling. LXL, YXD and CS produced and analyzed the data. LXL, YXD and RL wrote the manuscript. PL and JL revised the manuscript. All authors approved the final manuscript.

References

Alkan C, Sajjadian S and Eichler EE (2011) Limitations of next-generation genome sequence assembly. Nat Methods 8:61-65.

Asaf S, Waqas M, Khan AL, Khan MA, Kang SM, Imran QM, Shahzad R, Bilal S, Yun BW and Lee IJ (2017) The complete chloroplast genome of wild rice (*Oryza minuta*) and its comparison to related species. Front Plant Sci 8:304.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD *et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455-477.

Choe S, Lee E, Jin GN, Lee YH, Kim SY, Choi H, Chung H, Hwang BY and Kim S (2012) Genetic and chemical components analysis of Papaver setigerum naturalized in Korea. Forensic Sci Int 222:387-393.

Dagan T, Roettger M, Stucken K, Landan G, Koch R, Major P, Gould SB, Goremykin VV, Rippka R, and Tandeau de Marsac N *et al.* (2012) Genomes of *Stigonematalean cyanobacteria* (subsection V) and the evolution of oxygenic photosynthesis from prokaryotes to plastids. Genome Biol Evol 5:31-44.

Daniell H, Lee SB, Grevich J, Saski C, Quesada-Vargas T, Guda C, Tomkins J and Jansen RK (2006) Complete chloroplast genome sequences of *Solanum bulbocastanum*, *Solanum lycopersicum* and comparative analyses with other Solanaceae genomes. Theor Appl Genet 112:1503-1518.

Darling AC, Mau B, Blattner FR and Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394.

Dong WP, Xu C, Cheng T and Zhou SL (2013) Complete chloroplast genome of *Sedum sarmentosum* and chloroplast genome evolution in Saxifragales. PloS One 8:e77965.

Fan L, Cheng B and Hong L (1987) A preliminary study on species differences among *Papaver somniferum* L, *Papaver rhoeas* L and *Cannabis sativa* L by AFLP technique. Chin J Forensic Med 23:157–159.

Farmilo C, Rhodes H, Hart H and Taylor H (1953) Detection of morphine in *Papaver setigerum* DC. Bull Narc 5:26-31.

Frazer KA, Pachter L, Poliakov A, Rubin EM and Dubchak I (2004) VISTA: computational tools for comparative genomics. Nucleic Acids Res 32:W273-W279.

Fulton CC (1944) The opium poppy and other poppies. Government Printing Office, Washington, 85 p.

Greiner S, Sobanski J and Bock R (2015) Why are most organelle genomes transmitted maternally? Bioessays 37:80-94.

Greiner S, Wang X, Herrmann RG, Rauwolf U, Mayer K, Haberer G and Meurer J (2008) The complete nucleotide sequences of the 5 genetically distinct plastid genomes of *Oenothera*, subsection *Oenothera*: II. A microevolutionary view using bioinformatics and formal genetic data. Mol Biol Evol 25:2019-2030.

Hosokawa K, Shibata T, Nakamura I and Hishida A (2004) Discrimination among species of *Papaver* based on the plastid

*rpl*16 gene and the *rpl*16-*rpl*14 spacer sequence. Forensic Sci Int 139:195-199.

Jin JJ, Yu WB, Yang JB, Song Y, Yi TS and Li DZ (2018) GetOrganelle: a simple and fast pipeline for de novo assembly of a complete circular chloroplast genome using genome skimming data. bioRxiv:256479.

Kalis A (1979) Papaveraceae. Rev Palaeobot Palyno 28:A209-A260.

Katoh K and Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772.

Kim HW and Kim KJ (2016) Complete plastid genome sequences of *Coreanomecon hylomeconoides* Nakai (Papaveraceae), a Korea endemic genus. Mitochondrial DNA B 1:601-602.

Krause K (2012) Plastid genomes of parasitic plants: a trail of reductions and losses. In: Bullerwell CE (ed) Organelle genetics. Springer-Verlag, Heidelberg, pp 79-103.

La Valva V, Sabato S and Gigliano GS (1985) Morphology and alkaloid chemistry of *Papaver setigerum* DC. (Papaveraceae). Taxon:191-196.

Li P, Lu RS, Xu WQ, Ohitoma T, Cai MQ, Qiu YX, Cameron KM and Fu CX (2017) Comparative genomics and phylogenomics of East Asian tulips (*Amana*, Liliaceae). Front Plant Sci 8:451.

Librado P and Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451-1452.

Liu LX, Li R, Worth JRP, Li X, Li P, Cameron KM and Fu CX (2017) The complete chloroplast genome of Chinese bayberry (*Morella rubra*, Myricaceae): implications for understanding the evolution of Fagales. Front Plant Sci 8:968.

