Postprint: Phylogeny of Scrophulariaceae sensu lato Based on Plastid Genomics

Authors: Wang Xueqin, Song Weiwu, Ma Feilong, Gao Wenjing, Zhao Yan

Date: 2022-08-05T00:00:00+00:00

Abstract

This study analyzed the phylogenetic relationships of Scrophulariaceae in the broad sense using plastid genome data from 129 species representing 107 genera of Scrophulariaceae and related taxa publicly available in the GenBank database. Using protein-coding genes, we constructed a matrix and reconstructed phylogenetic trees using maximum likelihood (ML) and Bayesian inference. Based on both analytical methods, we obtained phylogenetic trees with completely consistent topology, showing high resolution and support. In the ML tree, the total number of branches was 129, of which 123 branches had support values $ $70%. The phylogenetic tree results indicated that (1) Scrophulariaceae in the broad sense is not a monophyletic group; the 51 species (37 genera) belonging to Scrophulariaceae s.l. are scattered across Orobanchaceae, Paulowniaceae, Wightiaceae, Mazaceae, Linderniaceae, Scrophulariaceae s.s., and Plantaginaceae. (2) Scrophulariaceae s.s. is a monophyletic group, which, in addition to Bontia, Calamphoreus, Diocirea, Eremophila, Glycocystis, Leucophyllum, Scrophularia, and Verbascum formerly belonging to Scrophulariaceae s.l., also includes Buddleja formerly belonging to Loganiaceae and Myoporum formerly belonging to Myoporaceae. (3) Lamiales is monophyletic, forming 14 well-supported monophyletic branches corresponding to 14 families (excluding Wightiaceae and Pedaliaceae, which each include only one species). Interfamilial relationships were well resolved, with Oleaceae being the earliest diverging lineage, and the remaining taxa together constituting the core Lamiales. Within the core Lamiales, the order of taxon divergence was Gesneriaceae, Plantaginaceae, Scrophulariaceae, and Linderniaceae. The remaining families clustered into two major clades: one major clade included five families, with Bignoniaceae, Verbenaceae, and Pedaliaceae forming a moderately supported subclade, in which Bignoniaceae and Verbenaceae are sister groups; Acanthaceae and Lentibulariaceae also formed a moderately supported subclade; the other major clade included six families, with the order of taxon divergence being Lamiaceae, Mazaceae, Wightiaceae + Phrymaceae, and Paulowniaceae and Orobanchaceae being sister groups and representing relatively derived taxa. Our results demonstrate that Scrophulariaceae in the traditional sense is not a natural group, and its included species and genera are scattered throughout Lamiales; plastid genomes are suitable for resolving the phylogenetic relationships of Scrophulariaceae s.l., and further studies with expanded sampling are warranted.

Full Text

Phylogeny of Scrophulariaceae sensu lato Based on Plastid Genomics

Wang Xueqin[1], Song Weiwu[2]*, Ma Feilong[3], Gao Wenjing[1], Zhao Yan[1]

[1]College of Life Science and Agronomy, Zhoukou Normal University, Zhoukou 466001, Henan, China
[2]School of Chemistry & Chemical Engineering, Zhoukou Normal University, Zhoukou 466001, Henan, China
[3]Key Laboratory of Plant Genetics and Molecular Breeding, Zhoukou Normal University, Zhoukou 466001, Henan, China

Funding: National Natural Science Foundation of China (81803388); Key Research Program of Henan Province (182102310617); Department of Science and Technology Planning Project of Henan Province (212102110142)

First Author: Wang Xueqin (1986-), Ph.D., Lecturer, Main research focus: plant systematics and evolution, (E-mail) wangxueqin0218@163.com
Corresponding Author: *Song Weiwu, Ph.D., Lecturer, Research direction: plant chemistry, (E-mail) wwsong@zknu.edu.cn

