

Complete chloroplast genome of a rare and endangered plant species *Phalaenopsis zhejiangensis*: genomic features and phylogenetic relationship within Orchidaceae

Ming Jiang, Yan Zhu, Qian Wu and Huijuan Zhang

College of Life Sciences, Taizhou University, Taizhou, PR China

ABSTRACT

Phalaenopsis zhejiangensis is a rare and endangered plant species with extremely small populations. The complete chloroplast (cp) genome of *P. zhejiangensis* was assembled, its structural organization was described and comparative genomic analyses was carried out. The cp genome of *P. zhejiangensis* is 143,547 bp in length, with a GC content of 37.2%, which includes a pair of inverted repeats (IRs) of 24,464 bp separated by a small single-copy region of 10,764 bp and a large single-copy region of 83,856 bp. The cp genome contains 126 genes, consisting of 80 protein-coding genes, 38 transfer RNAs, and eight ribosomal RNAs. Six protein-coding genes, including $\psi ndhB$ (two copies), $\psi ndhD$, $\psi ndhG$, $\psi ndhK$, and $\psi ndhI$, are identified as pseudogenes. Another six *ndh* genes, *ndhA*, *ndhC*, *ndhE*, *ndhF*, *ndhH*, and *ndhJ*, are missing from the plastid genome. A total of 41 cp simple sequence repeats (SSRs) were identified, including 40 mono-nucleotides and one di-nucleotides. Phylogenetic analysis revealed *P. zhejiangensis* was nested inside the *Phalaenopsis* species and sister to *P. wilsonii*. The assembly and analysis of *P. zhejiangensis* cp genome will provide essential data for further study of taxonomy and systematics of Orchidaceae.

ARTICLE HISTORY

Received 3 June 2021
Accepted 18 August 2021

KEYWORDS

Phalaenopsis zhejiangensis;
rare and endangered
species; chloroplast
genome; phylogenetic
analysis

Introduction

Orchidaceae, a family of monocotyledonous plants, is one of the most successful and highly evolved monocots, comprising 763 genera with approximately 28,000 species (Hadi et al. 2015; Christenhusz and Byng 2016). The genus *Phalaenopsis* belongs to tribe Vandaeae, which is commonly known as moth orchids, including more than 70 species widely distributed across Southeast Asia to northern Australia, with the majority in Indonesia and the Philippines (Ko et al. 2017; Gogoi and Rinya 2020; Lee et al. 2020). There are 12 *Phalaenopsis* species distributed in China, among which four species are endemics (Wu et al. 2009). *Phalaenopsis zhejiangensis* is a small epiphytic plant with slightly flattened roots, very short stems, basal leaves, and thinly textured flowers (Wu et al. 2009). *P. zhejiangensis* was transferred from *Doritis zhejiangensis*, while the latter was renamed from *Nothodoritis zhejiangensis* (Yukawa and Kita 2005), which was first found on a trunk of *Podocarpus macrophyllus* in 1970 at Xitianmu Mountain in Linan County, Zhejiang Province (Tsi 1989; Schuiteman 2012). *P. zhejiangensis* is a rare and endangered plant species in China, which is narrowly distributed in provinces of Zhejiang Province, Gansu, Anhui, and Shanxi, growing on the bark of trees or rock surfaces, with extremely small populations. As a result of habitat loss, *P. zhejiangensis* is listed as threatened with extinction. Fortunately, Zeng et al (2011) have established an effective propagation protocol for

large-scale propagation of this endangered orchid species. In recent years, several chloroplast (cp) genomes of *Phalaenopsis* plants, including *P. aphrodite*, *Phalaenopsis* 'Tiny Star', *P. lowii*, *P. lobbii*, and *P. manni*, were sequenced and assembled (Chang et al. 2006; Kim et al. 2016; Wang et al. 2019; Chen et al. 2020; Zhang et al. 2020). However, the cp genome of *P. zhejiangensis* has not yet been assembled. In our study, the complete cp genome sequence of *P. zhejiangensis* was assembled to provide new insights into taxonomy and systematics of Orchidaceae.

Materials and methods

Plant material and DNA extraction

Fresh leaves were collected from a plant nursery (28°65.762'N, 121°46.976'E) in Jiaojiang, Zhejiang Province, China. A voucher specimen was deposited at the Molecular Biology Laboratory in Taizhou University (<http://www.tzc.edu.cn>), Dr. Huijuan Zhang, zhanghj82@126.com) under a collection number of CHS2020029. Fresh leaves were ground into a fine powder in liquid nitrogen with a mortar and pestle, and genomic DNA was extracted following the CTAB-based protocol as described by Doyle and Doyle (1987).

