

Article

Severe Plastid Genome Size Reduction in a Mycoheterotrophic Orchid, *Danxiaorchis singchiana*, Reveals Heavy Gene Loss and Gene Relocations

Shiou Yih Lee ^{1,2}, Kaikai Meng ², Haowei Wang ², Renchao Zhou ², Wenbo Liao ², Fang Chen ³, Shouzhou Zhang ^{1,*} and Qiang Fan ^{2,*}

¹ Shenzhen Key Laboratory of Southern Subtropical Plant Diversity, FairyLake Botanical Garden, Shenzhen & Chinese Academy of Sciences, Shenzhen 518004, China; leesy3@mail.sysu.edu.cn

² State Key Laboratory of Biocontrol and Guangdong Provincial Key Laboratory of Plant Resources, School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China; mengkk@mail2.sysu.edu.cn (K.M.); lsswhw@foxmail.com (H.W.); zhrench@mail.sysu.edu.cn (R.Z.); lsslwb@mail.sysu.edu.cn (W.L.)

³ Administrative Committee of Danxiashan World Natural Heritage and Global Geopark, Shaoguan 521300, China; waynexiao800@gmail.com

* Correspondence: shouzhouz@szbg.ac.cn (S.Z.); fanqiang@mail.sysu.edu.cn (Q.F.)

Received: 10 March 2020; Accepted: 11 April 2020; Published: 17 April 2020



Abstract: *Danxiaorchis singchiana* (Orchidaceae) is a leafless mycoheterotrophic orchid in the subfamily Epidendroideae. We sequenced the complete plastome of *D. singchiana*. The plastome has a reduced size of 87,931 bp, which includes a pair of inverted repeat (IR) regions of 13,762 bp each that are separated by a large single copy (LSC) region of 42,575 bp and a small single copy (SSC) region of 17,831 bp. When compared to its sister taxa, *Cremastra appendiculata* and *Corallorhiza striata* var. *involuta*, *D. singchiana* showed an inverted gene block in the LSC and SSC regions. A total of 61 genes were predicted, including 21 tRNA, 4 rRNA, and 36 protein-coding genes. While most of the housekeeping genes were still intact and seem to be protein-coding, only four photosynthesis-related genes appeared presumably intact. The majority of the presumably intact protein-coding genes seem to have undergone purifying selection ($dN/dS < 1$), and only the *psaC* gene was positively selected ($dN/dS > 1$) when compared to that in *Cr. appendiculata*. Phylogenetic analysis of 26 complete plastome sequences from 24 species of the tribe Epidendreae had revealed that *D. singchiana* diverged after *Cr. appendiculata* and is sister to the genus *Corallorhiza* with strong bootstrap support (100%).

Keywords: Calypsoinae; evolution; gene loss; nonphotosynthetic; plastome; pseudogene

1. Introduction

Epidendroideae is the largest of the five subfamilies in the family Orchidaceae [1]. As this subfamily accounts for the majority of the orchid radiation and displays well-known orchid characteristics, the study of epidendroid diversification is deemed essential in orchid systematics. However, a portion of the members of this subfamily shows low photosynthetic ability and instead rely on mycoheterotrophy for nutrition [2]. As plants with mycoheterotrophic capabilities commonly have distinct anatomical, physiological, and genomic features, a reduction in plastid genome size is expected to occur due to changes in photosynthesis-related protein-coding genes, i.e., loss or pseudogenized [3]. Studies have revealed no significant changes in the plastome structure despite a reduction in the *ndh* complex [4]. Further investigations have proven that the evolution of heterotrophy occurs rapidly and independently across the genus [2,5]. Although studies on the plastome of selected mycoheterotrophic orchids have been performed by several plant researchers, the lack of comprehensive species samplings limits the

amount of available information on the mycoheterotrophic lineages in these selected orchids. Therefore, additional information on mycoheterotrophic orchids of different genera and species is required to better understand the peculiar evolution that leads to plastome reduction.

To date, the complete plastome of a total of 30 species in 12 genera of mycoheterotrophic orchids in the family Orchidaceae has been reported. However, only *Cyrtosia septentrionalis* and *Rhizanthella gardneri* represent the subfamilies Vanilloideae and Orchidoideae [6,7], respectively, while the other species with a complete plastome belong to the Epidendroideae subfamily [5,8–17]. Although these orchids are reported as mycoheterotrophs, they can be categorized into two types, namely, photosynthetic mycoheterotrophs (or semi-mycoheterotrophs, not to be confused with partial mycoheterotrophs), which often have reduced leaves and/or photosynthetic pigmentation observed in deformed leaves or stem tissues, and full (nonphotosynthetic) mycoheterotrophs, which lack chlorophyll and are thus unable to produce carbohydrates through photosynthesis [17]. Studies on fully mycoheterotrophic orchid genera can provide unexpected yet valuable information about the evolutionary transition of nonphotosynthetic plants from partial mycoheterotrophy to full mycoheterotrophy [16].

The recently described mycoheterotrophic orchid genus, *Danxiaorchis* (Calypsoinae, Epidendreae) contains only two recorded species, which are full mycoheterotrophs and are restricted to specific habitats. The type species, *Danxiaorchis singchiana* J. W. Zhai, F. W. Xing, and Z. J. Liu., known as the Danxia orchid, is a distinct lineage with morphological features such as a labellum with a large Y-shaped callus and two sacs at the base, and cylindrical, fleshy seeds. The Danxia orchid is confined to Mount Danxia in Guangdong, China, after which the genus was named [18]. According to the most recent phylogenetic analysis of *Danxiaorchis*, the genus is sister to *Corallorhiza* and *Oreorchis*, although *Oreorchis* could be paraphyletic to the other two genera [19].

Research on orchids, the largest family of flowering plants [20], has moved from phylogenetic analyses using short gene sequences to those using genome-scale data. Researchers have sought to understand the extraordinary diversification and genetic evolution of these higher plant species, especially the species that evolved full mycoheterotrophy and leaflessness. In this study, we conducted sequencing, assembly and gene annotation of the *D. singchiana* plastome. Our aim was to reveal the extent of reductive plastome evolution in *D. singchiana*, to perform comparative analyses between *D. singchiana* plastome and published plastomes of selected photosynthetic and nonphotosynthetic sister genera in subtribe Calypsoinae, and to construct a phylogenomic tree in order to clarify the molecular placement of *D. singchiana* in tribe Epidendreae at the genome level.

