The Chloroplast Genome Evolution of Venus Slipper (Paphiopedilum): IR Expansion, SSC Contraction, and Highly Rearranged SSC Regions

Yan-Yan Guo (guoyy@henau.edu.cn)
Henan Agricultural University
Jia-Xing Yang
Henan Agricultural University
Guo-Qiang Zhang
The National Orchid Conservation Center of China
Zhong-Jian Liu
Fujian Agriculture and Forestry University

Research Article

Keywords: Orchidaceae, Paphiopedilum, phylogenomics, plastome, boundary shift, IR/SSC boundary, gene loss, pseudogenization

Posted Date: March 3rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-257472/v1

Abstract

Background: *Paphiopedilum* is the largest genus of slipper orchids. Previous studies showed that the phylogenetic relationships of this genus are not well resolved, and sparse taxon sampling documented inverted repeat (IR) expansion and small single copy (SSC) contraction of the chloroplast genomes of *Paphiopedilum*. Here, we sequenced, assembled, and annotated 77 plastomes of *Paphiopedilum* species. The phylogeny based on the plastome resolved the relationships of the genus except for the phylogenetic position of two unstable species. We used phylogenetic and comparative genomic approaches to elucidate the plastome evolution of *Paphiopedilum*.

Results: The plastomes of *Paphiopedilum* have conserved genome structure and gene content except in the SSC region. The large single copy/inverted repeat (LSC/IR) boundaries are relatively stable, while the boundaries of inverted repeat/small single copy (IR/SSC) boundaries varied among species. Corresponding to the IR/SSC boundary shifts, the chloroplast genomes of the genus experienced IR expansion and SSC contraction. The IR region incorporated one to six genes of the SSC region. Unexpectedly, great variation in the size, gene order, and gene content of the SSC regions was found, especially in the subg. *Parvisepalum*. Furthermore, *Paphiopedilum* provides evidence for the ongoing degradation of the *ndh* genes in the photoautotrophic plants. The estimated substitution rates of the protein coding genes show accelerated rates of evolution in *clpP*, *psbH*, and *psbZ*. Genes transferred to the IR region due to the boundary shift also have higher substitution rates.

Conclusions: We found IR expansion and SSC contraction in the chloroplast genomes of *Paphiopedilum* with dense sampling, and the genus shows variation in the size, gene order, and gene content of the SSC region. This genus provides an ideal system to investigate the dynamics of plastome evolution.

Background

High throughput sequencing technology has made obtaining chloroplast genome (plastome) sequences more practical [1]. And several studies have used plastome data to address the phylogenetic relationships among land plants, chloroplast genome evolution, and patterns and rates of nucleotide substitutions [2–5]. These studies indicate that the chloroplast genome has striking variations in genome size, genome structure, and gene substitution rate.

Though the average chloroplast genome of land plants is 151 kb, the average inverted repeat (IR) region of land plants is 25 kb. Published studies show extensive variation in the plastome size and length of the IR region. The plastome size of *Pelargonium transvaalense* is 242,575 bp, with an IR region of 87,724 bp [6]. On the other hand, IR region loss has been detected in all lineages across land plants [7]. High throughout sequencing provides a good opportunity to test the plastome evolution in more groups.

Paphiopedilum (Venus slipper) is the largest genus of slipper orchids, mainly distributed in southeast Asia, with half of these species growing on islands. The sequenced plastomes of the *Paphiopedilum* genus show the expansion of the IR region, while the SSC regions are greatly reduced in size and gene

content [8–11]. Even the typical SSC genes (*ycf1*, *psaC*, and *ndhD*) transferred to the IR region in *Paphiopedilum armeniacum* [8]. However, present studies of *Paphiopedilum* plastomes are based on sparse taxon sampling, and the pattern of IR expansion/SSC contraction at the genus level are unknown. *Paphiopedilum* provides a unique opportunity to study the dynamics of the boundary shift impact on plastid genome structure and sequence evolution.

Additionally, there are still unresolved phylogenetic questions in the *Paphiopedilum* genus. For instance, present chloroplast markers cannot solve the deep phylogenetic relationship. Resolving relationships in *Paphiopedilum* will promote the speciation study of the genus. Recent phylogenetic analyses indicated widespread reticulate evolution of the genus and that sect. *Cochlopetalum* is not monophyly in the chloroplast gene tree [12, 13]. Interestingly, there is also systematic uncertainty of four taxa (*P. canhii*, *P. fairrieanum*, *P. hirsutissimum*, and *P. rungiyasanum*) with unusual morphologies that fall outside the established section/subgenera groupings [13, 14]. For example, *P. canhii* is a newly described species from Vietnam, proposed to the subgeneric status [15, 16]. Given that *Paphiopedilum* is a genus with reticulate evolution and clonal propagation, current phylogenetic study of the genus is rather limited and the effect of these events on the evolution of the chloroplast genome is unknown. *Paphiopedilum* is a suitable system to study the evolution of the chloroplast genome and to test whether whole plastome sequences can resolve the phylogeny of the genus.

In an effort to answer these unresolved questions, we used a genome skimming method to sequence 77 chloroplast genomes of the genus *Paphiopedilum*. We used comparative genomics to study the evolution of the *Paphiopedilum* chloroplast genome and included four *Paphiopedilum* plastome sequences reported in previous studies [8, 10, 17, 18]. We analysed the sequences of all the shared protein coding genes from 81 *Paphiopedilum* samples to study patterns of evolutionary rates of the chloroplast genome. The goals of this study are to 1) reconstruct the phylogeny relationships of the genus, especially the systematic positions of the phylogenetically unstable taxa (*P. canhii*, *P. fairrieanum*, *P. hirsutissimum*, and *P. rungiyasanum*), and test whether the chloroplast genome could resolve the recently diverged taxa of the genus; 2) characterize the chloroplast genome evolution pattern of *Paphiopedilum*; and 3) calculate the substitution rate of the coding genes, evaluate the impact of IR/SSC boundary shift, and test whether the genes transferred to the IR region due to boundary shift have decreased substitution rates.

Results

Plastomes of Paphiopedilum

We obtained 66 full plastome sequences. The other 11 plastome sequences had one or two gaps located in regions of high AT content within the three intergeneric regions (*trnS-trnG*, *trnE-trnT*, and *trnP-psaJ*). The obtained plastid genome sequences were deposited in GenBank (accession Nos. MN587749 – MN587825) (Table S1). The mean coverage depth of the sequenced plastomes was over 3000-fold (Table S1). We included four published plastome sequences, namely, *P. armeniacum* [8], *P. dianthum* [10], *P. malipoense* [18], and *P. niveum* [9], in subsequent analyses, yielding a total of 81 genomes of the genus

Paphiopedilum. The downloaded plastome sequence of *Phragmipedium longifolium* [8] was used as the outgroup.

The genome size of *Paphiopedilum* ranged from 152,130 bp in *P. tigrinum* to 164,092 bp in *P. emersonii* (Table S1). *P. emersonii* had the largest number of genes and one of the shortest SSC regions (660 bp). The plastid genome of the genus shows typical quadripartite structure, with two identical copies of IR separated by a LSC region and an SSC region (Fig. S1). Compared with the plastome of other angiosperms, *Paphiopedilum* has numerous expansions of IRs. The length of the IR region was enlarged to 31,743 bp – 37,043 bp, with the length of the IR region of eight samples even larger than 35 kb (Table S1), while the length of the SSC region contracted to 524 bp – 5916 bp (Fig. 1, Table S1). In addition, the SSC regions of *Paphiopedilum* are hotspots for gene transfer, loss, and rearrangement (Fig. 2–4). The size variation in the SSC region mainly results from the transfer of typical SSC genes to IR regions and the loss/pseudogenization of *ndh* genes. Subg. *Parvisepalum* has a relatively larger SSC region than the other species in the genus (Fig. 1–2, Table S1).

The gene order was conserved and composed of 127 to 134 genes, including 76 to 81 protein coding genes, 38 to 39 tRNA genes, eight rRNAs, three to eight pseudogenes, and 20 to 25 genes duplicated in the IR region (Table S1-S2). In addition to the duplication of the IR regions, *trnG-GCC* duplicated in *P. exul* and *P.* aff. *exul*, while *trnQ-UUG* duplicated in *P. charlesworthii*, *P. tigrinum*, and *P. barbigerum var. lockianum*. The two copies of *trnG-GCC* have one nucleotide variation, and the two copies of *trnQ-UUG* have eight nucleotide variations. Gene density ranged from 0.78 to 0.84, and *P. tigrinum* had the shortest plastome size and the highest gene density (Table S1). The GC content of the plastome genome ranged from 34.7–36.3%, and the GC content of the protein-coding genes ranged from 29.7–46.1%.

