Diverse trajectories of plastome degradation in holoparasitic *Cistanche* and the whereabouts of the lost plastid genes

Xiaoqing Liu[1], Weirui Fu[1], Yiwei Tang[1], Wenju Zhang[1], Zhiping Song[1], Linfeng Li[1], Ji Yang[1], Hong Ma[2,3], Jianhua Yang[4], Chan Zhou[5], Charles C. Davis[6] and Yuguo Wang[1,†]

[1] Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, Institute of Biodiversity Science, School of Life Sciences, Fudan University, Shanghai 200433, China

[2] State Key Laboratory of Genetic Engineering, School of Life Sciences, Institute of Plant Biology, Center for Evolutionary Biology, Fudan University, Shanghai 200433, China

[3] Department of Biology, Institute of Molecular Evolutionary Genetics, and the Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

[4] College of Pharmacy, The First Affiliated Hospital, Xinjiang Medical University, Urumqi 830011, China

[5] Department of Population and Quantitative Health Sciences, Massachusetts General Hospital, 55 Lake Ave North Worcester, MA 01605, USA

[6] Department of Organismic and Evolutionary Biology, Harvard University Herbaria, 22 Divinity Avenue, Cambridge, MA 02138, USA

[†] **Correspondence:** wangyg@fudan.edu.cn; Tel: 0086-21-31246697

The email address for each author:
Xiaoqing Liu, 15110700073@fudan.edu.cn;
Weirui Fu, 15210700090@fudan.edu.cn;
Yiwei Tang, 16210700094@fudan.edu.cn;
Wenju Zhang, wjzhang@fudan.edu.cn;

Zhiping Song, songzp@fudan.edu.cn;

Linfeng Li, lilinfeng@fudan.edu.cn;

Ji Yang, jiyang@fudan.edu.cn;

Hong Ma, hongma@fudan.edu.cn;

Jianhua Yang, yjh-yft@163.com;

Chan Zhou, zhouchan99@gmail.com;

Charles C. Davis, cdavis@oeb.harvard.edu;

Yuguo Wang, wangyg@fudan.edu.cn.

Highlight

Comparative plastome analysis of holoparasitic *Cistanche* and its relatives revealed the clade-specific pattern of plastome degradation in a single genus, and different genomic locations of the lost plastid genes.

Abstract

The plastid genomes (plastomes) of non-photosynthetic plants generally undergoes gene loss and pseudogenization. Despite massive plastomes reported in different parasitism types of the broomrape family (Orobanchaceae), more plastomes representing different degradation patterns in a single genus are expected to be explored. Here, we sequenced and assembled the complete plastomes of three holoparasitic *Cistanche* species (*C. salsa*, *C. tubulosa* and *C. sinensis*) and compared them with the available plastomes of Orobanchaceae. We identified that the diverse degradation trajectories under purifying selection existed among three *Cistanche* clades, showing obvious size differences on entire plastome, long single copy region and non-coding region, and different patterns of the retention/loss of functional genes. With few exception of putatively functional genes, massive plastid fragments which have been lost and transferred into the mitochondrial or nuclear genomes are nonfunctional. In contrast with the equivalents of the *Orobanche* species, some plastid-derived genes with diverse genomic locations are found in *Cistanche*. The early and initially diverged clades in different genera such as *Cistanche* and *Aphyllon* possess obvious patterns of plastome degradation, suggesting that such key lineages should be considered prior to comparative analysis of plastome evolution, especially in the same genus.

Keywords: *Cistanche*, plastome, degradation, pseudogenization, gene loss, intracellular gene transfer

Abbreviations:

IGT, intracellular gene transfer; mipt, mitochondrial plastid insertion; nupt, nuclear plastid insertion; ORFs, open reading frames; MRCA, the most recent common ancestor; IR, inverted repeat; LSC, long single copy; SSC, short single copy.

Introduction

Despite the overall stability in architecture, gene content, and gene order of the plastid genomes across most angiosperms, parasitic plants are exceptions with plastid genomes that tend to degrade in comparison with non-parasitic plants (dePamphilis and Palmer, 1990; Jansen and Ruhlman, 2012). There are dramatic differences in genome size and gene content between parasitic and photo-synthetic plants (Wicke *et al*., 2013; Cusimano and Wicke, 2016). In parasitic plants, holoparasites generally possess extremely special plastomes with a functional and physical reduction, due to the pseudogenization and massive loss of photosynthesis-associated genes (Wolfe *et al*., 1992a, 1992b; Delavault *et al*., 1996; Funk *et al*., 2007; McNeal *et al*., 2007, 2009). Some holoparasites may have even lost their entire plastid genomes, e.g. *Rafflesia lagascae* (Molina *et al*., 2014).

The broomrape family Orobanchaceae is the only family with species that span the full trophic spectrum of parasitism ranging from holoparasites to hemi-parasites and free-living nonparasites (Westwood *et al*., 2010), providing us a good system to study the patterns of plastome degradation and to compare various plastomes size among different species (Li *et al*., 2013; Wicke *et al.*, 2013; Cho *et al.*, 2015; Cusimano and Wicke, 2016; Fan *et al*., 2016; Roquet *et al*., 2016; Samigullin *et al*., 2016; Schneider *et al*., 2018). Wicke *et al.* (2013) studied the complete plastomes of ten photosynthetic and nonphotosynthetic parasites plus their nonparasitic sister group from Orobanchaceae and found that the plastomes of this family varied 3.5-fold in size. Among them, the plastome of *Conopholis americana* (45 kb) and *Phelipanche ramosa* (62 kb) had lost one inverted repeat (IR) region, while *P. purpurea* have a shortened IR region, extending only over the *ycf*2 gene. Cusimano and Wicke (2016) analyzed the plastomes of several parasites, focusing predominantly on the genus *Orobanche*, and found that the physical plastome reductions are proceeded by small deletions that accumulate over time. The IR region of *Orobanche gracilis* is the shortest among the *Orobanche* species reported, in spite of their other regions being similar (Wicke *et al*., 2013; Cusimano and Wicke, 2016). The sequences of increasing plastomes in Orobanchaceae have been available, however, comparative studies on gene loss or pseudogenization from plastomes of the same genus remain insufficient. It is unclear whether there are obvious differences in the evolution of plastomes among different clades of the other genera in Orobanchaceae, and whether their plastome degradation occurs in similar or different regions.

During the evolution of plastomes, massive genes or gene fragments have been lost or transferred into other genomes one or more times (Bock and Timmis, 2008; Lloyd *et al.*, 2012; Rice *et al.*, 2013; Bellot and Renner, 2015; Naumann *et al.*, 2015; Cusimano and Wicke, 2016; Su *et al.*, 2019). In Orobanchaceae, Cusimano and Wicke (2016) reported that some plastid genes transferred into its mitochondria and nuclei in the species of *Orobanche*. Here, we sequenced the plastomes of three holoparasitic *Cistanche* species: *Cistanche sinensis*, *Cistanche tubulosa*, and *Cistanche salsa*, and compared their plastome size and patterns of gene loss with those of other available species of holoparasitic Orobanchaceae (e.g. *Cistanche deserticola* and *Cistanche phelypaea*), and revealed the differences among the five *Cistanche* species in plastome size, structure and gene content, as well as the fate of lost plastid genes in the evolution of *Cistanche*. Our study aims to find out the similarity and difference of plastome degradation patterns in different lineages of the same genus of Orobanchaceae, and compare the evolutionary fates of different kinds of these lost genes.

Materials and Methods

Taxon sampling and DNA sequencing

The samples of three *Cistanche* species (*C. salsa*, *C. tubulosa* and *C. sinensis*) representing two sections within *Cistanche* were collected from Xinjiang and Ningxia of China. The voucher specimens were deposited in the Herbarium of Fudan University (FUS), Shanghai, China. Total genomic DNAs (gDNAs) were extracted from silica-gel dried tissue using the improved CTAB method (Doyle and Doyle, 1987) or the Plant Genomic DNA Kit (Tiangen Biotech Co., Beijing, China) following the manufacturer’s instructions. For each *Cistanche* species, an Illumina library with the insert size of 350 ± 50 bp was prepared from 5 µg of gDNA following the protocol of Bentley *et al.* (2008). All libraries were sequenced on a HiSeq2000 platform for 150 bp paired-end (PE) sequencing.

Plastome assembly, annotation and repeat analysis of Cistanche species

Illumina reads from the three *Cistanche* species were assembled using two different approaches: a mapping approach and a reference-assisted *de novo* assembly approach (Hahn *et al.*, 2013; Westbury *et al.*, 2017). The mapping approach allows us to obtain a relatively correct gene order for the three *Cistanche* plastomes according to their closest relatives, and the reference-assisted assembly approach enables us to obtain accurate nucleotide sequences of the three *Cistanche* plastomes.

In the mapping approach, the plastome of *C. phelypaea* (GenBank accession no. HG515538) was used as reference for obtaining the plasotmes of *C. tubulosa* and *C. sinensis*. For *C. salsa*, we used the plastome of *C. deserticola* (KC128846) as reference. Total clean reads of three *Cistanche* species were mapped to references using Bowtie2 v2.3.4.1 (Langmead and Salzberg, 2012). After replacing all the single nucleotide polymorphisms (SNPs) and Insertions and Deletions (InDels) with our data at the corresponding sites, we obtained the plastome sequences of the three *Cistanche* species.

In the reference-assisted *de novo* assembly approach, SOAPdenovo2 v2.04 (Luo *et al.*, 2012) was used to assemble the plastomes of the three *Cistanche* species. First, the plastome sequences of six Orobanchaceae species (*C. phelypaea*, HG515538; *C. deserticola*, KC128846; *Aphyllon californicum*, HG515539; *Orobanche gracilis*, HG803179; *Schwalbea americana*, HG738866; *Phelipanche ramosa*, HG803180) were used as reference bait sequences for extracting plastid reads. Next, all candidate plastid reads were extracted and collected by Magic-BLAST v2.1 (https://ncbi.github.io/magicblast/, Splice = T; other parameters were set as default) and Seqtk v1.0.1 software (https://github.com/lh3/seqtk.git). After that, all the extracted plastid reads were respectively used to *de novo* assemble the plastomes of the three *Cistanche* species using SOAPdenovo2. The preliminary assemble results with maximum scaffold N50 values were selected for further analysis. All contigs and scaffolds were sorted according to their collinearity with *C. deserticola* and *C. phelypaea* by Mummer v3.23 (Kurtz *et al.*, 2004) and the missing parts were regarded as gaps until the plastome sequences of the three *Cistanche* species were obtained. Gaps between contigs or scaffolds were closed with the total genomic data of each *Cistanche* species by GapCloser (a module of SOAPdenovo2, Luo *et al.*, 2012). The results from the two above approaches were combined. For the IR-single copy region connections and retained gaps in the intergenic spacers, primers were designed for polymerase chain reaction (PCR) amplification and sequence verification.

The reaction mixture included 6 μl of 10 × PCR buffer, 6.4 μl of 2.5 mM deoxynucleoside triphosphates, 5 μl of 2.5mM Mg^{2+}, 2.5 U of TaqE, 3 μl of 10 μM forward and reverse primers, and about 1 μg of template genomic DNA, with deionized water added to 50 μl. Each PCR program had a 4 min hot start at 94 °C, followed by denaturing at 94 °C for 30~60 s, annealing at Tm for 1~2 min; and extension at 72 °C for 2~3 min, for 35 cycles; one cycle of denaturing at 72 °C for 10 min. The TAIL-PCR method (Liu and Whittier, 1995; Liu and Huang, 1998) was also used for amplification of the downstream flanking region of *ycf*1 gene to obtain the complete short single copy (SSC) region of *C. sinensis*. Primers and programs for completing and verifying the plastomes of the three *Cistanche* species can be found in Supplementary Table S1-S4.

The plastomes of the three *Cistanche* species were preliminarily annotated using DOGMA (http://dogma.ccbb.utexas.edu/, Wyman *et al.*, 2004). The annotation results were further adjusted manually. To identify the initiation and termination sites of the protein-coding genes, open reading frames (ORFs) were identified by ORFfinder on the NCBI website (https://www.ncbi.nlm.nih.gov/orffinder/) and the corresponding sequences of closely related species of *Cistanche* were also applied as reference. Compared with the reference species, pseudogenes were determined by their short size and lack of start codons (Wicke *et al.*, 2013; Bellot and Renner, 2015). Protein-coding genes that were truncated at 5' end, lacking an unambiguously identifiable translation start, and could not translated into amino acid sequences were annotated as pseudogenes. Considering the conservation of plastid genes and the inherent properties of the splice sites (Black*,* 2003; Matlin *et al.*, 2005), the splicing sites at the exon boundary of the plastid genes in the three *Cistanche* species were predicted according to their closest relatives. The plastomes of the three *Cistanche* species were further annotated with Sequin v15.10 (www.ncbi.nlm.nih.gov/Sequin/index.html). The validated complete plastome sequences were deposited in GenBank (*C. salsa*, MK386640; *C. sinensis*, MK386641; *C. tubulosa*, MK386642). Graphical genome maps of the three *Cistanche* species were drawn by OGDraw (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html, Lohse *et al*., 2013). The forward and palindromic repeats longer than 20 bp and a Hamming distance of 3 in the plastid genomes of the three *Cistanche* species were identified and located using the online REPuter software (http://bibiserv.techfak.uni-bielefeld.de/reputer, Kurtz *et al.*, 2001). Repeats with e-value > 0.1 were not considered. The same REPuter analyzing process was run to assess the repeat number of *Cistanche* and the members of other closely related genera.

Estimation of the relative divergence time of different lineages from Cistanche and its closely related genera

To estimate the divergence time of the three *Cistanche* clades relative to the species of other closely related genera, thirty-two retained housekeeping plastid genes were used to construct the phylogenetic tree of Orobanchaceae. The DNA sequence matrixes of these genes were combined and aligned by nucleotide with MEGA5 (Tamura et al., 2011) and adjusted manually. All missing data were replaced with gaps. The phylogenetic tree was constructed using maximum likelihood (ML) method with RAxML v7.0.4 (Stamatakis, 2006), applying the GTRGAMMA model and 1,000 replications to evaluate the support of each branch. Divergence timing was analyzed by Bayesian Evolutionary Analysis by Sampling Trees (BEAST 1.8.2; Drummond *et al*., 2012). Owing to lacking of a reliable fossil record within Orobanchaceae, we followed the methods of Fu *et al.* (2017) to set two external fossil calibration points (the stem age of Solanaceae and the crown age of *Pedicularis* and *Olea*, Zanne *et al*., 2014) and execute further analyses to generate the final maximum clade credibility tree. FigTree 1.3.1 (Rambaut, 2006) was used to visualize the topology and node height with 95% highest posterior density (HPD).