Abstract

Phylogenetic relationships of Scrophulariaceae s.l. (sensu lato) and closely related groups were explored using 129 plastid genomes representing 107 genera downloaded from GenBank. Based on coding sequences (CDS), phylogenetic trees were reconstructed using maximum likelihood (ML) and Bayesian inference (BI) approaches. The ML and BI analyses shared identical topologies, with main clades consistently recovered with high support. Of 129 total nodes, 123 were supported by ML bootstrap values $ $70%. The resulting phylogenies revealed: (1) Scrophulariaceae s.l. is polyphyletic, with 51 species from 37 genera distributed across seven families: Orobanchaceae, Paulowniaceae, Wightiaceae, Mazaceae, Linderniaceae, Scrophulariaceae s.s. (sensu stricto), and Plantaginaceae. (2) Scrophulariaceae s.s. is monophyletic, comprising eight genera primarily from former Scrophulariaceae s.l. (Bontia, Calamphoreus, Diocirea, Eremophila, Glycocystis, Leucophyllum, Scrophularia, and Verbascum), plus Buddleja (originally Loganiaceae) and Myoporum (traditionally Myoporaceae). (3) Lamiales is monophyletic, with 14 well-supported monophyletic clades corresponding to families (Wightiaceae and Pedaliaceae, each represented by a single species, were excluded). Interfamilial relationships were well resolved: Oleaceae diverged first, with remaining families forming core Lamiales. Within core Lamiales, lineages diverged sequentially as Gesneriaceae, Plantaginaceae, Scrophulariaceae, and Linderniaceae. The remaining families formed two major clades: one containing Bignoniaceae, Verbenaceae, and Pedaliaceae (with Bignoniaceae and Verbenaceae as sister groups) plus Acanthaceae + Lentibulariaceae; the other containing six families with Lamiaceae, Mazaceae, Wightiaceae + Phrymaceae, and a sister pair of Paulowniaceae + Orobanchaceae as the most derived lineage. These results indicate that traditional Scrophulariaceae represents an unnatural assemblage distributed throughout Lamiales. Plastid phylogenomics shows strong potential for resolving relationships within Scrophulariaceae s.l., though expanded taxon sampling is needed.

Keywords: Scrophulariaceae sensu lato, phylogenomics, plastomes, Lamiales

Introduction

Scrophulariaceae (figwort family) belongs to Lamiales and has a global distribution. The family represents a large and heterogeneous assemblage lacking diagnostic characters to distinguish it from related groups. During the era of morphology-based classification, taxonomists assigned numerous lineages to Scrophulariaceae, making it once the most species-rich family in Lamiales. This absence of definitive synapomorphies suggested that Scrophulariaceae might not be monophyletic. The earliest molecular phylogenetic studies of Scrophulariaceae date to the late 1990s, when Olmstead & Reeves (1995) used two chloroplast fragments to test its monophyly and found it to be polyphyletic. Olmstead et al. (2001) subsequently analyzed three chloroplast genes and demonstrated that Scrophulariaceae s.l. comprises at least five monophyletic groups: Scrophulariaceae s.s. (sensu stricto), Veronicaceae, Orobanchaceae, Calceolariaceae, and Stilbaceae. Further research led to the disintegration of Scrophulariaceae s.l., with former members transferred to other families or recognized as new, smaller families such as Calceolariaceae, Linderniaceae, Mazaceae, Orobanchaceae, Paulowniaceae, Schlegeliaceae, Scrophulariaceae s.s., and Wightiaceae (Oxelman et al., 2005; Rahmanzadeh et al., 2005; Tank et al., 2006; Liu et al., 2020).

Scrophulariaceae s.s. can be divided into eight tribes: Aptosimeae, Buddlejeae, Hemimerideae, Leucophylleae, Limoselleae, Myoporeae, Scrophularieae, and Teedieae. Except for Scrophularieae, which is primarily distributed in the north temperate zone, most tribes are predominantly southern hemisphere lineages (Oxelman et al., 2005; Tank et al., 2006). According to the *Flora of China*, Scrophulariaceae includes approximately 220 genera and 4,500 species worldwide, with 61 genera and 681 species in China. The circumscription adopted in *Flora of China* corresponds to Scrophulariaceae s.l. as described above. Previous studies on Scrophulariaceae s.l. typically used few chloroplast fragments, yielding phylogenetic relationships with insufficient support for some clades. For example, the monophyly of Plantaginaceae, Scrophulariaceae, and Bignoniaceae received only moderate support, while relationships among Bignoniaceae, Verbenaceae, Acanthaceae, and Lentibulariaceae remained poorly resolved (Liu et al., 2020). Therefore, further investigation is warranted. This study uses plastid genome data from protein-coding genes to reconstruct phylogenetic trees for Scrophulariaceae s.l. and related groups, addressing the taxonomic placement of former Scrophulariaceae s.l. members and exploring interfamilial relationships within Lamiales.

