

The complete chloroplast genome sequence of an Endangered orchid species *Dendrobium huoshanense* (Orchidaceae) and phylogenetic analysis with related species

Bin Wang (✉ happywangbin2003@163.com)

East China University of Technology

Jun Chen

East China University of Technology

Min Li

East China University of Technology

Research Article

Keywords: *Dendrobium huoshanense*, chloroplast genome, genome assembly, phylogenetic analysis

Posted Date: April 11th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1527666/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The family Orchidaceae, including many medicinal plants, is a famous ornamental plant. While, *Dendrobium huoshanense* is an endangered Orchidaceae species mainly distributed in Huoshan county, Anhui province, China. Here, using the program MITObim v1.7, we assembled the complete chloroplast genome sequence of *D. huoshanense*, which consists of 153,188 bp, and a pair of inverted repeats (IR) of 26,525 bp. The genome also includes one small and one large single-copy region (SSC and LSC) of 14,524 bp and 85,614 bp, respectively. GC content of the genome is 40.0%, and it contains 129 genes coding for 76 proteins. Codon usage analysis showed that a total of 23,448 codons were identified, and the simple-sequence repeat sequences mainly contain A/T mononucleotide. Phylogenetic analyses revealed that *D. huoshanense* belongs to the *Dendrobium* of the Orchidaceae family, which is in accordance with the traditional classification. The research finding in this paper will be beneficial for further investigations on *Dendrobium huoshanense* from the aspect of evolution, and chloroplast genetic engineering.

Introduction

Dendrobium huoshanense is a most familiar traditional Chinese medicines, which belongs to the *Dendrobium* of the highly diverse Orchidaceae family. Due to its extremely demanding for growth environment, the *D. huoshanense* in wild is rare and endangered. Thus, *D. huoshanense* has become an endangered medicinal plant with an enormous pharmaceutical and medicinal value (Silva et al. 2015). So far, more and more complete chloroplast genome of the *Dendrobium* plants such as *Dendrobium pendulum*, *Dendrobium moniliforme*, and *Dendrobium bellatulum* have been reported (Gao et al., 2017; Wang et al., 2017; Zhang et al., 2018). Whereas, the information about *D. huoshanense* chloroplast genome content is limited.

In plant cells, the chloroplast is a very important organelle, and take a significant role in the process of photosynthesis, and also has a vital function in the biosynthesis of fatty acids and starch (H. et al., 2000). Through photosynthesis, the atmospheric CO₂ was converted into carbohydrates, and the light energy is stored as chemical energy in this process. Thus, the chloroplast is also called “energy converters” (Raveendar et al. 2015). Additionally, the plant chloroplast genome is also named chloroplast DNA, which is circular in shape, and contains double strands. Moreover, it encodes genes that are photosynthesis and specific essential metabolic pathways (Kong and Yang 2015).

In 1986, the complete chloroplast genomes of tobacco and liverwort were firstly published (Shinozaki et al. 1986). And then, with the development of molecular biological techniques and cost-effective next-generation sequencing approaches, more and more complete chloroplast genomes from different species have been reported. Moreover, in 2009, the Consortium for the Barcode of Life's Plant Working Group recommended the combination of two-locus chloroplast genes *rbcL* + *matK* as the plant “barcode” (Group 2009). Furthermore, the chloroplast genome was used as DNA barcoding (Provan et al. 2001) for phylogenetic analyses on account of the smaller and more conserved genomic sequence (Qian et al.,

2013). Thus, in recent studies, the complete chloroplast genome was widely used as a barcode for taxonomy and identification of plant systems (Hollingsworth et al., 2011; Xiwen et al., 2014).

Here, the complete chloroplast genome of *D. huoshanense* was reported. Then, the codon usage, simple sequence repeats and IR junctions of *D. huoshanense* were analyzed. And then, the complete chloroplast genome comparison of five *Dendrobium* species was analyzed. We also analyzed the phylogenetic relationship between *D. huoshanense* and other Orchidaceae categories. The major purpose is to contribute to furnishing resourced organelle genome of *Dendrobium*, which will be beneficial for species identification, germplasm diversity, and genetic engineering.

Materials And Methods

DNA extraction and sequencing

The fresh leaves of *D. huoshanense* for the total genome DNA extraction were collected in the glasshouse of National Engineering laboratory for Resource Developing of Endangered Chinese Crude Drugs in Northwest of China (108°53′30"E, 34°9′14"N). The *D. huoshanense* total genomic DNA was extracted, and then was used to construct shotgun libraries with the NEBNext[™] Ultra[™] DNA Library Prep Kit for Illumina[®]. The whole-genome sequencing was executed by Shanghai Genesky Biotechnologies Inc. (Shanghai, China) with an Illumina HiSeq 2500 Sequencing System (Illumina, San Diego, CA, USA).

Chloroplast genome assembly and annotation

After the chloroplast genome is sequenced, the whole raw reads were quality trimmed using CLC Genomics Workbench v. 7.5 under default parameters (Patel and Mukesh 2012). Then, the complete chloroplast genome was constructed with the trimmed reads using the program MITObim v. 1.7 (Christoph et al., 2013), and the *Dendrobium catenatum* chloroplast genome (accession number: KC771275) was used as the initial reference. And then, the program GENEIOUS R8 (Kearse et al., 2012) was used to annotate the assembled sequence by comparing it with the initial reference.