Comparisons of nucleotide substitution rates of plastid protein-coding genes of Cistanche species

Six functional protein-coding plastid genes (*mat*K, *rps*2, *rpl*16, *rps*4, *rps*7, and *rps*14) available in all sampling species of Orobanchaceae were used to check whether the relative nucleotide substitution rates have elevated in *Cistanche* against other lineages in Orobanchaceae. Substitution rate analyses were carried out by the Codeml program of PAML v1.3.1 (Yang, 2007; Xu and Yang, 2013). The sequences of a non-parasitic plant (*Lindenbergia philippensis*), nine hemiparasites representing seven genera (*Buchnera*, *Castilleja*, *Neobartsia*, *Pedicularis*, *Schwalbea*, *Striga* and *Triphysaria*), and 23 holoparasites representing eight genera (*Aphyllon*, *Boulardia*, *Cistanche*, *Conopholis*, *Epifagus*, *Lathraea*, *Orobanche* and *Phelipanche*) in Orobanchaceae were downloaded from GenBank and included in this analysis. Considering potential rate differences in specific lineages, the branch model was used to estimate rates of synonymous and non-synonymous substitutions (dS and dN) and the ω ratio (dN/dS). Using a likelihood ratio test (LRT) (Yang and Nielsen, 1998; Yang, 1998), we further tested whether the *Cistanche* lineage possesses different ω ratios from other lineages. For pairwise dN/dS evaluations in *Cistanche* species, we referred the method of Schelkunov *et al*. (2015) to perform the PAML analysis in a pairwise mode (runmode= -2). The initial dN/dS and tS/tV ratios were set to 0.5 and 2.0, respectively, with a codon frequency model F3×4.

Identification of genomic location of the lost plastid genes or fragments

To identify the genomic location of plastid genes or fragments lost from plastomes, 29 mitochondrial genomes of 19 families representing 14 orders and 100 plastomes of 21 families representing four orders were taken as reference bait sequences (Supplementary Tables S5, S6). All of the mitochondrial and plastid reads of the three *Cistanche* species were extracted according to these references by Magic-BLAST v2.1 and Seqtk v1.0.1. All the extracted reads were then combined as a read pool to carry out *de novo* assemble using SOAPdenovo2.

All the contigs and scaffolds were annotated against all the available data in GenBank using BLASTN v2.2.23 with an e-value ≤ 10^{-5}. The lost plastid genes or fragments from the assembled contigs or scaffolds of the three *Cistanche* species were searched. Contigs or scaffolds that were annotated as plastid genes or fragments with mitochondrial or nuclear flanking sequences were used to identify mitochondrial plastid insertions (mipts) or nuclear plastid insertions (nupts). Sequences of mipts and nupts for further phylogenetic analyses were deposited to NCBI under GenBank accession numbers (MK413701, MK413702, and MK413703). The lost plastid genes were also investigated through mapping the total genomic clean reads of the three *Cistanche* species to the corresponding genes of their closest relatives using Bowtie2 v2.3.4.1 (Langmead and Salzberg, 2012). The average coverage of multiple plastid, mitochondrial and nuclear genes in *C. tubulosa* were calculated to infer the locations of the lost plastid genes in the three *Cistanche* species.

Results

Structure and physical features of Cistanche plastomes

The sequence quality of the three *Cistanche* species and total clean reads are shown in Supplementary Table S7. The average coverage of the plastomes of *C. salsa*, *C. tubulosa* and *C. sinensis* are 2,731.25×, 1,507.09× and 3,025.83×, respectively. Detailed characteristics of the five *Cistanche* plastomes are listed in Table 1, the physical maps of the five *Cistanche* plastomes are shown in Fig. 1, and the circular maps of the three newly sequenced plastid genomes can be found in Supplementary Fig. S1. Phylogenetic relationships among these *Cistanche* species and closely related taxa are shown in Supplementary Fig. S2. Similar to the vast majority of angiosperms, the plastomes of the three *Cistanche* species are consisted of a long single copy (LSC) region, a SSC region and two IRs that separate the two single copy regions (Fig. 2). With a total length of 87,707 bp, *C. sinensis* possesses the smallest plastome of the five *Cistanche* species (Table 1). The plastid genomes of *C. salsa* and *C. tubulosa* are 101,776 bp and 94,123 bp, respectively. The GC content of *C. sinensis* is 37.95%, which is the highest among the five *Cistanche* species (Table 1).

In terms of the gene content in the plastomes of the five *Cistanche* species, there are 68 different types of genes in the plastome of *C. sinensis*, which is less than that of *C. salsa* (89 genes) and *C. tubulosa* (72 genes). There are 22 tRNA genes, four ribosomal RNA genes, 13 photosynthesis and energy production genes, 21 ribosomal protein and initiation factor genes, three RNA polymerase and intron maturase genes, and five other essential genes in *C. sinensis*. Twenty genes are duplicated by IRs in the plastome of *C. sinensis* (18 in the IRs of *C. salsa*, 22 in the IRs of *C. tubulosa*). Essential genes in the plastomes of the five *Cistanche* species including putatively functional gene with intact ORFs and structural RNAs (transfer and ribosomal RNAs) are shown in Table 2.

Three main clades of *Cistanche* has distinct gene status. The plastid coding gene *pet*G is the most typical example that it is putatively functional in *C. sinensis* but lost in the *C. tubulosa-C. phelypaea* clade and pseudogenized in the *C. deserticola-C. salsa* clade. Besides *pet*G, 16 genes including five degeneration types represented by *rpl*32, *psb*A, *ycf*1, *ndh*E, and *rps*3 can distinguish *C. sinensis* from other two clades. Four plastid genes (*atp*A/B/E/F) were consistently lost in the *C. tubulosa-C. phelypaea* clade but pseudogenized in other two clades, indicating that losses of these genes probably predated the differentiation of *C. tubulosa* and *C. phelypaea*, but postdated the differentiation of *C. sinensis* and *C. tubulosa*. Some plastid genes (*psa*I, *psb*B/D/L, *ycf*3 and *trn*V-UAC) show the different situation that they exist in the *C. deserticola-C. salsa* clade (pseudogenization or putatively functional gene) but lost in the other *Cistanche* species, suggesting that these genes exist at least before MRCA of the five *Cistanche* species. Some photosynthesis- and energy production-related genes such as *atp*H/I, *cem*A, *ndh*A/C/D/F/G/I/J/K, *pet*A/B/D/N, *psa*C, *psb*H/N/T, and *rpo*C1 are consistently lost from the plastomes of all five *Cistanche* species. Furthermore, the species within same clade (*C. deserticola* vs. *C. salsa* and *C. tubulosa* vs. *C. phelypaea*) tend to have consistent patterns of pseudogenization and gene loss. The clade-specific patterns of gene degeneration in the five *Cistanche* plastomes are shown in Table 2.

In the plastomes of the five *Cistanche* species, the repeat number of *C. sinensis* is the lowest, including 18 forward and 69 palindromic repeats, with at least 1,008 bp per repeat-unit with a sequence identity of more than 90%. As for the repeat density, *C. sinensis* is also the smallest (1/1008) among the five *Cistanche* species; *C. deserticola* is the biggest (1/446); and the rest three species range from low to high as follows: *C. tubulosa* (1/645), *C. salsa* (1/476), and *C. phelypaea* (1/465). Compared with the species of other closely related genera, the repeat density of *Cistanche sinensis* is smaller than that of *Orobanche* species, but it is larger than that of most *Aphyllon* species (excluding *A. californucum*). *Phelipanche* species possess the largest repeat density among these four genera.

Nucleotide substitution rate analyses of the retained plastid genes in Cistanche

As expected, analysis of relative nucleotide substitution rates based on a concatenated set of six plastid protein-coding genes (*mat*K, *rps*2, *rpl*16, *rps*4, *rps*7, and *rps*14) which shared in all sampling Orobanchaceae showed that there was no significant difference between the species of *Cistanche* and the members of *Aphyllon*, *Orobanche* and *Phelipanche.* Selectional strength was shifted towards a more neutral evolution in these genes of Orobanchaceae (ω between 0.0001 and 2.133), with all obligate parasites (including the hemi-parasitic *Schwalbea americana*) adopting a higher ω than the nonparasitic species, *Lindenbergia philippensis* as previous study of Cusimano and Wicke (2016). Detailed results are shown in Fig. 3 and Supplementary Table S8. Pairwise comparisons of these protein-coding genes of *Cistanche* and a photosynthetic *Lindenbergia philippensis* confirmed that the dN/dS value is lower than 1 (Fig. 4), strongly supporting the idea that these plastid genes are under purifying selection in the species of *Cistanche*.

Genomic location analysis of the lost plastid genes or fragments

In this study, we found that numerous lost plastid genes or fragments had been transferred into mitochondrion genome and nucleus through intracellular gene transfer (IGT) (Table 3). In *C. tubulosa*, the flanking regions of the plastid genes *pet*A/G and *rpo*C1 are mitochondrial sequences, supported by average coverage range of which ranged from 440.32× to 482.62×. *C. tubulosa* nested within Lamiales showing its vertical placement on the phylogenetic tree of these genes (Fig. 5). In addition, the plastid gene of *acc*D in *C. salsa* was suggested to have been transferred into nucleus because of its flanking sequences of nuclear origin (*C. salsa*-C1185, MK413701). Distinct from horizontal gene transfer (HGT), it had a vertical placement on the phylogenetic tree close to *C. deserticola* (Fig. 5, Table 3).

The average coverage of multiple plastid, mitochondrion, and nuclear genes in *C. tubulosa* were investigated using the total clean genome data and were found to range from 1026.11× (*rps*4) to 1540.20× (*rbc*L) for plastid genes, 113.37× (*sdh*4) to 385.53× (*mat*R) for mitochondrial genes, and usually less than 20× for nuclear genes (e.g. 0.95×, 7S globulin gene; 8.55×, alpha-tubulin gene). Based on the relatively clear ranges of different types of genes, genomic location of lost plastid genes of the three *Cistanche* species can be inferred (Table 3). The average coverage of the plastid gene *atp*H in *C. salsa* and *C. sinensis* is 1.63× and 14.27×, and it is inferred as being located in their nuclei; inferred from average coverage, the *atp*I gene were probably transferred into nucleus in *C. salsa* (13.28×), and mitochondrion in *C. tubulosa* (146.99×) and *C. sinensis* (428.14×); the average coverage of the *pet*A gene in *C. salsa* and *C. tubulosa* is 85.47× and 475.85×, which is close to the mitochondrion range, suggesting that they are located in mitochondrial genome. This is consistent with the assembly result (*C. tubulosa*-scaffold190, MK413703), the flanking region of which is mitochondrial sequence. Similarly, the average coverage of *rpo*C1 gene in *C. tubulosa* and *C. sinensis* is 107.23× and 44.60×, and is inferred to be located in the mitochondrial genome, one of which is also supported by the assembly result (*C. tubulosa*-scaffold133, MK413702).

Comparison of the relative divergence time of different lineages in Orobanchaceae

Molecular timing based on the plastid genomes can elucidate the relative occurrence order of different lineages in Orobanchaceae (Fig. 2). The *Cistanche* species are deeply diverged: the *C. sinensis* clade was the earliest-diverging lineage, followed by the *C. tubulosa*-*C. phelypaea* clade and the *C. salsa*-*C. deserticola* clade. Similarly, *Aphyllon californicum* was the basalmost clade of *Aphyllon*, followed by the *Aphyllon purpureum*-*Aphyllon fasciculatum* clade and the *Aphyllon epigalium*-*Aphyllon franciscanum* clade. The divergence time of *A. californicum* and other six *Aphyllon* species postdates that of *C. sinensis* and other four *Cistanche* species, but predates that of *C. salsa* and *C. tubulosa*. Similarly, the timing of the most recent common ancestor (MRCA) of *A. fasciculatum*-*A. franciscanum* are comparable to that of *C. salsa*-*C. tubulosa* and that of the *Orobanche* species, followed by the divergence of three *Phelipanche* species.

Discussion

The distinct plastome characteristics of three Cistanche *clades*

The plastome architecture is highly conserved in most flowering plants (Palmer, 1985). Generally, the plastome size and architecture are similar within the same genus, e.g. the photosynthetic and non-parasitic *Rehmannia* and holoparasitic *Phelipanche* (Wicke *et al*., 2016; Zeng *et al*., 2017). However, there are some exceptions in Orobanchaceae, e.g. the plastome sizes of *Orobanche gracilis* and *Aphyllon californicum* possessing extremely small and big plastome in their own genus, respectively (Cusimano and Wicke, 2016; Schneider *et al*., 2018). Among the five *Cistanche* species, there are clear clade-specific differences in plastome size. The *C. salsa*-*C. deserticola* clade and *C. sinensis* possessed the largest plastome and the smallest plastome, respectively, while the *C. phelypaea*-*C. tubulosa* clade possesses medium-sized plastome. Inferred from phylogenetic timing, *C. sinensis* was the earliest diverging clade, followed by the *C. tubulosa*-*C.*

phelypaea clade and the *C. salsa*-*C. deserticola* clade. Our results suggest the difference in divergence time of the species of *Cistanche* may be contributed to diverse degrees of plastome degradation, affecting plastome size of different clades.

The differences of plastome size in the three *Cistanche* clades are associated with their gene content. The shortened IR region of *O. gracilis* result in its smallest plastome among *Orobanche* species, and the largest plastome size of *A. californicum* among *Aphyllon* species is attributed largely to the expansion of its single copy regions (LSC and SSC). The diversity in plastome sizes of the *Cistanche* species can be attributed to the length changes of their LSC regions. The LSC region of *C. sinensis* is the shortest among the five *Cistanche* species, while the *C. salsa*-*C. deserticola* clade possesses the longest LSC region.