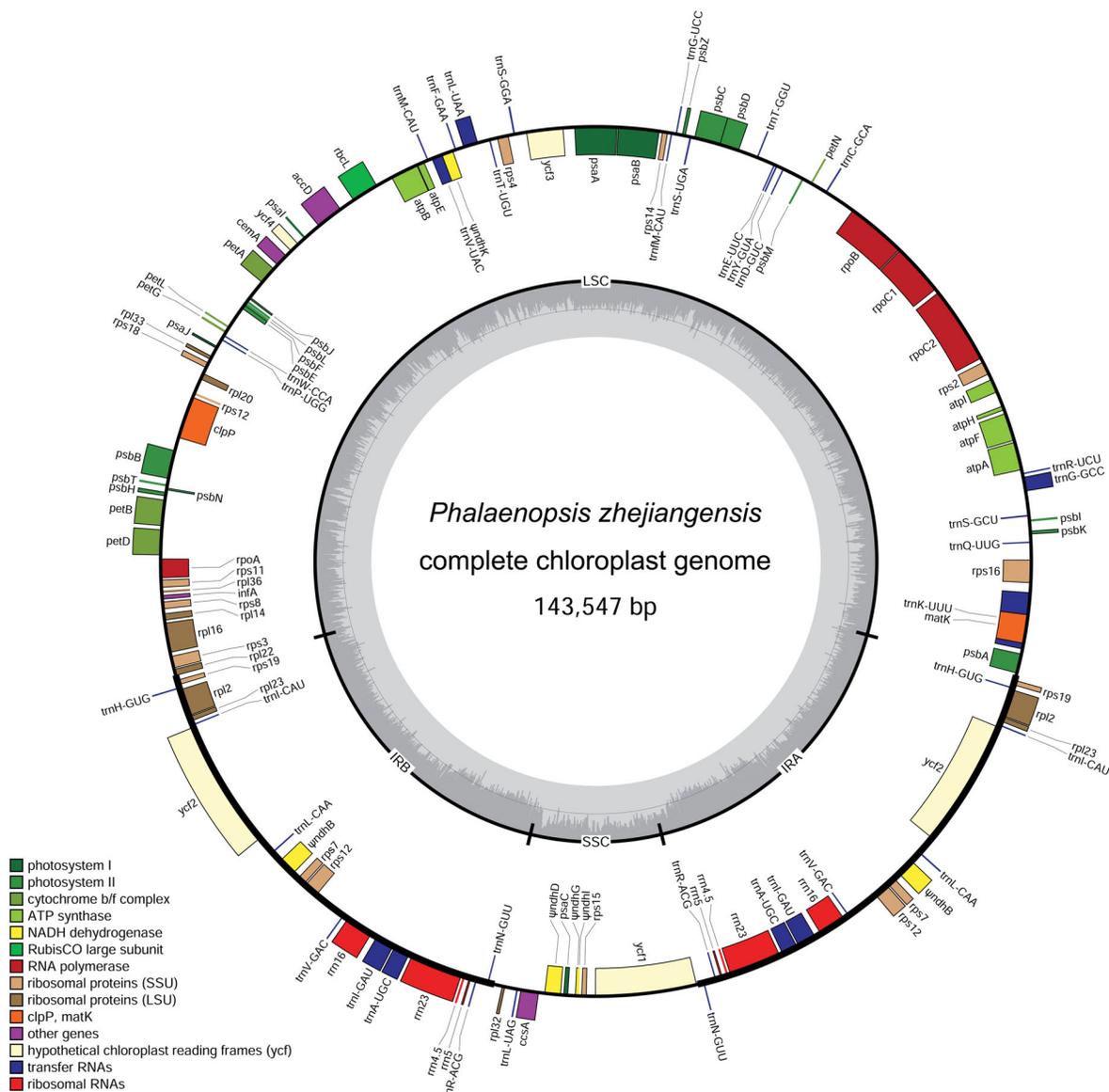


Figure 1. The chloroplast genome of *Phalaenopsis zhejiangensis*. From the center going outward, the four circles indicate scattered forward and reverse repeats, tandem repeats, microsatellite sequences identified, and gene structure of the plastome.

DNA sequencing and sequence assembly

Construction of a genomic DNA library was carried out according to the manufacturer's instructions, and high-throughput sequencing was then performed using the Illumina HiSeq X Ten system (Illumina, San Diego, CA). Approximately, 3.64 Gb raw data of 150 bp paired-end Illumina reads were generated, and 3.63 Gb high-quality clean reads were harvested by using the NGS QC Toolkit v2.3.3 (Patel and Jain 2012). Geneious Prime 2019 (Biomatters, Auckland, New Zealand) was employed to assemble the complete cp genome.

Chloroplast genome annotation

Annotation of the complete cp genome was performed with Geneious Prime 2019 by using *P. lobbii* as a reference

genome (NCBI accession number: MT830847) (Wang et al. 2019), and the borders of each gene were manually inspected and corrected where necessary. The circular cp genome map of *P. zhejiangensis* was drawn using the online program OrganellarGenomeDRAW (OGDRAW, <http://ogdraw.mpimp-golm.mpg.de>) with default settings (Lohse et al. 2013).

Simple sequence repeat analysis

Perl scripts of MicroSatellite Identification Tool (MISA) were used to identify and locate the potential simple sequence repeats (SSRs) loci in complete cp genome sequence of *P. zhejiangensis* (Thiel et al. 2003). To determine the presence of SSRs, 1–6 bp nucleotide motifs were considered, and the minimum repeat numbers were set as 10 repeat units for

Table 1. List of genes in the chloroplast genome of *Phalaenopsis zhejiangensis*.