Codon Usage, simple sequence repeats and comparative genome analyses

With the CodonW software, the distribution of codon usage with the RSCU ratio was detected (Sharp and Li 1987). Meanwhile, with the MISA software (Sebastian et al. 2017), the simple sequence repeats (SSRs) were investigated with the corresponding parameters (Li et al. 2013a). Additionally, with the mVISTA tool (Sebastian et al. 2017) and the Shuffle-LAGAN alignment algorithm, the whole-genome alignment for the chloroplast genomes of the four species including *Dendrobium strongylanthum* (GenBank:KR673323), *Dendrobium pendulum* (GenBank:KT695604), *Dendrobium officinale* (GenBank:KJ862886) and *Dendrobium huoshanense* (GenBank:KT630834) were performed and plotted with the *Dendrobium catenatum* (GenBank: KC771275) chloroplast genomes the initial reference.

Phylogenetic analyses

To determine the phylogenetic positions of *D. huoshanense*, we obtained 21 Orchidaceae species (Supplementary Table S1) complete chloroplast genomes sequences from the NCBI database, firstly. Then, we constructed a phylogenetic tree with the chloroplast genomes protein-coding genes of those species using the MEGA6 (Tamura et al., 2013) program under the neighbor-joining (NJ) method.

Results

The chloroplast genome characteristics of *D. huoshanense*

Once the *D. huoshanense* chloroplast genome was sequenced and annotated, we submitted it to the NCBI database, and got the GenBank number KT630834. The length of *D. huoshanense* chloroplast genome is 153,188 bp, and it has an 85,614 bp LSC (large single-copy) region and a 14,524 bp SSC (small single-copy) region, respectively. It also has a pair of 26,525 bp IR (inverted repeat) regions (Fig. 1). In addition, the *D. huoshanense* chloroplast genome nucleotide composition is asymmetrical including 30.0% A, 20.1% C, 22.9% G and 27.0% T, and the GC contents is 37.5% (Table 1). Meantime, the highest values of CG contents in the IR regions are 43.4%, and the values of CG contents in the LSC and SSC regions are 35.0% and 30.5%, respectively.

Table 1
Length of the whole chloroplast genome of *D. huoshanense* and its base composition

Regions	T(U) (%)	C (%)	A (%)	G (%)	Length (bp)
LSC	33.2	17.9	31.8	17.2	85,614
SSC	36.3	15.9	33.2	14.6	14,524
IRa	28.3	22.4	28.3	21.0	26,525
IRb	28.3	22.4	28.3	21.0	26,525
Total	31.8	19.0	30.7	18.5	153,188

Furthermore, we found 129 genes (111 species), including 76 protein-coding genes (69 species) in the *D. huoshanense* chloroplast genome. It also contains 38 transfer RNA genes (30 species) and eight ribosomal RNA genes (four species) (Table 2). We also found 17 genes have introns, including *atpF*, *ycf3*, *clpP*, *petB*, *petD*, *rpl2*, *ndhB*, *rps16*, *rps12*, *rpoC1* and *rpl16* and 6 tRNA genes (Table 3). Otherwise, we found that 14 genes only have a single intron, and *rps12*, *clpP* and *ycf3* contain 2 introns (Table 3).

Table 2
Gene contents in the chloroplast genomes of *D. huoshanense*

No	Group of Genes	Gene names	Total
1	ATP synthase	<i>atpA,atpB,atpE ,atpF ,atpH,atpI</i>	6
2	NADH dehydrogenase	<i>ndhA, ndhB (x2) ,ndhE, ndhF(x2), ndhG, ndhH ,ndhJ</i>	9
3	Cytochrome b/f complex	<i>petA, petB,petD ,petG, petL, petN</i>	6
4	Photosystem I	<i>psaA, psaB, psaC, psal, psaJ</i>	5
5	Photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbl, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>	15
6	RubisCO large subunit	<i>rbcl</i>	1
7	Ribosomal proteins (LSU)	<i>rpl14,rpl16, rpl2(x2),rpl20,rpl22,rpl23(x2),rpl32, rpl33, rpl36</i>	11
8	RNA polymerase	<i>rpoA,rpoB,rpoC1,rpoC2</i>	4
9	Ribosomal proteins (SSU)	<i>rps11,rps12(x2),rps14,rps15,rps16,rps18,rps19(x2) ,rps2,rps3,rps4, rps7(x2),rps8</i>	15
10	Ribosomal RNAs	<i>rrn16(x2),rrn23(x2) rrn4.5(x2), rrn5(x2)</i>	8
11	Transfer RNAs	<i>38tRNAs</i>	38
12	Proteins of unknown function	<i>ycf1, ycf2(x2) ycf3,ycf4</i>	5
13	Other genes	<i>accD, ccsA,cemA,clpP, infA, matK</i>	6

Table 3

Genes with introns in the chloroplast genomes of *D. huoshanense* as well as the lengths of the exons and introns.

Gene	Location	Exon I(bp)	Intron I(bp)	Exon II(bp)	Intron II(bp)	Exon III(bp)
<i>trnK(uuu)</i>	LSC	35	2815	37		
<i>trnG(ucc)</i>	LSC	31	671	59		
<i>trnL(uaa)</i>	LSC	35	800	50		
<i>trnV(uac)</i>	LSC	37	587	39		
<i>trnI(gau)</i>	IR	37	944	35		
<i>trnA(ugc)</i>	IR	38	801	35		
<i>rps16</i>	LSC	248	892	40		
<i>atpF</i>	LSC	410	924	145		
<i>rpoC1</i>	LSC	1611	762	453		
<i>ycf3</i>	LSC	153	723	230	747	124
<i>clpP</i>	LSC	231	966	292	672	71
<i>petB</i>	LSC	6	732	642		
<i>petD</i>	LSC	8	862	484		
<i>rpl16</i>	LSC	399	1197	9		
<i>rpl2</i>	IR	431	714	385		
<i>ndhB</i>	IR	756	699	777		
<i>rps12</i>	LSC	138	-	229	547	26

Codon Usage Analysis

According to the protein-coding genes sequences, we estimated the codon usage frequency of *D. huoshanense* chloroplast genome with the CodonW software. The result showed that there are 64 codons encoding 20 amino acids and 23,448 codons identified in the protein-coding genes of the *D. huoshanense* chloroplast genome (Supplementary Table S2). The results also indicated that in *D. huoshanense* chloroplast genome, the most universal amino acids is 2,369 encode leucine, and the least universal amino acid is 276 encode cysteine (Supplementary Table S2). In addition, the results revealed that the value of relative synonymous codon usage (RSCU) increased with the quantity of codons that code for a specific amino acid, such as leucine, serine and arginase (Supplementary Table S2). Furthermore, in addition to the tryptophan and methionine, most of amino acid codons have preferences (Supplementary Table S2).