The total length of noncoding regions including rRNAs, tRNAs and intergenic regions among three *Cistanche* clades also shows obvious differences (Table 1). Among three *Cistanche* clades, the size of noncoding region of *C. sinensis* is the shortest, whereas the *C. deserticola*-*C. salsa* clade and the *C. phelypaea*-*C. tubulosa* clade possesses the longest and medium-sized noncoding region, respectively.

Additionally, the extreme short intergenic region (Table 1) and the extreme small repeat number of *C. sinensis* are inferred to be attributed largely to its smallest plastome size among *Cistanche* species. Compared with other two *Cistanche* clades, *C. sinensis* possesses the largest GC content, as well as the smallest plastome with the shortest noncoding region, making it unique among the five *Cistanche* species.

In contrast to the complex gene contents among the closely related genera (Fig. 6), the plastid gene content of the three clades of *Cistanche* are distinctly different. The number of plastid genes vary from 60-80 in *Orobanche*, 61-68 in *Phelipanche* and 66-80 in *Aphyllon*. There is no obvious gene content pattern in different clades of these genera. In *Cistanche*, however, *C. sinensis* possesses the least number of plastid genes (68 genes) and *C. deserticola*-*C. salsa* clade possesses the most plastid genes (89 genes) (Table 1).

Diverse patterns of pseudogenization and gene loss in plastomes of Cistanche and its relatives

As far as the functional gene loss was concerned, both ancient and recent origins of gene loss or pseudogenization can be traced by the phylogenetic relationships among the *Cistanche* species and their relatives (Fig. 7). The loss of different genes happened in almost every clade, while pseudogenization occured only in the specific clades.

Analyzed from gene status, there are two forms of gene loss. Some genes such as *psbH* and *ndh*A are missing in the known plastid genome of the whole Orobanchaeae, suggesting that gene losses occurred anciently and the molecular timing was estimated about 24.91 Ma; similarly, the timing of gene losses of *ndh*J, *atp*I/H, and *rpo*C1 was very ancient, dating back to 19.01 Ma. But in another case, genes such as *pet*G, *psa*J and *rpo*A are still retained in certain *Cistanche* species, while they are pseudogenes or completely lost in other *Cistanche* species, suggesting that these losses are recent. Phylogenetic analysis of single gene (e.g., *pet*G) can reveal that the loss of these genes should be directly from the complete genes or via pseudogenes indirectly (Fig. 8).

As other Orobanchaceae genera, obvious plastome degeneration patterns within *Cistanche* could not be found if we didn’t make a distinction between pseudogenization and gene loss. However, there are a clear difference among the three *Cistanche* clades when we consider pseudogenization and gene loss separately, especially when we analyze character evolution of putatively functional gene, pseudogene and lost gene of different *Cistanche* clades in the phylogenetic framework (Fig. 9, Supplementary Fig. S4). The clade-specific pattern is most evident between *C. sinensis* and the clade of the remaining *Cistanche* species. The plastome size of the early and initially diverged clade represented by *C. sinensis* has obviously different from that of the latter.

Although not all genera exhibit clear clade specificity, we have still found similar clade-specific patterns in *Aphyllon* (Fig. 9, Supplementary Fig. S4)*. Aphyllon* represents another example that associates divergence time with extreme plastome size: as the earliest diverging species in *Aphyllon*, *A. californicum*, which divergent time separated from other species within a single genus is similar to that of *C. sinensis* (Fig. 2), possesses larger plastome size than the other reported *Aphyllon* species (Schneider *et al*., 2018).

Similar phenomena should exist in other taxa. Our study exemplifies that the well-differentiated parasitic lineages exhibit obvious differences of plastome degeneration among diverse clades within the same genus, indicating that exploring gene loss, pseudogenization and gene retention within the phylogenetic framework will help us understand the evolutionary process of plastome degeneration.

The fate of the lost plastid genes

Examples of lost plastid genes being integrated to mitochondrial and nuclear genome through IGT have been reported in some parasitic genera, e.g., *Pilostyles* (Apodanthaceae), and *Orobanche* (Orobanchaceae) (Bellot and Renner, 2015; Cusimano and Wicke, 2016). Incorporation of exogenous chloroplast-derived sequences of host into the mitochondrial genomes of parasitic plants through HGT is not rare, e.g. *Rafflesia* and *Sapria* (Rafflesiaceae) (Xi *et al.*, 2013; Molina *et al.*, 2014), and *Aphyllon* (Schneider *et al*., 2018), and the inverse mitochondrion-to-mitochondrion HGT from parasitic plant to host has been found in *atp*I (Mower *et al.*, 2004). However, no transfer of plastid-derived genes from the parasitic plant level to the host has been found. In Orobanchaceae, IGTs of only five *Orobanche* species (*O. austrohispanica*, *O. crenata*, *O. densiflora*, *O. gracilis*, and *O. rapum-genstae*) were studied (Cusimano and Wicke, 2016) and massive mipts and nupts were found in this genus. Similarly, we found 16 mipts and 19 nupts in three *Cistanche* species through investigating the average coverage of lost plastid genes and the assembly results.

Diverse IGT patterns can also be found in three *Cistanche* species, e.g. for *psb*C and *psb*D which transferred into their nucleus in *Orobanche* species, they had diverse genomic locations in *Cistanche* species: the IGT copies of *psb*C were located in the mitochondrial genome of *C. sinensis* and *C. salsa*, while it was found in the nuclear genome of *C. tubulosa*, the IGT copy of *psb*D was found located in the nucleus of *C. salsa*, mitochondrion of *C. tubulosa*, while no IGT copy was found in *C. sinensis*; *pet*A and *pet*G were located in the mitochondrial genome of *C. tubulosa* and *C. salsa*, while they were lost and transferred into the nucleus of *C. sinensis*, respectively; *rpo*C1 were located in the mitochondrion of *C. sinensis* and *C. tubulosa*, whereas it was lost or transferred into nucleus of *C. salsa*. It has been uncertain whether the different whereabouts of lost plastid genes in these species represents clade specificity. However, the intracellular transfer patterns of mipts and nupts show obvious differences between the species of *Cistanche* and *Orobanche*, e.g. the *psa*B copies were transferred into mitochondrion or nucleus of the *Orobanche* species, but no IGT copy of *psa*B was found in *Cistanche* species; the *rpl*23 copies had been transferred into nucleus in *Orobanche* species, but no IGT copy of *rpl*23 occurred in the plastomes of *Cistanche* species.

To investigate the function of these transferred genes above, we referred the standard of Bellot and Renner (2015) to analyze their gene length and internal stop codons. Two mipts (*pet*A and *rpo*A) and one nupt (*acc*D) were classified as pseudogenes for their truncated length compared with their functional ones in flowering plants (Fig. 5). In *C. tubulosa*, one mitochondrial scaffold contains possibly functional plastid gene-*pet*G, as inferred from its intact ORF which can be translated into complete amino acid sequence (Fig. 5) and its length is same as that of *Lindenbergia philippensis*. The plastid gene *atp*I of *C. sinensis* probably has function judged from the coverage reads but it needs further verification. Other mipts and nupts, mostly in the form of fragments, mean that they have lost their function in *Cistanche*, almost in line with previous studies on parasitic plants (Bellot and Renner, 2015; Cusimano and Wicke, 2016). This may also be a common phenomenon in angiosperms (Notsu *et al.*, 2002; Goremykin *et al.*, 2009; Alverson *et al.*, 2010; Rice *et al.*, 2013).

At present, only few available genomic data can be used for specific comparison of the obvious differences of interspecific IGTs. From these data, however, it is difficult to order the movements into the nucleus or mitochondrion accurately in most cases for the poor phylogenetic resolution. Nevertheless, the movements shared with a clade should be judged as ancient IGTs. For instance, *atp*I has been transferred into the nucleus or mitochondrion in three species of *Cistanche*. Interestingly, this gene has been lost at the MRCA of the clade of (*Cistanche*, (*Conopholis*, *Epifagus*)). Similar intracellular transfer of this gene may also occur in the latter two genera. We would like to see that more species will be added to the comparative analysis in the future to better understand the regularity of gene loss during the process of plastome reduction.

In sum, we revealed that the infrageneric clades of *Cistanche* with different divergence time possess distinct plastome size and gene content, leading to diverse trajectories of plastome degradation. As for plastome size and gene content, a similar phenomenon can be found among different clades of a heterotrophic orchid complex (even the same species) with different divergence times (Barrett *et al*., 2018). The clade-specific pattern similar to *Cistanche* were also found in the species of *Aphyllon*, suggesting that this pattern should also occur in the other taxa. As more genomic data accumulate for different lineages involving parasitic, heterotrophic and other plants (e.g. *Pilostyles*, Bellot and Renner, 2015; *Hydnora*, Naumann *et al*., 2015; *Epipogium*, Schelkunov *et al*., 2015; *Monotropa*, *Hypopitys* and *Pyrola*, Logacheva *et al*., 2016; *Cytinus*, Roquet *et al*., 2016; *Balanophora*, Su *et al*., 2019), such phenomena of plastid genes that were lost from plastomes and had been transferred into mitochondrion or nucleus may not be limited to the known taxa, additional cases of mipts and nupts in the closely related genera will continue to be discovered. Our comparative plastome analyses enlighten that if different divergent clades of the same genus were sampled densely in other lineages, more new findings will be undertaken to improve our understanding on plant plastome evolution and the fate of the lost plastid genes.

Acknowledgements

We are grateful to the members of Evolutionary Ecology Lab of Fudan University for useful discussions and critical comments on this manuscript. This research was financially supported by the National Natural Sciences Foundation grants of China (No. 31870202, No. 31370248) and the National Basic Research Program of China (973 programs, No. 2014CB954100).

Conflict of Interest

The authors declare no conflict of interests.

References

Alverson AJ, Wei X, Rice DW, Stern DB, Barry K, Palmer JD. 2010. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). Molecular Biology and Evolution **27,** 1436–1448.

Barrett CF, Wicke S, Sass C. 2018. Dense infraspecific sampling reveals rapid and independent trajectories of plastome degradation in a heterotrophic orchid complex. New Phytologist **218,** 1192–1204.

Bellot S, Renner SS. 2015. The plastomes of two species in the endoparasite genus *Pilostyles* (Apodanthaceae) each retain just five or six possibly functional genes. Genome Biology and Evolution **8,** 189–201.

Bentley DR, Balasubramanian S, Swerdlow HP, *et al*. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature **456,** 53–59.

Black DL. 2003. Mechanisms of alternative pre-messenger RNA splicing. Annual Review of Biochemistry **72,** 291–236.

Bock R, Timmis JN. 2008. Reconstructing evolution: gene transfer from plastids to the nucleus. BioEssays **30,** 556–566.

Cho WB, Choi IS, Choi BH. 2015. Development of microsatellite markers for the endangered *Pedicularis ishidoyana* (Orobanchaceae) using next-generation sequencing. Applications in Plant Sciences **3**, 1500083.

Cusimano N, Wicke S. 2016. Massive intracellular gene transfer during plastid genome reduction in nongreen Orobanchaceae. New Phytologist **210,** 680–693.

Delavault PM, Russo NM, Lusson NA, Thalouarn P. 1996. Organization of the reduced plastid genome of *Lathraea clandestina*, an achlorophyllous parasitic plant. Physiologia Plantarum **96,** 674–682.

dePamphilis CW, Palmer JD. 1990. Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. Nature **348,** 337–339.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin **19,** 11–15.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution **29**, 1969–1973.

Fan W, Zhu A, Kozaczek M, Shah N, Pabón-Mora N, González F, Mower JP. 2016. Limited mitogenomic degradation in response to a parasitic lifestyle in Orobanchaceae. Scientific Report **6,** 36285.

Fu W, Liu X, Zhang N, Song Z, Zhang W, Yang J, Wang Y. 2017. Testing the hypothesis of multiple origins of holoparasitism in Orobanchaceae: phylogenetic evidence from the last two unplaced holoparasitic genera, *Gleadovia* and *Phacellanthus*. Frontiers in Plant Science **8,** 1380.

Funk HT, Berg S, Krupinska K, Maier UG, Krause K. 2007. Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, *Cuscuta reflexa* and *Cuscuta gronovii.* BMC Plant Biology **7,** 45.

Goremykin VV, Salamini F, Velasco R, Viola R. 2009. Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. Molecular Biology and Evolution **26,** 99–110.

Hahn C, Bachmann L, Chevreux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Research **41**, e129.

Jansen RK, Ruhlman TA. 2012. Plastid genomes of seed plants. In: Bock R, Knoop V, eds. Genomics of Chloroplast and Mitochondria. Netherlands: Springer, **35,** 103–126.

Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. 2001. REPuter: The manifold applications of repeat analysis on a genomic scale. Nucleic Acids Research **29,** 4633–4642.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome Biology **5,** R12.

Langmead B, Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods **9,** 357–359.

Li X, Zhang TC, Qiao Q, Ren Z, Zhao J, Yonezawa T, Hasegawa M, Crabbe MJC, Li J, Zhong Y. 2013. 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.

Liu YG, Whittier RF. 1995. Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from PI and YAC clones for chromosome walking. Genomics **25,** 674–681.

Liu YG, Huang N. 1998. Efficient amplification of insert end sequences from bacterial artificial chromosome clones by thermal asymmetric interlaced PCR. Plant Molecular Biology Reporter**16,** 175–181.

Lloyd AH, Rousseau-Gueutin M, Timmis JN, Sheppard AE, Ayliffe MA. 2012. Promiscuous organeller DNA. In: Bock R, Knoop V, eds. Genomics of chloroplasts and mitochondria. Netherlands: Springer, 201–221.

Logacheva MD, Schelkunov MI, Shtratnikova VY, Matveeva MV, Penin AA. 2016. Comparative analysis of plastid genomes of non-photosynthetic Ericaceae and their photosynthetic relatives. Scientific Report **6,** 30042.

Lohse M, Drechsel O, Kahlau S, Bock R. 2013. OrganellarGenome DRAW-a suite of tools for generating physical maps of plastid and mitochondrial genomes visualizing expression data sets. Nucleic Acids Research **41,** W575–W581.

Luo R, Liu B, Xie YL, *et al.* 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* **1,** 18.

