Article

Molecular Evolution of Chloroplast Genomes of Orchid Species: Insights into Phylogenetic Relationship and Adaptive Evolution

Wan-Lin Dong [†], Ruo-Nan Wang [†], Na-Yao Zhang, Wei-Bing Fan, Min-Feng Fang and Zhong-Hu Li *

Key Laboratory of Resource Biology and Biotechnology in Western China, Ministry of Education,
College of Life Sciences, Northwest University, Xi'an 710069, China; dongwl@stumail.nwu.edu.cn (W.-L.D.);
wangruonan@stumail.nwu.edu.cn (R.-N.W.); 201620812@stumail.nwu.edu.cn (N.-Y.Z.);
201631689@stumail.nwu.edu.cn (W.-B.F.); fangmf@nwu.edu.cn (M.-F.F.)
* Correspondence: lizhonghu@nwu.edu.cn; Tel./Fax: +86-029-8830-2411
† These two authors contributed equally to this study.

Abstract: Orchidaceae is the 3rd largest family of angiosperms, an evolved young branch of monocotyledons. This family contains a number of economically-important horticulture and flowering plants. However, the limited availability of genomic information largely hindered the study of molecular evolution and phylogeny of Orchidaceae. In this study, we determined the evolutionary characteristics of whole chloroplast (cp) genomes and the phylogenetic relationships of the family Orchidaceae. We firstly characterized the cp genomes of four orchid species: *Cremastra appendiculata*, *Calanthe davidii*, *Epipactis mairei*, and *Platanthera japonica*. The size of the chloroplast genome ranged from 153,629 bp (*C. davidi*) to 160,427 bp (*E. mairei*). The gene order, GC content, and gene compositions are similar to those of other previously-reported angiosperms. We identified that the genes of *ndhC*, *ndhI*, and *ndhK* were lost in *C. appendiculata*, in that the *ndh I* gene was lost in *P. japonica* and *E. mairei*. In addition, the four types of repeats (forward, palindromic, reverse, and complement repeats) were examined in orchid species. *E. mairei* had the highest number of repeats (81), while *C. davidii* had the lowest number (57). The total number of Simple Sequence Repeats is at least 50 in *C. davidii*, and, at most, 78 in *P. japonica*. Interestingly, we identified 16 genes with positive selection sites (the *psbH*, *petD*, *petL*, *rpl22*, *rpl32*, *rpoC1*, *rpoC2*, *rps12*, *rps15*, *rps16*, *accD*, *ccsA*, *rbcL*, *ycf1*, *ycf2*, and *ycf4* genes), which might play an important role in the orchid species' adaptation to diverse environments. Additionally, 11 mutational hotspot regions were determined, including five non-coding regions (*ndhB* intron, *ccsA-ndhD*, *rpl33-rps18*, *ndhE-ndhG*, and *ndhF-rpl32*) and six coding regions (*rps16*, *ndhC*, *rpl32*, *ndhI*, *ndhK*, and *ndhF*). The phylogenetic analysis based on whole cp genomes showed that *C. appendiculata* was closely related to *C. striata* var. *vreelandii*, while *C. davidii* and *C. triplicate* formed a small monophyletic evolutionary clade with a high bootstrap support. In addition, five subfamilies of Orchidaceae, Apostasioideae, Cypripedioideae, Epidendroideae, Orchidoideae, and Vanilloideae, formed a nested evolutionary relationship in the phylogenetic tree. These results provide important insights into the adaptive evolution and phylogeny of Orchidaceae.

Keywords: adaptive variation; chloroplast genome; molecular evolution; Orchidaceae; phylogenetic relationship

1. Introduction

Orchidaceae is the biggest family of monocotyledons and the third largest angiosperm family, containing about five recognized subfamilies (Apostasioideae, Cypripedioideae, Epidendroideae,

Orchidoideae, and Vanilloideae) [1], with over 700 genera and 25,000 species [2–4]. The orchid species are generally distributed in tropical and subtropical regions in the world, while a few species are found in temperate zones. Many orchid species have important ornamental and flowering values, e.g., their flowers are characterized by labella and a column, and they are attractive to humans [5,6]. In recent years, due to the overexploitation and habitat destruction of orchid species, many wild population resources have become rare and endangered [7]. Presently, some scholars have mainly concentrated on the study of Orchidaceae for their morphology and medicinal value, and research on genomes has been relatively scarce [8,9]. Some studies showed that the two subfamilies, Apostasioideae and Cypripedioideae, were clustered into the two respective genetic clades based partial on chloroplast DNA regions and nuclear markers [4,10]. However, the major phylogenetic relationships among the five orchid subfamilies remain unresolved [11].

In recent years, the fast progress of next-generation sequencing technology has provided a good opportunity for the study of genomic evolution and interspecific relationships of organisms based on large-scale genomic dataset resources, such as complete plastid sequences [12,13]. The chloroplast (cp) is made up of multifunctional organelles, playing a critical role in photosynthesis and carbon fixation [5,14–16]. The majority of the cp genomes of angiosperms are circular DNA molecules, ranging from 120 to 160 kb in length, with highly-conserved compositions, in terms of gene content and gene order [17–20]. Generally, the typical cp genome is composed of a large single copy (LSC) region and a small single copy (SSC) region, which are separated by two copies of inverted repeats (IRa/b) [21–23]. Due to its maternal inheritance and conserved structure characteristics [24–27], the cp genomes can provide abundant genetic information for studying species divergence and the interspecific relationships of plants [28–31]. For example, based on complete cp genomes, some studies suggested that *Dactylorhiza viridis* diverged earlier than *Dactylorhiza incarnate* [12]; *Lepanthes* is was distinct from *Pleurothallis* and *Salpistele* [13]. In addition, some researchs based on one nuclear region (*ITS-1*) and five chloroplast DNA fragment variations revealed that *Bolusiella talbotii* and the congeneric *B. iridifolia* were clustered into an earlier diverged lineage [10]. However, up to now, the phylogenetic relationships of some major taxons (e.g., *Cremastra* and *Epipactis*) in the Orchidaceae family remain unclear.

In this study, the complete cp genomes of four orchid species (*Cremastra appendiculata*, *Calanthe davidii*, *Epipactis mairei*, and *Platanthera japonica*) were first assembled and annotated. Following this, we analyzed the differences in genome size, content, and structure, and the inverted repeats (IR) contraction and expansion, identifying the sequence divergence, along with variant hotspot regions and adaptive evolution through combination with other available orchid cp genomes. In addition, we also constructed the evolutionary relationships of the Orchidaceae family, based on the large number of cp genome datasets.

2. Results

2.1. The Chloroplast Genome Structures

In this study, the cp genomes of four species displayed a typical quadripartite structure and similar lengths, containing a pair of inverted repeats IR regions (IRa and IRb), one large single-copy (LSC) region, and one small single-copy (SSC) region (Figure 1, Table 2). The cp genome size ranged from 153,629 bp in *C. davidii* to 160,427 bp in *E. mairei*, with *P. japonica* at 154,995 bp and *C. appendiculata* at 155,320 bp. The length of LSC ranged from 85,979 bp (*P. japonica*) to 88,328 bp (*E. mairei*), while the SSC length and IR length ranged from 13,664 bp (*P. japonica*) to 18,513 bp (*E. mairei*), and from 25,956 bp (*C. davidii*) to 27,676 bp (*P. japonica*). In the four species, the GC contents of the LSC and SSC regions (about 34% and 40%) were lower than those of the IR regions (about 43%) (Table 1). There were 37 tRNA genes and eight rRNA genes that were identified in each orchid cp genome, but there were some differences in terms of protein-coding genes. In *C. davidii*, we annotated 86 protein-coding genes. There were no *ndhC*, *ndhI*, and *ndhK* genes in *C. appendiculata*. In *P. japonica* and *E. mairei*, the *ndhI* gene

