



# Article The Evolution of Mitochondrial Genomes between Two *Cymbidium* Sister Species: Dozens of Circular Chromosomes and the Maintenance and Deterioration of Genome Synteny

Xiaoling Li<sup>1,2</sup>, Mengqing Zhe<sup>1,2</sup>, Yiwei Huang<sup>1,2</sup>, Weishu Fan<sup>1</sup>, Junbo Yang<sup>1</sup> and Andan Zhu<sup>1,\*</sup>

<sup>1</sup> Germplasm Bank of Wild Species, Yunnan Key Laboratory of Crop Wild Relatives Omics, Kunning Institute of Botany, Chinese Academy of Sciences, Kunning 650201, China

<sup>2</sup> University of Chinese Academy of Sciences, Beijing 100049, China

\* Correspondence: zhuandan@mail.kib.ac.cn; Tel.: +86-0871-65238370

Abstract: Plant mitochondrial genomes (mitogenomes) exhibit fluid genome architectures, which could lead to the rapid erosion of genome synteny over a short evolutionary time scale. Among the species-rich orchid family, the leafy Cymbidium lancifolium and leafless Cymbidium macrorhizon are sister species with remarkable differences in morphology and nutritional physiology. Although our understanding of the evolution of mitochondria is incomplete, these sister taxa are ideal for examining this subject. In this study, the complete mitogenomes of C. lancifolium and C. macrorhizon, totaling 704,244 bp and 650,751 bp, respectively, were assembled. In the 2 mitogenomes, 38 protein-coding genes, 18 cis- and 6 trans-spliced introns, and approximately 611 Kb of homologous sequences are identical; overall, they have 99.4% genome-wide similarity. Slight variations in the mitogenomes of C. lancifolium and C. macrorhizon in repeat content (21.0 Kb and 21.6 Kb, respectively) and mitochondrial DNA of plastid origin (MIPT; 38.2 Kb and 37.5 Kb, respectively) were observed. The mitogenome architectures of C. lancifolium and C. macrorhizon are complex and comprise 23 and 22 mini-circular chromosomes, respectively. Pairwise comparisons indicate that the two mitogenomes are largely syntenic, and the disparity in chromosome numbers is likely due to repeat-mediated rearrangements among different chromosomes. Notably, approximately 93.2 Kb C. lancifolium mitochondrial sequences lack any homology in the C. macrorhizon mitogenome, indicating frequent DNA gains and losses, which accounts mainly for the size variation. Our findings provide unique insights into mitogenome evolution in leafy and leafless plants of sister species and shed light on mitogenome dynamics during the transition from mixotrophy to mycoheterotrophy.

Keywords: mitogenome; multi-chromosomes; genome synteny; recombination; nutritional modes

# 1. Background

Mitochondria are the centers of energy production and metabolic conversion within cells and play a vital role in plant productivity and development [1–3]. Plant mitochondrial genomes (mitogenomes) have been long recognized as evolving rapidly in structure and slowly in sequence [4,5]. In fact, sequence divergence among angiosperm mitogenomes is typically slow [6,7], approximately 4- and 20-fold slower than in their chloroplast and nuclear counterparts, respectively [8,9]; however, exceptions have been observed in several lineages [9–12]. The genome size of angiosperm mitogenomes can vary substantially between closely-related species or even among populations [13]. Angiosperm mitochondrial genomes typically range from 200 to 800 Kb in size, but genome sizes have exceeded 1 Mb in some species [14–16]. Gene order and genome structure are rarely conserved in angiosperm mitogenome is tightly associated with its high recombination rate, which is mediated mainly by large (e.g., >1 Kb) repeats [7,17,18]. The complexity of angiosperm mitogenome architectures



Citation: Li, X.; Zhe, M.; Huang, Y.; Fan, W.; Yang, J.; Zhu, A. The Evolution of Mitochondrial Genomes between Two *Cymbidium* Sister Species: Dozens of Circular Chromosomes and the Maintenance and Deterioration of Genome Synteny. *Genes* **2023**, *14*, 864. https://doi.org/10.3390/ genes14040864

Academic Editor: Zhiqiang Wu

Received: 1 March 2023 Revised: 30 March 2023 Accepted: 31 March 2023 Published: 3 April 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). is also substantiated by their in vivo structural diversity, including linear, circular, and branched structures [19].

With more than 28,000 species, Orchidaceae is one of the most diverse angiosperm families, accounting for approximately 9% of all vascular plants [20,21]. Within the Orchidaceae, the genus *Cymbidium* comprises about 70 species, which are distributed mainly from East and Southeast Asia to Australia, and exhibits an extraordinary diversity in lifestyles, including terrestrial, epiphytic, and lithophytic life forms [22]. The genus *Cymbidium* can be divided into three subgenera (subg.): *Cymbidium, Jensoa*, and *Cyperorchis* [22].

*C. lancifolium* Hook and *Cymbidium macrorhizon* Lindl are sister species in the Subg. *Jensoa*, which diverged from each other approximately 2.4 million years ago (Mya) [23–25]. The two species differ substantially in their morphological traits and nutritional modes. For example, *C. lancifolium* has leaves and is apparently a mixotrophic orchid species [26]. In contrast, *C. macrorhizon* lacks leaves, leading to increased dependence on fungal associations for nutrient acquisition [24,27].

To date, only a few mitochondrial genome sequences have been reported for the species-rich Orchidaceae family. Previous studies have indicated that the mitochondrial genome of the mycoheterotrophic orchid, *Gastrodia elata*, has markedly expanded to >1339 Kb, with many photosynthesis-related genes being lost [28]. The mitochondrial genome of the moth orchid (*Phalaenopsis aphrodite*) is approximately 576 Kb in size, with 27% and 9% of the sequence derived from the nucleus or plastid, respectively [29]. The newly reported mitochondrial genome of *Paphiopedilum micranthum* consists of 26 circular chromosomes with a total length of 447,368 bp, of which many are intracellular gene transfers [30].

