

RESEARCH

Open Access



Complete chloroplast genome structural characterization of two *Phalaenopsis* (Orchidaceae) species and comparative analysis with their alliance

Lei Tao^{1,2}, Hanning Duan¹, Kaifeng Tao¹, Yan Luo³, Qingqing Li^{2,4} and Lu Li^{1*}

Abstract

Background The taxonomy and infrageneric delimitation of *Phalaenopsis* Blume has been significantly disputed due to some overlapping morphological features between species related, which needed further evidence for clarification. The structural characterization of complete chloroplast genomes of *P. storbatiana* and *P. wilsonii* were analyzed and compared with those of related taxa to provide a better understanding of their genomic information on taxonomy and phylogeny.

Results It was shown that chloroplast genomes of *Phalaenopsis storbatiana* and *P. wilsonii* had a typical quadripartite structure with conserved genome arrangements and moderate divergence. The chloroplast genomes of *P. storbatiana* and *P. wilsonii* were 145,885 bp and 145,445 bp in length, respectively, and shared a similar GC content of 36.8%. Gene annotations of two species revealed 109 single-copy genes consistently. In addition, 20 genes duplicated in the inverted regions, 16 genes each possessed one or more introns, and five *ndh* (NA (D)H dehydrogenase) genes were observed in both. Comparative analysis of the total cp genomes of *P. storbatiana* and *P. wilsonii* with those of other six related *Phalaenopsis* species confirmed the stable sequence identity for coding and non-coding regions and higher sequence variation in SC regions than IR regions. Most of their protein-coding genes had a high degree of codon preference. Moreover, 45 genes were discovered with significantly positive selection. However, different amplifications in IR regions were observed in these eight species. Phylogenetic analysis based on CDS from 60 species representing main clades in Orchidaceae indicated that *Phalaenopsis* species including *P. storbatiana* and *P. wilsonii* formed a monophyletic clade with high bootstrap nested in tribe Vandaeae of Epidendroideae, which was consistent with those from previous studies.

Conclusions The results could provide insight into understanding the plastome evolution and phylogenetic relationships of *Phalaenopsis*.

Keywords *Phalaenopsis storbatiana*, *Phalaenopsis wilsonii*, Chloroplast genome, Orchidaceae, Structural characterization

*Correspondence:

Lu Li
lilu@swfu.edu.cn

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

The broader genus *Phalaenopsis* Blume (Aeridinae, Vandaeae, Epidendroideae, Orchidaceae) is consisted of about 40–45 species, which are collectively distributed from India to China, Korea, Japan, Thailand, Indochina, Malaysia, and Indonesia to the Philippines, Australia, and New Guinea [1]. There are 22 species recorded in China, including five endemic ones, which occurred in Southern China [2]. *Phalaenopsis wilsonii* Rolfe is endemic to China, while *P. stobartiana* Rchb.f. is distributed in South China and Burma [3]. Most *Phalaenopsis* species possess highly ornamental values and are used for breeding systems. However, the wild populations of *Phalaenopsis* have been decreasing due to their habitat fragmentation and over-exploration [1]. And then more attention should be paid to their biological conservation [2].

The phylogeny and infrageneric delimitation of *Phalaenopsis* has been disputed based on the morphological and molecular data available [4, 5]. *Phalaenopsis* was ever classified into two categories based on the presence of lip appendages [6] but into four groups according to the size of sepals and petals and the structure of the column and lip tip [7]. And then, it was divided into eight sections [8] or five subgenera with eight sections [9]. *Phalaenopsis wilsonii* and *P. stobartiana* were distinguished from other related species by flowers un-spurred [3]. Moreover, the intergeneric relationships were significantly confused with the alliance in the molecular phylogeny of Vandaeae investigated [10–13]. It was supported that a broad definition of *Phalaenopsis* was preferable, while a new infrageneric taxonomy encompassing four subgenera was proposed: *Parishianae* (H.R.Sweet) Christenson, *Phalaenopsis* (i.e., *Doritis* Lindl., *Kingidium* P.F.Hunt, and *Nothodoritis* Z.H.Tsi), *Hygrochilus* Pfitzer and *Ornithochilus* (Lindl.) Wall. ex Benth. [1, 14–16]. However, the phylogeny and taxonomy of *Phalaenopsis* have not been clear up to now. For example, it was suggested that the broad definition of *Phalaenopsis* should not include *Sedirea* Garay & H.R.Sweet [17, 18].

The chloroplast (cp) genome has been crucial in plant phylogenetics [19–21]. Compared with nuclear and mitochondrial gene sequences used in origin and phylogenetic relationships, the cp genomes are smaller, less prone to recombination, and have low rates of nucleotide substitutions [5, 22]. It has been frequently used in Orchidaceae for phylogenetics, which strongly supported the view that this family was comprised of five subfamilies [20, 22, 23]. It was implied that *Phalaenopsis* was related to *Neofinetia* Hu, *Pelatantheria* Ridl., and *Gastrochilus* D.Don and placed in Vandaeae based on 79 CDS and four nrDNA from the cp genomic data [24]. Meanwhile, it seemed that some infrageneric relationships of *Phalaenopsis* had been stable. Subgenus *Aphyllae* (H.R.Sweet) Christenson consistently consisted of *P. wilsonii*, *P. stobartiana*, *P.*

honghenensis F.Y.Liu, *P. minus* (Seidenf.), E.A. Christ., and *P. deliciosa* Rchb.f. based on a combined plastid sequence [9, 11, 12, 25]. However, *P. stobartiana* and *P. wilsonii* were also placed into subgenus *Parishianae* based on the broad definition of *Phalaenopsis* [2, 26].

Characterization of complete chloroplast genomes of *P. wilsonii* [27, 28] and *P. stobartiana* [25] were reported, but no additional genomic information was available. In this study, the structural and genomic information in detail were analyzed and compared with those of the related *Phalaenopsis* species downloaded from Genbank. The objectives of this study were: (1) to characterize and compare two complete chloroplast genome structures of *P. wilsonii* and *P. stobartiana* in detail, and (2) to provide further genomic information for a better understanding of phylogeny in *Phalaenopsis*.

Results

General data on the chloroplast genome

The structures of chloroplast genomes of two *Phalaenopsis* species were highly similar. The total sizes of two cp genomes were 145,885 bp (*P. stobartiana*) and 145,445 bp (*P. wilsonii*) (Fig. 1; Table 1). Same as those of most angiosperms, their chloroplast genome structures displayed a typical quadripartite structure with a large single-copy (LSC) region (85,349 bp, 85,076 bp), a small single-copy (SSC) region (10,596 bp, 10,473 bp), and two inverted repeats (IR) regions (24,970 bp, 24,948 bp). In both cp genomes, the amounts of GC contents in LSC, SSC, and IR regions were 36.8%, 28.2%, and 43.3%, respectively. Comparative analysis of both cp genomes consistently showed that the GC content in IR regions was higher than in LSC and SSC regions. The GC content of the three positions of the two cp genomes was very similar. However, the third letter GC (29.72%) content was lower than the first (45.47%, 45.40%) and second (37.85%, 37.86%) letter GC content (Table 2). Both cp genomes contained 127 genes, including 82 CDS, eight rRNAs, and 37 tRNAs. However, the length of the LSC was different. It was longer in *Phalaenopsis stobartiana* than in *P. wilsonii* (Table 1). Among these, there were 109 unique genes in each cp genome. The LSC region contained 63 CDS genes and 20 tRNA genes, whereas the SSC region comprised seven CDS genes and only one tRNA gene. Eight CDS (*ndhB*, *rpl2*, *rpl22*, *rpl23*, *rps7*, *rps12*, *rps19*, and *ycf2*), eight tRNA (*trnA-UGC*, *trnH-GUG*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, and *trnV-GAC*), and four rRNA (*rrn4.5*, *rrn5*, *rrn16*, and *rrn23*) genes were repeated in the IR regions (Table S1). There were 16 genes with introns, 13 genes (*trnV-UAC*, *trnL-UAA*, *trnI-GAC*, *trnG-UCC*, *trnA-UGC*, *rps16*, *rpoC1*, *rpl2*, *rpl16*, *petD*, *petB*, *ndhB*, and *atpF*) of which had only one intron, while the others (*clpP*, *ycf3*, *rps12*) had two introns (Table S1). Four of the 16 intron-containing genes

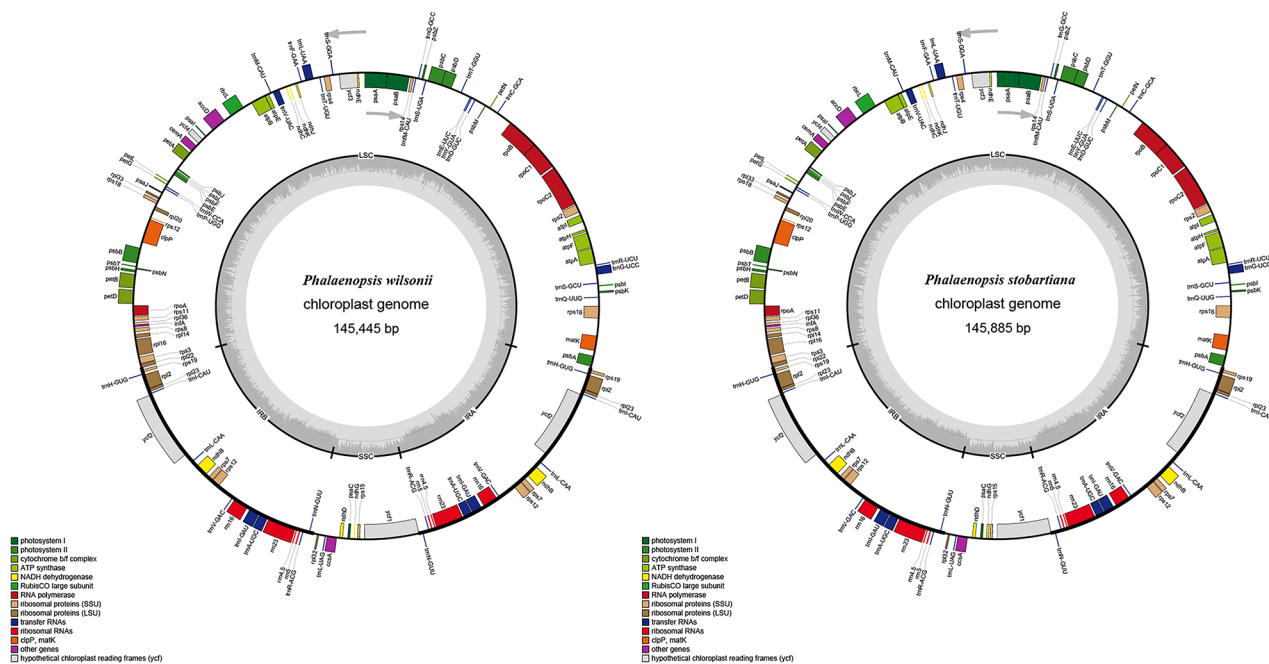


Fig. 1 The chloroplast genome maps of *Phalaenopsis stobartiana* and *P. wilsonii*. Internal genes were clockwise transcribed, while external genes were counterclockwise transcribed. The inside circle bright and dark gray coloring indicated the genome guanine-cytosine (GC) content