In addition, we found 17 genes containing introns, including six tRNA genes (*trnA-UGC*, *trnG-UCC*, *trnL-UAA*, *trnI-GAU*, *trnK-UUU*, and *trnV-UAC*) and 11 protein coding genes (*atpF*, *clpP*, *ndhB*, *petB*, *petD*, *rpl2*, *rpl16*, *rpoC1*, *rps12*, *rps16*, and *ycf3*). Eight of the protein coding genes contain one intron, while three of them (*clpP*, *rps12*, and *ycf3*) contain two introns (Table S2).

The LSC/IR_b boundary is relatively stable. While the LSC/IR_b junction is on *rpl22* in most species (76 of 81 samples), the LSC/IR_b junction is on *rps19* in *P. concolor* and *P. wenshenanse* × *P. bellatulum*, between *rpl22* and *rps19* in *P. rhizomatosum*, and between *rps19* and *trnH-GUG* in *P. hirsutissimum* (Fig. 4). Compared to the LSC/IR boundaries, the IR/SSC boundaries of *Paphiopedilum* varied among species (Fig. 4). Substantial variation in the SSC/IR boundary was mainly in subg. *Parvisepalum*. In most other samples (56 of 81 samples), one end of the SSC/IR junction was located in the intergeneric spacer region *trnL-ccsA*, near *trnL-UAG*, whereas the other junction of SSC/IR was located on the *ccsA* gene (Fig. 2).

The contraction of the SSC region resulted in the typical SSC genes being transferred to the IR region. One to six genes from the SSC region were transferred to the IR region. For example, *ycf1* was transferred to the IR region in all the sequenced samples, while Ψ*ndhD*, *psaC*, and *rps15* were incorporated into the IR

region in most species. The gene *ccsA* expanded in the IR region occasionally, and *trnL-UAG* was transferred to the IR region in *P. delenatii*, *P. dianthum*, and *P. parishii* (Fig. 2).

The genomic comparison demonstrates that the SSC region of *Paphiopedilum* differs greatly in gene content, gene order, and gene orientation (Fig. 2, 4, S2). The SSC regions of most species contain *trnL*, *rpl32*, and partial *ccsA*, while the SSC regions of five species are on the brink of losing, *P. appletonianum*, *P. barbigerum*, *P. emersonii*, *P. hirsutissimum*, and *P. villosum* only contain *trnL-UAG* in this region (Fig. 2). In addition, the genes *psaC* and Ψ*nadD* were preserved in the SSC region in six samples of subg. *Parvisepalum* (Fig. 2). In addition, there might be two copies of SSC with different directions in the same species [19]. Wang and Lanfear [20] used long-read sequencing to test the structural heteroplasmy in land plants and found the presence of chloroplast structural heteroplasmy in most land plant individuals, so the direction of the SSC region was not considered. Based on gene content and gene orientation, the SSC regions were classified into twelve types (Fig. S2), and type □ is the dominant type (56 of 81 samples) (Fig. 2). Type □ and type □ are identical in gene content but differ in the gene direction of *rpl32* and *trnL-UAG* (Fig. S2). Type □ and type □ are also identical in gene content, but in type IX, the two genes run in opposite directions, while in type X, the two genes run in the same direction (Fig. S2). Type □ and type □ both have *trnL-UAG*, but one nucleotide in type □ has shifted to the IR region. Subg. *Parvisepalum* has six types, whereas sect. *Cochlopetalum* has only one type (type □) (Fig. 2). When the SSC types are plotted on the phylogenetic tree, the result shows that the SSC types are not lineage-specific and that even the closely related species have distinct SSC types, such as species in subg. *Brachypetalum* and subg. *Parvisepalum* (Fig. 2). Surprisingly, the gene content of SSC regions has intraspecies variation. For example, one sample of *P. barbigerum* contains *trnL-UAG*, while the other sample contains *trnL-UAG* and *rpl32*. The two samples of *P. appletonianum* also have different SSC types (Fig. 2).

Gene gain and loss in *Paphiopedilum* samples was also analysed. Some of the gene losses are shared throughout the genus (e.g., some *ndh* genes), while other gene losses are lineage specific (Table S3). Most of the *ndh* genes were pseudogenized (Ψ*ndhD*, Ψ*ndhJ*, and Ψ*ndhK*) or lost (*ndhA*, *ndhC*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, and *ndhI*) from the *Paphiopedilum* species plastome, except for *ndhB*. Most of the samples (76 of 81) sequenced in this study retained an intact copy of *ndhB*. In addition, the complete open reading frame of *ndhJ* was preserved in 23 samples, including four samples of sect. *Cochlopetalum* and 19 samples of sect. *Paphiopedilum* (Table S3). The genes *ndhC* and *ndhK* were preserved as pseudogenes in subg. *Parvisepalum* and sect. *Concoloria* but lost in the other species (Table S3). In particular, in five species sequenced (*P. barbatum*, *P. dayanum*, *P. platyphyllum*, *P. sugiyamanum*, and *P. tigrinum*), all 11 of the *ndh* genes were lost or pseudogenized (Fig. 2, Table S3).

In addition to the pseudogenes found in the *ndh* genes, we found that premature termination induced pseudogenization of other protein coding genes (*cemA* and *ycf15*). Most of the annotated copies (38 of 48) of *cemA* are preserved as pseudogenes, while all the annotated copies of *ycf15* were retained as pseudogenes (Table S3). The pseudogenization of *cemA* is mainly due to the slippage of poly structure and the appearance of premature stop codons, while the pseudogenization of *ycf15* is due to the appearance of more than one premature stop codon.

The plastomes of *Paphiopedilum* also show multiple structural rearrangements. We found widespread structural variation in the SSC regions, especially in subg. *Parvisepalum*, including the inversion and recombination of the SSC genes and the shift of the IR/SSC boundary (Fig. 2–4). In particular, we found a 47 kb inversion spanning from *petN* to *clpP* in *P. fairrieanum*, which is absent in other species of *Paphiopedilum* (Figs. 3, S1).

Phylogenetic analyses

Phylogenetic analyses were performed on the three concatenated datasets. The results of ML and BI analyses are almost identical, despite being based on different matrices, so we used the result based on the whole plastomes directly. Relationships among the sections are consistent with previous studies (Fig. 1–2, Fig. S3) [12, 13]. The phylogeny analyses showed subg. *Parvisepalum* at the base of the tree, followed by subg. *Brachypetalum*, and then the five sections in subg. *Paphiopedilum*. The five sections grouped into two clades, with one clade formed by sect. *Cochlopetalum* (sect. *Coryopedilum*, sect. *Pardlopetalum*) and the other clade formed by sect. *Barbata* and sect. *Paphiopedilum*. Many terminal nodes in the tree are still unresolved, especially species in sect. *Paphiopedilum* (Fig. S3-S4).

Additionally, the four enigmatic species of *Paphiopedilum* all nested in subg. *Paphiopedilum*. *P. fairrieanum* clustered with the sect. *Paphiopedilum*-sect. *Barbata* clade with a high support value (BP = 99), and *P. canhii* clustered with the sect. *Cochlopetalum* (sect. *Coryopedilum*, sect. *Pardlopetalum*) clade (BP = 90); meanwhile, *P. rungsuriyanum* and *P. hirsutissimum* formed a clade with a medium support value (BP = 82), and the relationship with other branches is unresolved (Fig. S3). Removing the four taxa results in a tree with the same topology but with elevated support values for some of the branches (Fig. S4).

Nucleotide substitution rate analyses

Mean synonymous and nonsynonymous divergence was low (dN = 0.0471, dS = 0.1222) with the dS value almost three times that of dN (Fig. 5, Table S4). We detected signals of positive selection in *clpP* (dN/dS = 1.6561), *psbH* (dN/dS = 1.0318), and *psbZ* (dN/dS = 1.5768). Our analyses show that both dN and dS had significantly increased in *psbM* (dN = 0.3585, dS = 0.7519) (Fig. 5), which is induced by a frameshift mutation in *P. niveum* (KJ524105). Genes in the SC regions have higher substitution rates than genes in the IR regions. Unexpectedly, genes transferred to the IR region due to boundary shift did not show lower substitution rates (Fig. 5, Table S4).