In this study, we assembled the complete mitochondrial genomes of *C. lancifolium* and *C. macrorhizon* based on a combination of long- and short-read sequencing. The aims of this study include (1) a comparative analysis of mitochondrial genome architectures of leafy *C. lancifolium* and leafless *C. macrorhizon*; and (2) an investigation of the evolutionary mechanisms underlying the maintenance and deterioration of genome synteny in multi-circular molecules.

#### 2. Materials and Methods

#### 2.1. Plant Materials and Sequencing

Seedlings of two *Cymbidium* sister species (*C. lancifolium* and *C. macrorhizon*) were collected from the Orchid Germplasm Resource Garden of Kunming Institute of Botany and Changchong Mountain of Kunming (Yunnan, China), respectively (Figure S1). Total genomic DNA was extracted from fresh leaves of *C. lancifolium* by using a modified cetyltrimethylammonium bromide (CTAB) protocol [31]. For genome sequencing, a genomic library with 500 bp fragments was prepared for short-read sequencing in the pairedend sequencing mode on the DNBSEQ-T7 platform, and a 15-Kb insertion library was constructed for long-read sequencing on the Pacific Biosciences (PacBio) sequel II platform with HiFi mode (Table S1). For the leafless *C. macrorhizon*, total genomic DNA and shortreads were obtained from both scapes (flowering stems) and rhizomes by using the CTAB protocol and next-generation method, as described for *C. lancifolium*. (Figure S1). Long reads were generated on the PacBio sequel II platform by using the Continuous Long Reads (CLR) mode with a 20 Kb insertion library (Table S1).

#### 2.2. Mitogenome Assembly and Evaluation

Approximately 9 Gb Illumina short reads were used to assemble the draft mitogenomes of *C. lancifolium* and *C. macrorhizon* with the "-R 20 -k 21,45,65,85,105 -P 1,000,000 -F embplant\_mt" option implemented in GetOrganelle version 1.7.5 [32–35]. The assembly graphs were visualized by using Bandage based on the raw graphical fragment assembly (GFA) file produced by GetOrganelle [36], and the read depths of contigs as well as the links of graph edges were manually checked. Short (<1 Kb) contigs with low sequencing depths (<7×) or without direct links to other mitochondrial contigs were removed from further analyses.

The initial assembly outputs for *C. lancifolium* and *C. macrorhizon* contained 363 contigs and 228 contigs, with a total length of 843,278 bp and 787,850 bp, respectively (Figure S2). Then, Burrows–Wheeler Aligner (BWA)-mem [37] was used to identify putative mitochondrial DNA-derived PacBio long reads from approximately 14 Gb raw long reads by mapping to these draft assemblies. Canu v.2.2 [38] was used to assemble the mitochondrial genome of *C. lancifolium* and *C. macrorhizon* independently, based on the -pacbio-hifi and -pacbio modes, respectively. Ultimately, two complete mitochondrial genomes with multiple circular molecules were assembled for the two *Cymbidium* species. Pilon version 1.24 [39] was used to correct the mitogenome of *C. macrorhizon*, which was assembled by using long reads generated via the PacBio CLR mode.

To verify the integrity of the mitogenome structure of *C. lancifolium* and *C. macrorhizon*, the sequenced long reads were mapped to their corresponding assemblies by using maphifi and map-pb modes of minimap2 v.2.24 [40], index files were generated with SAMtools and check GAP via the integrative genomics viewer (IGV) [41,42], and the average read depth of each mitogenome was estimated with BEDTools v.2.30.0 [40,42]. To avoid mapping bias at the ends of each chromosome, we also redefined a new starting point using the command "restart -i 5000" implemented in SeqKit [43] to check the read coverage. For a final test, we mapped the initial draft mitochondrial assembles (based on Illumina short reads) to the complete mitogenomes (based on PacBio long reads) by using the asm10 mode of Minimap2. The mapping rates of *C. lancifolium* and *C. macrorhizon* contigs were 98.38% and 96.26%, respectively, giving strong cross-validation for the completeness of PacBio-based mitogenome assemblies.

### 2.3. Genome Annotation

The mitogenomes were annotated by using the "Live annotate & predict" tool in Geneious R7 [44]. Two mitochondrial genome references, *Arabidopsis thaliana* (NC\_037304) and *Nymphaea colorata* (NC\_037468), were downloaded from GenBank. Gene structure was manually checked to refine the in-frame start and stop codons and adjust the exon-intron boundaries. tRNAscan-SE [45] was used to identify mitochondrial tRNAs with default parameters. The mitogenome maps were drawn by using TBtools [46]. The assembled mitogenomes and their annotations were deposited in GenBank under the accession numbers OQ024442-OQ024464 and OQ029542-OQ029563.

#### 2.4. Identification of Repeats and Mitochondrial DNA of Plastid Origin

Simple sequence repeats (SSR) were analyzed by using MISA (2017, https://webblast. ipk-gatersleben.de/misa/ accessed on 20 March 2023) with the following parameters: (1/10) (2/5) (3/4) (4/4) (5/3) (6/3). Disperse repeats within and across chromosomes of each mitogenome were identified by using the basic local alignment search tool (BLAST)N v.2.12.0 to search against themselves with a minimum alignment length of 100 bp and minimal identity of 95%. The chloroplast genomes of *C. lancifolium* and *C. macrorhizon* were downloaded from the National Center for Biotechnology Information (NCBI) (MW582681 and MW582687). The plastid-derived sequences (mitochondrial DNA of plastid origin) were identified by mapping the *C. lancifolium* and *C. macrorhizon* plastomes to their corresponding mitogenomes with BLASTN by using the same criterion as for the disperse repeat analysis.

#### 2.5. Genome Similarity and Synteny Analysis

The whole-genome average nucleotide identity between *C. lancifolium* and *C. macrorhizon* mitogenomes was calculated by using FASTANI [47] with default parameters. Wholemitogenome alignment was performed by using BLASTN to identify homologous chromosomes. Based on the alignment results, we used SeqKit to reset the start site and manually sort the results, and Mauve v.2.4.0 [48] was used to determine and visualize genome synteny as locally collinear blocks with a minimal length of 1 Kb.