Methods

1.1 Data Collection

We downloaded plastid genome sequences (including chloroplast genomes) and protein-coding genes (coding sequences, CDS) for 129 species representing 107 genera of Scrophulariaceae and related groups from GenBank (Table 1). Species selection prioritized inclusion of all Scrophulariaceae s.l. members. Additional species from other Lamiales families were selected based on published literature, including Oleaceae, Gesneriaceae, Lentibulariaceae, Acanthaceae, Pedaliaceae, Verbenaceae, Bignoniaceae, Lamiaceae, Phrymaceae, and Orobanchaceae. Oleaceae does not belong to core Lamiales, and Gesneriaceae is phylogenetically distant from Scrophulariaceae s.l.; therefore, despite numerous published sequences for these families, we included relatively few representatives. For remaining families, we sampled extensively. Linderniaceae and Lentibulariaceae are represented by fewer species due to limited publicly available sequences. Three outgroup species were selected: *Catharanthus roseus* (Apocynaceae, KC561139.1), *Gentiana veitchiorum* (Gentianaceae, MG192310.1), and *Neolamarckia cadamba* (Rubiaceae, MG572117.1). Genome information for all ingroup species, including genome size and number of protein-coding genes, was compiled using Excel.

1.2 Matrix Construction

We constructed a phylogenetic matrix based on protein-coding genes for subsequent analyses. CDS from each species were sorted alphabetically by gene name, concatenated, and aligned using the online software MAFFT version 7 (https://mafft.cbrc.jp/alignment/server/) (Katoh et al., 2019) with default parameters. Alignments were manually adjusted using MEGA 7.0 software (Kumar et al., 2016), and matrix characteristics including length and number of variable sites were recorded.

1.3 Phylogenetic Analysis

We performed phylogenetic analyses using both maximum likelihood (ML) and Bayesian inference (BI) methods. ML analysis was conducted using RAxML ver. 8.2.12 (Stamatakis, 2014) on the CIPRES Science Gateway (Miller et al., 2010) with the GTRGAMMA model and rapid bootstrap analysis of 1,000 replicates (BS). BI analysis was performed using MrBayes v3.2.7a (Ronquist & Huelsenbeck, 2003) on CIPRES Science Gateway with the GTR+GAMMA model. BI parameters included: MCMC algorithm, 1,000,000 generations, sampling every 1,000 generations, with the first 25% of samples discarded as burn-in. The remaining samples were used to construct a majority-rule consensus tree and calculate posterior probabilities (PP) for each branch.

Results

2.1 Genome Size

We downloaded genome sequences for 132 species (including outgroups), comprising 127 complete plastid genomes and five partial genomes: *Neobartsia inaequalis*, *Callicarpa nudiflora*, *Calamphoreus inflatus*, *Glycocystis beckeri*, and *Linaria vulgaris* (Table 1). Among the partial genomes, *L. vulgaris* had the shortest sequence with only 13 CDS, while the other four species had relatively complete genomes with 82–87 CDS. Among complete plastid genomes, *Conopholis americana* (Orobanchaceae) had the fewest CDS (21), while *Anemopaegma acutifolium* (Bignoniaceae) had the most. With the exception of Orobanchaceae and Lentibulariaceae, all species had genome sizes ranging from 144–168 kb. Orobanchaceae exhibited substantial variation in genome size (45–160 kb), correlating with lifestyle: autotrophic lineages had larger genomes, while parasitic lineages lost photosynthesis-related genes (Cusimano & Wicke, 2016), resulting in smaller genomes and fewer CDS. Lentibulariaceae had smaller genomes (140–150 kb) and fewer CDS (72–88) due to loss of *ndh* genes associated with carnivory (Wicke et al., 2013).

2.2 Matrix Characteristics

Among the 132 species included, 14 had $ $50 CDS (13 of which were Orobanchaceae), while 118 species (89.39%) had $ $70 CDS. Matrix construction requires balancing the number of informative sites against the proportion of missing data. Using only shared CDS reduces missing data but may yield insufficient data and fewer informative sites. Previous studies demonstrate that larger matrices, despite containing high proportions of missing data (80–90%), improve phylogenetic resolution and support by increasing total informative sites (Roure et al., 2013; Tripp et al., 2017). Therefore, we used all available CDS for each species. This approach results in missing data for some species post-alignment but substantially increases sequence length. The final CDS matrix comprised 183,488 bp, with 55,152 variable sites (30.06%) and 32,932 informative sites (17.95%), and a missing data proportion of 60.83%.