Group of genes	Name of genes	Total number
Large subunit of ribosomal proteins	<i>rpl2</i> ($\times 2$)*, <i>rpl14</i> , <i>rpl16</i> *, <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> ($\times 2$), <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>	11
Small subunit of ribosomal proteins	<i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> ($\times 2$), <i>rps8</i> , <i>rps11</i> , <i>rps12</i> ($\times 2$)**, <i>rps14</i> , <i>rps15</i> , <i>rps16</i> *, <i>rps18</i> , <i>rps19</i> ($\times 2$)	15
DNA-dependent RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> *, <i>rpoC2</i>	4
Ribosomal RNA genes	<i>rrn4.5</i> ($\times 2$), <i>rrn5</i> ($\times 2$), <i>rrn16</i> ($\times 2$), <i>rrn23</i> ($\times 2$)	8
Transfer RNA genes	<i>trnA</i> -UGC ($\times 2$)*, <i>trnC</i> -GCA, <i>trnD</i> -GUC, <i>trnE</i> -UUC, <i>trnF</i> - GAA, <i>trnM</i> -CAU, <i>trnG</i> -GCC*, <i>trnG</i> -GCC, <i>trnH</i> -GUG ($\times 2$), <i>trnI</i> -CAU ($\times 2$), <i>trnI</i> -GAU ($\times 2$)*, <i>trnK</i> -UUU*, <i>trnL</i> -CAA ($\times 2$), <i>trnL</i> -UAA*, <i>trnL</i> -UAG, <i>trnM</i> -CAU, <i>trnN</i> -GUU ($\times 2$), <i>trnP</i> -UGG, <i>trnQ</i> -UUG, <i>trnR</i> -ACG ($\times 2$), <i>trnR</i> -UCU, <i>trnS</i> -GCU, <i>trnS</i> -GGA, <i>trnS</i> -UGA, <i>trnT</i> -GGU, <i>trnT</i> -UGU, <i>trnV</i> -GAC ($\times 2$), <i>trnV</i> -UAC*, <i>trnW</i> -CCA, <i>trnY</i> -GUA	38
Photosystem I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psal</i> , <i>psaj</i> , <i>ycf3</i> **	6
Photosystem II	<i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbH</i> , <i>psbl</i> , <i>psbj</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbN</i> , <i>psbT</i> , <i>psbZ</i>	15
Cytochrome b6/f complex	<i>petA</i> , <i>petB</i> *, <i>petD</i> *, <i>petG</i> , <i>petL</i> , <i>petN</i>	6
NADH dehydrogenase	ψ <i>ndhB</i> ($\times 2$), ψ <i>ndhD</i> , ψ <i>ndhG</i> , ψ <i>ndhI</i> , ψ <i>ndhK</i>	6
ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> *, <i>atpH</i> , <i>atpI</i>	6
Rubisco	<i>rbcl</i>	1
Translational initiation factor	<i>infA</i>	1
Maturase	<i>matK</i>	1
Protease subunit P	<i>clpP</i> **	1
Envelop membrane protein	<i>cemA</i>	1
Subunit acetyl-CoA carboxylate	<i>accD</i>	1
c-type cytochrome synthesis gene	<i>ccsA</i>	1
Conserved open reading frames	<i>ycf1</i> , <i>ycf2</i> ($\times 2$), <i>ycf4</i>	4
Total		126

(ψ) pseudogene; *one intron; **two introns.

Table 2. Details of nine *Phalaenopsis* chloroplast genomes.

Plant species	Accession number in NCBI	Genome length (bp)	GC contents (%)	LSC (bp)	SSC (bp)	IR (bp)
<i>Phalaenopsis aphrodite</i> subsp. <i>formosana</i>	AY916449	148,964	36.7	85,957	11,543	25,732
<i>Phalaenopsis equestris</i>	JF719062	148,959	36.7	85,967	11,300	25,846
<i>Phalaenopsis hybrid</i> 'Tiny Star'	NC_025593	148,918	36.7	85,885	11,523	25,755
<i>Phalaenopsis japonica</i>	NC_046808	146,942	36.8	84,882	10,568	25,746
<i>Phalaenopsis lobbii</i>	MT830847	144,607	36.6	83,482	11,687	24,719
<i>Phalaenopsis lowii</i>	NC_050652	146,834	36.9	84,616	10,473	25,633
<i>Phalaenopsis mannii</i>	NC_050940	148,596	36.7	85,300	11,640	25,828
<i>Phalaenopsis wilsonii</i>	MW218959	145,373	36.9	84,995	10,668	24,855
<i>Phalaenopsis zhejiangensis</i>	MZ326749	143,574	38.7	83,854	10,764	24,464

mono-nucleotides, six for di-nucleotides, and five for tri-, tetra-, penta-, and hexa-nucleotides.

Phylogenetic analysis

The complete cp genome sequences of 39 Orchidaceae species downloaded from NCBI were used for phylogenetic analysis, which included eight *Phalaenopsis* plants, six *Holcoglossum* species, and five *Cleisomeria* plants. The complete cp genome of *Coix lacryma-jobi* (FJ261955) was applied as an outgroup. A total of 41 genome sequences were aligned with a multiple sequence alignment program MAFFT v7.388 plugin in Geneious Prime 2019 under default settings (Kato and Standley 2013). The best-fit evolutionary model of DNA substitution for maximum-likelihood (ML) analysis was determined by using jModelTest 2.1.9 under the Akaike information criterion (AIC) (Darriba et al. 2012). A phylogenetic tree was generated by PhyML 3.1, with 1000 bootstrap replicates (Guindon et al. 2010).

Results

Genome sequencing and assembly

More than 12 million paired-end clean reads were obtained, and a circular cp genome was assembled with Geneious Prime. The complete cp plastome of *P. zhejiangensis* is 143,547 bp in length which includes a pair of inverted repeats (IRs) of 24,464 bp separated by a large single-copy (LSC) region of 83,855 bp and a small single-copy (SSC) region of 10,764 bp (Figure 1). The overall guanine–cytosine (GC) content of *P. zhejiangensis* cp genome is 36.8%, while the corresponding values of LSC, SSC, and IR sequences are 34.0%, 28%, and 43.5%, respectively.

