Comparative Analysis of the Complete Plastomes of *Apostasia wallichii* and *Neuwiedia singapureana* (Apostasioideae) Reveals Different Evolutionary Dynamics of IR/SSC Boundary among Photosynthetic Orchids

Zhitao Niu, Jiajia Pan, Shuying Zhu, Ludan Li, Qingyun Xue, Wei Liu and Xiaoyu Ding*

College of Life Sciences, Nanjing Normal University, Nanjing, China

Apostasioideae, consists of only two genera, *Apostasia* and *Neuwiedia*, which are mainly distributed in Southeast Asia and northern Australia. The floral structure, taxonomy, biogeography, and genome variation of Apostasioideae have been intensively studied. However, detailed analyses of plastome composition and structure and comparisons with those of other orchid subfamilies have not yet been conducted. Here, the complete plastome sequences of *Apostasia wallichii* and *Neuwiedia singapureana* were sequenced and compared with 43 previously published photosynthetic orchid plastomes to characterize the plastome structure and evolution in the orchids.

Unlike many orchid plastomes (e.g., *Paphiopedilum* and *Vanilla*), the plastomes of Apostasioideae contain a full set of 11 functional NADH dehydrogenase (*ndh*) genes. The distribution of repeat sequences and simple sequence repeat elements enhanced the view that the mutation rate of non-coding regions was higher than that of coding regions. The 10 loci—*ndhA* intron, *matK*-5′*tmK*, *clpP*-psbB, *rps8*-rpl14, *tmT-tmL*, 3′*tmK*-matK, *clpP* intron, *psbK-tmK*, *tmS-psbC*, and *ndhF*-rpl32—that had the highest degrees of sequence variability were identified as mutational hotspots for the *Apostasia* plastome. Furthermore, our results revealed that plastid genes exhibited a variable evolution rate within and among different orchid genus. Considering the diversified evolution of both coding and non-coding regions, we suggested that the plastome-wide evolution of orchid species was disproportional. Additionally, the sequences flanking the inverted repeat/small single copy (IR/SSC) junctions of photosynthetic orchid plastomes were categorized into three types according to the presence/absence of *ndh* genes. Different evolutionary dynamics for each of the three IR/SSC types of photosynthetic orchid plastomes were also proposed.

Keywords: Apostasioideae, plastome, hotspots, substitution rates, IR expansion/contraction
INTRODUCTION

Orchidaceae (orchids) is one of the most diverse and widespread families, comprising approximately 25,000 species in 880 genera (Chase et al., 2003). Their colorful flowers are often found in long sprays and stay fresh for months, giving orchids great ornamental as well as commercial value. According to molecular studies, orchids have been divided into five subfamilies: Apostasioideae, Vanilloideae, Cypripedioideae, Orchidoideae, and Epidendroideae (e.g., Givnish et al., 2015; Kim H.T. et al., 2015). Apostasioideae, the basal-most subfamily, consists of only two genera, Apostasia and Neuwiedia, with 18 species that are mainly distributed in Southeast Asia and northern Australia (Govaerts et al., 2012). These genera share some apomorphies with other orchids, e.g., enlarged petals (labella), a cylindrical structure of a fused gynoecium and anther (column) and small seeds, but they possess several unique characters, in particular their vessel cells in the root (Stern et al., 1993).

Because of their plesiomorphic floral characters and sister-group to the remaining orchids, members of the Apostasioideae have been intensively studied for their floral structure, taxonomy, biogeography, and genome variation. Kocyan and Endress (2001) analyzed the relationship between Apostasioideae and other orchids by studying their floral structure and development. The phylogenetic relationship within Apostasioideae was also evaluated based on ITS, trnL-trnF. The plastome sequences of Apostasia wallichii and Neuwiedia singapureana, which represent the genera Apostasia and Neuwiedia, respectively. The genome organization and the gene content and order were compared between the two species and with 45 previously reported photosynthetic orchid plastomes. Our aims were (1) to comparatively study the plastome structure and evolution of Apostasia and Neuwiedia, (2) to identify more powerful mutational hotspots for Apostasioideae, and (3) to evaluate the evolution of photosynthetic orchid plastomes and the mechanisms that underlie the shifts of their IR boundaries.

MATERIALS AND METHODS

DNA Extraction and Plastome Sequencing

Healthy leaves (2 g) were harvested from individuals of A. wallichii and N. singapureana grown in the greenhouse of Nanjing Normal University. The total DNA from the leaves was extracted using a Qiagen DNAeasy Plant Mini Kit (Qiagen, Germany) based on the manufacturer's protocol. The quality of the obtained DNA was examined using a NanoDrop 8000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, United States). Extracted DNA that met the criteria for analysis (concentration > 150 ng/µl, A260/A280 = 1.8–2.0, and A260/A230 > 1.7) was sequenced at 1Gene, Hangzhou (Hangzhou, China), using an Illumina HiSeq 4000 sequencer. The sequencing depth was 3.75 Gb of 150-bp paired-end reads for each species.