Table 1 The general genome characteristics of two *Phalaenopsis* species

Characteristics and Parameters	<i>P.stobartiana</i>	<i>P.wilsonii</i>
Total cp genome size (bp)	145,885	145,445
LSC length (bp)	85,349	85,076
SSC length (bp)	10,596	10,473
IR length (bp)	24,970	24,948
Total GC content (%)	36.8	36.8
GC content for LSC (%)	34.1	34.1
GC content for SSC (%)	28.2	28.2
GC content for IR (%)	43.3	43.3
Total number of genes	127	127
CDS genes	82	82
rRNAs genes	8	8
tRNAs genes	37	37

Table 2 The GC content of the three positions of two *Phalaenopsis* species

Species	1st letter GC	2nd letter GC	3rd letter GC
<i>P. stobartianna</i>	45.47%	37.85%	29.72%
<i>P. wilsonii</i>	45.50%	37.86%	29.72%

were in the IR regions, while 12 of the 16 genes spread across the LSC region. All the exons of tRNA genes in both segments were 20–50 bp in length. The *rpl16*, *petD*, and *petB* genes had one very short exon compared with other genes, while the *rpoC1* had one longer exon. In addition to the above, *rps12* was a unique trans-splicing

gene in which the first exon dispersed in the LSC region, but the second and third exons were in IR regions. Five *ndh* (NA (D)H dehydrogenase) genes (*ndhB/C/D/E/G/J/K*) were identified (Fig. 1, Table S2).

Codon usage analysis

Based on 82 coding sequences (CDS), codon usage frequency and relative synonymous codon usage (RSCU) were computed in both cp genomes. These CDS were composed of 23,281 (*Phalaenopsis wilsonii*) and 23,324 codons (*P. stobartiana*), respectively, and encoded 20 amino acids in the chloroplast genomes in them (Fig. 2, Table S3). The RSCU value of two chloroplast genomes was similar, with six codons for arginine (Arg) and leucine (Leu) and only one codon for methionine (Met) and tryptophan (Trp). Among them, leucine (Leu: 10.11%, 10.13%) was the amino acid that was utilized the most frequently, whereas cysteine (Cys: 1.19%, 1.19%) was the least ubiquitous amino acid in the two cp genomes. Except for methionine (Met) and tryptophan (Trp), practically all amino acids were encoded by 2–6 synonymous codons, according to the RSCU analysis. Relative synonymous codon usage was 1 for methionine (Met) and tryptophan (Trp). Thirty codons had RSCU>1, and 31 had RSCU<1. Almost CDS in *Phalaenopsis* species had the standard ATG start codon, but *rpl2* started with ATA/TAT. Among three stop codons, the TAA was the most common.

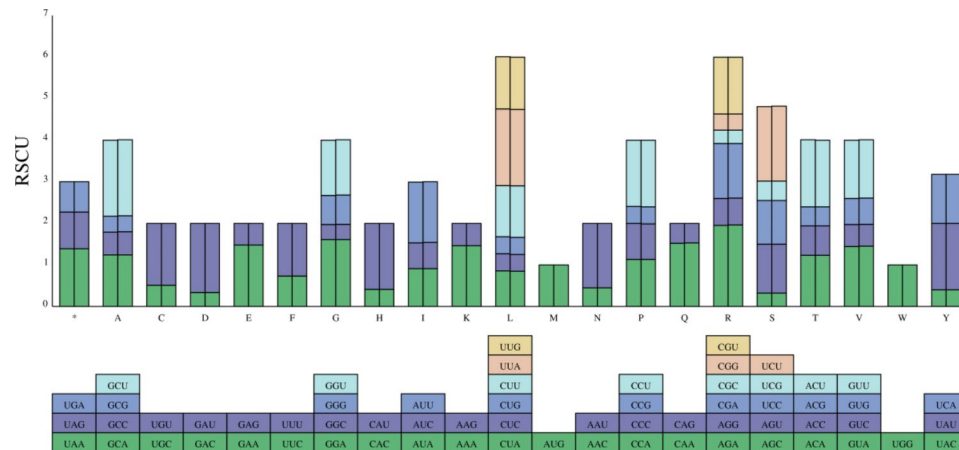


Fig. 2 Codon content of 20 amino acids and stop codons in all protein-coding genes of the cp genomes of two *Phalaenopsis* species. The histogram above each amino acid shows codon usage within *Phalaenopsis*. Colors in the column graph reflected codons in the same colors shown below the figure. RSCU: relative synonymous codon usage; A: alanine; R: arginine; N: asparagine; D: asparagine; C: cysteine; Q: glutamine; E: glutamic; G: glycine; H: histidine; L: leucine; I: isoleucine; K: lysine; M: methionine; F: phenylalanine; P: proline; Ser: serine; T: threonine; W: tryptophan; Y: tyrosine; V: valine. Left: *Phalaenopsis wilsonii*; Right: *Phalaenopsis stobartiana*

Repeat sequences analysis

In this study, 75 (*Phalaenopsis stobartiana*) and 73 (*P. wilsonii*) SSRs were identified in two cp genomes, with 54–55 mononucleotides (mono-), six dinucleotides (di-), four trinucleotides (tri-), seven tetranucleotides (tetra-), and two pentanucleotides (penta-) (Fig. 3: A, B). Only the pentanucleotide was present in the IR regions, with most SSRs dispersed in the LSC, SSC, and IR regions. According to statistical analysis, most SSRs were in the LSC (51–54) region, while just 2 SSRs were dispersed in the IR regions (Fig. 3: C, D). Repeat units were composed mainly of A or T; besides, the mononucleotides of the two cp genomes were A/T type rather than G/C type. Furthermore, the AAAT/ATTT type tetranucleotide was only found in *P. willsonii*.

Four different types of tandems were identified based on the complete genome sequence: complement (C), forward (F), palindromic (P), and reverse (R). Complete tandem content was the lowest, while the palindromic tandem content was the highest in the two cp genomes. However, the tandem sequence of the two cp genomes contained two different tandem forms (F and R) (Fig. 4: A, B). Except for *Phalaenopsis stobartiana*, there were almost tandem types in LSC region when comparing the two cp genomes. The F-type tandem was present in the LSC and IR regions; nevertheless, the P-type tandem also dispersed in the SSC region (Fig. 4: C, D). The consensus patterns in each of the two cp genomes ranged from two bp to more than 30 bp. Most of them were between 11 and 20 bp, according to the consensus pattern analysis (Fig. 4: E). The most tandem copy numbers of two cp genomes were between 2 and 4. However, there were no 8–10 copy numbers in the cp genome of *P.*

wilsonii (Fig. 4: F). The tandem repeat sequence exhibited an enrichment of A/T nucleotides.

IR expansion and contraction

Comparative analysis between the two species investigated and six sibling ones, cp genomes of *Phalaenopsis* were highly conserved structurally. Nevertheless, some structural variations were observed on these boundaries (LSC/IRb, IRb/SSC, SSC/IRa, IRa/LSC) (Fig. 5). The *rps19-trnN-trnH* was in the junction of IRa/LSC region, while *rps19-trnN* was in the junction of IRb/SSC region in *P. japonica* (Rchb.f.) Kocyan & Schuit., *P. equestris* (Schauer) Rchb.f., and *P. zhejiangensis* (Z.H.Tsi) Schuit. Moreover, the *rps19-trnH* was only in the junction of IRb/LSC in *P. mannii* Rchb.f. The *rps3-rpl22* was in the junction of IRb/LSC. The *rpl22* was expanded from LSC to IRb region in eight species with 31 and 37 bp but distanced to IRb region with 68 bp in *P. mannii*. The *rpl32* was in the SSC region in six species, except *P. equestris* and *P. zhejiangensis*. Besides, the *rpl32* was the lowest in *P. wilsonii*. The *psbA* was in the LSC region. The *ycf1* was expanded from SSC to the IRa region in five species ranging from 9 to 132 bp, while it distanced to IRa region with 31 and 41 bp in *P. japonica* and *P. zhejiangensis*. In addition, the *ycf1* was also in the IRb region, and *rpl2* was in the IRa region in *P. manni*.

Structural comparison and divergence hotspot identification analysis

Based on the annotation of *Phalaenopsis stobartiana* as the reference, the chloroplast genome sequences of eight *Phalaenopsis* species were compared by mVISTA (Fig. 6). In comparison to LSC and SSC regions, the IR regions were more conversant. In contrast, the