2.3 Phylogenetic Analysis of Scrophulariaceae s.l.

ML and BI analyses robustly resolved phylogenetic relationships within Scrophulariaceae and related groups, yielding stable, well-supported topologies. Support values were high: in the ML tree, 112 of 129 nodes (86.82%) had BS $ $95, 123 nodes (95.35%) had BS $ $70, and only three nodes (2.33%) had BS $ $50, all within Orobanchaceae. Topologically, Scrophulariaceae s.l. was not monophyletic [Figure 1: see original paper]. Fifty-one species from 37 genera were distributed across seven families: Orobanchaceae, Paulowniaceae, Wightiaceae, Mazaceae, Linderniaceae, Scrophulariaceae s.s., and Plantaginaceae. Eight genera remained in Scrophulariaceae s.s., while 29 were transferred out. Eleven genera moved from Scrophulariaceae s.l. to Orobanchaceae: *Brandisia*, *Castilleja*, *Centranthera*, *Euphrasia*, *Lindenbergia*, *Melampyrum*, *Pedicularis*, *Phtheirospermum*, *Rehmannia*, *Siphonostegia*, and *Triaenophora*. *Paulownia* was removed to form Paulowniaceae. *Wightia* was removed to establish Wightiaceae. *Dodartia*, *Lancea*, and *Mazus* were removed to form Mazaceae. *Torenia* was transferred to Linderniaceae. Twelve genera moved to Plantaginaceae: *Antirrhinum*, *Aragoa*, *Bacopa*, *Digitalis*, *Hemiphragma*, *Lagotis*, *Linaria*, *Littorella*, *Neopicrorhiza*, *Plantago*, *Veronica*, and *Veronicastrum*.

Scrophulariaceae s.s. was monophyletic (BS=100, PP=1.00) [Figure 1: see original paper], comprising 10 genera and 15 species. In addition to former Scrophulariaceae s.l. members (*Bontia*, *Calamphoreus*, *Diocirea*, *Eremophila*, *Glycocystis*, *Leucophyllum*, *Scrophularia*, and *Verbascum*), it included *Buddleja* (originally Loganiaceae) and *Myoporum* (traditionally Myoporaceae). Our sampling covered four tribes of Scrophulariaceae s.s.: Buddlejeae (two *Buddleja* species), Scrophularieae (two *Scrophularia* and two *Verbascum* species), Leucophylleae (one *Leucophyllum* species), and Myoporeae (six remaining genera). Except for Leucophylleae (single species), all tribes were monophyletic (BS=100, PP=1.00). Buddlejeae and Scrophularieae were sister groups, while Leucophylleae and Myoporeae formed another sister pair.

2.4 Interfamilial Relationships Within Lamiales

Lamiales was monophyletic (BS=100, PP=1.00), with 14 well-supported monophyletic clades corresponding to families (Wightiaceae and Pedaliaceae, each represented by one species, were excluded) [Figure 1: see original paper]. Among the 16 families, Oleaceae diverged first; the remaining families formed core Lamiales. Core Lamiales was monophyletic (BS=100, PP=1.00), with sequential divergence of Gesneriaceae, Plantaginaceae, Scrophulariaceae, and Linderniaceae. The remaining families formed two well-supported major clades (BS=100, PP=1.00). One clade contained five families: Bignoniaceae, Verbenaceae, and Pedaliaceae formed a moderately supported subclade (BS=88, PP=1.00) where Bignoniaceae and Verbenaceae were sisters; Acanthaceae + Lentibulariaceae formed another moderately supported subclade (BS=74, PP=0.99). The other clade comprised six families (Orobanchaceae, Paulowniaceae, Phrymaceae, Wightiaceae, Mazaceae, and Lamiaceae), with Lamiaceae diverging first, followed by Mazaceae, then Phrymaceae + Wightiaceae, and finally the sister pair Orobanchaceae + Paulowniaceae as the most derived lineage.