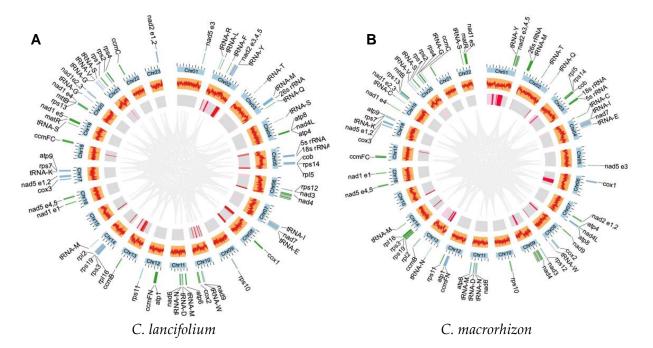
### 2.6. Identification and Verification of Species-Specific Mitochondrial Sequences

Homologous sequences were identified by using BLASTN searches against the *C. lancifolium* and *C. macrorhizon* mitogenomes with sequence identity >98% and a minimum alignment length of 100 bp. Species-specific mitochondrial sequences were identified by parsing the BLASTN alignment results. The evolution of mitochondrial sequence composition was measured by the maintenance and turnover of homologous sequences [49]. Moreover, homologous sequences were extracted from syntenic blocks and realigned with Muscle [50] to identify insertions/deletions (indels). Six large (each > 1 Kb in size) indels that were caused by the maintenance of *C. lancifolium*-specific sequences were identified, and these indels mostly accounted for the mitogenome size difference between *C. lancifolium* and *C. macrorhizon*. We validated these large indels by mapping *C. macrorhizon* short reads to the *C. lancifolium* mitogenome by using BWA and visualized them by using IGV [37,41,42].

# 3. Results

# 3.1. The Mitogenomes of Sister Species C. lancifolium and C. macrorhizon Comprise Dozens of Minicircular Chromosomes

De novo assemblies using PacBio long reads generated two complete mitogenomes with multiple circular molecules (Figure 1; Table 1). The mitogenome of the leafy *C. lancifolium* comprises 23 circular chromosomes that vary from 23,637 to 46,518 bp in size, totaling 704,244 bp (Table S2). The leafless *C. macrorhizon* mitogenome has a total length of 650,751 bp and consists of 22 circular chromosomes, ranging from 20,861 to 46,822 bp in size. Both the *C. lancifolium* and *C. macrorhizon* mitogenomes exhibit a 44% GC content, similar to other angiosperms [51]. Long reads were mapped back to their corresponding mitogenome to verify genome continuity, and the average depth of assembled mitogenomes was ~110× (Figure S3).



**Figure 1.** The complete mitogenomes of *C. lancifolium* and *C. macrorhizon*. (A) The mitogenome structure and features of the leafy *C. lancifolium*. (B) The mitogenome structure and features of the leafless *C. macrorhizon*. From outermost to innermost, circle I: mitochondrial genes transcribed clockwise and counterclockwise are shown in green and blue, respectively; circle II: mitochondrial chromosomes; circle III: GC content; circle IV: MIPT sequences; circle V: repeats across chromosomes.

	Size (bp)	Chromosomes	Protein-Coding Genes	MIPT (bp)	Repeats (bp)
C. lancifolium	704,244	23	38	53,587	20,951
C. macrorhizon	650,751	22	38	50,901	21,619

Table 1. General genome features of the assembled C. lancifolium and C. macrorhizon mitogenomes.

# 3.2. Maintenance of Gene and Intron Content in the C. lancifolium and C. macrorhizon Mitogenomes

The *C. lancifolium* and *C. macrorhizon* mitogenomes are nearly identical regarding gene and intron content (Table 2 and Figure 1). A total of 38 unique protein-coding genes and 3 rRNAs was identified in each mitogenome. In total, 8 *cis*- and 6 *trans*-spliced introns were found in 10 mitochondrial genes (*nad1*, *nad2*, *nad4*, *nad5*, *nad7*, *cox2*, *ccmFC*, *rpl2*, *rps3*, and *rps10*). Notably, *mttB* lacks the standard start codon (ATG), which is consistent with previous findings [9,52]. Furthermore, *nad1* likely starts with ACG but could be converted to AUG at the transcript level through RNA editing. In the mitogenomes of *C. lancifolium* and *C. macrorhizon*, 21 and 19 tRNAs, respectively, were annotated.

**Table 2.** Genes encoded in the *C. lancifolium* and *C. macrorhizon* mitogenomes. \* indicates mitochondrial genes with introns.

Category	Gene			
Complex I	nad1 * nad2 * nad3 nad4 * nad4L nad5 * nad6 nad7 * nad9			
Complex III	cob			
Complex IV	cox1 cox2* cox3			
Complex V	atp1 atp4 atp6 atp8 atp9			
Cytochrome c biogenesis	vtochrome c biogenesis ccmB ccmC ccmFN ccmFC *			
Ribosome large subunits	ome large subunits ppl2 * ppl5 ppl16			
Ribosome small subunits	rps1 rps2 rps3 * rps4 rps7 rps10 * rps11 rps12 rps13 rps14 rps19			
Others	matR mttB			
rRNA genes	rrn5 rrn18 rrn26			
tRNA genes	trnL-TAG trnF-GAA trnY-GTA trnT-TGT trnQ-TTG trnM-CAT trnT-TGT trnS-AGA trnE-TTG trnI-TAT trnW-CCA trnD-GTC trnN-GTT trnM-CAT trnM-CAT trnK-TTT trnS-GCT trnC-GCA trnS-TAC trnV-TAC trnG-TCC	trnY-GTA trnM-CAT trnT-TGT trnC-GCA trnE-TTC trnI-TAT trnW-CCA trnN-GTT trnM-CAT trnD-GTC trnN-GTT trnM-CAT trnK-TTT trnC-GCA trnG-TCC trnS-GCT trnV-TAC trnS-GCT trnQ-TTG		

