

# Complete chloroplast genomes of hemiparasitic genus *Cymbaria*: Insights into comparative analysis, development of molecular markers, and phylogenetic relationships

Yang Ma<sup>1</sup>, Jordi López-Pujol<sup>2</sup>, Dongqing Yan<sup>1</sup>, Zhen Zhou<sup>1</sup>, Zekun Deng<sup>1</sup>, and Jianming Niu<sup>1</sup>

<sup>1</sup>Inner Mongolia University

<sup>2</sup>Botanic Institute of Barcelona

October 9, 2023

## Abstract

*Cymbaria* L. is a small genus nearly endemic to the Mongolian Plateau with only two hemiparasitic species. *Cymbaria daurica* L. is an important herb “Xinba” in traditional Mongolian medicine. However, *C. daurica* is often confused with its sister species *Cymbaria mongolica* Maxim. because they only differ in anther morphology. In addition, phylogenetic placement and reductive evolution of this genus within the Orobanchaceae remain poorly resolved. Here, we firstly sequenced two *Cymbaria* chloroplast genomes and then conducted a comparative analysis among those of 54 Orobanchaceae species. We found that the chloroplast genomes of *Cymbaria* species were characterized by pseudogenization/loss events of eight *ndh* genes and a unique *rbcL-matK* inversion. Unlike the high-variability observed in holoparasites, *Cymbaria* and other hemiparasites exhibited high similarity to autotrophs in genome size, GC content, and intact genes. Notely, four pairs of specific DNA barcodes (*CymN1*, *CymN2*, *CymY*, and *CymR*) were developed to distinguish the herb *C. daurica* from its adulterant *C. mongolica*. Phylogenetic analyses revealed that genus *Cymbaria* and the *Schwalbea-Siphonostegia* clade were grouped into the tribe *Cymbarieae*. This tribe formed an independent sister clade to the remaining parasitic lineages. Furthermore, the diversification of monophyletic *Cymbaria* occurred during the late Miocene (6.72 Mya) in the Mongol-Chinese steppe region. These findings provide valuable information for regarding taxonomy, phylogeny and evolution of *Cymbaria* and can be used to validate the authenticity of the herb “Xinba”.

## RESEARCH ARTICLE

Complete chloroplast genomes of hemiparasitic genus *Cymbaria* : Insights into comparative analysis, development of molecular markers, and phylogenetic relationships

Yang Ma <sup>1</sup>, Jordi López-Pujol <sup>2, 3</sup>, Dongqing Yan <sup>1</sup>, Zhen Zhou <sup>1</sup>, Zekun Deng <sup>1</sup> and Jianming Niu <sup>1, 4, 5</sup>

<sup>1</sup> School of Ecology and Environment, Inner Mongolia University, Hohhot 010020, China

<sup>2</sup> Botanic Institute of Barcelona (IBB), CSIC-Ajuntament de Barcelona, Barcelona 08038, Spain

<sup>3</sup> Escuela de Ciencias Ambientales, Universidad Espíritu Santo (UEES), Samborondón 091650, Ecuador

<sup>4</sup> Ministry of Education Key Laboratory of Ecology and Resource Use of the Mongolian Plateau, Hohhot 010020, China

<sup>5</sup> Inner Mongolia Key Laboratory of Grassland Ecology and the Candidate State Key Laboratory of Ministry of Science and Technology, Hohhot 010020, China

Correspondence

Jianming Niu, School of Ecology and Environment, Inner Mongolia University, Hohhot 010020, China; Ministry of Education Key Laboratory of Ecology and Resource Use of the Mongolian Plateau, Hohhot 010020, China; Inner Mongolia Key Laboratory of Grassland Ecology and the Candidate State Key Laboratory of Ministry of Science and Technology, Hohhot 010020, China

Email: 111964401@imu.edu.cn

#### Funding information

Major Science and Technology Projects of Inner Mongolia Autonomous Region, Grant/ Award Number: 2019ZD008; National Natural Science Foundation of China, Grant/ Award Number: 31860106 and 32260304

#### Abstract

*Cymbaria* L. is a small genus nearly endemic to the Mongolian Plateau with only two hemiparasitic species. *Cymbaria daurica* L. is an important herb “Xinba” in traditional Mongolian medicine. However, *C. daurica* is often confused with its sister species *Cymbaria mongolica* Maxim. because they only differ in anther morphology. In addition, phylogenetic placement and reductive evolution of this genus within the Orobanchaceae remain poorly resolved. Here, we firstly sequenced two *Cymbaria* chloroplast genomes and then conducted a comparative analysis among those of 54 Orobanchaceae species. We found that the chloroplast genomes of *Cymbaria* species were characterized by pseudogenization/loss events of eight *ndh* genes and a unique *rbcL*-*matK* inversion. Unlike the high-variability observed in holoparasites, *Cymbaria* and other hemiparasites exhibited high similarity to autotrophs in genome size, GC content, and intact genes. Notably, four pairs of specific DNA barcodes (CymN1, CymN2, CymY, and CymR) were developed to distinguish the herb *C. daurica* from its adulterant *C. mongolica*. Phylogenetic analyses revealed that genus *Cymbaria* and the *Schwalbea*-*Siphonostegia* clade were grouped into the tribe Cymbarieae. This tribe formed an independent sister clade to the remaining parasitic lineages. Furthermore, the diversification of monophyletic *Cymbaria* occurred during the late Miocene (6.72 Mya) in the Mongol-Chinese steppe region. These findings provide valuable information for regarding taxonomy, phylogeny and evolution of *Cymbaria* and can be used to validate the authenticity of the herb “Xinba”.

#### Keywords

Orobanchaceae, Hemiparasite, Plastome, Mongolian medicinal herb, DNA barcoding, Reductive evolution

#### TAXONOMY CLASSIFICATION

Botany, Genomics, Phylogenetics, Taxonomy

#### 1 | Introduction

Plastids serve as crucial organelles that carry out photosynthesis in green plants. Generally, autotrophic plants are highly conserved in the chloroplast genomes (Wicke, Schneeweiss, dePamphilis, Müller, & Quandt, 2011). By contrast, heterotrophic plants including mycoheterotrophs and parasites exhibit varying degrees of degeneration in their chloroplast genomes (Graham, Lam, & Merckx, 2017). The family Orobanchaceae is recognized as the largest single family of parasites, consisting of 101 genera and over 2100 species with autotrophic, hemiparasitic, and holoparasitic lifestyles (Byng et al., 2016; Nickrent, 2020). Orobanchaceae is thus considered as an excellent model to elaborate the parasitic plants evolution (Westwood et al., 2012). Most previous studies have focused on autotrophs and holoparasites (Xiaoqing Liu et al., 2019; Wicke & Naumann, 2018; Zeng et al., 2017). However, limited investigations have been done on hemiparasites (Xin Li et al., 2021; R. Zhang et al., 2020). The hemiparasitic tribe Cymbarieae is widely distributed in Eurasia and comprises approximately 20 species in five genera (Schneeweiss, 2013). Cymbarieae was originally thought to be sister to various parasitic lineages in the family Orobanchaceae (Bennett & Mathews, 2006; McNeal, Bennett, Wolfe, & Mathews, 2013), while analysis of the tribe Orobanchaeae suggests that this original classification merits reconsideration (Xi Li, Feng, Randle, & Schneeweiss, 2019). Furthermore, although chloroplast genomes of *Schwalbea americana* (Wicke et al., 2013) and *Siphonostegia chinensis* (Gao et al., 2019; Jiang et al., 2022)

have been previously published, no systematic comparative analysis of chloroplast genomes of various groups within the tribe Cymbarieae has been conducted to date.

*Cymbaria* L. *sensu stricto* (i.e., excluding *Cymbochasma* Endl.) is the type genus of the tribe Cymbarieae and usually parasitizes on roots of *Stipa* (Poaceae) and *Caragana* (Fabaceae). It is now generally believed that this genus includes only two facultative hemiparasites (Zhao, 1999). *C. mongolica* Maxim. is endemic to the Loess Plateau of China, while *C. daurica* L. is a characteristic species of Mongolian Plateau steppe. According to the Chinese Pharmacopoeia and the Mongolian Medicinal Materials Standard, *C. daurica* is the only original plant species for elaborating the traditional Mongolian medicine, which is referred to as “Kanba-Arong” in Mongolian and “Xinba” in Chinese. It is widely used for treating several diseases, including pruritus, psoriasis, fetotoxicity, impetigo, and diabetes (Zhang et al., 2013). Its historical application dates back to the 19th-century Mongolian pharmaceutical classic known as *Mengyaozhengdian* (Fig. S1), which provides a systematic approach to its usage. The dried whole plant is included in several classical herbal formulations, such as “Siweixinbasan” and “Baweixinbasan.” Previous studies have shown that *C. daurica* contains 177 chemical components (Wu et al., 2020), and the most common are flavonoids (Z.-H. Li et al., 2014) and iridoid glycosides (Q. Wang, Bao, Hao, & Han, 2018). This herb has received much recent attention owing to its diverse bioactivities, including anti-diabetic (Gong et al., 2020), anti-inflammatory (J.-J. Guo, Liu, Zhu, Ren, & Liang, 2017), and anti-bacterial (Shi et al., 2020) activities. However, either accidental or intentional, the adulteration and substitution of *C. daurica* frequently occurs in China and Mongolia, due to its minor morphological differences with *C. mongolica* (Hu, 2018; Liang, Liang, Yao, Liu, & Li, 2016). Specifically, *C. daurica* has densely white sericeous anther locules that are 4–4.5 mm and apically pilose, whereas *C. mongolica* has pilose anther locules that are 3–3.6 mm and glabrous apically or occasionally with few hairs; thus, distinguishing between these two species is a major challenge in the non-flowering stage. This issue has an unforeseeable subsequent effect on the herb “Xinba”, jeopardizing its clinical use, potency, and safety.

In recent years, researchers have shifted their focus from morphological and chemical identification to using DNA-based molecular markers as a precise method to assess the authenticity of medicinal herbs. DNA barcodes has become particularly popular for identifying species of Chinese medicinal herbs owing to its accuracy and speed (Zhu, Liu, Qiu, Dai, & Gao, 2022). The universal barcodes include ITS, *matK*, *rbc L*, and *trn H-psb A*, either individually or in combination (CBOL Plant Working Group et al., 2009; China Plant BOL Group et al., 2011). However, these barcodes might be ineffective for complex taxonomic groups, especially for radically evolved and closely related taxa, because sufficient genetic variation is lacking (Z. F. Liu et al., 2022). Therefore, more reliable specific DNA barcodes are needed to distinguish the *C. daurica* from its adulterant *C. mongolica*, and ensure the authenticity of the herb “Xinba”.