Matlin AJ, Clark F, Smith CW. 2005. Understanding alternative splicing: towards a cellular code. Nature Reviews Molecular Cell Biology **6,** 1194–1200.

McNeal JR, Kuehl JV, Boore JL, dePamphilis CW. 2007. Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta.* BMC Plant Biology **7,** 57.

McNeal JR, Kuehl JV, Boore JL, Leebens-Mack J, dePamphilis CW. 2009. Parallel loss of plastid introns and their maturase in the genus *Cuscuta*. PLoS ONE **4**, e5982.

Molina J, Hazzouri KM, Nickrent D, *et al*. 2014. Possible loss of the chloroplast genome in the parasitic flowering plant *Rafflesia lagascae* (Rafflesiaceae). Molecular Biology and Evolution **31**, 793–803.

Mower JP, Stefanović S, Young GJ, Palmer JD. 2004. Gene transfer from parasitic to host plants. Nature **432**, 165–166.

Naumann J, Der JP, Wafula EK, *et al.* 2015. Detecting and characterizing the highly divergent plastid genome of the nonphotosynthetic parasitic plant *Hydnora visseri* (Hydnoraceae). Genome Biololgy and Evolution **8,** 345–363.

Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, Hirai A, Kadowaki K. 2002. The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. Molecular Genetics and Genomics **268**, 434–445.

Palmer JD. 1985. Comparative organization of chloroplast genomes. Annual Review of Genetics **19,** 325–354.

Rambaut A. 2006. FigTree. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh. Available online at: http://tree.bio.ed.ac.uk/software/ figtree/.

Rice DW, Alverson AJ, Richardson AO, *et al*. 2013. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. Science **342,** 1468–1473.

Roquet C, Coissac É, Cruaud C, *et al*. 2016. Understanding the evolution of holoparasitic plants: the complete plastid genome of the holoparasite *Cytinus hypocistis* (Cytinaceae). Annals of Botany **118**, 885–896.

Samigullin TH, Logacheva MD, Penin AA, Vallejo-Roman CM. 2016. Complete plastid genome of the recent holoparasite *Lathraea squamaria* reveals earliest stages of plastome reduction in Orobanchaceae. PLoS ONE **11,** e0150718.

Schelkunov MI, Shtratnikova VY, Nuraliev MS, Selosse MA, Penin AA, Logacheva MD. 2015. Exploring the limits for reduction of plastid genomes: A case study of the mycoheterotrophic orchids *Epipogium aphyllum* and *Epipogium roseum*. Genome Biology and Evolution **7,** 1179–1191.

Schneider AC, Chun H, Stefanović S, Baldwin BG. 2018. Punctuated plastome reduction and host-parasite horizontal gene transfer in the holoparasitic plant genus *Aphyllon*. Proceedings of the Royal Society B: Biological Sciences **285,** 20181535.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22,** 2688–2690.

Su HJ, Barkman TJ, Hao W, Jones SS, Naumann J, Skippington E, Wafula E, Hu JM, Palmer JD, dePamphilis CW. 2019. Novel genetic code and record-setting AT-richness in the highly reduced plastid genome of the holoparasitic plant *Balanophora*. Proceedings of the National Academy of Sciences of the United States of America **116,** 934-943.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution **28,** 2731–2739.

Westbury M, Baleka S, Barlow A, *et al.* 2017. A mitogenomic timetree for Darwin's enigmatic South American mammal *Macrauchenia patachonica*. Nature Communications **8**, 15951.

Westwood JH, Yoder JI, Timko MP, dePamphilis CW. 2010. The evolution of parasitism in plants. Trends in Plant Science **15,** 227–235.

Wicke S, Müller KF, dePamphilis CW, Quandt D, Wickett NJ, Zhang Y, Renner SS, Scheneeweiss GM. 2013. Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. Plant Cell **25,** 3711–3725.

Wicke S, Müller KF, dePamphilis CW, Quandt D, Bellot S, Schneeweiss GM. 2016. Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. Proceedings of the National Academy of Sciences of the United States of America **113,** 9045–9050.

Wolfe KH, Morden CW, Ems SC, Palmer JD. 1992a. Rapid evolution of the plastid translational apparatus in a nonphotosynthetic plant: Loss or accelerated sequence evolution of tRNA and ribosomal protein genes. Journal of Molecular Evolution **35,** 304–317.

Wolfe KH, Morden CW, Palmer JD. 1992b. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. Proceedings of the National Academy of Sciences of the United States of America **89,** 10648–10652.

Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics **20,** 3252–3255.

Xu, B. and Yang, Z. (2013) PAMLX: a graphical user interface for PAML. *Mol. Biol. Evol.*, **30**, 2723–2724.

Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate *Lysozyme* evolution. Molecular Biology and Evolution **15,** 568–573.

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution **24**, 1586–1591.

Yang Z, Nielsen R. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. Journal of Molecular Evolution **46,** 409–418.

Zanne AE, Tank DC, Cornwell WK, *et al*. 2014. Three keys to the radiation of angiosperms into freezing environments. Nature **506,** 89–92.

Zeng S, Zhou T, Han K, Yang Y, Zhao J, Liu ZL. 2017. The complete chloroplast genome sequences of six *Rehmannia* species. Genes **8,** 103.

Figure legends

Fig. 1. Physical maps of the plastid genomes of five *Cistanche* species. All genes were colored according to functional complexes. Functional genes and structural RNAs were shown in solid blocks. Pseudogenes were indicated by ψ, shown in dashed blocks.

Fig. 2. Comparison of boundary positions between single copy (large, LSC or small, SSC) and inverted repeat (IR) regions among the plastomes of five *Cistanche* species and other Orobanchaceae species. The location of inverted repeat region (IRa and IRb) was referred to Fig. 1. Numbers in red are obviously different from those of other species in the same genus. Maximum likelihood tree is analyzed based on 32 housekeeping genes available in these Orobanchaceae species. The expanded phylogenetic tree with bootstrap support values is shown in Supplementary Fig. S2.

Fig. 3. Nucleotide substitution rates and selectional regimes in *Cistanche* and their relatives. The proportion of sites under purifying selection (ω < 1), neutral evolution (ω = 1, but not significant), positive selection are shown by different colors (blue, grey and red); red branch means ω > 1 (*p*-value < 5e-02), but no further discussion in this study. Different mean ω values according to LRT test are illustrated as branch width, withthick(er) branches indicating a higher mean ω. Branch length reflects the number of substitution per site.

Fig. 4. dN/dS between the species of *Cistanche* and *Lindenbergia philippensis*. Whiskers show standard errors estimated by PAML.

Fig. 5. Gene maps (**A**) and phylogenetic relationships (**B**) of three mipts (*pet*A, *pet*G and *rpo*C1) and one nupt (*acc*D). The gene maps show the genome location of the transgenes. The orange box represents the mitochondrial sequence, indicating that the transgene is located in the mitochondrion genome; the green box represents the plastid-origin transgenes; and the red box represents the nuclei

sequence, indicating that the transgene is located in the nucleus. The transfer length of *pet*A, *pet*G, *rpo*C1, and *acc*D is 539 bp, 156 bp, 525 bp, and 279 bp, respectively.

Fig. 6. Gene content and annotation of plastomes of 25 Orobancheae species. The matrix shows the physically lost gene (black), pseudogenized gene (gray) or putatively functional gene with intact ORFs (white). An asterisk (*) indicates the gene was pseudogenized and putatively functional in different IR region, respectively.

Fig. 7. Inferred gene losses and pseudogenization in *Cistanche* and its relatives. In the maximum likelihood tree, the unambiguous gene loss and pseudogenization are shown below and above branches, respectively. Different colors represent different types of genes. Branch lengths of the tree are proportional to the maximum number of the lost genes or the pseudogenes, whose names are given along the branches. Triangles at the tip of each terminal branch simplifies the internal structure of three genera (*Aphyllon*, *Phelipanche* and *Orobanche*).

Fig. 8. Clade-specific degeneration pattern of petG gene in Cistanche and related species. The box above the branches indicates different gene status. Gene possess intact ORF is white, pseudogene is gray and lost gene is black.

Fig. 9. The simplified clade-specific degeneration patterns of protein-coding genes in the species of *Cistanche* and *Aphyllon*. Three different degeneration patterns were found in *Cistanche* clade based on phylogenetic trees. The expended trees of different genes are presented in Fig. 8 and Supplementary Fig. S4. Among these genes, *psb*D and *trn*V-UAC show the *C. deserticola-C. salsa* clade specific pattern; *ndh*E, *psb*A, *ycf*1, *ycf*2 and *trn*T-UGU show the *C. sinensis* clade specific degeneration/existence pattern; *atp*B shows the *C. phelypaea-C. tubulosa* clade-specific degeneration pattern. In *Aphyllon, atp*F, *rpo*A, *rps*12 and *trn*G-GCC show the *A. californicum* clade-specific degeneration/existence pattern; *psb*I shows the

A. epigalium-*A. franciscanum* clade-specific degeneration pattern; *psb*M shows diverse gene status in different branches. Black, gray and white box indicates the lost gene, pseudogene and putatively functional gene with intact ORFs, respectively.

Table 1. *Characteristics of five Cistanche plastomes*

	C. sinensis	*C. phelypaea*	*C. tubulosa*	*C. salsa*	*C. deserticola*
Total size (bp)	87,707	94,380	94,123	101,776	102,657
LSC size (bp)	26,435	31,196	31,017	46,579	48,350
SSC size (bp)	11,865	8,019	8,547	8,468	8,800
IR length (bp)	49,407	55,165	54,559	46,729	45,507
Size of coding regions (bp)	50,648	48,019	48,957	52,293	53,259
Size of protein-coding regions (bp)	15,347	28,438	29,538	22,806	29,303
Size of rRNA (bp)	9,049	9,048	9,048	9,940	9,044
Size of tRNA (bp)	4,884	5,095	3,999	7,763	7,185
Size in intergenic regions (bp)	23,517	32,228	32,249	31,931	33,166
No. of different genes	68	77	72	89	89
No. of different pseudogenes	22	19	20	35	28
No. of different protein-coding genes	23	24	22	20	27
No. of different tRNA genes	22	29	26	30	30
No. of different rRNA genes	4	4	4	4	4
No. of different genes duplicated by IR	20	21	22	18	18
Overall GC content (%)	37.95	36.56	36.53	37.26	36.78
GC content in protein-coding regions (%)	33.67	36.05	34.49	36.51	36.70
GC content in IGSs (%)	34.02	31.20	31.60	31.45	31.10
GC content in rRNA (%)	54.64	54.89	54.87	54.90	54.97
GC content in tRNA (%)	51.37	49.07	48.99	47.89	48.20

Table 2. *Statistics of the degenerate and putatively functional plastid genes in the five Cistanche species*

Gene Classes	Gene IDs
Putatively functional gene with intact ORFs in the five *Cistanche* species	*mat*K, *inf*A, *rpl*14, *rpl*20, *rpl*33, *rps*11, *rps*18, *rps*2/4/7/8
Structural RNAs (transfer and ribosomal RNAs) in the five *Cistanche* species	*trn*A-UGC, *trn*C-GCA, *trn*D-GUC, *trn*E-UUC, *trn*fM-CAU, *trn*M-CAU, *trn*G-UCC, *trn*I-CAU, *trn*I-GAU, *trn*K-UUU, *trn*L-CAA, *trn*L-UAG, *trn*N-GUU, *trn*P-UGG, *trn*R-ACG, *trn*S-GCU, *trn*S-GGA, *trn*V-GAC, *trn*W-CCA, *trn*Y-GUA, *rrn*16/23/4.5/5
Lost genes in the five *Cistanche* species	*atp*H/I, *cem*A*, ndh*A/C/D/F/G/I/J/K, *pet*A/B/D/N, *psa*C, *psb*H/N/T, *rpo*C1
Pseudogenized genes in the five *Cistanche* species	*ndh*B/H, *psa*A/B, *rbcL*, *rpoB*, *rpoC2*, *rpl23*
Lost genes in *C. sinensis* which exist in the other *Cistanche* species	*rpl*32, *trn*F-GAA, *trn*H-GUG, *trn*Q-UUG, *trn*S-UGA, *trn*T-UGU
Pseudogenized genes in C. sinensis but lost in the other *Cistanche* species	*ndh*E*, pet*L
Lost genes in *C. sinensis* but pseudogenized in the other *Cistanche* species	*psb*A/E/J/K, *ycf*4
Putatively functional gene in *C. sinensis* but pseudogenized in the other *Cistanche* species	*ycf*1
Pseudogenized genes in *C. sinensis* but exist with putative function in the other *Cistanche* species	*rps*3*, ycf*2
Lost genes in the *C. tubulosa-C. phelypaea* clade but pseudogenized in the other *Cistanche* species	*atp*A/B/E/F
Pseudogenized genes in the *C. deserticola-C. salsa* clade but lost in the other *Cistanche* species	*psa*I, *psb*B/D/ L*, ycf*3
Gene exists in the *C. deserticola-C. salsa* clade but lost in the other *Cistanche* species	*trn*V-UAC
Putatively functional gene exists in *C. sinensis* but lost in the *C. tubulosa-C. phelypaea* clade and pseudogenized in the *C. deserticola-C. salsa* clade	*pet*G

Table 3. *Statistics of mipts and nupts in the species of Orobanche and Cistanche*

	O_ aus	*O_cre*	*O_den*	*O_gra*	*O_rap*	*C_sin*	*C_tub*	*C_sal*
*atp*A	-	n	-	n	-	-	n	-
*atp*B	-	n	-	n	m	-	m	-
*atp*E	-	-	-	n	m, n	-	n	-
*atp*F	-	n	-	n	m	-	n	-
*atp*H	-	n	n	n	n	n	-	n
*atp*I	n	n	-	n	n	m	m	n
*psb*A	m, n	n	n	n	n	-	-	n
*psb*B	-	-	-	-	-	n	-	n
*psb*C	n	n	n	n	n	m	n	m
*psb*D	n	n	n	n	n	-	m	n
*psb*E	-	n	-	n	n	-	-	-
*psb*F	-	-	-	-	-	n	n	-
*psb*K	-	-	-	-	n	-	-	-
*psb*Z	n	n	-	n	n	m	-	-
*pet*A	-	n	-	-	n	-	m	m
*pet*G	-	-	-	-	-	-	m	m
*pet*N	n	-	-	-	-	-	-	-
*psa*A	-	n	n	n	-	-	-	-
*psa*B	m, n	n	m	n	n	-	-	-
*psa*C	-	-	-	n	-	-	-	-
*ndh*A	-	-	-	n	-	-	-	-
*ndh*B	-	n	-	n	-	-	-	-
*ndh*C	-	-	-	n	-	-	-	-
*ndh*G	-	-	-	n	-	-	-	n
*ndh*H	n	n	n	-	-	-	-	-
*ndh*I	-	n	-	-	-	-	-	-
*ndh*J	n	n	n	n	-	m	-	-
*ndh*K	-	-	-	n	-	-	-	-
*rpo*A	-	n	-	n	-	-	n	-
*rpo*A	-	n	-	n	n	-	-	-
*rpo*C1	-	n	n	n	n	m	m	-
*rpo*C2	-	n	-	n	-	-	-	m
*rpl*23	n	n	n	n	n	-	-	-
*rpl*32	-	-	-	-	-	n	-	-
*ycf*3	m	n	n	n	-	-	n	-
*ycf*4	-	n	n	n	n	-	-	-
*cem*A	-	-	-	n	-	-	-	-
*ccs*A	-	-	-	n	-	-	-	-
*acc*D	-	-	-	-	-	-	n	-
*rbc*L	-	-	n	n	-	-	m	-

"m" and "n" represent mipt and nupt, respectively. The symbol "-" means that neither mipt nor nupt is found in genomic data. The abbreviations of species name are as follows: *O_aus*, *Oroanche austrohispanica*; *O_cre*, *Orobanche crenata*; *O_den*, *Orobanche densiflora*; *O_gra*, *Orobanche gracilis*; *O_rap*, *Orobanche rapum-genistae*; *C_sin*, *Cistanche sinensis*; *C_ tub*, *Cistanche tubulosa*; *C_sal*, *Cistanche salsa*.