was lost (Tables 1 and 2). Fourteen out of the seventeen genes contained a single intron, while three (*clpP*, *ycf3*, and *rps12*) had two introns (Table 2).
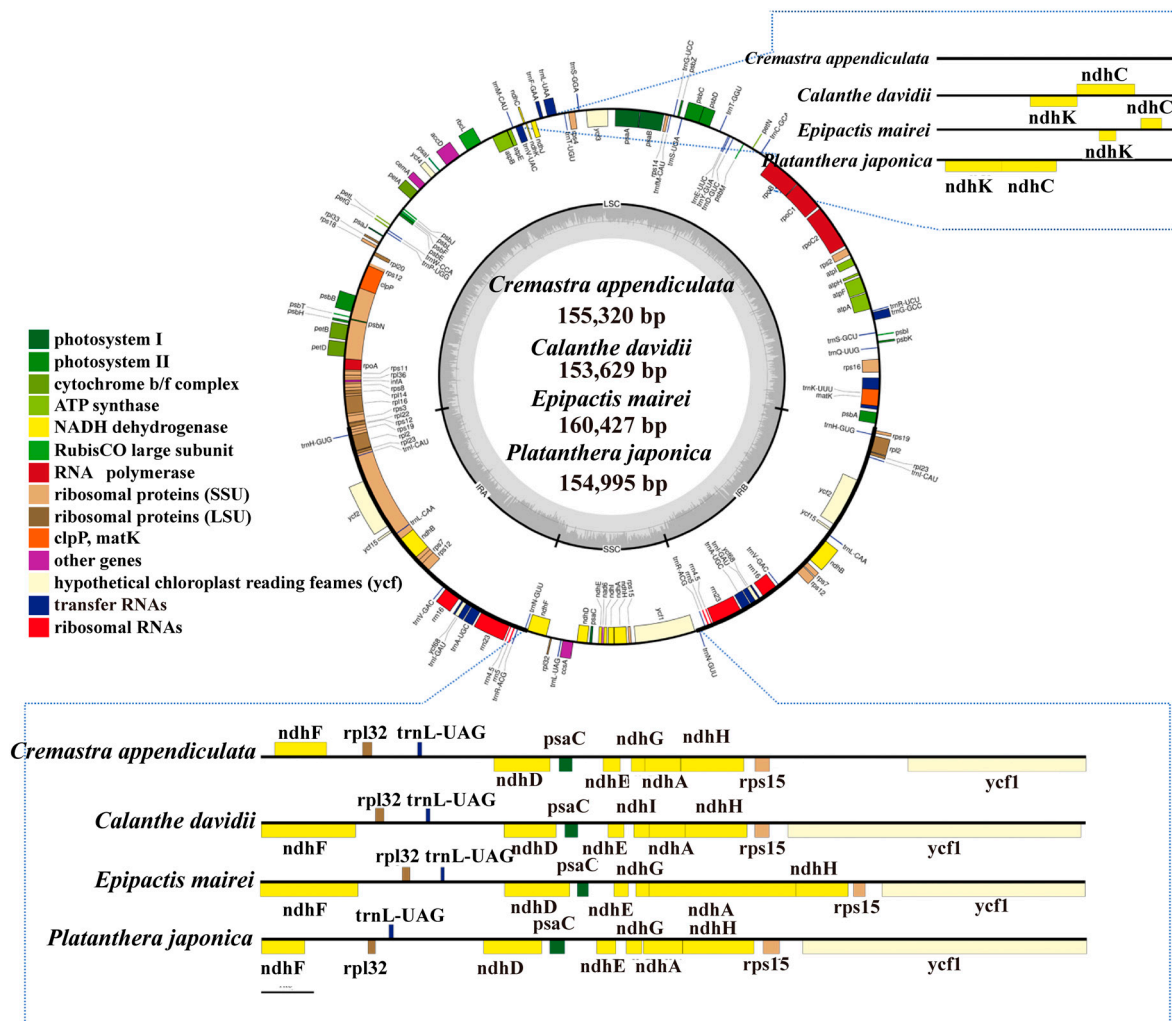
Figure 1. Chloroplast genome maps of the four orchid species. Gene locations outside of the outer rim are transcribed in the counter clockwise direction, whereas genes inside are transcribed in the clockwise direction. The colored bars indicate known different functional groups. The dashed gray area in the inner circle shows the proportional GC content of the corresponding genes. LSC, SSC and IR are large single-copy region, small single-copy region, and inverted repeat region, respectively.

Table 1. Comparison of chloroplast genome features in four orchid species.

Species	*Cremastra appendiculata*	*Calanthe davidii*	*Epipactis mairei*	*Platanthera japonica*
Accession number	MG925366	MG925365	MG925367	MG925368
Genome size (bp)	155,320	153,629	160,427	154,995
LSC length (bp)	87,098	86,045	88,328	85,979
SSC length (bp)	15,478	15,672	18,513	13,664
IR length (bp)	26,372	25,956	26,790	27,676
Coding (bp)	100,018	104,531	113,915	107,028
Non-coding (bp)	55,302	49,098	46,512	47,967
Number of genes	130 (0)	132 (19)	131 (19)	128 (17)
Number of protein-coding genes	83 (7)	86 (7)	85 (7)	85 (7)
Number of tRNA genes	38 (8)	38 (8)	38 (8)	38 (8)
Number of rRNA genes	8 (4)	8 (4)	8 (4)	8 (4)
GC content (%)	37.2	36.9	37.2	37
GC content in LSC (%)	34.5	34.5	34.9	34.2
GC content in SSC (%)	30.4	30.2	31.0	29
GC content in IR (%)	43.5	43.1	43.1	43.2
Mapped read number	551,680	324,741	230,968	322,259
Chloroplast coverage	544.9	217.4	216	313.6

The numbers in parenthesis indicate the genes duplicated in the IR regions.

Table 2. List of genes present in four orchid chloroplast genomes.

| Category of Genes	Group of Gene	Name of Gene	Name of Gene	Name of Gene	Name of Gene	Name of Gene
	Ribosomal RNA genes	*rrn16* [(×2)]	*rrn2* [(×2)]	*rrn4.5* [(×2)]	*rrn5* [(×2)]	
Self-replication	Transfer RNA genes	*trnA-UGC* [*,(×2)] *trnfM-CAU* *trnI-GAU* [*,(×2)] *trnM-CAU* *trnR-UCU* *trnT-UGU*	*trnC-GCA* *trnG-GCC* [*] *trnK-UUU* [*] *trnN-GUU* [(×2)] *trnS-GCU* *trnV-GAC* [(×2)]	*trnD-GUC* *trnG-UCC* *trnL-CAA* [(×2)] *trnP-UGG* *trnS-GGA* *trnV-UAC* [(×2)]	*trnE-UUC* *trnH-GUG* [(×2)] *trnL-UAA* [*] *trnQ-UUG* *trnS-UGA* *trnW-CCA*	*trnF-GAA* *trnI-CAU* [(×2)] *trnL-UAG* *trnR-ACG* [(×2)] *trnT-GGU* *trnY-GUA*
	Small subunit of ribosome	*rps2* *rps11* *rps18*	*rps3* *rps12* [**,(×2)] *rps19* [(×2)]	*rps4* *rps14*	*rps7* [(×2)] *rps15*	*rps8* *rps16* [*]
	Large subunit of ribosome	*rpl2* [*,(×2)] *rpl23* [(×2)]	*rpl14* *rpl32*	*rpl16* [*] *rpl33*	*rpl20* *rpl36*	*rpl22*
	DNA-dependent RNA polymerase	*rpoA*	*rpoB*	*rpoC1* [*]	*rpoC2*	
	Translational initiation factor	*infA*				
Genes for photosynthesis	Subunits of NADH-dehydrogenase	*ndhA* [*] *ndhF* *ndhK* [a]	*ndhB* [*,(×2)] *ndhG*	*ndhC* [a] *ndhH*	*ndhD* *ndhI* [a,c,d]	*ndhE* *ndhJ*
	Subunits of photosystem I	*psaA* *ycf3* [**]	*psaB* *ycf4*	*psaC*	*psaI*	*psaJ*
	Subunits of photosystem II	*psbA* *psbF* *psbL*	*psbB* *psbH* *psbM*	*psbC* *psbI* *psbN*	*psbD* *psbJ* *psbT*	*psbE* *psbK* *psZ*
	Subunits of cytochrome b/f complex	*petA* *petN*	*petB*[*]	*petD* [*]	*petG*	*petL*
	Subunits of ATP synthase	*atp A* *atpI*	*atp B*	*atp E*	*atp F* [*]	*atp H*
	Subunits of rubisco	*rbcL*				
Other genes	Maturase	*matK*				
	Protease	*clpP* [**]				
	Envelope membrane protein	*cemA*				
	Subunit of acetyl-CoA carboxylase	*accD*				
	C-type cytochrome synthesis gene	*ccsA*				
Genes of unknown function	Conserved open reading frames	*ycf1*	*ycf2* [(×2)]			

[a] gene is no in *Cremastra appendiculata*; [c] gene is not in *Epipactis mairei*; [d] gene is not in *Platanthera japonica*; * Gene contains one intron; ** gene contains two introns; [(×2)] indicates that the number of the repeat unit is 2.

2.2. Repeat Structure and Simple Sequence Repeats

Repeats in cp genomes were analyzed using REPuter (Figure 2a and Table S2). *E. mairei* had the greatest number, including 46 forward, 31 palindromic, three reverse repeats, and 1 complement repeat. This was followed by *C. appendiculata* with 43 forward, 33 palindromic, and 2 reverse repeats. *P. japonica* had 42, 21, 1, and 1 forward, palindromic, reverse, and complement repeats. *C. davidii* had the least number, with only 30 forward and 27 palindromic repeats. The comparison analyses revealed that most of the repeats were 30–90 bp, and that the longest repeats, with a length of 309 bp, were detected in the *E. mairei* cp genome (Figure 2b). Most of the repeats were distributed in non-coding regions. There were 9% repeats in coding sequence and intergenic spacer parts (CDS-IGS) in *E. mairei*, but none in *C. appendiculata* (Figure 2c). The highest number of tandem repeats was 53 in *E. mairei*, and the lowest was 29 in *C. davidii* (Table S3). The total number of SSRs was 51 in *C. appendiculata*, 50 in *C. davidii*, 58 in *E. mairei*, and 78 in *P. japonica* (Table S4). Only one six compound, SSR, was found in *C. appendiculata* (Figure 3a). A large proportion of SSRs were found in the LSC region, and we did not identify C/G mononucleotide repeats, while the majority of the dinucleotide repeat sequences were comprised of AT/TA repeats (Figure 3b).
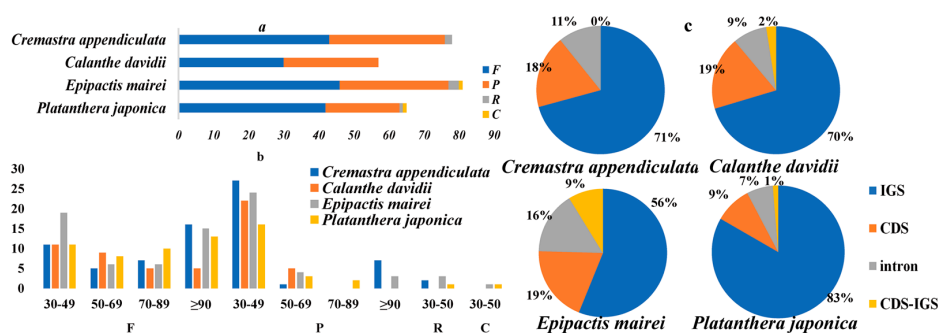
Figure 2. Maps of repeat sequence analyses. Repeat sequences in *C. appendiculata*, *C. davidii*, *E. mairei*, and *P. japonica* chloroplast genomes. (**a**) Number of the four repeat types, F, P, R, and C, indicate the repeat type (F: forward, P: palindrome, R: reverse, and C: complement, respectively). (**b**) Frequency of the four repeat types by length. (**c**) Repeat distribution among four different regions: IGS: intergenic spacer, CDS: coding sequence, intron and CDS-IGS part in CDS and part in IGS.
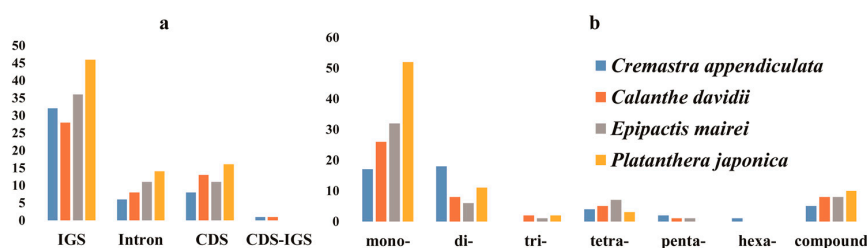
Figure 3. The distribution maps of simple sequence repeats (SSR) in *C. appendiculata*, *C. davidii*, *E. mairei*, and *P. japonica* chloroplast genomes. (**a**) Classification of SSRs in four orchid species. IGS, intergenic spacer; CDS, coding sequence, CDS-IGS, part in CDS and part in IGS. (**b**) Classification of SSRs by repeat type. mono-, mononucleotides; di-, dinucleotides; tri-, trinucleotides; tetra-, tetranucleotides; penta-, pentanucleotides; and hexa-, hexanucleotides.