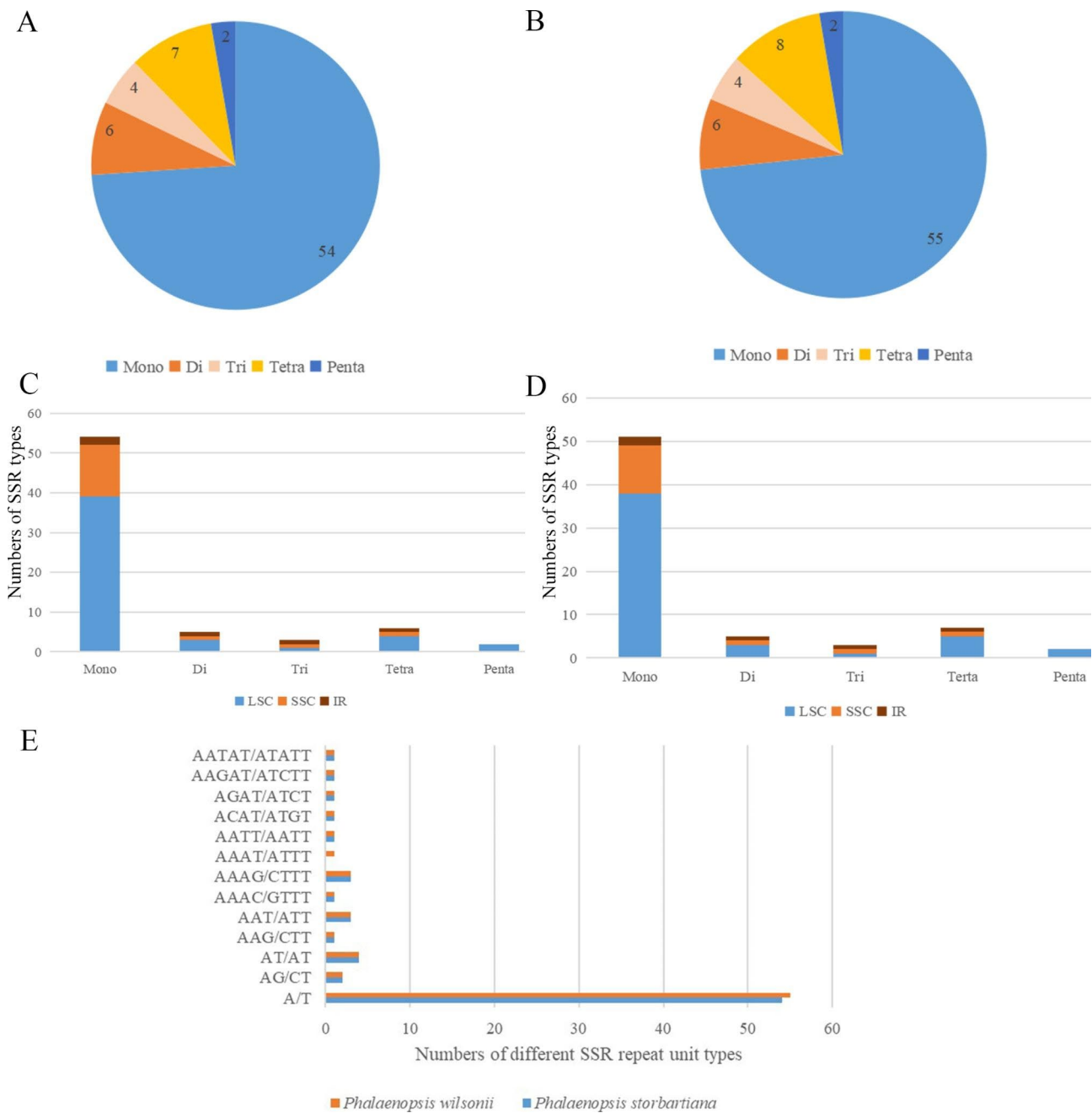


Fig. 3 SSRs analysis of two chloroplast genomes of *Phalaenopsis*. **A:** The number of SSRs distributed in different copy regions of *P. wilsonii*; **B:** The number of SSRs distributed in different copy regions of *P. stobartiana*; **C:** The number of SSR types of *P. wilsonii*; **D:** The number of SSR types of *P. stobartiana*; **E:** The number of different SSR repeat unit types

non-coding regions (CNS) were more diverse than the coding regions. The exons of the *ycf1* gene were the highest polymorphism. Moreover, the rRNA genes were highly conserved compared with other genes.

Examining CDS DNA polymorphism (Pi) revealed that the Pi value of LSC and SSC regions was greater than that of the IR regions, demonstrating that the latter were more varied. Three CDS stood out from the rest in terms of their higher Pi values: *matK* (0.01225), *psbK* (0.01434),

and *ycf1* (0.01901) (Fig. 7: A Table S5). There were two locations with high Pi values (>0.05) for the IGS, including *psbE_petL* (0.05805) and *rrn16_trnI-GAU* (0.23387). The Pi value in IGS ranged from 0.00 to 0.23 with an average of 0.024 and from 0.00 to 0.019 with an average of 0.005938 in CDS (Fig. 7: B, Table S4).

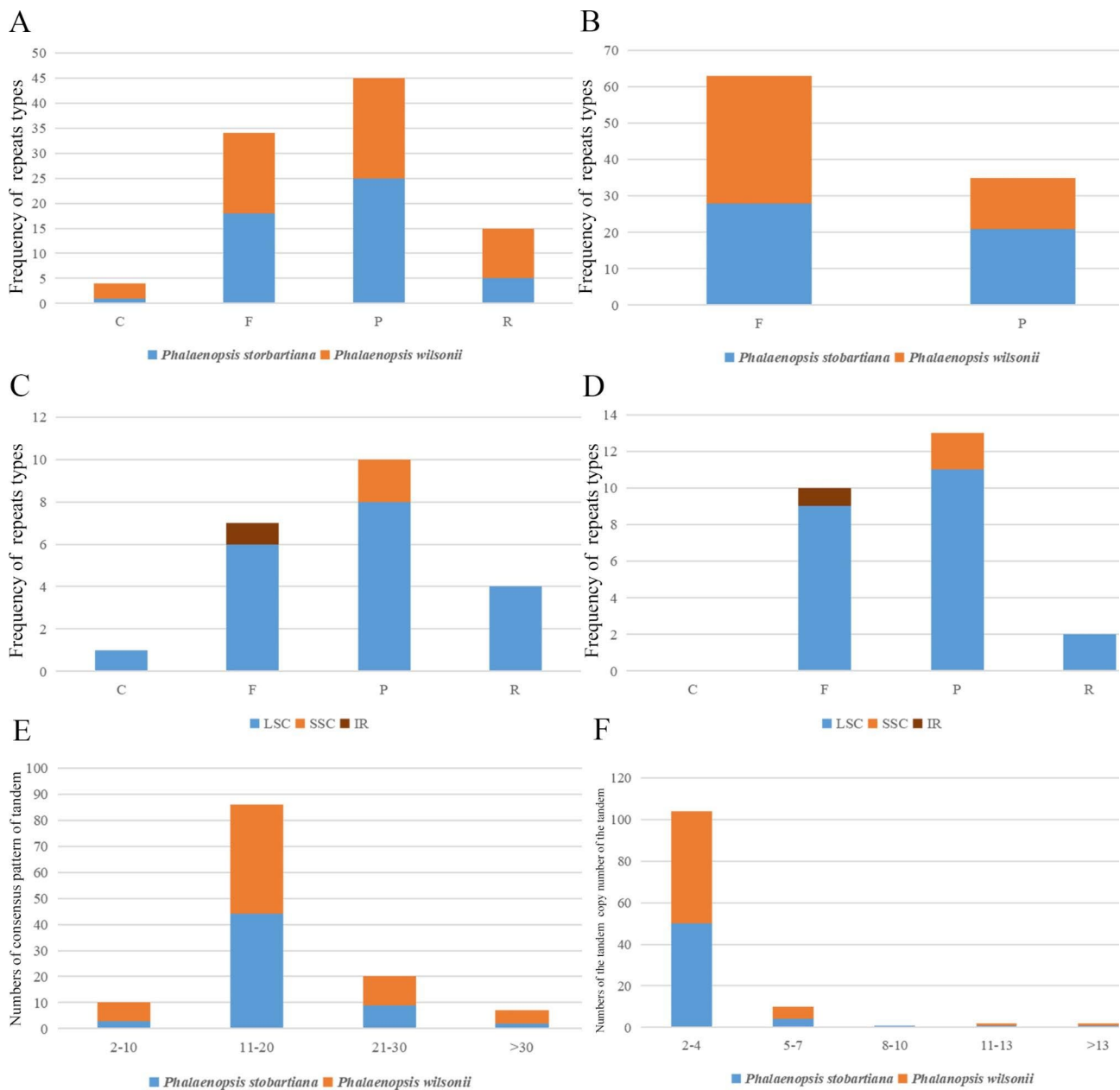


Fig. 4 Analysis of repeat sequences of two chloroplast genomes of *Phalaenopsis*. **A:** The frequency of four repeats types of complete genome sequences; **B:** The frequency of repeats types of tandem sequences; **C:** The frequency of repeats types of *P. wilsonii*; **D:** The frequency of repeats types of *P. stobartiana*; **E:** The consensus pattern of tandem; **F:** The copy number of the tandem. Abbreviations: C, complement; F, forward; P, palindromic; R, reverse

Positive selection analysis

The Bayes Empirical Bayes (BEB) method identified 45 significantly under positive selection genes, with the *psaI*, *rps3*, *rps18*, and *ycf2* genes having two significant positive selection sites. Other genes had just one substantial positive selection site aside. The number of positive selections of genes in LSC and SSC regions was higher than in IR (Table 3, Table S6).

Phylogenetic analysis

A Maximum-likelihood (ML) phylogenetic tree was constructed based on 51 single-copy CDS sequences of 60 species representing main clades in Orchidaceae, with *Iris domestica* (L.) Goldblatt & Mabb. and *Molineria capitulata* (Lour.) Herb. as outgroups, to shed fresh light on the phylogenetic position of *Phalaenopsis*. The ML tree (Fig. 8) showed that all taxon sampled formed five significant main clades corresponding to five subfamilies in Orchidaceae. Furthermore, two species investigated and six other taxa from *Phalaenopsis* were formed as a

Inverted Repeats

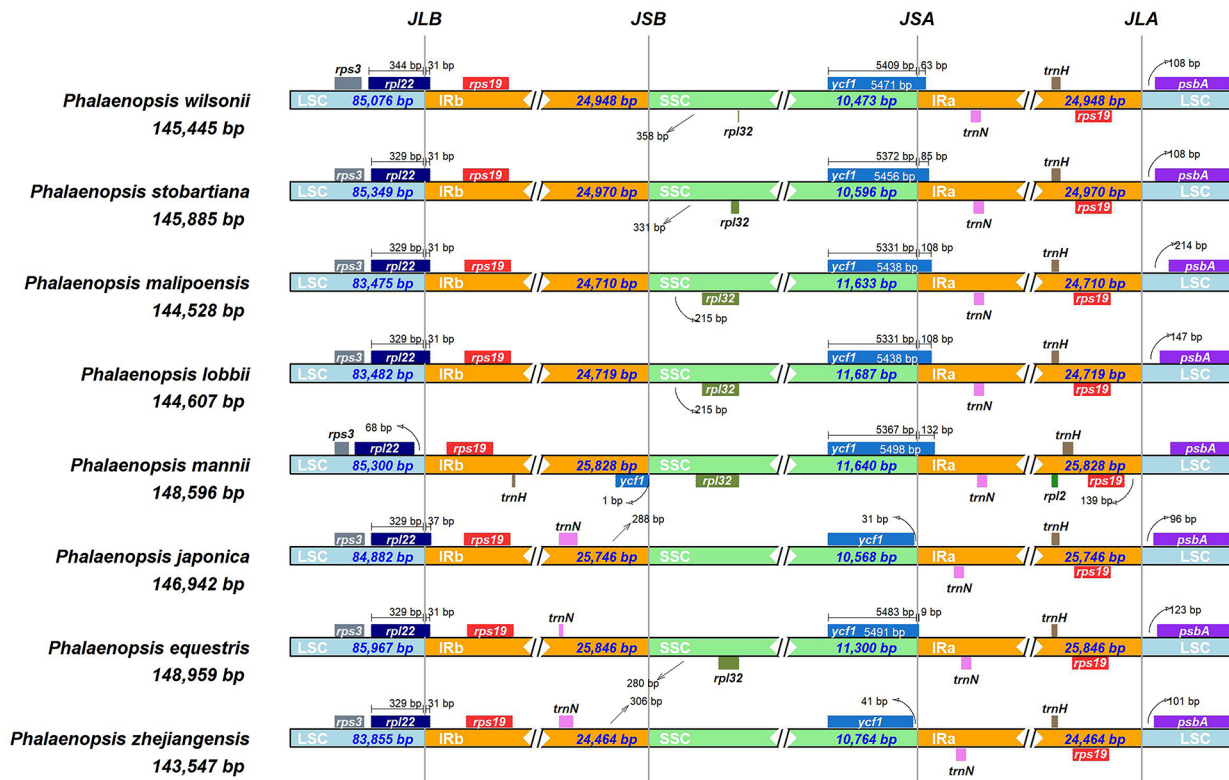


Fig. 5 Comparison of the borders of LSC, SSC, and IR regions among eight sequenced Orchidaceae chloroplast genomes. Genes transcribed forward were shown above the lines, whereas genes transcribed reverse were shown below the lines. Gene lengths in the corresponding regions were displayed above the boxes of gene names. The number of bp represented by the arrow showed genes away from a specific region of the chloroplast genome. JLB (LSC/IRb), JSB (IRb/SSC), JSA (SSC/IRa), and JLA (IRa/LSC) denoted the junction sites between each corresponding two regions on the chloroplast genome

clade with strong support (with SH-aLRT: 99.9%, and UFBoot: 100%), which was related to *Vandopsis* Pfitzer, *Vanda* R.Br., and *Holcoglossum* Schltr. in the tribe Vandaeae of subfamily Epidendroideae. It was indicated that *P. stobartiana* and *P. wilsonii* were grouped into a clade with strong support (with SH-aLRT: 100% and UFBoot: 100%).