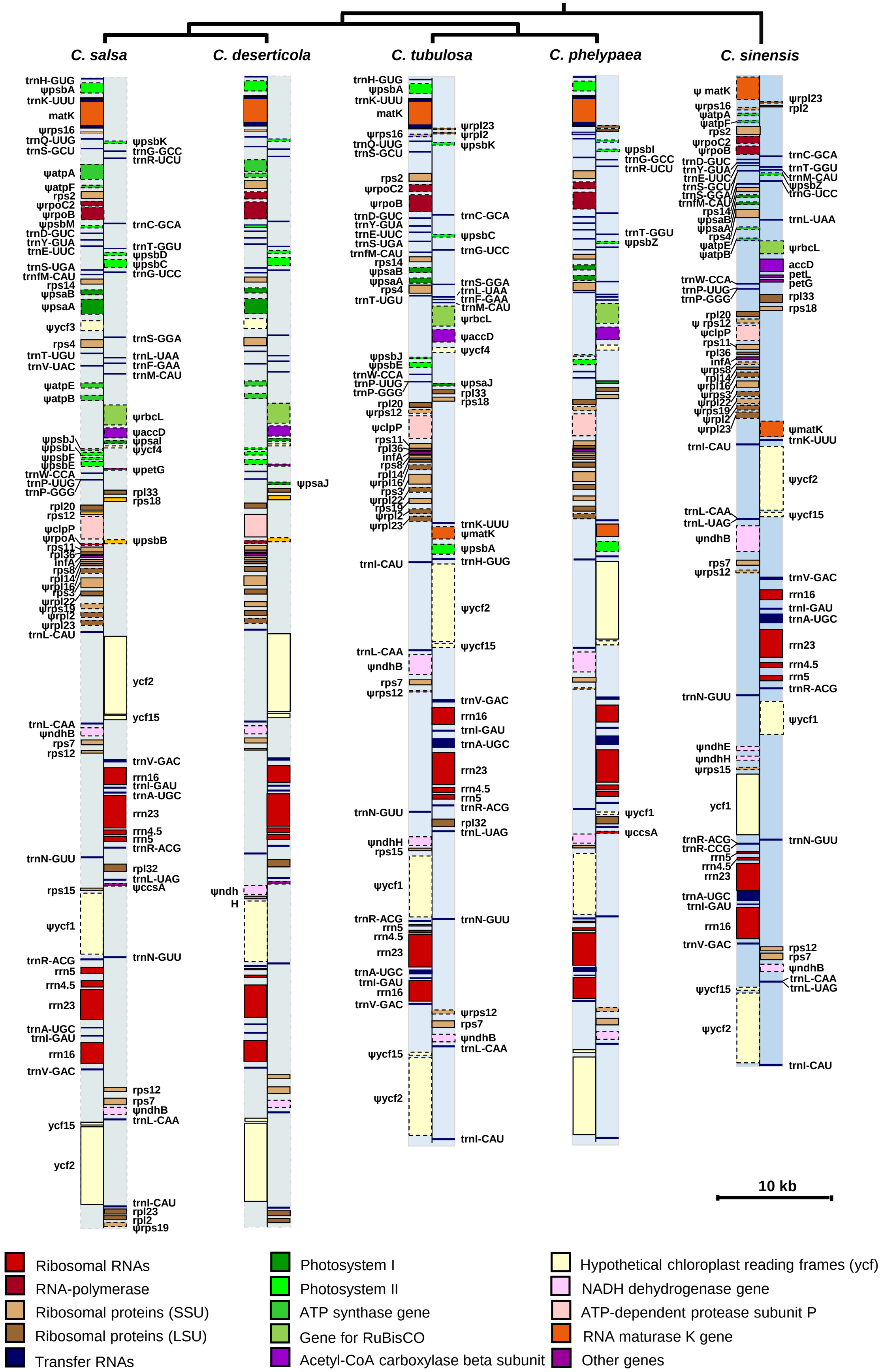

C. salsa
C. deserticola
C. tubulosa
C. phelypaea
C. sinensis
trnH-GUG
ψpsbA
trnK-UUU
matK
ψrps16
trnQ-UUG
trnS-GCU
ψatpA
ψatpF
rps2
ψrpoC2
ψrpoB
ψpsbM
trnD-GUC
trnY-GUA
trnE-UUC
trnS-UGA
trnfM-CAU
rps14
ψpsaB
ψpsaA
ψycf3
rps4
trnT-UGU
trnV-UAC
ψatpE
ψatpB
ψpsbJ
ψpsbL
ψpsbF
ψpsbE
trnW-CCA
trnP-UGG
trnP-GGG
rpl20
rps12
ψclpP
ψrpoA
rps11
rpl36
infA
rps8
rpl14
ψrpl16
rps3
ψrpl22
ψrps19
ψrpl2
ψrpl23
trnL-CAU
trnL-CAA
ψndhB
rps7
rps12
trnN-GUU
rps15
ψycf1
trnR-ACG
rrn5
rrn4.5
rrn23
trnA-UGC
trnI-GAU
rrn16
trnV-GAC
ycf15
ycf2
ψpsbK
trnG-GCC
trnR-UCU
trnC-GCA
trnT-GGU
ψpsbD
ψpsbC
trnG-UCC
trnS-GGA
trnL-UAA
trnF-GAA
trnM-CAU
ψrbcL
ψaccD
ψpsaI
ψycf4
ψpetG
rpl33
rps18
ψpsbB
ycf2
ycf15
trnV-GAC
rrn16
trnI-GAU
trnA-UGC
rrn23
rrn4.5
rrn5
trnR-ACG
rpl32
trnL-UAG
ψccsA
trnN-GUU
rps12
rps7
ψndhB
trnL-CAA
trnI-CAU
rpl23
rpl2
ψrps19
ψpsaJ
ψndh
H
trnH-GUG
ψpsbA
trnK-UUU
matK
ψrps16
trnQ-UUG
trnS-GCU
rps2
ψrpoC2
ψrpoB
trnD-GUC
trnY-GUA
trnE-UUC
trnS-UGA
trnfM-CAU
rps14
ψpsaB
ψpsaA
rps4
trnT-UGU
ψpsbJ
ψpsbE
trnW-CCA
trnP-UUG
trnP-GGG
rpl20
ψrps12
ψclpP
rps11
rpl36
infA
rps8
rpl14
ψrpl16
rps3
ψrpl22
ψrps19
ψrpl2
ψrpl23
trnI-CAU
trnL-CAA
ψndhB
rps7
ψrps12
trnN-GUU
ψndhH
rps15
ψycf1
trnR-ACG
rrn5
rrn4.5
rrn23
trnA-UGC
trnI-GAU
rrn16
trnV-GAC
ψycf15
ψycf2
ψrpl23
ψrpl2
ψpsbK
trnC-GCA
ψpsbC
trnG-UCC
trnS-GGA
trnL-UAA
trnF-GAA
trnM-CAU
ψrbcL
ψaccD
ψycf4
ψpsaJ
rpl33
rps18
trnK-UUU
ψmatK
ψpsbA
trnH-GUG
ψycf2
ψycf15
trnV-GAC
rrn16
trnI-GAU
trnA-UGC
rrn23
rrn4.5
rrn5
trnR-ACG
rpl32
trnL-UAG
trnN-GUU
ψrps12
rps7
ψndhB
trnL-CAA
trnI-CAU
ψpsbI
trnG-GCC
trnR-UCU
trnT-GGU
ψpsbZ
ψycf1
ψccsA
ψ matK
ψrps16
ψatpA
ψatpF
rps2
ψrpoC2
ψrpoB
trnD-GUC
trnY-GUA
trnE-UUC
trnS-GCU
trnS-GGA
trnfM-CAU
rps14
ψpsaB
ψpsaA
rps4
ψatpE
ψatpB
trnW-CCA
trnP-UUG
trnP-GGG
rpl20
ψ rps12
ψclpP
rps11
rpl36
infA
ψrps8
rpl14
ψrpl16
ψrps3
ψrpl22
ψrps19
ψrpl2
ψrpl23
trnI-CAU
trnL-CAA
trnL-UAG
ψndhB
rps7
ψrps12
trnN-GUU
ψndhE
ψndhH
ψrps15
ycf1
trnR-ACG
trnR-CCG
rrn5
rrn4.5
rrn23
trnA-UGC
trnI-GAU
rrn16
trnV-GAC
ψycf15
ψycf2
ψrpl23
rpl2
trnC-GCA
trnT-GGU
trnM-CAU
ψpsbZ
trnG-UCC
trnL-UAA
ψrbcL
accD
petL
petG
rpl33
rps18
ψmatK
trnK-UUU
ψycf2
ψycf15
trnV-GAC
rrn16
trnI-GAU
trnA-UGC
rrn23
rrn4.5
rrn5
trnR-ACG
ψycf1
trnN-GUU
rps12
rps7
ψndhB
trnL-CAA
trnL-UAG
trnI-CAU
10 kb
Ribosomal RNAs
RNA-polymerase
Ribosomal proteins (SSU)
Ribosomal proteins (LSU)
Transfer RNAs
Photosystem I
Photosystem II
ATP synthase gene
Gene for RuBisCO
Acetyl-CoA carboxylase beta subunit
Hypothetical chloroplast reading frames (ycf)
NADH dehydrogenase gene
ATP-dependent protease subunit P
RNA maturase K gene
Other genes