Plastome Assembly, Annotation, and Comparison

The raw sequencing reads were quality trimmed with an error probability of <0.05 and de novo assembled using CLC Genomic Workbench 6.0.1 (CLC Bio, Aarhus, Denmark). Contigs of <30× sequencing depths were discarded. Since the remaining contigs may contain the information not only from chloroplast genome, but also from nuclear genome or from mitochondrial genome, these contigs were searched using NCBI Blastn against the plastome sequences of A. odorata (NC_030722). Matched contigs with E-values of <10–10 were designated plastomic contigs. The gaps between plastomic contigs were closed by obtaining amplicons with specific primers and directly sequencing the amplicons. Junctions between the LSC/SSC and IRs were
amplified and confirmed by PCR assays. Genes were annotated using DOGMA (Wyman et al., 2004) and tRNAscan-SE 1.21 (Schattner et al., 2005). The exact boundaries of the annotated genes were confirmed by their alignment with their orthologous genes from published orchid plastomes.

The plastome sequences of *A. odorata* (NC_030722), *N. singapureana* (KM 244735) have been published in GenBank of NCBI. However, detailed analyses of plastome composition and structure and comparisons with those of other orchid species have not yet been conducted. Therefore, in this study, they were downloaded and compared with our newly sequenced two plastomes of *A. wallichii* (LC199394) and *N. singapureana* (LC199503) by using the mVISTA software (Frazer et al., 2004). The *N. singapureana* (KM 244735) was used as the reference.

Repeat Sequence and SSR Element Analysis

The size and location of repeat sequences, including palindromic, reverse and direct repeats, within the newly sequenced plastomes of *A. wallichii* (LC199394) and *N. singapureana* (LC199503) were identified using REPuter software (Kurtz et al., 2001). The repeats were identified according to the following conditions: (1) hammering distance of 3, (2) sequence identity ≥90%, and (3) minimum repeat size ≥30 bp. The simple sequence repeat (SSR) elements were detected using the tool GMATo (Wang et al., 2013), and the criteria of the “Min-length” for mononucleotide SSRs, dinucleotide SSRs, and multi-nucleotide SSRs were set to be greater than 8, 5, and 3 units, respectively.

Estimation of Sequence Divergence

For mutational hotspot selection, at least two complete plastomes from different species within the study genus should be available (e.g., Ahmed et al., 2013; Shaw et al., 2014). However, only the plastome of *N. singapureana* was sequenced in *Neuwiedia* genus. Therefore, the plastomes of *A. wallichii* and *A. odorata* were used to screen for the most informative regions for the genus *Apostasia*. The sequences of protein-coding genes and non-coding loci, including intergenic spacers and introns, were retrieved from the plastomes of *A. wallichii* and *A. odorata*. Non-coding loci <150 bp were discarded. The sequence alignments were conducted using MUSCLE (Edgar, 2004) implemented in Mega 5.2 (Tamura et al., 2011). The sequences of protein-coding genes were aligned with the Align Codons option using the default parameters. The sequences of non-coding loci were first aligned with the default parameters and then realigned with the “Refining” option. The gaps located at the 5′- and 3′-ends of the alignments were excluded. The yielded alignments were used to count the pairwise nucleotide substitutions and the insertion and deletion (InDel) events using DnaSP v5 (Librado and Rozas, 2009). The sequence variability (SV) was calculated according to the method of Shaw et al. (2005, 2014) and Downie and Jansen (2015): SV% = (the number of nucleotide mutations + the number of InDel events)/(the number of conserved sites + the number of nucleotide mutations + the number of InDel events) × 100.

Substitution Rates and Positive Selection Analysis

The sequences of 66 protein-coding genes were retrieved from the 45 orchid plastomes and the outgroup plastome of *Lilium longiflorum* (Supplementary Table S1). The sequence alignments of these genes were separately performed and then concatenated to generate a data set with 44,898 characters. The synonymous (ds) and non-synonymous (dn) substitution rates were estimated with the CodeML program of PAML 4.8 (Yang, 2007). The parameters were set to the following: seftype = 1, runmodel = −2. The SV for each non-coding locus of 10 genera from five orchid subfamilies: Epidendroideae (*Cymbidium*, *Phalaenopsis*, *Masdevallia*, *Dendrobium*, and *Bletilla*), Orchidoideae (*Goodyera*), Cypripedioideae (*Paphiopedilum* and *Cyripedium*), Vanilloideae (*Vanilla*), and Apostasioideae (*Apostasia*) were calculated in Niu et al. (2017) and this study. In order to evaluate the plastome-wide variation of substitution rates among orchid species, the synonymous (ds) and non-synonymous (dn) substitution rates for each protein-coding gene of those 10 genera were estimated in this study.