Considering the problematic taxonomy of *Phalaenopsis*, a phylogenetic tree was created using the Maximum-likelihood (ML) method based on the *matK* sequence of 14 *Phalaenopsis* species, with four *Papilionanthe* Schltr. species and a *Holcoglossum* species as the outgroups (Fig. 9). It was shown that *Phalaenopsis stobartiana* and *P. wilsonii*, which were related to *P. zhejiangensis*, were assigned to the section *Aphyllae* subgenus *Parishianae* with strong support (SH-aLRT: 98.6% and UFBoot: 100%).

Discussion

In this study, the complete chloroplast genomes of *Phalaenopsis stobartiana* and *P. wilsonii* were sequenced and compared with those of other related orchid species to learn more about the cp genomic information and the molecular phylogeny of *Phalaenopsis*.

The chloroplast genomes of *Phalaenopsis* were highly similar [27, 29–34], which was confirmed by new data presented here. The chloroplast genomes of *P. stobartiana* and *P. wilsonii* presented a typical quadripartite circular structure with the LSC and SSC regions divided by the IR regions, which were similar to the other orchids and most of the angiosperms with no significant differences [35, 36]. The genome size was different from the previous research, and 82 CDS were annotated in these two cp genomes, but 73, 74, and 76 CDS were annotated in previous research [25, 27, 28]. The annotation of the *ndh* CDS caused this difference. *P. wilsonii* and *P. stobartiana* contained *ndh* B, C, D, E, G, J, and K, while other *Phalaenopsis* species lacked *ndh* genes or *ndh* pseudogenes [24].

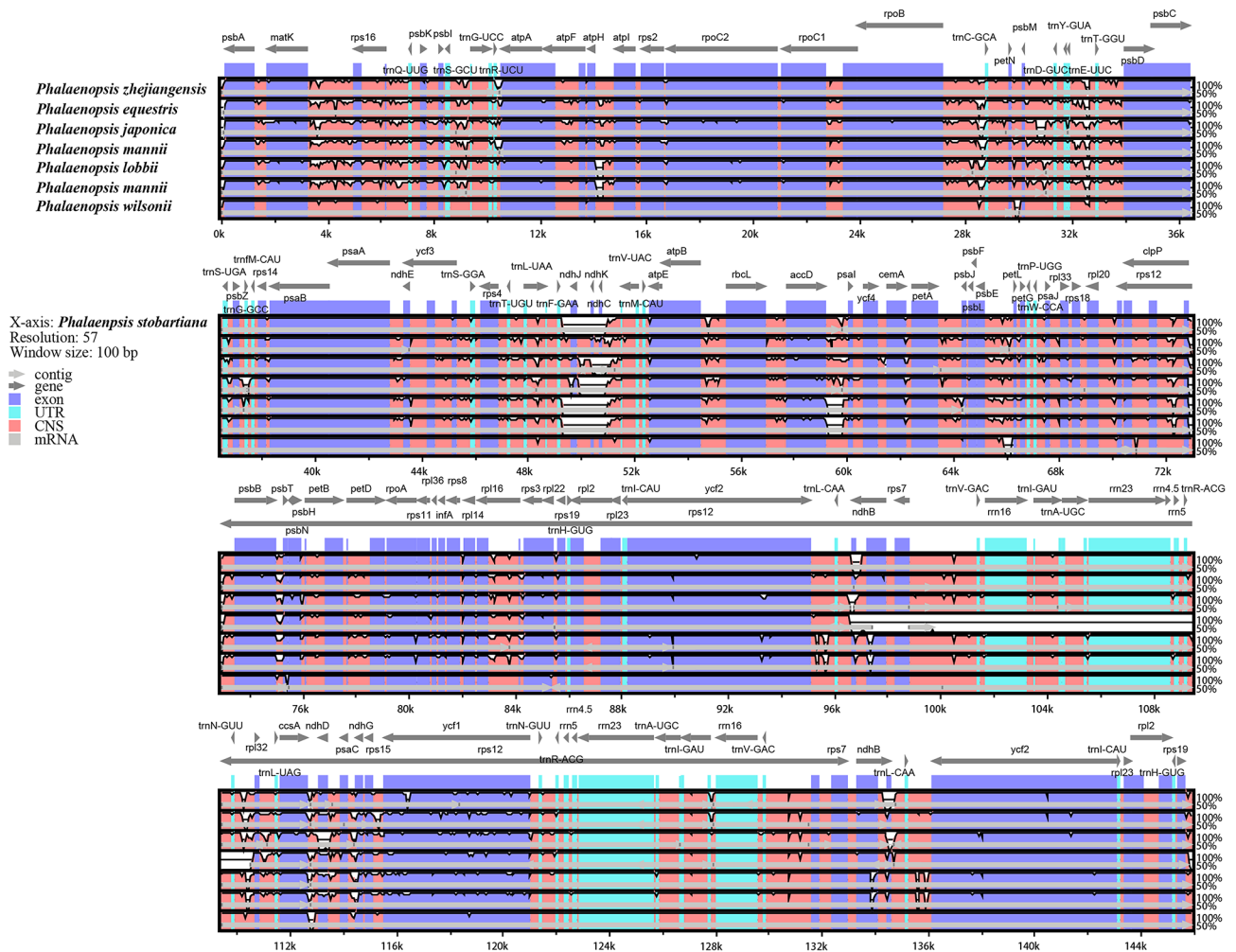


Fig. 6 Sequence alignment of eight *Phalaenopsis* chloroplast genomes using mVISTA. The vertical scale indicates the percentage of identity, ranging from 50 to 100%. The horizontal axis indicated the coordinates within the chloroplast genome. Genome regions were color coded as exon, intron, and conserved non-coding sequences (CNS) and mRNA.

Eleven *ndh* genes in cp genomes encode the NAD(p)H dehydrogenase [37]. The Apostasioideae is *ndh*-complete, Vanilloideae is *ndh*-deleted, and Cyripedioideae, Orchidoideae, and Epidendroideae are both *ndh*-complete and *ndh*-deleted, indicating that a whole functioning set of the gene was present in the common ancestor of orchids, according to earlier research [38]. In some photoautotrophic plants, the NDH complex is unnecessary [38, 39].

In phylogenetic and relationship analysis, simple sequence repeats (SSRs), also known as microsatellites, are shorter tandem repeats made up of 1–6 bp repeat units widely dispersed across the chloroplast genome [40–43]. The most frequent SSRs in the chloroplast genomes of *Phalaenopsis stobartiana* and *P. wilsonii* were mononucleotide repeats. As well as other plants, the chloroplast SSRs are almost always composed of short poly-A or poly-T repeats, and the mononucleotide repeats are typically the most common forms [44–48]. Additionally, the GC content of IR regions was much

higher than that of the LSC and SSC regions, and these characteristics were also revealed in other plant species [49, 50]. This phenomenon is caused by the presence of rRNA (*rrn4.5*, *rrn5*, *rrn16*, and *rrn23*) and tRNA genes in this region, which is the same as other Orchidaceae chloroplast genomes [40, 51–53].

Codon formation is necessary to convert genetic information from mRNA to protein [54], and codon bias, especially the third base usage pattern, is strongly connected [55]. It has been demonstrated that the GC composition influences the utilization of codons and amino acids and that the GC content of a third codon base (GC3) is thought to represent codon usage trends most closely [56]. Regarding *Phalaenopsis* species, the GC content was similar in this study but varies greatly among plant species [57]. The third-letter GC content of the two *Phalaenopsis* cp genomes was much lower than the first and second-letter GC content, and the findings were also reported in other plant species [58]. According to the