Fig. 1

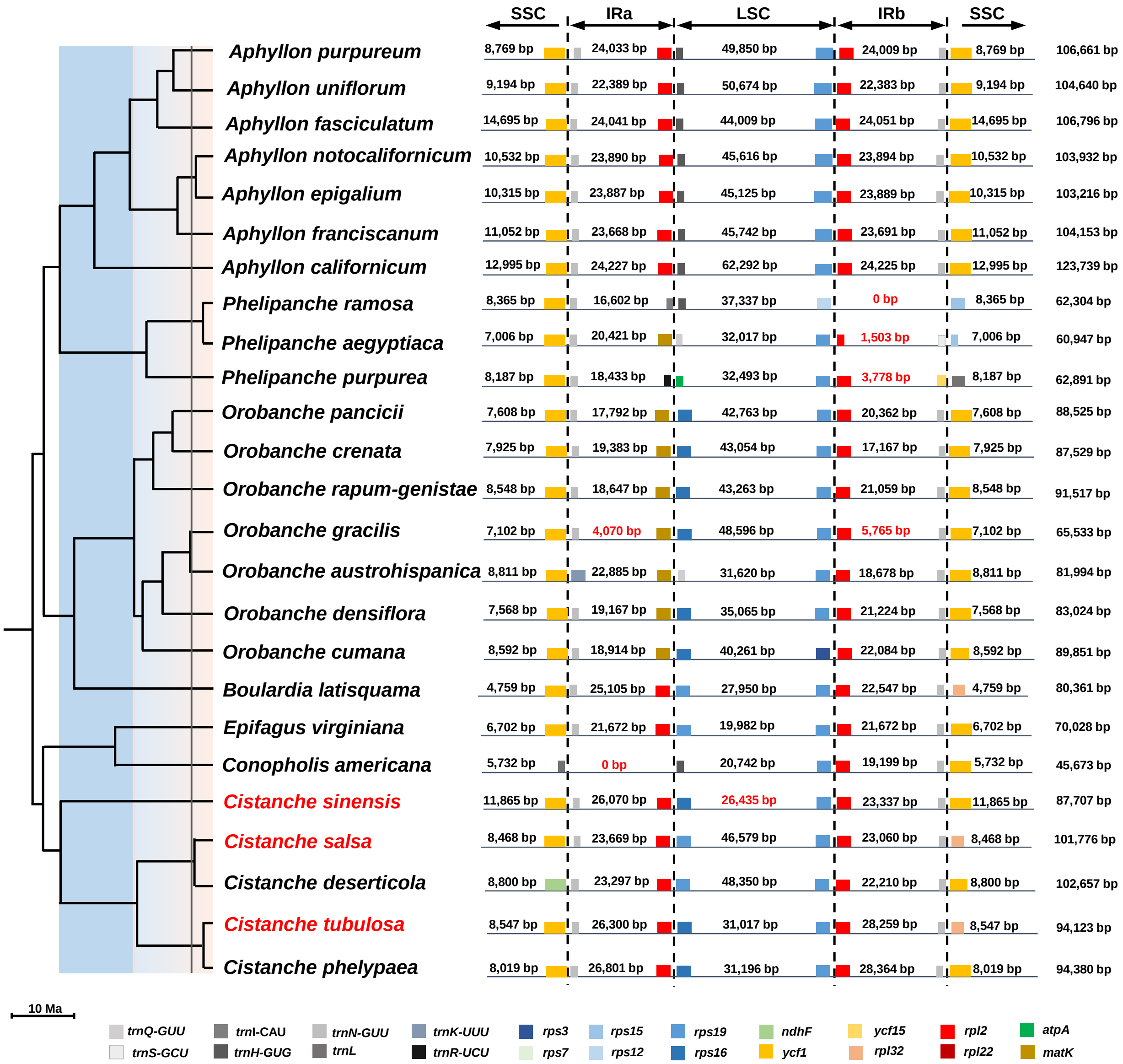

SSC IRa LSC IRb SSC
Aphyllon purpureum 8,769 bp 24,033 bp 49,850 bp 24,009 bp 8,769 bp 106,661 bp
Aphyllon uniflorum 9,194 bp 22,389 bp 50,674 bp 22,383 bp 9,194 bp 104,640 bp
Aphyllon fasciculatum 14,695 bp 24,041 bp 44,009 bp 24,051 bp 14,695 bp 106,796 bp
Aphyllon notocalifornicum 10,532 bp 23,890 bp 45,616 bp 23,894 bp 10,532 bp 103,932 bp
Aphyllon epigalium 10,315 bp 23,887 bp 45,125 bp 23,889 bp 10,315 bp 103,216 bp
Aphyllon franciscanum 11,052 bp 23,668 bp 45,742 bp 23,691 bp 11,052 bp 104,153 bp
Aphyllon californicum 12,995 bp 24,227 bp 62,292 bp 24,225 bp 12,995 bp 123,739 bp
Phelipanche ramosa 8,365 bp 16,602 bp 37,337 bp 0 bp 8,365 bp 62,304 bp
Phelipanche aegyptiaca 7,006 bp 20,421 bp 32,017 bp 1,503 bp 7,006 bp 60,947 bp
Phelipanche purpurea 8,187 bp 18,433 bp 32,493 bp 3,778 bp 8,187 bp 62,891 bp
Orobanche pancicii 7,608 bp 17,792 bp 42,763 bp 20,362 bp 7,608 bp 88,525 bp
Orobanche crenata 7,925 bp 19,383 bp 43,054 bp 17,167 bp 7,925 bp 87,529 bp
Orobanche rapum-genistae 8,548 bp 18,647 bp 43,263 bp 21,059 bp 8,548 bp 91,517 bp
Orobanche gracilis 7,102 bp 4,070 bp 48,596 bp 5,765 bp 7,102 bp 65,533 bp
Orobanche austrohispanica 8,811 bp 22,885 bp 31,620 bp 18,678 bp 8,811 bp 81,994 bp
Orobanche densiflora 7,568 bp 19,167 bp 35,065 bp 21,224 bp 7,568 bp 83,024 bp
Orobanche cumana 8,592 bp 18,914 bp 40,261 bp 22,084 bp 8,592 bp 89,851 bp
Boulardia latisquama 4,759 bp 25,105 bp 27,950 bp 22,547 bp 4,759 bp 80,361 bp
Epifagus virginiana 6,702 bp 21,672 bp 19,982 bp 21,672 bp 6,702 bp 70,028 bp
Conopholis americana 5,732 bp 0 bp 20,742 bp 19,199 bp 5,732 bp 45,673 bp
Cistanche sinensis 11,865 bp 26,070 bp 26,435 bp 23,337 bp 11,865 bp 87,707 bp
Cistanche salsa 8,468 bp 23,669 bp 46,579 bp 23,060 bp 8,468 bp 101,776 bp
Cistanche deserticola 8,800 bp 23,297 bp 48,350 bp 22,210 bp 8,800 bp 102,657 bp
Cistanche tubulosa 8,547 bp 26,300 bp 31,017 bp 28,259 bp 8,547 bp 94,123 bp
Cistanche phelypaea 8,019 bp 26,801 bp 31,196 bp 28,364 bp 8,019 bp 94,380 bp
10 Ma
trnQ-GUU
trnS-GCU
trnI-CAU
trnH-GUG
trnN-GUU
trnL
trnK-UUU
trnR-UCU
rps3
rps7
rps15
rps12
rps19
rps16
ndhF
ycf1
ycf15
rpl32
rpl2
rpl22
atpA
matK

Fig. 2

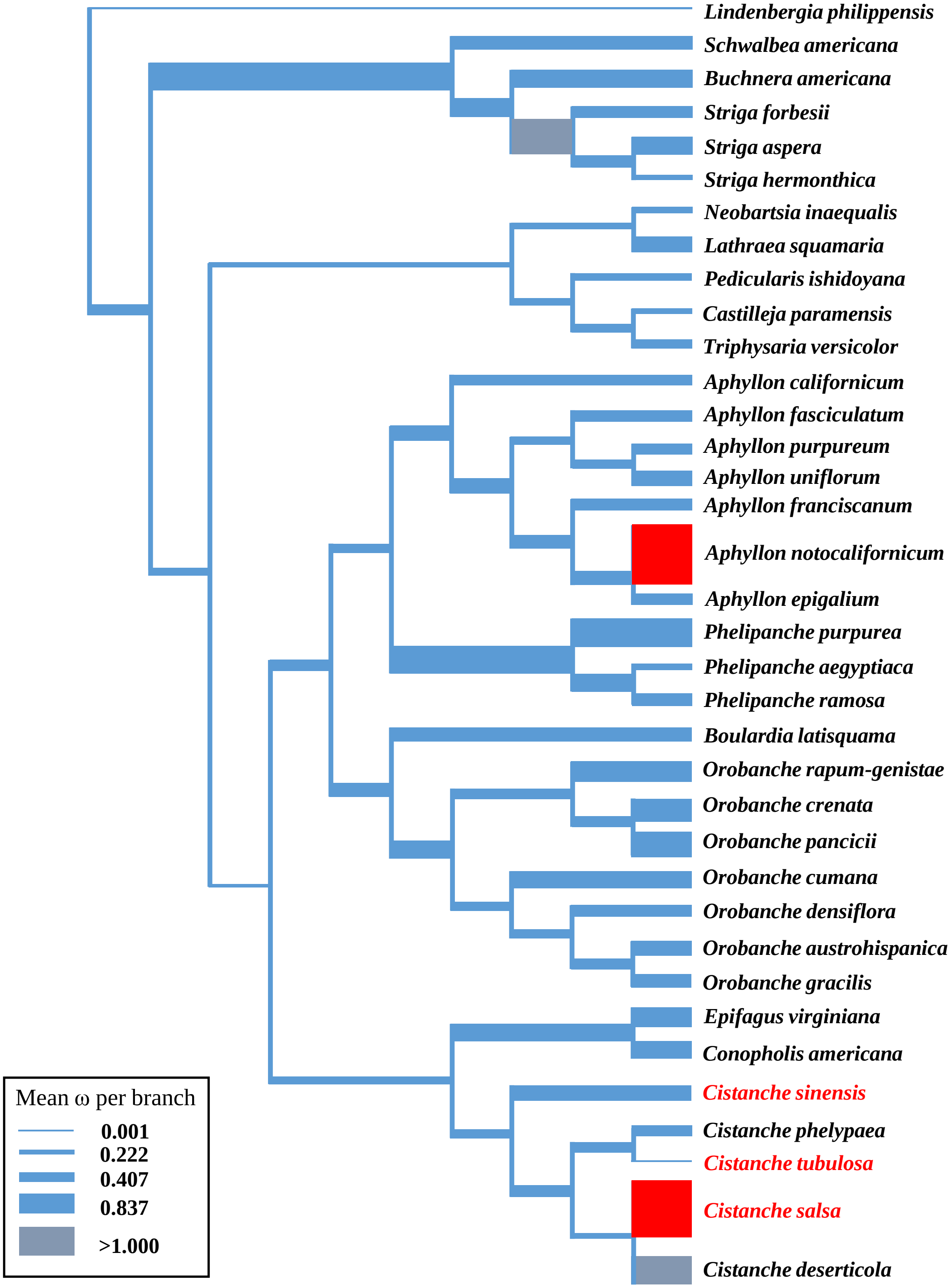
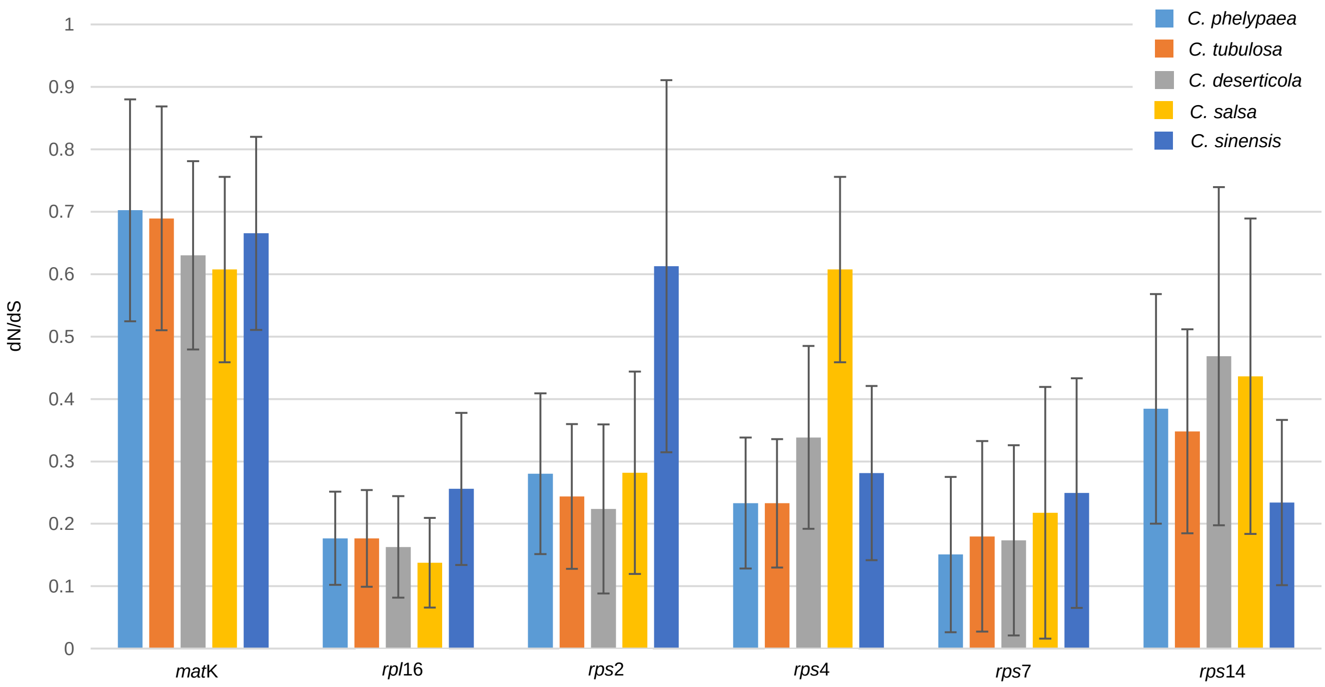

Lindenbergia philippensis
Schwalbea americana
Buchnera americana
Striga forbesii
Striga aspera
Striga hermonthica
Neobartsia inaequalis
Lathraea squamaria
Pedicularis ishidoyana
Castilleja paramensis
Triphysaria versicolor
Aphyllon californicum
Aphyllon fasciculatum
Aphyllon purpureum
Aphyllon uniflorum
Aphyllon franciscanum
Aphyllon notocalifornicum
Aphyllon epigalium
Phelipanche purpurea
Phelipanche aegyptiaca
Phelipanche ramosa
Boulardia latisquama
Orobanche rapum-genistae
Orobanche crenata
Orobanche pancicii
Orobanche cumana
Orobanche densiflora
Orobanche austrohispanica
Orobanche gracilis
Epifagus virginiana
Conopholis americana
Cistanche sinensis
Cistanche phelypaea
Cistanche tubulosa
Cistanche salsa
Cistanche deserticola
Mean ω per branch
0.001
0.222
0.407
0.837
>1.000

Fig. 3



1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
dN/dS
*mat*K
*rpl*16
*rps*2
*rps*4
*rps*7
*rps*14
C. phelypaea
C. tubulosa
C. deserticola
C. salsa
C. sinensis