The selective pressure on the 66 genes from the 45 orchid species was also examined. Firstly, maximum likelihood tree was constructed using RAxML 8.0.2 (Stamatakis, 2014) based on the concatenated data set with a GTR+GAMMA model. The genes of *L. longiflorum* were used as outgroup. Secondly, the selective pressure on the 66 genes was analyzed using the site models in the CodeML program. LRT P-values were determined for three pairs of site models: M1a vs. M2a, M0 vs. M3, and M7 vs. M8.

Statistical Analysis

Mesquite v. 3.02 (Maddison and Maddison, 2011) was employed to evaluate the correlation between the shift of IR boundaries and the retained *ndh* gene length. The boundary of IRs that was used for the comparison was a region from the 5′ end of *ycf1* to the junction between the IR and SSC (Figure 6). Statistical analyses with Spearman and Mann–Whitney tests were performed using SPSS Statistics 20.0.

RESULTS

Sequencing and Plastome Assembly

A total of approximately 3.75 Gb of 150 bp pair-end reads for *A. wallichii* (3.83 Gb) and *N. singapureana* (3.67 Gb) were obtained from the Illumina paired-end sequencing. The *de novo* assembly produced 25,033 contigs for *A. wallichii*, 19,755 contigs for *N. singapureana*. Those contigs were searched against the plastome sequences of *A. odorata* (NC_030722). Then, 52 contigs were obtained with E-values <10−10 and mean coverage depth >30× for *A. wallichii*, 87 contigs for *N. singapureana*. Among them, five contigs that longer than 18 kb resulting in a nearly complete draft genome for *A. wallichii*. Eleven contigs that longer than 9 kb were used for assembling the plastome of *N. singapureana*. Gaps between contigs (5 gaps in *A. wallichii* plastome and 11 gaps in *N. singapureana* plastome) were closed by obtaining amplicons from PCR procedures. After
assembly and gap closure, two complete plastomes were obtained. Furthermore, the four junctions between the LSC/SSC and IRs of each plastome were also confirmed by PCR amplification.

Plastome Features of Two Newly Sequenced Orchids
The newly sequenced plastomes of *A. wallichii* (LC199394) and *N. singapureana* (LC199503) were circular and contained 156,126 and 161,068 bp, respectively (Figure 1). The plastomes displayed the typical quadripartite structure, consisting of a pair of IRs (26,452–26,991 bp) separated by the LSC (83,035–89,031 bp) and SSC (20,187–18,058 bp) regions (Table 1). The gene content of the two plastomes was relatively conserved with identical complements, each containing 79 unique protein-coding genes, 30 unique tRNA genes, and four unique rRNA genes. The overall AT content was 63.93–64.01%, indicating nearly identical levels among the orchid plastomes. The AT content was 66.70–66.79%, 71.26–70.83%, and 56.81–67.14% in the LSC, SSC, and IR regions, respectively (Table 1). Notably, the region from *trnS-GCU* to *trnS-GGA* was reversed in the two *Apostasia* plastomes. The four plastomes of Apostasioideae were compared and plotted using mVISTA, with *N. singapureana* (KM 244735) as the reference (Supplementary Figure S1). As expected, the SC regions were more divergent than the IR regions. Furthermore, the non-coding regions (intergenic spacers and introns) exhibited higher divergence levels than coding regions.

Repeat and SSR Analysis
We employed REPuBer to analyze the repeat sequences of the *A. wallichii* and *N. singapureana* plastomes. The total number of repeats was 38 in *A. wallichii* and 42 in *N. singapureana*. *A. wallichii* contained 15 forward repeats, 18 palindrome repeats, and 5 reverse repeats (Figure 2A and Supplementary Table S2). The length of these repeats ranged from 30 to 156 bp, and a copy length of 30–49 bp was most common (32 repeats), followed by >90 bp (three repeats), 50–69 bp (two repeats), and 70–89 bp (one repeat). *N. singapureana* contained 18 forward repeats, 16 palindrome repeats, and 8 reverse repeats. The length of the 42 repeats ranged from 31 to 140 bp, with a copy length of 30 to 49 bp being most common (31 repeats), while >90 bp was the least common (one repeat). In addition, our analyses revealed that the proportions of repeats located in non-coding regions were higher than in coding regions. For example, the proportion of the repeats located in non-coding regions of the *N. singapureana* plastome was 79.76%, while the repeats located in coding regions only accounted for 16.67%.

FIGURE 1 | Plastome map of *Neuwiedia singapureana*. The plastome of *Apostasia wallichii* has an identical structure to that of *N. singapureana*, except for an inversion from *trnS-GCU* to *trnS-GGA*. The genes outside and inside the circle are transcribed clockwise and counterclockwise, respectively.
The SSRs contained in the two plastomes were also counted (Figure 2B). The total number of SSRs identified in the A. wallichii and N. singapureana plastomes was 225 and 234, respectively. Among them, A- and T-monomer nucleotides were the most abundant SSRs in both plastomes, while the G- or C-repeats were rare. The number of dinucleotide repeats was slightly higher than other repeats, such as tri-, tetra-, and penta-nucleotides. Similar to the repeat sequences, the ratios of SSRs located in non-coding regions were higher than in coding regions (Mann–Whitney two-sided, \( P < 0.05 \)).