Liu LX, Li P, Zhang HW and Worth J (2018a) Whole chloroplast genome sequences of the Japanese hemlocks, *Tsuga diversifolia* and *T. sieboldii*, and development of chloroplast microsatellite markers applicable to East Asian *Tsuga*. J Forest Res 23:318-323.

Liu LX, Wang YW, He PZ, Li P, Lee J, Soltis DE and Fu CX (2018b) Chloroplast genome analyses and genomic resource development for epilithic sister genera *Oresitrophe* and *Mukdenia* (Saxifragaceae), using genome skimming data. BMC Genomics 19:235.

Lohse M, Drechsel O, Kahlau S and Bock R (2013) OrganellarGenomeDRAW-a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res 41:W575-W581.

Lu F, Cheng BW, Li H, Zeng FM, Hong JY, Wen YB, Jiao DQ, Li LS, Zhao WS and Fang P (2008) A preliminary study on species differences among *Papaver somniferum* L., *Papaver rhoeas* L. and *Cannabis sativa* L. by AFLP technique. Chin J Forensic Med 23:157-159.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q and Liu Y (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience 1:18.

Miller MA, Pfeiffer W and Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Proceedings of the Gateway Computing Environments Workshop (GCE). IEEE, New Orleans, pp 1-8.

Moore MJ, Soltis PS, Bell CD, Burleigh JG and Soltis DE (2010) Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proc Natl Acad Sci USA 107:4623-4628.

Pignatti S (1982) Flora d'italia. Edagricole, Bologna, vol. 1, 790 pp.

Raubeson L and Jansen R (2005) Chloroplast genomes of plants. In: Henry RJ (ed) Plant diversity and evolution: genotypic and phenotypic variation in higher plants. CABI Publishing, Wallingford, pp 45-68.

Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boore JL and Jansen RK (2007) Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. BMC Genomics 8:174.

Rezaei Osalou A, Daneshvar Rouyandezagh S, Alizadeh B, Er C and Sevimay CS (2013) A comparison of ice cold water pretreatment and α-bromonaphthalene cytogenetic method for identification of *Papaver* species. Sci World J 2013:608650.

Ronquist F and Huelsenbeck J (2003) MrBayes: Bayesian phylogenetic inference under mixed models, 3.0 b4. Bioinformatics 19:1572-1574.

Schattner P, Brooks AN and Lowe TM (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res 33:W686-W689.

Shi C, Liu Y, Huang H, Xia EH, Zhang HB and Gao LZ (2013) Contradiction between plastid gene transcription and function due to complex posttranscriptional splicing: an exemplary study of *ycf*15 function and evolution in angiosperms. PLoS One 8:e59620.

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchishinozaki K *et al.* (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J 5:2043.

Straub SC, Parks M, Weitemier K, Fishbein M, Cronn RC and Liston A (2012) Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. Am J Bot 99:349-364.

Thomson RC, Wang IJ and Johnson JR (2010) Genome-enabled development of DNA markers for ecology, evolution and conservation. Mol Ecol 19:2184-2195.

Tsuji S, Ueda K, Nishiyama T, Hasebe M, Yoshikawa S, Konagaya A, Nishiuchi T and Yamaguchi K (2007) The chloroplast genome from a lycophyte (microphyllophyte), *Selaginella uncinata*, has a unique inversion, transpositions and many gene losses. J Plant Res 120:281-290.

Twyford AD and Ness RW (2017) Strategies for complete plastid genome sequencing. Mol Ecol Resour 17:858-868.

Wicke S, Schneeweiss GM, Depamphilis CW, Kai FM and Quandt D (2011) The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol 76:273.

Young HA, Lanzatella CL, Sarath G and Tobias CM (2011) Chloroplast genome variation in upland and lowland switchgrass. PloS One 6:e23980.

Zhang CJ and Cheng CG (2009) Identification of *Papaver somniferum* L. and *Papaver rhoeas* using DSWT-FTIR-RBFNN. Spectrosc Spect Anal 29:1255-1259.

Zhang S, Liu Y, Wu Y, Cao Y and Yuan Y (2015) Screening potential DNA barcode regions of genus *Papaver*. Zhongguo Zhong Yao Za Zhi 40:2964-2969.

Zhou J, Cui Y, Chen X, Li Y, Xu Z, Duan B, Li Y, Song J and Yao H (2018) Complete chloroplast genomes of *Papaver rhoeas* and *Papaver orientale*: Molecular structures, comparative analysis, and phylogenetic analysis. Molecules 23:437.

Supplementary Material

The following online material is available for this article:
Table S1 - The GenBank accession numbers for other seven species used in comparative chloroplast genome analyses.
Table S2 - The variable sites information of seven chloroplast (cp) regions between *P. somniferum* and *P. setigerum*.

Associate Editor: Marcela Uliano Silva