Genome features of *Phalaenopsis zhejiangensis*

The complete cp plastome of *P. zhejiangensis* is comprised of 126 genes, including 74 protein-coding genes, 38 tRNA genes, eight rRNA genes, and six pseudogenes. All the six

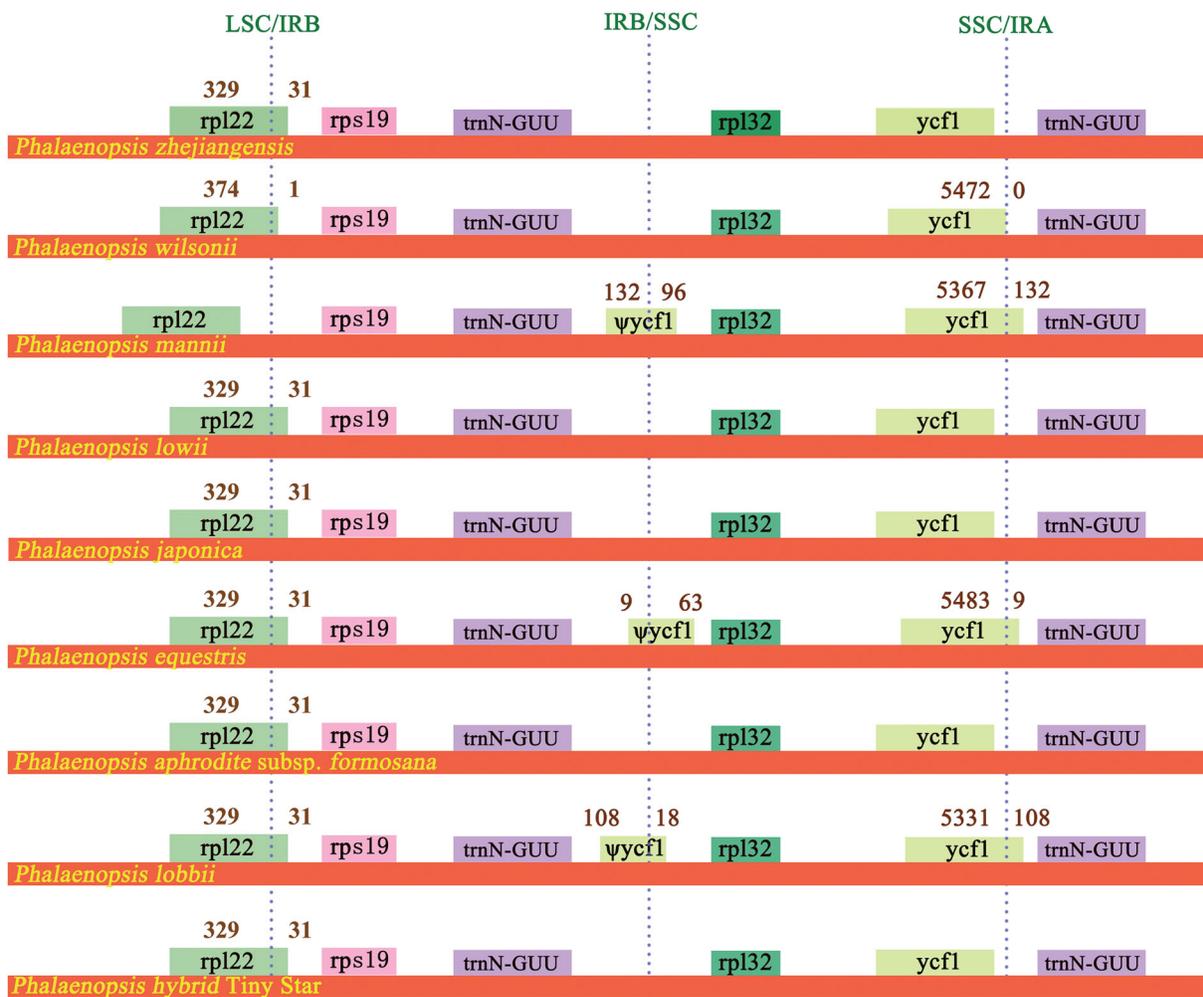


Figure 2. Structural variations at inverted-repeat and single-copy borders in nine *Phalaenopsis* chloroplast genomes. The figure features are not to scale regarding sequence length.

pseudogenes are *ndhs*, including $\psi ndhD$, $\psi ndhG$, $\psi ndhI$, $\psi ndhK$, and two copies of $\psi ndhB$. Additionally, other six *ndh* genes have been lost in the plastome, these are *ndhA*, *ndhC*, *ndhE*, *ndhF*, *ndhH*, and *ndhJ*. Six tRNA genes (*trnA-UGC*, *trnH-GUG*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, and *trnV-GAC*) and four rRNA genes (*rnr4.5*, *rnr5*, *rnr16*, and *rnr23*) are present in two copies. Among the protein-coding genes, eight genes including *rpl16*, *rps16*, *rpoC1*, *petB*, *petD*, *atpF*, and two copies of *rpl2* contain a single intron, while three genes, *rps12*, *ycf3*, and *clpP*, contain two introns. Moreover, six tRNA genes, *trnA-UGC*, *trnG-GCC*, *trnI-GAU*, *trnK-UUU*, *trnL-UAA*, and *trnV-UAC*, harbored a single intron have been identified (Table 1).

IR expansion and contraction

The IR region of *P. zhejiangensis* is 24,464 bp in length, while the other eight *Phalaenopsis* IRs included in our study are longer than *P. zhejiangensis*, they range from 24,719 bp to 25,846 bp (Table 2). Comparisons of LSC, IRB, SSC, and IRA junction boundaries among nine *Phalaenopsis* cp genomes were performed and presented in Figure 2. In *Phalaenopsis* plastome, LSC/IRB boundary junctions lie within *rpl22* gene except *P. manni*, which locates in the intergenic region

between *rpl22* and *rps19*. Eight *rpl22* genes extend 31 bp into IRB regions, while the *rpl22* in *P. wilsonii* extends only one base pair into its IRB region (Figure 2). The IRB/SSC junctions of *P. manni*, *P. equestris*, and *P. lobbii* locate within the $\psi ycf1$ pseudogenes; however, no $\psi ycf1$ is identified in other six *Phalaenopsis* cp genomes including *P. zhejiangensis*. The SSC/IRA junctions of *P. wilsonii*, *P. manni*, *P. equestris*, and *P. lobbii* lie within *ycf1* genes, which extend 0 bp, 132 bp, 9 bp, and 108 bp into IRA regions, respectively while the SSC/IRA junctions of *P. zhejiangensis*, *P. lowii*, *P. japonica*, *P. aphrodite subsp. formosana*, and the hybrid 'Tiny Star' locate between *ycf1* and *trnN-GUU*.