#### 3.3. Repeat Content and Mitochondrial DNA of Plastid Origin (MIPT)

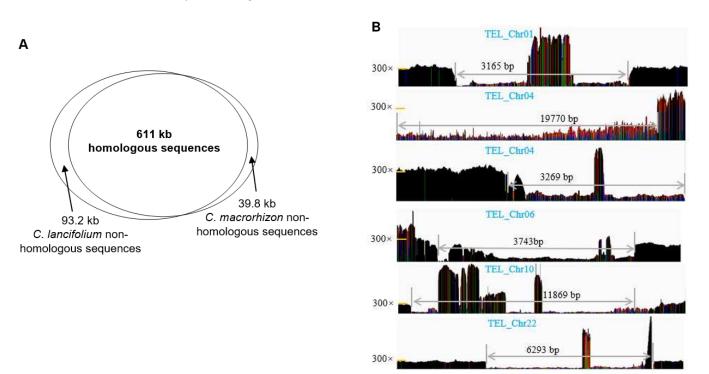
A total of 130 and 128 tandem repeats (TRs) was detected in the mitochondrial genomes of *C. lancifolium* and *C. macrorhizon* (Table S3), respectively; mononucleotide TRS (~46%) were the most common, followed by trinucleotide (~27%) and dinucleotide (~22%) TRs. Dispersed repeats, ranging from 107 bp to 1338 bp, are prevalent, accounting for approximately 2.97% and 3.34% of the *C. lancifolium* and *C. macrorhizon* mitogenomes (Table 1), respectively. Intriguingly, nearly 65% of dispersed repeats are remarkably similar in sequence within each species. These repeats can be categorized into two distinct classes (group A and group B) based on their conserved regions (CRs). Group A contains 18 repetitive sequences from 18 chromosomes of the *C. lancifolium* mitogenome that share an identical 212 bp CR (Figure S4A) and 19 repetitive sequences from 19 chromosomes in *C. macrorhizon* mitogenomes have 6 repeats that share conserved regions of 149 and 148 bp, respectively (Figures S4B and S5B).

Large chloroplast genome segments have transferred into the mitogenomes of the two *Cymbidium* species. A total of 38,185 bp and 37,518 bp MIPT fragments was identified in the mitogenomes of *C. lancifolium* and *C. macrorhizon*, respectively, totaling approximately 5–6% of the genome sizes (Table 1, Figure S6). The identified MIPT fragments vary substantially in length between the *C. lancifolium* and *C. macrorhizon* mitogenomes, with the two largest fragments being 9.4 Kb and 11.6 Kb, respectively. The insertion of these MIPT fragments into

these circular chromosomes is not uniformly distributed. For example, 3 MIPT fragments, which were 15.0 Kb in total length, with the largest insertion being 11.6 Kb, were identified in *C. macrorhizon* Chr01 (chromosome size = 46.8 Kb), and a total of 12.7 Kb MIPT sequences was identified in *C. lancifolium* Chr02 (chromosome size = 45.7 Kb). In contrast, there are 13 and 11 chromosomes of *C. lancifolium* and *C. macrorhizon* mitogenomes, respectively, without any MIPT insertions.

# *3.4. Sequence and Genome Synteny Evolution between the C. lancifolium and C. macrorhizon Mitogenomes*

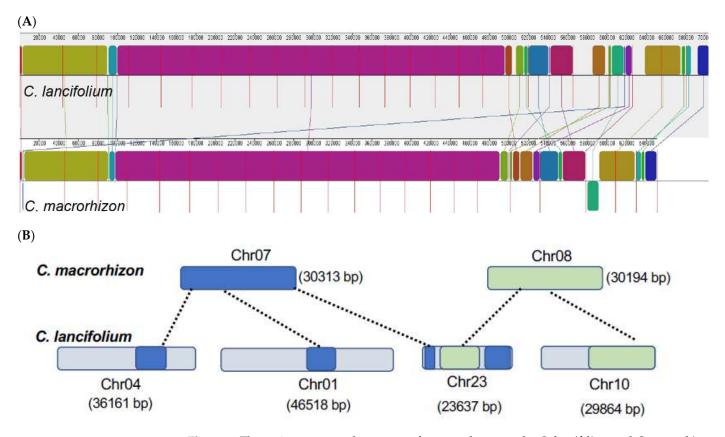
The whole-mitogenome average nucleotide identity is 99.4% between *C. lancifolium* and *C. macrorhizon*, which is consistent with their sister species relationship and relatively recent divergence time (approximately 2.4 Mya) [24]. Homologous sequences of approximately 611 Kb were shared in both species (Figure 2A), covering 86.8% of the *C. lancifolium* mitogenome and 93.9% of the *C. macrorhizon* mitogenome. However, the variation in mitogenome sizes was mainly due to the deletion of several large DNA segments in *C. macrorhizon* and/or DNA gains in *C. lancifolium*. For example, approximately 48 Kb mitochondrial sequences from five *C. lancifolium* chromosomes (Chr01, Chr04, Chr06, Chr10, and Chr12) lacked any homology in *C. macrorhizon* (Figure 2B). These non-homologous sequences were nested in other homologous sequences, leading to insertions and deletions (indels) during genome alignment. To verify this, we examined all 6 large (>1 Kb) indels, ranging from 3.2 to 19.8 Kb. All six of these regions are located in intergenic regions and are continuous in both mitogenomes. As expected, inter-specific mapping of the *C. macrorhizon* reads to the *C. lancifolium* mitogenome revealed that these indel regions were less covered by reads (Figure 2B).