Simple sequence repeats analyses

Simple sequence repeats (SSRs), generally including 1–6 nucleotide repeat units, also named as microsatellites, are tandem repeated DNA sequences, which are ubiquitous among the genomes (Powell et al. 1995). SSRs are universally used for molecular markers in species identification, phylogenetic investigations and population genetics for the high level of polymorphism (Jiao et al. 2012; Xue et al. 2012; Yang et al. 2011). Using the microsatellite identification tool (MISA), there were 165 SSRs identified, and the largest SSRs in number is mononucleotide throughout these SSRs, a total of 105 were found (Table 4). A/T mononucleotide repeats (96.1%) were the most common SSRs, and the majority of dinucleotide repeat sequences were the AT/TA repeats (55.6%) (Table 3). These results are in accord with the previous studies where proportions of polythymine (polyT) and polyadenine (polyA) were higher than polyguanine (polyG) and polycytosine (polyC) within chloroplast SSRs in many plants (Kuang et al. 2011; Zhou et al. 2017).

Table 4
Types and amounts of SSRs in the *D. huoshanense* chloroplast genomes

SSR Type	Repeat Unit	Amount	Ratio(%)
Mono	A/T	105	96.3%
	C/G	4	3.7%
Di	AC/GT	3	5.6%
	AG/CT	21	38.9%
	AT/AT	30	55.5%
Tri	AAC/GTT	8	11.3%
	AAG/CTT	28	39.4%
	AAT/ATT	12	16.9%
	ACC/GGT	2	2.8%
	ACT/AGT	4	5.6%
	AGC/CTG	4	5.6%
	AGG/CCT	6	8.5%
	ATC/ATG	7	9.9%
Tetra	AAAG/CTTT	3	50.0%
	AATT/AATT	1	16.7%
	ACAG/CTGT	1	16.7%
	AGAT/ATCT	1	16.7%
Penta	AATCC/ATTGG	1	50.0%
	ATATC/ATATG	1	50.0%
Hexa	ACAGAT/ATCTGT	1	100%

IR junctions and comparative genome analyses

Based on the previously study, the contraction and expansion at the borders of the inverted repeat regions are widely events during evolution, and stand for the primary reasons for size variation of the chloroplast genome (Yang et al. 2010; Raubeson et al. 2007). We compared the junctions of the LSC and IR regions of four species including *Nicotiana tabacum* (GenBank: NC_001879), *Arabidopsis thaliana* (GenBank: NC_000932), *Dendrobium huoshanense* and the reference *Dendrobium catenatum*. The results showed

that the SSC/IRa border was situated in the CDS of the *ycf1* gene in the four chloroplast genomes, and the *trnH* genes were not located in the LSC regions in *D. huoshanense* and *D. catenatum* (Fig. 2). These results are in line with those described for members of Orchidaceae (Jing et al. 2014).

With the mVISTA program, the complete chloroplast genome of *D. huoshanense* was compared to those of *D. pendulum* (Wang et al. 2017), *D. strongylanthum* (Jing et al. 2015), *D. officinale* (Yang et al. 2016) and *D. catenatum* (Jing et al. 2014) (Fig. 3). The finding revealed that the two IR districts were less conserved than the SSC and LSC districts, and the four rRNA genes were more conserved than other genes. In addition, the results also showed that coding districts appeared a higher conserved than non-coding districts, and the most divergent districts were localized in the intergenic spacers throughout the five chloroplast genomes (Fig. 3).

Phylogenetic analyses

Due to the quickly development of high-throughput sequencing technology, much more complete chloroplast plant genomes have been reported (Curci et al. 2015), and the plant chloroplast genomes provide abundant resources for the phylogenetic studies, taxonomic and evolutionary (Jansen et al. 2008; Moore et al. 2007; Qian et al. 2013). Furthermore, more and more phyloevolutionary relationships at almost any taxonomic level were successfully resolved, depending on the protein-coding genes or whole chloroplast genomes (Li et al. 2013b). Orchidaceae family is a largest family in the Kingdom Plantae, and *Dendrobium* is the most important medicinal plant of this family. According to the chloroplast sequences of *Dendrobium* plants, much information about the phylogenetic relationships of this category have been reported (Jing et al. 2015; Wang et al. 2017; Yang et al. 2016).

Therefore, to identify the phylogenetic positions of *D. huoshanense* within Orchidaceae family, 21 species from *Phalaenopsis*, *Corallorhiza*, *Cymbidium*, *Dendrobium* as well as *Paphiopedilum* were used to construct a phylogenetic tree with the neighbor-joining (NJ) method. The results showed that the bootstrap values for all of the clades subgroups were high, suggesting that genes in the same subgroups may share a similar origin, and *D. huoshanense* is the part of *Dendrobium* of Orchidaceae family, which is in accord with the traditional systematic plant classification (Fig. 4).