2. Results

2.1. Plastome Size and Structure

With next-generation sequencing, a total of 32,313,250 raw reads were obtained. Raw reads were directly fed into the pipeline without quality filtering or trimming in order to obtain the maximum amount of useful data. To speed up the assembly of plastid genomes, we chose only the first 2.5 million sequences in each paired-end dataset for downstream assembly. Finally, a total of 71,364 aligned reads were acquired, and 65,892 assembled reads were incorporated in the plastome assembly, with an average coverage depth of 122 times per site. The complete plastome of *D. singchiana* has a quadripartite structure and is 87,931 bp in length (Figure 1), containing a pair of inverted repeats (IRs; 13,762 bp each) that separate the large single-copy (LSC; 42,575 bp) region and the small single-copy (SSC; 17,831 bp) region. The total GC content of the plastome is 34.54%.

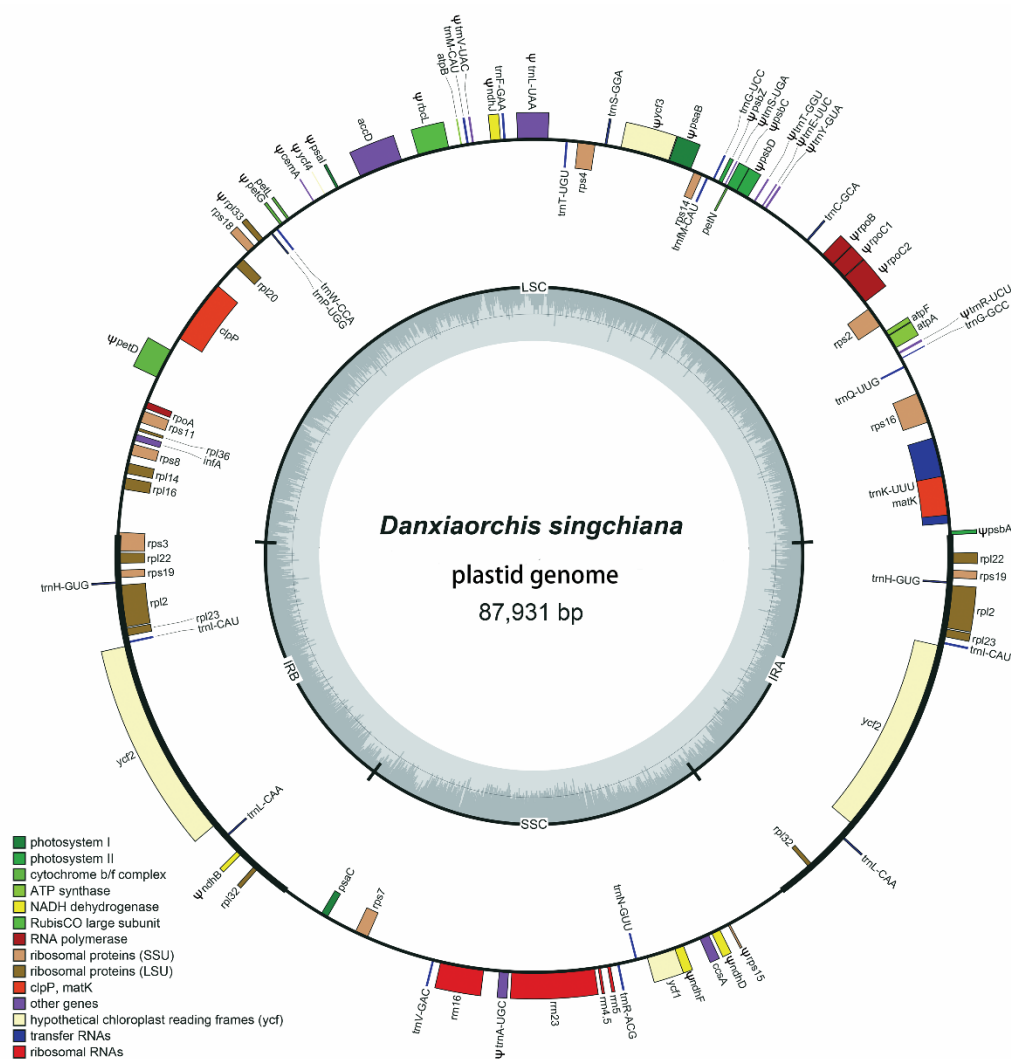


Figure 1. Plastid genome map of *Danxiaorchis singchiana*. Genes shown inside circles are transcribed in a clockwise direction, and those outside the circle are transcribed in a counterclockwise direction. Pseudogenes are indicated with ψ . The figure legend shows the gene functions based on color codes.

2.2. Gene Content and Order

For *D. singchiana*, we predicted a total of 61 genes, among which 36 were protein-coding genes, 21 were tRNA genes, and four were ribosomal RNA genes. However, nine of them were duplicates in the IR regions (Table 1). *Cremastra appendiculata* is the closest relative with a known plastome sequence available in the NCBI GenBank database. A comparison between the plastomes of *D. singchiana* and *Cr. appendiculata* revealed that 27 genes were pseudogenized, while 25 genes were lost (Table 1). Based on the classification scheme used in a previous study [5], the genes in the plastome were further grouped into five categories, namely, housekeeping, photosynthesis-related, NADPH dehydrogenase, plastid-encoded RNA polymerase, and ATP synthase. Among the housekeeping genes, 12 of 61 (19.7%) were recorded as pseudogenes or lost, with eight tRNA pseudogenes for the tRNA genes (*trnA*-UGC, *trnE*-UUC, *trnL*-UAA, *trnR*-UCU, *trnS*-UGA, *trnT*-GGU, *trnV*-UAC, and *trnY*-GUA) and one for the small subunit of the ribosome (*rps15*), while three tRNA genes (*trnD*-GUC, *trnL*-UAG, and *trnS*-GCU) were not found in the plastome. Among the photosynthesis-related genes, 28 of 31 (90.3%) were recorded as pseudogenes or lost, with the remaining presumably intact genes including the photosystem I subunit *psaC*, and the cytochrome *b₆f* subunits *petL* and *petN*. For the plastid-encoded RNA polymerase, only the subunit *rpoA* remained presumably intact. Interestingly, all the genes

in the NADPH dehydrogenase (*ndh*), ATP synthase (*atp*), and photosystem II (*psb*) subunits of the photosynthesis-related genes were either pseudogenized or lost in *D. singchiana* plastome. By using progressiveMAUVE to detect the presence of large-scale evolutionary events among *Cr. appendiculata*, *Corallorhiza striata* var. *involuta*, and *D. singchiana*, we found two inversions relative to the conserved gene order in angiosperm plastomes in *D. singchiana* but none in *Co. striata* var. *involuta* (Figure 2). The inversions each occurred in the LSC (2.97-kb in length) and SSC (6.9 kb in length) regions, with the LSC inversion including only the protein-coded gene for *petN* and pseudogenes *psbC*, *trnE-UUC*, *trnS-UGA* *trnT-GGU*, and *trnY-GUA* and the SSC inversion consisting of the presumably intact genes *trnN-GUU* and *ycf1* and pseudogenes for *ccsA*, *ndhF*, *ndhD*, and *rps15*.