**Figure 2.** The maintenance and turnover of homologous sequences in the *C. lancifolium* and *C. macrorhizon* mitogenomes. (**A**) The relative composition of homologous and non-homologous sequences in the two mitogenomes. (**B**) Validation of *C. lancifolium* non-homologous sequences by using *C. macrorhizon* short reads. The gray arrows highlight the six largest indels caused by the maintenance of *C. lancifolium*-specific sequences.

The mitogenome architectures of *C. lancifolium* and *C. macrorhizon* are generally consistent at the macrosynteny level (Figure 3). Specifically, 20 out of the 22 total chromo-

somes in *C. macrorhizon* had a 1:1 homologous relationship with individual chromosome in *C. lancifolium* (Table S4). Despite the largely maintained genome synteny between these 2 sister species, a total of 14 rearrangements has occurred between the *C. lancifolium* and *C. macrorhizon* mitogenomes. The most notable genomic rearrangements were found in Chr07 and Chr08. Chr08 in *C. macrorhizon* was homologous to parts of Chr10 and Chr23 in *C. lancifolium*, while Chr07 has a tripartite network of homologous relationships to partial Chr01, Chr04, and Chr23 sequences in *C. lancifolium* (Figure 3B).



**Figure 3.** The maintenance and turnover of synteny between the *C. lancifolium* and *C. macrorhizon* mitogenomes. (**A**) The overall whole-genome synteny between the two mitogenomes. Syntenic blocks were identified by using progressiveMauve implemented in Mauve. Upper: *C. lancifolium;* lower: *C. macrorhizon*. (**B**) Inter-chromosomal synteny relationship between the *C. lancifolium* and *C. macrorhizon* mitogenomes.

# 4. Discussion

Plant mitochondrial DNAs (mtDNAs) exhibit striking structural diversity within and among species, with architectures that can shift rapidly between linear, circular, and branched structures in vivo [19,53,54]. This study revealed multi-chromosomal structures of mitogenomes within two *Cymbidium* sister species, each consisting of >20 mini-circular chromosomes (20–47 Kb; Figure 1). Multi-chromosomal mitogenomes have independently evolved in several lineages, such as *Cucumis* [55], *Silene* [12], *Ajuga* [56], and a parasitic plant by the name of *Rhopalocnemis phalloides* [57]. Notably, for these multi-chromosomal mitogenomes, both the chromosome numbers and their sizes can differ substantially in closely related species and even among populations. For example, the mitogenome of *Silene noctiflora* exhibits differences in chromosome numbers (59 and 63) and sizes (6.73 Mb and 7.14 Mb) between accessions [12,58]. Rapid gain or loss of several entire chromosomes has been suggested as a crucial evolutionary mechanism accounting for these disparities [58].

Our syntenic analysis of the mitogenomes of the sister species, *C. lancifolium* and *C. macrorhizon*, provided an alternative mechanism for differences in chromosome number and size. Chromosome-scale syntenic relationships could rapidly erode via inter-

chromosomal rearrangements, resulting in disparity of chromosomal numbers (Figure 3B). Inter-chromosomal rearrangements are presumably mediated by recombinationally-active repeated sequences, which are widespread in vascular mitogenomes [54,59,60]. We found that an identical large (1256 bp) repeat pair is shared by Chr07 and Chr08 in the *C. macrorhizon* mitogenome, suggesting that these chromosomes might be recombination hotspots.

Intriguingly, two types of conserved sequences that are shared by nearly all chromosomes were dominant among the identified dispersed repeats (Figures S4 and S5). It is unclear why and how such repeats have been maintained or accumulated in plant mitogenomes. A recent study in the parasitic plant *R. phalloides* has identified an 896 bp conserved region that is shared by all chromosomes and could act as the origin of replication by forming stem–loop secondary structures [57]. If this mitochondrial replication mechanism is not unique to parasitic plants, it is reasonable to propose that similar adaptive roles might affect the maintenance of these accumulated repeats.

The leafless *C. macrorhizon* differs considerably in morphology and nutritional mode from its sister species *C. lancifolium* (Figure S1). However, the mitogenomes of these two species are generally consistent in sequence composition (gene and intron content) and macrosynteny (except for the few rearranged chromosomes). This result mirrors our findings of minor genetic differences between the chloroplast genomes of these two morphologically distinct sister species [61]. The evolution of the genomes of *C. macrorhizon* organelles contrasts sharply with the findings for a fully mycoheterotrophic plant, *G. elata*, which has a dramatic reduction in gene content of its organelle genomes [28]. This difference might suggest that *C. macrorhizon* is at an early stage of a transition to a fully mycoheterotrophic lifestyle.

The mitogenomes of *C. lancifolium* and *C. macrorhizon* are conserved in terms of sequence organization. The multi-chromosome structure is consistent with the recently reported mitogenome of *P. micranthum*, which consists of 26 circular chromosomes, ranging in size from 5973 bp to 32,281 bp [30]. For other orchid species, it has been reported that the mitogenome is fragmented, suggesting that the multi-circular structure of the mitogenome is likely common in species within the orchid family; plastid-derived DNA fragments represented 5% and 6% of the *C. lancifolium* and *C. macrorhizon* mitogenome, respectively, with similar findings reported for *P. aphrodite* (9%) and *P. micranthum* (10%) [29,30]. Frequent inter-chromosomal recombination allows the mitochondrial genome to easily become a receptor for the lateral transfer of foreign genomes, which is consistent with the large number of recombination events that have occurred in the mitogenomes of different orchid species.