Discussion

With the development of Next Generation Sequencing (NGS), we could get more biological information for the species identification, molecular genetic markers and evolution within and between different species (W. et al., 1995; Grassi et al., 2002; Leonie et al., 2011; Straub et al., 2011). Therefore, the complete chloroplast genome also could furnish abundant genetic information and molecular markers that are valuable tools to solve obscure phylogenetic relationships among land plants (Luo et al., 2014; Alzahrani et al., 2021).

Here, using the Illumina sequencing platform and GENEIOUS program, we firstly sequenced and assembled the complete chloroplast genome of *D. huoshanense*. Reported *Dendrobium* complete

chloroplast genomes range in size from 156,612 to 156,781 bp (Wang et al., 2017; Yan-xia et al., 2018; Zhang et al., 2018), and the size of assembled chloroplast genome of *D.huoshanense* is 153,188 bp, which is consistent with those reported previously in plants of the same species. The average GC content of the *D. huoshanense* chloroplast genome is 40.0%, similar to other *Dendrobium* species (Gao et al.; Wang et al., 2017; Zhang et al., 2018). It suggested that this method which was used for assembling the plant complete chloroplast genome is feasible. Furthermore this method is very more convenient and efficient than other methods.

In this newly determined chloroplast genome, 129 predicted genes were found. There are 42 genes are involved in photosynthesis, of which 6 encode different subunits of ATP synthase, 9 for the subunits of the NADH-oxidoreductase, 6 for the cytochrome b6/f complex, 5 for photosystem I, 15 for photosystem II, and 1 for the Rubisco. In addition, there are 38 tRNA genes, 36 ribosomal subunit genes (15 small subunits and 11 large subunits), 8 rRNA genes and 4 RNA polymerase genes. Furthermore, 5 genes encode miscellaneous proteins which were involved in different functions, and 6 genes were of unknown function (Table 5). Those results also showed that the genome organization appeared to be more conserved with unique gene sequence as discovered previously in *Dendrobium* species (Wang et al., 2017; Yan-xia et al., 2018; Zhang et al., 2018).

Table 5
Genes present in the *D. huoshanense* chloroplast genome.

Group	Name of genes	Numbers
Photosystem I	<i>psaA, psaB, psaC, psal, psaJ</i>	5
Photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbl, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>	15
Cytochrome b6/f	<i>petA, petB, petD, petG, petL, petN</i>	6
ATP synthase	<i>atpA, atpB, atpE, atpF, atpH, atpI</i>	6
NADH dehydrogenase	<i>ndhA, ndhB (×2), ndhD, ndhE, ndhF, ndhG, ndhH, ndhJ</i>	9
Rubisco	<i>rbcL</i>	1
RNA polymerase	<i>rpoA, rpoB, rpoC1, rpoC2</i>	4
Small subunits of ribosome	<i>rps2, rps3, rps4, rps7(×2), rps8, rps11, rps12(×2), rps14, rps15, rps16, rps18, rps19(×2)</i>	15
Large subunits of ribosome	<i>rpl2(×2), rpl14, rpl16, rpl20, rpl22, rpl23(×2), rpl32, rpl33, rpl36</i>	11
Other genes	<i>accD, ccsA, cemA, clpP, infA, matK</i>	6
Miscellaneous proteins coding gene	<i>ycf1, ycf2(×2), ycf3, ycf4</i>	5
tRNAs	<i>trnA-UGC(×2), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnFM-CAU, trnG-GCC, trnG-UCC, trnH-GUG(×2), trnI-CAU(×2), trnI-GAU(×2), trnK-UUU, trnL-CAA(×2), trnL-UAA, trnL-UAG, trnM-CAU, trnN-GUU(×2), trnP-UGG, trnQ-UUG, trnR-ACG(×2), trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC, trnV-GAC, trnV-UAC, trnW-CCA, trnY-GUA</i>	38
rRNAs	<i>rrn4.5(×2), rrn5(×2), rrn16(×2), rrn23(×2)</i>	8

Additionally, we also found that there are 23,448 codons encoding the genes in the *D. huoshanense* chloroplast genome, and the most universal codons are the coding for the amino acids Leucine, which has been previously reported in some complete chloroplast genomes of plants (Liu et al., 2018; Alzahrani et al., 2021). Furthermore, there are 165 simple sequence repeats (SSRs) markers found in the 153,188 bp sequence of the *D. huoshanense* chloroplast genome. In other words, the observed frequency of SSRs was approximately 1/928.4 bp of chloroplast genome. As previously study, the SSRs were also only observed in the non-coding region of the chloroplast genome (Raveendar et al., 2015).

With the mVISTA program, we examined the degree of DNA sequence divergence in the four-chloroplast genome. In addition to the genes of *ycf1*, *rps19*, *rpoC2*, and *atpF*, many different non-coding regions among the four chloroplast genomes. We also found that the gene-coding regions are more highly

conserved than those of their non-coding districts, which was similar to the other *Dendrobium* chloroplast genome. Otherwise, the IR-LCS and IR-SSC boundaries of four chloroplast genome plants were compared. The results displayed that *N. tabacum* has the greatest chloroplast genome from the four plants, while, *D. catenatum* has the smallest chloroplast genome in size. The largest LSC region in *N. tabacum* and the smallest in *A. thaliana*, The largest SSC region in *N. tabacum* and the smallest in *D. catenatum*, the largest IR region in *D. huoshanense* and the smallest in *N. tabacum*.

Phylogenetic relationships based on the chloroplast genomes protein-coding genes placed all samples into five main clades, where every family is in a separate clade (Fig. 4). The results advised that all species in the same clade could be clustered into the same genus, which indicated that the relationships and phylogenetic positions of this family could be resolved by the chloroplast genome. Nevertheless, it should use different methods to analyze the phylogeny for accurately clarify the evolution of the Orchidaceae family (Fig. 4). This study could offer a reference for plant classification among *Dendrobium*, and other genus with using the chloroplast genome.