2.3. IR Contraction and Expansion

We examined the differences between inverted repeat and single-copy (IR/SC) boundary regions among 20 orchid genera, which were classified into several different types (Figure 4). First, the *rps19* gene crossed the large single-copy and inverted repeat b (LSC/IRb) regions within the two parts for eighteen Orchidaceae genera. In *C. crispate* and *C. appendiculata*, the *rps19* gene existed only in the IRb

region. Second, in 12 genera, the *ndhF* gene and the *ycf1* pseudogene overlapped in the IRb/SSC region. In *C. appendiculata* and *Dendrobium strongylanthum*, the *ndhF* gene was complete in the SSC region, 8–35 bp away from the IRb region. In *C. crispate*, *E. pusilla*, and *Phalaenopsis equestris*, the *rpl32* gene was in the SSC region instead of the *ndhF* gene, 280–464 bp away from the IRb region. For the 17 genera mentioned above, the *ycf1* gene crossed the SSC/IRa region. In *C. edavidii* and *Bletilla ochracea*, the *ndhF* gene crossed the IRb/SSC region, and the *ycf1* gene was complete in the SSC region, 101 and 4 bp away from the IRa region. The *trnH-GUG* genes were all located in the LSC region, which was 231 to 1390 bp away from the LSC/IRa boundary. Most specifically, in *Vanilla planifolia*, the *ccsA* gene crossed the IRb/SSC region, as we did not find the *ndhF* and *ycf1* genes where they should be. The SSC/IRa borders were located between the *rpl32* and *ycf1* genes. Thirdly, all 20 genera had the same IRa/LSC borders: the *rps19* gene in the IRa region and the *psbA* gene in the LSC region.

Figure 4. Comparison of the borders of LSC, SSC, and IR regions in 20 orchid complete chloroplast genomes.

2.4. Sequence Divergence and Mutational Hotspot

The whole chloroplast genome sequences of *C. appendiculata*, *C. davidii*, *E. mairei*, and *P. japonica* were compared to 16 other species, using mVISTA [32] (Figures 5 and 6, and Table S5). The comparison analyses showed a high sequence similarity across the cp genomes, with a sequence identity of 82.0%. Interestingly, the proportions of variability in the non-coding regions (introns and intergenic spacers) ranged from 6.77% to 100% with a mean value of 45.97%, i.e., values that are twice as high as in the coding regions (where the range was from 5.80% to 61.76% with a mean value of 24.68%). Five regions within the non-coding regions (*ndhB* intron, *ccsA-ndhD*, *rpl33-rps18*, *ndhE-ndhG*, and *ndhF-rpl32*) and six regions within the coding parts (*rps16*, *ndhC*, *rpl32*, *ndhI*, *ndhK*, and *ndhF*) showed greater levels of variations (percentage of variability >80% and 50%, respectively). In particular, the *ndhB* intron and *ccsA-ndhD* showed a variable percentage of 100%.