### 5. Conclusions

We accurately assembled the mitogenomes of *C. lancifolium* and *C. macrorhizon* by using a combination of short-read and long-read data. Both mitogenomes have multi-circular architectures, which are composed of more than 20 sub-genomes. Furthermore, a large number of fragments have been transferred from the chloroplast genome, reflecting the complex and dynamic evolution of the mitogenome of orchid species. The protein-coding genes in both species are consistent, and the mitogenome-wide average nucleic acid identity is 99.4%, with more than 86% homologous sequences, which is consistent with the evolutionary relationship of these sister species. However, species-specific sequences are still present, and we identified the insertion/deletion of several large fragments, which directly led to the difference in mitogenome size. In addition, we suggest that the difference in the number of chromosomes resulted from intermolecular or intramolecular rearrangements. Interestingly, we identified two sets of homologous repeats that appeared 6–19 times in the mitochondrial genomes of both species. However, while the reasons for the large accumulation and maintenance of such repeats by species are currently unclear, we proposed an adaptive role for these homologous repeats. The multi-circle structure of the mitochondrial genome of Orchidaceae is not the first to be reported, and in addition to the current fragmented mitochondrial genome in Orchidaceae, we speculate that multi-circle structures may be common in the mitochondrial genome of orchid species. Information

on the mitochondrial genome of orchid species is largely unavailable. Further studies are needed to unravel the evolution of the mitogenome of orchids.

Supplementary Materials: The following supporting information can be downloaded at: https://www.action.com/actionals //www.mdpi.com/article/10.3390/genes14040864/s1, Figure S1. Plant materials used in this study. (A) The morphology of C. lancifolium. (B) Leaves of a C. lancifolium seedling were used for sequencing. (C) The morphology of C. macrorhizon. (D) Scapes and rhizomes of a C. macrorhizon seedling were used for sequencing. Figure S2. Diagrams of Illumina-based draft assemblies. (A) C. lancifolium mitogenome assembly. (B) C. macrorhizon mitogenome assembly. Contigs with high read coverage (>50×) are highlighted in green. Figure S3. Read coverage of the *C. lancifolium* and *C. macrorhizon* mitogenomes. (A) Read mapping plot of the C. lancifolium mitogenome. (B) Read mapping plot of the C. macrorhizon mitogenome. Figure S4. Evolution of boundary sequences of C. lancifolium repeats. (A) Alignment of boundary sequences surrounding a 212 bp conserved block of 18 homologous repeats. (B) Alignment of boundary sequences surrounding a 149 bp conserved block of 6 homologous repeats. Figure S5. Evolution of boundary sequences of C. macrorhizon repeats. (A) Alignment of boundary sequences surrounding a 140 bp conserved block of 19 homologous repeats. (B) Alignment of boundary sequences surrounding a 149 bp conserved block of 6 homologous repeats. Figure S6. Identification and distribution of mitochondrial DNA of plastid origin (MIPT). (A) MIPT sequences identified in chromosomes of the C. lancifolium mitogenome. (B) MIPT sequences identified in chromosomes of the C. macrorhizon mitogenome. CP: chloroplast; Chr: chromosome. Red lines link homologous sequences with >95% similarity, and blue lines link homologous sequence with >90% similarity. Table S1. Sampled species and sequencing data for this study. Table S2. Lengths of mitogenome chromosomes. Table S3. Summary of the identified simple sequence repeats. Table S4. Syntenic relationship among C. lancifolium and C. macrorhizon chromosomes.

**Author Contributions:** A.Z. and J.Y. designed and conceived this study; X.L., M.Z., Y.H. and W.F. performed experiments and analyses; X.L. and A.Z. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study is jointly supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB31000000), the Project for Innovation Team of Yunnan Province (Grant No. 202105AE160012), the Science and Technology Basic Resources Investigation Program of China (Grant No. 2021FY100200), and Yunnan Fundamental Research Projects (Grant NO. 202101AT070172).

**Data Availability Statement:** All the assembled and annotated mitogenomes were submitted to GenBank under accessions OQ024442-OQ024464 and OQ029542-OQ029563.

Conflicts of Interest: The authors declare no conflict of interest.

### References

- 1. Sagan, L. On the origin of mitosing cells. J. Theor. Biol. 1967, 14, 255–274. [CrossRef] [PubMed]
- Gras, D.E.; Mansilla, N.; Rodríguez, C.; Welchen, E.; Gonzalez, D.H. Arabidopsis thaliana surfeit1-like genes link mitochondrial function to early plant development and hormonal growth responses. *Plant J.* 2020, *103*, 690–704. [CrossRef]
- Shtolz, N.; Mishmar, D. The mitochondrial genome-on selective constraints and signatures at the organism, cell, and single mitochondrion levels. *Front. Ecol. Evol.* 2019, 7, 342. [CrossRef]
- Jacoby, R.P.; Li, L.; Huang, S.; Lee, C.P.; Millar, A.H.; Taylor, N.L. Mitochondrial composition, function and stress response in plants. J. Integr. Plant Biol. 2012, 54, 887–906. [CrossRef] [PubMed]
- Palmer, J.D.; Herbon, L.A. Plant mitochondrial DNA evolved rapidly in structure, but slowly in sequence. J. Mol. Evol. 1988, 28, 87–97. [CrossRef]
- 6. Wolfe, K.H.; Li, W.H.; Sharp, P.M. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear dnas. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 9054–9058. [CrossRef] [PubMed]
- Fan, W.; Liu, F.; Jia, Q.; Du, H.; Chen, W.; Ruan, J.; Lei, J.; Li, D.-Z.; Mower, J.; Zhu, A. Fragaria mitogenomes evolve rapidly in structure but slowly in sequence and incur frequent multinucleotide mutations mediated by micro-inversions. *New Phytol.* 2022, 236, 745–759. [CrossRef] [PubMed]
- Drouin, G.; Daoud, H.; Xia, J. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol. Phylogen. Evol.* 2008, 49, 827–831. [CrossRef]
- Zhu, A.; Guo, W.; Jain, K.; Mower, J.P. Unprecedented heterogeneity in the synonymous substitution rate within a plant genome. *Mol. Biol. Evol.* 2014, 31, 1228–1236. [CrossRef]