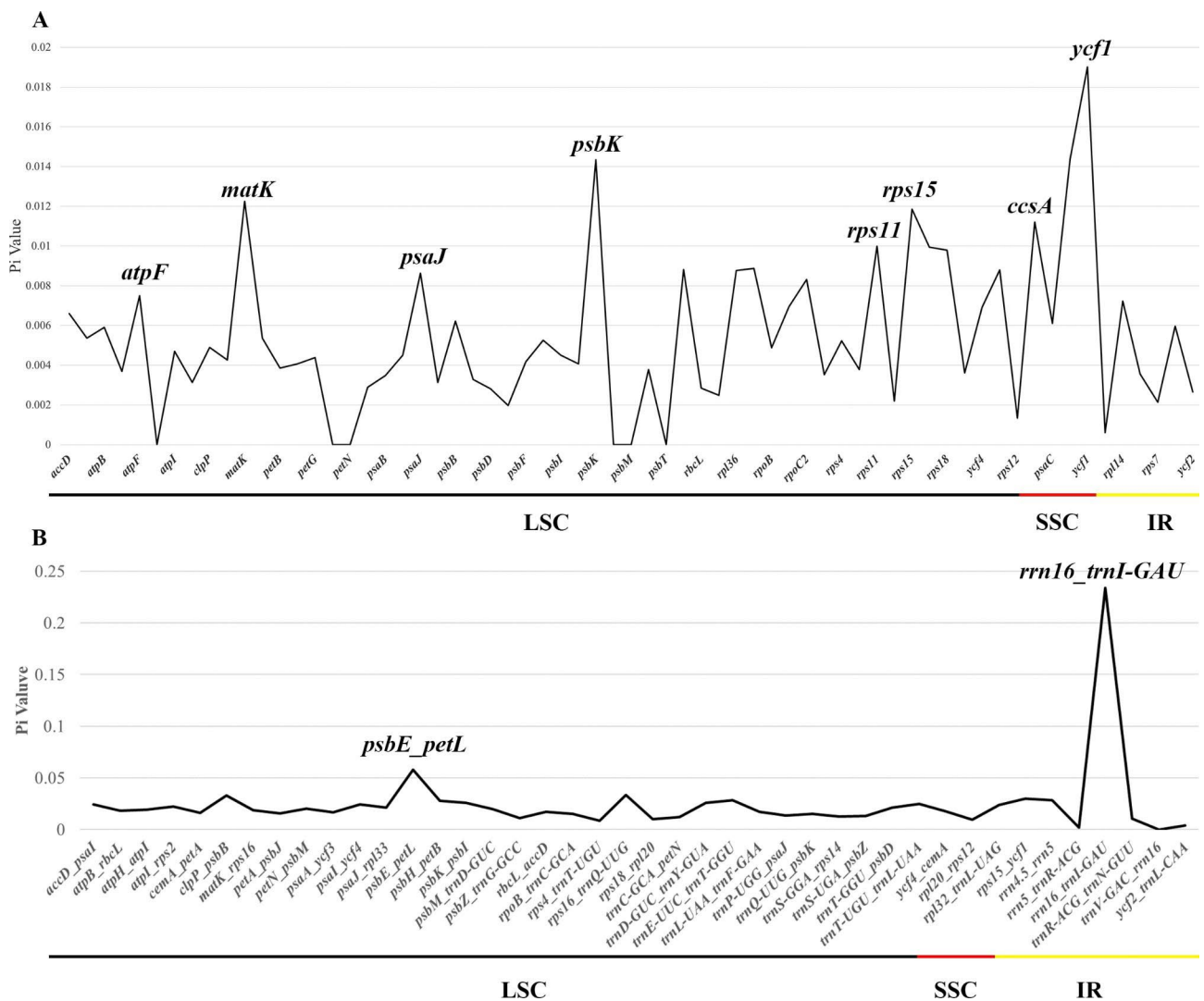


Fig. 7 Sliding window analysis of eight cp genomes. **(A)** Comparison of the nucleotide variability (PI) among CDS regions; **(B)** Comparison of the nucleotide variability among IGS regions. X-axis: position of the midpoint of a window; Y-axis: nucleotide diversity of each window. The colored lines at the bottom delineate these gene locations in different regions

RSCU analysis, six codons encoded leucine and arginine; however, only one codon encoded methionine and tryptophan, which was also reported in other Orchidaceae [40, 59].

The IR region of the chloroplast genome is thought to be the most conservative section. Nevertheless, its boundaries have frequently contracted and expanded related to the chloroplast genome evolution, which is the primary cause of the variation in chloroplast genome length [60–62]. In contrast to basal angiosperms and eudicots, most monocots have *trnH-rps19* clusters in each IR region [63]. However, in this study, the *trnH-rps19* clusters were only located in IRa region, which was consistent with *Platanthera ussuriensis* (Regel) Maxim. [64], and *Paphiopedilum henryanum* Braem [59]. The occurrence of the *trnH-rps19* gene cluster in the IR of

most monocots has been claimed to be evidence of a duplication event that occurred before the divergence of monocot lineages, and fluxes in the IR borders have been suggested to implicate the taxonomic relationships among angiosperms [41, 63]. Furthermore, *Phalaenopsis zhejiangensis* and *P. japonica* were consistent with *P. aphrodite* Rchb.f. [34], of which the *ycf1* gene was only in the SSC region. In contrast, in other species, the *ycf1* gene spanned the SSC and IRa regions.

The divergent regions could offer valuable data for DNA barcoding and phylogenetic research, which are used as molecular markers in phylogenetic reconstruction studies [65, 66]. In this study, the nucleotide sequence of non-coding regions was more varied than the coding regions, which was generally consistent with other Orchidaceae chloroplast genomes [35, 40, 64].

Table 3 The positive selection analysis of two *Phalaenopsis* species

M8							
Region	Gene name	Positive sites	Pr($\omega > 1$)	Region	Gene name	Positive sites	Pr($\omega > 1$)
LSC	<i>atpE</i>	135	1.000**	LSC	<i>psbM</i>	35	1.000**
LSC	<i>atpH</i>	82	1.000**	LSC	<i>psbN</i>	44	1.000**
LSC	<i>atpI</i>	248	1.000**	LSC	<i>psbT</i>	36	1.000**
SSC	<i>ccsA</i>	322	1.000**	LSC	<i>rbcl</i>	488	0.999**
LSC	<i>cemA</i>	230	1.000**	LSC	<i>rpl14</i>	123	1.000**
LSC	<i>infA</i>	78	1.000**	LSC	<i>rpl16</i>	136	1.000**
LSC	<i>petA</i>	321	1.000**	LSC	<i>rpl20</i>	118	1.000**
LSC	<i>petB</i>	216	1.000**	LSC	<i>rpl33</i>	67	1.000**
LSC	<i>petD</i>	164	1.000**	LSC	<i>rpl36</i>	38	1.000**
LSC	<i>petG</i>	38	1.000**	LSC	<i>rps3</i>	58 E	0.950*
LSC	<i>petL</i>	32	1.000**			219	1.000**
LSC	<i>petN</i>	30	1.000**	LSC	<i>rps4</i>	202	1.000**
LSC	<i>psaI</i>	37	1.000**	LSC	<i>rps8</i>	132	1.000**
LSC	<i>psaJ</i>	21 S	0.973*	LSC	<i>rps12</i>	124	1.000**
		45	1.000**	LSC	<i>rps14</i>	101	1.000**
LSC	<i>psbA</i>	354	1.000**	LSC	<i>rps18</i>	27 Q	0.964*
LSC	<i>psbB</i>	509	0.992**			102	1.000**
LSC	<i>psbC</i>	474	1.000**	LSC	<i>ycf4</i>	185	1.000**
LSC	<i>psbD</i>	354	1.000**	SSC	<i>rpl32</i>	58	1.000**
LSC	<i>psbE</i>	84	1.000**	IR	<i>rpl23</i>	94	1.000**
LSC	<i>psbF</i>	40	1.000**	IR	<i>rps7</i>	156	1.000**
LSC	<i>psbH</i>	74	1.000**	IR	<i>rps19</i>	93	1.000**
LSC	<i>psbI</i>	37	1.000**			562 I	0.969*
LSC	<i>psbJ</i>	41	1.000**			563 P	0.978*
LSC	<i>psbL</i>	39	0.999**			687 S	0.950*

*p > 95%; ** p > 99%

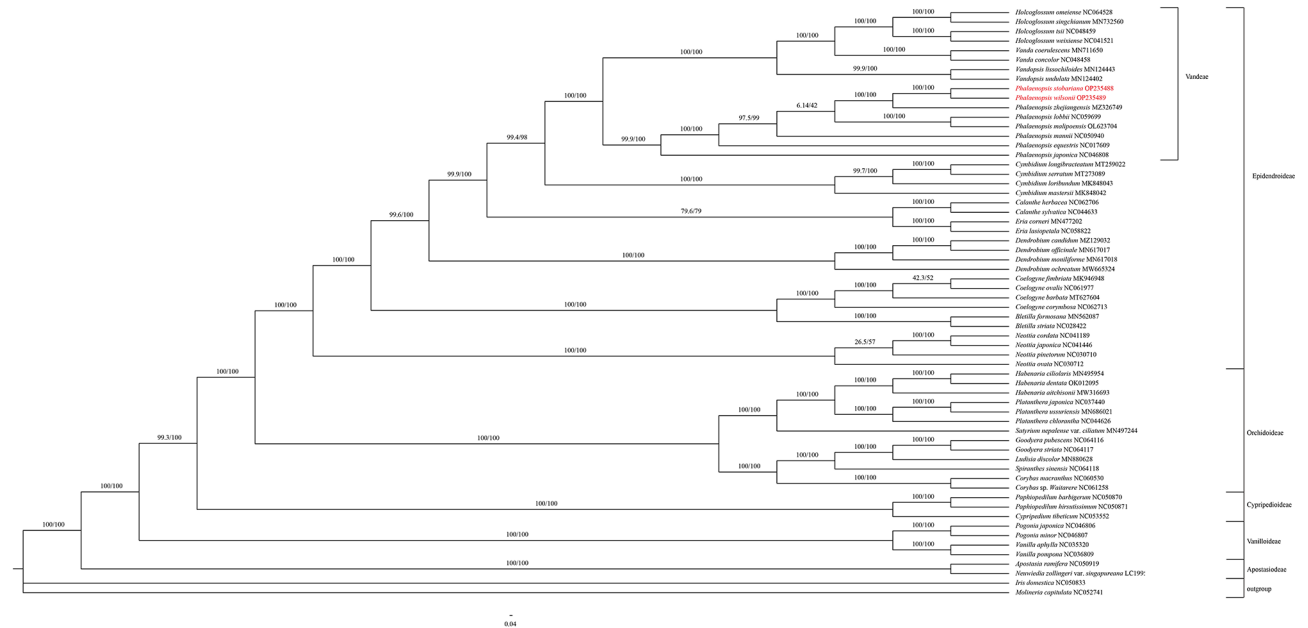


Fig. 8 Phylogenetic tree reconstructed of Orchidaceae using Maximum-likelihood (ML) method based on 51 single-copy CDS sequences of 60 orchid species, with *Iris domestica* and *Molineria capitulata* as outgroup

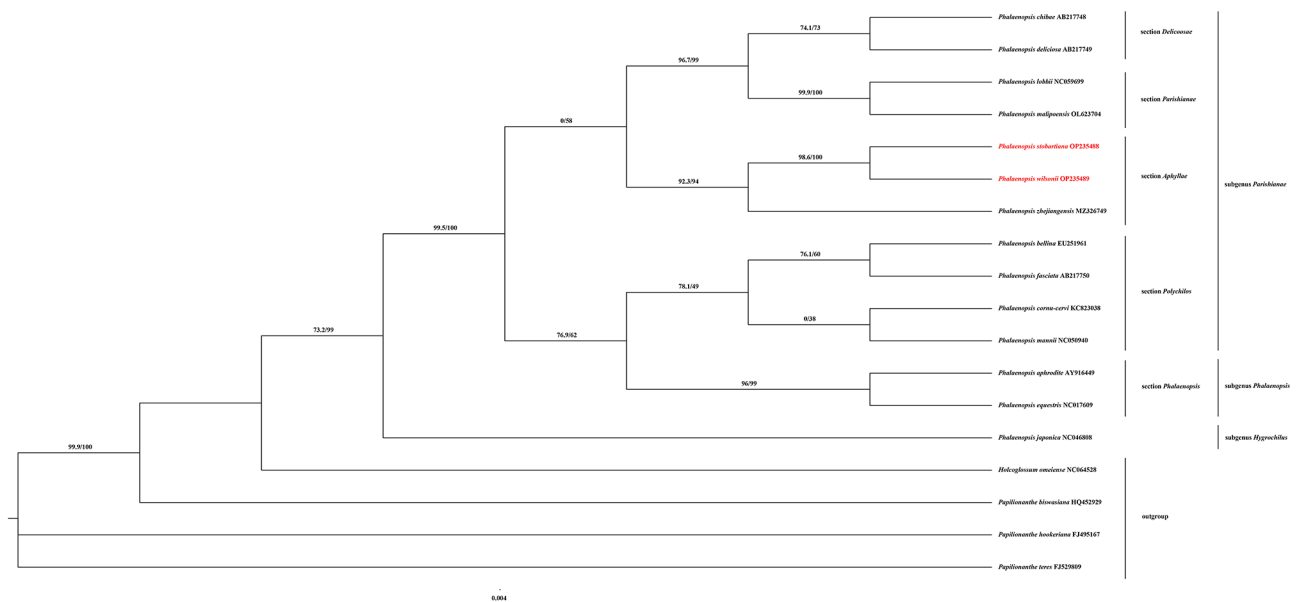


Fig. 9 Maximum-likelihood phylogenetic tree of *Phalaenopsis* based on *matK* gene of 14 *Phalaenopsis* species, with four *Papilionanthe* species and one *Holcoglossum* species as outgroup

Additionally, the CDS region analysis revealed that the genes *matK*, *psbK*, and *ycf1* had much higher Pi values. Of these, *matK* and *ycf1* have been employed as DNA markers for phylogenetic studies [66]. The *psbK* genes may be helpful for the phylogenetic analysis of chloroplast genomes in the NCBI database. In this research, *psbE_petL*, and *rrn16_trnI-GAU* also have the highest degree of variability, indicating a diversity of highly variable sequences in the chloroplast genome of orchids. However, *trnS_trnG*, *psaC_ndhE*, *clpP_psbB*, *rpl16* intron et al. were the highest degree of variability in *Phalaenopsis*, and *rpl32_trnL*, *trnE_trnT* et al. were the highest degree of variability in *Cymbidium* Sw. [67].

