



Transit From Autotrophism to Heterotrophism: Sequence Variation and Evolution of Chloroplast Genomes in Orobanchaceae Species

Ruiting Zhang[†], Bei Xu[†], Jianfang Li, Zhe Zhao, Jie Han, Yunjing Lei, Qian Yang, Fangfang Peng and Zhan-Lin Liu^{*}

Key Laboratory of Resource Biology and Biotechnology in Western China (Ministry of Education), College of Life Science, Northwest University, Xi'an, China

OPEN ACCESS

Edited by:

Lifeng Zhu,
Nanjing Normal University, China

Reviewed by:

Joo-Hwan Kim,
Gachon University, South Korea
Hengchang Wang,
Wuhan Botanical Garden, Chinese
Academy of Sciences, China

*Correspondence:

Zhan-Lin Liu
liuzl@nwu.edu.cn

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 11 March 2020

Accepted: 07 September 2020

Published: 06 October 2020

Citation:

Zhang R, Xu B, Li J, Zhao Z,
Han J, Lei Y, Yang Q, Peng F and
Liu Z-L (2020) Transit From
Autotrophism to Heterotrophism:
Sequence Variation and Evolution
of Chloroplast Genomes
in Orobanchaceae Species.
Front. Genet. 11:542017.
doi: 10.3389/fgene.2020.542017

The family Orobanchaceae including autotrophic, hemiparasitic, and holoparasitic species, is becoming a key taxa to study the evolution of chloroplast genomes in different lifestyles. But the early evolutionary trajectory in the transit from autotrophism to hemiparasitism still maintains unclear for the inadequate sampling. In this study, we compared 50 complete chloroplast genomes in Orobanchaceae, containing four newly sequenced plastomes from hemiparasitic *Pedicularis*, to elucidate the sequence variation patterns in the evolution of plastomes. Contrasted to the sequence and structural hypervariabilities in holoparasites, hemiparasitic plastomes exhibited high similarity to those of autotrophs in gene and GC contents. They are generally characterized with functional or physical loss of *ndh*/tRNA genes and the inverted small-single-copy region. Gene losses in Orobanchaceae were lineage-specific and convergent, possibly related to structural reconfiguration and expansion/contraction of the inverted region. Pseudogenization of *ndh* genes was unique in hemiparasites. At least in *Pedicularis*, the *ndhF* gene might be most sensitive to the environmental factors and easily pseudogenized when autotrophs transit to hemiparasites. And the changes in gene contents and structural variation potentially deeply rely on the feeding type. Selective pressure, together with mutational bias, was the dominant factor of shaping the codon usage patterns. The relaxed selective constraint, potentially with genome-based GC conversion (gBGC) and preferential codon usage, drive the fluctuation of GC contents among taxa with different lifestyles. Phylogenetic analysis in Orobanchaceae supported that parasitic species were single-originated while holoparasites were multiple-originated. Overall, the comparison of plastomes provided a good opportunity to understand the evolution process in Orobanchaceae with different lifestyles.

Keywords: Orobanchaceae, plastomes, hemiparasites, gene loss, selection, phylogeny

INTRODUCTION

Chloroplasts are the key organelles for green plants, where light energy can be converted into chemical energy through photosynthesis. Chloroplast DNAs of land plants have widely been sequenced and typically have highly conserved structure with a large and a small single-copy region (LSC and SSC, respectively) separated from each other by two identical inverted repeat (IR)

regions (Jansen and Ruhlman, 2012). Generally, chloroplast genomes (plastomes) range in size from 120 to 160 kb with 120–130 genes, many of which encode proteins, tRNAs, and rRNAs essential for photosynthesis (Williams et al., 2015). Research works on plastomes can provide valuable information not only to reconstruct the phylogenetic relationships among complex taxa due to their high resolution (Huang et al., 2014), but also to understand the evolutionary history of plants with different lifestyles. For non-photosynthetic plants retrieving nutrients partially or exclusively from the hosts, photosynthesis is not the only process to obtain energy. Genes involved in photosynthesis are no longer required and free from selective constraints. Therefore, the plastomes in heterotrophs have undergone various changes in genome size and structure, such as pseudogenization, gene loss and genome rearrangements (Cusimano and Wicke, 2016; Graham et al., 2017). For example, the plastomes of heterotrophic species in Ericaceae are highly reduced in size and have lost all photosynthesis-related genes (Braukmann et al., 2017). Large-scale structural reconfigurations have occurred in the holoparasitic species of Orobanchaceae (Wicke et al., 2016). Loss of IRs and *ndh* genes and accelerated substitution rates of protein-coding genes were reported in a hemiparasitic herb *Cassytha filiformis* (Wu et al., 2017).

The broomrape family, Orobanchaceae, contains about 90 genera and more than 2000 species with autotrophic, hemiparasitic and holoparasitic lifestyles, becoming a model for tracing the evolutionary trajectories of non-photosynthetic plant genomes (Westwood et al., 2012). Except for the non-parasitic genera of *Lindenbergia*, *Triaenophora*, and *Rehmannia*, the rest can be classified as holoparasites or hemiparasites (Byng et al., 2016). Hemiparasites have green leaves and are capable of photosynthesis while holoparasites without photosynthetic tissue depend on their hosts for all their requirements. Hemiparasites include two groups: facultative and obligate. Contrasting to the obligate, facultative hemiparasites can fulfill its lifecycle without a host. Phylogenetic works supported that parasitism in Orobanchaceae was single originated and there were at least three times of the origin in holoparasitic taxa (Samigullin et al., 2016). Recently, the evolution of plastomes in Orobanchaceae species was investigated (Wicke et al., 2011, 2013) and hypothesis was proposed that stepwise gene loss followed by the functional loss of photosynthesis was convergent in the degradation route of plastomes (Wicke and Naumann, 2018). At the beginning of heterotrophs, non-essential or stress-relevant genes (*ndh*) are non-functional or lost, followed by primary photosynthesis-related genes (*pet*, *psa*), and non-essential housekeeping genes in the process of plastome degradation. When photosynthesis-unrelated metabolic genes and other housekeeping genes are deleted, the degradation reaches its final stage (Wicke and Naumann, 2018). But the evolutionary hypothesis was conducted by plastome comparisons mainly between autotrophs and holoparasites. The reduction of plastomes might be diverse (Liu et al., 2020) and lineage-specific (Wicke et al., 2013). As the early stage from autotrophs to heterotrophs, the evolutionary trajectories of hemiparasitic plastomes are still not well-known. More hemiparasitic species should be necessarily involved to investigate the evolutionary patterns of plastomes.

Pedicularis is the largest lineage of photosynthetic root hemiparasitic plants with about 600 species, widely distributed in the northern temperate zone (Yu et al., 2015). Most of species in the genus are facultative and narrowly distributed. The diversified morphological characters made classification of the genus still unresolved (Yu et al., 2015). Research works are mainly focused on its phylogenetic position and interspecific relationships. Interestingly, *Pedicularis* species are considered as primitive parasites with very wide host ranges. Majority of species in the genus are potentially both parasitic and mycorrhizal. Some could live well without dominant hosts in artificial condition (Li and Guan, 2008). The slight dependence on hosts made *Pedicularis* a good model for heterotrophism. Studies on plastomes in the genus potentially shed new lights on variation pattern at the first stage of genome evolution in Orobanchaceae. In this study, we newly sequenced four *Pedicularis* plastomes, and performed comparative genome analysis with their autotrophic or holoparasitic relatives, together with 13 publicly available hemiparasitic plastome data in Orobanchaceae. Our aims were mainly to characterize the hemiparasitic plastomes, to uncover the sequence divergence and structural changes, and investigate the early stage of the evolutionary trajectories of plastomes in Orobanchaceae with variable lifestyles.

MATERIALS AND METHODS

Sample Collection, Sequencing, Genome Assembly, and Annotation

Four *Pedicularis* species were collected from their natural habitats [*P. resupinata* L. (voucher No. 15-0327), *P. muscicola* Maxim. (voucher No. 15-0345), *P. oederi* Vahl (voucher No. 15-0381), 107.71°E, 34.03°N; *P. longiflora* Rudolph (voucher No. 16-0012), 91.10°E, 30.78°N]. All specimen were deposited in Herbarium of Northwest University. Next generation sequencing was constructed by Illumina HiSeq 2500 platform in Biomarker Technologies, Inc. (Beijing, China). Raw reads were quality-trimmed using the NGS QC Toolkit v2.3.3 (Patel and Jain, 2012) allowing a maximum of 2% ambiguous base calls. The clean reads were mapped to the plastome sequence of *Castilleja paramensis* (KT959111) using the software MITObim v1.8 with the default parameters (Hahn et al., 2013). The assembled contigs were generated into a consensus sequence in GENEIOUS R8 (Biomatters, Ltd., Auckland, New Zealand), then annotated using the program DOGMA (Wyman et al., 2004) and manually corrected by comparing them with the reference plastome in GENEIOUS R8 program.