Table 1. List of presumably intact genes in the plastome of *Danxiaorchis singchiana* when compared to that of *Cremastra appendiculata* (MH356724). Genes that were duplicated in the inverted repeats (IR) regions are marked with the symbol *.

Type of Genes	Group of Gene	Gene Name	
Housekeeping	Large subunit of ribosome	<i>rpl2*</i> ; <i>rpl14</i> ; <i>rpl16</i> ; <i>rpl20</i> ; <i>rpl22*</i> ; <i>rpl23*</i> ; <i>rpl32*</i> ; <i>rpl33</i> ; <i>rpl36</i>	
		<i>rps2</i> ; <i>rps3*</i> ; <i>rps4</i> ; <i>rps7</i> ; <i>rps8</i> ; <i>rps11</i> ; <i>rps12</i> ; <i>rps14</i> ; <i>rps16</i> ; <i>rps18</i> ; <i>rps19*</i>	
	Small subunit of ribosome	<i>rrn4.5</i> ; <i>rrn5</i> ; <i>rrn16</i> ; <i>rrn23</i>	
		<i>trnC-GCA</i> ; <i>trnF-GAA</i> ; <i>trnM-CAU</i> ; <i>trnG-GCC</i> ; <i>trnG-UCC</i> ; <i>trnH-GUG*</i> ; <i>trnI-CAU*</i> ; <i>trnI-GAU</i> ; <i>trnK-UUU</i> ; <i>trnL-CAA*</i> ; <i>trnM-CAU</i> ; <i>trnN-GUU</i> ; <i>trnP-UGG</i> ; <i>trnQ-UUG</i> ; <i>trnR-ACG</i> ; <i>trnS-GGA</i> ; <i>trnT-UGU</i> ; <i>trnV-GAC</i> ; <i>trnW-CCA</i>	
	Transfer RNA genes	<i>infA</i>	
		<i>matK</i>	
	Photosynthesis-related	Translational initiation factor	<i>accD</i>
		Maturase	<i>clpP</i>
		Subunit of acetyl-CoA carboxylase	<i>ycf1</i> ; <i>ycf2*</i>
		Subunit of protease Clp	<i>psaC</i>
Component of TIC complex		<i>petL</i> ; <i>petN</i>	
Plastid-encoded RNA polymerase	Subunits of photosystem I	<i>rpoA</i>	
	Subunits of cytochrome <i>b₆f</i>		
	Subunits of the plastid-encoded RNA polymerase		

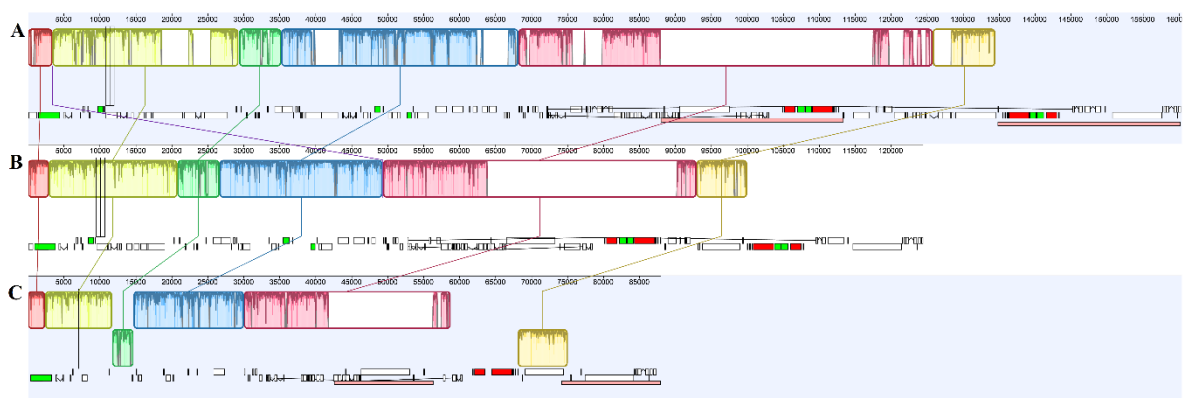


Figure 2. Plastome sequence alignment between (A) *Cremastra appendiculata*, (B) *Corallorhiza striata* var. *involuta*, and (C) *Danxiaorchis singchiana*, using progressiveMAUVE [21].

2.3. Genome Comparison and Selection Pressure Analyses

When compared to the other 25 complete plastomes from tribe Epidendreae, the plastome of *D. singchiana* displayed the smallest total size (Table S1). More precisely, the lengths of its LSC region and IR region were the shortest across all species that were included in the study, but not that of its SSC region. The SSC region of *D. singchiana* is intermediate in length compared with the SSC region lengths of the other orchid species and shorter than the SSC region of *Anathalis obovata*,

Cr. appendiculata, *Masdevallia coccinea*, and *Masdevallia picturata*. In the dN/dS analysis, the overall dN/dS for the concatenated dataset was less than 1.0, which was due to signals of purifying selection (Table S2). When analyzed separately, the 25 genes displayed purifying selection, whereby changes in amino acid residues in the genes that could favor an excess of synonymous versus nonsynonymous substitutions were prevented. The chloroplast gene *psaC*, displayed positive selection, whereby the protein sequences were altered, in *D. singchiana*, whereas the other four groups of genes did not exhibit any nonsynonymous (K_a) or synonymous (K_s) values indicative of selection due to the constraints of the used model.

2.4. Phylogenomic Analysis

The maximum-likelihood (ML) tree was based on the complete plastome sequences of 26 orchid species in tribe Epidendreae and demonstrated that *D. singchiana* diverged after *Cr. appendiculata* and was sister to the genus *Corallorhiza* (Figure 3). This molecular placement was supported by strong bootstrap values at almost all branch nodes.