Discussion

3.1 Polyphyly and Disintegration of Scrophulariaceae s.l.

Our plastid phylogenomic analyses reconstructed relationships within Scrophulariaceae and related groups with high support, confirming that Scrophulariaceae s.l. is polyphyletic. Its species are distributed across Orobanchaceae, Paulowniaceae, Wightiaceae, Mazaceae, Linderniaceae, Scrophulariaceae s.s., and Plantaginaceae, consistent with previous chloroplast and nuclear gene studies (Olmstead et al., 2001; Oxelman et al., 2005; Rahmanzadeh et al., 2005; Tank et al., 2006; Liu et al., 2020).

Traditionally, Orobanchaceae comprised parasitic herbs. Our study transferred 11 former Scrophulariaceae s.l. genera to Orobanchaceae. *Rehmannia* and *Triaenophora* were sister taxa forming a basal Orobanchaceae lineage, consistent with Albach et al. (2009) based on chloroplast fragments and ITS. Olmstead et al. (2001) using three chloroplast fragments suggested transferring *Lindenbergia*, *Castilleja*, *Melampyrum*, and *Pedicularis* to Orobanchaceae. Bennett & Mathews (2006) using nuclear gene *PHYA* supported placement of *Phtheirospermum*, *Siphonostegia*, and *Euphrasia* in Orobanchaceae. McNeal et al. (2013) using two chloroplast and two nuclear genes resolved the position of *Centranthera* within Orobanchaceae.

For *Paulownia*, Olmstead et al. (2001) demonstrated it was neither Scrophulariaceae nor related to Bignoniaceae, supporting its independent lineage status. For *Wightia*, Liu et al. (2020) using six chloroplast genes established its position in Lamiales and recommended recognition of Wightiaceae.

Lancea and *Mazus*, primarily distributed in eastern and southeastern Asia, were suggested by Beardsley & Olmstead (2002) to be transferred from Scrophulariaceae to Phrymaceae as tribe Mazoideae. Subsequently, Albach et al. (2009) using four chloroplast fragments and ITS recommended removing them from Phrymaceae. Deng et al. (2019) using chloroplast and nuclear data supported Mazaceae as a monophyletic group comprising *Dodartia*, *Lancea*, and *Mazus*.

Olmstead et al. (2001) proposed transferring *Antirrhinum* and *Digitalis* from Scrophulariaceae s.l. to form Veronicaceae. Based on this, Albach et al. (2005) further recommended reassigning Veronicaceae as Plantaginaceae, with *Antirrhinum*, *Aragoa*, *Bacopa*, *Digitalis*, *Hemiphragma*, *Lagotis*, *Linaria*, *Littorella*, *Plantago*, *Veronica*, and *Veronicastrum* placed in Plantaginaceae. *Lagotis* and *Neopicrorhiza* have long been considered Plantaginaceae members but were not sampled in these studies; Zhang et al. (2019) and Cheng et al. (2020) using chloroplast genomes supported their placement in Plantaginaceae.

Phylogenetic relationships within Scrophulariaceae s.s. are largely consistent with previous studies (Oxelman et al., 2005; Tank et al., 2006): Buddlejeae and Scrophularieae are closely related, as are Leucophylleae and Myoporeae, with improved support (BS=100, PP=1.00). Given our limited sampling of Scrophulariaceae s.s. (only four of eight tribes and few species), detailed discussion is not warranted.

3.2 Interfamilial Relationships Within Lamiales

Species from Scrophulariaceae s.l. in our study are distributed across seven families: Orobanchaceae, Paulowniaceae, Wightiaceae, Mazaceae, Linderniaceae, Scrophulariaceae s.s., and Plantaginaceae. Plantaginaceae occupies a relatively basal position in Lamiales, while Orobanchaceae is derived, making Scrophulariaceae s.l. phylogeny intrinsically linked to broader Lamiales relationships. Our study includes 16 Lamiales families, with relationships largely congruent with previous research. Liu et al. (2020) used six chloroplast fragments to investigate Wightiaceae placement and reconstruct Lamiales phylogeny, recovering Scrophulariaceae s.s. as monophyletic with moderate support (BS=78, PP=0.94) within core Lamiales, closely related to Plantaginaceae, Stilbaceae, and Byblidaceae + Linderniaceae. Our plastid phylogenomic topology is largely consistent with Liu et al. (2020) but shows substantially improved support, with mean support of 96.06%, 86.82% of nodes having BS $ $95, and 95.35% having BS $ $70. Only nodes within Lamiaceae had BS <50. Interfamilial relationships are well resolved, with only three nodes showing BS <95: the clade comprising Bignoniaceae + (Verbenaceae + Pedaliaceae) (BS=88, PP=1.00), the Acanthaceae + Lentibulariaceae clade (BS=74, PP=0.99), and the Mazaceae-containing clade (BS=82, PP=0.99). These relationships warrant further investigation.