### Non-synonymous (dn) and Synonymous (ds) Substitution Rates among Orchid Species

Among the plastomes of all orchid species, the estimated synonymous (ds) and non-synonymous (dn) substitution rates were 0.2415–0.3439 (substitutions per ds sites) and 0.0387–0.0784 (substitutions per dn site), separately (Figure 4). The plastomes of four subfamilies (Epidendroideae, Orchidoideae, Cypripedioideae, and Apostasioideae) showed diversified substitution rates, with dn substitution rates ranging from 0.0387 to 0.0467 and ds substitution rates ranging from 0.2415 to 0.2927, while Vanilloideae (dn: 0.0753–0.0784, ds: 0.3383–0.3439) exhibited significantly higher substitution rates than those of the other orchid subfamilies (Mann–Whitney two-sided, \( P < 0.01 \)).

Moreover, the dn and ds rates were also calculated for each protein-coding genes (Figure 5). The result displayed a diversified dn and ds values among orchid genera. (1) Genes estimated with the highest dn (ds) rate might not have the highest value for ds (dn). For example, the gene of rpl22 has the highest dn value in Apostasia, whereas, the ds value was the lowest among all the orchid genera. (2) In one orchid genus, the substitutions rates among different genes were variable. For example, in the Vanilla plastomes, the dn rate of clpP was the highest among all protein-coding genes. The dn rate of psbB, the gene adjacent to clpP, was lower than most genes. In addition, both of them have showed a low degree of ds rates. (3) The substitution rates of

### Sequence Divergence of Apostasia Plastomes

We identified 79 coding genes and 89 syntenic non-coding regions longer than 150 bp in the two Apostasia plastomes. The SV for these loci was calculated. The SV value for each coding gene except for a few variable genes (ycf1, rps16, rpl20, and psal) was below 2.5%, with an average of 0.86% (Supplementary Table S3). The SV in non-coding regions was significantly higher, by approximately 2.6-fold, than that in coding regions (Mann–Whitney two-sided, \( P < 0.05 \)).

Furthermore, the top 10 loci—ndhA intron, matK-5’trnK, clpP-psbB, rps8-rpl14, trnT-trnL, 3’trnK-matK, clpP intron, psbK-trnK, trnS-psbC, and ndhF-rpl32—listed in Figure 3 were identified as mutational hotspots for the Apostasia plastome that can be used for phylogenetic analyses.

### TABLE 1 | Characteristics of the newly sequenced two Apostasioid Orchid Plastomes.

| Species           | Accession | Plastome length (bp) | LSC length (bp) | SSC length (bp) | IR length (bp) | AT content (%) | AT content in LSC (%) | AT content in SSC (%) | AT content in IR (%) |
|-------------------|-----------|----------------------|-----------------|-----------------|---------------|---------------|-----------------------|-----------------------|----------------------|
| Apostasia wallichii | LC199394  | 156,126              | 83,035          | 20,187          | 26,452        | 63.93         | 66.7                  | 71.26                 | 56.81                |
| Neuwiedia singapureana | LC199503  | 161,068              | 89,031          | 18,058          | 26,991        | 64.01         | 66.79                 | 70.83                 | 57.14                |

### Figure 2

**A** Number of repeat sequences identified by REPuter. **B** Number of SSR elements determined by GMATo.
Comparison of Sequences Flanking IR/SSC Junctions among Photosynthetic Orchid Plastomes

Sequences flanking IR/SSC junctions vary among photosynthetic orchid plastomes (e.g., Yang et al., 2013; Kim H.T. et al., 2015; Lin et al., 2015; Niu et al., 2017). To assess the variability of IR/SSC boundaries, we sampled and categorized 34 available plastomes from 19 genera into three types according to the presence or absence of ndh genes. As shown in Figure 6, type A, which has all 11 functional ndh genes, contains four genera from Epidendroideae (Calanthe, Elleanthus, Masdevallia, and Sobralia), three genera from Orchidoideae (Goodyera, Habenaria, and Ludisia), one genus from Cypripedioideae (Cypripedium), and the two genera from Apostasioideae (Apostasia and Neuwiedia). Type B has no ndh genes in the SSC region and includes two genera from Cypripedioideae (Paphiopedilum) and Vanilloideae (Vanilla). Type C, which has independently lost the ndh genes in SSC regions, consists of six genera from Epidendroideae (Cymbidium, Cattleya, Oncidium, Dendrobium, Phalaenopsis, and Bletilla) and one genus from Cypripedioideae (Phragmipedium). In addition, we determined the degree of IR expansion/contraction based on the length of a region from the 5′ end of the ycf1 gene to the IR and SSC junction. The IR lengths after expansion/contraction were significantly different among the three types of plastomes (Mann–Whitney two-sided, all P < 0.05).