In addition, we performed a MAUVE [33] alignment of the 20 orchid chloroplast genomes. The *C. appendiculata* genome is shown at the top as the reference genome (Figure 7). These species maintained a consistent sequence order in most of the genes. However, in *B. ochracea* and *C. faberi*, the *psbM* gene was in front of the *petN*, while the others were upside-down. *Bletilla* and *Cymbidium* actually had the nearest relationship.
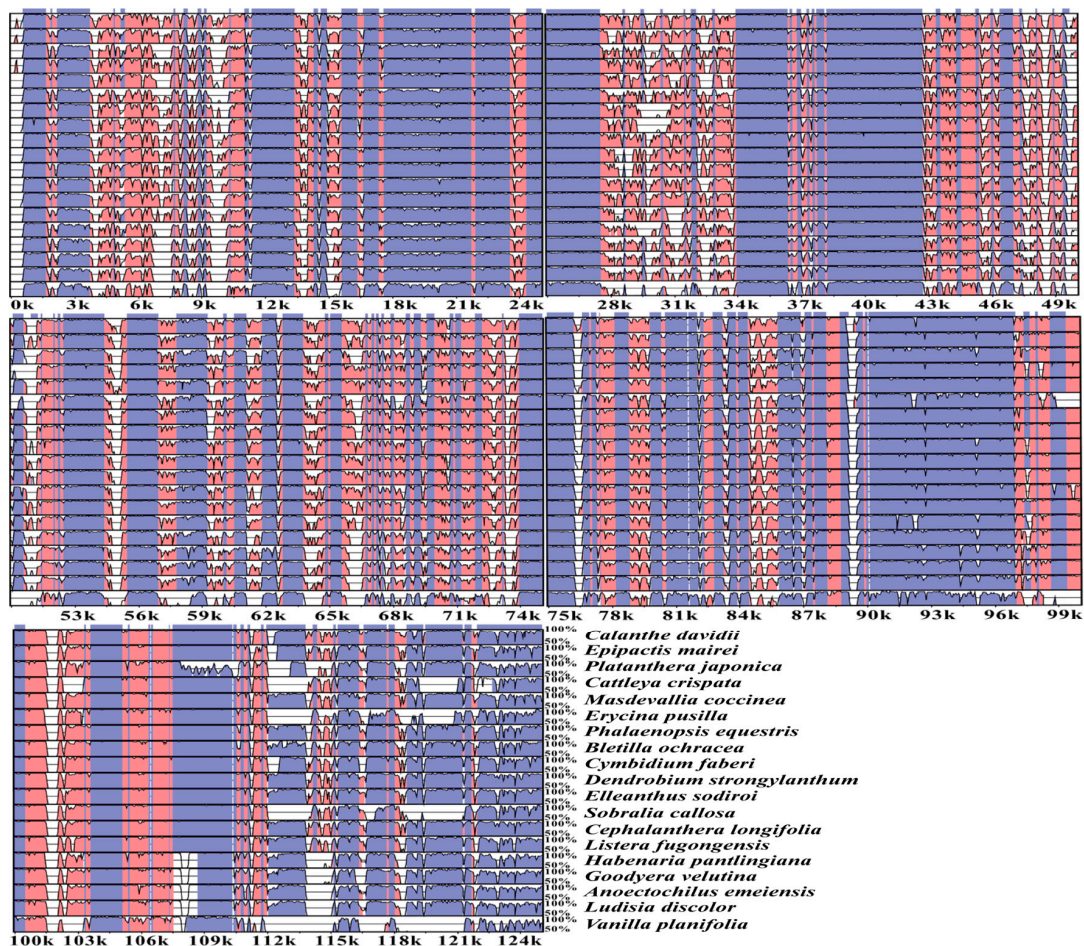
Figure 5. Sequence alignment of chloroplast genomes of 20 orchid species. Sequence identity plot comparing the chloroplast genomes with *C. appendiculata* as a reference using mVISTA. The red color-coded as intergenic spacer regions. The blue color-coded as gene regions. A cut-off of 70% identity was used for the plots, and the *Y*-scale represents the percent identity between 50% and 100%.
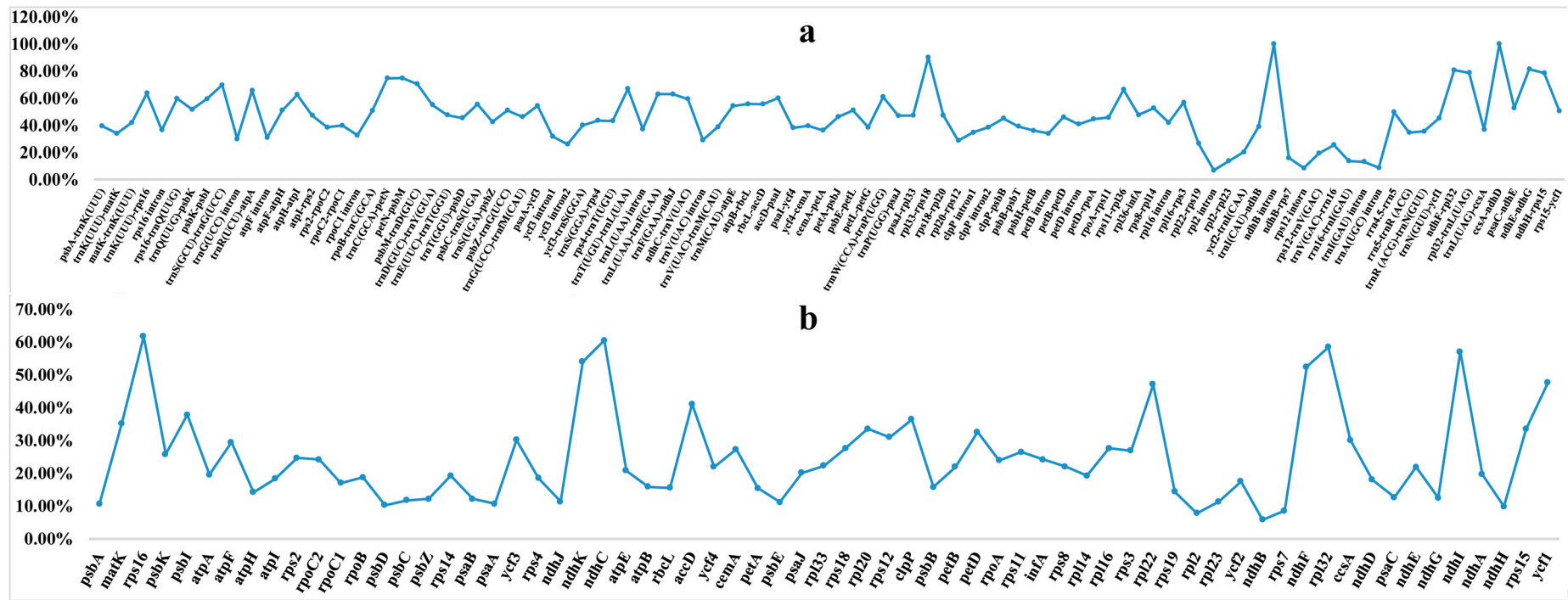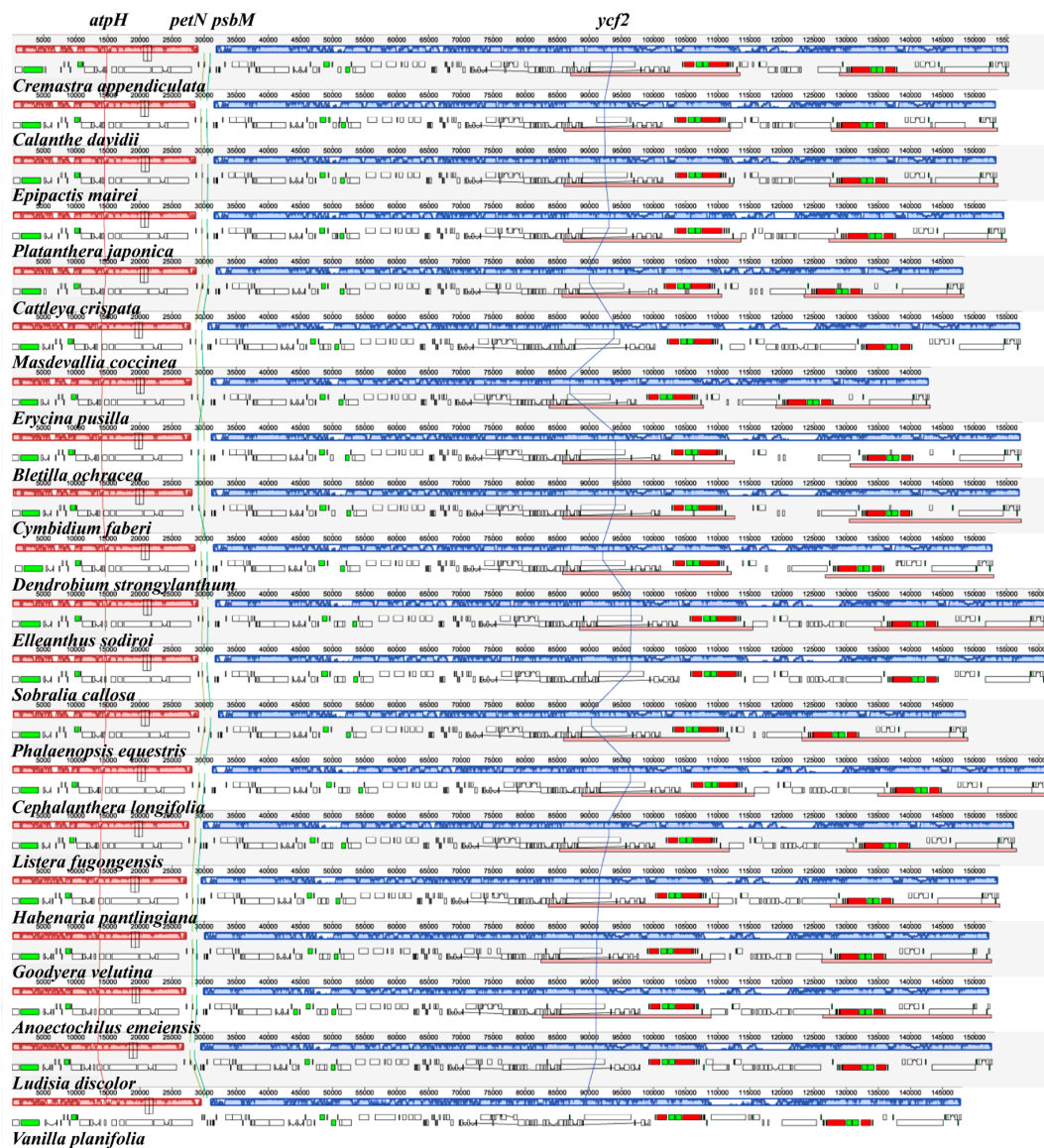
Figure 6. Percentages of variable sites in homologous regions across the 20 orchids with complete chloroplast genomes. (**a**) The introns and spacers (IGS); and (**b**) protein coding sequences (CDS).

Figure 7. MAUVE genome alignments of the 20 orchid chloroplast genomes, with *C. appendiculata* set as a reference genome. The corresponding colored boxes indicate locally-collinear blocks, which present homologous gene clusters. The red vertical line is the location of *atpH* gene. The yellow vertical line is the location of *petN* gene. The green vertical line is the location of *psbM* gene. The blue vertical line is the location of *ycf2* gene.

2.5. Gene Selective Analysis

We compared the rate of nonsynonymous (dN) and synonymous (dS) substitutions for 68 common protein-coding genes between *C. appendiculata*, *C. davidii*, *E. mairei*, and *P. japonica* with 16 other Orchidaceae species (Table S6). Sixteen genes with positive selection sites were identified (Table S7). These genes included one subunit of the photosystem II gene (*psbH*), two genes for cytochrome b/f complex subunit proteins (*petD* and *petL*), two genes for ribosome large subunit proteins (*rpl22* and *rpl32*), two DNA-dependent RNA polymerase genes (*rpoC1* and *rpoC2*), three genes for ribosome small subunit proteins (*rps12*, *rps15*, and *rps16*), and *accD*, *ccsA*, *rbcL*, *ycf1*, *ycf2*, and *ycf4* genes. Interestingly, the *ycf1* gene possesses 13 and 15 positive selective sites, followed by *accD* (8, 10), *rbcL* (4, 7), *ycf2* (2, 3), *rpoC1* (2, 4), *rpoC2* (1, 2), *rpl22* (1, 2), *rps16* (1, 2), *rpl32* (1, 1), *rps12* (1, 1), *ccsA* (0, 2), *petD* (0, 1), *petL* (0, 1), *psbH* (0, 1), and *ycf4* (0, 1). What is more, the likelihood ratio tests (LRTs) of variables under different

models were compared in the site-specific models, M0 vs. M3, M1 vs. M2 and M7 vs. M8, in order to support the sites under positive selection (*p* < 0.01) (Table S7).