Conclusions

D. huoshanense is a traditional Chinese herbal medicine with important pharmaceutical and economic value. Here, we assemble the complete chloroplast genome of *D. huoshanense*, and revealed the genome characteristics and gene content. These studies enhance the genomic information for *Dendrobium* plants, and contribute to the study of germplasm diversity. These data also provide a valuable resource of labels for future research on Orchidaceae family. In addition, the genome sequence in chloroplast also offers more information on functional protein variability about the chloroplast.

Declarations

Acknowledgements The authors are thankful to the National Engineering Laboratory for Resource Development of Endangered Crude Drugs in Northwest China, for providing plant material to carry out this research work.

Funding This study was carried out with the support of the Jiangxi province project Education Fund (Project no: GJJ170434), National Natural Science Foundation of China (Project no: 21968001 and 41562021)

Author Contributions BW initiated the project, and BW and JC conducted experiments and drafted the manuscript. Bioinformatic analyses were performed by ML.

Availability of Data and Materials All the data and plant material are available with the corresponding author.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. Alzahrani, D., Albokhari, E., Yaradua, S. and Abba, A., 2021. Complete chloroplast genome sequences of *Dipterygium glaucum* and *Cleome chrysantha* and other Cleomaceae Species, comparative analysis and phylogenetic relationships. *Saudi J Biol Sci* 28, 2476–2490.
2. Christoph, H., Lutz, B. and Bastien, C., 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Research* 41, e129.
3. Curci, P.L., Domenico, D.P., Donatella, D., Vendramin, G.G., Gabriella, S. and Jean-Fran?ois, P., 2015. Complete Chloroplast Genome of the Multifunctional Crop Globe Artichoke and Comparison with Other Asteraceae. *Plos One* 10, e0120589-.
4. Gao, Y.X., Zhou, Y.-y., Xie, Y., Feng, L. and Shen, S.-g. The complete chloroplast genome sequence of an endangered Orchidaceae species *Dendrobium monilforme* and its phylogenetic implications. *Conservation Genetics Resources*.
5. Gao, Y.X., Zhou, Y.Y., Xie, Y., Feng, L. and Shen, S.G., 2017. The complete chloroplast genome sequence of an endangered Orchidaceae species *Dendrobium monilforme* and its phylogenetic implications. *Conservation Genetics Resources*.
6. Grassi, F., Labra, M., Scienza, A. and Imazio, S., 2002. Chloroplast SSR markers to assess DNA diversity in wild and cultivated grapevines. *Vitis -Geilweilerhof-* 41, 157–158.
7. Group, C.P.W., 2009. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America* 106, 12794–12797.
8. H., E., Neuhaus, and, M., J. and Emes, 2000. NONPHOTOSYNTHETIC METABOLISM IN PLASTIDS. *Annual Review of Plant Physiology & Plant Molecular Biology*.
9. Hollingsworth, P.M., Graham, S.W. and Little, D.P., 2011. Choosing and Using a Plant DNA Barcode. *Plos One* 6, e19254.
10. Jansen, R.K., Cai, Z., Raubeson, L.A., Daniell, H. and Boore, J.L., 2008. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proceedings of the National Academy of Sciences* 104, 19369–19374.
11. Jiao, Y., Jia, H.M., Li, X.W., Chai, M.L. and Gao, Z.S., 2012. Development of simple sequence repeat (SSR) markers from a genome survey of Chinese Bayberry (*Myrica rubra*). *BMC Genomics* 13, 201.
12. Jing, Li, Chen, Chen, Zhe-Zhi and Wang, 2015. The complete chloroplast genome of the *Dendrobium strongylanthum* (Orchidaceae: Epidendroideae): Mitochondrial DNA Part A: Vol 27, No 4. *Mitochondrial DNA Part A*.
13. Jing, Luo, Bei-Wei, Hou, Zhi-Tao, Niu, Wei, Liu, Qing-Yun and Xue, 2014. Comparative Chloroplast Genomes of Photosynthetic Orchids: Insights into Evolution of the Orchidaceae and Development of Molecular Markers for Phylogenetic Applications. *PLoS ONE* 9, e99016.

14. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S. and Duran, C., 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–9.
15. Kong, W. and Yang, J., 2016. The complete chloroplast genome sequence of *Morus mongolica* and a comparative analysis within the Fabidae clade. *Curr Genet* 62, 165–72.
16. Kuang, D.Y., Wu, H., Wang, Y.L., Gao, L.M. and Lu, L., 2011. Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): implication for DNA barcoding and population genetics. *Genome* 54, 663–673.
17. Leonie, D., Barbara, G., Yuri, L., Yavuz, A., Thomas, C.A.W. and Klaas, V., 2011. The Complete Chloroplast Genome of 17 Individuals of Pest Species *Jacobaea vulgaris*: SNPs, Microsatellites and Barcoding Markers for Population and Phylogenetic Studies. *DNA Research* 18, 93–105.
18. Li, X., Gao, H., Wang, Y., Song, J., Henry, R., Wu, H., Hu, Z., Yao, H., Luo, H., Luo, K., Pan, H. and Chen, S., 2013a. Complete chloroplast genome sequence of *Magnolia grandiflora* and comparative analysis with related species. *Sci China Life Sci* 56, 189–98.
19. Li, X., Zhang, T.-C., Qiao, Q., Ren, Z., Zhao, J., Yonezawa, T., Hasegawa, M., Crabbe, M.J.C., Li, J. and Zhong, Y., 2013b. Complete chloroplast genome sequence of holoparasite *Cistanche deserticola* (Orobanchaceae) reveals gene loss and horizontal gene transfer from its host *Haloxylon ammodendron* (Chenopodiaceae). *PloS one* 8, e58747.
20. Liu, H.-J., Ding, C.-H., He, J., Cheng, J.I.N., Pei, L.Y. and Xie, L.E.I., 2018. Complete chloroplast genomes of *Archiclematis*, *Naravelia* and *Clematis* (Ranunculaceae), and their phylogenetic implications. *Phytotaxa* 343, 214.
21. Luo, J., Hou, B.W., Niu, Z.T., Liu, W., Xue, Q.Y. and Ding, X.Y., 2014. Comparative chloroplast genomes of photosynthetic orchids: insights into evolution of the Orchidaceae and development of molecular markers for phylogenetic applications. *PLoS One* 9, e99016.
22. Moore, M.J., Bell, C.D., Soltis, P.S. and Soltis, D.E., 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proc Natl Acad Sci U S A* 104, 19363–19368.
23. Patel, R.K. and Jain, M., 2012. NGS QC Toolkit: A Toolkit for Quality Control of Next Generation Sequencing Data. *Plos One* 7, e30619.
24. Powell, Morgante, McDevitt, G. G., Vendramin, A. J. and Rafalski, 1995. Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. *Proceedings of the National Academy of Sciences of the United States of America*.
25. Provan, J. and Al, E., 2001. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends in Ecology & Evolution* 16, 142–147.
26. Qian, J., Song, J., Gao, H., Zhu, Y., Xu, J., Pang, X., Yao, H., Sun, C., Li, X. and Li, C., 2013. The Complete Chloroplast Genome Sequence of the Medicinal Plant *Salvia miltiorrhiza*. *Plos One* 8, e57607.
27. Raubeson, L.A., Peery, R., Chumley, T.W., Dziubek, C., Fourcade, H.M., Boore, J.L. and Jansen, R.K., 2007. Comparative chloroplast genomics: analyses including new sequences from the angiosperms

- Nuphar advena and Ranunculus macranthus. BMC Genomics 8, 174.
28. Raveendar, S., Na, Y.W., Lee, J.R., Shim, D., Ma, K.H., Lee, S.Y. and Chung, J.W., 2015. The complete chloroplast genome of *Capsicum annuum* var. *glabriusculum* using Illumina sequencing. *Molecules* 20, 13080–8.
 29. Sebastian, Beier, Thomas, Thiel, Münch, Uwe, Scholz, Martin and Mascher, 2017. MISA-web: a web server for microsatellite prediction. *Bioinformatics*.
 30. Sharp, P.M. and Li, W.H., 1987. The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic acids research* 15, 1281–1295.
 31. Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T. and Sugiura, M., 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *Plant Molecular Biology Reporter* 5, 2043–2049.
 32. Silva, J.A.T.D., Tsavkelova, E.A., Ng, T.B., Parthibhan, S. and Zeng, S., 2015. Asymbiotic in vitro seed propagation of *Dendrobium*. *Plant Cell Reports* 34.
 33. Straub, S.C.K., Parks, M., Weitemier, K., Fishbein, M. and Liston, A., 2011. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany* 99, 349–364.
 34. Tamura, K., Stecher, G., Peterson, D., Filipiński, A. and Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology & Evolution* 44, 2725.
 35. W., Powell, M., Morgante, R., McDevitt, G., G., Vendramin and J., 1995. Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. *Proceedings of the National Academy of Sciences* 92, 7759–7763.
 36. Wang, L.R., Wang, S.Q., Niu, J.F., Wang, S.P. and Wang, Z.Z., 2017. The complete chloroplast genome sequence of an endangered species, *Dendrobium pendulum* Roxb. (Orchidaceae). *Conservation Genetics Resources*.
 37. Xiwen, Li, Yang, Yang, Robert, J., Henry, Maurizio, Rossetto and Yitao, 2014. Plant DNA barcoding: from gene to genome. *Biological Reviews*.
 38. Xue, J., Wang, S. and Zhou, S.L., 2012. Polymorphic chloroplast microsatellite loci in *Nelumbo* (Nelumbonaceae). *American Journal of Botany* 99.
 39. Yan-xia, Gao, Yu-yan, Zhou, Ying, Xie, Li, Feng, Shi-gang and Shen, 2018. The complete chloroplast genome sequence of an endangered Orchidaceae species *Dendrobium moniliforme* and its phylogenetic implications. *Conservation Genetics Resources*.
 40. Yang, A.H., Zhang, J.J., Yao, X.H. and Huang, H.W., 2011. Chloroplast microsatellite markers in *Liriodendron tulipifera* (Magnoliaceae) and cross-species amplification in *L. chinense*. *American Journal of Botany* 98.
 41. Yang, M., Zhang, X., Liu, G., Yin, Y. and Yu, J., 2010. The Complete Chloroplast Genome Sequence of Date Palm (*Phoenix dactylifera* L.). *PLoS ONE* 5, e12762.

42. Yang, P., Zhou, H., Qian, J., Xu, H., Shao, Q., Li, Y. and Yao, H., 2016. The complete chloroplast genome sequence of *Dendrobium officinale*. *Dna Sequence* 27, 1262–1264.
43. Zhang, Y.-J., Ma, C., Feng, Y., Cheng, X. and Song, J., 2018. The complete chloroplast genome sequence of an Endangered orchid species *Dendrobium bellatulum* (Orchidaceae). *Mitochondrial DNA Part B* 3, 233–234.
44. Zhou, J., Chen, X., Cui, Y., Sun, W., Li, Y., Wang, Y., Song, J. and Yao, H., 2017. Molecular Structure and Phylogenetic Analyses of Complete Chloroplast Genomes of Two *Aristolochia* Medicinal Species. *Int J Mol Sci* 18.