In type A plastomes, the sequences flanking the IR/SSC junctions have similar genic architectures (Figure 6). Their junctions of IR_A/SSC are located approximately 1 kb downstream of the 5′ end of ycf1. Conversely, the boundaries of IR/SSC among the four type B species have expanded to variable positions (Figure 6). The different IR contraction lengths between the ndh-containing and ndh-lacking groups (Mann–Whitney two-sided, P < 0.05) indicated that the full set of ndh genes plays a role in IR/SSC junction stability. Among the orchid species from type C, the junctions of IR/SSC were variable due to the expansion/contraction of IRs (Figure 6). For example, the junction of IR_A/SSC was located at the spacer region adjacent to rpl32 in most species or to the 3′ end of ndhF among Cymbidium and Dendrobium, while the junctions of IR_A/SSC were located in ycf1, ranging from 9 to 664 bp. The correlation analysis based on the phylogenetic method suggested that the expansion/contraction length of IRs in type C is strongly correlated with the retained ndh gene length (Supplementary Figure S2, \( y = 0.136x, R^2 = 0.847, P < 0.05 \)).

DISCUSSION

Architecture of Plastid Genomes in Apostasioideae

Complete plastome sequencing and comparative genome analyses have revealed the high conservation of orchid plastomes in size, structure, and gene order and content (e.g., Chang et al., 2006; Jheng et al., 2012; Yang et al., 2013; Kim H.T. et al., 2015; Lin et al., 2015). The structure of Apostasioideae plastomes is
similar to that of most orchid plastomes, with the only difference being a large inversion from trnS-GCU to trnS-GGA in the Apostasia plastomes. Plastome rearrangement was also reported for the plastome of Cyripedium, which has an inversion from atpA to petG (Lin et al., 2015). Local disruptions of gene order occur in orchid plastomes because of the loss of ndh genes. A character map reconstructed by Kim H.T. et al. (2015) suggested that the ndh genes have experienced independent loss in the orchid genera but were present in the common ancestor of orchids. The full set of 11 functional ndh genes observed in both of these Apostasioideae plastomes confirmed their results.

Larger repeats are generally rare among angiosperm plastomes and are considered to be more prevalent in those plastomes that have undergone major changes in genome organization (Chumley et al., 2006; Haberle et al., 2008; Weng et al., 2014). Ahmed et al. (2012) reported the association between repeats, InDels and substitution imply that the regions, which have repeat sequences, tend to exhibit higher sequence variation. In the repeat analysis of the newly sequenced plastomes, 38 and 42 repeats of 30 bp or longer were identified in A. wallichii and N. singapureana plastomes, respectively. Most of these repeats are located in the non-coding regions, including some highly variable regions, such as 5’trnK-matK, ndhF-rpl32, and rps15-ycf1. These findings enhance the view that the evolution of non-coding regions was higher than that of coding regions. The repeat sequences have also played an important role in rearrangement in plastomes (Downie and Jansen, 2015; Asaf et al., 2016; Choi et al., 2016; Zhang et al., 2016). However, several repeats occurred in the same genes (ycf1 and ycf2) or genes with similar functions (pasA/pasB and trnS-GCU/trnS-UGA/trnS-GGA). These repeated elements are similar in number, size, and location to those of other angiosperms whose plastomes are structurally unrearranged (Ruhlman et al., 2006; Yi et al., 2012). Therefore, we suggest that these repeats may have been caused by replication slippage, which generates improper sequence recombination (Palmer, 1991; Downie and Jansen, 2015). Additionally, the SSRs contained in A. wallichii and N. singapureana were counted. These new resources will be potentially useful for the phylogenetic and population genetics studies in the Apostasia and Neuwiedia genera.

Disproportional Plastome-Wide Variation among Orchid Species

Plastid non-coding regions evolve through accumulated nucleotide substitutions and microstructural mutations such as InDels. However, Kelchner (2000) suggested that nucleotide substitutions and InDels would not be randomly distributed throughout plastomes. Morton (1995) and Decker-Walters et al. (2004) proposed that the occurrence of nucleotide substitutions and microstructural mutations depends strongly on sequence context. The extents of SV, InDel events, and substitution rates are related to low GC content, which suggests that mutational hotspots in orchid plastomes are accompanied by biased AT
compositions. Indeed, the GC contents of the top 10 mutational hotspots of *Apostasia* plastome were lower than that of most other loci. AT-based mutational hotspots are also observed in *taro* (Ahmed et al., 2012, 2013), *rice* (Yamane et al., 2006), and cycad (Wu and Chaw, 2015). AT richness contributes to potentially deleterious recombination and replication errors, which may explain its impact on patterns of substitution and the distribution of InDels.
The top 10 loci—ndhA intron, matK-5′trnK, clpP-psbB, rps8-rpl14, trnT-trnL, 3′trnK-matK, clpP intron, psbK-trnK, trnS-psbC, and ndhF-rpl32—that have the highest degrees of SV were identified as mutational hotspots for the Apostasia plastome. Compared to the mutational hotspots identified in Cymbidium, Phalaenopsis, and other orchid genera (Yang et al., 2013; Shaw et al., 2014; Niu et al., 2017), the top 10 mutational hotspots are variable among different orchid genera. Although different criteria and methods could affect the identification of the hotspots, our findings in this study suggest that mutational hotspots in orchid plastomes were genus-specific.