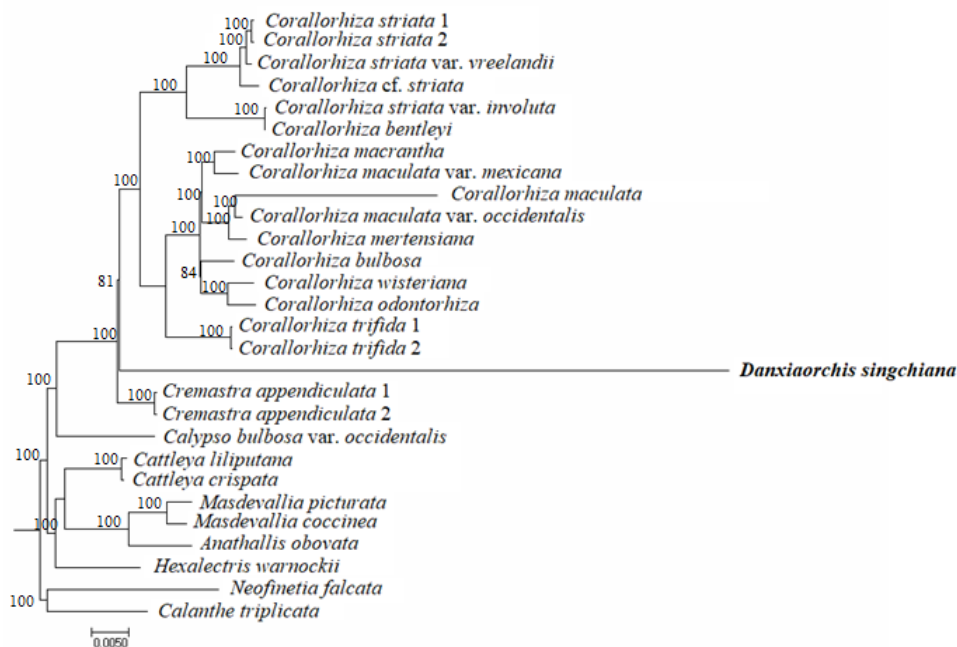


Figure 3. Maximum-likelihood (ML) tree of 26 complete plastome sequences derived from the 23 orchid species in tribe Epidendreae. The GenBank accession numbers for the orchid species used in this study are listed in Table S1. Two closely-related species, *Calanthe triplicata* (tribe Collabieae) and *Neofinetia falcata* (tribe Vandeeae), were included as outgroups.

3. Discussion

In general, land plants contain a chloroplast (cp) genome between 120 and 160 kb in size, while differences in cp genome size are mostly influenced by the expansion/contraction of IR regions due to changes in the amount of repeated DNA and/or changes in sequence complexity [22]. This is true for *D. singchiana* as the IR sequence length was halved compared to the IR sequence length in the other species of the same subtribe (Table S1). However, its SSC sequence length was slightly larger than that of the other mycoheterotrophic orchid species in the same subtribe but shorter than that of other photosynthetic orchid species in the same tribe. *Cremastra appendiculata*, which is the closest relative of *D. singchiana* with a known plastome, contains 129 genes, including 86 protein-coding, 35 tRNA, and 8 rRNA genes. Upon comparison, we found that 50, 14, and four protein-coding, tRNA, and rRNA genes, respectively, were lost in *D. singchiana*. As *D. singchiana* diverged after *Cr. appendiculata*, it experienced a 52.7% reduction in its total gene abundance. The major loss of presumably intact

genes mainly occurred for the photosynthesis-related, *ndh*, plastid-encoded RNA polymerase, and *atp* genes, while the amount of gene loss was minimal among the housekeeping genes. In our study, the presence of pseudogenes was based on gene predictions computed with the gene annotation tool GeSeq, whereby precise annotation was based on the similarity of the protein sequence with a proper start codon [23]. The plastome sequence of *D. singchiana* was aligned against the plastome sequence of *Cr. appendiculata* and crosschecked for possible pseudogenized gene sequences and confirmed gene losses. On the basis of our observation, pseudogenes are present among protein-coding loci in the plastome for three reasons that cause them to be putatively nonintact: (1) premature stop codons within the gene sequence, (2) frequent deletion of gene sequences at both ends, and (3) the occurrence of unexpected single nucleotide polymorphisms (SNPs) in the start codon that result in its loss.

Losses of *ndh* genes in mycoheterotrophic orchids are common, as demonstrated in several genera in family Orchidaceae and in Calypsoinae, and consistent *ndh* gene loss was detected in the genus *Corallorhiza* [5] (Figure S1). It was proven that some photoautotrophic and all heterotrophic plants lack a functional *ndh* complex and thus, are likely to be missing plastid-encoded *ndh* genes [24]. Furthermore, multiple independent degradation pathways may exist in the orchid family according to the presence of variations in the status of *ndh* genes among the orchid plastomes that have been reported thus far [4]. The literature had shown that in some plant species, however, *ndh* genes have been transferred to the mitochondrial genome, usually in a truncated form or with large deletions [25]. The loss of *ndh* genes is strongly correlated with the instability of the SSC/IR junction in Orchidaceae, in which the species with presumably intact *ndhF* genes exhibit a part of the *ycf1* sequence (approximately 1 kb) located within the IR, while the *ndhF* gene is located very near to the other end of the SSC/IRA junction [7]. For species with *ndhF* gene deletions, the *ycf1* gene was observed to have varying sequence lengths over the SSC/IRA junction, to be completely placed within the IR region, or to be completely in the SSC region but with its sequence end less than 21 bp away from the SSC/IRA junction [25]. However, in *D. singchiana*, we observed that the deletion of *ndh* genes severely shifted the IR boundary, resulting in extensive structural rearrangement whereby the *ycf1* gene was completely placed within the SSC region, which was 3,940 bp away from the SSC/IRB junction. On the other hand, the pseudogenized *ndhF* gene was found to be located at the 3' end, before the *ycf1* gene (Figure 1). This unique translocation of the *ycf1* gene was also accompanied by the 6.9-kb gene block inversion detected in the SSC region (Figure 2). As inversions of the LSC often correspond to sequence alterations in IR regions [26], the occurrence of similar events could be proposed for the SSC region of *D. singchiana*. Genetic inversions are usually found in plastomes of parasitic and mycoheterotrophic plants but rarely in photosynthetic plants [4,7,12,27,28]. Inversions in plastomes could result from homologous recombination between repeated sequences and could lead to rearrangement of gene order [29]. In this study, the two inversions that were detected in the *D. singchiana* plastome were found to be unique to the species. Such a phenomenon might lead to unexpected findings, and the mechanism of the natural development of genetic inversions in the species could be related to its unique lifestyle. Nevertheless, in-depth studies are needed to obtain a better understanding of this phenomenon.