SSR analysis

Totally, *P. zhejiangensis* cp genome contains 41 SSRs (Table 3). Among the SSRs identified herein, A/T mononucleotide repeats account for the largest proportion of 97.56%. Only one di-nucleotide repeat was observed, and no tri-, tetra-, penta-, or hexa-nucleotides presented in the plastome. Most SSRs are distributed in LSC (70.73%), followed by SSC (24.39%), and the proportion in IRs is less than 5%. Among 41 SSRs, 25 loci lie in intergenic regions, four in introns, and 12 in coding regions or pseudogenes.

Table 3. Repeat sequences and their distribution within *Phalaenopsis zhejiangensis* chloroplast genome.

Number	Motif	Size	Start	End	Location
1	(T)10	10	4245	4254	IGS
2	(A)10	10	7693	7702	IGS
3	(T)10	10	9184	9193	IGS
4	(T)10	10	9599	9608	IGS
5	(T)17	17	14,113	14,129	IGS
6	(T)12	12	17,053	17,064	IGS
7	(A)16	16	27,913	27,928	IGS
8	(T)10	10	28,756	28,765	IGS
9	(A)11	11	32,765	32,775	IGS
10	(T)10	10	33,072	33,081	IGS
11	(A)10	10	33,357	33,366	IGS
12	(A)10	10	37,057	37,066	IGS
13	(A)11	11	47,194	47,204	IGS
14	(A)10	10	47,875	47,884	IGS
15	(A)11	11	56,140	56,150	IGS
16	(T)12	12	62,492	62,503	IGS
17	(A)10	10	64,380	64,389	IGS
18	(T)10	10	64,612	64,621	IGS
19	(T)10	10	71,393	71,402	IGS
20	(T)10	10	79,492	79,501	IGS
21	(T)10	10	83,505	83,514	IGS
22	(T)10	10	98,214	98,223	IGS
23	(A)11	11	108,703	108,713	IGS
24	(T)10	10	113,355	113,364	IGS
25	(A)10	10	129,180	129,189	IGS
26	(A)10	10	111,423	111,432	<i>ψndhD</i>
27	(A)10	10	111,579	111,588	<i>ψndhD</i>
28	(TA)8	16	111,722	111,737	<i>ψndhD</i>
29	(T)10	10	49,573	49,582	<i>ψndhK</i>
30	(A)12	12	74,999	75,010	<i>petB</i> intron
31	(T)10	10	82,386	82,395	<i>rpl16</i> intron
32	(A)11	11	23,345	23,355	<i>rpoC1</i> intron
33	(T)11	11	69,574	69,584	<i>clpP</i>
34	(A)10	10	69,867	69,876	<i>rps12</i>
35	(T)10	10	9805	9814	<i>trnG-GCC</i> intron
36	(T)12	12	114,415	114,426	<i>ycf1</i>
37	(T)11	11	115,271	115,281	<i>ycf1</i>
38	(T)13	13	116,348	116,360	<i>ycf1</i>
39	(T)10	10	117,578	117,587	<i>ycf1</i>
40	(T)10	10	117,695	117,704	<i>ycf1</i>
41	(A)13	13	45,453	45,465	<i>ycf3</i>

IGS: intergenic spacer region.

Phylogenetic analysis

Statistical selection of the best-fit model of nucleotide substitution was carried out by the jModelTest program, and GTR+G+I was turned out to be the best-fitted. To evaluate the phylogenetic relationship of *P. zhejiangensis* within Orchidaceae, a phylogenetic tree using the complete plastome sequences of 41 plant species were generated (Figure 3). The resulting ML tree had very high bootstrap support for the majority of clades, and the phylogenetic tree was shown to be consistent with the traditional morphology-based taxonomy of Orchidaceae. Nine plants from the genus *Phalaenopsis* formed a well-supported monophyletic clade with 100% bootstrap value, and *P. zhejiangensis* was sister to *P. wilsonii*, with a support value of 97%.

Discussion

The classification of Orchidaceae is a very complex problem for its bewildering diversity and morphological parallelism, and also for botanical neglect (Dressler 1993). Orchidaceae were divided into five subfamilies, including Apostasioideae, Vanilloideae, Cypripedioideae, Orchidoideae and

Epidendroideae, and Apostasioideae are monophyletic and sister to other subfamilies (Dixon et al. 2003; Tsai et al. 2013). The subfamily Orchidoideae is comprised of a large group of orchids, including four tribes (Diurideae, Cranichideae, Codonorchideae, and Orchideae) with around 3600 plant species that largely share terrestrial habits (Bateman et al. 2003; Serna-Sánchez et al. 2021). The genus *Phalaenopsis* belongs to tribe Orchideae, which consists of about 62 genera (Inda et al. 2012). The morphology of *P. zhejiangensis* is similar to both *Doritis* and *P. lowii*, with the characteristic feature of pollinia similar to that of *Doritis*, and its shapes of petals and columns similar to *P. lowii* (Tsi 1999; Christenson 2001). *P. zhejiangensis* was initially named *N. zhejiangensis* based on morphology, and then renamed as *D. zhejiangensis* (Yukawa and Kita 2005; Schuiteman 2012). Molecular analyses using ITS, *trnL* intron as well as *trnL-F* spacer prove that *N. zhejiangensis* is species nested in genus *Phalaenopsis* (Deng et al. 2015). Our molecular phylogenetic analyses confirm its systematic position, which indicates that *P. zhejiangensis* is sister to *P. wilsonii*, with a high support value.