Sequence Divergence and Rearrangement Analysis

To identify the sequence divergence of autotrophic, hemiparasitic, and holoparasitic species of Orobanchaceae, four new *Pedicularis* plastomes, together with the 46 publicly available cpDNA sequence from GenBank, including 8 autotrophs, 13 hemiparasitic species and 25 holoparasitic species, were used in the evolutionary analyses. All taxa in this study are listed

in the **Supplementary Table S1** with their GenBank accession numbers. Sequences were aligned using MAFFT (Katoh and Standley, 2013) in GENEIOUS R8 with the default parameters set and manually edited when necessary. Sequence divergence was investigated with DnaSP v5 (Librado and Rozas, 2009) and MEGA v7.0.18 (Kumar et al., 2016). Sliding window analysis was conducted to generate nucleotide diversity (π) using program DnaSP. The step size was set to 200 bp with a window length of 600 bp. Mauve Alignment (Darling et al., 2004) was performed to reveal the rearrangement among plastome of these species.

Codon Usage Analysis

The program CodonW (Sharp and Li, 1987)¹ was employed for the analysis of synonymous codon usage in the protein coding regions (>300 bp). We tabulated two codon usage measures: the effective number of codons (N_c) and frequency of the GC at the third synonymous position (GC3s) (Ivanova et al., 2017). Estimation of the standard effective number of codon (ENc) was tabulated using the equation $ENc = 2 + s + 29/[s^2 + (1 - s)^2]$, where s denoted GC3s (Wright, 1990). We also calculated ENc ratio ($ENc_{exp} - ENc_{obs} / ENc_{exp}$), to estimate the difference between observed and expected ENc values (Kawabe and Miyashita, 2003).

Selection Pressure Analysis

To calculate the non-synonymous (dN) and synonymous (dS) substitution rates and their ratio ($\omega = dN/dS$), the nine protein-coding genes shared in 50 Orobanchaceae species were extracted and aligned separately. The value of dN , dS and ω for each protein-coding gene were calculated using site-specific model implemented in the codeml package (seqtype = 1, model = 0, NSsites = 1, 2, 7, 8) of PAML v4.7 (Yang, 2007), holding the maximum likelihood (ML) tree constructed with plastomes by using software RAxML (Stamatakis, 2014). Adaptive evolution of each gene was determined by computing likelihood ratio tests (LRTs). To check the presence of positively selected sites, we examined the LRTs for positive selection of two pairwise models: (1) M1a (Nearly Neutral; $\omega_0 < 1$, $\omega_1 = 1$) against M2a (Positive Selection; $\omega_0 < 1$, $\omega_1 = 1$, $\omega_2 > 1$), while (2) M7 (Neutral; β) against M8 (Selection; β and ω). The Bayes empirical Bayes (BEB) procedure was used to identify the positively selected sites with the degree of freedom (df) of two. Only candidate sites of positive selection with significant support values from posterior probability ($\omega > 1$; $p \geq 0.99$; BEB) identified by M2 and M8 were considered further (Nielsen and Yang, 1998; Yang and Nielsen, 2002).

We used RELAX (Wertheim et al., 2015), which is available from the Datamonkey webserver², to test the changes in the relative selection intensity on all most shared genes in Orobanchaceae. Given test and reference branches in a codon-based phylogenetic framework, RELAX can calculate a selection intensity parameter value (k) and its statistical significance, with $k > 1$ showing intensified selection and $k < 1$ indicating relaxed selection.

¹<http://codonw.sourceforge.net/>

²<https://www.datamonkey.org/RELAX>

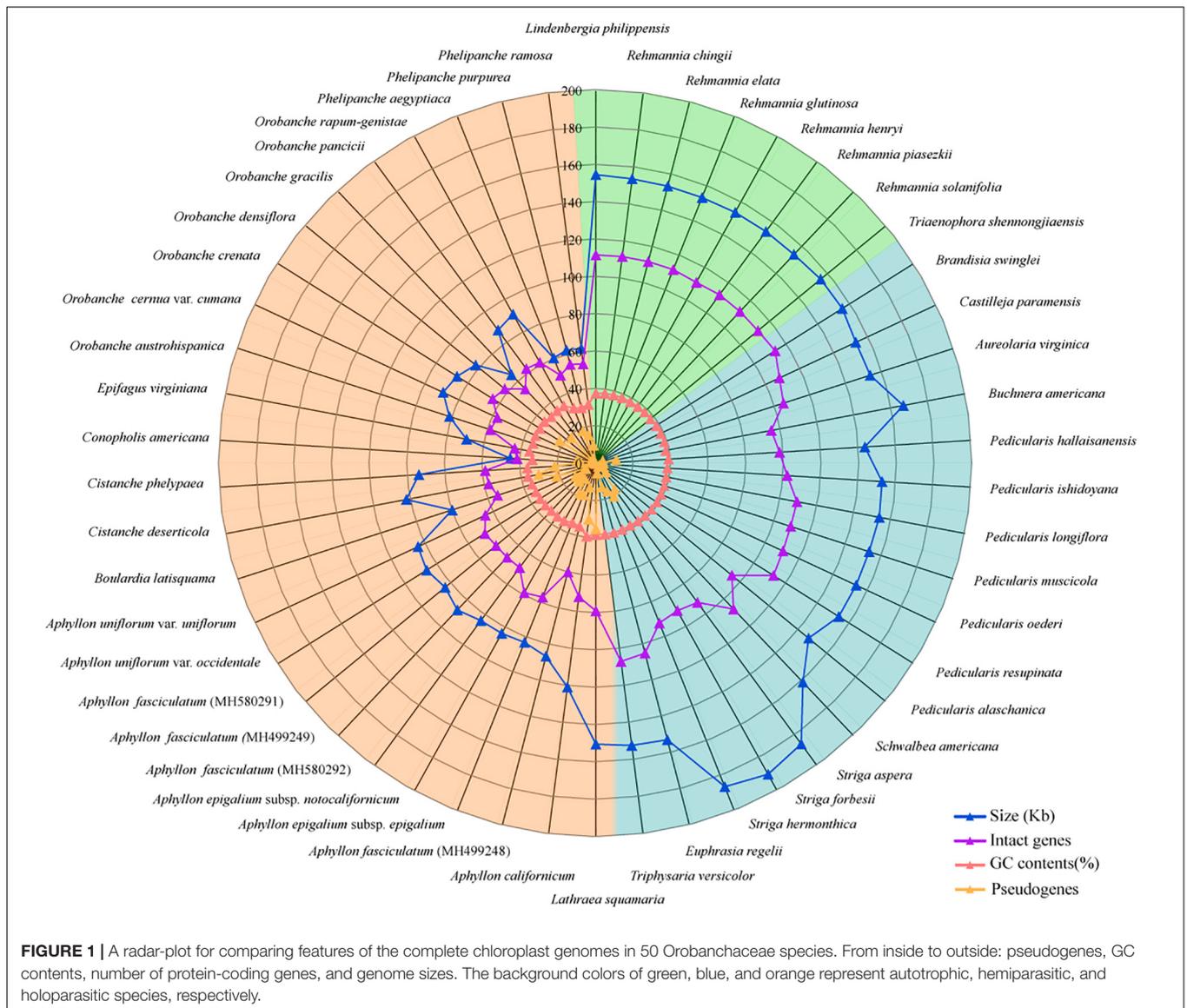
Phylogenetic Analyses

Plastomes of fifty representative species in Orobanchaceae were used for the reconstruction of phylogenetic tree with two Lamiaceae species, *Tectona grandis* (NC_020431) and *Rosmarinus officinalis* (NC_027259), as outgroup. Because of the highly variable molecular evolutionary rates among the different plastome regions, phylogenetic relationship analyses were performed using nine conserved protein-coding genes (*rps2*, *rps4*, *rps7*, *rps8*, *rps14*, *rpl2*, *rpl16*, *rpl36*, and *matK*) in Orobanchaceae. We determined the best-fitting evolutionary model using Modeltest v3.7 (Posada and Crandall, 1998) with the Akaike Information Standard (AIC) score (Posada and Buckley, 2004). PartitionFinder 2 (Lanfear et al., 2017) was used to estimate optimal partitioning schemes by grouping similar data blocks from the output of the remaining datasets. Phylogenetic analysis with maximum likelihood (ML) method was carried out using software RAxML v8.2 (Stamatakis, 2014) with 1000 bootstrap replicates. Bayesian inference (BI) analyses were performed with program MrBayes v3.1.2 (Ronquist and Huelsenbeck, 2003). The Markov chain Monte Carlo (MCMC) algorithm was run for 10,000,000 generations with trees sampled every 10,000 generations for each data partition. The first 25% of trees from all runs were discarded as burn-in, and the remaining trees were used to construct the majority-rule consensus tree. FigTree v.1.31 was used to visualize the tree topology (Rambaut, 2006). The phylogenetic relationships among hemiparasitic species in Orobanchaceae were further identified by using the ML method with 55 common intact genes.