Because of the absence of almost all photosynthesis-related genes in the plastome, the loss-of-function in plastid-encoded *atp* genes is expected. Even so, most nonphotosynthetic orchid species that experienced a substantial loss of photosynthesis-related genes still retain the *atp* complex, such as *Corallorhiza* spp. (Figure S1). The *atp* complex is responsible for ATP synthesis in photosynthesis, also called photosynthetic phosphorylation or photophosphorylation [30]. Cyclic photophosphorylation requires only a light source for complete photosynthesis and is driven by photosystem I, while noncyclic photophosphorylation is linked to oxygen evolution, water oxidation, and acceptor reduction [31] and requires both photosystem I and II [32]. As a result of few to no photosynthesis-related genes being found in the *D. singchiana* plastome, the *atp* complex was reduced and experienced gene loss. On the other hand, the retention of the *atp* complex in *Corallorhiza* could be regulated by an unknown, photosynthesis-independent mechanism. Such a mechanism was proposed

to function in maintaining a proton gradient between the thylakoid lumen and stroma, which is a criterion for the Tat-dependent protein translocation system [33].

Reductive evolution could be a consequence of a nonphotosynthetic lifestyle [34], in which the mycoheterotrophic plant relies on the microbiome community in its rhizome for consistent nutrient uptake. Although the plastome of *D. singchiana* experienced extensive gene loss, the conservation of essential plastid genes plays an indispensable role in cellular viability and thus ensures the plant's survival [35]. In a comprehensive review, a total of 52 genes were identified to be essential to heterotrophic growth [36]. Among these genes, nine were lost from the *D. singchiana* plastome, namely, *trnA*-UGC, *trnD*-GUC, *trnE*-UUC, *trnL*-UAA, *trnL*-UAG, *trnR*-UCU, *trnS*-GCU, *trnV*-UAC, and *trnY*-GUA, which were all in the transfer RNA gene family. At least 25 transfer RNA genes are required for a plant plastome to be functional [37]. A study predicted that the *trnE* gene, which is essential for haem biosynthesis [38], will cling to the plastome even when plant cell lines experience plastome deletion [39]. However, the *trnE* gene is predicted to be a pseudogene in the *D. singchiana* plastome. The ribosomal RNA genes, on the other hand, were still all presumably intact in the *D. singchiana* plastome. This suggests that the plastome in *D. singchiana* remains functional; however, the essential transfer RNA genes were incomplete. Therefore, translation of *D. singchiana* plastid mRNAs is probably to involve tRNAs of cytosolic origin [40], suggesting that tRNA import may occur in *D. singchiana* plastome to warrant its functionality.

The first phylogenetic analysis of *D. singchiana* was performed against 71 genera of Orchidaceae using a combined dataset of three gene/loci, including the two protein-coding plastid genes *matK*, and *rbcL* and the nuclear ribosomal DNA internal transcribed spacer (ITS) region [18]. Although the finding revealed that *Danxiaorchis* is a sister species of *Yuania*, the inclusion of the partial *rbcL* gene sequence of *Yuania japonica* in the phylogenetic analysis was questionable. Thus, a comprehensive phylogenetic analysis of 37 species in 14 genera of subtribe Calypsoinae was carried out using the combined *matK* and ITS dataset [19]. The new phylogenetic tree revealed that *D. singchiana* is a sister species to the *Oreorchis*+*Corallorhiza* clade under a strong Bayesian posterior probability (0.90) and moderate ML bootstrap support (77%). The same analysis criteria were also used when reporting the molecular placement of *Danxiaorchis yangii*, a novel congener that was discovered a few years after *D. singchiana*, on Mount Jinggang of Hunan province, China [41]. Finally, the results of the phylogenetic analysis carried out using the complete genome sequence of 26 species in tribe Epidendreae were congruent with the phylogenetic relationships in subtribe Calypsoinae proposed in a previous study [19]. Although a minor inconsistency in molecular placement within the *Corallorhiza* clade was observed, we concluded that the ML tree constructed using the complete plastome sequences was capable of depicting the phylogenetic relationships among orchid species in subtribe Calypsoinae and is, therefore, useful at the subfamily level.

Identifying a suitable gene region for phylogenetic analysis in Orchidaceae is rather challenging due to gene absence and pseudogenes in some orchid species. In fact, pseudogenes are not favorable in phylogenetic analysis in some orchid species [42]. As a result, the application of *matK* and ITS in the construction of a phylogenetic tree for Epidendreae is promising [19]; *rbcL* is a pseudogene in mycoheterotrophic orchids, such as *D. singchiana* and *Y. japonica*. However, the *matK* sequence was not found in *Yuania amagiensis* based on a dataset of NGS sequences [19]. The absence of the *matK* gene in *Y. amagiensis* may be individual-specific, as different gene numbers were detected in the same mycoheterotrophic/photosynthetic orchid, e.g., *Co. striata* and *Cr. appendiculata*. We suggest performing additional NGS sequencing to clarify the presence of the *matK* gene in *Y. amagiensis* in order to provide better resolution for mycoheterotrophic orchids in Epidendreae.

Mount Danxia, from which *D. singchiana* originated, is speculated to have formed 6 million years ago (Mya) [43]. Researchers have suggested that the unique geological conditions and relative environmental isolation of this area give rise to rare and endemic species with restricted habitats [44]. The presence of *D. singchiana* could be a good example of genetic isolation and speciation leading to the evolution of the mycoheterotrophic lifestyle due to its distinctive habitat. To date, there is

limited information on the molecular dating of *D. singchiana*, which is based on three mitochondrial protein-coding genes, namely, *atp1*, *matR*, and *nad5* [45]. The divergence time between *Cr. appendiculata* and *D. singchiana* was speculated to have been approximately 11.06 Mya. To clarify the relationship between the formation of Mount Danxia and the speciation of *D. singchiana*, sequencing of its only congener, *D. yangii*, could provide better insights into the evolution traits of this mycoheterotrophic species and aid in deciphering the plastome evolution in *Danxiaorchis* as a whole.