Discussion

The phylogeny of Paphiopedilum

Our phylogeny based on the whole plastome resolves the relationship among sections, which is largely consistent with the most recent phylogenies obtained from partial cytoplasmic DNA markers (Fig. S3) [12, 13]. The phylogeny resolved monophyly of sect. *Cochlopetalum* with high bootstrap support (BP = 96), while this clade was poorly resolved in previous studies [12, 13]. The four enigmatic species all fall

outside the established sections. *P. fairrieanum* clusters with sect. *Paphiopedilum*-sect. *Barbata* clade with a high support value (BP = 100), *P. canhii* clusters with sect. *Cochlopetalum*-(sect. *Coryopedilum*, sect. *Pardlopetalum*) clade with a medium support value (BP = 90) (Fig. S3). However, the phylogenetic position of the other two enigmatic taxa (*P. hirsutissimum* and *P. rungiyasanum*) clustered together with a medium support value (BP = 82) and formed parallel relationships with *P. fairrieanum*, sect. *Paphiopedilum*, and sect. *Barbata* (Fig. S3). Interestingly, the support values of several clades increased after the removal of these species. For example, the sister relationship between sect. *Paphiopedilum* and sect. *Barbata* increased to 100 and the support value of sect. *Cochlopetalum* increased to 99 (Fig. S4). The uncertain phylogenetic positions of the enigmatic taxa might be related to the heteroplasmy found in the chloroplast genomes (unpublished data). The short branch lengths in sect. *Paphiopedilum* suggest their recent radiation, which means species in this section are in their preliminary stage of speciation. On the other hand, the long branch length represents the differentiation from the closely related species. For example, the long branch of *P. druryi* might be due to its isolated distribution in south India.

The chloroplast genome evolution of Paphiopedilum

The largest (*P. emersonii*, 164,092 bp) and smallest (*P. tigrinum*, 152,130 bp) plastome sizes differ by ~ 12 kb, and the gene number ranges from 127 to 134 genes (Table S1). The genome size and gene number variation are mainly due to the IR expansion and the loss of the *ndh* genes, and the genome size is larger than the average plastome size (151 kb) [21]. Compared with other sequenced chloroplast genomes of photoautotrophic orchid, such as *Dendrobium* [22] and *Holcoglossum* [23], the plastomes of *Paphiopedilum* contain more size variation and structure variation at the genus level (Fig. 1, Table S1). In addition, the GC content of the genus (34.7 − 36.3%) is also more variable and lower at the genus level than in other sequenced genera in Orchidaceae, e.g., *Cymbidium* (36.8 − 37%) [24], *Dendrobium* (37.31 − 37.68%) [22], and *Holcoglossum* (35.3 − 35.5%) [23]. The lower GC content might relate to the extremely lengthy AT-repeat regions in the genus. We found AT repeats more than 100 bp long in some species, especially in the 11 samples with gaps in the LSC regions (Table S1).

The LSC/IR_b boundary of *Paphiopedilum* is relatively stable; LSC/IR_b resided on *rpl22* in most species, while the IR/SSC boundary of the genus is different from those in other species of orchids and variable at the genus level (Fig. 4). Owing to the instability of the IR/SSC boundaries, the chloroplast genome of *Paphiopedilum* experienced IR expansion and SSC contraction. The typical land plant IR is 25 kb [25], while the IR region of *Paphiopedilum* is 32–37 kb, and the IRs account for approximately 40–45% of the plastomes. The increased length in the IR regions results from the IR/SSC boundary shifts.

A large IR expansion (several kilobases) has also been reported in other angiosperm lineages, including *Acacia* and *Inga* [17], *Erodium* [26], *Passiflora* [27, 28], and *Pelargonium* [6, 29]. In *Pelargonium*, IR expanded towards both the LSC and the SSC regions. However, the IR expansions in *Paphiopedilum* mainly included former SSC genes. IR expansion into the SSC region has also been reported in other species, e.g., *Musa acuminate* [30], *Pedicularis ishidoyana* [31], and *Vanilla* [9, 11, 32].

The large IRs of the plastomes are hypothesized to contribute to plastome stabilization because their absence often coincides with severe changes in gene order [33]. IRs are thought to stabilize the plastome through homologous recombination-induced repair mechanisms [34]. However, large IRs have no impact on the plastome stability of the genus. With the shifts in the IR/SSC boundaries, the SSC regions in *Paphiopedilum* have shortened and variable SSC lengths (524–5913 bp). The tightening of the SSC regions is associated with the loss of *ndh* genes and the transfer of former SSC genes to the IR region. For example, the 524 bp SSC region of *P. hirsutissimum* is smaller than other sequenced plastomes in Orchidaceae. The strong shrinking of SSC regions was also documented in other unlinked photoautotrophic plants, such as *Annona cherimola* (2966 bp) [35], *Asarum* (0–14 bp) [36], *Lamprocapnos spectabilis* (1645 bp) [37], *P. ishidoyana* (27 bp) [31], *Pelargonium* (c. 6.7 kb) [6], and *Vanilla* (c. 2 kb) [9, 11, 32]. Interestingly, the other genera that experienced SSC contraction have a relatively stable SSC length at the genus level, e.g., all the sequenced plastomes of *Pelargonium* shared a 6.7 kb SSC region [6], while the length of the SSC region of *Paphiopedilum* is more variable. The SSC-contracted plastomes of *Paphiopedilum* have unique features compared with other tightened SSC lineages, including gene content variation, length variation, and gene rearrangement (Fig. 2–4), according to the sequenced chloroplast genomes of other genera of slipper orchids [8, 9, 38], suggesting that SSC contraction occurred after *Paphiopedilum* split from its sister clade.

In addition to size variation, the SSC region of *Paphiopedilum* experienced extensive gene rearrangement (Fig. 3). The SSC regions of most species in sect. *Concoloria*, sect. *Cochlopetalum*, sect. *Coryopedilum*, sect. *Paphiopedilum*, and sect. *Barbata* retain *trnL-UAG*, *rpl32* and a truncated *ccsA* fragment (*trnL-UAG*(+)-*rpl32*(+)-*ccsA*-P(+), type □, approximately 1.8 kb). In the remaining species, the gene content, gene order, and length of the region are highly variable (524–5913 bp) (Fig. 1–2, S2; Table S1). In other angiosperm, there are two gene clusters in the SSC region, namely, *rpl32*(+)-*trnL-UAG*(+)-*ccsA*(+) and *ndhD*(−)-*psaC*(−)-*rps15*(−). In Apostasioideae [39], *Cypripedium* [9, 38], *Cymbidium* [24], and *Phalaenopsis aphrodite* [40], the genes on the two clusters are in opposite orientations. However, the original *rpl32*(+)-*trnL-UAG*(+)-*ccsA* (+) linkage was only preserved in *P. armeniacum* [8]. In other species of the genus, the above gene linkages were broken and changed to *trnL-UAG*(+)-*rpl32*(+)-*ccsA*(+) and *trnL-UAG*(−)-*rpl32*(−)-*ccsA*(−) in most cases, which means that there is inversion and recombination in the SSC region. Gene rearrangement bring *rpl32* and *ccsA* (which are normally separated in the plastome) next to each other (Fig. 2, 4). In addition, we found a 47 kb inversion in the LSC region of *P. fairieanum* (Fig. 3, S1c). A 61 kb inversion in the LSC region was previously reported in *Cypripedium formosanum* [9]. But the mechanisms for these inversions are unknown.

Complete gene duplications (of *rpl32*, *trnL-UAG*, *ccsA*, Ψ*ndhD*, *psaC*, *rps15*, and *ycf1*) were documented due to the expansion of the IR region. Additionally, gene duplication outside of the IR regions was also documented. We found that *trnG-GCC* was duplicated in the LSC in two species, while *trnQ-UUG* was duplicated in the LSC in three species, and the two duplication events both occurred in sect. *Paphiopedilum*. The second copies of *trnG-GCC* and *trnQ-UUG* are identical to sequences from other orchids, indicating that the two copies might be horizontally transferred, especially *trnQ-UUG*. The two

copies of the gene have eight nucleotide variations. The duplication of *trnQ-UUG* has also been documented in other studies [41, 42].