RESULTS

Chloroplast Genome Structure of Four *Pedicularis* Species

Illumina sequencing technology produced sufficient data (with coverage > 150X) for plastome assembly from 17,462,980 (*P. muscicola*) to 23,876,956 (*P. oederi*) paired-end reads. The plastomes of four *Pedicularis* species sequenced here (GenBank accession numbers: MH703578 ~ MH703581) had a typical quadripartite structure with high similarity in sizes (from 152,907 bp in *P. muscicola* to 153,547 bp in *P. longiflora*) (**Supplementary Table S1** and **Figure 1**). They also have the identical GC contents (about 38%) and good collinear structure (not shown here). Each *Pedicularis* plastome contained 133 genes, including 37 tRNA genes and eight rRNA genes, of which 115 occurred as single copy genes and 18 genes were duplicated in the IR regions (**Supplementary Tables S1, S2**). The number of protein-coding genes (PCGs) and pseudogenes varied from species to species. There were 88 PCGs and no pseudogene in *P. resupinata*. In contrast, two pseudogenes (*ndhD* and *ndhF*) existed in *P. oederi*, and three pseudogenes (*ndhD*, *ndhF*, and *ndhH*) were found in *P. muscicola* and *P. longiflora* (**Supplementary Table S2**). These pseudogenes had multiple predicted frameshift mutations and premature stop codons located at different sites. We compared IR/LSC and IR/SSC junctions in our *Pedicularis* plastomes and



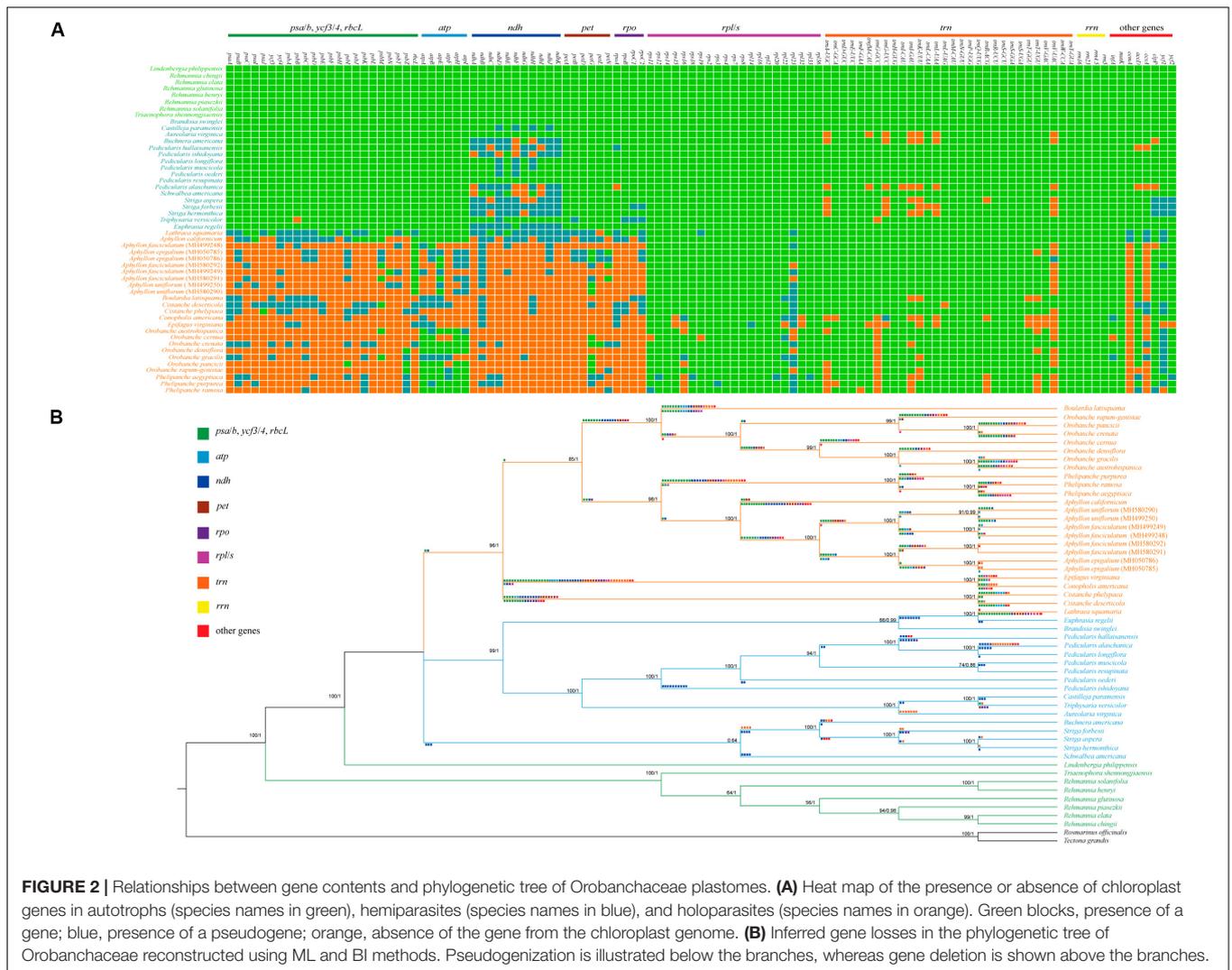
found that two copies of *ycf1* gene spanned the SSC/IR junctions, while the *ndhF* gene crossed the SSC region into IRa region with some variable nucleotides in *Pedicularis* species, except *P. resupinata* (**Supplementary Figure S1**). Contrasting to the relatively stable SC/IR boundaries in the four newly sequenced species, plastomes of *P. alaschanica*, *P. hallaisanensis* had prominent shifts of SSC/IRa boundaries due to loss of *ndh* genes (see below) and *P. ishidoyana* had the shortest SSC region (27 bp) for the expansion of IRs (**Supplementary Figure S1**).

General Features of Chloroplast Genomes in Orobanchaceae

A total of 50 plastomes of Orobanchaceae were utilized to analyze the variation of sequence and structure, including 8 autotrophs, 17 hemiparasites, and 25 holoparasites species (**Supplementary Table S1** and **Figure 1**). The largest plastome

was 190 kb (*Striga forbesii*), about 4.2 fold larger than the smallest plastome of *Conopholis americana* (46 kb). Autotrophs and hemiparasites generally have similar genome sizes, varying from 143 kb (*P. hallaisanensis*) to 160 kb (*Schwalbea americana*) with unexpected larger plastomes in *Striga* species. A significant reduction of genome sizes was observed in the holoparasites plastomes with a range from 46 kb (*Conopholis americana*) to 150 kb (*Lathraea squamaria*). The GC contents of plastomes also dramatically decreased in holoparasites (35.0% in average), in contrast to the slightly but significantly different GC values in hemiparasites (38.2% in average) and autotrophs (37.9% in average) (**Supplementary Figure S2**).

When functional genes simply counted, only 26 genes, including nine PCGs (*rps2*, *rps4*, *rps7*, *rps8*, *rps14*, *rpl2*, *rpl16*, *rpl36*, and *matK*), four rRNAs genes, and 13 tRNA genes were commonly present in all 50 plastomes of Orobanchaceae species (**Figure 2A**). Comprising 133 genes, autotrophs had



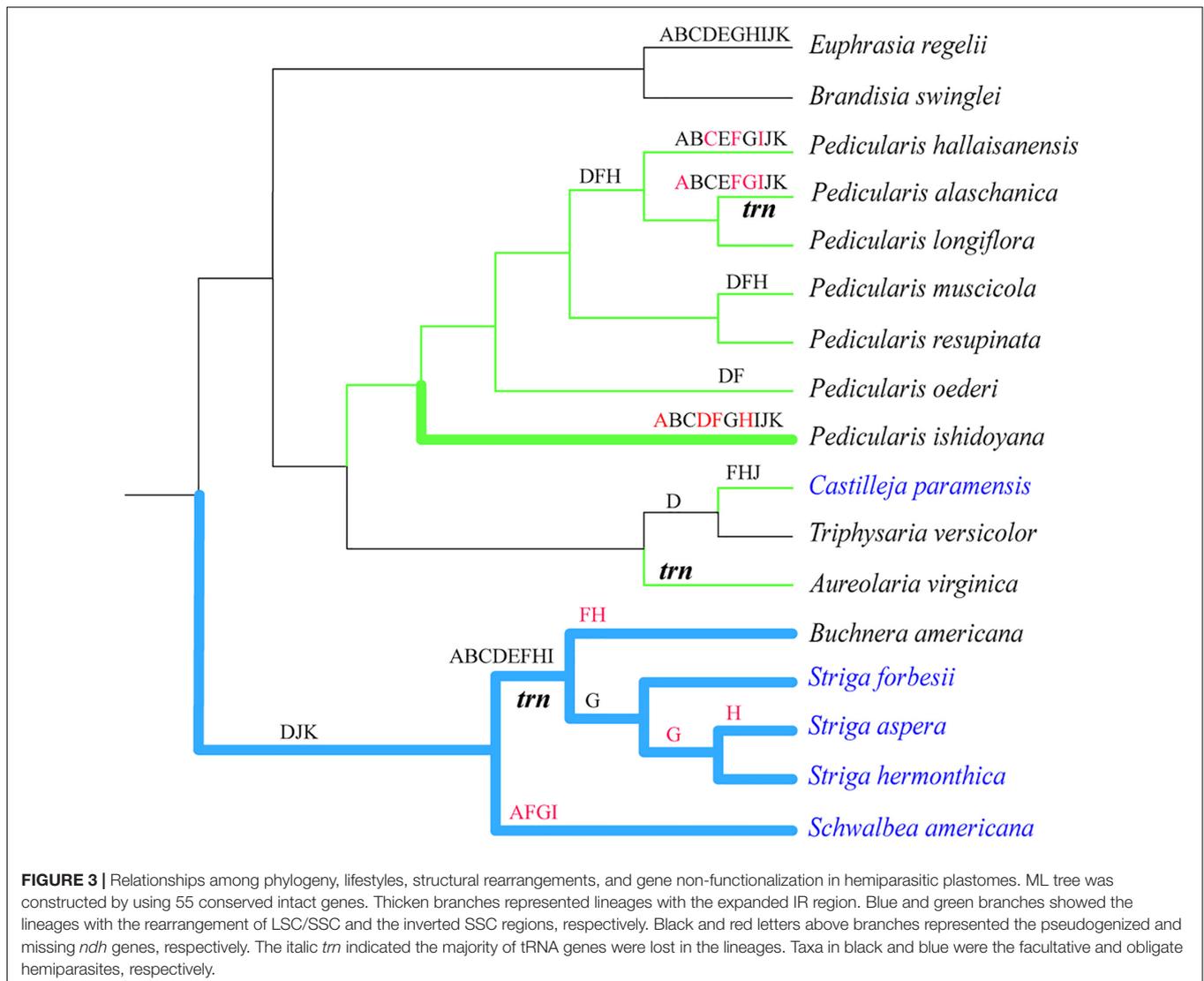
no gene loss and pseudogenes. Pseudogenization and loss of the *ndh* genes were commonly observed in the hemiparasitic plants. The functional or physical deletion of tRNA gene family, generally together with other housekeeping genes, is unique at the genus level (Figure 3). For holoparasites, further gene loss and non-functionalization were commonly found with most of the photosynthetic genes, such as *psa/psb* (photosystem I and II), *pet* (cytochrome b6/f complex), *ccsA* (heme attachment factor), *cemA* (inner membrane protein for CO₂ uptake), *ycf3/4* (both photosystem assembly factors), and *ndh* (Figure 2A). Phylogenetic tree of the 50 Orobanchaceae species reconstructed by all most conserved plastome genes indicated that gene loss was highly lineage-specific (Figures 2B, 3).