Shaw et al. (2014) concluded that a plastid region identified as highly variable in one group might not be consistently the most variable across other groups, which suggested that the evolution rates of the non-coding regions among different plant lineages were diversified. In line with this, the current research also revealed that the mutational hotspots were genus-specific among different orchid genera. These results indicated that the evolution rates of the non-coding regions among orchid genera were variable. Moreover, the coding genes also exhibited a diversified evolution rate not only in one orchid specie, but also among different orchid genera. Therefore, we proposed that there is a disproportional plastome-wide variation of substitution rates among orchid species.

Although the substitution rates among protein-coding genes are variable in photosynthetic orchid plastomes, some housekeeping genes related to photosynthesis showed evidence of positive selection pressure. As positive selection pressure would directly affect photosynthetic efficiency, whether it can benefit orchid adaptation is worthy of further investigations.

**Dynamic Evolution of Three IR/SSC Types in Photosynthetic Orchid Plastomes**

The expansion and contraction of the border position of the IR can be used to provide relationship evidence in phylogenetic analysis and have been used successfully in resolving relationships of major clades in ferns (Gao et al., 2009), Pinaceae (Lin et al., 2010), Poaceae (Guisinger et al., 2010), Apiaceae (Downie and Jansen, 2015), and many monocots (Wang et al., 2008). However, although overall genomic structures and gene orders are highly conserved, orchid plastomes exhibited obvious differences at the IR/SSC boundaries, which cannot readily be used in a phylogenetic analysis. Moreover, the ndh genes in SSC regions have been independently lost across orchid genera (e.g., Kim H.T. et al., 2015; Lin et al., 2015; Niu et al., 2017). The variations in the sequences flanking IR/SSC junctions and the independent loss of ndh genes have attracted the intense attention of researchers, leading to the publication of numerous plastomic comparative studies. For example, Luo et al. (2014) suggested dividing the orchid plastomes into four types based on distinct characteristics at the IR/SC junctions. Kim H.T. et al. (2015) proposed that the instability of the IR/SSC junctions in orchid was strongly correlated with the deletion of the ndhF gene. However, there is still a lack of information on the mechanism underlying the variations in the sequences flanking the IR/SSC junctions of orchid plastomes.

In the present study, we categorized 34 available photosynthetic orchid plastomes from 19 genera into three types according to the presence of ndh genes. The different IR contraction lengths between the plastomes from type A and type B and the correlation between the expansion/contraction length of IRs and the retained ndh gene length in type C lead us to conclude that the variations in sequences flanking the IR/SSC junctions of photosynthetic orchid plastomes are associated with the loss of ndh genes. Furthermore, we propose different evolutionary dynamics for each of the three IR/SSC types of orchid plastomes. For the plastomes in type A, the border position of the IR and SSC was stable. The IRA and SSC first experienced an expansion from 991 bp in Apostasia to 1462 bp in Cypripedium and then contracted progressively to downstream of the 5′ end of ycf1, while the junction of IRB and SSC was located between 17 bp adjacent to ndhF and 76 bp near the 5′ end of ndhF. For the plastomes in type B, the boundaries of IR/SSC have expanded to variable positions due to the loss of ndh genes. The sequences flanking the IR/SSC junctions vary among the type C plastomes. The correlation analysis based on phylogenetics indicated that the variations in IR/SSC junctions are strongly correlated with the retained ndh gene length. However, more molecular data still needs to be collected for intensifying our understanding of the dynamic evolution of sequence variations among IR/SSC junctions.

**AUTHOR CONTRIBUTIONS**

XD designed the study. ZN, SZ, JP, and WL performed the experiments. ZN, JP, LL, and QX analyzed the data. ZN wrote the manuscript.

**FUNDING**

This work was supported by grants from the National Natural Science Foundation of China (Grant No. 31170300 and No. 31670330) and the Priority Academic Program Development of Jiangsu Higher Education Institutions to XD.

**SUPPLEMENTARY MATERIAL**

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpls.2017.01713/full#supplementary-material

**FIGURE S1** | Sequence identity plots among the plastomes of A. wallichii (LC199394), A. odorata (NC_030722), and N. singapureana (LC199503) with the N. singapureana (KM244735) sequence as the reference. The region from trnS-GCU to trnS-GCA in A. wallichii and A. odorata plastomes is reversed for comparison.