Figure Caption

Figure 1. Phylogenetic tree of Scrophulariaceae and closely related groups based on protein-coding genes. Numbers represent maximum likelihood bootstrap support (BS, first value) and Bayesian posterior probability (PP, second value); nodes with BS=100 and PP=1.00 are not shown; indicates species originally placed in Scrophulariaceae s.l.

References

ALBACH DC, MEUDT HM, OXELMAN B, 2005. Piecing together the "new" Plantaginaceae [J]. Amer J Bot, 92 (2): 297-315.

ALBACH DC, YAN K, JENSEN SR, et al., 2009. Phylogenetic placement of Triaenophora (formerly Scrophulariaceae) with some implications for the phylogeny of Lamiales [J]. Taxon, 58(3):749-756.

BEARDSLEY PM, OLMSTEAD RG, 2002. Redefining Phrymaceae: the placement of Mimulus, tribe Mimuleae, and Phryma [J]. Amer J Bot, 89(7):1093-1102.

BENNETT JR, MATHEWS S, 2006. Phylogeny of the parasitic plant family Orobanchaceae inferred from phytochrome A [J]. Amer J Bot, 93 (7): 1039-1051.

CHENG JP, ZHANG YM, QIAN ZG, et al., 2020. Complete chloroplast genome sequences of Lagotis yunnanensis (Scrophulariaceae): an Endangered species endemic to the Hengduan Mountains region [J]. Mitochondrial DNA Part B, 5 (1): 897-898.

CUSIMANO N, WICKE S. 2016. Massive intracellular gene transfer during plastid genome reduction in nongreen Orobanchaceae [J]. New Phytologist, 210(2): 680-693.

DENG T, LIN N, HUANG X, et al., 2019. Phylogenetics of Mazaceae (Lamiales), with special reference to intrageneric relationships within Mazus [J]. TAXON, 68 (5):1037-1047.

KATOH K, ROZEWICKI J, YAMADA KD, 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization [J]. Brief Bioinform, 20(4):1160-1166.

KUMAR S, STECHER G, TAMURA K, 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets [J]. Mol Biol Evol, 33 (7):1870-1874.

LIU B, TAN YH, LIU S, et al., 2020. Phylogenetic relationships of Cyrtandromoea and Wightia revisited: A new tribe in Phrymaceae and a new family in Lamiales [J]. J Syst Evol, 58(1): 1-17.

MCNEAL JR, BENNETT J R, WOLFE A D, et al., 2013. Phylogeny and origins of holoparasitism in Orobanchaceae [J]. Amer J Bot, 100(5): 971-983.

MILLER MA, PFEIFFER WT, SCHWARTZ T, 2010. Creating the CIPRES science gateway for inference of large phylogenetic trees [C]. 2010 Gateway Computing Environments Workshop (GCE 2010). New Orleans: 1-8.

OLMSTEAD RG, DEPAMPHILIS CW, WOLFE AD, et al., 2001. Disintegration of the Scrophulariaceae [J]. Amer J Bot, 88(2): 348-361.

OLMSTEAD RG, REEVES PA, 1995. Evidence for the polyphyly of Scrophulariaceae based on chloroplast rbcL and ndhF sequences [J]. Ann Miss Bot Gard, 82(2):176-193.

OXELMAN B, KORNHALL P, OLMSTEAD RG, et al., 2005. Further disintegration of Scrophulariaceae [J]. Taxon, 54(2): 411-425.

RAHMANZADEH R, MULLER K, FISCHER E, et al., 2005. The Linderniaceae and Gratiolaceae are further lineages distinct from the Scrophulariaceae (Lamiales) [J]. Plant Biol, 7(1): 67-78.