Sequence Divergence and Rearrangement in Orobanchaceae Plastomes

Sequence divergence estimated by nucleotide diversity showed that sequence variation increased from the autotrophs

($\pi = 0.0147$) to heterotrophs (the holoparasites had the highest, $\pi = 0.1430$) (Supplementary Table S3). This tendency was also confirmed by sliding window analysis (Supplementary Figure S3). The autotrophic plants held relatively conserved plastomes ($\pi = 0 \sim 0.069$), of which *psbE-petL*, were the most variable region (Supplementary Figure S3A); For the hemiparasitic plants, the range of nucleotide diversity was much wider ($\pi = 0.0025 \sim 0.4177$) with the most variable region of the *ndh* gene cluster (*ndhH*, *ndhA*, *ndhI*, *ndhG*, and *ndhE*) (Supplementary Figure S3B); Due to excessively alignment gaps in plastomes of holoparasitic species, net number of nucleotide substitution per site was limited and sliding window analysis could not be successfully performed. This result indicated that contrast to the low level of variation in the most parts of plastomes of autotrophs and hemiparasites, all regions of the holoparasitic plastomes harbored high variation of nucleotide acids.

Linear rearrangement comparisons across the plastome sequences (one IR region removed) of 50 Orobanchaceae species were conducted by aligning with *Lindenbergia philippensis*



as the reference. All autotrophs presented the completely identical collinear plastomes (data not shown). By contrast, heterotrophs exhibited lineage-specific structural variation from good collinearity to high reconfiguration. Among the hemiparasitic species, plastomes in *Brandisia swinglei* and *Triphysaria versicolor* had similar gene orders to autotrophs while others had partly or completely inverted and rearranged SSC regions to a variable extent (Supplementary Figure S4A). In contrast to the stability of the LSC and IR regions in most hemiparasites, a 5 kb inversion was found in the LSC region of *Schwalbea americana* and frequent large-scale structural reconfigurations existed in LSCs of *Buchnera americana* and three *Striga* species. In holoparasitic plants, plastomes of *Lathraea squamaria* and species in *Aphyllon* genus showed a good collinearity to autotrophs with genome reduction and occasionally inverted sequences in SSC. But high structural reconfigurations, such as inversion, a large scale of reduction or even loss of the IR region were usually found in the majority of holoparasitic taxa (Supplementary Figure S4B).

Codon Usage and Selection Events in Protein-Coding Genes

A notable tendency of high A/T contents, especially the third codon position (GC3s, with the mean value of 0.25), was commonly found in all plastomes (Supplementary Table S1 and Supplementary Figure S4B). The effective number of codons (ENC) is widely used to measure the codon bias. The ENC values varied from 33 to 61, demonstrating a significant difference in the codon usage bias (CUB) with extreme bias or equal usage among genes in Orobanchaceae plastomes. Genes in hemiparasites usually presented the highest values of GC3s and ENC while holoparasites had the lowest (Supplementary Figure S4B), indicating the stronger preference of codon usage of AT in holoparasitic cp genes. The ENC-plot distribution (ENC versus GC3s) showed that ENC values of most genes were far away from the expected ENC curve, suggesting that these genes were not solely constrained by the compositional constraints (Supplementary Figure S5) and other factors (e.g.,

translational selection) might influence the features of codon usage. The possibility of natural selection shaping the codon usage of the photosynthesis genes (*psa/b*, *ycf3/4*, *rbcL*) was also supported by their narrow distribution of GC3s (Kawabe and Miyashita, 2003). Genes below or above the ENC curve are likely under positive or negative selection pressure for codon usage (Wright, 1990). Therefore, these photosynthesis genes in autotrophs and hemiparasites had potentially experienced a selection in opposite direction for codon usage. The frequency distribution of ENC ratios showed that there was a simple peak of $-0.05\sim 0.15$ in most chloroplast genes, except the *ndh* genes in holoparasites. The major genes of heterotrophic species usually had ENC ratio values of $-0.05\sim 0.05$. By contrast, autotrophs harbored higher ENC ratio values (0.10–0.15) in core photosynthesis genes (*pet*) and ATP synthase (*atp*) genes, suggesting that they had slightly smaller ENC values than expected, and other factors apart from compositional constrain, such as selection pressure, influenced CUB as implied by ENC plots (data not shown).

To identify genes under the positive selection, nine protein-coding genes shared among 50 Orobanchaceae species were extracted to calculate the dN/dS ratio (ω) using KaKs_calculator (Zhang et al., 2006). The dN/dS ratios of *rps2*, *rps14*, and *rpl36* gene were almost above 1.0 in all 50 species, indicating that these genes were potentially under the pressure of positive selection in Orobanchaceae (Supplementary Figure S6). The genes of *rps4*, *rps7*, *rps8*, *rpl2*, *rpl16*, and *matK* were subjected to purifying selection ($dN/dS < 1$). When the site-specific selection events were identified by using the dN/dS analysis, both LRTs models (M1 vs. M2; M7 vs. M8) supported *rps2* gene had potentially undergone the positive selection with one selective site (Supplementary Table S4).

We used the codon-based phylogenetic framework RELAX to test the change of selection pressure. Among the shared nine genes of Orobanchaceae species with different lifestyles, a significantly relaxed selection intensity of *rpl2* gene was found between autotrophs and heterotrophs ($k = 0.45$, $p < 0.01$). There was no significant relaxed or intensified selection existing between hemiparasites and holoparasites ($p = 0.86$).

Phylogenetic Analysis

The phylogenetic tree based on nine shared genes in Orobanchaceae with ML and BI methods had an identical topological structure with moderate-to-high support values. The monophyletic Orobanchaceae was well-supported (Figure 2B). The non-parasitic taxa of *Triaenophora* and *Rehmannia* formed a basal lineage and the third autotrophic taxon, *Lindenbergia philippensis*, was sister to all other parasite species as supported in the previous study (Wicke et al., 2013). The holoparasitic *Lathraea squamaria* was clustered with hemiparasites, demonstrating the multiple origins of the holoparasites (Samigullin et al., 2016). All hemiparasitic species were grouped into two major lineages. *Schwalbea* was close to *Striga* and *Buchnera*, but with weak support values. The genus *Pedicularis* was monophyletic and sister to *Castilleja-Triphysaria-Aureolaria* clade. Interspecific relationships within *Pedicularis* were also determined and the clearly basal position

of *P. ishidoyana* was identified in both ML and BI phylogenetic trees (Figures 2B, 3).

DISCUSSION

Plastome Size and Structure Rearrangement

Parasitic species in Orobanchaceae were proven to be a monophyletic group (Ree, 2005; Yu et al., 2015) and became a good model for the plastome evolutionary study. In this work, we compared the complete chloroplast genomes of 50 species to understand the evolutionary transition in the early stage from autotrophs to holoparasites. Plants in the family exhibited a high variation of plastomes in size and structure. Genome reduction and multiple rearrangements mainly occurred in the holoparasitic species (Figure 1 and Supplementary Figure S4B) and well-discussed (Wicke et al., 2013; Wicke and Naumann, 2018). In Santalales, hemiparasites generally characterized by the reduction of chloroplast genome (Li et al., 2017; Shin and Lee, 2018; Wicke and Naumann, 2018). But most of the hemiparasitic plastomes in Orobanchaceae were similar to those of autotrophs in genome size, GC and gene contents (Figure 1 and Supplementary Table S1) (Wicke et al., 2013; Zeng et al., 2017). This stable structure indicated their plastomes were just at the very early stage of transition from autotrophs to holoparasites (Wicke and Naumann, 2018). Slight plastome reduction were observed in *P. hallaisanensis* and *P. alaskanica*, which was essentially due to the loss of *ndh* genes (Figure 2 and Supplementary Figure S4). By contrast, recent works in *Buchnera americana* and three *Striga* species uncovered an unexpected enlargement of plastomes due to the expansion of the inverted repeats into the single copy region (Frailey et al., 2018). This increase of plastome size was rather rare and its effects were still unknown.