Here, we provide the chloroplast genome sequences of these two *Cymbaria* species and then conducted a comparative analysis of these two newly sequenced chloroplast genomes with those of 52 other Orobanchaceae species. Our specific objectives were to (1) characterize *Cymbaria* chloroplast genomes, (2) develop the specific molecular markers as DNA barcodes, and (3) investigate the evolution and phylogeny of *Cymbaria*. These findings will provide key insights on the taxonomic identification, phylogenetic placement and reductive evolution of hemiparasitic genus *Cymbaria*, and valuable genetic tools to validate the authenticity of the traditional Mongolian medicine “Xinba.”

## 2. | Materials and Methods

### 2.1. | Sampling, sequencing, assembly, and annotation

Wild *C. mongolica* and *C. daurica* plants were collected from Binzhou, Shaanxi Province (35°14'38.63"N, 108deg13'06.11"E) and Xilingol, Inner Mongolia Autonomous Region (43deg28'13.92"N, 116deg47'07.76"E) in China. One sample of young fresh leaves per species was collected in a liquid nitrogen tank. The habitat and altitude were also recorded. Voucher specimens were stored in the Herbarium of Inner Mongolia University. Total genomic DNA was extracted using the modified CTAB method (Allen, Flores-Vergara, Krasynanski, Kumar, & Thompson, 2006). Illumina sequencing platform was used to generate raw reads.

The assembly was performed using clean data through SPAdes v. 3.14.0 (Bankevich et al., 2012) and GetOrganelle (Jin et al., 2020). The graphic visualization and assembly quality was evaluated by Bandage (Wick, Schultz, Zobel, & Holt, 2015) and GetOrganelle (Jin et al., 2020), respectively. Annotations were performed using GeSeq (Tillich et al., 2017) and manually adjusted in Geneious (Kearse et al., 2012). After identifying the boundaries by BLAST, the sequences were submitted to GenBank and visualized as a single circular maps using the Organellar Genome DRAW tool (Greiner, Lehwark, & Bock, 2019).

## 2.2. | Comparative analysis of chloroplast genomes

A comparative analysis of Orobanchaceae chloroplast genomes was conducted including eight autotrophs, 21 hemiparasites, and 25 holoparasites (Table S1). The genome sizes, GC contents, and intact genes were investigated using PhyloSuite (D. Zhang et al., 2020) and drawn with RadarMap package in R software (Team, 2015). A heatmap of genes in the chloroplast genome was generated using TBtools (Chen et al., 2020). The PAML package v4.0 was used to calculate the nonsynonymous (Ka) and synonymous (Ks) substitution rates and their ratio ( $\omega = \text{Ka/Ks}$ ) (Yang, 2007). Contractions and expansions of the boundaries were detected using Irscope (Amiryousefi, Hyvönen, & Poczai, 2018). The Mauve program was used to align sequences against the *Schwalbea americana* reference sequence (Darling, Mau, & Perna, 2010). CodonW (<http://codonw.sourceforge.net/>) was used to calculate RSCU. REPuter (Kurtz et al., 2001) was employed to detect repeat sequences and simple sequence repeats (SSRs) were identified by MISA (Beier, Thiel, Münch, Scholz, & Mascher, 2017).

## 2.3. | Development and validation of DNA barcodes

Structural comparisons were conducted using mVISTA (Frazer, Pachter, Poliakov, Rubin, & Dubchak, 2004) with *C. mongolica* as a reference. Sliding window analysis was conducted to identify hypervariable regions using DnaSP (Librado & Rozas, 2009). Primer3web v 4.1.0 (<https://primer3.ut.ee/>) was used to design DNA barcoding primers based on the hypervariable regions, and these were verified using seven individuals from different regions of each species (Table S4). PCR was conducted in 25  $\mu\text{L}$  reactions with 12.5  $\mu\text{L}$  of 2 $\times$ EasyTaq PCR SuperMix, 1.0  $\mu\text{L}$  of each primer (0.4  $\mu\text{M}$ ), 1  $\mu\text{L}$  of template DNA, and 9.5  $\mu\text{L}$  of ddH<sub>2</sub>O. A SimpliAmp Thermal Cycler (Applied Biosystems, Carlsbad, CA, USA) was used to conduct all PCR reactions with the following thermal cycling conditions: 94 @C for 5 min; 30 cycles of 94 @C for 30 s, a specific annealing temperature (T<sub>m</sub>) for 30 s, and 72 @C for 30 s; and 72 @C for 10 min. Agarose gel electrophoresis (1.5%) was used to visualize PCR products. The DNA fragments were purified and sequenced by Biomarker Technologies Co., Ltd.

## 2.4. | Phylogenetic analyses and divergence time estimation

To identify early divergence events within Orobanchaceae, phylogeny was inferred from 54 Orobanchaceae species with the two Paulowniaceae species *Tectona grandis* (GenBank accession number: NC\_020098) and *Paulownia tomentosa* (GenBank accession number: MK875778) as outgroups. Phylogenetic trees were generated using three datasets: (1) complete chloroplast genome sequences; (2) coding DNA sequences (CDSs); and (3) shared CDSs (*rps 2*, *rps 7*, *rps 8*, *rps 14*, *rpl 2*, *rpl 16*, *rpl 36*, and *mat K*). The standard phylogenetic workflow in PhyloSuite v1.2.2 (D. Zhang et al., 2020) was used as follows: alignment of multiple sequences with MAFFT v7.313 (Katoh & Standley, 2013), optimization of aligned regions with Gblocks v0.91b (Castresana, 2000), concatenation of datasets with the Concatenate Sequence function, selection of the optimal partitioning scheme with PartitionFinder2, and phylogenetic inference with IQ-TREE v1.6.8 (Nguyen, Schmidt, Von Haeseler, & Minh, 2015) and MrBayes v3.2.6 (Drummond, Suchard, Xie, & Rambaut, 2012). Phylogeny was inferred using Maximum likelihood and Bayesian inference methods. Phylogenetic trees visualized using iTOL v6 (Letunic & Bork, 2021).

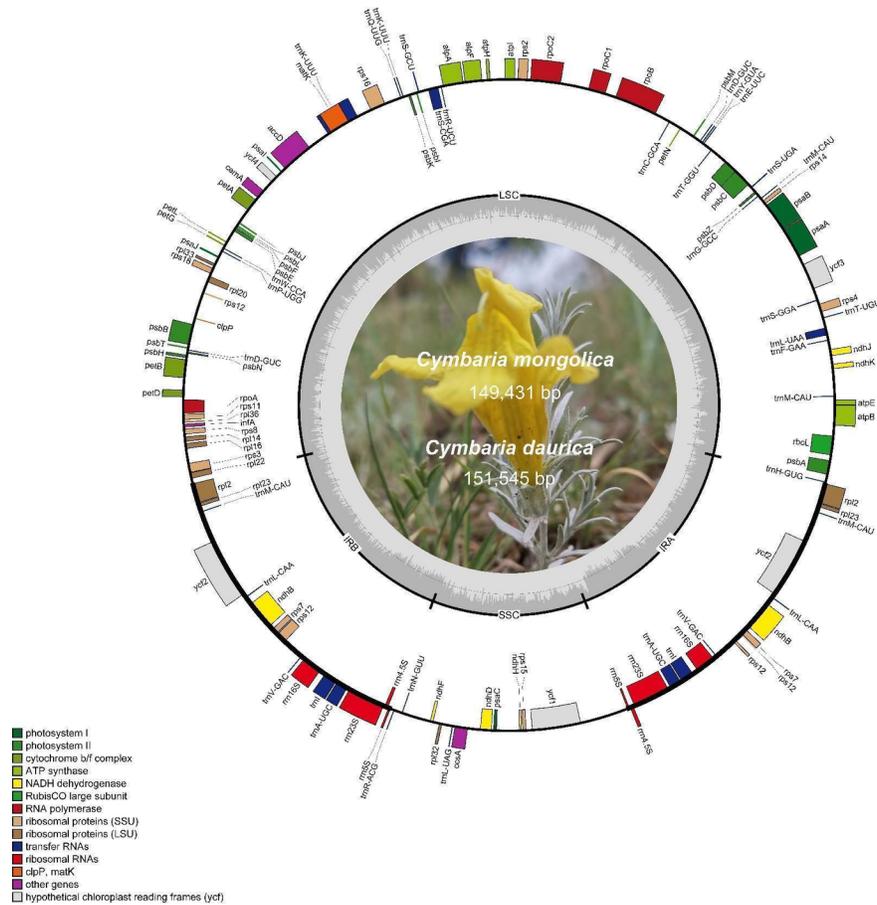
BEAST v1.7 (Drummond et al., 2012) was used to estimate divergence times based on the shared concatenated CDSs. As no reliable fossil record is available, the crown age for Orobanchaceae (C) was constrained as  $56 \pm 10$  Mya using a normal model in TimeTree (<http://timetree.org/>). The results of PartitionFinder were used to determine the nucleotide substitution model of the unlinked subsets. The following parameters were used in BEAUti software: “Lognormal relaxed clock (Uncorrelated)” for “Clock model” and “Speciation: Yule

Process” for “Tree model.” Three independent MCMC chains were carried out under the same parameters. Each MCMC was run for 50,000,000 generations, and sampling was conducted every 5,000 generations. Convergence was evaluated using Tracer v.1.6 (<http://beast.bio.ed.ac.uk/Tracer>). The LogCombiner program was used to combine the three log files. The TreeAnnotator program was used to generate the maximum clade credibility (MCC) tree. The effective sample sizes in the combined three log outputs were greater than 200, which indicated that the MCMC sampling was adequate. FigTree v1.4.3 (Rambaut, 2017) was used to visualize the MCC tree with 95% highest posterior density intervals.

### 3. | Results

#### 3.1. | Characteristics of *Cymbaria* chloroplast genomes

Illumina sequencing yielded more than 20 million bp clean reads. The chloroplast genomes of *C. mongolica* and *C. daurica* displayed a typical quadripartite structure with sizes of 149,431 bp (38.0% GC content) and 151,545 bp (38.2% GC content), respectively. Each chloroplast genome comprised an LSC region (86,595 bp and 87,376 bp), an SSC region (16,962 bp and 17,825 bp), and two IR regions (22,937 bp and 23,172 bp). The GC content in the IR region (44.1%) was higher than that in the LSC (35.9–36.2%) and SSC (32.2–32.7%) regions.