**FIGURE S2** | Relationship between the contraction length of IRs and the retained ndh gene length in group C. The phylogenetic comparative method indicated that the expansion/contraction of IRs is significantly correlated with the loss of ndh genes (γ = 0.136; R² = 0.847; P < 0.05).
Ahmed, I., Biggs, P. J., Matthews, P. J., Collins, L. J., Hendy, M. D., and Lockhart, P. J. (2012). Mutational dynamics of aroid chloroplast genomes. *Genome Biol. Evol.* 4, 1316–1323. doi: 10.1093/gbe/evs110

Ahmed, I., Matthews, P. J., Biggs, P. J., Naeem, M., McLenachan, P. A., and Lockhart, P. J. (2013). Identification of chloroplast genome loci suitable for high-resolution phylogeographic studies of *Colocasia esculenta* (L.) Schott (Araceae) and closely related taxa. *Mol. Ecol. Resour.* 13, 929–937. doi: 10.1111/1755-0998.12128

Asaf, S., Khan, A. L., Khan, A. R., Waqas, M., Kang, S. M., Khan, M. A., et al. (2016). Complete chloroplast genome of *Nicotiana obtokia* and its comparison with related species. *Front. Plant Sci.* 7:843. doi: 10.3389/fpls.2016.00843

Chang, C. C., Lin, H. C., Lin, I. P., Chow, T. Y., Chen, H. H., Chen, W. H., et al. (2006). The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. *Mol. Biol. Evol. 23, 279–291*. doi: 10.1093/molbev/msj029

Chase, M. W., Cameron, K. M., Barrett, R. L., and Freudenstein, J. V. (2003). “DNA data and Orchidaceae systematics: a new phylogenetic classification,” in *Orchid Conservation*, eds K. W. Dixon, S. P. Kell, R. L. Barrett, and P. J. Cribb (Kota Kinabalu: Natural History Press), 69–89.

Choi, K. S., Chung, M. G., and Park, S. J. (2016). The complete chloroplast genome sequences of three Veronicaceae species (*Plantaginaceae*): comparative analysis and highly divergent regions. *Front. Plant Sci.* 7:355. doi: 10.3389/fpls.2016.00355

Chumley, T. W., Palmer, J. D., Mower, J. P., Fourcade, H. M., Calie, P. J., Boore, J. L., et al. (2006). The complete chloroplast genome sequence of *Pelargonium x hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Mol. Biol. Evol. 23, 2175–2190*. doi: 10.1093/molbev/msl089

Decker-Walters, D. S., Chung, S. M., and Staub, J. E. (2004). Plastid sequence evolution: a new pattern of nucleotide substitutions in the Cucurbitaceae. *J. Mol. Evol.* 58, 606–614. doi: 10.1007/s00239-004-2585-z

Delannoy, E., Fujii, S., Colas des Francs-Small, C., Brundrett, M., and Small, I. (2011). Raptant gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary changes on plastid genomes. *Mol. Biol. Evol. 28, 2077–2086*. doi: 10.1093/molbev/msr028

Downie, S. R., and Jansen, R. K. (2015). A comparative analysis of whole plastid genomes from the Aipaeas: expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. *Biotrop. Syst. 40, 336–351*. doi: 10.1006/0363644156X86607

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M., and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* 32, W273–W279. doi: 10.1093/nar/gkh458

Gao, L., Yi, X., Yang, Y. X., Su, Y. J., and Wang, T. (2009). Complete chloroplast genome sequence of *Phalaenopsis aphrodite* (Orchidaceae) based on ITS, trnL-F and *matK* sequences. *Ann. Mo. Bot. Gard.* 96, 499–527. doi: 10.1609/annbot.01904.10033–3

Kolansowska, M., Mystkowski, K., Kras, M., Dudek, M., and Konowalik, K. (2016). Evolution of the climatic tolerance and postglacial range changes of the most primitive orchids (Apostasioideae) within Sundaland, Wallacea and Sahul. *PeerJ* 4:e2384. doi: 10.7717/peerj.2384

Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., and Giegerich, R. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 29, 4633–4642. doi: 10.1093/nar/29.22.4633

Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi: 10.1093/bioinformatics/btp187

Lin, C. P., Huang, J. P., Wu, C. S., Hsu, C. Y., and Chaw, S. M. (2010). Comparative chloroplast genomics reveals the evolution of Pinaceae genera and subfamilies. *Genome Biol. Evol.* 2, 504–517. doi: 10.1093/gbe/eqv036

Lin, C. S., Chen, J. J., Huang, Y. T., Chan, M. T., Daniell, H., Chang, W. J., et al. (2015). The location and translocation of ndh genes of chloroplast origin in the Orchidaceae family. *Sci. Rep.* 5:9940. doi: 10.1038/srep09940

Lin, C. S., Chen, J. J., Chiu, C. C., Hsiao, H. C. W., Yang, C. J., Jin, X. H., et al. (2017). Concomitant loss of NDH complex-related genes within chloroplast and nuclear genomes in some orchids. *Plant J.* 90, 994–1006. doi: 10.1111/tjp.13525