All plastomes of hemiparasites in this study had a typical quadripartite structure with slight expansions and contractions in the IR/SC boundary region with an exception as mentioned above. Structural reconfiguration of plastomes was gradually increased and lineage specific. *Brandisia swinglei*, *Triphysaria versicolor*, and *Euphrasia regelii* had no large rearrangement compared to the autotrophic plastomes. All these hemiparasites are facultative and heavily depend on photosynthesis. No or a small scale of rearrangement is expected in these species (Frailey et al., 2018). Sequence inversions and rearrangements in the SSC region were the common feature in the rest hemiparasitic plastomes (Supplementary Figure S4A). These large inversions have also been reported in other angiosperms, such as Asteraceae (Jansen and Palmer, 1987) and Leguminosae (Jansen et al., 2008). The chloroplast genome was hypothesized with a flip-flop recombination and two equally chloroplast structural haplotypes occurred in most land plant individuals (Wang and Lanfear, 2019). But the phenomenon that nearly all hemiparasites shared the inverted SSC might not be simply ascribed to the random capture of two haplotypes when sequence reads were assembled. And further studies are necessarily needed to identify its universality. In addition to the inverted SSC region,

the clade of *Schwalbea-Buchnera-Striga* also had inversion and rearrangements of LSC fragments (Frailey et al., 2018), which implied their specificity in the phylogenetic position and the evolutionary trajectory of plastomes.

Pseudogenization and Gene Losses

From autotrophs to heterotrophs, the genome size and the gene number in plastomes decreased gradually along with the increasing number of the pseudogenization and gene loss (Figure 1). The reduction is mainly due to the loss of genes on a variable scale. Pseudogenes and gene losses were not found in autotrophs of Orobanchaceae, but appeared occasionally in most of hemiparasites, and became rather common in holoparasites (Figure 2) (Wicke et al., 2016). Several models of the plastome degradation have been proposed to explain the physical or functional change linked with the transition to heterotrophy. *Ndh* genes are often the first non-functional or lost genes due to parasitism, followed by the photosynthesis-related and -unrelated apparatus. The housekeeping genes maintain functional till the final stage of the chloroplast reduction (Barrett and Davis, 2012; Barrett et al., 2014; Naumann et al., 2016; Wicke et al., 2016; Graham et al., 2017; Wicke and Naumann, 2018). Our data well-supported the hypothesis. Holoparasitic species in Orobanchaceae exhibited serious functional or physical gene loss including photosynthesis genes (*atp*, *ndh*, *psa/b*, *ycf3/4*, *rbcL*, *pet*, *ccsA*, *cemA*), housekeeping genes (*rpo*, *rpl/s*), and tRNA genes. By contrast, some hemiparasitic *Pedicularis* species harbored none or a few *ndh* pseudogenes (Figure 2), indicating the earliest stage of parasites.

Ndh genes are composed of 11 genes (*ndhA* ~ *ndhK*), encoding the thylakoid NAD(P)H dehydrogenase complex together with the nuclear genes *ndhL* ~ *ndhO*. The complex plays a critically important role in photosynthesis. Meanwhile, their mutation rates are high and sensitive to environmental conditions and stress (Silva et al., 2016). For heterotrophs, once the parasitic system was set, the mode of carbon uptake and utilization changed and *ndh* genes became dispensable. The process in the functional loss of *ndh* genes is still unknown. In *Pedicularis*, the number of pseudogenized or lost *ndh* genes varied from none (*P. resupinata*) to all (*P. hallaisanensis* and *P. alaschanica*) (Figure 3). According to the hypothesis of the degradation of plastomes in a stepwise manner, dispensable *ndh* genes should be pseudogenized successfully before they were physically deleted (Wicke et al., 2016; Wicke and Naumann, 2018). Pseudogenes could be generated by premature stop codons and frameshift mutations (Balakirev and Ayala, 2003; Poliseño et al., 2010). In the newly studied *Pedicularis* species, *ndh* pseudogenes were verified for the presence of several internal stop codons. Among the absent member of *ndh* gene family, *ndhD* and *ndhF* were pseudogenized or lost in all seven *Pedicularis* species except *P. resupinata* (Figure 3). The non-functionalization in *Pedicularis* plastome might have opportunities to start with the pseudogenization of *ndhD* and *ndhF* genes. The clues are also found in other parasitic groups. These two genes are also identified as commonly inactive genes in the lineage of *Striga* (Frailey et al., 2018). The losses of *ndhF* are ancestrally shared in the Santalales (Shin and Lee, 2018). The *ndhF* gene is generally

located in the junction of SSC/IRa (Supplementary Figure S1). The frequent contraction and expansion of the IRs, a common evolutionary event in angiosperms (Kode et al., 2005), might affect the stability of IRs and their immediate neighborhood, leading to the increase of mutation rates and further loss of the neighbor *ndh* genes (Wicke et al., 2011; Li et al., 2017). The inversion and rearrangement of the SSC region in hemiparasites could potentially accelerate this procedure (Supplementary Figure S4). In reverse, the loss of the *ndh* genes is likely to lead to the structural changes of plastomes. Contrasted to species with intact or pseudogenized *ndh* genes, those with lost *ndh* genes had undergone a slight shift of IR/SSC (*P. hallaisanensis* and *P. alaschanica*) or expansion IRs (*P. ishidoyana*) (Figure S4). The loss of *ndh* genes, especially *ndhF*, might play every important role in the shift of the IR/SC boundary of plastomes (Kim et al., 2015; Cho et al., 2018). But the assumption of pseudogenization starting with *ndhF* in *Pedicularis* should be tested further in a wide range of heterotrophs. More works, especially functional and physiological analysis, are necessarily needed in future to uncover the evolutionary trajectories of *ndh* genes.

In heterotrophs, gene loss is tightly linked to the dependent extents of parasites on the hosts (Cusimano and Wicke, 2016; Wicke and Naumann, 2018). Their different reliance to the host potentially shaped the evolution of chloroplast genome to some extent (Barrett et al., 2014; Naumann et al., 2016; Wicke et al., 2016; Wicke and Naumann, 2018). Although *Pedicularis* is commonly considered as the root hemiparasitic group, little is known about its parasitic habit. The recent study uncovering the coexistence of haustorial and mycoheterotrophic parasites demonstrated that *Pedicularis* had complex nutrient uptake strategy (Li and Guan, 2008). Unfortunately, the parasitic habit of *Pedicularis* species involved in this study retains unknown. Justified from no or quite a few pseudogenes, these *Pedicularis* species might slightly rely on the host plants. But the functional and/or physical loss of genes is not determined by heterotrophy *per se* (Wicke et al., 2011, 2014), i.e., facultative hemiparasites might not necessarily harbor less non-functional genes than the obligate. Structural variation as mentioned above could also affect the evolution of *ndh* genes. For example, *P. ishidoyana* had much more lost or pseudogenized *ndh* genes than most of the facultative species (Figure 3).

The multiple origins of the facultative species were found in Orobanchaceae as verified in orchids (Barrett and Davis, 2012), carnivorous plants (Silva et al., 2016), even gymnosperms (Lin et al., 2010). In the monophyletic genus of *Pedicularis* (Ree, 2005; Robart et al., 2015), species are phylogenetically independent to non-functionalization or loss genes of *ndh*. For instance, in the phylogenetic tree, *P. ishidoyana*, a basal lineage in *Pedicularis*, lost all *ndh* genes (except *ndhE*), while *P. resupinata* contained all functional *ndh* genes (Figure 3). These results showed that parallel pseudogenization or loss of *ndh* genes frequently occurred in *Pedicularis*, as reported in previous works (Frailey et al., 2018; Shin and Lee, 2018). When phylogenetic relationships were separately analyzed with each gene, the loss was derived directly from the intact gene or indirectly from the pseudogene as proposed by Liu et al. (2020).

Contrasting to *Pedicularis*, the clade of *Schwalbea-Buchnera-Striga* exhibited lineage-related pattern of gene loss. Most *ndh* genes are lost early in the lineage, but *ndhG* and *ndhH* are only absent in the terminal branches, indicating *ndh* genes sequences might retain despite the loss of function (Frailey et al., 2018). In addition, *Buchnera-Striga* clade also lost several tRNA genes. This degradation pattern is associated with rearrangement of SSC/LSC and contraction/expansion of IR regions (Figure 3). The basal *Schwalbea* had the lowest structural variation and gene loss while the terminal *Striga* showed the highest (Figures 2, 3 and Supplementary Figure S4). This group would potentially be a key lineage for analysis of the gradual degradation pattern in plastome evolution. This phylogeny-related pattern of gene loss was also found in Santalales. Genes of *ndhF*, *ndhI*, and *ndhJ* are commonly absent in all Santalales species. *NdhA*, *ndhC*, *ndhD*, *ndhE*, *ndhG*, *ndhH*, and *ndhK* are lost in *Viscum* of Viscaceae (Shin and Lee, 2018).