RONQUIST F, HUELSENBECK JP, 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models [J]. Bioinformatics, 19(12): 1572-1574.

ROURE B, BAURAIN D, PHILIPPE H, 2013. Impact of missing data on phylogenies inferred from empirical phylogenomic data sets [J]. Mol Biol Evol. 30 (1):197-214.

STAMATAKIS A, 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies [J]. Bioinformatics, 30(9): 1312-1313.

TANK DC, BEARDSLEY PM, KELCHNER SA, et al., 2006. Review of the systematics of Scrophulariaceae s.l. and their current disposition [J]. Austral Syst Bot, 19(4): 289-307.

TRIPP EA, TSAI YH, ZHUANG YB, et al., 2017. RADseq dataset with 90% missing data fully resolves recent radiation of Petalidium (Acanthaceae) in the ultra-arid deserts of Namibia [J]. Ecol Evol, 7(19):7920-7936.

WICKE S, SCHAFERHOFF B, DEPAMPHILIS CW, et al., 2013. Disproportional plastome-wide increase of substitution rates and relaxed purifying selection in genes of carnivorous Lentibulariaceae [J]. Mol Biol Evol, 31(3):529-545.

ZHANG YM, QIAN ZG, ZHANG AL, et al., 2019. The complete plastid genome sequence of Neopicrorhiza scrophulariiflora (Plantaginaceae): an endangered species endemic to the Himalayas regions [J]. Mitochondrial DNA B, 4(2): 1-2.

Table 1. Samples included in study, with summary of GenBank accession number and plastome statistics

Species name	GenBank Accession Number	Sequence length	Number of protein-coding genes
Orobanchaceae			
Aeginetia indica	MN529629.1		
Aphyllon epigalium subsp. *notocalifornicum*	MH050786.1		
A. fasciculatum	MH580292.1		
Brandisia swinglei	MK381315.1		
Boulardia latisquama	HG514460.1		
Castilleja paramensis	KT959111.1		
Centranthera grandiflora	MW262988.1		
Cistanche deserticola	KC128846.1		

Species name	GenBank Accession Number	Sequence length	Number of protein-coding genes
C. phelypaea	HG515538.1		
Conopholis americana	HG514459.1		
Epifagus virginiana	M81884.1		
Euphrasia regelii	MK070895.1		
Lathraea squamaria	KM652488.1		
Lindenbergia philippensis	HG530133.1		
Melampyrum koreanum	MW463054.1		
M. roseum	MN075942.1		
Neobartsia inaequalis	KF922718		
Orobanche austrohispanica	KT387721.1		
O. crenata	HG515537.1		
Pedicularis hallaisanensis	MG770330.1		
P. ishidoyana	KU170194.1		
Phelipanche purpurea	HG515536.1		
P. ramosa	HG803180.1		
Phelypaea coccinea	MK922354.1		
Phtheirospermum japonicum	MN075943.1		
Rehmannia elata	KX636161.1		
R. glutinosa	KX636157.1		
Schwalbea americana	HG738866.1		

Species name	GenBank Accession Number	Sequence length	Number of protein-coding genes
Siphonostegia chinensis	MK113828.1		
Triaenophora shennongjiaensis	MH071405.1		
Triphysaria versicolor	MN075944.1		
Paulowniaceae			
Paulownia coreana	KP718622.1		
P. tomentosa	KP718624.1		
Phrymaceae			
Erythranthe lutea	KU705476.1		
Phryma leptostachya subsp. *asiatica*	MK381317.1		
Wightiaceae			
Wightia speciosissima	MK381318.1		
Mazaceae			
Dodartia orientalis	MW238404		
Lancea hirsuta	MG551489.1		
L. tibetica	MF593117.1		
Mazus pumilus	MG642817.1		
Lamiaceae			
Ajuga reptans	KF709391.1		
Callicarpa nudiflora	MK783316.1		
Caryopteris mongholica	MF346535.1		
Dracocephalum palmatum	KU958581.1		