The sizes of previously published *Phalaenopsis* cp genomes ranged from 144,607 bp (*P. lobbii*) to 148,964 bp (*P. phalaenopsis aphrodite* subsp. *formosana*) (Chang et al. 2006; Kim et al. 2016; Wang et al. 2019; Chen et al. 2020; Zhang et al. 2020). In our present study, the complete plastid genome of *P. zhejiangensis* is only 143,574 bp in length, which turns out to be the smallest among all known cp genome sequences in genus *Phalaenopsis*. The cp genomes of most flowering plants range from 120 kb to 160 kb, and the difference in genome size is mainly caused by contraction and expansion of the IR region (Goulding et al. 1996; Ingvarsson et al. 2003; Yao et al. 2020). The IR sizes of published *Phalaenopsis* plastid genomes are 24,719 (*P. lobbii*) (Wang et al. 2019) to 25,846 bp (*P. equestris*) (Jheng et al. 2012). In our present study, the length of *P. zhejiangensis* IR sequence is 24,464 bp, which is smaller than that of *P. lobbii*.

Plastid gene loss and pseudogenization are common phenomena in parasitic, semi-parasitic, and saprophytic plants (Molina et al. 2014; Li and Zheng 2018; Nie et al. 2019). In semi-parasitic plants of *Macrosolen cochinchinensis*, *M. tricolor*, and *M. bibracteolatus*, the *infA* and all the *ndh* genes were lost among the three species, and two genes, *ycf1* and *rpl2*, were found to be pseudogenes (Nie et al. 2019). *Gastrodia elata* is a saprophytic plant with extremely small cp genome in which many genes related to photosynthesis are missing, and it contains only 28 genes, including 20 protein-coding genes, three rRNAs, and five tRNAs (Park et al. 2020). However, cp gene loss and pseudogenization are rarer in photosynthetic species, because the plastid gene cannot simply be discarded (Magee et al. 2010; Wicke et al. 2011; Daniell et al. 2016). Interestingly, pseudogenization, truncation, and deletion of *ndh* genes in cp genomes are common in some cp genomes of photosynthetic orchid plants (Kim et al. 2015). Orchids like *Apostasia odorata*, *Sobralia callosa*, *Paphiopedilum armeniacum*, and *Phragmipedium longifolium* retain the complete set of *ndh* genes, whereas most of the *ndh* genes in *P. equestris*, *Dendrobium officinale*, and *D. catenatum* were found to be lost in their plastids (Kim et al. 2015; Lin et al. 2017). In this study, six *ndh* genes are