2.6. Phylogenetic Relationship

In this study, the maximum likelihood (ML) analysis suggested that *C. appendiculata* and the congeneric *C. davidii* clustered into the Epidendroideae subfamily clade with high bootstrap support, and that *E. mairei* and *P. japonica* clustered into Orchidoideae subfamily (Figure 8). Interestingly, five subfamilies of Orchidaceae, Apostasioideae, Cypripedioideae, Epidendroideae, Orchidoideae, and Vanilloideae have a nested evolutionary relationship in the ML tree. Meanwhile, *C. appendiculata* was closely-related to *C. striata* var. *vreelandii*, *C. davidii*, and *C. triplicate*, which formed a small evolutionary clade with a high bootstrap. *P. japonica* and *Habenaria pantlingiana* had a relatively-closer affinity in the Orchidoideae subfamily.
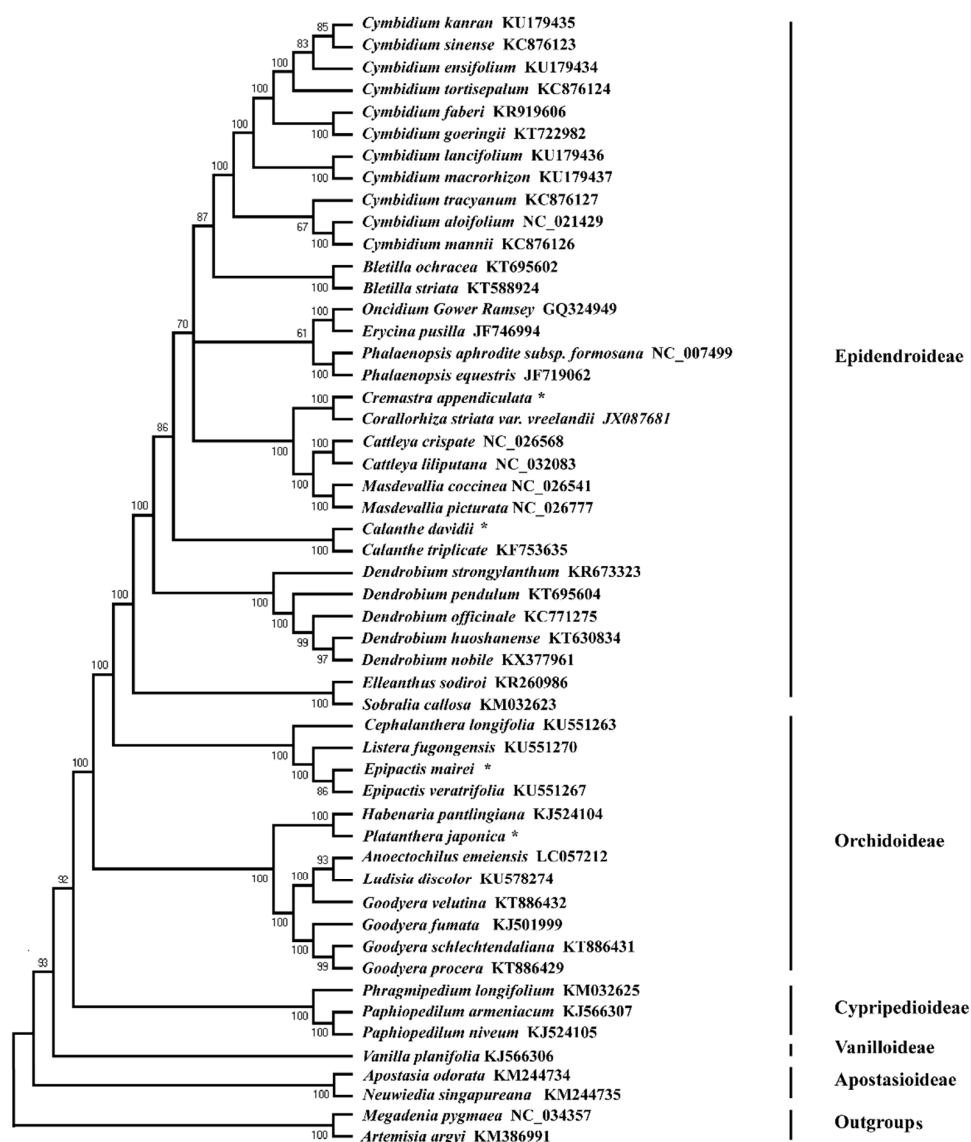
Figure 8. Cladogram of 54 nucleotide sequences of complete chloroplast genomes of orchid species based on the GTRGAMMA model with maximum likelihood (ML) analysis. * The newly generated chloroplast genomes of orchid species.

3. Discussion

3.1. Sequence Variation

In this study, we first determined the whole chloroplast genomes of four orchid species. The cp genome size of *C. davidii* was shorter than that of others, which might be the result of the expansion and contraction of the border positions between the IR and SC regions [21–23]. In addition, the GC contents of the LSC and SSC regions in all the orchid species were much lower than those of the IR regions, which possibly resulted from four rRNA genes (*rrn16*, *rrn23*, *rrn4.5*, and *rrn5*) sequences in the IR regions. In addition, we identified some obvious differences in the protein-coding genes for the orchid chloroplast genomes, despite that the cp genomes of land plants are generally considered to be highly conserved [34]. Interestingly, there were no *ndhC*, *ndhI*, and *ndhK* genes in *C. appendiculata*. In *P. japonica* and *E. mairei*, the *ndhI* gene was lost. Previous studies also found that some orchid species had lost the *ndh* gene, which encodes the subunits of the nicotinamide-adenine dinucleotid (NADH) dehydrogenase-like complex proteins [35–37]. The loss of this gene might have hindered cyclic electron flow around photosystem I and affected the plant photosynthesizing [35–38]. In addition, some studies suggested that the different Orchidaceae species harbored a variable loss or retention of *ndh* genes [35]. For example, *Cymbidium* has the *ndhE*, *ndhJ*, and *ndhC* genes [39], and *Oncidium* has the *ndhB* gene [40]. Nevertheless, the mechanisms that underlie the variable loss or retention of *ndh* genes in orchids remain unclear [11,41].

In addition, we identified 233 SSRs in four orchid species (*C. appendiculata*, *C. davidii*, *E. mairei*, and *P. japonica*); 77.68% of SSRs were distributed in the IGS and intron regions. Generally, microsatellites consist of 1–6 nucleotide repeat units, which are widely distributed across the entire genome and have a great influence on genome recombination and rearrangement [42,43]. The large amount of SSRs also have been identified in *Forsythia suspense* [44], *Dendrobium nobile*, *Dendrobium officinale*, and so on [45]. The majority of these SSRs consisted of mono- and di-nucleotide repeats. Tri-, tetra-, and penta-nucleotide repeat sequences were detected at a much lower frequency in these orchid species and in other organisms [46,47].

Meanwhile, our analyses revealed that the mutational hotspots among orchid genera were highly variable. A diversity of IR contraction and expansion, along with the high level of mutational hotspots, revealed that Orchidaceae had experienced a complex evolution process. Interestingly, in the orchid species, the two IR regions were less divergent than the LSC and SSC regions. Five regions within the non-coding regions (*ndhB* intron, *ccsA-ndhD*, *rpl33-rps18*, *ndhE-ndhG*, and *ndhF-rpl32*) and six regions within the coding regions (*rps16*, *ndhC*, *rpl32*, *ndhI*, *ndhK*, and *ndhF*) showed greater levels of variations (percentage of variability >80% and 50%, respectively). These regions can be used as potential DNA barcodes for the further study of phylogenetic relationships, species identification, and population genetics.

3.2. Adaptive Evolution

We used the site-specific model (seqtype = 1, model = 0, NSsites = 0, 1, 2, 3, 7, 8), one of the codon substitution models, to estimate the selection pressure [48]. Sixteen genes with positive selection sites were identified in these orchid species. These genes included one subunit of the photosystem II gene (*psbH*), two genes for cytochrome b/f complex subunit proteins (*petD* and *petL*), two genes for ribosome large subunit proteins (*rpl22* and *rpl32*), two DNA-dependent RNA polymerase genes (*rpoC1* and *rpoC2*), three genes for ribosome small subunit proteins (*rps12*, *rps15*, and *rps16*), and *accD*, *ccsA*, *rbcL*, *ycf1*, *ycf2*, and *ycf4* genes. We found that the genes with positive selection sites can be divided into four categories: Subunits of photosystem (*psbH* and *ycf4*), subunits of cytochrome (*petD*, *petL*, and *ccsA*), subunits of ribosome (*rpl22*, *rpl32*, *rps12*, *rps15*, and *rps16*) and others (*rpoC1*, *rpoC2*, *accD*, *rbcL*, *ycf1*, and *ycf2*). The plastid *accD* gene, which encodes the β-carboxyl transferase subunit of acetyl-CoA carboxylase, is an essential and required component for plant leaf development [49–53]. In this study, 10 positively-selected sites were identified in *accD* genes for orchid species, suggesting that the *accD*

gene played a possible pivotal role in the adaptive evolution of orchids. What is more, the *ycf1* gene is also essential for almost all plant lineages [5,54], except for Gramineae, which lost the *ycf1* gene in its cp genomes [55]. Additionally, *ycf1* is one of the largest cp genes, encoding a component of the chloroplast's inner envelope membrane protein translocon [56]. This gene, which is also highly variable in terms of phylogenetic information at the level of species, has also been shown to be subject to positive selection with 15 sites, as has also been identified in many plant lineages [57–59]. In addition, we found that the *rbcL* gene possessed seven sites under positive selection in orchid species. Generally, *rbcL* is the gene for the Rubisco large subunit protein, and as the result of enzymatic activity of Rubisco, which is an important component as a modulator of photosynthetic electron transport [60,61]. Current research has revealed that positive selection of the *rbcL* gene in land plants may be a common phenomenon [62]. Additionally, the *rbcL* gene is also widely used in the phylogenetic analysis of land plants [63]. In conclusion, these results showed that multiple factors, several of them interconnected (positive selection, heterogeneity environments), have possibly contributed to orchid diversification and adaptation. For example, some positively-selected sites that were identified (e.g., *rbcL*, *ycf1*, and *accD*) were associated in a significant manner with environment adaptation, including factors such as temperature, light, humidity, and atmosphere [49]. Additionally, epiphytism in orchid species is a key innovation which should help generate and maintain high levels of plant diversity. On the other hand, the tropical distributions of orchid species might have increased the rates of speciation relative to those outside of the tropics as a result of more stable climates (e.g., the lack of glaciation and suitable temperatures), the greater habitat area, and together, this possibly provided a greater opportunity for the co-evolution of plants and their mutualists, and for greater adaptation [49,58,59].