The ratio of substitution rates at synonymous and non-synonymous sites (dN/dS, ω) had been used to determine adaptive signals among species and infer the processes of evolution [68, 69]. Additionally, it could suggest that environmental factors impacted the evolution of chloroplast genomes, which was the primary cause of the divergence of many genes in cp chloroplast [70]. In this study, 45 genes were identified significantly under positive selection. Among them, *atpF*, *atpH*, *petL*, and *rps4* genes were also found in other orchids [40, 71, 72]; the *atpE* and *petF* were also found in other plant species [73, 74]. Moreover, these genes could be used for identification and phylogenetic research for orchids.

The structural features of the chloroplast genome would aid in understanding plant phylogeny [30–32, 52, 53, 75]. Moreover, protein-coding regions and conserved sequences are informative for phylogeny and taxonomy [76]. It was confirmed that Orchidaceae was divided into five subfamilies, and *Phalaenopsis* species were grouped

into a stable clade in Vandeeae of Epidendroideae based on CDS presented here was consistent with previous data available [5]. In addition, the plastid *matK* gene has been one of the most valuable single loci for plant phylogenetics at both shallow and deep stages of evolution [77–80].

Phalaenopsis was divided into five [9] or four subgenera in a broad definition of *Phalaenopsis* [1, 2, 15]. A ML phylogenetic tree based on *matK* sequence from 14 *Phalaenopsis* species with related taxon in Vandeeae as an outgroup presented here indicated *P. stobartiana* and *P. wilsonii* were grouped into a clade of section *Aphyllae*, subgenus *Parishianae*, together with a newly recorded species of *P. zhejiangensis*. It was congruent with the latest research on *Phalaenopsis* [2]. It seemed that *P. stobartiana* was more closely to *P. wilsonii* than *P. zhejiangensis*.

There were exhibited similar floral features in *Phalaenopsis stobartiana* and *P. wilsonii* by the presence of an inconspicuous spur and a nipple-shaped structure beneath the posterior callus [3]. However, they were distinguished by different mid-lobe of the labellum. The mid-lobe of the labellum was not obcordate without a terminal notch in *P. stobartiana*, but obcordate with an acentralapical fleshy knob in *P. wilsonii*[9].

The support of some internal nodes was low based on *matK* sequence (Fig. 9). Similarly, there were still some branches of the previous studies based on other plastid and/or nrDNA sequences that were also less supported [12, 26, 81]. The taxonomy and phylogeny of *Phalaenopsis* remained unclear and needed to be clarified by more data [1, 2].

Conclusion

Complete chloroplast genomes of *P. willsonii* and *P. stobartiana* were sequenced and analyzed, including the general genome structure, codon usage, repeat sequences, IR boundaries, DNA polymorphism, positive selection suites, and phylogenetic position. These cp genomic data were compared with those of the other six *Phalaenopsis* species available. It was confirmed that the cp genomic feature of *Phalaenopsis* was almost congruent and highly conserved, which could be used to understand the plastome evolution and evolutionary relationships of *Phalaenopsis*.

Methods and materials

Ethical statement

No specific permits were required for the collection of specimens for this study. This research was carried out in compliance with the relevant laws of China.

Plant materials and chloroplast genome sequencing

Leaf samples of *P. stobartiana* (Cultivar No. 0020180019) and *P. wilsonii* (Cultivar No. 0020172683) were cultivated and obtained from the Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Yunnan. The specimen was deposited in the Herbarium of Southwest Forestry University (HSFU, Lulu20180015, lulu@swfu.edu.cn). Total genomic DNA from fresh leaves was extracted by using the TiangenDNA kit (TIANGEN, China). An Illumina paired-end DNA library was constructed using the IlluminaTruSeq Library Preparation Kit (San Diego, CA, USA), following the manufacturer's instructions. The library was sequenced by the Illumina HiSeq 2500 sequencing platform (Illumina, CA, USA) at Personal Biotechnology Co., Ltd (Shanghai, China).

Chloroplast genome assembly and annotation

The two complete chloroplast genome from the clean reads was assembled by the GetOrganelle version 1.7.7.0 [82] and annotated the new sequences using the Geneious Prime version 2020.0.4 [83]. The complete chloroplast genomes sequences of *P. stobartiana* and *P. wilsonii* were submitted to GenBank (Accession number: OP235488 and OP235489). The circular genome maps were drawn by the OGDRAW program (<https://chlorobox.mpimp-golm.mpg.de/OGDraw>) [69].

Sequence analysis and statistics

The repeat sequences were analyzed by REPuter (<https://bibiserv.cebitec.uni-bielefeld.de/reputer/>) [84], which included forward (F), reverse (R), complement (C) and palindromic (P) repeat with maximal repeat size set to 50 bp, minimal repeat size set to 30 bp, and hamming distance set to 8 [31]. In addition to the above, the tandem repeat sequences were detected by Tandem

Repeats Finder with default parameters (<http://tandem.bu.edu/trf/trf.html>) [85]. By setting the minimum number of repeats to 10, 5, 4, 3, and 3 for mononucleotide (mono-), dinucleotide (din-), trinucleotide (tri-), tetranucleotide (tetra-), pentanucleotide (penta-), and hexanucleotide (hexan-), respectively, simple sequence repeats (SSR), a tract of repetitive DNA that typically ranges in length from 1 to 6 nucleotides, were detected by via MISA (<https://webblast.ipk-gatersleben.de/misa/index.php?action=1>) [86, 87]. Codon usage was analyzed by MEGA11 software [88], and the relative synonymous codon usage (RSCU) and amino acid frequencies were calculated with default settings [89]. The RSCU analysis was performed using JSHYCloud (<http://cloud.genepioneer.com:9929>). In addition, the GC content of the three positions was analyzed by CUSP on EMBOSS program (<http://emboss.toulouse.inra.fr/cgi-bin/emboss/cusp>) [90].

Sequence divergence and genome comparison

The pairwise alignments and sequence divergence of *Phalaenopsis wilsonii* and *P. stobartiana* with six other *Phalaenopsis* species were performed by the mVISTA with Shuffle-LAGAN mode (https://genome.lbl.gov/cgi-bin/VistaInput?num_seqs=2) [91]. Using the web tool IRSCOPE (<https://irscope.shinyapps.io/irapp/>), the contraction and extension of the IR borders between the four major areas (LSC/IRa/SSC/IRb) of the eight chloroplast genome sequences were performed [92].

Positive selection analysis

The CDS sequences were extracted by PhyloSuite version 1.2.2 [93], and the single-copy CDS sequences were aligned by MAFFT version 7 [94]. The phylogenetic tree based on CDS was platformed by MEGA 11 with Neighbor-Joining (NJ) methods [88]. The non-synonymous (dN) and synonymous (dS) substitution rates were calculated by the CodeML algorithm implemented in EasyCodeML [95] and selected the M8 mode for selection suites to detect the protein-coding genes under selection in the eight *Phalaenopsis* species.

Phylogenetic analysis

Sixty complete chloroplast genomes were chosen to build phylogenetic trees to determine the position of *P. wilsonii* and *P. stobartiana* within Orchidaceae. The complete chloroplast genome sequences of 58 orchid species were downloaded from the NCBI database, representing all five subfamilies of Orchidaceae (Orchidoideae, Epidendroideae, Cyripedioideae, Vanilloideae, and Apostasioideae). As Orchidaceae were sister to all other Asparagales [96, 97], two species from Iridaceae (*Iris domestica* (L.) Goldblatt & Mabb.) and Hypoxidaceae (*Molinieria capitulata* (Lour.) Herb.) were selected

as outgroups. These single-CDS sequences (Table S7) were extracted by PhyloSuite version 1.2.2 [93], aligned by MAFFT version 7 [94], trimmed by Gblocks [98], and concatenated by plugins in PhyloSuite version 1.2.2 [93]. The Maximum-Likelihood (ML) tree was performed in GTR+F+R3 mode based on CDS sequences by IQ-TREE 2 with 5000 ultrafast bootstrap (UFBoot) and 5000 SH-aLRT [99–101]. The 14 *matK* gene sequences, marked complete CDS, were downloaded from the NCBI database, extracted by PhyloSuite version 1.2.2 [93], and trimmed by Gblocks [98]. The phylogenetic tree based on *matK* gene sequence was constructed by IQ-TREE 2 in K3Pu+F+G4 mode, with 5000 ultrafast bootstrap (UFBoot) and 5000 SH-aLRT [99–101], with four *Papilionanthe* species and a *Holcoglossum* species as outgroups. The taxonomic system was adopted based on the broad definition of *Phalaenopsis* 2018 [2].

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09448-5>.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3
Supplementary Material 4
Supplementary Material 5
Supplementary Material 6

Acknowledgements

We thank Dr. Fei Zhao from Kunming Institute of Botany, CAS for suggestions and for revising the article and Associate Professor Yuxiao Zhang from Southwest Forestry University for providing the computer server.

Authors' contributions

LT, HD, and KT collaborated on the analysis and writing of this manuscript. YL provided the material. LL undertook the formal identification of the plant material. LL and QL contributed to the design and editing of this manuscript. All authors reviewed and approved the final manuscript.

Funding

This study was supported by the National Nature Science Foundation of China (NSFC 32060049).

Data Availability

The datasets generated or analyzed during the current study are available in the NCBI BioProject (PRJNA861671, SRA SRR20710655, and SRA20710656).