4. Materials and Methods

Fresh aboveground tissue of *D. singchiana* was collected from Mount Danxia, Guangdong Province, China. Genomic DNA was extracted using a modified CTAB protocol [46]. A genomic library with an insert size of ~350 bp was prepared using a TruSeq DNA Sample Prep Kit (Illumina, USA) and sequenced on an Illumina NovaSeq platform (Illumina, USA) to obtain pair-end reads of 150 bp. Upon removal of adapter sequences using NGSToolkit [47], the raw reads were subjected to *de novo* assembly using NOVOPlasty 2.7.2 [48], in which the *rbcl* gene sequence of *D. singchiana* (Genbank accession: JX293187) was used as the initial seed. A single contig was obtained at the end of the assembly process. The assembled genome was initially annotated and its junction points between the LSC, SSC, and IRs were identified using the online annotation tool GeSeq [23]; then, the genome was manually checked for annotation errors. The plastome sequence was deposited in NCBI GenBank under accession number MN584923. Plastome alignment was conducted using the web-based alignment software MAFFT [49] with 26 complete plastomes of 23 species in the Epidendreae tribe downloaded from NCBI GenBank (Table S1), and the locations of pseudogenes were manually identified by crosschecking the unannotated gene sequences. The size of the plastome was compared with that of the other 23 species, and a circular plastome map was visualized in OGDRAW 1.3.1 [50].

Based on recent phylogenetic analysis and classification of the Calypsoinae subtribe, *Cremastra* is believed to be the closest genus to *Danxiaorchis*, and *Corallorhiza* and *Oreorchis* diverged after *Danxiaorchis* [19]. Subsequently, plastome comparative analysis was conducted for *D. singchiana* against *Cr. appendiculata* and *Co. striata* var *involuta*, whereby the plastomes were aligned using progressiveMAUVE [21], and the presence of large-scale evolutionary events such as rearrangement and inversion between *D. singchiana* and its sister genera was determined.

The ratio of nonsynonymous to synonymous substitutions (dN/dS) of protein-coding regions was calculated for *Cr. appendiculata* and *D. singchiana* (Table S1). Plastome alignment was conducted using MAFFT [49], and protein-coding genes present in both species were included in the analysis, using PAML 4.7 [51] to estimate the selection pressure acting on these genes. The Yang and Nielsen codon frequency model (yn00) was selected, with the parameters for the initial ratio of transitions to transversion frequency (K) set at $0.3 < K < 0.7$. Calculations were carried out in two ways: (1) with a concatenated dataset containing 30 protein-coding genes and (2) with all 30 protein-coding genes analyzed separately.

For phylogenomic tree construction, the plastome sequence of *D. singchiana* and 25 complete plastomes of 23 species in tribe Epidendreae were included. Two species, *Neofinetia falcata* (tribe Vandaeae) and *Calanthe triplicata* (tribe Collabieae) were included as outgroups (Table S1). Plastomes were aligned using MAFFT [49] and trimmed using trimAl v1.2 [52] with the gappyout method in order to reduce the systematic errors produced by poor alignment, while an ML analysis was conducted using RAxML [53] available in the CIPRESS Science Gateway web portal (<https://www.phylo.org/portal2/login!input.action>) [54], using the general time-reversible (GTR) with gamma distribution (+G) (=GTR+G) nucleotide substitution model and 1000 bootstrap replicates for each branch node.

Supplementary Materials: Supplementary materials can be found at <http://www.mdpi.com/2223-7747/9/4/521/s1>. Table S1: General information, GenBank accession numbers and plastome sizes of orchid species used in this study; Table S2: List of protein-coding genes used in the calculation of nonsynonymous to synonymous substitution ratio (dN/dS or K_a/K_s); Figure S1: Gene content heatmap of the 111 selected genes of 26 complete plastome sequences derived from the 23 orchid species under the tribe Epidendreae. All authors have read and agreed to the published version of the manuscript.

Author Contributions: Conceptualization, S.Y.L., K.M., W.L.; methodology, S.Y.L.; software, K.M.; validation, S.Y.L. and R.Z.; formal analysis, S.Y.L.; resources, H.W. and F.C.; writing—original draft preparation, S.Y.L.; writing—review and editing, R.Z., W.L. and Q.F.; supervision, W.L., S.Z. and Q.F.; funding acquisition, S.Y.L., W.L., Q.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Shenzhen Key Laboratory of Southern Subtropical Plant Diversity, grant number SYSU-3300-20190702-00, the Fourth National Survey on Material Medical Resources Program from the State Administration of Traditional Chinese Medicine of China, grant numbers 2017-152-003 and 2018-523-001, and the Foundation of Administrative Committee of Danxia Mountain World Natural Heritage and Global Geopark, grant number 20163300071021059.

Acknowledgments: The authors thank the administration of Danxia Mountain National Nature Reserve for granting permission and assistance in specimen collection.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- Govaerts, R.; Bernet, P.; Kratochvil, K.; Gerlach, G.; Carr, G.; Alrich, P.; Pridgeon, A.M.; Pfahl, J.; Campacci, M.A.; Baptista, D.H.; et al. World checklist of Orchidaceae. Available online: <http://apps.kew.org/wcsp/> (accessed on 20 December 2019).
- Merckx, V.S.F.T.; Mennes, C.B.; Peay, K.G.; Geml, J. Evolution and diversification. In *Mycoheterotrophy*; Merckx, V.S.F.T., Ed.; Springer: New York, NY, USA, 2013; pp. 215–244.
- Wolfe, K.H.; Katz-Downie, D.S.; Morden, C.W.; Palmer, J.D. Evolution of the plastid ribosomal RNA operon in a nongreen parasitic plant: Accelerated sequence evolution, altered promoter structure, and tRNA pseudogenes. *Plant Mol. Biol.* **1992**, *18*, 1037–1048. [[CrossRef](#)] [[PubMed](#)]
- Barrett, C.F.; Freudenstein, J.V.; Li, J.; Mayfield-Jones, D.R.; Perez, L.; Pires, J.C.; Santos, C. Investigating the path of plastid genome degradation in an early-transitional clade of heterotrophic orchids, and implications for heterotrophic angiosperms. *Mol. Biol. Evol.* **2014**, *31*, 3095–3112. [[CrossRef](#)] [[PubMed](#)]
- Barrett, C.F.; Wicke, S.; Sass, C. Dense infraspecific sampling reveals rapid and independent trajectories of plastome degradation in a heterotrophic orchid complex. *New Phytol.* **2018**, *218*, 1192–1204. [[CrossRef](#)] [[PubMed](#)]
- Delannoy, E.; Fujii, S.; Colas des Francs-Small, C.; Brundrett, M.; Small, I. Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes. *Mol. Biol. Evol.* **2011**, *28*, 2077–2086. [[CrossRef](#)]
- Kim, Y.K.; Jo, S.; Cheon, S.H.; Joo, M.J.; Hong, J.R.; Kwak, M.H.; Kim, K.J. Extensive losses of photosynthesis genes in the plastome of a mycoheterotrophic orchid, *Cyrtosia septentrionalis* (Vanilloideae: Orchidaceae). *Genome Biol. Evol.* **2019**, *11*, 565–571. [[CrossRef](#)] [[PubMed](#)]
- Logacheva, M.D.; Schelkunov, M.I.; Penin, A.A. Sequencing and analysis of plastid genome in mycoheterotrophic orchid *Neottia nidus-avis*. *Genome Biol. Evol.* **2011**, *3*, 1296–1303. [[CrossRef](#)]
- Barrett, C.F.; Davis, J.I. The plastid genome of the mycoheterotrophic *Corallorhiza striata* (Orchidaceae) is in the relatively early stages of degradation. *Am. J. Bot.* **2012**, *99*, 1513–1523. [[CrossRef](#)]
- Schelkunov, M.I.; Shtratnikova, V.Y.; Nuraliev, M.S.; Selosse, M.A.; Penin, A.A.; Logacheva, M.D. Exploring the limits for reduction of plastid genomes: A case study of the mycoheterotrophic orchids *Epipogium aphyllum* and *Epipogium roseum*. *Genome Biol. Evol.* **2015**, *7*, 1179–1191. [[CrossRef](#)]
- Feng, Y.L.; Wicke, S.; Li, J.W.; Han, Y.; Lin, C.S.; Li, D.Z.; Zhou, T.T.; Huang, W.C.; Huang, L.Q.; Jin, X.H. Lineage-specific reductions of plastid genomes in an orchid tribe with partially and fully mycoheterotrophic species. *Genome Biol. Evol.* **2016**, *8*, 2164–2175. [[CrossRef](#)]
- Barrett, C.F.; Kennedy, A.H. Plastid genome degradation in the endangered, mycoheterotrophic, North American orchid *Hexalectris warnockii*. *Genome Biol. Evol.* **2018**, *10*, 1657–1662. [[CrossRef](#)]
- Huo, X.; Zhao, Y.; Qian, Z.; Liu, M. Characterization of the complete chloroplast genome of *Eulophia zollingeri*, an endangered orchid in China. *Conserv. Genet. Resour.* **2018**, *10*, 817–819. [[CrossRef](#)]
- Kim, H.T.; Shin, C.H.; Sun, H.; Kim, J.H. Sequencing of the plastome in the leafless green mycoheterotroph *Cymbidium macrorhizon* helps us to understand an early stage of fully mycoheterotrophic plastome structure. *Plant Syst. Evol.* **2018**, *304*, 245–258. [[CrossRef](#)]