Most of the *ndh* genes were pseudogenized or lost from the genus, but the functional *ndhJ* and *ndhB* were preserved in some lineages (Table S3). Considering that *ndhA*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, and *ndhI* were also lost in *P. longifolium* [8] whereas they were preserved in *C. formosanum* [9] and *C. japonicum* [38], we infer that the loss of these genes occurred in the ancestors of the conduplicate genera. The pseudo copies of *ndhK* and *ndhC* were preserved in most samples of subg. *Parvisepalum* and subg. *Brachypetalum* (Table S3) and lost in subg. *Paphiopeilum*, suggesting that *ndhK* and *ndhC* were lost in the ancestor of subg. *Paphiopeilum*. Considering the retention of *ndhJ* and *ndhB* in the genus (Table S3), we infer that the loss/pseudogenization of *ndhB* and *ndhJ* probably occurred rather recently because they are repeatedly retained in the terminal taxa.

Furthermore, four of the five species with all 11 *ndh* genes pseudogenized or lost are distributed on islands, and the other one (*P. tigrinum*) is distributed on the mainland. The five species are all photoautotrophic plants. In addition, the five species belong to three sections (Fig. 2) and are nested in different clades, suggesting that there are four independent *ndh* loss/pseudogenization events at the genus level. The above evidence provided clues that the *ndh* complex is undergoing degradation in this genus, and the degradation of the *ndh* genes is not lineage specific, suggesting a very recent pseudogenization of some *ndh* genes. The independent degradation of *ndh* genes was also found in *Cymbidium* [43]. In addition, the loss of the full set of plastid *ndh* genes in other photoautotrophic orchids, such as *Erycina pusilla* [44], *Holcoglossum* [23], *P. aphrodite* [40], and *Vanilla planifolia* [9], has also been reported in previous studies. Lin *et al*. [45] even proposed that the loss of *ndh* complexes probably increased the transition of the life history from photoautotrophic to heterotrophic.

Kim *et al*. [8] proposed that the *ndhF* gene strongly correlated with the IR/SSC junction stability, whereas the result of Niu *et al*.'s [11] study indicated that the *ndh* genes strongly related to the IR/SSC junction stability. Though the IR/SSC boundary is different from other orchids but relatively stable in most *Paphiopedilum* species, we infer that *ndhF* or *ndh* is not correlated with the IR/SSC junction stability at the genus level. The *ndhF* gene was lost in all the sequenced species of *Paphiopedilum*, and the five species that lost all 11 functional *ndh* genes have stable IR/SSC junctions (Fig. 2). We infer that there are other underlying mechanisms related to the structure variations in the genus.

The substitution rates of the chloroplast genes

The substitution rate of the genus is low (dN = 0.0471, dS = 0.1222) and varied among genes, with the dS value almost three times that of dN (Fig. 5, Table S4). We detected signals of positive selection in *clpP* (dN/dS = 1.6561), *psbH* (dN/dS = 1.0318), and *psbZ* (dN/dS = 1.5768). In several angiosperm lineages, such as *Acacia* and *Inga* [17], *Passiflora* [28], and *Silene* [46], positive selection of *clpP*, which is involved in protein metabolism, has been identified. Two other positively selected genes are involved in photosynthesis: *psbH* was under positive selection in the shade tolerant *Oryza* [47], and psbZ showed positive selection in *Rhododendron pulchrum* [48].

Previous studies have revealed that genes in the IR region have lower substitution rates than genes in the SC region [29, 49, 50]. Genes in the IR region also have relatively lower substitution rates compared to SC region genes in orchids [24]. Unexpectedly, *ycf1* moved to the IR region but still had a high substitution rate (dN = 0.0821, dS = 0.1140), and its substitution rate is higher than that the other genes in the IR region. Even *psaC* (dN = 0, dS = 0.0825) and *rps15* (dN = 0.0388, dS = 0.1183), which transferred to the IR region, have higher divergence than the other genes in the IR region. Furthermore, *ccsA* (dN = 0.0932, dS = 0.2543) and *rpl32* (dN = 0.1222, dS = 0.3193) were mostly retained in the SSC region and have higher divergence than other genes in the SSC region (Fig. 5). Weng *et al*. [6] found similar phenomena in *Pelargonium*, where the expanded IR genes in plastid genomes did not have lower substitution rates. Interestingly, the divergence of photosynthesis-related genes is relatively lower than the divergence of other genes (Fig. 5).

Conclusion

Analysing the chloroplast genome is a key way to study the molecular evolution of orchids. However, previously published orchid studies mainly focused on heterotrophic species and the subfamily Epidendroideae. We used comparative chloroplast genome sequences to reveal the evolutionary history of *Paphiopedilum*. The genome size of *Paphiopedilum* is slightly larger than that of other angiosperm, and all the species in *Paphiopedilum* experienced IR expansion and SSC contraction. The expansion of the IR region ranged from ~ 7 kb to ~ 12 kb, and the IR expansion in *Paphiopedilum* is associated with the transfer of former SSC genes to the IR region. The comparison of plastomes showed that IR expansion and *ndh* loss contributed to the size variation in the genus. Because of the variation in the IR/SSC boundaries, the SSC region of *Paphiopedilum* chloroplast genomes showed massive variation in length and gene content. *Paphiopedilum* provides the best opportunity to study the dynamics of chloroplast genome evolution and to test the rate of heterogeneity related to the variation in the IR/SSC boundaries. Though genome rearrangement is mainly reported in the mitochondrial genome (e.g. [51]), this study provides evidence of highly active genome rearrangements in the SSC regions of *Paphiopedilum* with dense sampling. The *ndh* genes of *Paphiopedilum* experienced varying degrees of degradation, and five species have lost/pseudogenized all 11 *ndh* genes. However, the mechanisms underlying the loss/pseudogenization of these genes are unknown. This study sheds light on the tempo and mode of evolutionary changes that occurred in the chloroplast genomes across *Paphiopedilum* and provides a good example of how to study the chloroplast genome evolution of Orchidaceae.

Methods

Taxon sampling and library construction

Plant materials sequenced in this study were collected from the National Orchid Conservation & Research Center of Shenzhen (NOCC) (Table S1), including four species downloaded from GenBank. A total of 81 samples representing 63 species and covering seven major clades of *Paphiopedilum*, four natural hybrids

and one man-made hybrid were also included. Based on previous phylogenetic studies, *P. longifolium* was chosen as the outgroup.

Total genomic DNA was extracted from fresh leaves using the CTAB method [52]. Paired-end libraries with 350 bp inserts were constructed for this study, and sequencing was performed at Novogene (Beijing, China) on the Hiseq/Miseq platform (150 bp).

Sequence assembling and annotation

High-quality filtered paired-end reads were *de novo* assembled into contigs using CLC Genomics Workbench v.10.1.1 (with different word sizes and bubble sizes) (https://www.qiagenbioinformatics.com/) or NOVOPlasty3.5 (K-mer = 39, *rbcL* from *P. armeniacum* was used as the seed sequence) [53]. Contigs were merged in Geneious v.11.1.2 (Biomatters, Inc.) to build draft plastomes. We then mapped reads to draft plastomes in Geneious v.11.1.2 to check the ambiguous regions. Annotation of the plastomes was performed in DOGMA [54], coupled with manual corrections in Geneious v.11.1.2 (Biomatters, Inc.).

Plastome maps were generated with OrganellarGenomeDRAW [55]. The boundaries of IR and SC regions were defined by REPuter [56] with default settings. We calculated GC content in Geneious v.11.1.2. Then, changes in genome structure were visualized using progressive Mauve [57] with IRa removed. Protein coding regions (CDS), rRNA, tRNA, intergenic spacers (IGS), and intron sequences were extracted and aligned to generate original CDS, rRNA, tRNA, IGS, and intron alignments.

Phylogenetic analysis

For phylogenetic analyses, 68 protein-coding gene sequences (*ndhB* included), four rRNA gene sequences, 30 tRNA gene sequences, and 87 intergenic spacer regions were aligned with Muscle (Table S2, S5) [58] and refined manually. Then, they were concatenated into different datasets. We obtained three concatenated datasets: 1) protein-coding sequences (CDS) and rDNA; 2) intergenic spacer (IGS), intron, and tRNA; and 3) whole plastomes (CDS + IGS + intron + rRNA + tRNA). The poorly aligned regions were removed with Gblocks [59] with the default settings.