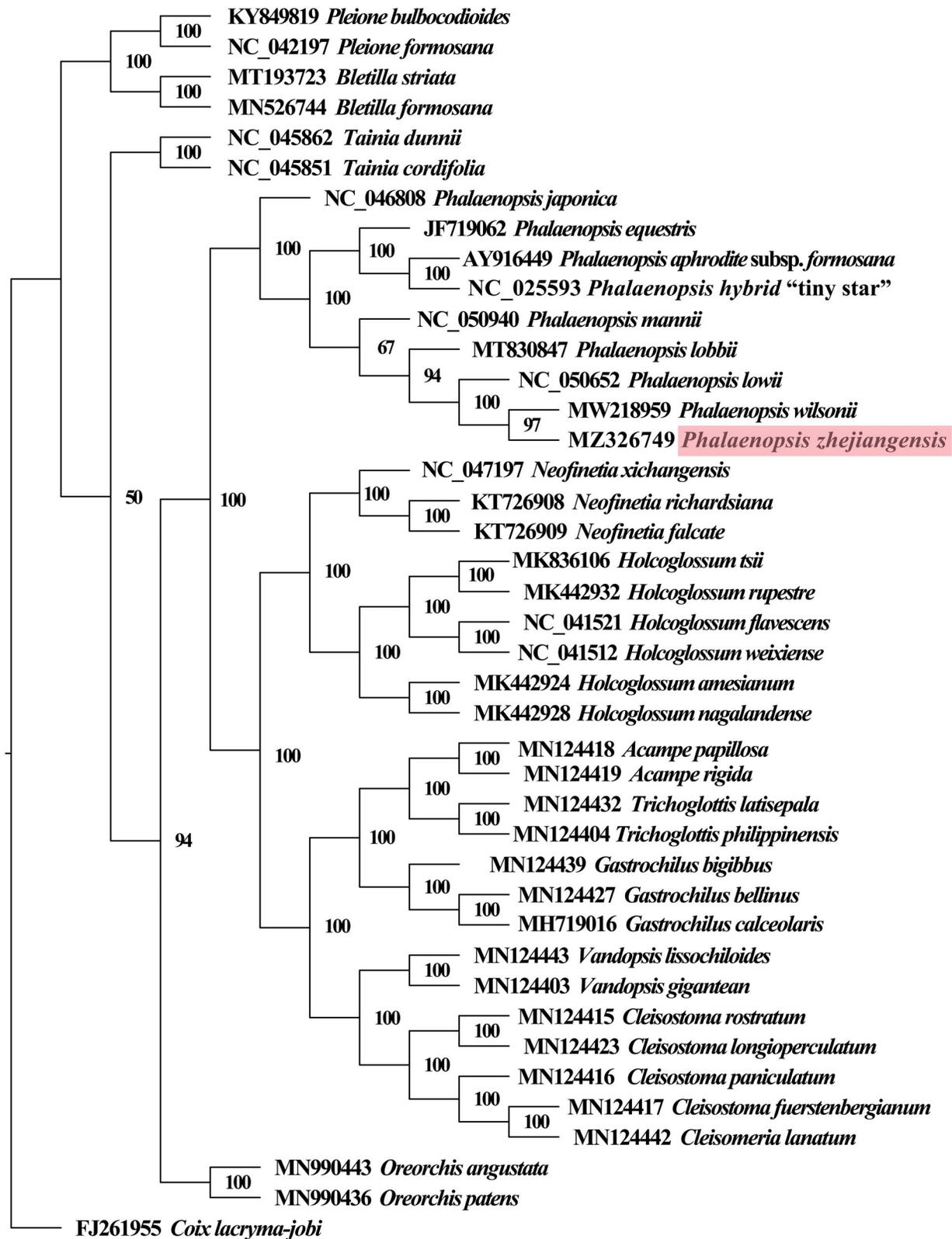


Figure 3. The maximum-likelihood tree inferred from 41 Orchidaceae complete chloroplast genomes under GTR + G+I model with maximum-likelihood value (Llk)=-604333.07383, Akaike information criterion (AIC)=1208844.14766, and Bayesian information criterion (BIC)=1209758.57750.

pseudogenized, and another six *ndh* genes are missing from the plastid genome of *Phalaenopsis zhejiangensis*.

Authors contributions

Conceived and designed the experiments: MJ and HJZ. Performed the experiments: MJ, WQ, and ZY. Analyzed the data: HZW and LQP. Contributed to reagents/materials/analysis tools: MJ and WQ. Wrote the paper: MJ and HJZ.

Disclosure statement

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

Funding

This work was supported by 211 Talent Training Fund of Taizhou under Grant number [2012].

Data availability statement

The data that support the findings of this study are openly available in GenBank of NCBI at <https://www.ncbi.nlm.nih.gov/nucleotide/MZ326749>. The associated BioProject, SRA, and Bio-Sample numbers are PRJNA734444, SRR14710950, and SAMN19491762, respectively.

References

- Bateman RM, Hollingsworth PM, Preston J, Yi-Bo LUO, Pridgeon AM, Chase MW. 2003. Molecular phylogenetics and evolution of Orchidaceae and selected Habenariaceae (Orchidaceae). *Bot J Linn Soc.* 142(1):1–40.
- Chang CC, Lin HC, Lin IP, Chow TY, Chen HH, Chen WH, Cheng CH, Lin CY, Liu SM, Chang CC, et al. 2006. The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. *Mol Biol Evol.* 23(2):279–291.
- Chen B, Zhang Y, Cao Y, Zheng Y, Wei Z, Zhao K, Peng D, Zhou Y. 2020. Chloroplast characterizations of a *Phalaenopsis* native to China, *Phalaenopsis manni* (Orchidaceae). *Mitochondrial DNA Part B.* 5(3): 3725–3726.
- Christenhusz MJM, Byng JW. 2016. The number of known plants species in the world and its annual increase. *Phytotaxa.* 261(3):201–217.
- Christenson EA. 2001. *Phalaenopsis*—a monograph. Portland: Timber Press.
- Daniell H, Lin CS, Yu M, Chang WJ. 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 17(1):134.
- Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods.* 9(8):772.
- Deng H, Zhang G-Q, Liu Z-J, Wang Y. 2015. A new species and a new combination of *Phalaenopsis* (Orchidaceae: Epidendroideae: Aeridinae): evidence from morphological and DNA analysis. *Phytotaxa.* 238(3):243–254. doi:10.11646/phytotaxa.238.3.3.
- Dixon KW, Kell SP, Barrett RL, Cribb PJ. 2003. *Orchid conservation*. Kota Kinabalu: Sabah Natural History Publications.
- Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull.* 19:11–15.
- Dressler RL. 1993. *Phylogeny and classification of the orchid family*. Cambridge: Cambridge University Press.
- Gogoi K, Rinya K. 2020. *Phalaenopsis arunachalensis* sp. nov. (Orchidaceae: Epidendroideae: Aeridinae) A new epiphytic orchid from Arunachal Pradesh, Northeast India. *Lankesteriana.* 20(3):275–280.
- Goulding SE, Olmstead RG, Morden CW, Wolfe KH. 1996. Ebb and flow of the chloroplast inverted repeat. *Mol Gen Genet.* 252(1–2):195–206.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 59(3): 307–321.
- Hadi H, Razali SN, Awadh AI. 2015. A comprehensive review of the cosmecological benefits of *Vanda* species (Orchidaceae). *Nat Prod Commun.* 10(8):1483–1488.
- Inda LA, Pimentel M, Chase MW. 2012. Phylogenetics of tribe Orchideae (Orchidaceae: Orchidoideae) based on combined DNA matrices: inferences regarding timing of diversification and evolution of pollination syndromes. *Ann Bot.* 110(1):71–90.
- Ingvarsson PK, Sarah R, Taylor DR. 2003. Molecular evolution of insertions and deletion in the chloroplast genome of silene. *Mol Biol Evol.* 20(11):1737–1740.
- Jheng CF, Chen TC, Lin JY, Chen TC, Wu WL, Chang CC. 2012. The comparative chloroplast genomic analysis of photosynthetic orchids and developing DNA markers to distinguish *Phalaenopsis* orchids. *Plant Sci.* 190:62–73.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30(4):772–780.
- Kim HT, Kim JS, Moore MJ, Neubig KM, Williams NH, Whitten WM, Kim JH. 2015. Seven new complete plastome sequences reveal rampant independent loss of the *ndh* gene family across orchids and associated instability of the inverted repeat/small single-copy region boundaries. *PLOS One.* 10(11):e0142215.
- Kim GB, Kwon Y, Yu HJ, Lim KB, Seo JH, Mun JH. 2016. The complete chloroplast genome of *Phalaenopsis* “Tiny Star”. *Mitochondrial DNA Part A.* 27(2):1300–1302.
- Ko YZ, Shih HC, Tsai CC, Ho HH, Liao PC, Chiang YC. 2017. Screening transferable microsatellite markers across genus *Phalaenopsis* (Orchidaceae). *Bot Stud.* 58(1):48.
- Lee YI, Tseng YF, Lee YC, Chung MC. 2020. Chromosome constitution and nuclear DNA content of *Phalaenopsis* hybrids. *Sci Hortic.* 262: 109089.
- Li B, Zheng Y. 2018. Dynamic evolution and phylogenomic analysis of the chloroplast genome in Schisandraceae. *Sci Rep.* 8(1):9285.
- Lin CS, Chen JJW, Chiu CC, Hsiao HCW, Yang CJ, Jin XH, Leebens-Mack J, de Pamphilis CW, Huang YT, Yang LH, et al. 2017. Concomitant loss of NDH complex-related genes within chloroplast and nuclear genomes in some orchids. *Plant J.* 90(5):994–1006.
- Lohse M, Drechsel O, Kahlau S, Bock R. 2013. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res.* 41(Web Server issue):W575–W581.
- Magee AM, Aspinall S, Rice DW, Cusack BP, Sémon M, Perry AS, Stefanović S, Milbourne D, Barth S, Palmer JD, et al. 2010. Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genome Res.* 20(12):1700–1710.
- Molina J, Hazzouri KM, Nickrent D, Geisler M, Meyer RS, Pentony MM, Flowers JM, Pelsner P, Barcelona J, Inovejas SA, et al. 2014. Possible loss of the chloroplast genome in the parasitic flowering plant *Rafflesia lagascae* (Rafflesiaceae). *Mol Biol Evol.* 31(4):793–803.
- Nie L, Cui Y, Wu L, Zhou J, Xu Z, Li Y, Li X, Wang Y, Yao H. 2019. Gene losses and variations in chloroplast genome of parasitic plant *Macrosolen* and phylogenetic relationships within Santalales. *Int J Mol Sci.* 20(22):5812.
- Park J, Suh Y, Kim S. 2020. A complete chloroplast genome sequence of *Gastrodia elata* (Orchidaceae) represents high sequence variation in the species. *Mitochondrial DNA Part B.* 5(1):517–519.
- Patel RK, Jain M. 2012. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLOS One.* 7(2):e30619.
- Schuiteman A. 2012. *Nothodoritis zhejiangensis* transferred to *Phalaenopsis*. *Renziana.* 2:48–50.