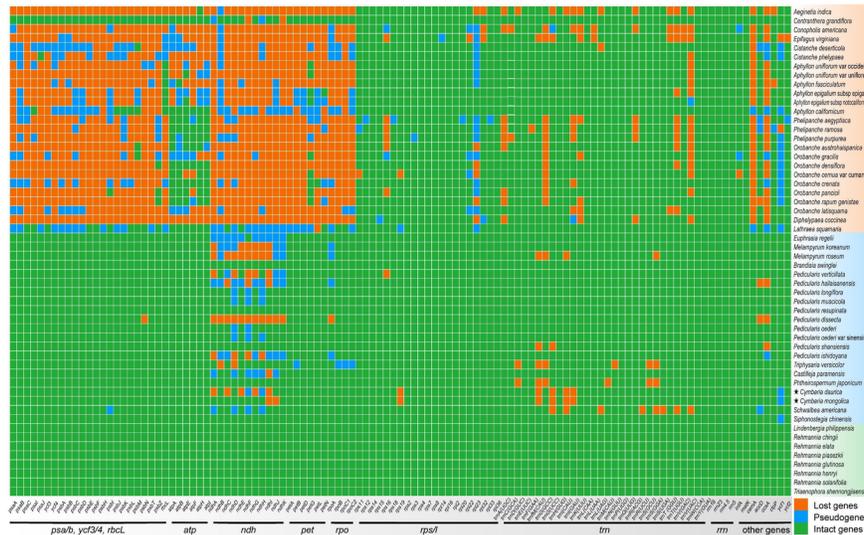


**Figure 1.** Circular maps of *Cymbaria* chloroplast genomes.

The chloroplast genome sequences of *C. mongolica* and *C. daurica* contained 105 and 100 intact genes, respectively (Fig. 1; Table S1). *Cymbaria* contained 26 tRNA and four rRNA genes, while the number of PCGs and pseudogenes varied between two species. In contrast to the autotrophic *Lindenbergia philippensis*

, *C. mongolica* encoded 75 PCGs, and one pseudogene and lacked seven genes while *C. daurica* encoded 70 PCGs, and three pseudogenes and lacked ten genes. Among the 15 intron-containing genes, 12 genes contained one intron, and the remaining three genes harbored two introns (Table S1). In addition, purifying or neutral selection was detected on all PCGs, with the exception of *ycf 2*, which was under positive selection. No pronounced differences were detected in the boundary regions. An inversion of large gene blocks (*rbc L-mat K*) was identified in the LSC region according to the Mauve alignment (Fig. S2). Four pairs of palindromic repeats (123, 67, 48, and 37 bp) and three pairs of palindromic repeats (190, 152, and 121 bp) were detected at both ends of the inverted region of *C. mongolica* (54,654 bp size) and *C. daurica* (55,904 bp size), respectively.

We estimated the sequence divergence and gene content for 54 chloroplast genomes within the family Orobanchaceae (Fig. 2; Table S2). The genome sizes of the autotrophs (153,622–155,319 kb, mean: 154,213 kb) and hemiparasites (142,733–160,910 kb, mean: 151,152 kb) were similar; however, the genome sizes of the autotrophs were higher than those of the holoparasites (45,673–150,504 kb, mean: 87,505 kb) (Fig. S3). The GC content was lower in holoparasites (mean: 34.93%) than in those of hemiparasites (38.25% on average) and autotrophs (mean: 37.90%). Autotrophs contained all intact genes; the number of intact genes was lower in hemiparasites than in autotrophs, and this decrease primarily stemmed from the pseudogenization/loss of *ndh* genes (Fig. 2; Table S2). Moreover, non-functionalization and gene loss were observed in most photosynthetic genes (e.g., *psa /psb ,ycf 3/4, ndh* , and *cem A*) in holoparasites.

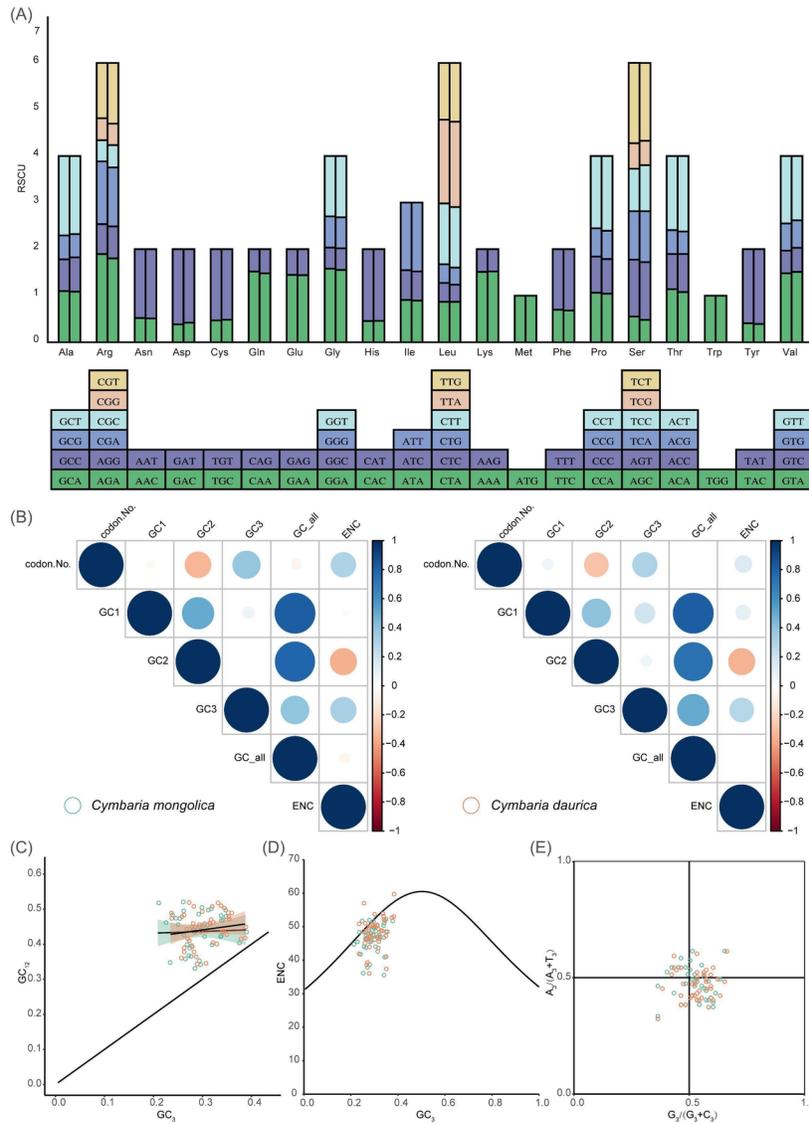


**Figure 2.** Heat map depicting the chloroplast gene content across 54 Orobanchaceae species. Blocks in green, blue, and orange indicate intact genes, pseudogenes, and lost genes, respectively. Background colors in green, blue, and orange indicate autotrophs, hemiparasites, and holoparasites, respectively.

### 3.2. | Codon usage bias

The PCGs in the *C. mongolica* and *C. daurica* comprised 18,482 and 15,896 codons, respectively. Leucine (Leu, 10.50% and 10.62%) was the most common amino acid, and cysteine (Cys, 1.09% and 1.07%) was the least common amino acid. A total of 30 codons (RSCU > 1) were A/T-ending codons, with the exception of UUG (Fig. 3A). The range of the ENC in *C. mongolica* and *C. daurica* was 35.24–56.04 and 35.75–59.41, respectively, which suggested weak codon usage bias. ENC was significantly negatively correlated with GC2 in both species; a significant positive correlation between ENC and GC3 was only detected in *C. mongolica* (Fig. 3B). There was no significant correlation between GC3 and GC12 according to neutral plot analysis (Fig. 3C); the regression coefficient was 0.049 and 0.193 in *C. mongolica* and *C. daurica* , respectively. Most genes were below and around the standard curve according to ENC plot analysis (Fig. 3D), and the ENC

ratio was from 0.05 to 0.15. PR2-plot analysis (Fig. 3E) showed that  $T > A$  and  $G > C$  in the base usage frequency. A total of 16 codons were identified as preferred codons, of which 10 were shared by the two species.

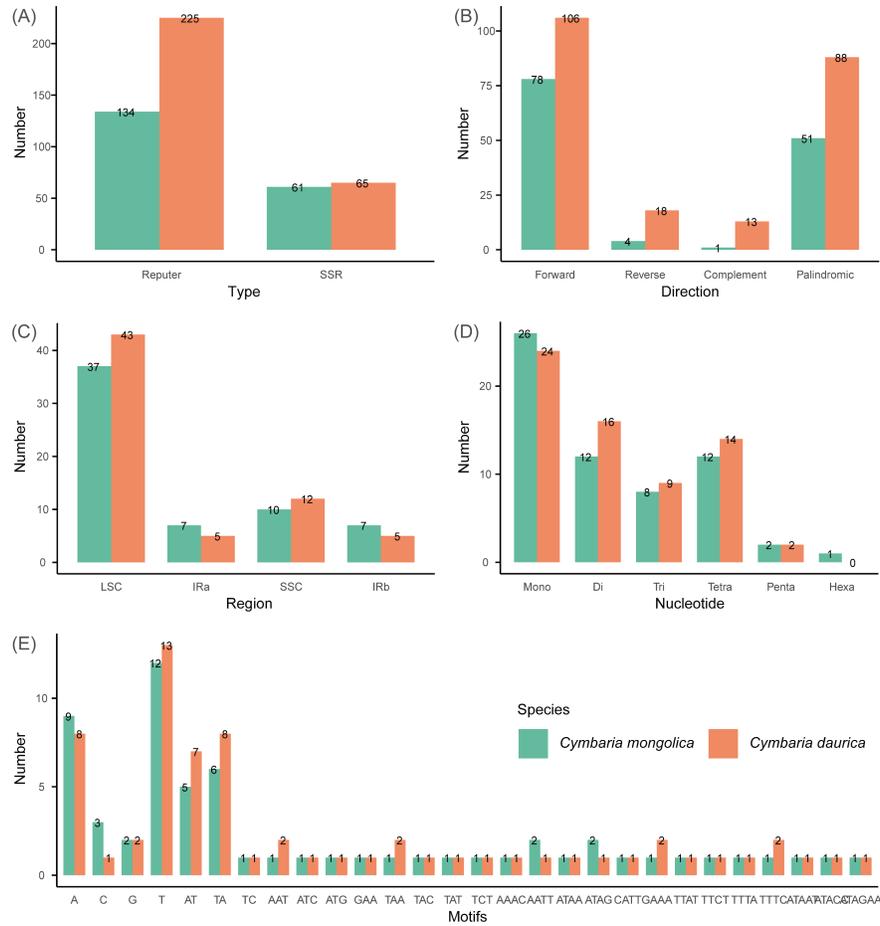


**Figure 3.** Codon usage bias of *Cybaria* chloroplast genomes. (A) RSCU. (B) Correlation heatmap. (C) Neutral plot analysis. (D) PR2 plot analysis. (E) ENC plot analysis. Circle colors of green and orange represent *C. mongolica* (left) and *C. daurica* (right), respectively.