3.3. Phylogenetic Relationship

In this study, the maximum likelihood (ML) tree obtained high bootstrap support values, which had 33 nodes with 100% bootstrap support, with 36 of the 46 nodes having values ≥95%. The phylogenetic analyses based on complete cp genomes, suggested that five subfamilies of Orchidaceae (Apostasioideae, Cypripedioideae, Epidendroideae, Orchidoideae, and Vanilloideae) have a nested evolutionary relationship (Figure 8). Apostasioideae is the earliest diverging subfamily of orchids. Some recent molecular studies have shown that the five subfamilies had formed their respective five monophylies [11,41,49]. The generic relationships of the five subfamilies found in our analyses are basically congruent with those of recent studies. However, our finding that Orchidoideae is a nested subfamily is different from the studies of Kim et al. [41] and Givnish et al. [49]. They reconstructed ML trees using the concatenated coding sequences of plastid genes, resulting in large amounts of missing data for these orchid taxa. In this study, we sampled these newly orchid species (*C. longifolia*, *L. fugongensis*, *E. mairei*, and *E. veratrifolia*) to construct a more widespread Orchidaceae phylogenetic tree, through which we obtained the different species relationships. However, some molecular phylogenetic studies, to date, have failed to identify the placement of Cypripedioideae and Vanilloideae [8,25,64,65]. Recently, Givnish et al. [49] and Niu et al. [11] reconstructed ML trees from 39 and 53 orchids species, respectively, using the sequence variations in 75 genes and 67 genes from the plastid genomes. Their results showed that five orchid subfamilies clustered into the five monophyletic clades: Epidendroideae–Orchidoideae–Cypripedioideae–Vanilloideae–Apostasioideae. However, the current study found that *C. appendiculata* and the congeneric *C. davidii* clustered into the Epidendroideae subfamily clade, and that *E. mairei* and *P. japonica* clustered into the Orchidoideae subfamily. These results were largely consistent with traditional morphological evidence [66–68]. However, the inconsistent phylogenetic relationships for the five subfamilies may be due to the differences in the collected samples used in different studies [11,49,64,65], which need to be further explored by sampling a much higher number of orchid species.

4. Materials and Methods

4.1. Plant Material, DNA Extraction, Library Construction, and Sequencing

Fresh leaf tissues were collected from *Cremastra appendiculata*, *Calanthe davidii*, *Epipactis mairei*, and *Platanthera japonica* in the Qinling Mountains, Shaanxi Province, China. The leaves were cleaned and preserved in a −80 °C refrigerator at Northwest University. The voucher specimens of the four species materials were deposited into the Northwest University Herbarium (NUH). The total genomic DNA was isolated using the modified Cetyltrimethyl Ammonium Bromide (CTAB) method [69], which added the EDTA buffer (Amresco, Washington, DC, USA) (1.0 mol/L Tris-HCl (Amresco, Washington, DC, USA) (pH 8.0), 0.5 mol/L EDTA-Na$_2$ (Amresco, Washington, DC, USA), 5.0 mol/L NaCl) solution before isolating the high-quality DNA with the CTAB solution (1.0 mol/L Tris-HCl (pH 8.0), 0.5 mol/L EDTA-Na$_2$, 2% CTAB). Following this, we constructed a pair-end (PE) library with 350 bp insert size fragments using TruSeq DNA sample preparation kits (Sangon, Shanghai, China). Subsequently, we sequenced at least 4.5 GB of clean data for each orchid species. The detailed next-generation sequencing was conducted on the Illumina Hiseq 2500 platform by Sangon Biotech (Shanghai, China).

4.2. Chloroplast Genome Assembly and Annotation

First, we used the software, NGSQCToolkit v2.3.3 [70], to trim the low-quality reads. After removing the low-quality sequences, the clean reads were assembled using MIRA v4.0.2 [70] and MITObim v1.8 [71] with the cp genome of a closely-related species, *Dendrobium nobile* (KX377961), as reference. The programs, DOGMA (http://dogma.ccbb.utexas.edu/) [72] and Geneious v8.0.2 [73] were used to annotate the chloroplast genome. Finally, we obtained four high-quality, complete chloroplast genome sequences. The Circle maps of the four species were drawn using OGDRAW v1.1 [74].

4.3. Repeat Sequence Analyses

The REPuter program (Available online: https://bibiserv.cebitec.uni-bielefeld.de/reputer/manual.html) was used to identify repeats, including forward, reverse, palindrome, and complement sequences. The maximum computed repeats and the minimal repeat size were limited to 50 and 30, respectively, with a Hamming distance equal to 3 [75]. The tandem repeats finder welcome page (http://tandem.bu.edu/trf/trf.html) was used to identify tandem repeats sequences [76]. The alignment parameters match, mismatch, and indels, were 2, 7, and 7, respectively. The minimum alignment score to report repeat, maximum period size and maximum TR array Size (bp, millions) are limited to 80, 500, and 2, respectively. A Perl script MISA (MIcroSAtellite identification tool, http://pgrc.ipk-gatersleben.de/misa/) was used to search for simple sequence repeat (SSR or microsatellite) loci in the chloroplast genomes [77]. Tandem repeats of 1–6 nucleotides were viewed as microsatellites. The minimum number of repeats were set to 10, 5, 4, 3, 3, and 3, for mono-, di-, tri-, tetra-, penta-, and hexanucleotides, respectively.

4.4. Genome Structure and Mutational Hotspot

In order to compare the genome structures and divergence hotspots in a broad manner, we used 16 cp genomes (available in Genbank https://www.ncbi.nlm.nih.gov/) representing each orchid genus, and added the four newly-sequenced ones (Table 3). The boundaries between the IR and SC regions of *C. appendiculata*, *C. davidii*, *E. mairei*, and *P. japonica* and other 16 sequences were compared and analyzed. Meanwhile, the whole-genome alignment of the chloroplast genomes of the 20 species were performed and plotted using the mVISTA program [32]. Following this, we selected the regions within non-coding and coding regions that had a greater level of variation (percentage of variability >80% and 50%, respectively) as mutational hotspots. The formula was as follows: percentage of

variable = (number of nucleotide substitutions + the number of indels)/(the length of aligned sites−the length of indels + the number of indels) × 100%.

Table 3. List of taxa sampled in the study and species accessions numbers (GenBank).

Subfamily	Species	Accession Number
Orchidaceae subfamily. Epidendroideae	*Cattleya crispata*	KP168671
	Cremastra appendiculata	MG925366
	Masdevallia coccinea	KP205432
	Erycina pusilla	JF746994
	Phalaenopsis equestris	JF719062
	Bletilla ochracea	KT695602
	Cymbidium faberi	KR919606
	Calanthe davidii	MG925365
	Dendrobium strongylanthum	KR673323
	Elleanthus sodiroi	KR260986
	Sobralia callosa	KM032623
Orchidaceae subfamily. Orchidoideae	*Epipactis mairei*	MG925367
	Cephalanthera longifolia	KU551263
	Listera fugongensis	KU551270
	Platanthera japonica	MG925368
	Habenaria pantlingiana	KJ524104
	Goodyera velutina	KT886432
	Anoectochilus emeiensis	LC057212
	Ludisia discolor	KU578274
Orchidaceae subfamily. Vanilloideae	*Vanilla planifolia*	KJ566306

4.5. Gene Selective Pressure Analysis

The codon substitution models in the Codeml program, PAML3.15 [46] were used for calculating the non-synonymous (dN) and synonymous (dS) substitution rates, along with their ratios (ω = dN/dS). We analyzed all CDS gene regions, except *ndh*, due to there being too many losses there. These unique CDS gene sequences were separately extracted and aligned using Geneious v8.0.2 [73]. A maximum likelihood phylogenetic tree was built based on the complete cp genomes of the 20 species using RAxML [78]. We used the site-specific model (seqtype = 1, model = 0, NSsites = 0, 1, 2, 3, 7, 8) to estimate the selection pressure [79]. This model allowed the ω ratio to vary among sites, with a fixed ω ratio in all the branches. Comparing the site-specific model, M1 (nearly neutral) vs. M2 (positive selection), M7 (β) vs. M8 (β and ω) and M0 (one-ratio) vs. M3 (discrete) were calculated in order to detect positive selection [79].

4.6. Phylogenetic Analysis

In order to deeply detect the evolutionary relationship of the Orchidaceae family, 50 available complete chloroplast genomes were downloaded from the NCBI Organelle Genome Resources database (Table S1). In addition, *Artemisia argyi* and *Megadenia pygmaea* were used as outgroups. In total, 54 nucleotide sequences of complete chloroplast genomes were aligned using MAFFT [73]; the detailed parameters were as follows: 200 PAM/K = 2 and 1.53 gap open penalty [73]. The choice of the best nucleotide sequence substitution model (GTRGAMMA model) was determined using the Modeltest v3.7 [80]. We constructed a maximum likelihood phylogenetic tree based on these complete plastomes using MAGA7 [34] with 1000 bootstrap replicates under the GTRGAMMA model [80].

References

1. 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: Kota Kinabalu, Malaysia, 2003; pp. 69–89.

2. Dressler, R.L. *The Orchids: Natural History and Classification*; Harvard University Press: Cambridge, MA, USA, 1990.

3. Chase, M.W. Classification of Orchidaceae in the age of DNA data. *Curtis's Bot. Mag.* **2005**, *22*, 2–7. [CrossRef]

4. 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] [PubMed]

5. Raubeson, L.A.; Jansen, R.K. Chloroplast genomes of plants. In *Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants*; Henry, R.J., Ed.; CAB International: Wallingford, UK, 2005; pp. 45–68.

6. Van den Berg, C.; Goldman, D.H.; Freudenstein, J.V.; Pridgeon, A.M.; Cameron, K.M.; Chase, M.W. An overview of the phylogenetic relationships within Epidendroideae inferred from multiple DNA regions and recircumscription of Epidendreae and Arethuseae (Orchidaceae). *Am. J. Bot.* **2005**, *92*, 13–24. [CrossRef] [PubMed]

7. Mendonca, M.P.; Lins, L.V. *Revisao das Listas das Especies da Flora eda Fauna Ameaçadas de Extincao do Estado de Minas Gerais*; Fundacao Biodiversitas: BeloHorizonte, Brazil, 2007.