Species name	GenBank Accession Number	Sequence length	Number of protein-coding genes
Galeopsis tetrahit	KY562586.1		
Haplostachys haplostachya	KU724133.1		
Lamium album	KY562589.1		
Lavandula angustifolia	KT948988.1		
Leonurus japonicus	MG673937.1		
Mentha longifolia	KU956042.1		
M. spicata	MG256495.1		
Ocimum basilicum	KY623639.1		
Origanum vulgare subsp. *vulgare*	JX880022.1		
Perilla citriodora	KT220690.1		
Phyllostegia velutina	KU724134.1		
Pogostemon stellatus	KP718620.1		
Premna microphylla	KM981744.1		
Prunella vulgaris	MG589640.1		
Rosmarinus officinalis	NC_{027259}.1		
Salvia japonica	KY646163.1		
S. miltiorrhiza	HF586694.1		
Scutellaria baicalensis	MF521632.1		
Stachys byzantina	KU724141.1		

Species name	GenBank Accession Number	Sequence length	Number of protein-coding genes
Stenogyne bifida	KU724132.1		
Tectona grandis	HF567869.1		
Teucrium mascatense	MH325132.1		
Bignoniaceae			
Adenocalymma pedunculatum	MG008313.1		
Amphilophium carolinae	MK163625.1		
Anemopaegma acutifolium	MF460826.1		
Dolichandra cynanchoides	MG831874.1		
Neojobertia candolleana	MG008316.1		
Pleonotoma albiflora	MG831876.1		
Tanaecium tetragonolobum	KR534325.1		
Tecomaria capensis	MG831880.1		
Verbenaceae			
Aloysia citrodora	KY085903.1		
Duranta erecta	MW525381.1		
Glandularia tenera	MW538952.1		
Lippia origanoides	MK248831.1		
Verbena officinalis	MW328640.1		
Pedaliaceae			
Sesamum indicum	JN637766.2		
Acanthaceae			

Species name	GenBank Accession Number	Sequence length	Number of protein-coding genes
Andrographis paniculata	KF150644.2		
Aphelandra knappiae	MH909777.1		
Clinacanthus nutans	MH778102.1		
Echinacanthus lofouensis	MF490441.1		
Strobilanthes cusia	MG874806.1		
Lentibulariaceae			
Genlisea aurea	MF593121.1		
Pinguicula ehlersiae	HG803178.1		
Utricularia foliosa	KY025562.1		
Linderniaceae			
Torenia benthamiana	MK789686.1		
T. concolor	MK789685.1		
Scrophulariaceae			
Bontia daphnoides	MN044637.1		
Buddleja alternifolia	MN623351.1		
B. colvilei	MH411147.1		
Calamphoreus inflatus	MN044643.1		
Diocirea violacea	MN044644.1		
Eremophila gibbifolia	MN044640.1		
E. oppositifolia	MN044645.1		
Glycocystis beckeri	MN044639.1		
Leucophyllum frutescens	MN044638.1		

Species name	GenBank Accession Number	Sequence length	Number of protein-coding genes
Myoporum bontioides	MN044642.1		
M. laetum	MN044641.1		
Scrophularia buergeriana	KP718626.1		
S. dentata	MF861202.1		
Verbascum chinense	MT610040.1		
V. phoeniceum	MN893301.1		
Plantaginaceae			
Antirrhinum majus	MW877560.1		
Aragoa abietina	MW877561.1		
A. cleefii	MW877562.1		
Bacopa monnieri	MN736955.1		
Callitriche palustris	MW774642.1		
Digitalis lanata	KY085895.1		
Hemiphragma heterophyllum	MN383191.1		
Hippuris vulgaris	MW044609.1		
Lagotis brevituba	MW182582.1		
L. yunnanensis	MN752238.1		
Linaria vulgaris	MT984536.1		
Littorella uniflora	MW877563.1		
Neopicrorhiza scrophulariiflora	MK986819.1		
Plantago maritima	KR297244.1		

Species name	GenBank Accession Number	Sequence length	Number of protein-coding genes
P. media	KR297245.1		
Veronica nakaiana	KT633216.1		
V. persica	KT724052.1		
Veronicastrum axillare	MW244757.1		
V. sibiricum	KT724053.1		
Gesneriaceae			
Boea hygrometrica	JN107811.1		
Chirita eburnea	MF177038.1		
Oleaceae			
Abeliophyllum distichum	KT274029.1		
Chionanthus parkinsonii	MG255752.1		
Forsythia suspensa	MF579702.1		

Note: indicates species originally placed in Scrophulariaceae s.l.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.