Figures

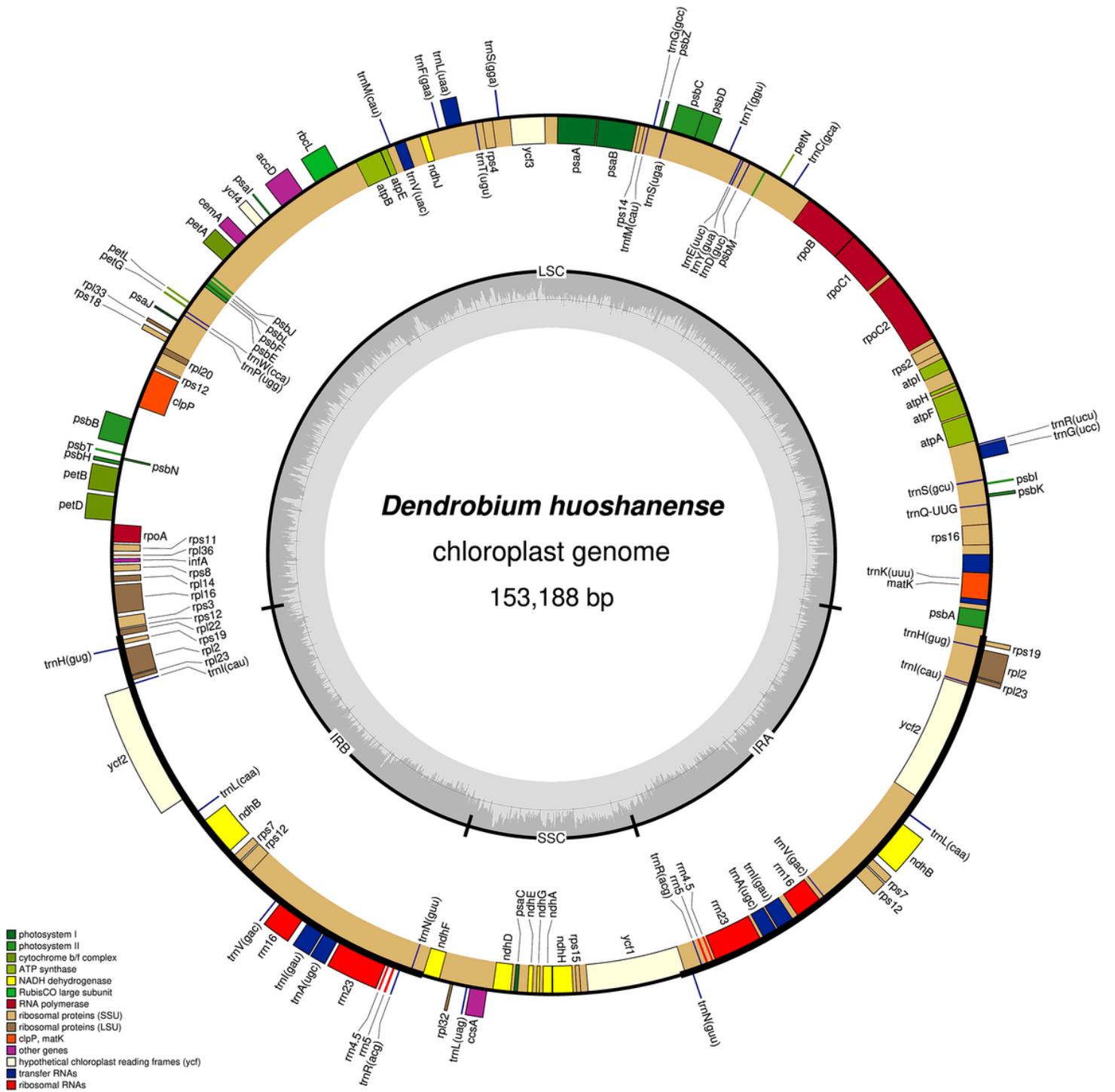


Figure 1

Gene map of the *Dendrobium huoshanense* chloroplast genome.