### 3.3. | Repetitive sequence variation

A total of 134 repeats, consisting of 78 forward, four reverse, one complementary, and 51 palindrome repeats, were detected from *C. mongolica* chloroplast genome. Meanwhile, a total of 225 repeats, including 106 forward, 18 reverse, 13 complementary, and 88 palindrome repeats, were identified from *C. daurica* chloroplast genome (Fig. 4A; Fig. 4B). The size of approximately 90% of the repeats ranged from 30 bp to 70 bp. *C. mongolica* and *C. daurica* contained 61 and 65 SSRs, and most were present in the LSC region (37 and 43 SSRs), respectively (Fig. 4A; Fig. 4C). Hexa-nucleotide SSRs were only detected in *C. mongolica*; the

remaining SSRs were identified in both species (Fig. 4D). Mono-nucleotides were the most plentiful, followed by di- and tetra-nucleotides. The mononucleotide motifs A/T had the highest proportion, accounting for 34.4% in *C. mongolica* and 32.3% in *C. daurica* (Fig. 4E).



**Figure 4.** Comparison of the Repeats and SSRs in *Cymbaria* chloroplast genomes. (A) Repeats and SSRs. (B) Repeat types. (C) Occurrences of SSRs. (D) SSR types. (E) SSR motif types. Bar colors of green and orange correspond to *C. mongolica* (left) and *C. daurica* (right), respectively.

### 3.4. | Development and validation of DNA barcodes

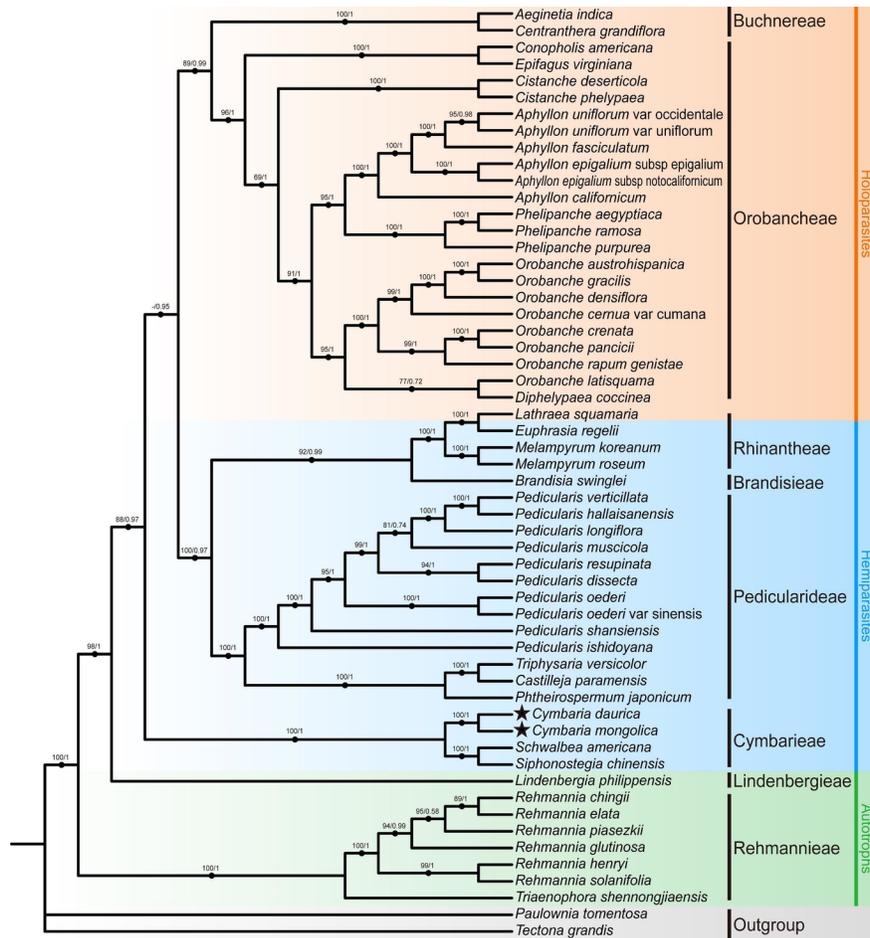
High conservation with some degree of divergence was observed in the two chloroplast genomes (Figure 5). Most of the sequence differences were observed in the non-coding regions. Nucleotide diversity ( $P_i$ ) was 0.02099, and higher divergence was observed in the LSC and SSC regions (Fig. S4). We also detected several divergence hotspot regions ( $P_i > 0.05$ ), including *trn* M-CAU-*ndh* C, *psa* A, *mat* K, *acc* D-*psa* I, *yef* 4-*cem* A, *rpl* 32-*trn* L-UAG, *ndh* D-*ndh* G, *rps* 15-*yef* 1, *rrn* 23S, and *trn* A-UGC-*trn* E-UUC. We designed four specific DNA markers (CymN1, CymN2, CymY, and CymR), and they were validated using sequences from different regions from seven individuals of each species (Table S3; Table S4). DNA fragments varying in length were obtained from *C. mongolica* and *C. daurica* using each primer, and both species could be distinguished via agarose gel electrophoresis (Fig. 6). Likewise, the same results were observed from the sanger sequencing alignment and chromatogram of the amplification of DNA barcodes. Overall, these findings suggest that the four pairs of DNA barcodes could be used to distinguish among *Cymbaria* species.



**Figure 6.** Agarose gel electrophoresis, chromatogram, and alignment of sequences obtained via Sanger sequencing of the amplified DNA barcodes. The indel makers include (A) CymN1, (B) CymN2, (C) CymY, and (D) CymR. The lanes correspond to the PCR products amplified from seven individuals of *C. mongolica* (left) and *C. daurica*(right). The red squares correspond to the Indel regions. The original uncropped image is presented in Figure S5.

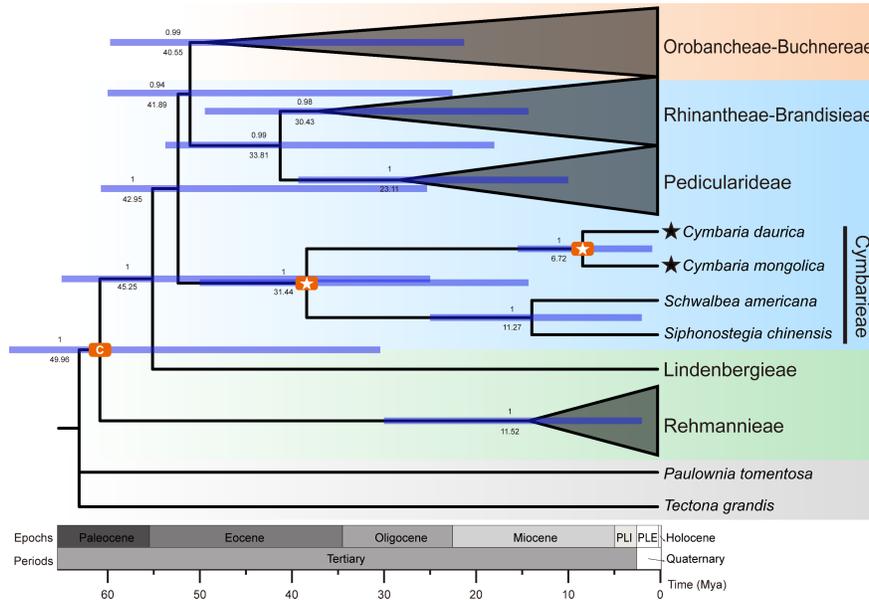
### 3.5. | Phylogenetic relationships and divergence times

Maximum likelihood (ML) and Bayesian inference (BI) phylogenetic analyses yielded highly consistent topologies for the three datasets (Fig. 7). The monophyly of Orobanchaceae was strongly supported. The tribes Rehmannieae, Lindenbergieae, Cymbarieae, Pedicularideae, Brandisieae, Rhinanthaeae, Orobancheae, and Buchnereae corresponded to well-supported clades. The hemiparasitic tribe Cymbarieae was a clade sister to the other parasitic lineages. *C. mongolica* and *C. daurica* were grouped into the monophyletic genus *Cymbaria*, which comprised a clade sister to the *Schwalbea* -*Siphonostegia* clade. Divergence time analyses (Fig. 8) revealed that the common ancestor of the Orobanchaceae originated in the early Eocene (49.96 Mya), and parasites diverged from autotrophic plants in the mid-Eocene (42.95 Mya). The emergence of Cymbarieae predates the mid-Oligocene (31.44 Mya). Moreover, the diversification of *Cymbaria* was estimated to occur around the late Miocene (6.72 Mya).



**Figure 7.** Phylogenetic relationship inferred from ML and BI based on shared protein-coding genes of 54 Orobanchaceae species. Numbers above the branches indicate bootstrap values (left) and posterior

probabilities (right). Background colors of gray, green, blue, and orange indicate outgroups, autotrophs, hemiparasites, and holoparasites, respectively. The stars ( ) indicate the two newly sequenced *Cymbaria* species.



**Figure 8.** Divergence time estimated from the MCC tree in BEAST. Node numbers indicate posterior probabilities (below) and mean divergence times (above). Node bars represent the 95% HPD interval (blue bar). Background colors of gray, green, blue, and orange indicate outgroups, autotrophs, hemiparasites, and holoparasites, respectively. The stars ( ) indicate the two newly sequenced *Cymbaria* species.

## 4. | Discussion

### 4.1. | Pseudogenization/loss events of *ndh* genes and the unique *rbc* L-mat K inversion

It is acknowledged that the lifestyle transition from autotrophy to heterotrophy triggers the degradation of chloroplast genomes (Wicke & Naumann, 2018). Contrasting with the hypervariability of holoparasites, *Cymbaria* species and other Orobanchaceae hemiparasites exhibit high similarity to autotrophs in length, GC content, and intact genes. It has been confirmed that holoparasites have the characteristic of chloroplast genome reduction (Wicke et al., 2013). The high variability of holoparasites is explained by increases in pseudogenization and gene loss. However, patterns of variation in hemiparasites are diverse. The chloroplast genomes of hemiparasites in the family Orobanchaceae were more similar to those of autotrophs, which is in contrast to the reductions in the genome sizes of hemiparasites in the order Santalales (X. Guo, Zhang, Fan, Liu, & Ji, 2021; Y. Li et al., 2017; Shin & Lee, 2018). This might be attributed in part to GC-biased gene conversion and mutational biases, which suggests that sophisticated mechanisms contribute to the stability (Niu et al., 2017).

Angiosperms typically possess 113 plastid genes, consisting of 79 functional PCGs, 30 tRNA, and four rRNA genes (Wicke et al., 2011). Within the Orobanchaceae family, pseudogenes and gene losses were largely absent in autotrophs, occasionally observed in most hemiparasites, and common in nearly all holoparasites. This can be explained by the tendency for the chloroplast genomes of parasites to be reduced in size (Naumann et al., 2016; Wicke & Naumann, 2018). The chloroplast NAD(P)H-dehydrogenase complex comprises 11 *ndh* genes (Ma, Liu, Bai, & Yong, 2021), and the pseudogenization or loss of these genes represents the initial stage of reductive evolution (Wicke & Naumann, 2018). The results of our study indicated that *C. mongolica* lost genes (*ndh* I, *ndh* J) and that *C. daurica* contained pseudogenes (*ndh* F, *ndh* H) and lost

genes (*ndh A*, *ndh C*, *ndh E*, *ndh G*, *ndh I*), indicating that these two species are in the initial phase of the autotroph-to-heterotroph transition. A previous study has confirmed that a hemiparasitic lifestyle can lead to an increase in the pseudogenization/loss of *ndh* genes (Xin Li et al., 2021). This can be explained to some extent by the facultative root hemiparasitic lifestyle of the two *Cymbaria* species. The degradation of *ndh* genes affects several morphological and physiological traits and enhances the adaptation of plants to environmental stress (Sabater, 2021).