Maximum likelihood (ML) trees were inferred from RAXml 8.2.4 [60] under the GTRCAT model with 1000 bootstrap replicates. Bayesian inferences (BI) were performed with MrBayes [61]. The evolutionary models of the BI method were calculated in MrModeltest v2 [62], and GTR + I + G was selected as the best model under the AIC. For the Bayesian inference, one cold and three incrementally heated Markov chain Monte Carlo (MCMC) chains were run for 10,000,000 cycles and repeated twice to avoid spurious results. One tree per 1000 generations was sampled, with a burn-in of the first 300 samples for each run. ML and BI analyses were performed on the CIPRES Science Gateway platform [63].

Nucleotide substitution rate analyses

We used the CODEML programme in PAML v. 4.9 (model = 0) [64] to calculate the average nonsynonymous substitution rate (dN) and synonymous substitution rate (dS) for 67 protein coding

genes by the F3X4 codon model. Gapped regions were excluded for rate estimation (cleandata = 1). The ML tree inferred from whole plastome was used as the input tree.

Abbreviations

AIC: Akaike information criterion; AT: Adenosine Thymine; BI: Bayesian inference; BP: bootstrap probability; bp: Base pair; BS: Branch support; CDS: Protein-coding sequences; cp: Chloroplast; CTAB: Cetyltrimethylammonium bromide; dN: Non-synonymous; dS: synonymous mutation; GC: Guanine-cytosine; GTR: General time reversible; IGS: Intergenic sequences; IR: Inverted repeat; IR_a: Inverted repeat A; IR_b: Inverted repeat B; IRs: Inverted repeat regions; kb: kilobase; LSC: Large single copy; ML: Maximum likelihood; rRNAs: Ribosomal RNAs; SC: single copy; SSC: Small single copy; tRNAs: Transfer RNAs;

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All plastomes generated in this study are deposited in NCBI database (https://www.ncbi.nlm.nih.gov/) (GenBank accession Nos. MN587749 – MN587825, see Table S1), and all the datasets supported the conclusion are available at Dryad Digital Repository. These data will remain private until the related manuscript has been accepted. All other data generated in this manuscript are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by grants from National Natural Science Foundation of China (No. U1804117); National Key R&D Program of China (Nos. 2019YFD1000400, 2018YFD1000401, and 2019YFD1001000); and Key scientific research projects of Henan Province (No. 17A180023). The funders had no role in the

design of the study, analysis of the data nor in writing the manuscript.

Author contributions

YYG, ZJL conceived and designed the study. YYG, JXY, and GQZ analyzed the data. YYG wrote the paper. All authors have read and approved the final manuscript.

Acknowledgements

The authors thank Mr Wen-Hui Rao, Mr Jie Huang and Miss Xin-Yi Wu for sample collection.

References

1. Wilkinson MJ, Szabo C, Ford CS, Yarom Y, Croxford AE, Camp A et al. Replacing Sanger with Next Generation Sequencing to improve coverage and quality of reference DNA barcodes for plants. Sci Rep. 2017; 7:46040.
2. Jansen RK, Ruhlman TA. Plastid genomes of seed plants. In: Bock R, Knoop V. Editors. *Genomics of Chloroplasts and Mitochondria.* New York: Springer; 2012. p. 103–126.
3. Tonti-Filippini J, Nevill PG, Dixon K, Small I. What can we do with 1000 plastid genomes? Plant J. 2017; 90(4):808–18.
4. Barrett CF, Sinn BT, Kennedy AH. Unprecedented parallel photosynthetic losses in a heterotrophic orchid genus. Mol Biol Evol. 2019; 36(9):1884–901.
5. Barrett CF, Wicke S, Sass C. Dense infraspecific sampling reveals rapid and independent trajectories of plastome degradation in a heterotrophic orchid complex. New Phytol. 2018; 218(3):1192–204.
6. Weng ML, Ruhlman TA, Jansen RK. Expansion of inverted repeat does not decrease substitution rates in *Pelargonium* plastid genomes. New Phytol. 2017; 214(2):842–51.
7. Mohanta TK, Khan A, Khan A, Abd_Allah EF, Al-Harrasi A. Gene loss and evolution of the plastome. BioRxiv. 2019, 10.1101/676304:676304.
8. Kim HT, Kim JS, Moore MJ, Neubig KM, Williams NH, Whitten WM et al. 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. 2015; 10(11):e0142215.
9. Lin C-S, Chen JJ, 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.
10. Hou N, Wang G, Zhu Y, Wang L, Xu J. The complete chloroplast genome of the rare and endangered herb *Paphiopedilum dianthum* (Asparagales: Orchidaceae). Conserv Genet Resour. 2018; 10(4):709–12.

11. Niu Z, Xue Q, Zhu S, Sun J, Liu W, Ding X. The complete plastome sequences of four orchid species: insights into the evolution of the Orchidaceae and the utility of plastomic mutational hotspots. Front Plant Sci. 2017; 8:715.
12. Chochai A, Leitch IJ, Ingrouille MJ, Fay MF. Molecular phylogenetics of *Paphiopedilum* (Cypripedioideae; Orchidaceae) based on nuclear ribosomal ITS and plastid sequences. Bot J Linn Soc. 2012; 170(2):176–96.
13. Guo Y-Y, Luo Y-B, Liu Z-J, Wang X-Q. Reticulate evolution and sea-level fluctuations together drove species diversification of slipper orchids (*Paphiopedilum*) in Southeast Asia. Mol Ecol. 2015; 24(11):2838–55.
14. Yap JW: Molecular and Genome Evolution in the Malesian Slipper Orchids (*Paphiopedilum* section *Barbata*). London: Queen Mary University of London; 2016.
15. Bream GJ, Gruss O. *Paphiopedilum* subgenus *Megastaminodium* Braem & Gruss, a new subgenus to accommodate *Paphiopedilum canhii*. Orchid Digest. 2011; 75(3):164–5.
16. Górniak M, Szlachetko DL, Kowalkowska AK, Bohdanowicz J, Canh CX. Taxonomic placement of *Paphiopedilum canhii* (Cypripedioideae; Orchidaceae) based on cytological, molecular and micromorphological evidence. Mol Phylogenet Evol. 2014; 70(1):429–41.
17. Dugas DV, Hernandez D, Koenen EJ, Schwarz E, Straub S, Hughes CE et al. Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions, and accelerated rate of evolution in *clpP*. Sci Rep. 2015; 5:16958.
18. Li L-Q, Huang J, Chen L-J, Zhang Q-Y, Chen J-B. The complete chloroplast genome of *Paphiopedilum malipoense* (Orchidaceae). Mitochondrial DNA B. 2019; 4(2):2617–8.
19. Palmer JD, Nugent JM, Herbon LA. Unusual structure of geranium chloroplast DNA: A triple-sized inverted repeat, extensive gene duplications, multiple inversions, and two repeat families. Proc Natl Acad Sci USA. 1987; 84(3):769–73.
20. Wang W, Lanfear R. Stable and widespread structural heteroplasmy in chloroplast genomes revealed by a new long-read quantification method. BioRxiv. 2019, 10.1101/692798:692798.
21. Ruhlman TA, Jansen RK. Aberration or analogy? The atypical plastomes of Geraniaceae. In: Chaw S-M, Jansen RK editors. *Advances in Botanical Research.* vol. 85: Elsevier; 2018. p. 223–262.
22. Niu Z, Zhu S, Pan J, Li L, Sun J, Ding X. Comparative analysis of *Dendrobium* plastomes and utility of plastomic mutational hotspots. Sci Rep. 2017; 7:2073.
23. Li Z-H, Ma X, Wang D-Y, Li Y-X, Wang C-W, Jin X-H. Evolution of plastid genomes of *Holcoglossum* (Orchidaceae) with recent radiation. BMC Evol Biol. 2019; 19(1):63.
24. Yang J-B, Tang M, Li H-T, Zhang Z-R, Li D-Z. Complete chloroplast genome of the genus *Cymbidium*: lights into the species identification, phylogenetic implications and population genetic analyses. BMC Evol Biol. 2013; 13(1):84.
25. Ruhlman TA, Jansen RK. The plastid genomes of flowering plants. In: Maliga P. Editors. *Chloroplast Biotechnology: Methods and Protocols.* New York: Humana Press; 2014. p. 3–38.