15. Yuan, Y.; Jin, X.; Liu, J.; Zhao, X.; Zhou, J.; Wang, X.; Wang, D.; Lai, C.; Xu, W.; Huang, J.; et al. The *Gastrodia elata* genome provides insights into plant adaptation to heterotrophy. *Nat. Commun.* **2018**, *9*, 1615. [[CrossRef](#)] [[PubMed](#)]
16. Barrett, C.F.; Sinn, B.T.; Kennedy, A.H. Unprecedented parallel photosynthetic losses in a heterotrophic orchid genus. *Mol. Biol. Evol.* **2019**, *36*, 1884–1901. [[CrossRef](#)]
17. Shevtsov, S.; Murik, O.; Zer, H.; Weinstein, O.; Keren, N.; Fragman-Sapir, O.; Ostersetzer-Biran, O. The complete plastid genome sequence and the photosynthetic activity of the putative mycoheterotrophic orchid *Limodorum abortivum*. *Isr. J. Plant Sci.* **2019**, *66*, 69–88. [[CrossRef](#)]
18. Zhai, J.W.; Zhang, G.Q.; Chen, L.J.; Xiao, X.J.; Liu, K.W.; Tsai, W.C.; Hsiao, Y.Y.; Tian, H.Z.; Zhu, J.Q.; Wang, M.N.; et al. A new orchid genus, *Danxiaorchis*, and phylogenetic analysis of the tribe Calypsoeae. *PLoS ONE* **2013**, *8*, e60371. [[CrossRef](#)]
19. Freudenstein, J.V.; Yukawa, T.; Luo, Y.B. A reanalysis of relationships among Calypsoinae (Orchidaceae: Epidendroideae): Floral and vegetative evolution and the placement of *Yoania*. *Syst. Bot.* **2017**, *42*, 17–26. [[CrossRef](#)]
20. Chase, M.W.; Cameron, K.M.; Barrett, R.L.; Freudenstein, J.V. DNA data and Orchidaceae systematics: A new phylogenetic classification. In *Orchid Conservation*; Dixon, K.W., Kell, S.P., Barrett, R.L., Cribb, P.J., Eds.; Natural History Publications: Sabah, Malaysia, 2003; pp. 69–89.
21. Darling, A.E.; Mau, B.; Perna, N.T. progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* **2010**, *5*, e11147. [[CrossRef](#)]
22. Palmer, J.D. Contrasting modes and tempos of genome evolution in land plant organelles. *Trends Genet.* **1990**, *6*, 115–120. [[CrossRef](#)]
23. Tillich, M.; Lehwark, P.; Pellizzer, T.; Ulbricht-Jones, E.S.; Fischer, A.; Bock, R.; Greiner, S. GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **2017**, *45*, 6–11. [[CrossRef](#)]
24. Lin, C.S.; Chen, J.J.; Chiu, C.C.; Hsiao, H.C.; Yang, C.J.; Jin, X.H.; Leebens-Mack, J.; de Pamphilis, C.W.; Huang, Y.T.; Yang, L.H.; et al. Concomitant loss of NDH complex-related genes within chloroplast and nuclear genomes in some orchids. *Plant J.* **2017**, *90*, 994–1006. [[CrossRef](#)] [[PubMed](#)]
25. Lin, C.S.; Chen, J.J.; Huang, Y.T.; Chan, M.T.; Daniell, H.; Chang, W.J.; Hsu, C.T.; Liao, D.C.; Wu, F.H.; Lin, S.Y.; et al. The location and translocation of *ndh* genes of chloroplast origin in the Orchidaceae family. *Sci. Rep.* **2015**, *5*, 9040. [[CrossRef](#)] [[PubMed](#)]
26. Wicke, S.; Müller, K.F.; de Pamphilis, C.W.; Quandt, D.; Wickett, N.J.; Zhang, Y.; Renner, S.S.; Schneeweiss, G.M. Mechanisms of functional and physical genome reduction in photosynthetic and nonphotosynthetic parasitic plants of the broomrape family. *Plant Cell* **2013**, *25*, 3711–3725. [[CrossRef](#)] [[PubMed](#)]
27. Logacheva, M.D.; Schelkunov, M.I.; Nuraliev, M.S.; Samigullin, T.H.; Penin, A.A. The plastid genome of mycoheterotrophic monocot *Petrosavia stellaris* exhibits both gene losses and multiple rearrangements. *Genome Biol. Evol.* **2014**, *6*, 238–246. [[CrossRef](#)]
28. Lim, G.S.; Barrett, C.F.; Pang, C.C.; Davis, J.I. Drastic reduction of plastome size in the mycoheterotrophic *Thismia tentaculata* relative to that of its autotrophic relative *Tacca chantrieri*. *Am. J. Bot.* **2016**, *103*, 1129–1137. [[CrossRef](#)] [[PubMed](#)]
29. Wicke, S.; Schneeweiss, G.M.; Depamphilis, C.W.; Müller, K.F.; Quandt, D. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol. Biol.* **2011**, *76*, 273–297. [[CrossRef](#)]
30. Arnon, D.I.; Allen, M.B.; Whatley, F.R. Photosynthesis by isolated chloroplasts. *Nature* **1954**, *174*, 394–396.
31. Arnon, D.I.; Whatley, F.R.; Allen, M.B. Triphosphopyridine nucleotide as a catalyst of photosynthetic phosphorylation. *Nature* **1957**, *180*, 182–185.
32. Hill, R.; Bendall, F. Function of the two cytochrome components in chloroplasts: A working hypothesis. *Nature* **1960**, *186*, 136–137.
33. Kamikawa, R.; Tanifuji, G.; Ishikawa, S.A.; Ishii, K.I.; Matsuno, Y.; Onodera, N.T.; Ishida, K.I.; Hashimoto, T.; Miyashita, H.; Mayama, S.; et al. Proposal of a twin arginine translocator system-mediated constraint against loss of ATP synthase genes from nonphotosynthetic plastid genomes. *Mol. Biol. Evol.* **2015**, *32*, 2598–2604. [[CrossRef](#)]
34. Oshima, K.; Kakizawa, S.; Nishigawa, H.; Jung, H.Y.; Wei, W.; Suzuki, S.; Arashida, R.; Nakata, D.; Miyata, S.; Ugaki, M.; et al. Reductive evolution suggested from the complete genome sequence of a plant-pathogenic phytoplasma. *Nat. Genet.* **2004**, *36*, 27–29. [[CrossRef](#)] [[PubMed](#)]