Declarations

Ethics approval and consent to participate

The study was conducted the plant material that complies with relevant institutional, national, and international guidelines and legislation. *P. wilsonii* (Cultivar No. 0020172683) and *P. stobartiana* (Cultivar No. 0020180019) were cultivated in Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Biological Conservation, Southwest Forestry University, Kunming, Yunnan 650224, China

²Department of Life Science, Southwest Forestry University, Kunming, Yunnan 650224, China

³Department of Horticulture and Gardening, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, Yunnan 666303, China

⁴Kunming Xianghao Technology Co. Ltd., Kunming, Yunnan 650204, China

Received: 13 October 2022 / Accepted: 14 June 2023

Published online: 27 June 2023

References

- Pridgeon AM, Cribb PJ, Chase MW, Rasmussen FN. Epidendroideae (Part Three). Genera Orchidacearum. Great Clarendon Street, Oxford, OX2 6DP, United Kingdom: Oxford University Press; 2014. 234.
- Liu Z, Wang J, Gruss O, Lan S. The genus *Phalaenopsis* in World. Higher Education Press; 2022.
- Chen X, Wood JJ. *Phalaenopsis* Blume. In: Flora of China. Beijing: Sciences Press and St. Louis: Missouri Botanical Garden Press; 2009. p. 478–83.
- Tsai C-C, Chen Y-KH, Chen C-H, Weng I-S, Tsai C-M, Lee S-R, et al. Cultivar identification and genetic relationship of mango (*Mangifera indica*) in Taiwan using 37 SSR markers. *Sci Hort*. 2013;164:196–201.
- Chase MW, Cameron KM, Freudenstein JV, Pridgeon AM, Salazar G, van den Berg C, et al. An updated classification of Orchidaceae. *Bot J Linn Soc*. 2015;177:151–74.
- Reichenbach HG. Eine neue *Phalaenopsis*. In: Eine neue *Phalaenopsis*. Hamburger Garten- und Blumenzeitung. p. 114–7.
- Rolfe RA. A revision of the genus *Phalaenopsis*. *Gard Chron* ns 1886;26:372.
- Sweet HR. The genus *Phalaenopsis*. *Orchid Digest*; 1980.
- Christenson EA, Alliance IP. *Phalaenopsis*: a Monograph. Timber Press; 2001.
- Padolina J, Linder CR, Simpson BB. A phylogeny of *Phalaenopsis* using multiple chloroplast markers. *Selbyana*. 2005;26:23–7.
- Tsai CC, Huang SC, Chou CH. Molecular phylogeny of *Phalaenopsis* Blume (Orchidaceae) based on the internal transcribed spacer of the nuclear ribosomal DNA. *Plant Syst Evol*. 2005;256:1–16.
- Tsai CC, Chiang YC, Huang SC, Chen CH, Chou CH. Molecular phylogeny of *Phalaenopsis* Blume (Orchidaceae) on the basis of plastid and nuclear DNA. *Plant Syst Evol*. 2010;288:77–98.
- Yukawa T, Tanaka N, Murata J. *Doritis natmataungensis* (Orchidaceae), a New Species from Myanmar. *Acta Phytotaxonomica et Geobotanica*. 2010;60:167–70.
- Cribb PJ, Schuiteman A. *Phalaenopsis*: Classif Renziana. 2012;2:14–40.
- Kocyan A, Schuiteman A. New combinations in Aeriinae (Orchidaceae). *Phytotaxa*. 2014;161:61–85.
- Li M, Gruss O, Liu Z-J. Nomenclature changes in *Phalaenopsis* subgen. *Hygrochilus* (Orchidaceae; Epidendroideae; Vandeeae) based on DNA evidence. *Phytotaxa*. 2016;275:55–61.
- Hidayat T, Weston P, Yukawa T, Ito M, Rice R. Phylogeny of Subtribe Aeriinae (Orchidaceae) inferred from DNA sequences data: advanced analyses including Australasian Genera. *Jurnal Teknologi*. 2013;59.
- Topik H, Yukawa T, Ito M. Molecular phylogenetics of subtribe Aeriinae (Orchidaceae): insights from plastid *matK* and nuclear ribosomal ITS sequences. *J Plant Res*. 2005;118:271–84.
- Tang J, Xia H, Cao M, Zhang X, Zeng W, Hu S, et al. A comparison of Rice Chloroplast Genomes. *Plant Physiol*. 2004;135:412–20.
- Yao X, Tang P, Li Z, Li D, Liu Y, Huang H. The first complete chloroplast genome sequences in Actinidiaceae: genome structure and comparative analysis. *PLoS ONE*. 2015;10:e0129347.
- Yang Y, Zhou T, Duan D, Yang J, Feng L, Zhao G. Comparative analysis of the complete chloroplast genomes of five *Quercus* Species. *Front Plant Sci*. 2016;7.

22. Xiao-Ming Z, Junrui W, Li F, Sha L, Hongbo P, Lan Q, et al. Inferring the evolutionary mechanism of the chloroplast genome size by comparing whole-chloroplast genome sequences in seed plants. *Sci Rep.* 2017;7:1555.
23. Li Y, Li Z-H, Schuiteman A, Chase M, Jianwu L, Huang W-C, et al. Phylogenomics of Orchidaceae based on plastid and mitochondrial genomes. *Mol Phylogenet Evol.* 2019;139:106540.
24. Kim Y-K, Jo S, Cheon S-H, Kwak M, Kim Y-D, Kim K-J. Plastome evolution and phylogeny of subtribe Aeridinae (Vandaeae, Orchidaceae). *Mol Phylogenet Evol.* 2020;144:106721.
25. Cao Y-H, Hu M-J, Tong Y, Zhang Y-P, Zheng R-Y, Zhao K, et al. Basic chloroplast genome characterization of *Phalaenopsis stobartiana* (Orchidaceae) from China. *Mitochondrial DNA Part B.* 2022;7:257–8.
26. Deng H, Zhang G-Q, Liu Z-J, Wang Y. A new species and a new combination of *Phalaenopsis* (Orchidaceae: Epidendroideae: Aeridinae): evidence from morphological and DNA analysis. *Phytotaxa.* 2015;238:243–54.
27. Xia K, Liu D-K, Wang J-Y. The complete chloroplast genome sequence of *Phalaenopsis wilsonii* (Orchidaceae). *Mitochondrial DNA Part B.* 2021;6:33303–5.
28. Fan Z-F, Yu D-Y, Ma C-L. The complete chloroplast genome sequence of *Phalaenopsis wilsonii* Rolfe, a vulnerable wild moth orchid species (Orchidaceae). *Mitochondrial DNA Part B.* 2021;6:2903–5.
29. Cai J, Liu X, Vanneste K, Proost S, Tsai W-C, Liu K-W, et al. The genome sequence of the orchid *Phalaenopsis equestris*. *Nat Genet.* 2015;47:65–72.
30. Kim G-B, Kwon Y, Yu H-J, Lim K-B, Seo J-H, Mun J-H. The complete chloroplast genome of *Phalaenopsis "Tiny star"*. *Mitochondrial DNA Part A.* 2016;27:1300–2.
31. Zhang Y, Chen B, Zheng Y, Cao Y, Wei Z, Zhao K, et al. Characterization of the complete chloroplast genome of *Phalaenopsis lobbii* (Orchidaceae), an important horticultural plant in China. *Mitochondrial DNA Part B.* 2020;5:3468–9.
32. Chen B, Zhang Y, Cao Y, Zheng Y, Wei Z, Zhao K, et al. Chloroplast characterizations of a *Phalaenopsis* native to China, *Phalaenopsis mannii* (Orchidaceae). *Mitochondrial DNA Part B.* 2020;5:3707–8.
33. Wang J-Y, Liu Z-J, Zhang G-Q, Peng C-C. The complete chloroplast genome sequence of *Phalaenopsis lowii* (Orchidaceae). *Mitochondrial DNA Part B.* 2019;4:3569–70.
34. Chang C-C, Lin H-C, Lin I-P, Chow T-Y, Chen H-H, Chen W-H, et al. The Chloroplast Genome of *Phalaenopsis aphrodite* (Orchidaceae): comparative analysis of evolutionary rate with that of Grasses and its phylogenetic implications. *Mol Biol Evol.* 2006;23:279–91.
35. Jiang K, Miao L-Y, Wang Z-W, Ni Z-Y, Hu C, Zeng X-H et al. Chloroplast Genome Analysis of two Medicinal *Coelogyne* spp. (Orchidaceae) Shed Light on the genetic information, comparative Genomics, and Species Identification. *Plants.* 2020;9.
36. Liu H-Y, Yu Y, Deng Y-Q, Li J, Huang Z-X, Zhou S-D. The Chloroplast Genome of *Lilium henrici*: genome structure and comparative analysis. *Molecules.* 2018;23.
37. Lin C-S, Chen JJW, Chiu C-C, Hsiao HCW, Yang C-J, Jin X-H, et al. Concomitant loss of *NDH* complex-related genes within chloroplast and nuclear genomes in some orchids. *Plant J.* 2017;90:994–1006.
38. Lin C-S, Chen JJW, Huang Y-T, Chan M-T, Daniell H, Chang W-J, et al. The location and translocation of *ndh* genes of chloroplast origin in the Orchidaceae family. *Sci Rep.* 2015;5:9040.
39. Liu D-K, Tu X-D, Zhao Z, Zeng M-Y, Zhang S, Ma L, et al. Plastid phylogenomic data yield new and robust insights into the phylogeny of *cleisostoma–gastrochilus* clades (Orchidaceae, Aeridinae). *Mol Phylogenet Evol.* 2020;145:106729.
40. Jiang H, Tian J, Yang J, Dong X, Zhong Z, Mwachala G, et al. Comparative and phylogenetic analyses of six Kenya *Polystachya* (Orchidaceae) species based on the complete chloroplast genome sequences. *BMC Plant Biol.* 2022;22:177.
41. Agrama HA, Tuinstra MR. Phylogenetic diversity and relationships among *sorghum* accessions using SSRs and RAPDs. *Afr J Biotechnol.* 2003;2:334–40.
42. Madhumati B. Potential and application of molecular markers techniques for plant genome analysis. *Int J Pure App Biosci.* 2014;2:169–88.
43. Li X, Zhao Y, Tu X, Li C, Zhu Y, Zhong H, et al. Comparative analysis of plastomes in Oxalidaceae: phylogenetic relationships and potential molecular markers. *Plant Divers.* 2021;43:281–91.
44. Vu H-T, Tran N, Nguyen T-D, Vu Q-L, Bui M-H, Le M-T, et al. Complete chloroplast genome of *Paphiopedilum delenatii* and phylogenetic Relationships among Orchidaceae. *Plants.* 2020;9:61.
45. Djedid IK, Terzaghi M, Brundu G, Cicatelli A, Laouar M, Guarino F et al. Genetic diversity and differentiation of Eleven *Medicago* Species from Campania Region revealed by Nuclear and Chloroplast Microsatellites markers. *Genes.* 2022;13.
46. NAVASCUÉS M, EMERSON BC. Chloroplast microsatellites: measures of genetic diversity and the effect of homoplasy. *Mol Ecol.* 2005;14:1333–41.
47. Provan J, Powell W, Hollingsworth PM. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends Ecol Evol.* 2001;16:142–7.
48. YANG J, ZHU Z, FAN Y, ZHU F, CHEN Y, NIU Z, et al. Comparative plastomic analysis of three *Bulbophyllum* medicinal plants and its significance in species identification. *Acta Pharm Sinica.* 2020;12:2736–45.
49. Alzahrani DA, Yaradua SS, Albokhari EJ, Abba A. Complete chloroplast genome sequence of *Barleria prionitis*, comparative chloroplast genomics and phylogenetic relationships among Acanthoideae. *BMC Genomics.* 2020;21:393.
50. Hu S, Sablok G, Wang B, Qu D, Barbaro E, Viola R, et al. Plastome organization and evolution of chloroplast genes in *Cardamine* species adapted to contrasting habitats. *BMC Genomics.* 2015;16:306.
51. Dong W-L, Wang R-N, Zhang N-Y, Fan W-B, Fang M-F, Li Z-H. Molecular evolution of Chloroplast Genomes of Orchid Species: insights into phylogenetic relationship and adaptive evolution. *Int J Mol Sci.* 2018;19.
52. Ma Q, Li S, Bi C, Hao Z, Sun C, Ye N. Complete chloroplast genome sequence of a major economic species, *Ziziphus jujuba* (Rhamnaceae). *Curr Genet.* 2017;63:117–29.
53. Wang S, Yang C, Zhao X, Chen S, Qu G-Z. Complete chloroplast genome sequence of *Betula platyphylla*: gene organization, RNA editing, and comparative and phylogenetic analyses. *BMC Genomics.* 2018;19:1–15.
54. Qiu S, Zeng K, Slotte T, Wright S, Charlesworth D. Reduced efficacy of natural selection on Codon usage Bias in Selfing *Arabidopsis* and *Capsella* Species. *Genome Biol Evol.* 2011;3:868–80.
55. SHANG, Ming-zhao LF, Jin-ping HUA. Kun-bo. Analysis on Codon usage of Chloroplast Genome of *Gossypium hirsutum*. *Scientia Agricultura Sinica.* 2011;44:245–53.
56. Chen L, Liu T, Yang D, Nong X, Xie Y, Fu Y, et al. Analysis of codon usage patterns in *Taenia pisiformis* through annotated transcriptome data. *Biochem Biophys Res Commun.* 2013;430:1344–8.
57. Ng P-K, Lin S-M, Lim P-E, Liu L-C, Chen C-M, Pai T-W. Complete chloroplast genome of *Gracilaria firma* (Gracilariaceae, Rhodophyta), with discussion on the use of chloroplast phylogenomics in the subclass Rhodmeniophycidae. *BMC Genomics.* 2017;18:40.
58. Gao Y, Lu Y, Song Y, Jing L. Analysis of codon usage bias of *WRKY* transcription factors in *Helianthus annuus*. *BMC Genomic Data.* 2022;23:46.
59. Liu H, Ye H, Zhang N, Ma J, Wang J, Hu G, et al. Comparative analyses of chloroplast genomes provide Comprehensive Insights into the adaptive evolution of *Paphiopedilum* (Orchidaceae). *Horticulturae.* 2022;8:391.
60. Yi X, Gao L, Wang B, Su Y-J, Wang T. The complete chloroplast genome sequence of *Cephalotaxus oliveri* (Cephalotaxaceae): evolutionary comparison of *Cephalotaxus* chloroplast DNAs and insights into the loss of inverted repeat copies in gymnosperms. *Genome Biol Evol.* 2013;5:688–98.
61. Zhang H, Li C, Miao H, Xiong S. Insights from the complete chloroplast genome into the evolution of *Sesamum indicum* L. *PLoS ONE.* 2013;8:e80508.
62. Daniell H, Lin C-S, Yu M, Chang W-J. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 2016;17:134.
63. Wang R-J, Cheng C-L, Chang C-C, Wu C-L, Su T-M, Chaw S-M. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evol Biol.* 2008;8:36.
64. Han C, Ding R, Zong X, Zhang L, Chen X, Qu B. Structural characterization of *Platanthera ussuriensis* chloroplast genome and comparative analyses with other species of Orchidaceae. *BMC Genomics.* 2022;23:84.
65. Činkulov M, Popovski Z, Porcu K, Tanaskovska B, Hodžić A, Bytyqi H, et al. Genetic diversity and structure of the West Balkan Pramenka sheep types as revealed by microsatellite and mitochondrial DNA analysis. *J Anim Breed Genet.* 2008;125:417–26.
66. Menezes APA, Resende-Moreira LC, Buzatti RSO, Nazareno AG, Carlsen M, Lobo FP, et al. Chloroplast genomes of *Byrsonima* species (Malpighiaceae): comparative analysis and screening of high divergence sequences. *Sci Rep.* 2018;8:2210.
67. Shaw J, Shafer HL, Leonard OR, Kovach MJ, Schorr M, Morris AB. Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: the tortoise and the hare IV. *Am J Bot.* 2014;101:1987–2004.
68. Williams MJ, Zapata L, Werner B, Barnes CP, Sottoriva A, Graham TA. Measuring the distribution of fitness effects in somatic evolution by combining clonal dynamics with dN/dS ratios. *Elife.* 2020;9.