Fig. 4

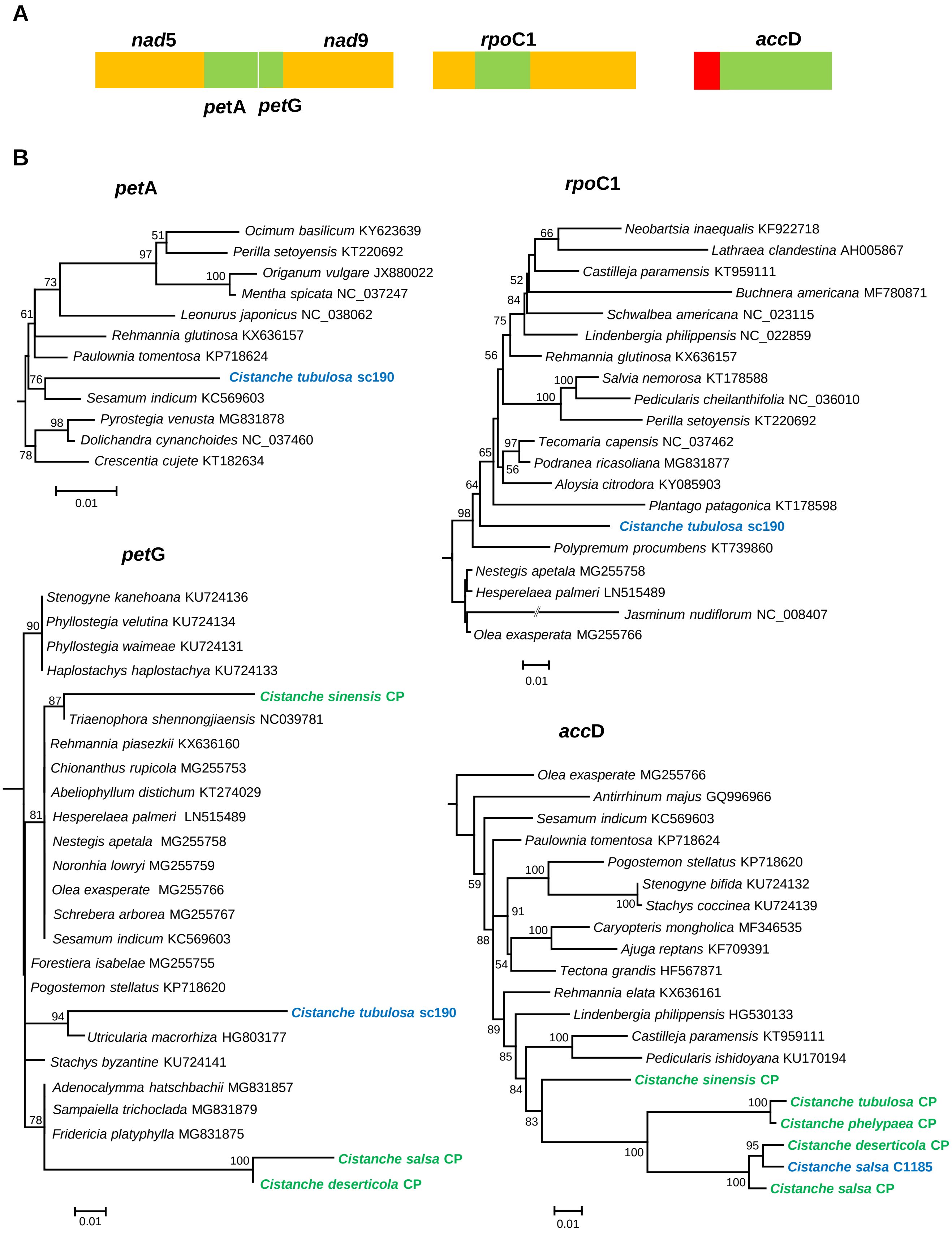

A
nad5
nad9
petA
petG
rpoC1
accD
B
petA
Ocimum basilicum KY623639
Perilla setoyensis KT220692
Origanum vulgare JX880022
Mentha spicata NC_037247
Leonurus japonicus NC_038062
Rehmannia glutinosa KX636157
Paulownia tomentosa KP718624
Cistanche tubulosa sc190
Sesamum indicum KC569603
Pyrostegia venusta MG831878
Dolichandra cynanchoides NC_037460
Crescentia cujete KT182634
51
97
100
73
61
76
98
78
0.01
rpoC1
Neobartsia inaequalis KF922718
Lathraea clandestina AH005867
Castilleja paramensis KT959111
Buchnera americana MF780871
Schwalbea americana NC_023115
Lindenbergia philippensis NC_022859
Rehmannia glutinosa KX636157
Salvia nemorosa KT178588
Pedicularis cheilanthifolia NC_036010
Perilla setoyensis KT220692
Tecomaria capensis NC_037462
Podranea ricasoliana MG831877
Aloysia citrodora KY085903
Plantago patagonica KT178598
Cistanche tubulosa sc190
Polypremum procumbens KT739860
Nestegis apetala MG255758
Hesperelaea palmeri LN515489
Jasminum nudiflorum NC_008407
Olea exasperata MG255766
66
52
84
75
56
100
100
97
56
65
64
98
0.01
petG
Stenogyne kanehoana KU724136
Phyllostegia velutina KU724134
Phyllostegia waimeae KU724131
Haplostachys haplostachya KU724133
Cistanche sinensis CP
Triaenophora shennongjiaensis NC039781
Rehmannia piasezkii KX636160
Chionanthus rupicola MG255753
Abeliophyllum distichum KT274029
Hesperelaea palmeri LN515489
Nestegis apetala MG255758
Noronhia lowryi MG255759
Olea exasperate MG255766
Schrebera arborea MG255767
Sesamum indicum KC569603
Forestiera isabelae MG255755
Pogostemon stellatus KP718620
Cistanche tubulosa sc190
Utricularia macrorhiza HG803177
Stachys byzantine KU724141
Adenocalymma hatschbachii MG831857
Sampaiella trichoclada MG831879
Fridericia platyphylla MG831875
Cistanche salsa CP
Cistanche deserticola CP
90
87
81
94
78
100
0.01
accD
Olea exasperate MG255766
Antirrhinum majus GQ996966
Sesamum indicum KC569603
Paulownia tomentosa KP718624
Pogostemon stellatus KP718620
Stenogyne bifida KU724132
Stachys coccinea KU724139
Caryopteris mongholica MF346535
Ajuga reptans KF709391
Tectona grandis HF567871
Rehmannia elata KX636161
Lindenbergia philippensis HG530133
Castilleja paramensis KT959111
Pedicularis ishidoyana KU170194
Cistanche sinensis CP
Cistanche tubulosa CP
Cistanche phelypaea CP
Cistanche deserticola CP
Cistanche salsa C1185
Cistanche salsa CP
100
100
59
91
100
88
54
89
100
85
84
83
100
100
95
100
0.01

Fig. 5

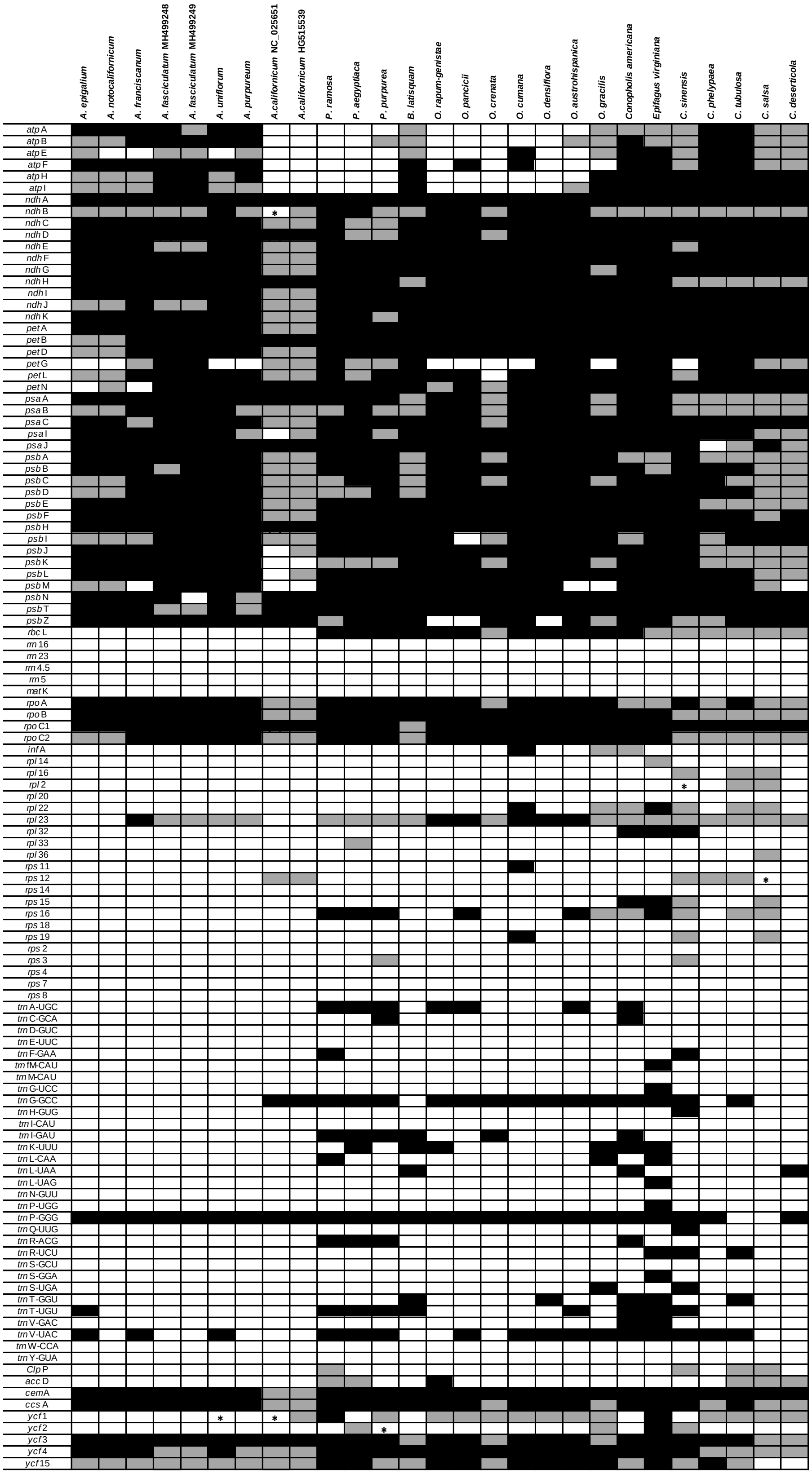

A. epigalium
A. notocalifornicum
A. franciscanum
A. fasciculatum MH499248
A. fasciculatum MH499249
A. uniflorum
A. purpureum
A.californicum NC_025651
A.californicum HG515539
P. ramosa
P. aegyptiaca
P. purpurea
B. latisquam
O. rapum-genistae
O. pancicii
O. crenata
O. cumana
O. densiflora
O. austrohispanica
O. gracilis
Conopholis americana
Epifagus virginiana
C. sinensis
C. phelypaea
C. tubulosa
C. salsa
C. deserticola
atp A
atp B
atp E
atp F
atp H
atp I
ndh A
ndh B
ndh C
ndh D
ndh E
ndh F
ndh G
ndh H
ndh I
ndh J
ndh K
pet A
pet B
pet D
pet G
pet L
pet N
psa A
psa B
psa C
psa I
psa J
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
rbc L
rrn 16
rrn 23
rrn 4.5
rrn 5
mat K
rpo A
rpo B
rpo C1
rpo C2
inf A
rpl 14
rpl 16
rpl 2
rpl 20
rpl 22
rpl 23
rpl 32
rpl 33
rpl 36
rps 11
rps 12
rps 14
rps 15
rps 16
rps 18
rps 19
rps 2
rps 3
rps 4
rps 7
rps 8
trn A-UGC
trn C-GCA
trn D-GUC
trn E-UUC
trn F-GAA
trn fM-CAU
trn M-CAU
trn G-UCC
trn G-GCC
trn H-GUG
trn I-CAU
trn I-GAU
trn K-UUU
trn L-CAA
trn L-UAA
trn L-UAG
trn N-GUU
trn P-UGG
trn P-GGG
trn Q-UUG
trn R-ACG
trn R-UCU
trn S-GCU
trn S-GGA
trn S-UGA
trn T-GGU
trn T-UGU
trn V-GAC
trn V-UAC
trn W-CCA
trn Y-GUA
Clp P
acc D
cem A
ccs A
ycf 1
ycf 2
ycf 3
ycf 4
ycf 15
*