8. Cameron, K.M.; Chase, M.W.; Whitten, W.M.; Kores, P.J.; Jarrell, D.C.; Albert, V.A.; Yukawa, T.; Hills, H.G.; Goldman, D.H. A phylogenetic analysis of the Orchidaceae: Evidence from *rbcL* nucleotide. *Am. J. Bot.* **1999**, *86*, 8–24. [CrossRef]

9. Van den Berg, C.; Higgins, W.E.; Dressler, R.L.; Whitten, W.M.; Soto-Arenas, M.; Chase, M.W. A phylogenetic study of *Laeliinae* (Orchidaceae) based on combined nuclear and plastid DNA sequences. *Ann. Bot.* **2009**, *104*, 17–30. [CrossRef] [PubMed]

10. Verlynde, S.; D'Haese, C.A.; Plunkett, G.M.; Simo-Droissart, M.; Edwards, M.; Droissart, V.; Stévart, T. Molecular phylogeny of the genus *Bolusiella* (Orchidaceae, Angraecinae). *Plant Syst. Evol.* **2017**, *304*, 269–279. [CrossRef]

11. Niu, Z.T.; Xue, Q.Y.; Zhu, S.Y.; Sun, J.; Liu, W.; Ding, X.Y. 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*, 1–11. [CrossRef] [PubMed]

12. Bateman, R.M.; Rudall, P.J. Clarified relationship between *Dactylorhiza viridis* and *Dactylorhiza iberica* renders obsolete the former genus *Coeloglossum* (Orchidaceae: Orchidinae). *Kew Bull.* **2018**, *73*, 1–17. [CrossRef]

13. Wilson, M.; Frank, G.S.; Lou, J.; Pridgeon, A.M.; Vieira-Uribe, S.; Karremans, A.P. Phylogenetic analysis of *Andinia* (Pleurothallidinae; Orchidaceae) and a systematic re-circumscription of the genus. *Phytotaxa* **2017**, *295*, 101–131. [CrossRef]

14. Neuhaus, H.E.; Emes, M.J. Nonphotosynthetic metabolism in plastids. *Annu. Rev. Plant Biol.* **2000**, *51*, 111–140. [CrossRef] [PubMed]

15. Rodríguezezpeleta, N.; Brinkmann, H.; Burey, S.C.; Roure, B.; Burger, G.; Löffelhardt, W.; Bohnert, H.J.; Philippe, H.; Lang, B.F. Monophyly of primary photosynthetic eukaryotes: Green plants, red algae, and glaucophytes. *Curr. Biol.* **2005**, *15*, 1325–1330. [CrossRef] [PubMed]

16. Yap, J.Y.; Rohner, T.; Greenfield, A.; Van Der Merwe, M.; McPherson, H.; Glenn, W.; Kornfeld, G.; Marendy, E.; Pan, A.Y.; Wilton, A.; et al. Complete chloroplast genome of the Wollemi pine (*Wollemia nobilis*): Structure and evolution. *PLoS ONE* **2015**, *106*, 126–128. [CrossRef] [PubMed]

17. Wicke, S.; Schneeweiss, G.M.; 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] [PubMed]

18. Dong, W.; Liu, H.; Xu, C.; Zuo, Y.J.; Chen, Z.J.; Zhou, S.L. A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: A case study on ginsengs. *BMC Genet.* **2014**, *15*, 138–145. [CrossRef] [PubMed]

19. Zhang, Y.; Li, L.; Yan, T.L.; Liu, Q. Complete chloroplast genome sequences of Praxelis (*Eupatorium catarium* Veldkamp), an important invasive species. *Gene* **2014**, *549*, 58–69. [CrossRef] [PubMed]

20. Xu, C.; Dong, W.P.; Li, W.Q.; Lu, Y.Z.; Xie, X.M.; Jin, X.B.; Shi, J.; He, K.; Suo, Z. Comparative analysis of six *Lagerstroemia* complete chloroplast genomes. *Front. Plant Sci.* **2017**, *8*, 15–26. [CrossRef] [PubMed]

21. Jer, J.D. Plastid chromosomes: Structure and evolution. In *Cell Culture and Somatic Cell Genetics in Plants, the Molecular Biology of Plastids 7A*; Vasil, I.K., Bogorad, L., Eds.; Academic Press: San Diego, CA, USA, 1991; pp. 5–53.

22. Bendich, A.J. Circular chloroplast chromosomes: The grand illusion. *Plant Cell* **2004**, *16*, 1661–1666. [CrossRef] [PubMed]

23. Jansen, R.K.; Raubeson, L.A.; Boore, J.L.; Pamphilis, C.W.; Chumley, T.W.; Haberle, R.C.; Wyman, S.K.; Alverson, A.J.; Peery, R.; Herman, S.J.; et al. Methods for obtaining and analyzing whole chloroplast genome sequences. *Methods Enzymol.* **2015**, *395*, 348–384.

24. Burke, S.V.; Grennan, C.P.; Duvall, M.R. Plastome sequences of two new world bamboos-*Arundinaria gigantea* and *Cryptochloa strictiflora* (Poaceae)-extend phylogenomic understanding of Bambusoideae. *Am. J. Bot.* **2012**, *99*, 1951–1961. [CrossRef] [PubMed]

25. Civan, P.; Foster, P.G.; Embley, M.T.; Séneca, A.; Cox, C.J. Analyses of charophyte chloroplast genomes help characterize the ancestral chloroplast genome of land plants. *Genome Biol. Evol.* **2014**, *6*, 897–911. [CrossRef] [PubMed]

26. Guo, W.; Grewe, F.; Cobo-Clark, A.; Fan, W.; Duan, Z.; Adams, R.P.; Schwarzbach, A.E.; Mower, J.P. 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*, 580–590. [CrossRef] [PubMed]

27. Ruhfel, B.R.; Gitzendanner, M.A.; Soltis, P.S.; Soltis, D.E.; Burleigh, J.G. From algae to angiosperms-inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evol. Biol.* **2014**, *14*, 385–399. [CrossRef] [PubMed]

28. Moore, M.J.; Bell, C.D.; Soltis, P.S.; Soltis, D.E. Using plastid genome-scale data to resolve enigmatic relation-ships among basal angiosperms. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19363–19368. [CrossRef] [PubMed]

29. Huang, H.; Shi, C.; Liu, Y.; Mao, S.Y.; Gao, L.Z. Thirteen Camellia chloroplast genome sequences determined by high-throughput sequencing: Genome structure and phylogenetic relationships. *BMC Evol. Biol.* **2014**, *14*, 4302–4315. [CrossRef] [PubMed]

30. Walker, J.F.; Zanis, M.J.; Emery, N.C. Comparative analysis of complete chloroplast genome sequence and inversion variation in *Lasthenia burkei* (Madieae, Asteraceae). *Am. J. Bot.* **2014**, *101*, 722–729. [CrossRef] [PubMed]

31. Oldenburg, D.J.; Bendich, A.J. The linear plastid chromosomes of maize: Terminal sequences, structures, and implications for DNA replication. *Curr. Genet.* **2015**, *62*, 1–12. [CrossRef] [PubMed]

32. Frazer, K.A.; Pachter, L.; Poliakov, A.; Rubin, E.M.; Dubchak, I. VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* **2004**, *32*, 273–279. [CrossRef] [PubMed]

33. Doose, D.; Grand, C.; Lesire, C. MAUVE Runtime: A Component-Based Middleware to Reconfigure Software Architectures in Real-Time. In Proceedings of the IEEE International Conference on Robotic Computing (IRC), Taichung, Taiwan, 10–12 April 2017; pp. 208–211.

34. Kumar, S.; Stecher, G.; Tamura, K. Mega7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [CrossRef] [PubMed]

35. Clegg, M.T.; Gaut, B.S.; Learn, G.H., Jr.; Morton, B.R. Rates and patterns of chloroplast DNA evolution. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 6795–6801. [CrossRef] [PubMed]

36. Delannoy, E.; Fujii, S.; Colas des Francs-Small, C.; Brundrett, M.S. Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes. *Mol. Biol. Evol.* **2011**, *28*, 2077–2086. [CrossRef] [PubMed]

37. 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] [PubMed]

38. 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] [PubMed]

39. 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*, 84. [CrossRef] [PubMed]

40. Wu, F.H.; Chan, M.T.; Liao, D.C.; Hsu, C.T.; Lee, Y.W.; Daniell, H.; Duvall, M.R.; Lin, C.S. Complete chloroplast genome of *Oncidium Gower Ramsey* and evaluation of molecular markers for identification and breeding in Oncidiinae. *BMC Plant Biol.* **2010**, *10*, 68. [CrossRef] [PubMed]

41. Kim, H.T.; Kim, J.S.; Moore, M.J.; Neubig, K.M.; Williams, N.H.; Whitten, W.M.; Kim, J.H. Seven new complete plastome sequences reveal rampant independent loss of the *ndh* gene family across orchids and associatedinstability of the inverted repeat/small single-copy region boundaries. *PLoS ONE* **2015**, *10*, e0142215.