- Serna-Sánchez MA, Pérez-Escobar OA, Bogarín D, Torres-Jimenez MF, Alvarez-Yela AC, Arcila-Galvis JE, Hall CF, de Barros F, Pinheiro F, Dodsworth S, et al. 2021. Plastid phylogenomics resolves ambiguous relationships within the orchid family and provides a solid timeframe for biogeography and macroevolution. *Sci Rep.* 11(1):6858.
- Thiel T, Michalek W, Varshney RK, Graner A. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet.* 106(3): 411–422.
- Tsai WC, Fu CH, Hsiao YY, Huang YM, Chen LJ, Wang M, Liu ZJ, Chen HH. 2013. OrchidBase 2.0: comprehensive collection of Orchidaceae floral transcriptomes. *Plant Cell Physiol.* 54(2):e7.
- Tsi ZH. 1989. New taxa of Orchidaceae from China. *Bull Bot Res.* 9(2): 21–32.
- Tsi ZH, Chen SC, Luo YB, Zhu GH. 1999. *Flora Reipublicae Popularis Sinicae*. Vol. 19. Beijing: Science Press.
- Wang JY, Liu ZJ, Zhang GQ, Peng CC. 2019. The complete chloroplast genome sequence of *Phalaenopsis lowii* (Orchidaceae). *Mitochondrial DNA Part B.* 4(2):3569–3570.
- Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol.* 76(3–5): 273–297.
- Wu ZY, Raven PH, Hong DY. 2009. *Flora of China*. Vol. 25. Beijing and St. Louis: Science Press and Missouri Botanical Garden Press.
- Yao J, Zhao F, Xu Y, Zhao K, Quan H, Su Y, Hao P, Liu J, Yu B, Yao M, et al. 2020. Complete chloroplast genome sequencing and phylogenetic analysis of two *Dracocephalum* plants. *Biomed Res Int.* 2020:4374801.
- Yukawa T, Kita K. 2005. *Doritis zhejiangensis* (Z.H.Tsi). *Acta Phytotax Geobot.* 56(2):157.
- Zeng S, Chen Z, Wu K, Zhang J, Bai C, Teixeira da Silva JA, Duan J. 2011. Asymbiotic seed germination, induction of calli and protocorm-like bodies, and in vitro seedling development of the rare and endangered *Nothodoritis zhejiangensis* Chinese orchid. *HortScience.* 46(3): 460–465.
- Zhang Y, Chen B, Zheng Y, Cao Y, Wei Z, Zhao K, Zhou Y. 2020. Characterization of the complete chloroplast genome of *Phalaenopsis lobbii* (Orchidaceae), an important horticultural plant in China. *Mitochondrial DNA Part B.* 5:3450–3451.