Fig. 6

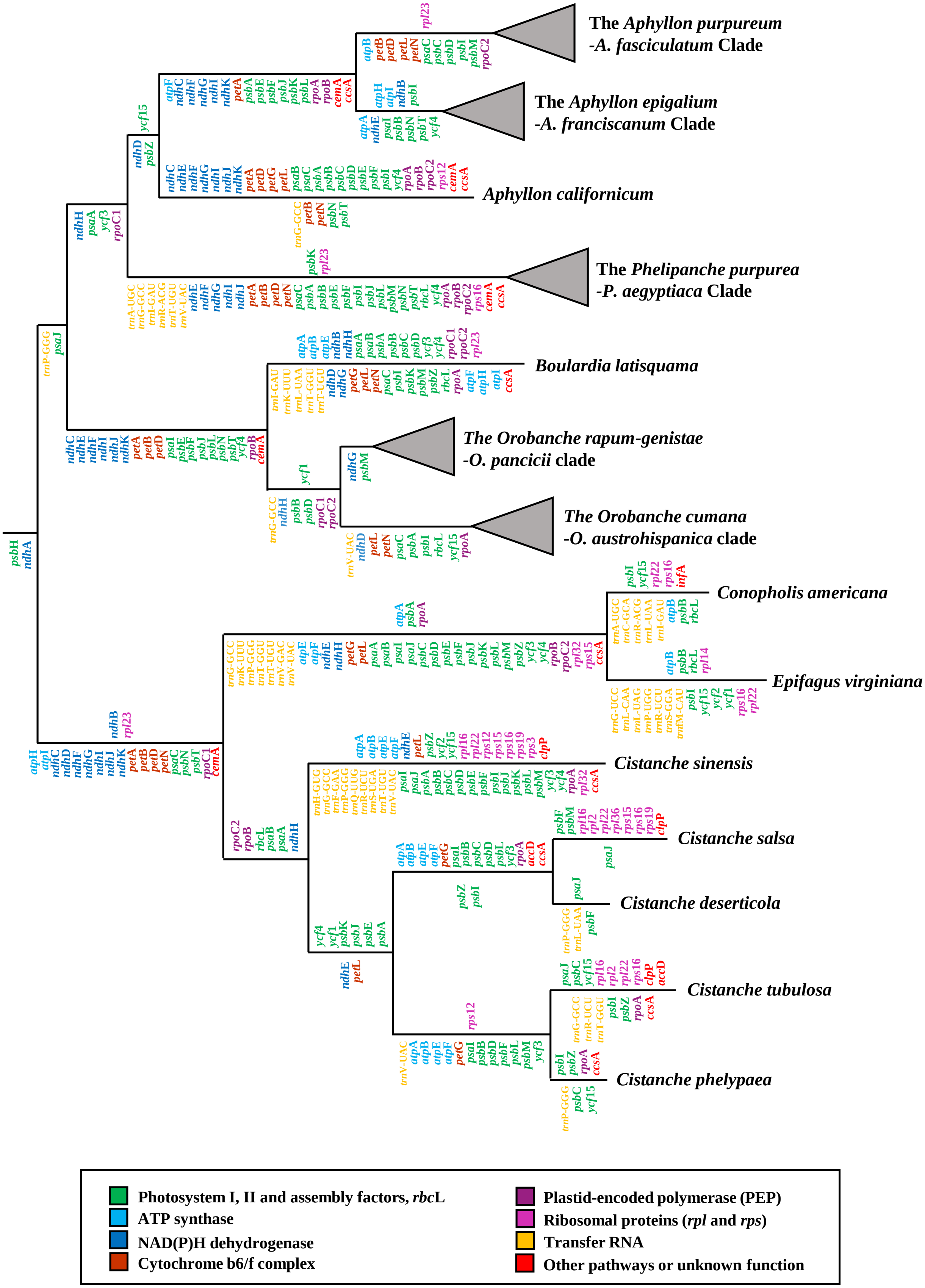

The *Aphyllon purpureum* -*A. fasciculatum* Clade
The *Aphyllon epigalium* -*A. franciscanum* Clade
Aphyllon californicum
The *Phelipanche purpurea* -*P. aegyptiaca* Clade
Boulardia latisquama
The Orobanche rapum-genistae -*O. pancicii* clade
The Orobanche cumana -*O. austrohispanica* clade
Conopholis americana
Epifagus virginiana
Cistanche sinensis
Cistanche salsa
Cistanche deserticola
Cistanche tubulosa
Cistanche phelypaea
psbH ndhA
trnP-GGG psaJ
ndhH psaA ycf3 rpoC1
ndhD psbZ
ycf15
atpF ndhC ndhF ndhG ndhI ndhJ ndhK petA psbA psbE psbF psbJ psbK psbL rpoA rpoB cemA ccsA
rpl23
atpB petB petD petL petN psaC psbC psbD psbI psbM rpoC2
atpH atpI ndhB psbI
atpA ndhE psaI psbB psbN psbT ycf4
ndhC ndhE ndhF ndhG ndhI ndhJ ndhK petA petD petG petL psaB psaC psbA psbB psbC psbD psbE psbF psbI ycf4 rpoA rpoB rpoC2 rps12 cemA ccsA
trnG-GCC petB petN psbN psbT
psbK rpl23
trnA-UGC trnG-GCC trnI-GAU trnR-ACG trnT-UGU trnV-UAC ndhE ndhF ndhG ndhI ndhJ petA petB petD petN psaC psaA psbB psbE psbF psbI psbJ psbL psbM psbN psbT rbcL ycf4 rpoA rpoB rpoC2 rps16 cemA ccsA
atpA atpB atpE ndhB ndhH psaA psaB psbA psbB psbC psbD ycf3 ycf4 rpoC1 rpoC2 rpl23
trnI-GAU trnK-UUU trnL-UAA trnT-GGU trnT-UGU ndhD ndhG petG petL petN psaC psbI psbK psbM psbZ rbcL rpoA atpF atpH atpI ccsA
ndhC ndhE ndhF ndhI ndhJ ndhK petA petB petD psaI psbE psbF psbJ psbL psbN psbT ycf4 rpoB cemA
ycf1
trnG-GCC ndhH psbB psbD rpoC1 rpoC2
ndhG psbM
trnV-UAC ndhD petL petN psaC psbA psbI rbcL ycf15 rpoA
ndhB rpl23
atpH atpI ndhC ndhD ndhF ndhG ndhI ndhJ ndhK petA petB petD petN psaC psbN psbT rpoC1 cemA
atpA psbA rpoA
trnG-GCC trnK-UUU trnP-GGG trnT-GGU trnT-UGU trnV-GAC trnV-UAC atpE atpF ndhE ndhH petG petL psaA psaB psaI psaJ psbC psbD psbE psbF psbJ psbK psbL psbM psbZ ycf3 ycf4 rpoB rpoC2 rpl32 rps15 ccsA
psbI ycf15 rpl22 rps16 infA
trnA-UGC trnC-GCA trnR-ACG trnL-UAA trnI-GAU atpB psbB rbcL
atpB psbB rbcL rpl14
trnG-UCC trnL-CAA trnL-UAG trnP-UGG trnR-UCU trnS-GGA trnfM-CAU psbI ycf15 ycf2 ycf1 rps16 rpl22
rpoC2 rpoB rbcL psaB psaA ndhH
atpA atpB atpE atpF ndhE petL psbZ ycf2 ycf15 rpl16 rpl22 rps12 rps15 rps16 rps19 rps3 clpP
trnH-GUG trnG-GCC trnF-GAA trnP-GGG trnQ-UUG trnR-UCU trnS-UGA trnT-UGU trnV-UAC psaI psaJ psbA psbB psbC psbD psbE psbF psbI psbJ psbK psbL psbM ycf3 ycf4 rpoA rpl32 ccsA
ycf4 ycf1 psbK psbJ psbE psbA
ndhE petL
atpA atpB atpE atpF petG psaI psbB psbC psbD psbL ycf3 rpoA accD ccsA
psbZ psbI
psbF psbM rpl16 rpl2 rpl22 rpl36 rps15 rps16 rps19 clpP
psaJ
psaJ
trnP-GGG trnL-UAA psbF
rps12
trnV-UAC atpA atpB atpE atpF petG psaI psbB psbD psbF psbL psbM ycf3
psaJ psbC ycf15 rpl16 rpl2 rpl22 rps16 clpP accD
trnG-GCC trnR-UCU trnT-GGU psbI psbZ rpoA ccsA
psbI psbZ rpoA ccsA
trnP-GGG psbC ycf15
🟩 Photosystem I, II and assembly factors, *rbc*L
🟦 ATP synthase
🟦 NAD(P)H dehydrogenase
🟥 Cytochrome b6/f complex
🟪 Plastid-encoded polymerase (PEP)
🟪 Ribosomal proteins (*rpl* and *rps*)
🟨 Transfer RNA
🟥 Other pathways or unknown function

Fig. 7

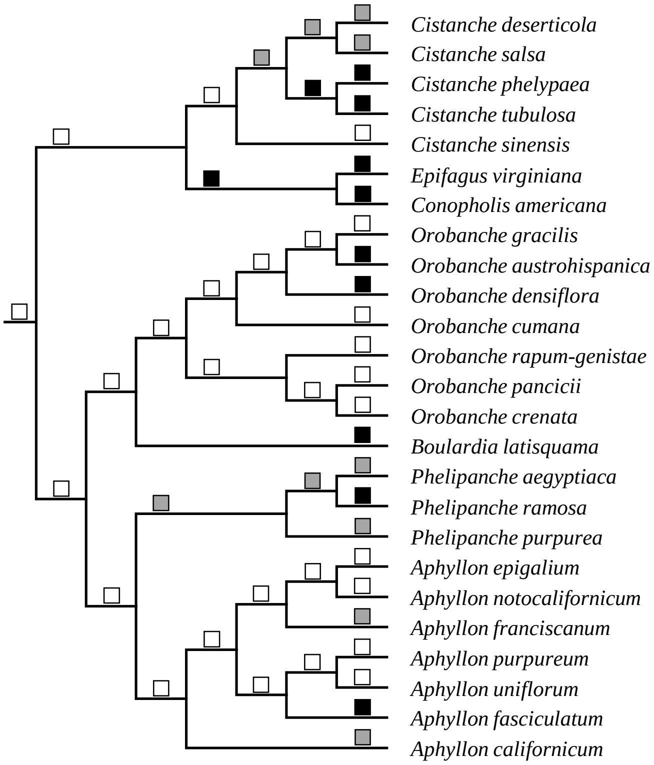

Cistanche deserticola
Cistanche salsa
Cistanche phelypaea
Cistanche tubulosa
Cistanche sinensis
Epifagus virginiana
Conopholis americana
Orobanche gracilis
Orobanche austrohispanica
Orobanche densiflora
Orobanche cumana
Orobanche rapum-genistae
Orobanche pancicii
Orobanche crenata
Boulardia latisquama
Phelipanche aegyptiaca
Phelipanche ramosa
Phelipanche purpurea
Aphyllon epigalium
Aphyllon notocalifornicum
Aphyllon franciscanum
Aphyllon purpureum
Aphyllon uniflorum
Aphyllon fasciculatum
Aphyllon californicum

Fig. 8

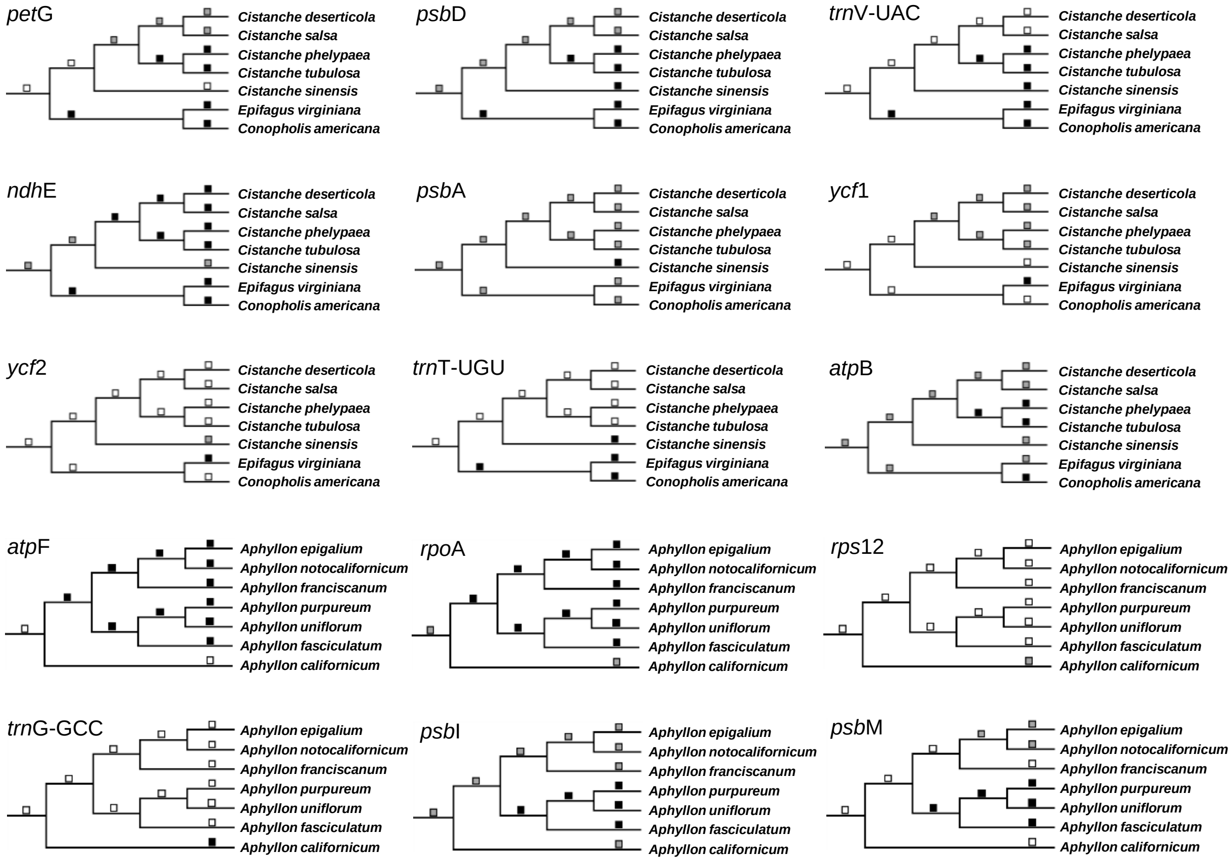

petG
Cistanche deserticola
Cistanche salsa
Cistanche phelypaea
Cistanche tubulosa
Cistanche sinensis
Epifagus virginiana
Conopholis americana
psbD
Cistanche deserticola
Cistanche salsa
Cistanche phelypaea
Cistanche tubulosa
Cistanche sinensis
Epifagus virginiana
Conopholis americana
trnV-UAC
Cistanche deserticola
Cistanche salsa
Cistanche phelypaea
Cistanche tubulosa
Cistanche sinensis
Epifagus virginiana
Conopholis americana
ndhE
Cistanche deserticola
Cistanche salsa
Cistanche phelypaea
Cistanche tubulosa
Cistanche sinensis
Epifagus virginiana
Conopholis americana
psbA
Cistanche deserticola
Cistanche salsa
Cistanche phelypaea
Cistanche tubulosa
Cistanche sinensis
Epifagus virginiana
Conopholis americana
ycf1
Cistanche deserticola
Cistanche salsa
Cistanche phelypaea
Cistanche tubulosa
Cistanche sinensis
Epifagus virginiana
Conopholis americana
ycf2
Cistanche deserticola
Cistanche salsa
Cistanche phelypaea
Cistanche tubulosa
Cistanche sinensis
Epifagus virginiana
Conopholis americana
trnT-UGU
Cistanche deserticola
Cistanche salsa
Cistanche phelypaea
Cistanche tubulosa
Cistanche sinensis
Epifagus virginiana
Conopholis americana
atpB
Cistanche deserticola
Cistanche salsa
Cistanche phelypaea
Cistanche tubulosa
Cistanche sinensis
Epifagus virginiana
Conopholis americana
atpF
Aphyllon epigalium
Aphyllon notocalifornicum
Aphyllon franciscanum
Aphyllon purpureum
Aphyllon uniflorum
Aphyllon fasciculatum
Aphyllon californicum
rpoA
Aphyllon epigalium
Aphyllon notocalifornicum
Aphyllon franciscanum
Aphyllon purpureum
Aphyllon uniflorum
Aphyllon fasciculatum
Aphyllon californicum
rps12
Aphyllon epigalium
Aphyllon notocalifornicum
Aphyllon franciscanum
Aphyllon purpureum
Aphyllon uniflorum
Aphyllon fasciculatum
Aphyllon californicum
trnG-GCC
Aphyllon epigalium
Aphyllon notocalifornicum
Aphyllon franciscanum
Aphyllon purpureum
Aphyllon uniflorum
Aphyllon fasciculatum
Aphyllon californicum
psbI
Aphyllon epigalium
Aphyllon notocalifornicum
Aphyllon franciscanum
Aphyllon purpureum
Aphyllon uniflorum
Aphyllon fasciculatum
Aphyllon californicum
psbM
Aphyllon epigalium
Aphyllon notocalifornicum
Aphyllon franciscanum
Aphyllon purpureum
Aphyllon uniflorum
Aphyllon fasciculatum
Aphyllon californicum

Fig. 9