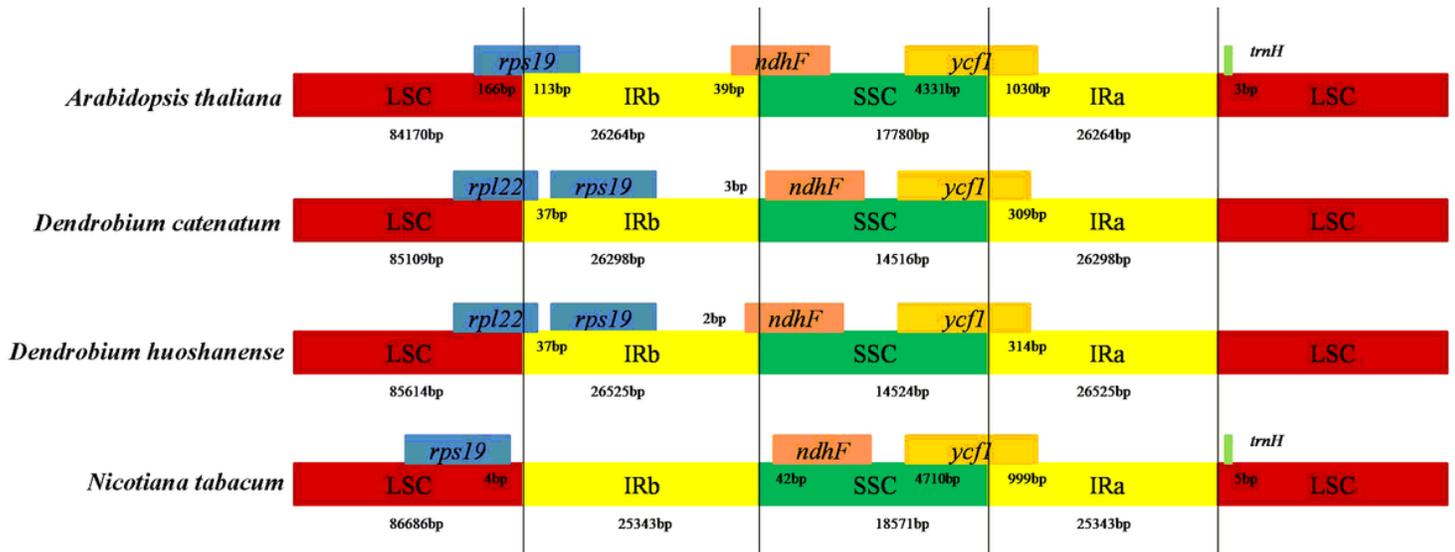


Figure 2

Comparison of the borders of LSC, SSC and IR regions among four chloroplast genomes. Number above the gene features means the distance between the ends of genes and the borders sites.

Figure 3

Sequence identity plot comparing the five chloroplast genomes with *D. catenatum* as a reference by using mVISTA. Grey arrows and thick black lines above the alignment indicate genes with their orientation and the position of the IRs, respectively. A cut-off of 70% identity was used for the plots, and the Y-scale represents the percent identity ranging from 50% to 100%.

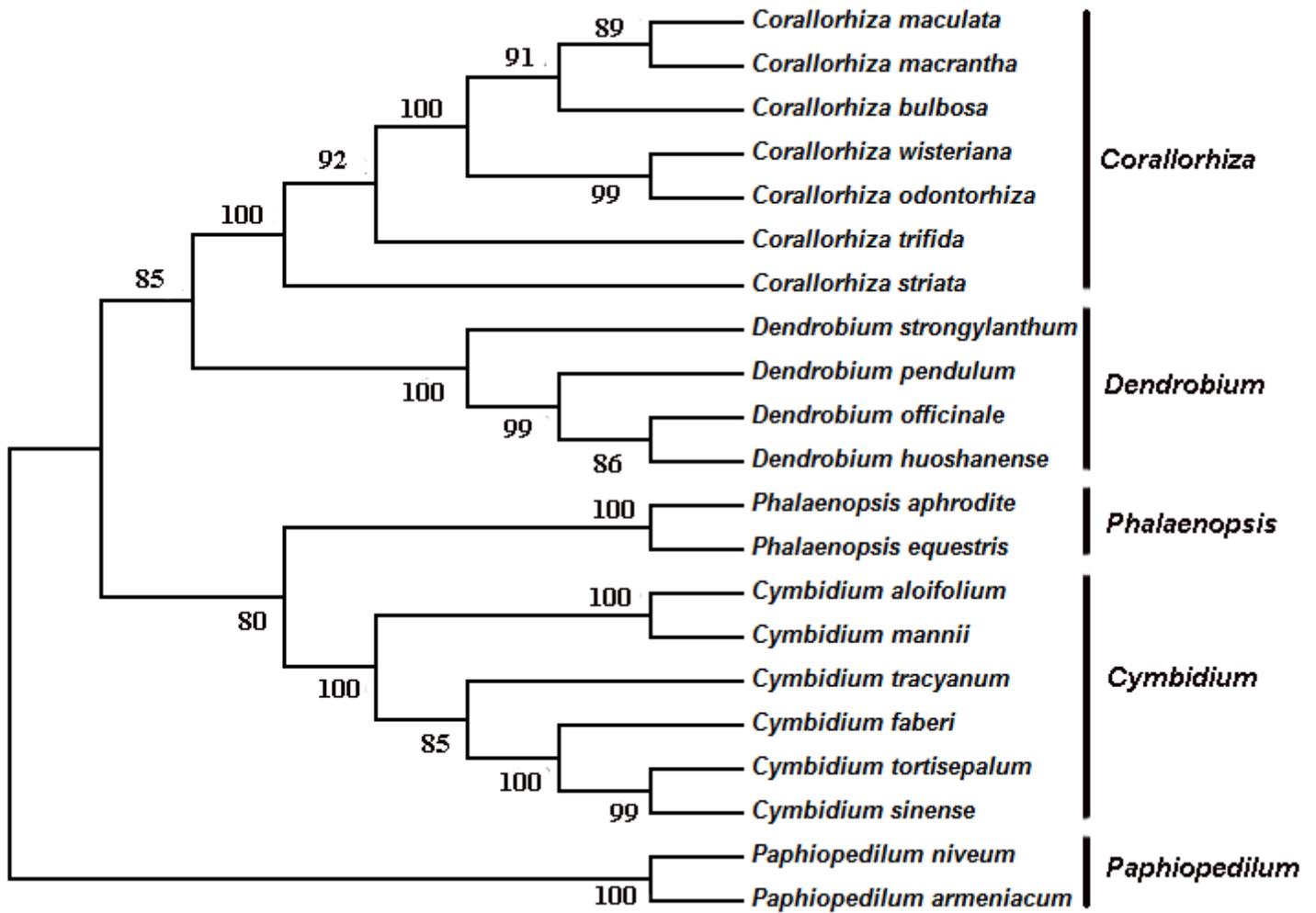


Figure 4

Phylogenetic tree of 21 Orchidaceae species, based on the complete chloroplast genome. The tree was generated by neighbor-joining (NJ) method using the MEGA6 program. Bootstrap values are indicated on the branches.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS1.docx](#)
- [TableS2.docx](#)