Luo, J., Hou, B. W., Niu, Z. T., Liu, W., Xue, Q. Y., and Ding, X. Y. (2014). Comparative chloroplast genomes of photosynthetic orchids: insights into evolution of the Orchidaceae and development of molecular markers for phylogenetic applications. *PLOS ONE* 9:e99016. doi: 10.1371/journal.pone.0099016

Maddison, W. P., and Maddison, D. R. (2011). Mesquite: A Modular System for Evolutionary Analysis. Version 3.2. Available: http://mesquiteproject.org

Morton, B. R. (1995). Neighboring base composition and transversion/transition bias in a comparison of rice and maize chloroplast noncoding regions. *Proc. Natl. Acad. Sci. U.S.A.* 92, 9717–9721. doi: 10.1073/pnas.92.21.9717

Niu, Z. T., Xue, Q. Y., Zhi, S. Y., Sun, J., Liu, W., and Ding, X. Y. (2017). 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.* 8:715. doi: 10.3389/fpls.2017.00715

Palmer, J. D. (1991). “Plastid chromosomes: structure and evolution,” in *Cell Culture and Somatic Cell Genetics of Plants*, Vol. 7A, eds L. Bogorad and J. K. Vasil (San Diego, CA: Academic Press), 5–53.

Raubeson, L. A., and Jansen, R. K. (2005). “Chloroplast genomes of plants,” in *Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants*, eds K. A. Kubitzke and T. H. R. C. Schrire (Berlin: Springer), 103–126.

Ruijtenbeek, P. J., Naeem, M., McLenachan, P. A., and Lockhart, P. J. (2013). Identification of chloroplast genome loci suitable for high-resolution phylogeographic studies of *Colocasia esculenta* (L.) Schott (Araceae) and closely related taxa. *Mol. Ecol. Resour.* 13, 929–937. doi: 10.1111/1755-0998.12128

Jansen, R. K., and Ruhilman, T. A. (2012). “Plastid genomes in seed plants,” in *Genomics of Chloroplasts and Mitochondria*, eds R. Bock and V. Knoop (Dordrecht: Springer), 103–126. doi: 10.1007/978-94-007-2920-9_5
Ruhlman, T., Lee, S. B., Jansen, R. K., Hostetler, J. B., Tallon, L. J., Town, C. D., et al. (2006). Complete plastid genome sequence of *Daucus carota*: implications for biotechnology and phylogeny of angiosperms. *BMC Genomics* 7:222. doi: 10.1186/1471-2164-7-222

Schattner, P., Brooks, A. N., and Lowe, T. M. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* 33, W686–W689. doi: 10.1093/nar/gki366

Shaw, J., Lickey, E. B., Beck, J. T., Farmer, S. B., Liu, W., Miller, J., et al. (2005). The *Cycas taitungensis* cpDNA: insights into cpDNA evolution and phylogeny of extant seed plants. *Mol. Biol. Evol.* 24, 1366–1379. doi: 10.1093/molbev/msm059

Wu, C. S., Jansen, R. K., and Boore, J. L. (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20, 3252–3255. doi: 10.1093/bioinformatics/bth352

Yamane, K., Yano, K., and Kawahara, T. (2006). Pattern and rate of Indel evolution inferred from whole chloroplast intergenic regions in sugarcane, maize and rice. *DNA Res.* 13, 197–204. doi: 10.1093/dnares/dsl012

Yang, J. B., Tang, M., Li, H. T., Zhang, Z. R., and Li, D. Z. (2013). Complete chloroplast genome of the genus *Cymbidium*: lights into the species identification, phylogenetic implications and population genetic analyses. *BMC Evol. Biol.* 13:84. doi: 10.1186/1471-2148-13-84

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088

Zhang, Y., Du, L., Liu, A., Chen, J., Wu, L., Hu, W., et al. (2016). The complete chloroplast DNA sequence of *Eleutherococcus senticosus* (Araliaceae); comparative evolutionary analyses with other three asterids. *Mol. Cells* 33, 497–508. doi: 10.1007/s10059-012-2281-6

Zhang, Y., Lu, P., and Luo, Z. (2013). GMATo: a novel tool for the identification and analysis of microsatellites in large genomes. *Bioinformation* 9, 541–544. doi: 10.6026/97320630009541

Weng, M. L., Blazier, J. C., Govindu, M., and Jansen, R. K. (2014). Reconstruction of the ancestral plastid genome in *geraniaceae* reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Mol. Biol. Evol.* 31, 645–659. doi: 10.1093/molbev/msu257

Wu, C. S., and Chaw, S. M. (2015). Evolutionary stasis in cycad plastomes and the first case of plastome GC-biased gene conversion. *Genome Biol. Evol.* 7, 2000–2009. doi: 10.1093/gbe/evv125