26. Blazier JC, Jansen RK, Mower JP, Govindu M, Zhang J, Weng M-L et al. Variable presence of the inverted repeat and plastome stability in *Erodium*. Ann Bot. 2016; 117(7):1209–20.
27. Rabah SO, Shrestha B, Hajrah NH, Sabir MJ, Alharby HF, Sabir MJ et al. *Passiflora* plastome sequencing reveals widespread genomic rearrangements. J Syst Evol. 2019; 57(1):1–14.
28. Shrestha B, Weng M-L, Theriot EC, Gilbert LE, Ruhlman TA, Krosnick SE et al. Highly accelerated rates of genomic rearrangements and nucleotide substitutions in plastid genomes of *Passiflora* subgenus *Decaloba*. Mol Phylogenet Evol. 2019; 138(9):53–64.
29. Zhu A, Guo W, Gupta S, Fan W, Mower JP. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. New Phytol. 2016; 209(4):1747–56
30. Martin G, Baurens F-C, Cardi C, Aury J-M, D'Hont A. The complete chloroplast genome of banana (*Musa acuminata*, Zingiberales): insight into plastid Monocotyledon evolution. PLoS ONE. 2013; 8(6):e67350.
31. Cho W-B, Choi B-H, Kim J-H, Lee D-H, Lee J-H. Complete plastome sequencing reveals an extremely diminished SSC region in hemiparasitic *Pedicularis ishidoyana* (Orobanchaceae). Ann Bot Fenn. 2018; 55(1–3):171–83.
32. Amiryousefi A, Hyvönen J, Poczai P. The plastid genome of Vanillon (*Vanilla pompona*, Orchidaceae). Mitochondrial DNA B. 2017; 2(2):689–91.
33. Palmer JD, Thompson WF. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. Cell. 1982; 29(2):537–50.
34. Wicke S, Schneeweiss G, dePamphilis C, Müller K, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 2011; 76(3):273–97.
35. Blazier JC, Ruhlman TA, Weng M-L, Rehman SK, Sabir JS, Jansen RK. Divergence of RNA polymerase α subunits in angiosperm plastid genomes is mediated by genomic rearrangement. Sci Rep. 2016; 6:24595.
36. Sinn BT, Sedmak DD, Kelly LM, Freudenstein JV. Total duplication of the small single copy region in the angiosperm plastome: rearrangement and inverted repeat instability in *Asarum*. Am J Bot. 2018; 105(1):71–84.
37. Park S, An B, Park S. Reconfiguration of the plastid genome in *Lamprocapnos spectabilis*: IR boundary shifting, inversion, and intraspecific variation. Sci Rep. 2018; 8(1):13568.
38. Kim JS, Kim HT, Kim J-H. The largest plastid genome of monocots: a novel genome type containing AT residue repeats in the slipper orchid *Cypripedium japonicum*. Plant Mol Biol Rep. 2015; 33(5):1210–20.
39. Niu Z, Pan J, Zhu S, Li L, Xue Q, Liu W et al. Comparative analysis of the complete plastomes of *Apostasia wallichii* and *Neuwiedia singapureana* (Apostasioideae) reveals different evolutionary dynamics of IR/SSC boundary among photosynthetic orchids. Front Plant Sci. 2017; 8:1713.
40. 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(2):279–91.

41. Guo W, Grewe F, Cobo-Clark A, Fan W, Duan Z, Adams RP et al. Predominant and substoichiometric isomers of the plastid genome coexist within *Juniperus* plants and have shifted multiple times during Cupressophyte evolution. Genome Biol Evol. 2014; 6(3):580–90.
42. 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(4):688–98.
43. Kim HT, Chase MW. Independent degradation in genes of the plastid *ndh* gene family in species of the orchid genus *Cymbidium* (Orchidaceae; Epidendroideae). PLoS ONE. 2017; 12(11):e0187318.
44. Pan I-C, Liao D-C, Wu F-H, Daniell H, Singh ND, Chang C et al. Complete chloroplast genome sequence of an orchid model plant candidate: *Erycina pusilla* apply in tropical *Oncidium* breeding. PLoS ONE. 2012; 7(4):e34738.
45. 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(5):994–1006.
46. Erixon P, Oxelman B. Whole-gene positive selection, elevated synonymous substitution rates, duplication, and indel evolution of the chloroplast *clpP1* gene. PLoS ONE. 2008; 3(1):e0001386.
47. Gao L-Z, Liu Y-L, Zhang D, Li W, Gao J, Liu Y et al. Evolution of *Oryza* chloroplast genomes promoted adaptation to diverse ecological habitats. Commun Biol. 2019; 2(1):278.
48. Shen J, Li X, Zhu X, Huang X, Jin S. The Complete Plastid Genome of *Rhododendron pulchrum* and Comparative Genetic Analysis of Ericaceae Species. Forests. 2020; 11(2):158.
49. Wolfe KH, Li WH, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA. 1987; 84(24):9054–8.
50. Perry AS, Wolfe KH. Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. J Mol Evol. 2002; 55(5):501–8.
51. Cole LW, Guo W, Mower JP, Palmer JD. High and variable rates of repeat-mediated mitochondrial genome rearrangement in a genus of plants. Molecular Biology and Evolution. 2018; 35(11):2773–85.
52. Doyle J, Doyle J. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 1987; 19(1):11–5.
53. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 2017; 45(4):e18.
54. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. Bioinformatics. 2004; 20(17):3252–5.
55. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3. 1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019; 47(W1):W59–64.

56. 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(22):4633–42.
57. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS ONE. 2010; 5(6):e0011147.
58. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinform. 2004; 5(1):113.
59. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol. 2007; 56(4):564–77.
60. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006; 22(21):2688–90.
61. Ronquist F, Teslenko M, Mark P, Ayres DL, Darling A, Hohna S. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012; 61(3):539–42.
62. Nylander JAA. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden. 2004.
63. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments Workshop (GCE): 14 Nov. 2010 2010; New Orleans, LA*. 1–8.
64. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007; 24(8):1586–91.