The unique inversion including *rbc L-mat K* in the LSC region, which most likely stems from a palindromic repeat-mediated rearrangement. Inversions of the LSC fragments have also been observed in *Schwalbea americana* (Wicke et al., 2013) and *Siphonostegia chinensis* (Jiang et al., 2022); this distinct evolutionary mechanism among Orobanchaceae members might explain the unique phylogenetic position of the tribe Cymbarieae. This inversion has also been observed in *Codonopsis pilosula* subsp. *tangshen* (Yue et al., 2022) and *Avena sativa* (Q. Liu et al., 2020).

The codon usage bias of the two *Cymbaria* species might be affected by natural selection and mutation, as has been observed in several other angiosperms (Q.-F. Lu, Luo, & Huang, 2020). The number of repeats plays a key role in maintaining the stability of in several angiosperms chloroplast genome (Jansen, Saski, Lee, Hansen, & Daniell, 2010). The chloroplast genome of *C. daurica* has more repetitive sequences than that of *C. mongolica*, suggesting that the stability of the former might be higher than that of the latter. Our findings suggest that A/T mononucleotide SSRs were dominant, and this is consistent with the high prevalence of AT richness (Xia Liu, Li, Yang, & Zhou, 2018).

#### 4.2. | Specific DNA barcodes for distinguishing herb *C. daurica* from its adulterant *C. mongolica*.

Traditional Mongolian medicine continues to receive much clinical attention because of its distinctive properties and herb resources (Z. F. Liu et al., 2022). *C. daurica* has been used to treat pruritus, psoriasis, fetotoxicity, impetigo, and diabetes. The incidence of *C. daurica* adulterated with its sister species *C. mongolica* is increasing, and this poses a threat to the clinical efficacy of the herb. The high similarity in morphology between *C. daurica* and *C. mongolica* is the root cause of this problem. Distinguishing between these two *Cymbaria* species is exceedingly difficult because they only differ in anther morphology (Zhang et al., 2013).

Both morphological and microscopic characteristics are currently used to identify *Cymbaria* species (Z.-W. Wang, Dong, Wu, & Li, 2012). However, both of these methods rely on the features of the anther locule, which are only visible during the short flowering period, and specialists are required to obtain accurate identifications. Several divergence hotspot regions were first identified through sequence divergence and nucleotide variability. We then developed and validated four pairs of specific DNA barcodes to distinguish between these two species. Each DNA barcode had at least one Indel locus and several SNP loci. Some previous studies have suggested that specific barcodes are superior to universal barcodes for identifying morphologically similar species (Fang, Dai, Liao, Zhou, & Liu, 2022; G. Y. Lu et al., 2022). Overall, these four pairs of specific DNA barcodes could be used to accurately and rapidly distinguish between *C. daurica* and *C. mongolica* without the need to evaluate the morphological characteristics of the anther during flowering nor specialized training.

#### 4.3. | Climate aridification and increasing host accelerate the diversification of *Cymbaria*

Traditional Orobanchaceae has been merged with all hemiparasitic genera as well as a few holoparasitic genera formerly placed in Scrophulariaceae (Bennett & Mathews, 2006; dePamphilis, Young, & Wolfe, 1997; Fischer, 2004; McNeal et al., 2013; Wolfe, Randle, Liu, & Steiner, 2005; Young, Steiner, & dePamphilis, 1999) and the three autotrophic genera previously considered as sister taxa to Orobanchaceae (Schneeweiss, 2013; Xia, Li, Wen, & Wang, 2021). Moreover, the two new tribes Brandisieae and Pterygielleae have been proposed (Jiang et al., 2022; Yu et al., 2018). To date, Orobanchaceae includes nine well-supported clades corresponding to nine tribes, i.e., the two autotrophic tribes Rehmannieae and Lindenbergeae, and the seven parasitic tribes Cymbarieae, Buchnereae, Orobancheae, Brandisieae, Pedicularideae, Rhinanthaeae, and Pterygielleae. The topology of the major clades and autotroph–parasite sister relationships revealed by

our phylogenetic analyses (Fig. 5) were generally consistent with previous findings.

The hemiparasitic tribe Cymbarieae is distinguished from other tribes in the family Orobanchaceae by the presence of bracteoles, a tubular calyx that is weakly dorsiventral, a highly two-lipped corolla, and anthers with two mostly rounded and equal thecae (Fischer, 2004). Cymbarieae has traditionally been considered sister to all other parasitic lineages (Bennett & Mathews, 2006; McNeal et al., 2013). However, a recent study has challenged this classification after finding that the holoparasitic tribe Orobancheae was sister to all other parasitic members (Xi Li et al., 2019). Our findings are consistent with the traditional classification of the Cymbarieae. The Cymbarieae was found to be sister to the remaining parasitic lineages, and the hemiparasites evolved earlier than the holoparasites; this is consistent with the progressive nature of the evolution of increased host dependence (Xu et al., 2022). This inconsistency in the placement of Cymbarieae might stem from phylogenetic sources of error, such as incomplete lineage sorting or deep coalescence. Resolving these early nodes will require a coalescent approach that involves many genes with different histories.

Within parasitic plants, the hemiparasitic tribe Cymbarieae diverged from the remaining lineages in the mid-Oligocene (31.44 Mya), which was most likely induced by global climate cooling and the retreat of the Tethys Sea during the Eocene–Oligocene Transition at 34 Mya (Abels, Dupont-Nivet, Xiao, Bosboom, & Krijgsman, 2011). The emergence and divergence of hemiparasites might be attributed to the rapid expansion of grasslands during the Oligocene (Torsvik & Cocks, 2016), which would have provided them with opportunities to exploit host plants. *Cymbaria* species diversified in the late Miocene (6.72 Mya), which was driven by the final uplift of the Qinghai–Tibetan Plateau, the onset of the East Asian monsoon, and the large accumulation of dust in the Loess Plateau from 10 to 7 Mya [77–79]. Both climate aridification and the increase in host steppe vegetation (Hurka et al., 2019) likely accelerated the adaptive evolution of *Cymbaria* species in the Mongol–Chinese steppe region.

## 5. | Conclusions

We characterized two *Cymbaria* chloroplast genome and conducted a comparative analysis using chloroplast genomes across 54 Orobanchaceae species. The chloroplast genomes of *C. mongolica* and *C. daurica* had a typical quadripartite structure, and their total lengths were 149,431 bp and 151,545 bp, respectively. Although the chloroplast genomes of holoparasites are hypervariable, the genome size, GC content, and intact genes of *Cymbaria* species, other hemiparasites, and autotrophs are highly similar. The pseudogenization/loss of *ndh* genes might be associated with the facultative root hemiparasitic habits of *Cymbaria*. The *rbcL-matK* inversion in the LSC region most likely stemmed from a palindromic repeat-mediated rearrangement. Specific DNA barcodes were developed using four pairs of primers (CymN1, CymN2, CymY, and CymR) that amplified sequences from the divergent hotspot regions to distinguish the traditional Mongolian herb *C. daurica* from its adulterant *C. mongolica*. The genus *Cymbaria* and the *Schwalbea-Siphonostegia* clade were clustered in the tribe Cymbarieae. This tribe comprised an independent clade sister to the remaining parasitic lineages, which is in contrast to the relationships hypothesized in a recent study. The monophyletic genus *Cymbaria* diversified during the late Miocene period (6.72 Mya). The aridification of the climate and increases in host steppe vegetation likely promoted the adaptive evolution of *Cymbaria* species in the Mongol–Chinese steppe region. Our results provide key information for clarifying the taxonomic identification, phylogenetic placement, and reductive evolution of *Cymbaria*; our findings will also help assessments of the authenticity of the traditional Mongolian medicine “Xinba.”

## Author contributions

Conceptualization, Yang Ma; Data curation, Yang Ma; Formal analysis, Yang Ma; Funding acquisition, Jianming Niu; Investigation, Yang Ma, Zhen Zhou and Zekun Deng; Methodology, Yang Ma; Project administration, Jianming Niu; Resources, Yang Ma, Zhen Zhou and Zekun Deng; Software, Yang Ma; Supervision, Jianming Niu; Validation, Yang Ma; Visualization, Yang Ma; Writing – original draft, Yang Ma; Writing – review & editing, Yang Ma, Jordi López-Pujol, Dongqing Yan and Jianming Niu. All authors have read and agreed to the published version of the manuscript.

## Acknowledgments

This research was funded by the Chinese Scholarship Council, Major Science and Technology Projects of Inner Mongolia Autonomous Region (2019ZD008), and National Natural Science Foundation of China (31860106, 32260304). We also thank Qingcheng Light for providing computing resources and TopEdit LLC for the linguistic editing and proofreading during the preparation of this manuscript.

## CONFLICT OF INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data Availability STATEMENT

Data contained within the article are openly available in GeneBank (NC\_064388 and NC\_064104).