35. Fleischmann, T.T.; Scharff, L.B.; Alkatib, S.; Hasdorf, S.; Schöttler, M.A.; Bock, R. Nonessential plastid-encoded ribosomal proteins in tobacco: A developmental role for plastid translation and implications for reductive genome evolution. *Plant Cell* **2011**, *23*, 3137–3155. [[CrossRef](#)] [[PubMed](#)]
36. Scharff, L.B.; Bock, R. Synthetic biology in plastids. *Plant J.* **2014**, *78*, 783–798. [[CrossRef](#)] [[PubMed](#)]
37. Alkatib, S.; Scharff, L.B.; Rogalski, M.; Fleischmann, T.T.; Matthes, A.; Seeger, S.; Schöttler, M.A.; Ruf, S.; Bock, R. The contributions of wobbling and superwobbling to the reading of the genetic code. *PLoS Genet.* **2012**, *8*, e1003076. [[CrossRef](#)]
38. Kumar, A.M.; Schaub, U.; Söll, D.; Ujwal, M.L. Glutamyl-transfer RNA: At the crossroad between chlorophyll and protein biosynthesis. *Trends Plant Sci.* **1996**, *1*, 371–376. [[CrossRef](#)]
39. Barbrook, A.C.; Howe, C.J.; Purton, S. Why are plastid genomes retained in non-photosynthetic organisms? *Trends Plant Sci.* **2006**, *11*, 101–108. [[CrossRef](#)]
40. Marechal-Drouard, L.; Weil, J.H.; Dietrich, A. Transfer RNAs and transfer RNA genes in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1993**, *44*, 13–32. [[CrossRef](#)]
41. Yang, B.; Xiao, S.; Jiang, Y.; Luo, H.; Xiong, D.; Zhai, J.; Li, B. *Danxiaorchis yangii* sp. nov. (Orchidaceae: Epidendroideae), the second species of *Danxiaorchis*. *Phytotaxa* **2017**, *306*, 287–295. [[CrossRef](#)]
42. Luo, J.; Hou, B.W.; Niu, Z.T.; Liu, W.; Xue, Q.Y.; Ding, X.Y. Comparative chloroplast genomes of photosynthetic orchids: Insights into evolution of the Orchidaceae and development of molecular markers for phylogenetic applications. *PLoS ONE* **2014**, *9*, e99016. [[CrossRef](#)]
43. Huang, J. On the distribution of Danxia landforms in China. *Econ. Geogr.* **1999**, *19*, 31–35.
44. Xu, J.; Jiang, X.L.; Deng, M.; Westwood, M.; Song, Y.G.; Zheng, S.S. Conservation genetics of rare trees restricted to subtropical montane cloud forests in southern China: A case study from *Quercus arbutifolia* (Fagaceae). *Tree Genet. Genomes* **2016**, *12*, 90. [[CrossRef](#)]
45. Li, Y.X.; Li, Z.H.; Schuiteman, A.; Chase, M.W.; Li, J.W.; Huang, W.C.; Hidayat, A.; Wu, S.S.; Jin, X.H. Phylogenomics of Orchidaceae based on plastid and mitochondrial genomes. *Mol. Phylogenet. Evol.* **2019**, *139*, 106540. [[CrossRef](#)] [[PubMed](#)]
46. Doyle, J.J.; Doyle, J.L. Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochem. Bull.* **1987**, *19*, 11–15.
47. Patel, R.K.; Jain, M. NGS QC Toolkit: A toolkit for quality control of next generation sequencing data. *PLoS ONE* **2012**, *7*, e30619. [[CrossRef](#)]
48. Dierckxsens, N.; Mardulyn, P.; Smits, G. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* **2016**, *45*, e18.
49. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [[CrossRef](#)]
50. 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*, w59–w64. [[CrossRef](#)]
51. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [[CrossRef](#)]
52. Capella-Gutierrez, S.; Silla-Martinez, J.M.; Gabaldon, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **2009**, *25*, 1972–1973. [[CrossRef](#)]
53. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [[CrossRef](#)]
54. Miller, M.A.; Pfeiffer, W.; Schwartz, T. Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees. In *Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, LA, USA, 14 November 2010*; Institute of Electrical and Electronics Engineers: Piscataway, NJ, USA, 2010; pp. 1–8.