69. Kryazhimskiy S, Plotkin JB. The population genetics of dN/dS. *PLoS Genet.* 2008;4:e1000304.
70. Zuo L-H, Shang A-Q, Zhang S, Yu X-Y, Ren Y-C, Yang M-S, et al. The first complete chloroplast genome sequences of *Ulmus* species by de novo sequencing: genome comparative and taxonomic position analysis. *PLoS ONE.* 2017;12:e0171264.
71. Carbonell-Caballero J, Alonso R, Ibañez V, Terol J, Talon M, Dopazo J. A phylogenetic analysis of 34 Chloroplast Genomes elucidates the Relationships between Wild and domestic species within the Genus *Citrus*. *Mol Biol Evol.* 2015;32:2015–35.
72. Tang H, Tang L, Shao S, Peng Y, Li L, Luo Y. Chloroplast genomic diversity in *Bulbophyllum* section *Macrocaulia* (Orchidaceae, Epidendroideae, Malaxideae): insights into species divergence and adaptive evolution. *Plant Divers.* 2021;43:350–61.
73. Giang VNL, Waminal NE, Park H-S, Kim N-H, Jang W, Lee J, et al. Comprehensive comparative analysis of chloroplast genomes from seven *Panax* species and development of an authentication system based on species-unique single nucleotide polymorphism markers. *J Ginseng Res.* 2020;44:135–44.
74. Wu Y, Liu F, Yang D-G, Li W, Zhou X-J, Pei X-Y, et al. Comparative chloroplast genomics of *Gossypium* species: insights into repeat sequence variations and phylogeny. *Front Plant Sci.* 2018;9:376.
75. Jheng C-F, Chen T-C, Lin J-Y, Chen T-C, Wu W-L, Chang C-C. The comparative chloroplast genomic analysis of photosynthetic orchids and developing DNA markers to distinguish *Phalaenopsis* orchids. *Plant Sci.* 2012;190:62–73.
76. Bobik K, Burch-Smith TM. Chloroplast signaling within, between and beyond cells. *Front Plant Sci.* 2015;6.
77. Hilu KW, Borsch T, Müller K, Soltis DE, Soltis PS, Savolainen V, et al. Angiosperm phylogeny based on < 011 > matK sequence information. *Am J Bot.* 2003;90:1758–76.
78. Döring E, Schneider J, Hilu KW, Röser M. Phylogenetic Relationships in the *Aveneae/Poaeae* Complex (Pooideae, Poaceae). *Kew Bull.* 2007;62:407–24.
79. Soltis DE, Soltis PS. Choosing an Approach and an appropriate gene for phylogenetic analysis. In: Soltis DE, Soltis PS, Doyle JJ, editors. *Molecular systematics of plants II: DNA sequencing.* Boston, MA: Springer US; 1998. pp. 1–42.
80. Hilu KW, Black C, Diouf D, Burleigh JG. Phylogenetic signal in matK vs. trnK: a case study in early diverging eudicots (angiosperms). *Mol Phylogenet Evol.* 2008;48:1120–30.
81. Yukawa T, Kita K, Handa T, Hidayat T, Ito M. Molecular phylogenetics of *Phalaenopsis* (Orchidaceae) and allied Genera: re-evaluation of generic concepts. *Acta Phytotaxonomica et Geobotanica.* 2005;56:141–61.
82. Jin J-J, Yu W-B, Yang J-B, Song Y, Yi T-S, Li D-Z. GetOrganelle: a simple and fast pipeline for de novo assembly of a complete circular chloroplast genome using genome skimming data. *BioRxiv.* 2018;4:256479.
83. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 2012;28:1647–9.
84. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 2001;29:4633–42.
85. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 1999;27:573–80.
86. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. *Bioinformatics.* 2017;33:2583–5.
87. Thiel T, Michalek W, Varshney R, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L). *Theor Appl Genet.* 2003;106:411–22.
88. Kumar S, Nei M, Dudley J, Tamura K. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform.* 2008;9:299–306.
89. Bylaiah S, Shedole S, Suresh KP, Gowda L, Patil SS, Indrabalan UB. Analysis of Codon usage Bias in Cya, Lef, and pag genes exists in px01 plasmid of *Bacillus anthracis*. *ICT analysis and applications.* Springer; 2022. 1–9.
90. Rice P, Longden I, Bleasby A. EMBOSS: the european molecular biology open software suite. *Trends Genet.* 2000;16:276–7.
91. Brudno M, Malde S, Poliakov A, Do CB, Couronne O, Dubchak I, et al. Global alignment: finding rearrangements during alignment. *Bioinformatics.* 2003;19(suppl1):i54–62.
92. Amiryousefi A, Hyvönen J, Pocza P. IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics.* 2018;34:3030–1.
93. Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, et al. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol Ecol Resour.* 2020;20:348–55.
94. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30:772–80.
95. Gao F, Chen C, Arab DA, Du Z, He Y, Ho SYW. EasyCodeML: a visual tool for analysis of selection using CodeML. *Ecol Evol.* 2019;3:3891–8.
96. Pires J, Maureira I, Givnish T, Systma K, Seberg O, Peterson G, et al. Phylogeny, genome size, and chromosome evolution of Asparagales. *Aliso.* 2006;22:287–304.
97. Givnish T, Pires J, Graham S, McPherson M, Prince L, Patterson T, et al. Phylogenetic Relationships of Monocots based on the highly informative Plastid Gene *ndhF*. *Aliso.* 2006;22:28–51.
98. Talavera G, Castresana J. Improvement of Phylogenies after removing Divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol.* 2007;56:564–77.
99. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 2020;37:1530–4.
100. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the Ultrafast bootstrap approximation. *Mol Biol Evol.* 2018;35:518–22.
101. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 2017;14:587–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.