## ORCID

Yang Ma <https://orcid.org/0000-0003-2566-4869>

## References

- Abels, H. A., Dupont-Nivet, G., Xiao, G., Bosboom, R., & Krijgsman, W. (2011). Step-wise change of Asian interior climate preceding the Eocene–Oligocene Transition (EOT). *Palaeogeography, Palaeoclimatology, Palaeoecology*, *299* (3-4), 399-412. doi:10.1016/j.palaeo.2010.11.028
- Allen, G. C., Flores-Vergara, M., Krasynanski, S., Kumar, S., & Thompson, W. (2006). A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nature protocols*, *1* (5), 2320-2325. doi:10.1038/nprot.2006.384
- Amiryousefi, A., Hyvönen, J., & Poczai, P. (2018). IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics (Oxford, England)*, *34* (17), 3030-3031. doi:10.1093/bioinformatics/bty220
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., . . . Prjibelski, A. D. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology*, *19* (5), 455-477. doi:10.1089/cmb.2012.0021
- Beier, S., Thiel, T., Münch, T., Scholz, U., & Mascher, M. (2017). MISA-web: A web server for microsatellite prediction. *Bioinformatics*, *33* (16), 2583-2585. doi:10.1093/bioinformatics/btx198
- Bennett, J. R., & Mathews, S. (2006). Phylogeny of the parasitic plant family Orobanchaceae inferred from phytochrome A. *American Journal of Botany*, *93* (7), 1039-1051.
- Byng, J., Chase, M., Christenhusz, M., Fay, M., Judd, W., Mabberley, D., . . . Stevens, P. (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society*, *181* (1), 1-20. doi:10.1111/boj.12385
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution*, *17* (4), 540-552. doi:10.1093/oxfordjournals.molbev.a026334
- CBOL Plant Working Group, Hollingsworth, P. M., Forrest, L. L., Spouge, J. L., Hajibabaei, M., Ratnasingham, S., . . . Erickson, D. L. (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences*, *106* (31), 12794-12797. doi:10.1073/pnas.0905845106
- Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., & Xia, R. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant*, *13* (8), 1194-1202. doi:10.1016/j.molp.2020.06.009

- China Plant BOL Group, Li, D. Z., Gao, L. M., Li, H. T., Wang, H., Ge, X. J., . . . Chen, S. L. (2011). Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceedings of the National Academy of Sciences*, *108* (49), 19641-19646. doi:10.1073/pnas.1104551108
- Darling, A., Mau, B., & Perna, N. (2010). progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS One*, *5*, e11147. doi:10.1371/journal.pone.0011147
- dePamphilis, C., Young, N., & Wolfe, A. (1997). Evolution of plastid gene rps2 in a lineage of hemiparasitic and holoparasitic plants: Many losses of photosynthesis and complex patterns of rate variation. *Proceedings of the National Academy of Sciences of the United States of America*, *94* (14), 7367-7372. doi:10.1073/pnas.94.14.7367
- Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular biology and evolution*, *29* (8), 1969-1973. doi:10.1093/molbev/mss075
- Fang, H., Dai, G. N., Liao, B. B., Zhou, P., & Liu, Y. L. (2022). Application of chloroplast genome in the identification of *Phyllanthus urinaria* and its common adulterants. *Frontiers in Plant Science*, *13* (1099856). doi:10.3389/fpls.2022.1099856
- Fischer, E. (2004). Scrophulariaceae. In *Flowering Plants. Dicotyledons* (pp. 333-432): Springer.
- Frazer, K., Pachter, L., Poliakov, A., Rubin, E., & Dubchak, I. (2004). VISTA: computational tools for Comparative genomics. *Nucleic acids research*, *32*(suppl.2), W273-W279. doi:10.1093/nar/gkh458
- Gao, J., Jin, Y.-H., Li, Y.-Q., Qian, Y.-X., Zhang, C., Wang, R., & Qi, Z. (2019). The complete chloroplast genome sequence of *Siphonostegia chinensis* Benth. (Orobanchaceae). *Mitochondrial DNA Part B*, *4* (1), 732-733. doi:10.1080/23802359.2018.1564384
- Gong, X., Wang, J., Zhang, M., Wang, P., Wang, C., Shi, R., . . . Li, M. (2020). Bioactivity, compounds isolated, chemical qualitative, and quantitative analysis of *Cymbaria daurica* extracts. *Front Pharmacol*, *11*, 48. doi:10.3389/fphar.2020.00048
- Graham, S., Lam, V., & Merckx, V. (2017). Plastomes on the edge: the evolutionary breakdown of mycoheterotroph plastid genomes. *The New phytologist*, *214* (1), 48-55. doi:10.1111/nph.14398
- Greiner, S., Lehwark, P., & Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3. 1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic acids research*, *47* (W1), W59-W64. doi:10.1093/nar/gkz238
- Guo, J.-J., Liu, H., Zhu, Y.-H., Ren, J.-X., & Liang, Y.-H. (2017). Study on anti-inflammatory and analgesic effects of mongolian medicine *Cymbaria dahurica* extract. *China Pharmacy*, *28* (1), 64-67. doi:10.6039/j.issn.1001-0408.2017.01
- Guo, X., Zhang, G., Fan, L., Liu, C., & Ji, Y. (2021). Highly degenerate plastomes in two hemiparasitic dwarf mistletoes: *Arceuthobium chinense* and *A. pini* (Viscaceae). *Planta*, *253* (6), 1-16. doi:10.1007/s00425-021-03643-y
- Hu, S.-L. (2018). The modern research progress of mongolian medicine *cymbaria dahurica* L. *Chinese Journal of Ethnomedicine and Ethnopharmacy*, *27* (13), 31-32.
- Huang, C., Li, S., Guo, W., Zhang, Z., Meng, X., Li, X., . . . Zhang, C. (2023). *Cymbaria daurica* L.: A Mongolian herbal medicine for treating eczema via natural killer cell-mediated cytotoxicity pathway. *Journal of Ethnopharmacology*, *308*, 116246. doi:10.1016/j.jep.2023.116246
- Hurka, H., Friesen, N., Bernhardt, K.-G., Neuffer, B., Smirnov, S., Shmakov, A., & Blattner, F. (2019). The Eurasian steppe belt: Status quo, origin and evolutionary history *22* (3), 5-71. doi:10.14258/turczaninowia.22.3.1

- Jansen, R., Saski, C., Lee, S.-B., Hansen, A., & Daniell, H. (2010). Complete Plastid Genome Sequences of Three Rosids (Castanea, Prunus, Theobroma): Evidence for At Least Two Independent Transfers of rpl22 to the Nucleus. *Molecular biology and evolution*, *28* (1), 835-847. doi:10.1093/molbev/msq261
- Jiang, N., Dong, L.-N., Yang, J.-B., Tan, Y., Wang, H., Randle, C., . . . Yu, W.-B. (2022). Herbarium phylogenomics: resolving the generic status of the enigmatic Pseudobartsia (Orobanchaceae). *Journal of Systematics and Evolution*, *60* (5), 1218-1228. doi:10.1111/jse.12829
- Jin, J., Yu, W.-B., Yang, J.-B., Song, Y., dePamphilis, C., Yi, T., & Li, D.-Z. (2020). GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome biology*, *21* (1), 241. doi:10.1186/s13059-020-02154-5
- Katoh, K., & Standley, D. (2013). MAFFT Multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular biology and evolution*, *30* (4), 772-780. doi:10.1093/molbev/mst010
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., . . . Duran, C. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28* (12), 1647-1649. doi:10.1093/bioinformatics/bts199
- Kurtz, S., Choudhuri, J., Ohlebusch, E., Schleiermacher, C., Stoye, J., & Giegerich, R. (2001). REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic acids research*, *29* (22), 4633-4642. doi:10.1093/nar/29.22.4633
- Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic acids research*, *49* (W1), W293-W296. doi:10.1093/nar/gkab301
- Li, X., Feng, T., Randle, C., & Schneeweiss, G. M. (2019). Phylogenetic relationships in Orobanchaceae inferred from Low-copy nuclear genes: Consolidation of major clades and identification of a novel position of the non-photosynthetic Orobanche clade sister to all other parasitic Orobanchaceae. *Frontiers in Plant Science*, *10* , 902. doi:10.3389/fpls.2019.00902
- Li, X., Yang, J.-B., Wang, H., Song, Y., Corlett, R., Yao, X., . . . Yu, W.-B. (2021). Plastid ndh pseudogenization and gene loss in a recently derived lineage from the largest hemiparasitic plant genus Pedicularis (Orobanchaceae). *Plant and Cell Physiology*, *62* (6), 971-984. doi:10.1093/pcp/pcab074
- Li, Y., Zhou, J.-G., Chen, X.-L., Cui, Y.-X., Xu, Z.-c., Li, Y.-H., . . . Yao, H. (2017). Gene losses and partial deletion of small single-copy regions of the chloroplast genomes of two hemiparasitic Taxillus species. *Scientific Reports*, *7* (1), 1-12. doi:10.1038/s41598-017-13401-4
- Li, Z.-H., Long, P., Bai, S., Yang, D.-W., Zhu, H., Cui, Z.-H., . . . Li, M.-H. (2014). Chemical constituents from Cymbaria dahurica L. (Scrophulariaceae). *Biochemical Systematics and Ecology*, *57* , 11-14. doi:10.1016/j.bse.2014.07.012
- Liang, Y.-H., Liang, J., Yao, H.-Y., Liu, H., & Li, D.-W. (2016). Advances in studies on the chemical constituents in Xinba and their pharmacological activities. *Lishizhen Medicine And Materia Medica Research*, *27* (5), 1201-1204. doi:10.3969/j.issn.1008-0805.2016.05. 071
- Librado, P. J. R., & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics (Oxford, England)*, *25* (11), 1451-1452. doi:10.1093/bioinformatics/btp187
- Liu, Q., Li, X., Li, M., Xu, W., Schwarzacher, T., & Heslop-Harrison, J. S. (2020). Comparative chloroplast genome analyses of Avena: insights into evolutionary dynamics and phylogeny. *BMC Plant Biology*, *20* (1), 1-20. doi:10.1186/s12870-020-02621-y
- Liu, X., Fu, W., Tang, Y., Zhang, W., Song, Z., Li, L., . . . Wang, Y. (2019). Diverse trajectories of plastome degradation in holoparasitic Cistanche and the whereabouts of the lost plastid genes. *Journal of experimental botany*, *71* . doi:10.1093/jxb/erz456