Codon Usage, GC Content, and Selective Pressure

Variable rates of molecular evolution are often associated with life history in heterotrophs (Smith and Donoghue, 2008). With the transition to a parasitic lifestyle, the chloroplast genome will experience relaxed selection pressure, accelerated evolutionary rate, and reduced GC contents during the evolution process (Wicke et al., 2016; Wu et al., 2017; Shin and Lee, 2018; Wicke and Naumann, 2018). According to our results, holoparasites exhibited higher structural variation and lower GC contents in plastomes in contrast to their autotrophic relatives while hemiparasitic plants harbored similar or slightly higher GC contents than autotrophs (Figure 1 and Supplementary Figure S2). One hypothesis is that these changes in heterotrophs are random byproducts of transition of lifestyle because plastomes are not directly involved in the parasite/host interaction (Wicke and Naumann, 2018). The alternative is that along with the reduction of genome size, especially the loss of genes with higher GC contents, the decrease of GC contents is becoming a general trait in heterotrophic plastomes (Wicke et al., 2013). The lowest GC contents in holoparasitic plastome could be well-explained by this hypothesis. However, unexpected slightly high GC values of hemiparasitic plastomes, mainly in LSC, were found in our study (Supplementary Figure S2A). This was not simply ascribed to the sampling bias of limited autotrophs involved in the study when a generally low GC value (<38%) was found in autotrophs of the relative families (Phrymaceae and Paulowniaceae). The slightly increased GC contents were closely linked to the inverted SSC in hemiparasites. The exteriorly poor GC contents in the *Schwalbea-Buchnera-Striga* clade were ascribed to the low GC values in IRs (Supplementary Table S1). IR is a hotspot region of mutation and rearrangement. Frequent contraction/expansion makes a generally AT-riched IR (Wicke et al., 2013) and produces an unstable structure in plastome. The process of GC-based gene conversion (gBGC) prefers repairing DNA mismatches with GC bases and elevating GC contents over evolutionary time (Niu et al., 2017). It shapes both the plastome-wide mutations and GC contents as a complementary mechanism

for maintaining the structure and function of plastome with rearranged sequences.

CUB toward AT driven by ultraviolet-radiation and DNA methylation was found in all Orobanchaceae species, also commonly in the chloroplast genomes of land plants (Ossowski et al., 2010; Qi et al., 2015). CUB was primarily determined by the dynamic balance associated with mutations and natural selection, as well as the structure and functions of genes, gene expression, and many other factors in the evolutionary process (Bulmer, 1991; Chiapello et al., 1998; Wong et al., 2002). In our study, the scattered distribution of ENc values in many genes, far from the expected curves, would not be ascribed to mutation pressure alone, implying other factors, for example, natural selection might have effects on CUB as reported in many plants (Liu and Xue, 2005) (Supplementary Figure S5). This selective pressure is considered to promote the translation efficiency for the chloroplast genomes (Suzuki and Morton, 2016; Zhang et al., 2018). Most genes in autotrophic plastomes were possibly driven by selective pressures to adapt the efficient functional express in the process of photosynthesis. But in heterotrophs, the changes in the mode of nutrients acquisition relaxed the selective constraints and compositional mutation of GC to AT transition (Krasovec and Filatov, 2019) became the main factor shaping CUB, finally resulting in the prominent decrease of the GC3s usage as observed in holoparasitic plastomes and supported by the non-significant shift of ENc ratios (data not shown). In our study, we observed an unexpected slight higher GC3s in hemiparasites than those in autotrophs. This finding was associated with the potentially negative or positive selection for photosynthesis genes in autotrophic and hemiparasitic plastomes, respectively (Supplementary Figure S5). We hypothesized that along with the change of lifestyles, the opposite selection in hemiparasites perhaps drove a shift of CUB toward GC. The changes in the direction of selection for codon usage was also found diatoms (Krasovec and Filatov, 2019). But it is still unknown that to what extent, the interaction of mutation pressure and natural selection influence on the patterns of codon usage. Therefore, this hypothesis of the preferential codon usage under natural selection should be further investigated with more measures. An alternative was that the increase of GC3 was a balance of mutation pressure and CUB toward AT. As discussed above in GC content variation, gBGC might elevate GC contents across the whole plastome and possibly counteracted the effect of CUB toward AT. Because no mutation rate and gBGC were examined here, the hypothesis needs to test in future.

Relaxed selective constraint was usually considered as the main factor driving gene more divergent in heterotrophs than in autotrophs (Wicke et al., 2011). The codon-based phylogenetic framework was used to test the selective pressure between autotrophic and heterotrophic plastomes in Orobanchaceae. We uncovered that *rpl2* gene had undergone relaxed selection, indicating an early accelerated base substitution rate along with the evolutionary history of heterotrophs in the lineage (Wicke et al., 2011). The gene *rpl2* located in the LSC/IRb junction and might be affected by the expansion/contraction of IR. And the relaxed purifying selection is potentially linked to the lowered photosynthesis rate in heterotrophs (Wicke et al., 2014). We

analyzed the selective pressures of the functional genes retained in all Orobanchaceae species and identified three genes (*rps2*, *rps14*, and *rpl36*) were subject to positive selection and only *rps2* gene had positive selection sites (**Supplementary Table S6** and **Supplementary Figure S6**). These genes encode the large or small subunit ribosomal protein together with other genes to ensure the process of initiation of protein synthesis. The genes of *rps2* and *rps14* are essential components for the plastid translation. Despite the non-essential component for ribosome, the loss of *rpl36* would cause a severe mutation phenotype (Tiller and Bock, 2014). Consequently, these genes (especially *rps2*) under natural selection play an important role in the evolution and divergence of Orobanchaceae.

Phylogenetic Relationships

Phylogenetic trees of Orobanchaceae using chloroplast or ITS gene fragments indicated that there was a deep split between heterotrophic and autotrophic plants, i.e., parasitic species were monophyletic (Olmstead et al., 2001). Our analysis based on plastomes drew a similar conclusion (**Figure 2**). The holoparasitic *Lathraea squamaria* is always clustered with hemiparasites, suggesting the multiple origins of holoparasitic species (Young et al., 1999; Olmstead et al., 2001; McNeal et al., 2013). *Pedicularis* is monophyletic and related to the clade of *Aureolaria-Triphysaria-Castilleja* (**Figures 2, 3**). This is not completely consistent with previous studies (Robart et al., 2015; Yu et al., 2018), mainly due to limited hemiparasitic plastomes available in our study. The future work on plastomes involving more basal branches, such as *Agalinis*, *Seymeria* and *Phtheirospermum*, will shed further lights on the evolutionary history of Orobanchaceae. As for *Pedicularis*, recent molecular phylogenetic works showed that most groups based on morphological traits were not monophyletic (Tkach et al., 2014; Robart et al., 2015; Yu et al., 2018). In this study, *P. ishidoyana* was the basal taxon, which was consistent with morphological traits and molecular data (Cho et al., 2018). But it is hard to compare the interspecific relationships of other species to the previous phylogenetic studies (Tkach et al., 2014; Yu et al., 2018) for the lack of plastomes in most *Pedicularis* species. More samples are urgently needed to clarify to phylogenetic relationships of this divergent genus in future.

CONCLUSION

The changes in lifestyle could trigger the degradation of plastomes in heterotrophs. Here, we compared the plastomes of 50 Orobanchaceae species in order to elucidate the sequence variation pattern in the transition from autotrophic to holoparasitic lifestyles. Plastomes of hemiparasites exhibited much more similarities to the autotrophic relatives with a few pseudogenized/lost *ndh* and tRNA genes, suggesting that these species were just at the very early stage of transition from autotrophs to heterotrophs. Both structural rearrangement and relaxed selective constraints contributed to the changes of GC content and gene order along with the transit of lifestyles. Due to the clade specific evolution of plastomes in many aspects,

more samples from the monophyletic hemiparasitic genus, such as *Pedicularis* and *Striga*, are necessarily needed to investigate the divergently evolutionary trajectories of chloroplast genomes in Orobanchaceae.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the GenBank accession numbers: MH703578–MH703581.

AUTHOR CONTRIBUTIONS

Z-LL and RZ conceived and designed the experiments. JL, QY, and FP performed the experiments. RZ and BX analyzed the data. ZZ, JH, and YL prepared the samples. RZ wrote the manuscript. Z-LL helped to revise the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was financially supported by the National Natural Science Foundation of China (31670219 and 31870321), National Basic Resources Survey Project of China (2017FY100200), and Graduate Student Independent Innovation Project of Northwest University (YZZ17165).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.542017/full#supplementary-material>

Supplementary Figure 1 | Comparison of the borders of LSC, SSC and IR regions in the plastomes of *Pedicularis* species with *Lindenbergia philippensis* as the reference. Ψ indicates a pseudogene. This figure is not to scale.

Supplementary Figure 2 | GC (A) and GC3s (B) contents in different regions of plastomes. Significant test values were labeled (ns, non-significant; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

Supplementary Figure 3 | Sliding window analysis of the whole chloroplast genomes of autotrophs (A) and hemiparasites (B). X-axis: position of the midpoint of a window, Y-axis: nucleotide diversity of each window.