Figures

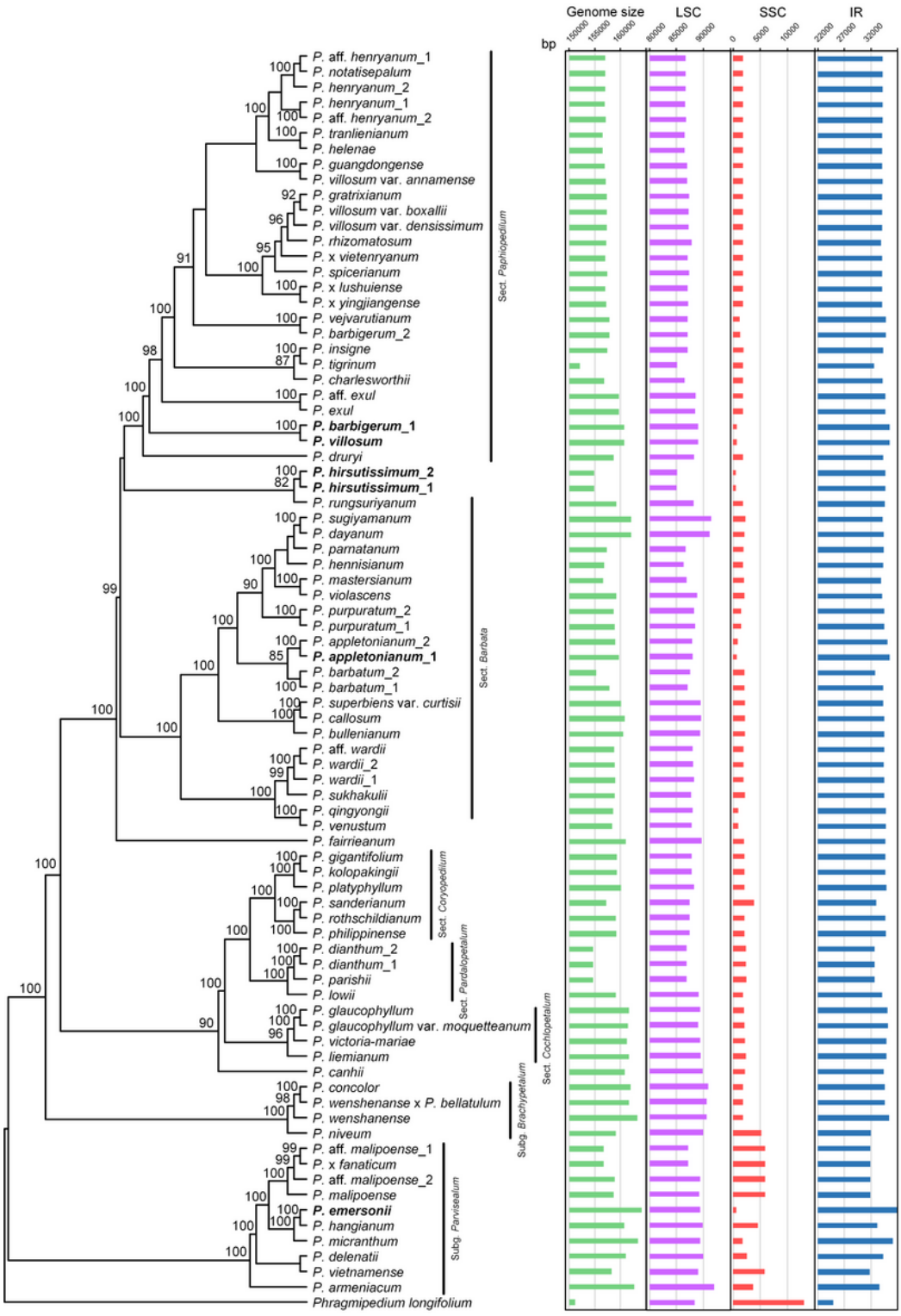

Figure 1

The length of plastid genome size, LSC, SSC, and IR regions were plotted on the ML tree of Paphiopedilum. LSC, SSC, and IR regions were scaled differently. The six samples in bold only have trnL-UAG preserved in the SSC region.

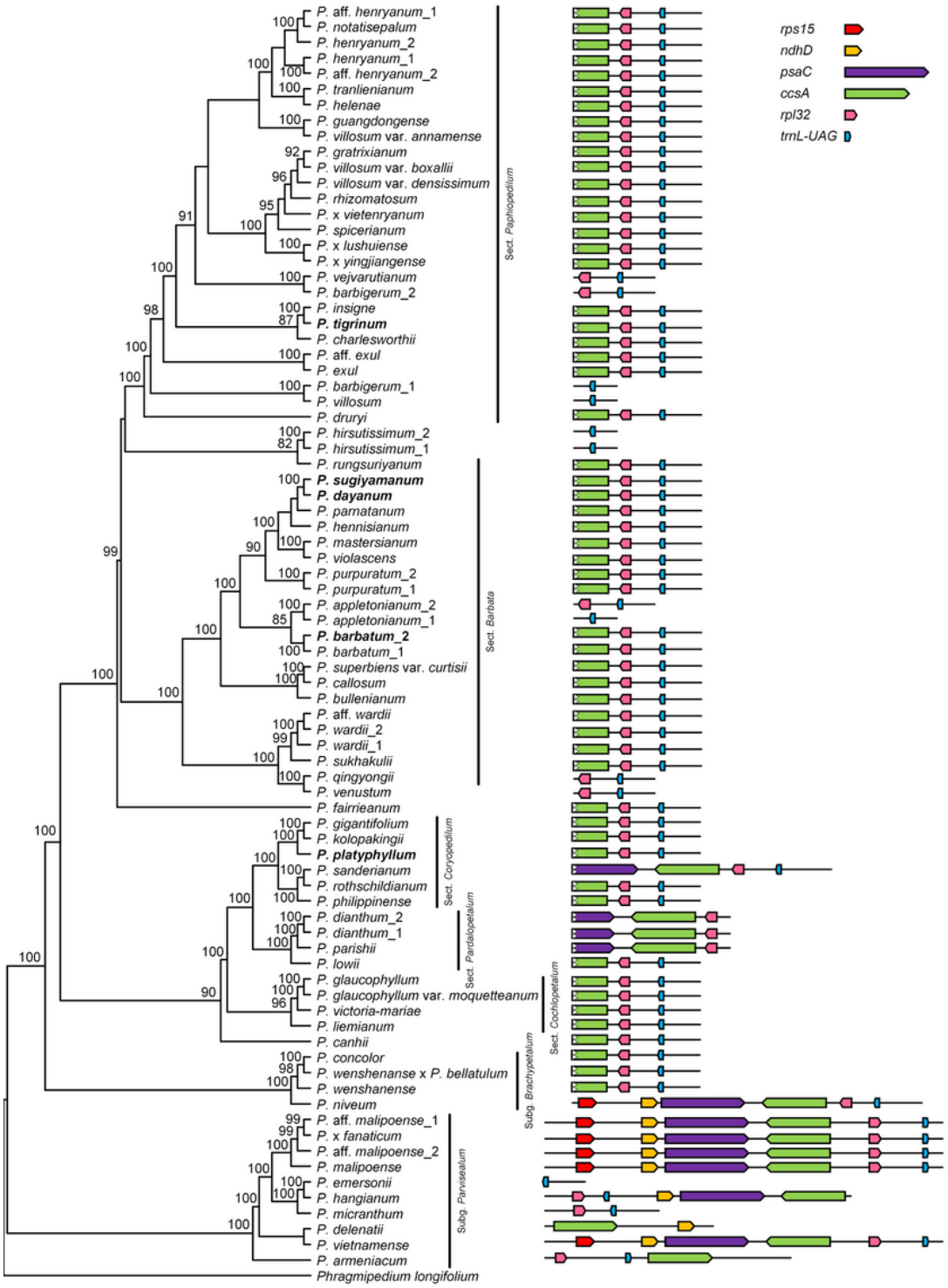

Figure 2

The variation in SSC regions of Paphiopedilum. The scaled representations of SSC types are plotted on the ML tree. The five species in bold lost all the functional ndh genes.

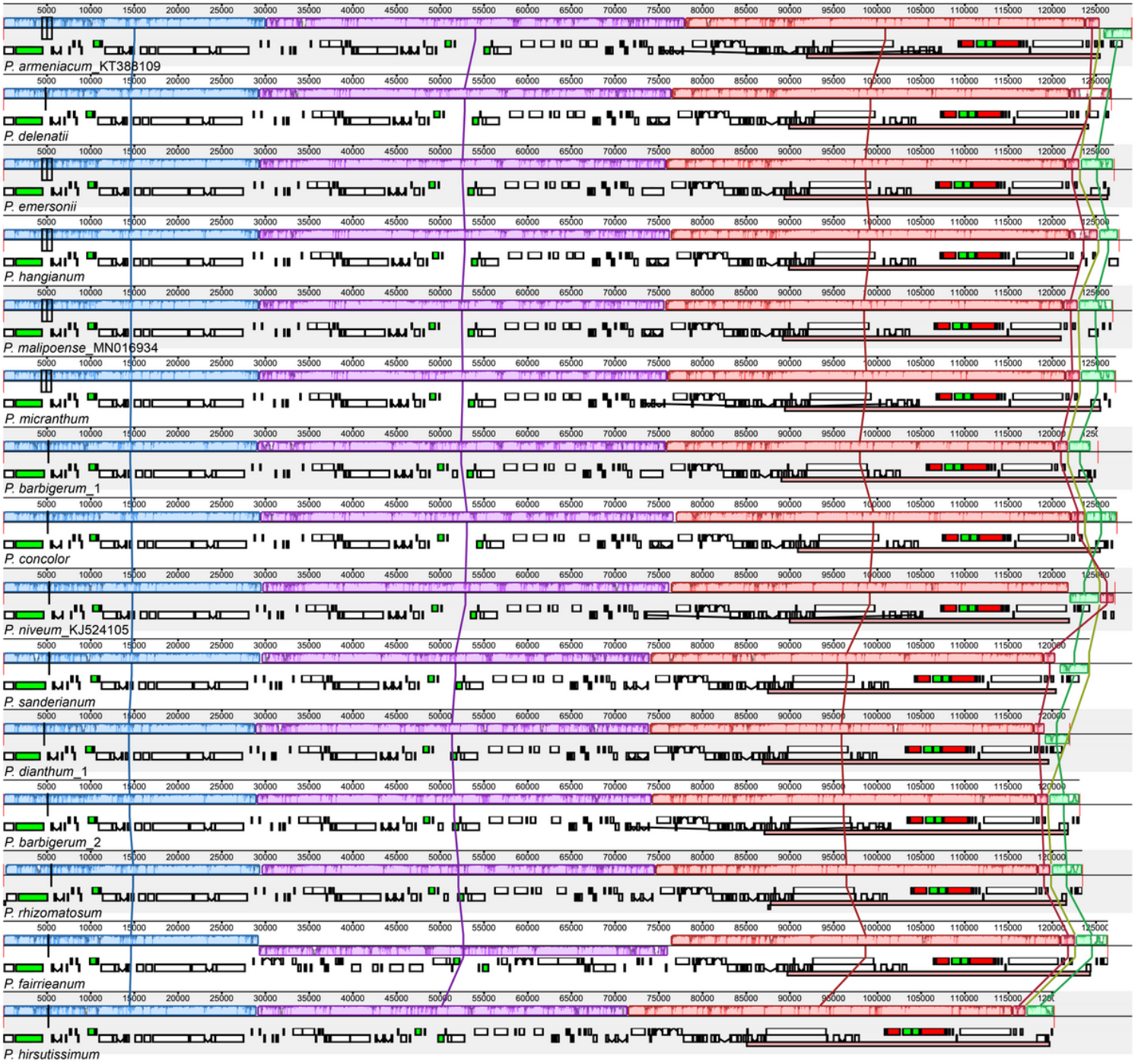

Figure 3

Extent of the gene rearrangements of 15 Paphiopedilum chloroplast genomes. Locally collinear blocks of the sequences are colour-coded and connected by lines.