- Liu, X., Li, Y., Yang, H., & Zhou, B. (2018). Chloroplast genome of the folk medicine and vegetable plant *Talinum paniculatum* (Jacq.) Gaertn.: Gene organization, comparative and phylogenetic analysis. *Molecules (Basel, Switzerland)*, *23* (4), 857. doi:10.3390/molecules23040857
- Liu, Z. F., Ma, H., Zhang, X. Y., Ci, X.-Q., Li, L., Hu, J. L., . . . Conran, J. G. (2022). Do taxon-specific DNA barcodes improve species discrimination relative to universal barcodes in Lauraceae? *Botanical Journal of the Linnean Society*, *199* (4), 741-753. doi:10.1093/botlinnean/boab089
- Lu, G. Y., Qiao, J. J., Wang, L., Liu, H., Wu, G., Zhu, Y., . . . Qin, M. J. (2022). An integrated study of *Viola* Herba (*Viola philippica*) and five adulterants by morphology, chemical compositions and chloroplast genomes: insights into its certified plant origin. *Chinese Medicine*, *17* (1), 1-18. doi:10.1186/s13020-022-00585-9
- Lu, Q.-F., Luo, W.-H., & Huang, Z.-H. (2020). Codon usage bias of chloroplast genome from two species of *Firmiana Marsili*. *Guihaia*, *40* (2), 173-183. doi:10.11931/guihaia.gxzw201811035
- Ma, M., Liu, Y., Bai, C., & Yong, J. W. H. (2021). The significance of chloroplast NAD (P) H dehydrogenase complex and its dependent cyclic electron transport in photosynthesis. *Frontiers in Plant Science*, *12*, 661863. doi:10.3389/fpls.2021
- McNeal, J. R., Bennett, J. R., Wolfe, A. D., & Mathews, S. (2013). Phylogeny and origins of holoparasitism in Orobanchaceae. *American Journal of Botany*, *100* (5), 971-983. doi:10.3732/ajb.1200448
- Naumann, J., Der, J., Wafula, E., Jones, S., Wagner, S., Honaas, L., . . . dePamphilis, C. (2016). Detecting and characterizing the highly divergent plastid genome of the nonphotosynthetic parasitic plant *Hydnora visseri* (Hydnoraceae). *Genome Biology and Evolution*, *8* (2), 345-363. doi:10.1093/gbe/evv256
- Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution*, *32* (1), 268-274. doi:10.1093/molbev/msu300
- Nickrent, D. (2020). Parasitic angiosperms: How often and how many? *Taxon*, *69* (1), 5-27. doi:10.1002/tax.12195
- Niu, Z., Xue, Q., Wang, H., Xie, X., Zhu, S., Liu, W., & Ding, X. (2017). Mutational biases and GC-biased gene conversion affect GC content in the plastomes of *Dendrobium* Genus. *International Journal of Molecular Sciences*, *18* (11), 2307. doi:10.3390/ijms18112307
- Rambaut, A. (2017). FigTree-version 1.4. 3, a graphical viewer of phylogenetic trees. *Computer program distributed by the author, website: <http://tree.bio.ed.ac.uk/software/figtree>*.
- Sabater, B. (2021). On the edge of dispensability, the chloroplast *ndh* genes. *International Journal of Molecular Sciences*, *22* (22), 12505. doi:10.3390/ijms222212505
- Schneeweiss, G. M. (2013). Phylogenetic relationships and evolutionary trends in Orobanchaceae. In *Parasitic Orobanchaceae* (pp. 243-265). Berlin, Heidelberg: Springer.
- Shi, R.-Y., Zhang, M.-Y., Agula, Gong, X., Zhang, C.-H., & Li, M.-H. (2020). Antimicrobial activities of the extracts from *Cymbaria daurica* L. *Journal of Medicine and Pharmacology of Chinese Minorities*, *26* (03), 22-24. doi:DOI:10.16041/j.cnki.cn15-1175.2020.03.014
- Shin, H. W., & Lee, N. (2018). Understanding plastome evolution in hemiparasitic Santalales: Complete chloroplast genomes of three species, *Dendrotrophe varians*, *Helixanthera parasitica*, and *Macrosolen cochinchinensis*. *PLoS One*, *13* (7), e0200293. doi:10.1371/journal.pone.0200293
- Team, R. (2015). RStudio: Integrated development for R. RStudio. *RStudio, Inc., Boston, MA URL <http://www.rstudio.com>*, *42* (14), 84.

- Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., & Greiner, S. (2017). GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic acids research*, *45* (W1), W6-W11. doi:10.1093/nar/gkx391
- Torsvik, T. H., & Cocks, L. R. M. (2016). *Earth History and Palaeogeography*. Cambridge: Cambridge University Press.
- Wang, Q., Bao, Y., Hao, J., & Han, J. (2018). HPLC analysis of six iridoid glycosides from *Cymbaria dahurica* L. and their structural elucidation. *Journal of Food Biochemistry*, *42* (2). doi:10.1111/jfbc.12470
- Wang, Z.-W., Domg, Y.-H., Wu, G.-D., & Li, M.-H. (2012). Microscopic identification of *Cymbaria* powder. *Journal Of Baotou Medical College*, *28* (5), 7-8. doi:10.16833/j.cnki.jbmc.2012.05.005
- Westwood, J., Cw, d., Das, M., Fernández-Aparicio, M., Honaas, L., Timko, M., . . . Yoder, J. (2012). The parasitic plant genome project: New tools for understanding the biology of *Orobanche* and *Striga*. *Weed Science*, *60* (2), 295-306. doi:10.2307/41497638
- Wick, R., Schultz, M., Zobel, J., & Holt, K. (2015). Bandage: Interactive visualization of de novo genome assemblies. *Bioinformatics (Oxford, England)*, *31* (20), 3350-3352. doi:10.1093/bioinformatics/btv383
- Wicke, S., Müller, K., Pamphilis, C., Quandt, D., Wickett, N., Zhang, Y., & Schneeweiss, G. (2013). Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. *The Plant cell*, *25* (10), 3711-3725. doi:10.1105/tpc.113.113373
- Wicke, S., & Naumann, J. (2018). Molecular evolution of plastid genomes in parasitic flowering plants. In *Advances in Botanical Research* (Vol. 85, pp. 315-347): Academic Press.
- Wicke, S., Schneeweiss, G., dePamphilis, C., Müller, K., & Quandt, D. (2011). The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Molecular Biology*, *76* (3), 273-297. doi:10.1007/s11103-011-9762-4
- Wolfe, A., Randle, C., Liu, L., & Steiner, K. (2005). Phylogeny and biogeography of Orobanchaceae. *Folia Geobotanica*, *40* (2), 115-134. doi:10.1007/BF02803229
- Wu, Q.-H., Li, B.-T., Zhu, S.-L., Xiao, X., Zhang, X.-Q., & Tu, J. (2020). Research on network pharmacology of Mongolian medicine *Cymbaria* in treatment of type 2 diabetes. *China Journal of Chinese Materia Medica*, *45* (8), 1764-1771. doi:10.19540/j.cnki.cjcm.20191213.401
- Xia, Z., Li, C.-C., Wen, J., & Wang, Y.-Z. (2021). Rehmannieae or Rehmanniaceae? Evidence from plastome sequences and floral morphology. *Botanical Journal of the Linnean Society*, *196* (2), 145-162. doi:10.1093/botlinnean/boaa105
- Xu, Y., Zhang, J., Ma, C., Lei, Y., Shen, G., Jin, J., . . . Wu, J. (2022). Comparative genomics of orobanchaceous species with different parasitic lifestyles reveals the origin and stepwise evolution of plant parasitism. *Molecular Plant*, *15* (8), 1384-1399. doi:10.1016/j.molp.2022.07.007
- Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular biology and evolution*, *24* (8), 1586-1591. doi:10.1093/molbev/msm088
- Young, N., Steiner, K., & dePamphilis, C. (1999). The evolution of parasitism in Scrophulariaceae/Orobanchaceae: plastid gene sequences refute an evolutionary transition series. *Annals of the Missouri Botanical Garden*, *86* (4), 876-893. doi:10.2307/2666173
- Yu, W.-B., Randle, C., Lu, L., Wang, H., Yang, J.-B., dePamphilis, C., . . . Li, D.-Z. (2018). The hemiparasitic plant *Phtheirospermum* (Orobanchaceae) is polyphyletic and contains cryptic species in the Hengduan Mountains of Southwest China. *Frontiers in Plant Science*, *9*, 142. doi:10.3389/fpls.2018.00142
- Yue, J., Ni, Y., Jiang, M., Chen, H., Chen, P., & Liu, C. (2022). Characterization of *Codonopsis pilosula* subsp. tangshen plastome and comparative analysis of codonopsis species. *PLoS One*, *17* (8), e0271813.

doi:10.1371/journal.pone.0271813

Zeng, S., Zhou, T., Han, K., Yang, Y., Zhao, J., & Liu, Z.-L. (2017). The Complete Chloroplast Genome Sequences of Six *Rehmannia* Species. *Genes*, 8, 103. doi:10.3390/genes8030103

Zhang, C.-H., Yao, X., Hasi, B.-T., Hu, Y.-L., Li, Z.-H., Wu, G.-D., & Li, M.-H. (2013). Research progress of Mongolian Xinba. *Modern Chinese Medicine*, 15 (12), 1068-1072.

Zhang, D., Gao, F., Jakovlić, I., Zou, H., Zhang, J., Li, W., & Wang, G. (2020). PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Molecular Ecology Resources*, 20 (1), 348-355. doi:10.1111/1755-0998.13096

Zhang, R., Xu, B., Li, J., Zhao, Z., Han, J., Lei, Y., . . . Liu, Z.-L. (2020). Transit from autotrophism to heterotrophism: Sequence variation and evolution of chloroplast genomes in Orobanchaceae species. *Frontiers in Genetics*, 11, 542017. doi:10.3389/fgene.2020.542017

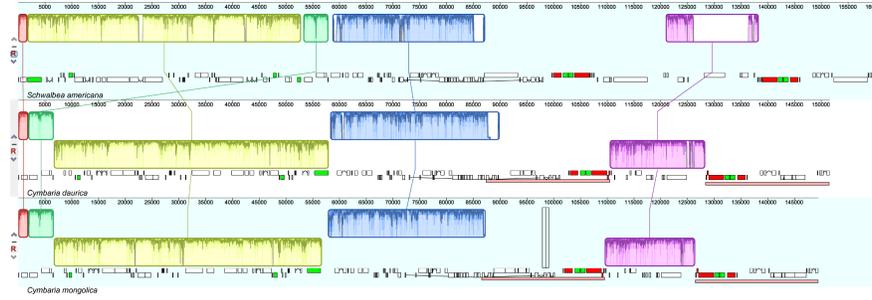
Zhao, Y.-Z. (1999). The species research and its floristic analysis of plants of Cymbaria. *Acta Scientiarum Naturalium Universitatis NeiMongol*, 30 (3), 351-353.

Zhu, S., Liu, Q. Z., Qiu, S. M., Dai, J. P., & Gao, X. X. (2022). DNA barcoding: an efficient technology to authenticate plant species of traditional Chinese medicine and recent advances. *Chinese Medicine*, 17 (1), 1-17. doi:10.1186/s13020-022-00655-y