- 10. Cho, Y.; Mower, J.P.; Qiu, Y.-L.; Palmer, J.D. Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. *Proc. Natl. Acad. Sci. USA* 2004, 101, 17741–17746. [CrossRef] [PubMed]
- Mower, J.P.; Touzet, P.; Gummow, J.S.; Delph, L.F.; Palmer, J.D. Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. *BMC Evol. Biol.* 2007, 7, 135. [CrossRef]
- Sloan, D.B.; Alverson, A.J.; Chuckalovcak, J.P.; Wu, M.; McCauley, D.E.; Palmer, J.D.; Taylor, D.R. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *PLoS Biol.* 2012, 10, e1001241. [CrossRef] [PubMed]
- 13. Ward, B.L.; Anderson, R.S.; Bendich, A.J. The mitochondrial genome is large and variable in a family of plants (cucurbitaceae). *Cell* **1981**, *25*, 793–803. [CrossRef] [PubMed]
- Alverson, A.J.; Wei, X.; Rice, D.W.; Stern, D.B.; Barry, K.; Palmer, J.D. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (cucurbitaceae). *Mol. Biol. Evol.* 2010, 27, 1436–1448. [CrossRef] [PubMed]
- 15. Gualberto, J.M.; Mileshina, D.; Wallet, C.; Niazi, A.K.; Weber-Lotfi, F.; Dietrich, A. The plant mitochondrial genome: Dynamics and maintenance. *Biochimie* 2014, 100, 107–120. [CrossRef]
- Rice, D.W.; Alverson, A.J.; Richardson, A.O.; Young, G.J.; Sanchez-Puerta, M.V.; Munzinger, J.; Barry, K.; Boore, J.L.; Zhang, Y.; Depamphilis, C.W.; et al. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. *Science* 2013, 342, 1468–1473. [CrossRef]
- 17. Allen, J.O.; Fauron, C.M.; Minx, P.; Roark, L.; Oddiraju, S.; Lin, G.N.; Meyer, L.; Sun, H.; Kim, K.; Wang, C.; et al. Comparisons among two fertile and three male-sterile mitochondrial genomes of maize. *Genetics* **2007**, 177, 1173–1192. [CrossRef]
- Cole, L.W.; Guo, W.; Mower, J.P.; Palmer, J.D. High and variable rates of repeat-mediated mitochondrial genome rearrangement in a genus of plants. *Mol. Biol. Evol.* 2018, 35, 2773–2785. [CrossRef]
- 19. Kozik, A.; Rowan, B.A.; Lavelle, D.; Berke, L.; Schranz, M.E.; Michelmore, R.W.; Christensen, A.C. The alternative reality of plant mitochondrial DNA: One ring does not rule them all. *PLoS Genet.* **2019**, *15*, e1008373. [CrossRef]
- 20. Chase, M.W.; Cameron, K.M.; Freudenstein, J.V.; Pridgeon, A.M.; Salazar, G.; van den Berg, C.; Schuiteman, A. An updated classification of orchidaceae. *Bot. J. Linn. Soc.* 2015, 177, 151–174. [CrossRef]
- Givnish, T.J.; Spalink, D.; Ames, M.; Lyon, S.P.; Hunter, S.J.; Zuluaga, A.; Iles, W.J.D.; Clements, M.A.; Arroyo, M.T.K.; Leebens-Mack, J.; et al. Orchid phylogenomics and multiple drivers of their extraordinary diversification. *Proc. R. Soc. B* 2015, 282, 20151553. [CrossRef] [PubMed]
- 22. Puy, D.D.; Cribb, P. The Genus Cymbidium; Royal Botanic Gardens: London, UK, 2007; pp. 71–73.
- 23. Ogura-Tsujita, Y.; Yokoyama, J.; Miyoshi, K.; Yukawa, T. Shifts in mycorrhizal fungi during the evolution of autotrophy to mycoheterotrophy in *Cymbidium* (orchidaceae). *Am. J. Bot.* **2012**, *99*, 1158–1176. [CrossRef] [PubMed]
- 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]
- Zhang, L.; Huang, Y.W.; Huang, J.L.; Ya, J.D.; Zhe, M.Q.; Zeng, C.X.; Zhang, Z.R.; Zhang, S.B.; Li, D.Z.; Li, H.T.; et al. DNA barcoding of cymbidium by genome skimming: Call for next-generation nuclear barcodes. *Mol. Ecol. Resour.* 2023, 23, 424–439. [CrossRef]
- 26. Motomura, H.; Selosse, M.-A.; Martos, F.; Kagawa, A.; Yukawa, T. Mycoheterotrophy evolved from mixotrophic ancestors: Evidence in *Cymbidium* (orchidaceae). *Ann. Bot.* **2010**, *106*, 573–581. [CrossRef]
- 27. Suetsugu, K.; Ohta, T.; Tayasu, I. Partial mycoheterotrophy in the leafless orchid *Cymbidium macrorhizon. Am. J. Bot.* **2018**, *105*, 1595–1600. [CrossRef] [PubMed]
- 28. 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]
- Chen, T.-C.; Su, Y.-Y.; Wu, C.-H.; Liu, Y.-C.; Huang, C.-H.; Chang, C.-C. Analysis of mitochondrial genomics and transcriptomics reveal abundant RNA edits and differential editing status in moth orchid, *Phalaenopsis aphrodite* subsp. *formosana. Sci. Hortic.* 2020, 267, 109304. [CrossRef]
- 30. Yang, J.-X.; Dierckxsens, N.; Bai, M.-Z.; Guo, Y.-Y. Multichromosomal mitochondrial genome of paphiopedilum micranthum: Compact and fragmented genome, and rampant intracellular gene transfer. *Int. J. Mol. Sci.* **2023**, *24*, 3976. [CrossRef] [PubMed]
- 31. Doyle, J.J.; Doyle, J.L. A rapid DNA isolation proceudre for small quantities of fresh leaf tissue. *Phytochem. Bull.* **1987**, *19*, 11–15.
- 32. Jin, J.-J.; Yu, W.-B.; Yang, J.-B.; Song, Y.; Depamphilis, C.W.; Yi, T.-S.; Li, D.-Z. Getorganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* 2020, *21*, 241. [CrossRef] [PubMed]
- Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. Blast+: Architecture and applications. BMC Bioinform. 2009, 10, 421. [CrossRef] [PubMed]
- Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.; Dvorkin, M.; Kulikov, A.; Lesin, V.; Nikolenko, S.; Pham, S.; Prjibelski, A.; et al. Spades: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 2012, 19, 455–477. [CrossRef]
- 35. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with bowtie 2. Nat. Methods 2012, 9, 357–359. [CrossRef] [PubMed]
- Wick, R.R.; Schultz, M.B.; Zobel, J.; Holt, K.E. Bandage: Interactive visualization of de novogenome assemblies. *Bioinformatics* 2015, *31*, 3350–3352. [CrossRef]