Supplementary Figure 4 | Collinearity of plastomes of autotrophs (A), hemiparasites (B), and holoparasites (C) in Orobanchaceae using the MAUVE algorithm with *Lindenbergia philippensis* as the reference. Local collinear blocks are colored in this figure to indicate syntenic regions, while histograms within each block represent the degree of sequence similarity.

Supplementary Figure 5 | ENc-plots (ENc vs. GC3s) for different gene groups of 50 plastomes in Orobanchaceae.

Supplementary Figure 6 | The *dN/dS* ratio of *rpl36*, *rps2* and *rps14* gene in Orobanchaceae plastomes. Green – autotrophic species, blue – hemiparasitic species, and orange – holoparasitic species.

Supplementary Table 1 | Chloroplast genome characteristics of 50 taxa in Orobanchaceae.

Supplementary Table 2 | Gene contents of four *Pedicularis* plastomes.

Supplementary Table 3 | The number of variable sites and nucleotide diversity in plastomes of Orobanchaceae.

Supplementary Table 4 | Positive selection sites estimated with Maximum Likelihood parameter in *rps2* gene.

REFERENCES

- Balakirev, E. S., and Ayala, F. J. (2003). Pseudogenes: are they "Junk" or functional DNA? *Annu. Rev. Genet.* 37, 123–151. doi: 10.1146/annurev.genet.37.040103.103949
- Barrett, C. F., and Davis, J. I. (2012). The plastid genome of the mycoheterotrophic *Corallorhiza striata* (Orchidaceae) is in the relatively early stages of degradation. *Am. J. Bot.* 99, 1513–1523. doi: 10.3732/ajb.1200256
- Barrett, C. F., Specht, C. D., Leebens-Mack, J., Stevenson, D. W., Zomlefer, W. B., and Davis, J. I. (2014). Resolving ancient radiations: can complete plastid gene sets elucidate deep relationships among the tropical ginger (Zingiberales)? *Ann. Bot.* 113, 119–133. doi: 10.1093/aob/mct264
- Braukmann, T. W. A., Broe, M. B., Stefanovic, S., and Freudenstein, J. V. (2017). On the brink: the highly reduced plastomes of nonphotosynthetic Ericaceae. *New Phytol.* 216, 254–266. doi: 10.1111/nph.14681
- Bulmer, M. (1991). The selection-mutation-drift theory of synonymous codon usage. *Genetics* 129, 897–907.
- Byng, J. W., Chase, M. W., Christenhusz, M. J. M., Fay, M. F., Judd, W. S., Mabberley, D. J., et al. (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linnean Soc.* 181, 1–20. doi: 10.1111/boj.12385
- Chiappello, H., Lisacek, F., Caboche, M., and Henaut, A. (1998). Codon usage and gene function are related in sequences of *Arabidopsis thaliana*. *Gene* 209, GC1–GC38.
- Cho, W. B., Choi, B. H., Kim, J. H., Lee, D. H., and Lee, J. H. (2018). Complete plastome sequencing reveals an extremely diminished SSC region in hemiparasitic *Pedicularis ishidoyana* (Orobanchaceae). *Ann. Bot. Fennici* 55, 171–183. doi: 10.5735/085.055.0122
- Cusimano, N., and Wicke, S. (2016). Massive intracellular gene transfer during plastid genome reduction in nongreen Orobanchaceae. *New Phytol.* 210, 680–693. doi: 10.1111/nph.13784
- Darling, A. C. E., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704
- Frailey, D. C., Chaluvadi, S. R., Vaughn, J. N., Coatney, C. G., and Bennetzen, J. L. (2018). Gene loss and genome rearrangement in the plastids of five hemiparasites in the family Orobanchaceae. *BMC Plant Biol.* 18:30. doi: 10.1186/s12870-018-1249-x
- Graham, S. W., Lam, V. K. Y., and Merckx, V. S. F. T. (2017). Plastomes on the edge: the evolutionary breakdown of mycoheterotroph plastid genomes. *New Phytol.* 214, 48–55. doi: 10.1111/nph.14398
- Hahn, C., Bachmann, L., and Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Res.* 41:e129. doi: 10.1093/nar/gkt371
- Huang, H., Shi, C., Liu, Y., Mao, S. Y., and Gao, L. Z. (2014). Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. *BMC Evol. Biol.* 14:151. doi: 10.1186/1471-2148-14-151
- Ivanova, Z., Sablok, G., Daskalova, E., Zahmanova, G., Apostolova, E., Yahubyan, G., et al. (2017). Chloroplast genome analysis of resurrection tertiary relict *Haberlea rhodopensis* highlights genes important for desiccation stress response. *Front. Plant Sci.* 8:204. doi: 10.3389/fpls.2017.00204
- Jansen, R. K., and Palmer, J. D. (1987). Chloroplast DNA from lettuce and *Barnadesia* (Asteraceae): structure, gene localization, and characterization of a large inversion. *Curr. Genet.* 11, 553–564. doi: 10.1007/bf00384619
- Jansen, R. K., and Ruhlman, T. A. (2012). "Plastid genomes of seed plants," in *Genomics of Chloroplasts and Mitochondria*, eds R. Bock and V. Knoop (Dordrecht: Springer), 103–126. doi: 10.1007/978-94-007-2920-9_5
- Jansen, R. K., Wojciechowski, M. F., Sanniyasi, E., Lee, S. B., and Daniell, H. (2008). Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of *rps12* and *clpP* intron losses among legumes (Leguminosae). *Mol. Phylogenet. Evol.* 48, 1204–1217. doi: 10.1016/j.ympev.2008.06.013
- Katoh, K., and Standley, D. M. (2013). MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kawabe, A., and Miyashita, N. T. (2003). Patterns of codon usage bias in three dicot and four monocot plant species. *Genes Genet. Syst.* 78, 343–352. doi: 10.1266/ggs.78.343
- Kim, H. T., Kim, J. S., Moore, M. J., Neubig, K. M., Williams, N. H., Whitten, W. M., et al. (2015). Seven new complete plastome sequences reveal rampant independent loss of the *ndh* gene family across orchids and associated instability of the inverted repeat/small single-copy region boundaries. *PLoS One* 10:e0142215. doi: 10.1371/journal.pone.0142215
- Kode, V., Mudd, E. A., Iamtham, S., and Day, A. (2005). The tobacco plastid accD gene is essential and is required for leaf development. *Plant J.* 44, 237–244. doi: 10.1111/j.1365-313x.2005.02533.x
- Krasovec, M., and Filatov, D. A. (2019). Evolution of codon usage bias in diatoms. *Genes* 10:894. doi: 10.3390/genes10110894
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., and Calcott, B. (2017). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* 34, 772–773.
- Li, A. R., and Guan, K. Y. (2008). Arbuscular mycorrhizal fungi may serve as another nutrient strategy for some hemiparasitic species of *Pedicularis* (Orobanchaceae). *Mycorrhiza* 18, 429–436. doi: 10.1007/s00572-008-0196-z
- Li, Y., Zhou, J. G., Chen, X. L., Cui, Y. X., Xu, Z. C., Li, Y. H., et al. (2017). Gene losses and partial deletion of small single-copy regions of the chloroplast genomes of two hemiparasitic *Taxillus* species. *Sci. Rep.* 7: 12834.
- Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187
- Liu, C. P., Huang, J. P., Wu, C. S., Hsu, C. Y., and Chaw, S. M. (2010). Comparative chloroplast genomics reveals the evolution of Pinaceae genera and subfamilies. *Genome Biol. Evol.* 2, 504–517. doi: 10.1093/gbe/evq036
- Liu, Q. P., and Xue, Q. Z. (2005). Comparative studies on codon usage pattern of chloroplasts and their host nuclear genes in four plant species. *J. Genet.* 84, 55–62. doi: 10.1007/bf02715890
- Liu, X., Fu, W., Tang, Y., Zhang, W., Song, Z., Li, L., et al. (2020). Diverse trajectories of plastome degradation in holoparasitic *Cistanche* and genomic location of the lost plastid genes. *J. Exp. Bot.* 71, 877–892. doi: 10.1093/jxb/erz456
- McNeal, J. R., Bennett, J. R., Wolfe, A. D., and Mathews, S. (2013). Phylogeny and origins of holoparasitism in Orobanchaceae. *Am. J. Bot.* 100, 971–983. doi: 10.3732/ajb.1200448
- Naumann, J., Der, J. P., Wafula, E. K., Jones, S. S., Wagner, S. T., Honaas, L. A., et al. (2016). Detecting and characterizing the highly divergent plastid genome of the nonphotosynthetic parasitic plant *Hydnora visseri* (Hydnoraceae). *Genome Biol. Evol.* 8, 345–363. doi: 10.1093/gbe/evv256
- Nielsen, R., and Yang, Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148, 929–936.
- Niu, Z. T., Xue, Q. Y., Wang, H., Xie, X. Z., Zhu, S. Y., Liu, W., et al. (2017). Mutational biases and gc-biased gene conversion affect gc content in the plastomes of *Dendrobium* genus. *Int. J. Mol. Sci.* 18:2307. doi: 10.3390/ijms18112307
- Olmstead, R. G., DePamphilis, C. W., Wolfe, A. D., Young, N. D., Elisons, W. J., and Reeves, P. A. (2001). Disintegration of the Scrophulariaceae. *Am. J. Bot.* 88, 348–361. doi: 10.2307/2657024
- Ossowski, S., Schneeberger, K., Lucas-Lledo, J. I., Warthmann, N., Clark, R. M., Shaw, R. G., et al. (2010). The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* 327, 92–94. doi: 10.1126/science.1180677