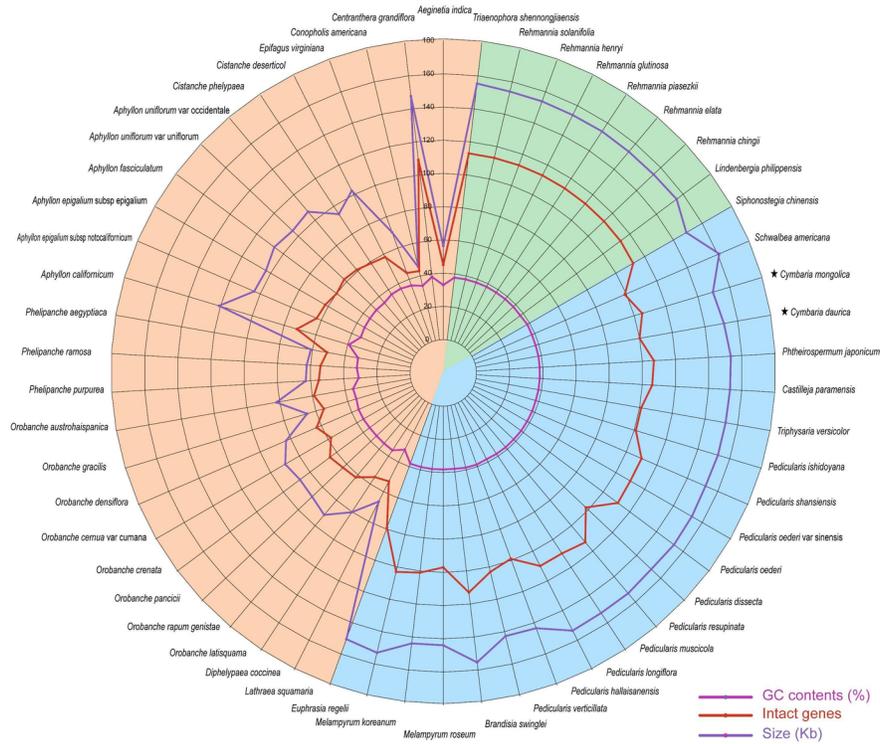
APPENDIX



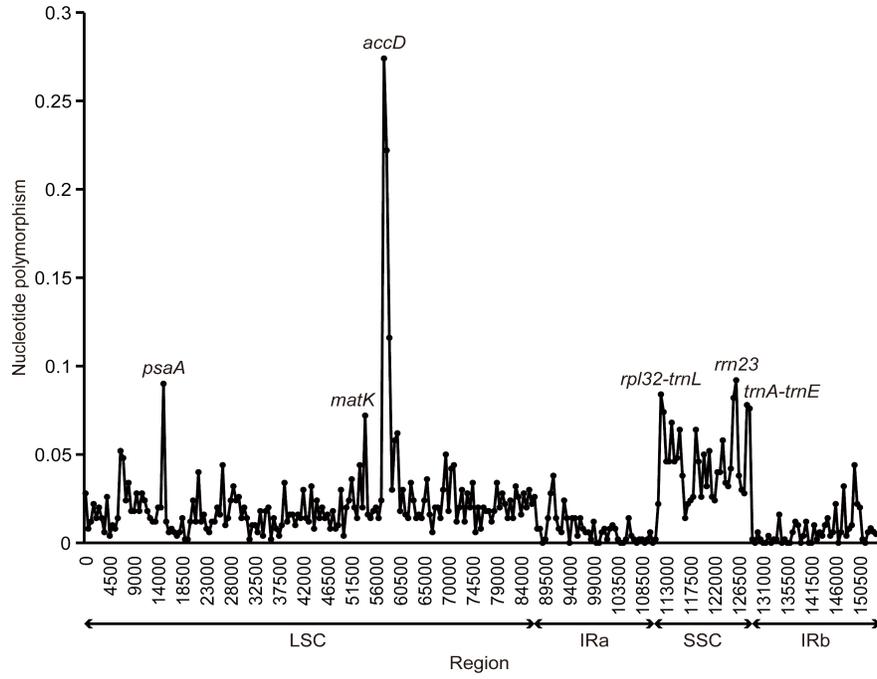
Figure S1. Mongolian pharmaceutical classic literature describing the herb *C. daurica* (Huang et al., 2023).



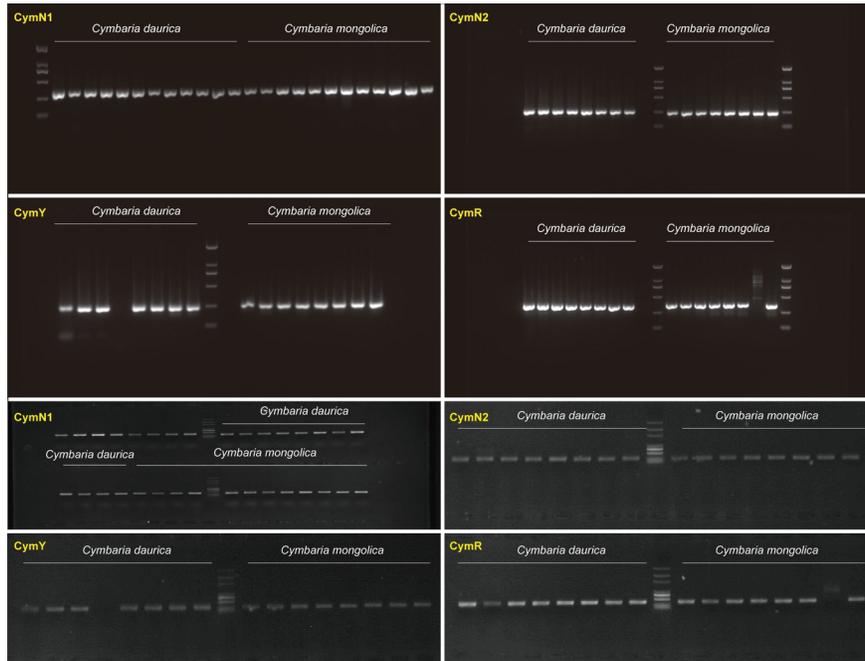
**Figure S2.** Collinearity of two *Cymbaria* chloroplast genomes using Mauve program with *Schwalbea americana* as the reference.



**Figure S3.** Radar chart of complete chloroplast genome comparison of 54 Orobanchaceae species. From outside to inside: genome size, intact genes, and GC content. Background colors of green, blue, and orange represent autotroph, hemiparasite, and holoparasite species, respectively.



**Figure S4.** Sliding window analysis of two *Cymbaria* chloroplast genomes.



**Figure S5.** The original and uncropped agarose gel electrophoresis of the amplification of DNA barcodes

**Table S1.** Gene contents in two *Cymbaria* chloroplast genomes.

Category	Gene group	Gene name
----------	------------	-----------



NC_027838	Holoparasites	<i>Lathraea_squamaria</i>	150.504	38.1	79	4
NC_045041	Hemiparasites	<i>Euphrasia_regelii</i>	153.026	38.2	103	6
NC_057523	Hemiparasites	<i>Melampyrum_koreanum</i>	143.865	38.2	101	6
NC_053791	Hemiparasites	<i>Melampyrum_roseum</i>	143.896	38.1	97	6
NC_042954	Hemiparasites	<i>Brandisia_swinglei</i>	155.344	38.1	113	7
MT040753	Hemiparasites	<i>Pedicularis_verticillata</i>	142.733	38.6	103	6
NC_037433	Hemiparasites	<i>Pedicularis_hallaisanensis</i>	143.469	38.7	99	6
NC_046852	Hemiparasites	<i>Pedicularis_longiflora</i>	153.547	38.1	110	7
NC_046853	Hemiparasites	<i>Pedicularis_muscicola</i>	152.907	38.3	110	7
NC_046397	Hemiparasites	<i>Pedicularis_resupinata</i>	153.145	38.4	113	7
NC_056312	Hemiparasites	<i>Pedicularis_dissecta</i>	152.120	38.3	98	6
NC_046854	Hemiparasites	<i>Pedicularis_oederi</i>	153.139	38.4	111	7
MW770457	Hemiparasites	<i>Pedicularis_oederi_var_sinensis</i>	152.770	38.3	110	7
NC_058762	Hemiparasites	<i>Pedicularis_shansiensis</i>	151.902	38.3	110	7
NC_029700	Hemiparasites	<i>Pedicularis_ishidoyana</i>	152.571	38.1	101	6
NC_053793	Hemiparasites	<i>Triphysaria_versicolor</i>	152.583	38.2	101	7
NC_031805	Hemiparasites	<i>Castilleja_paramensis</i>	152.926	38.2	106	7
NC_053792	Hemiparasites	<i>Phtheirospermum_japonicum</i>	153.397	38.1	107	7
NC_064388	Hemiparasites	<i>Cymbaria_daurica</i>	151.545	38.2	100	7
NC_064104	Hemiparasites	<i>Cymbaria_mongolica</i>	149.431	38	105	7
NC_023115	Hemiparasites	<i>Schwalbea_americana</i>	160.910	38.1	99	7
NC_046038	Hemiparasites	<i>Siphonostegia_chinensis</i>	148.961	38.4	112	7
NC_022859	Autotrophs	<i>Lindenbergia_philippensis</i>	155.103	37.8	113	7
NC_033534	Autotrophs	<i>Rehmannia_chingii</i>	154.055	38	113	7
NC_034312	Autotrophs	<i>Rehmannia_elata</i>	153.772	38	113	7
NC_034311	Autotrophs	<i>Rehmannia_piasezkii</i>	153.952	37.9	113	7
NC_034308	Autotrophs	<i>Rehmannia_glutinosa</i>	153.622	38	113	7
NC_034309	Autotrophs	<i>Rehmannia_henryi</i>	153.890	37.9	113	7
NC_034310	Autotrophs	<i>Rehmannia_solanifolia</i>	153.989	37.9	113	7
NC_039781	Autotrophs	<i>Triaenophora_shennongjiaensis</i>	155.319	37.7	113	7

Note: The statistics of gene numbers here refer to 113 unique plastid genes in the typical angiosperms. The duplicated genes are not included.

**Table S3.** Four pairs of primers for amplifying DNA barcodes.

Name	Sequence (5' to 3')	Position
CymN1_F	agacccattctagaactaaca	<i>ndhD-ndhG</i>
CymN1_R	agtcgctaaatctgcacaagt	
CymN2_F	gaccgacccgaaaattgat	<i>ndhD-ndhG</i>
CymN2_R	tcccaccttgtaaattaggaaa	
CymY_F	atcatgcgcataaccagtagc	<i>rps15-ycf1</i>
CymY_R	agctactatgttcacagaagca	
CymR_F	ggcaattgtgaactctcggg	<i>rpl32-trnL-UAG</i>
CymR_R	tcttctcagtagctcagcg	

**Table S4.** The list of sample numbers of the samples used in the validation of DNA barcodes.

Species	Sample No.	Location collected	Longitude	Latitude
---------	------------	--------------------	-----------	----------

<i>Cymbaria mongolica</i>	Cm1	Binzhou, Shaanxi Province, China	108°13'06.11"E	35°1
	Cm2	Helan, Ningxia Hui Autonomous Region, China	105°59'35.67"E	38°4
	Cm3	Helan, Ningxia Hui Autonomous Region, China	105°59'35.67"E	38°4
	Cm4	Tianzhu, Gansu Province, China	103°04'44.68"E	37°0
	Cm5	Tianzhu, Gansu Province, China	103°04'44.68"E	37°0
	Cm6	Xining, Qinghai Province, China	101°49'33.59"E	36°3
	Cm7	Xining, Qinghai Province, China	101°49'33.59"E	36°3
<i>Cymbaria_aurica</i>	Cd1	Xilingol, Inner Mongolia Autonomous Region, China	116°47'07.76"E	43°2
	Cd2	Yanqing, Hebei province, China	115°57'34.71"E	40°2
	Cd3	Yanqing, Hebei province, China	115°57'34.71"E	40°2
	Cd4	Yijinhuol Banner, Inner Mongolia Autonomous Region, China	110°11'42.41"E	39°3
	Cd5	Yijinhuol Banner, Inner Mongolia Autonomous Region, China	110°11'42.41"E	39°3
	Cd6	Daqing, Heilongjiang province, China	124°32'09.86"E	46°2
	Cd7	Daqing, Heilongjiang province, China	124°32'09.86"E	46°2

---