- 37. Li, H. Aligning sequence reads, clone sequences and assembly contigs with bwa-mem. arXiv 2013, arXiv:1303.3997.
- Koren, S.; Walenz, B.P.; Berlin, K.; Miller, J.R.; Bergman, N.H.; Phillippy, A.M. Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 2017, 27, 722–736. [CrossRef]
- Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K.; et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* 2014, 9, e112963. [CrossRef]
- 40. Li, H. New strategies to improve minimap2 alignment accuracy. Bioinformatics 2021, 37, 4572–4574. [CrossRef]
- 41. Robinson, J.T.; Thorvaldsdottir, H.; Turner, D.; Mesirov, J.P. Igv.Js: An embeddable javascript implementation of the integrative genomics viewer (igv). *Bioinformatics* **2022**, *39*, btac830. [CrossRef] [PubMed]
- 42. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al. Twelve years of samtools and bcftools. *GigaScience* **2021**, *10*, giab008. [CrossRef]
- 43. Shen, W.; Le, S.; Li, Y.; Hu, F. Seqkit: A cross-platform and ultrafast toolkit for fasta/q file manipulation. *PLoS ONE* 2016, 11, e0163962. [CrossRef]
- Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; 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, 28, 1647–1649. [CrossRef] [PubMed]
- Lowe, T.M.; Chan, P.P. Trnascan-se on-line: Integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* 2016, 44, W54–W57. [CrossRef] [PubMed]
- Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.; Xia, R. Tbtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 2020, *13*, 1194–1202. [CrossRef]
- 47. Jain, C.; Rodriguez, R.L.; Phillippy, A.M.; Konstantinidis, K.T.; Aluru, S. High throughput ani analysis of 90k prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **2018**, *9*, 5114. [CrossRef]
- Darling, A.C.E.; Mau, B.; Blattner, F.R.; Perna, N.T. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 2004, 14, 1394–1403. [CrossRef] [PubMed]
- 49. Guo, W.; Grewe, F.; Fan, W.; Young, G.J.; Knoop, V.; Palmer, J.D.; Mower, J.P. *Ginkgo* and *welwitschia* mitogenomes reveal extreme contrasts in gymnosperm mitochondrial evolution. *Mol. Biol. Evol.* **2016**, *33*, 1448–1460. [CrossRef]
- 50. Edgar, R.C. Muscle: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [CrossRef] [PubMed]
- 51. Mower, J.; Sloan, D.; Alverson, A. *Plant Mitochondrial Genome Diversity: The Genomics Revolution*; Springer: Vienna, Austria, 2012; Volume 1, pp. 123–144.
- 52. Siculella, L.; Pacoda, D.; Treglia, S.; Gallerani, R.; Ceci, L.R. Gtg as translation initiation codon in the apocytochrome b gene of sunflower mitochondria. *DNA Seq.* **1996**, *6*, 365–369. [CrossRef] [PubMed]
- 53. Bendich, A.J. Structural analysis of mitochondrial DNA molecules from fungi and plants using moving pictures and pulsed-field gel electrophoresis. *J. Mol. Biol.* **1996**, 255, 564–588. [CrossRef]
- Sloan, D.B. One ring to rule them all? Genome sequencing provides new insights into the 'master circle' model of plant mitochondrial DNA structure. *New Phytol.* 2013, 200, 978–985. [CrossRef]
- Alverson, A.J.; Rice, D.W.; Dickinson, S.; Barry, K.; Palmer, J.D. Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. *Plant Cell* 2011, 23, 2499–2513. [CrossRef]
- Liu, F.; Fan, W.; Yang, J.B.; Xiang, C.L.; Mower, J.P.; Li, D.Z.; Zhu, A. Episodic and guanine-cytosine-biased bursts of intragenomic and interspecific synonymous divergence in ajugoideae (lamiaceae) mitogenomes. *New Phytol.* 2020, 228, 1107–1114. [CrossRef] [PubMed]
- Yu, R.; Sun, C.; Zhong, Y.; Liu, Y.; Sanchez-Puerta, M.V.; Mower, J.P.; Zhou, R. The minicircular and extremely heteroplasmic mitogenome of the holoparasitic plant *Rhopalocnemis phalloides*. *Curr. Biol.* 2022, 32, 470–479.e475. [CrossRef]
- 58. Wu, Z.; Cuthbert, J.M.; Taylor, D.R.; Sloan, D.B. The massive mitochondrial genome of the angiosperm *Silene noctiflora* is evolving by gain or loss of entire chromosomes. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 10185–10191. [CrossRef]
- 59. Lonsdale, D.M.; Hodge, T.P.; Fauron, C.M.-R. The physical map and organisation of the mitochondrial genome from the fertile cytoplasm of maize. *Nucleic Acids Res.* **1984**, *12*, 9249–9261. [CrossRef] [PubMed]
- 60. Palmer, J.D.; Shields, C.R. Tripartite structure of the brassica campestris mitochondrial genome. *Nature* **1984**, 307, 437–440. [CrossRef]
- 61. Zhe, M.; Zhang, L.; Liu, F.; Huang, Y.; Fan, W.; Yang, J.; Zhu, A. Plastid RNA editing reduction accompanied with genetic variations in cymbidium, a genus with diverse lifestyle modes. *Plant Divers.* **2022**, *44*, 316–321. [CrossRef] [PubMed]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.