- Patel, R. K., and Jain, M. (2012). NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* 7:e30619. doi: 10.1371/journal.pone.0030619
- Poliseno, L., Salmena, L., Zhang, J. W., Carver, B., Haveman, W. J., and Pandolfi, P. P. (2010). A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* 465, 1033–U1090.
- Posada, D., and Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: advantages of akaike information criterion and Bayesian approaches over likelihood ratio tests. *Syst. Biol.* 53, 793–808. doi: 10.1080/10635150490522304
- Posada, D., and Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14, 817–818. doi: 10.1093/bioinformatics/14.9.817
- Qi, Y. Y., Xu, W. J., Xing, T., Zhao, M. M., Li, N. N., Yan, L., et al. (2015). Synonymous codon usage bias in the plastid genome is unrelated to gene structure and shows evolutionary heterogeneity. *Evol. Bioinform.* 11, 65–77.
- Rambaut, A. (2006). *FigTree*. Edinburgh: Institute of Evolutionary Biology, University of Edinburgh.
- Ree, R. H. (2005). Phylogeny and the evolution of floral diversity in *Pedicularis* (Orobanchaceae). *Int. J. Plant Sci.* 166, 595–613. doi: 10.1086/430191
- Robart, B. W., Gladys, C., Frank, T., and Kilpatrick, S. (2015). Phylogeny and biogeography of North American and Asian *Pedicularis* (Orobanchaceae). *Syst. Bot.* 40, 229–258. doi: 10.1600/036364415x686549
- Ronquist, F., and Huelsenbeck, J. P. (2003). MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574. doi: 10.1093/bioinformatics/btg180
- Samigullin, T. H., Logacheva, M. D., Penin, A. A., and Vallejo-Roman, C. M. (2016). Complete plastid genome of the recent holoparasite *Lathraea squamaria* reveals earliest stages of plastome reduction in Orobanchaceae. *PLoS One* 11:e0150718. doi: 10.1371/journal.pone.0150718
- Sharp, P. M., and Li, W. H. (1987). The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15, 1281–1295. doi: 10.1093/nar/15.3.1281
- Shin, H. W., and Lee, N. S. (2018). Understanding plastome evolution in hemiparasitic Santalales: complete chloroplast genomes of three species, *Dendrotrophe varians*, *Helixanthera parasitica*, and *Macrosolen cochinchinensis*. *PLoS One* 13:e0200293. doi: 10.1371/journal.pone.0200293
- Silva, S. R., Diaz, Y. C. A., Penha, H. A., Pinheiro, D. G., Fernandes, C. C., Miranda, V. F. O., et al. (2016). The Chloroplast genome of *Utricularia reniformis* sheds light on the evolution of the ndh gene complex of terrestrial carnivorous plants from the Lentibulariaceae family. *PLoS One* 11:e0165176. doi: 10.1371/journal.pone.0165176
- Smith, S. A., and Donoghue, M. J. (2008). Rates of molecular evolution are linked to life history in flowering plants. *Science* 322, 86–89. doi: 10.1126/science.1163197
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Suzuki, H., and Morton, B. R. (2016). Codon adaptation of plastid genes. *PLoS One* 11:e0154306. doi: 10.1371/journal.pone.0154306
- Tiller, N., and Bock, R. (2014). The translational apparatus of plastids and its role in plant development. *Mol. Plant* 7, 1105–1120. doi: 10.1093/mp/ssu022
- Tkach, N., Ree, R. H., Kuss, P., Roser, M., and Hoffmann, M. H. (2014). High mountain origin, phylogenetics, evolution, and niche conservatism of arctic lineages in the hemiparasitic genus *Pedicularis* (Orobanchaceae). *Mol. Phylogenet. Evol.* 76, 75–92. doi: 10.1016/j.ympev.2014.03.004
- Wang, W., and Lanfear, R. (2019). Long-reads reveal that the chloroplast genome exists in two distinct versions in most plants. *Genome Biol. Evol.* 11, 3372–3381.
- Wertheim, J. O., Murrell, B., Smith, M. D., Kosakovsky Pond, S. L., and Scheffler, K. (2015). RELAX: detecting relaxed selection in a phylogenetic framework. *Mol. Biol. Evol.* 32, 820–832. doi: 10.1093/molbev/msu400
- Westwood, J. H., dePamphilis, C. W., Das, M., Fernandez-Aparicio, M., Honaas, L. A., Timko, M. P., et al. (2012). The parasitic plant genome project: new tools for understanding the biology of *Orobanche* and *Striga*. *Weed Sci.* 60, 295–306.
- Wicke, S., Muller, K. F., de Pamphilis, C. W., Quandt, D., Wickett, N. J., Zhang, Y., et al. (2013). Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. *Plant Cell* 25, 3711–3725. doi: 10.1105/tpc.113.113373
- Wicke, S., Muller, K. F., dePamphilis, C. W., Quandt, D., Bellot, S., and Schneeweiss, G. M. (2016). Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. *Proc. Natl. Acad. Sci. U.S.A.* 113, 9045–9050. doi: 10.1073/pnas.1607576113
- Wicke, S., and Naumann, J. (2018). Molecular evolution of plastid genomes in parasitic flowering plants. *Plastid Genome Evol.* 85, 315–347. doi: 10.1016/bs.abr.2017.11.014
- Wicke, S., Schaferhoff, B., dePamphilis, C. W., and Muller, K. F. (2014). Disproportional plastome-wide increase of substitution rates and relaxed purifying selection in genes of carnivorous Lentibulariaceae. *Mol. Biol. Evol.* 31, 529–545. doi: 10.1093/molbev/mst261
- Wicke, S., Schneeweiss, G. M., dePamphilis, C. W., Muller, K. F., and Quandt, D. (2011). The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol. Biol.* 76, 273–297. doi: 10.1007/s11103-011-9762-4
- Williams, A. V., Boykin, L. M., Howell, K. A., Nevill, P. G., and Small, I. (2015). The complete sequence of the *Acacia ligulata* chloroplast genome reveals a highly divergent clpP1 gene. *PLoS One* 10:e0138367. doi: 10.1371/journal.pone.0125768
- Wong, G. K. S., Wang, J., Tao, L., Tan, J., Zhang, J. G., Passey, D. A., et al. (2002). Compositional gradients in Gramineae genes. *Genome Res.* 12, 851–856. doi: 10.1101/gr.189102
- Wright, F. (1990). The 'effective number of codons' used in a gene. *Gene* 87, 23–29. doi: 10.1016/0378-1119(90)90491-9
- Wu, C. S., Wang, T. J., Wu, C. W., Wang, Y. N., and Chaw, S. M. (2017). Plastome evolution in the sole hemiparasitic genus laurel dodder (*Cassytha*) and insights into the plastid phylogenomics of Lauraceae. *Genome Biol. Evol.* 9, 2604–2614. doi: 10.1093/gbe/evx177
- Wyman, S. K., Jansen, R. K., and Boore, J. L. (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20, 3252–3255. doi: 10.1093/bioinformatics/bth352
- Yang, Z. H. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088
- Yang, Z. H., and Nielsen, R. (2002). Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* 19, 908–917. doi: 10.1093/oxfordjournals.molbev.a004148
- Young, N. D., Steiner, K. E., and dePamphilis, C. W. (1999). The evolution of parasitism in Scrophulariaceae/Orobanchaceae: plastid gene sequences refute an evolutionary transition series. *Ann. Missouri Bot. Garden* 86, 876–893. doi: 10.2307/2666173
- Yu, W. B., Liu, M. L., Wang, H., Mill, R. R., Ree, R. H., Yang, J. B., et al. (2015). Towards a comprehensive phylogeny of the large temperate genus *Pedicularis* (Orobanchaceae), with an emphasis on species from the Himalaya-Hengduan Mountains. *BMC Plant Biol.* 15:176. doi: 10.1186/s12870-015-0547-9
- Yu, W. B., Randle, C. P., Lu, L., Wang, H., Yang, J. B., depamphilis, C. W., et al. (2018). The hemiparasitic plant *Phtheirospermum* (Orobanchaceae) is polyphyletic and contains cryptic species in the Hengduan Mountains of Southwest China. *Front. Plant Sci.* 9:142. doi: 10.3389/fpls.2018.00142
- Zeng, S. Y., Zhou, T., Han, K., Yang, Y. C., Zhao, J. H., and Liu, Z. L. (2017). The complete chloroplast genome sequences of six Rehmannia species. *Genes* 8:103. doi: 10.3390/genes8030103
- Zhang, R. Z., Zhang, L., Wang, W., Zhang, Z., Du, H. H., Qu, Z., et al. (2018). Differences in codon usage bias between photosynthesis-related genes and genetic system-related genes of chloroplast genomes in cultivated and wild Solanum species. *Int. J. Mol. Sci.* 19:3142. doi: 10.3390/ijms19103142
- Zhang, Z., Li, J., Zhao, X. Q., Wang, J., Wong, G. K., and Yu, J. (2006). KaKs-calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinform.* 4, 259–263. doi: 10.1016/s1672-0229(07)60007-2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Zhang, Xu, Li, Zhao, Han, Lei, Yang, Peng and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.