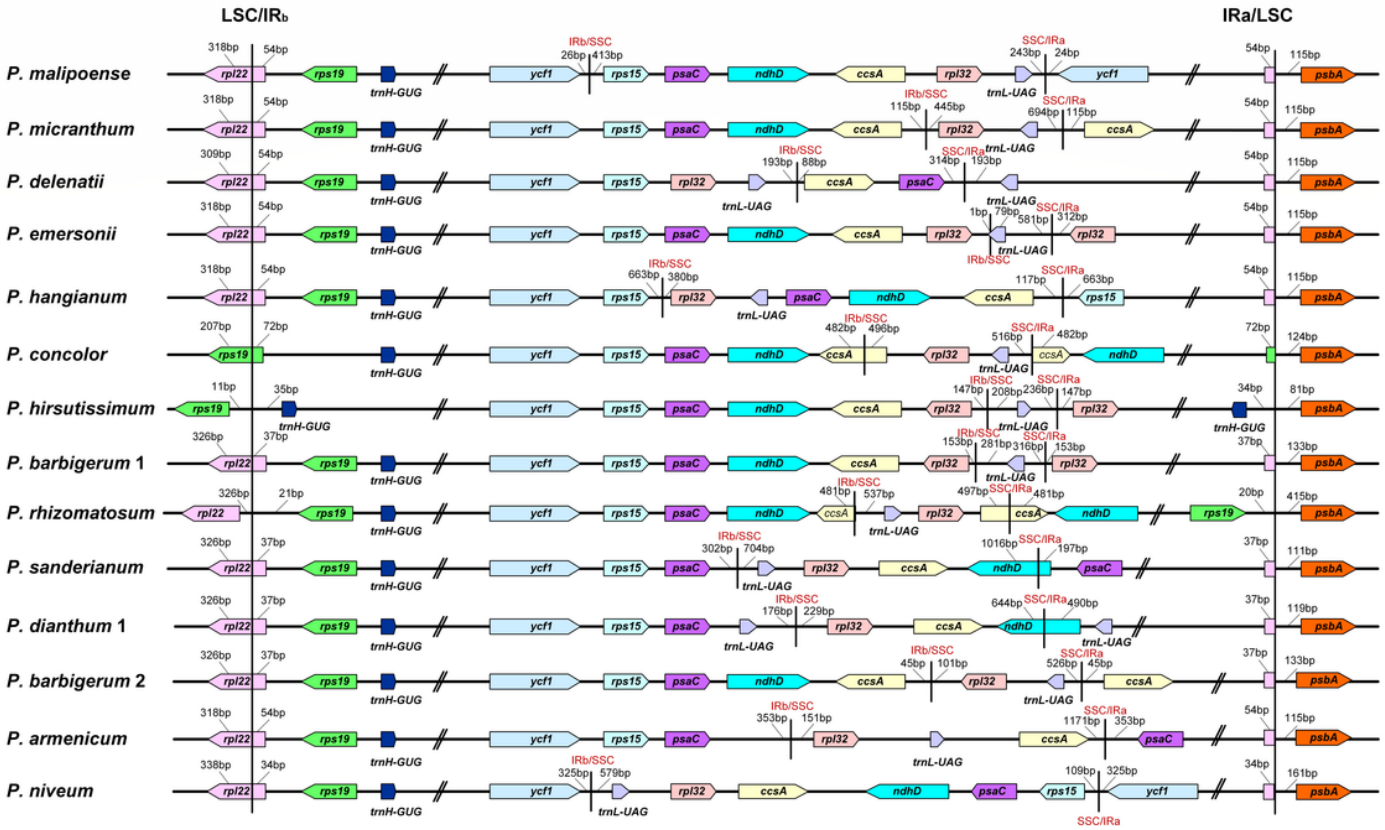

Figure 4

Shift of the LSC/IR boundaries and IR/SSC boundaries of Paphiopedilum.

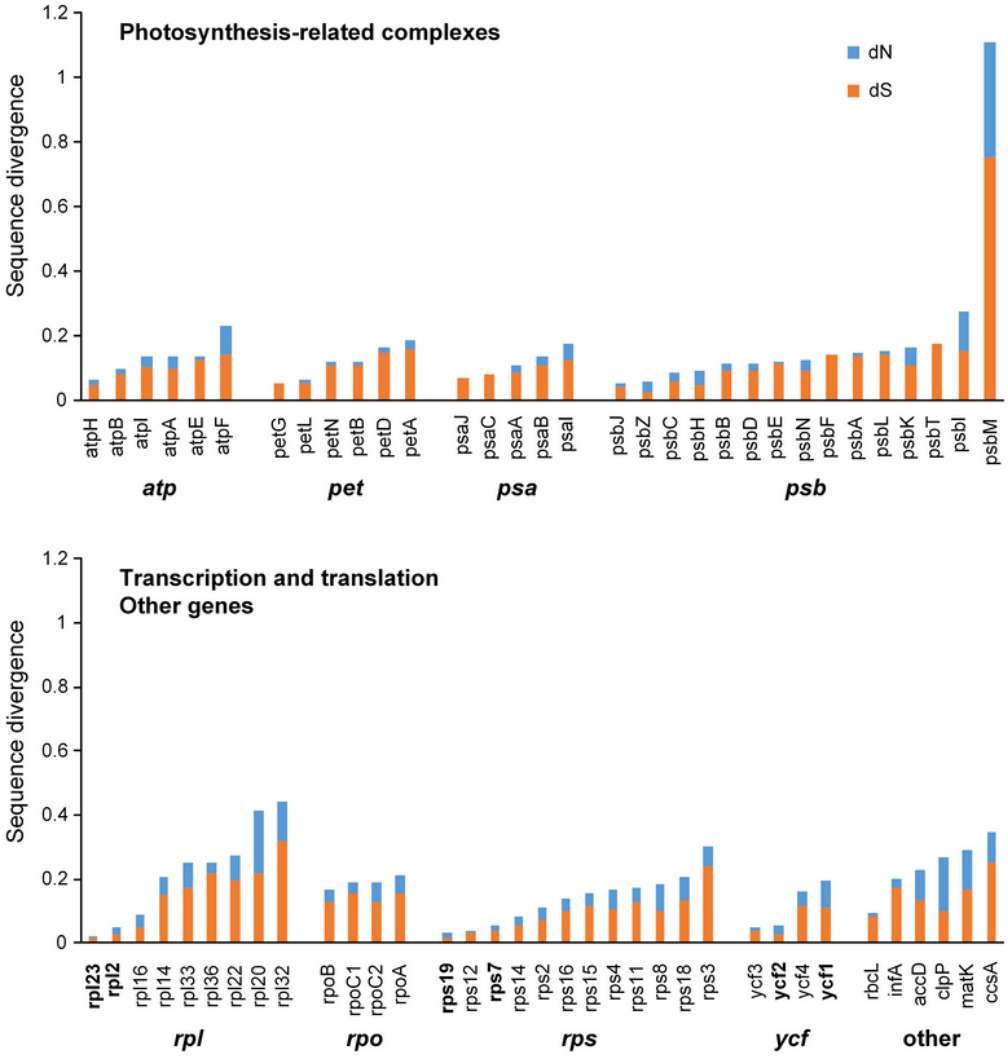

Figure 5

Synonymous (dN) and nonsynonymous (dS) substitution rates of Paphiopedilum chloroplast genes grouped by complex and function. Gene names blackened are genes in the IR region.

Supplementary Files

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

- TableS1S5.pdf
- Fig.S1S4.pdf