42. Ni, L.; Zhao, Z.; Xu, H.; Chen, S.; Dorje, G. The complete chloroplast genome of *Gentiana straminea* (Gentianaceae), an endemic species to the Sino-Himalayan subregion. *Gene* **2016**, *577*, 281–288. [CrossRef] [PubMed]

43. Ni, L.; Zhao, Z.; Xu, H.; Chen, S.; Dorje, G. Chloroplast genome structures in *Gentiana* (Gentianaceae), based on three medicinal alpine plants used in Tibetan herbal medicine. *Curr. Genet.* **2017**, *63*, 241–252. [CrossRef] [PubMed]

44. Wang, W.B.; Yu, H.; Wang, J.H.; Lei, W.J.; Gao, J.H.; Qiu, X.P.; Wang, J.S. The complete chloroplast genome sequences of the medicinal plant *Forsythia suspensa* (Oleaceae). *Int. J. Mol. Sci.* **2017**, *18*, 2288. [CrossRef] [PubMed]

45. Kanga, J.Y.; Lua, J.J.; Qiua, S.; Chen, Z.; Liu, J.J.; Wang, H.Z. *Dendrobium* SSR markers play a good role in genetic diversity and phylogenetic analysis of Orchidaceae species. *Sci. Hortic.* **2015**, *183*, 160–166. [CrossRef]

46. Song, Y.; Wang, S.; Ding, Y.; Xu, J.; Li, M.F.; Zhu, S.; Chen, N. Chloroplast genomic resource of Paris for species discrimination. *Sci. Rep.* **2017**, *7*, 3427–3434. [CrossRef] [PubMed]

47. Yu, X.Q.; Drew, B.T.; Yang, J.B.; Gao, L.M.; Li, D.Z. Comparative chloroplast genomes of eleven *Schima* (Theaceae) species: Insights into DNA barcoding and phylogeny. *PLoS ONE* **2017**, *12*, e0178026. [CrossRef] [PubMed]

48. Wang, B.; Jiang, B.; Zhou, Y.; Su, Y.; Wang, T. Higher substitution rates and lower dN/dS for the plastid genes in Gnetales than other gymnosperms. *Biochem. Syst. Ecol.* **2015**, *59*, 278–287. [CrossRef]

49. Givnish, T.J.; Spalink, D.; Ames, M.; Lyon, S.P.; Hunter, S.J.; Zuluaga, A.; Iles, W.J.; Clements, M.A.; Arroyo, M.T.; Leebens-Mack, J.; et al. Orchid phylogenomics and multiple drivers of their extraordinary diversification. *Proc. Biol. Sci. B* **2015**, *282*, 2108–2111. [CrossRef] [PubMed]

50. Sasaki, Y.; Hakamada, K.; Suama, Y.; Nagano, Y.; Furusawa, I.; Matsuno, R. Chloroplast encoded protein as a subunit of acetyl-COA carboxylase in pea plant. *J. Biol. Chem.* **1993**, *268*, 25118–25123. [PubMed]

51. Konishi, T.; Shinohara, K.; Yamada, K.; Sasaki, Y. Acetyl-CoA carboxylase in higher plants: Most plants other than Gramineae have both the prokaryotic and the eukaryotic forms of this enzyme. *Plant Cell Physiol.* **1996**, *37*, 117–122. [CrossRef] [PubMed]

52. Kode, V.; Mudd, E.A.; Iamtham, S.; Day, A. The tobacco plastid *accD* gene is essential and is required for leaf development. *Plant J.* **2005**, *44*, 237–244. [CrossRef] [PubMed]

53. Nakkaew, A.; Chotigeat, W.; Eksomtramage, T.; Phongdara, A. Cloning and expression of a plastid-encoded subunit, beta-carboxyltransferase gene (*accD*) and a nuclear-encoded subunit, biotin carboxylase of acetyl-CoA carboxylase from oil palm (*Elaeis guineensis* Jacq.). *Plant Sci.* **2008**, *175*, 497–504. [CrossRef]

54. Drescher, A.; Ruf, S.; Calsa, T.J.; Carrer, H.; Bock, R. The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *Plant J.* **2000**, *22*, 97–104. [CrossRef] [PubMed]

55. Asano, T.; Tsudzuki, T.; Takahashi, S.; Shimada, H.; Kadowaki, K. Complete nucleotide sequence of the sugarcane (*Saccharum officinarum*) chloroplast genome: A comparative analysis of four monocot chloroplast genomes. *DNA Res.* **2004**, *11*, 93–99. [CrossRef] [PubMed]

56. Kikuchi, S.; Bédard, J.; Hirano, M.; Hirabayashi, Y.; Oishi, M.; Imai, M.; Takase, M.; Ide, T.; Nakai, M. Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science* **2013**, *339*, 571–574. [CrossRef] [PubMed]

57. Greiner, S.; Wang, X.; Herrmann, R.G.; Rauwolf, U.; Mayer, K.; Haberer, G.; Meurer, J. The complete nucleotide sequences of the 5 genetically distinct plastid genomes of *Oenothera*, subsection *Oenothera*:

II. A microevolutionary view using bioinformatics and formal genetic data. *Mol. Biol. Evol.* **2008**, *25*, 2019–2030. [CrossRef] [PubMed]

58. Carbonell-Caballero, J.; Alonso, R.; Ibañez, V.; Terol, J.; Talon, M.; Dopazo, J. A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus *Citrus*. *Mol. Biol. Evol.* **2015**, *32*, 2015–2035. [CrossRef] [PubMed]

59. Hu, S.; Sablok, G.; Wang, B.; Qu, D.; Barbaro, E.; Viola, R.; Li, M.; Varotto, C. Plastome organization and evolution of chloroplast genes in *Cardamine* species adapted to contrasting habitats. *BMC Genom.* **2015**, *16*, 1. [CrossRef] [PubMed]

60. Allahverdiyeva, Y.; Mamedov, F.; Mäenpää, P.; Vass, I.; Aro, E.M. Modulation of photosynthetic electron transport in the absence of terminal electron acceptors: Characterization of the *rbcL* deletion mutant of tobacco. *Biochim. Biophys. Acta Bioenerg.* **2005**, *1709*, 69–83. [CrossRef] [PubMed]

61. Piot, A.; Hackel, J.; Christin, P.A.; Besnard, G. One-third of the plastid genes evolved under positive selection in PACMAD grasses. *Planta* **2018**, *247*, 255–266. [CrossRef] [PubMed]

62. Kapralov, M.V.; Filatov, D.A. Widespread positive selection in the photosynthetic Rubisco enzyme. *BMC Evol. Biol.* **2007**, *7*, 73–82. [CrossRef] [PubMed]

63. Ivanova, Z.; Sablok, G.; Daskalova, E.; Zahmanova, G.; Apostolova, E.; Yahubyan, G.; Baev, V. Chloroplast genome analysis of resurrection tertiary relict *Haberlea rhodopensis* highlights genes important for desiccation stress response. *Front. Plant Sci.* **2017**, *8*, 1–15. [CrossRef] [PubMed]

64. Shi, Y.; Yang, L.F.; Yang, Z.Y.; Ji, Y.H. The complete chloroplast genome of *Pleione bulbocodioides* (Orchidaceae). *Conserv. Genet. Resour.* **2017**, 1–5. [CrossRef]

65. Górniak, M.; Paun, O.; Chase, M.W. Phylogenetic relationships with Orchidaceae based on a low-copy nuclear-coding gene, Xdh: Congruence with organellar and nuclear ribosomal DNA results. *Mol. Phylogenet. Evol.* **2010**, *56*, 784–795. [CrossRef] [PubMed]

66. 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]

67. Rasmussen, F.N. The families of the monocotyledones—Structure, evolution and taxonomy. In *Orchids*; Dahlgren, R., Cliford, H.T., Yeo, P.F., Eds.; Springer: Berlin/Heidelberg, Germany, 1985; pp. 249–274.

68. Szlachetko, D.L. Systema orchidalium. *Fragm. Florist. Geobot. Pol.* **1995**, *3*, 1–152.

69. Doyle, J.J. A rapid DNA isolation procedure from small quantities of fresh leaf tissues. *Phytochem. Bull.* **1987**, *19*, 11–15.

70. Chevreux, B.; Pfisterer, T.; Drescher, B.; Driesel, A.J.; Müller, W.E.; Wetter, T.; Suhai, S. Using the mira EST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res.* **2004**, *14*, 1147–1159. [CrossRef] [PubMed]

71. Hahn, C.; Bachmann, L.; Chevreux, B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads-abaiting and iterative mapping approach. *Nucleic Acids Res.* **2013**, *41*, e129. [CrossRef] [PubMed]

72. Wyman, S.K.; Jansen, R.K.; Boore, J.L. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* **2004**, *20*, 3252–3255. [CrossRef] [PubMed]

73. Kearse, M.; Moir, R.; Wilson, A.; Steven, S.H.; Matthew, C.; Shane, S.; Simon, B.; Alex, C.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *12*, 1647–1649. [CrossRef] [PubMed]

74. Lohse, M.; Drechsel, O.; Kahlau, S.; Bock, R. Organellar Genome DRAW-a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res.* **2013**, *41*, 575–581. [CrossRef] [PubMed]

75. Kurtz, S.; Schleiermacher, C. REPuter: Fast computation of maximal repeats incomplete genomes. *Bioinformatics* **1999**, *15*, 426–427. [CrossRef] [PubMed]

76. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *27*, 573. [CrossRef] [PubMed]

77. Thiel, T.; Michalek, W.; Varshney, R.K.; Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **2003**, *106*, 411–422. [CrossRef] [PubMed]

78. Stamatakis, A. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **2006**, *22*, 2688–2690. [CrossRef] [PubMed]

79. Yang, Z.; Nielsen, R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **2002**, *19*, 908–917. [CrossRef] [PubMed]

80. Posada, D.; Crandall, K.A. Modeltest: Testing the model of DNA substitution. *Bioinformatics* **1998**, *14*, 817–818. [CrossRef] [PubMed]