Preprints.org

Article

Not peer-reviewed version

Ancient Horizontal Gene Transfers from Plastome to Mitogenome of a Nonphotosynthetic Orchid, *Gastrodia pubilabiata* (Epidendroideae, Orchidaceae)

Young-Kee Kim , Sangjin Jo , Se-Hwan Cheon , Ja-Ram Hong , Ki-Joong Kim *

Posted Date: 30 June 2023

doi: 10.20944/preprints202306.2204.v1

Keywords: Orchidaceae; Plastome; Mitogenome; Horizontal gene transfer; Mycoheterotorphy

Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 June 2023 doi:10.20944/preprints202306.2204.v1

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.

Article

Ancient Horizontal Gene Transfers from Plastome to Mitogenome of a Nonphotosynthetic Orchid, *Gastrodia pubilabiata* (Epidendroideae, Orchidaceae)

Young-Kee Kim [1], Sangjin Jo [2], Se-Hwan Cheon [1], Ja-Ram Hong [1] and Ki-Joong Kim [1,*]

[1] Division of Life Sciences, Korea University, Seoul, South Korea; gimyoung2@korea.ac.kr (Y.-K.K.); cheonsh@korea.ac.kr (S.-H.C.); ghdwkfka@korea.ac.kr (J.-R.H.); kimkj@korea.ac.kr (K.-J.K.)

[2] International Biological Material Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon, South Korea; sjjo0925@kribb.re.kr (S.J.)

* Correspondence: kimkj@korea.ac.kr

Abstract: *Gastrodia pubilabiata* is a nonphotosynthetic and mycoheterotrophic orchid belonging to subfamily Epidendroideae. Compared to other typical angiosperm species, the plastome of *G. pubilabiata* is dramatically reduced in size to be only 30,698 base pairs (bp). This reduction has led to the loss of most photosynthesis-related genes and some housekeeping genes in the plastome, which now only contains 19 protein coding genes, three tRNAs, and three rRNAs. This study decoded the entire mitogenome of *G. pubilabiata*, which consisted of 44 contigs with a total length of 867,349 bp. Its mitogenome contained 38 protein coding genes, nine tRNAs, and three rRNAs. To determine possible gene transfer events between the plastome and the mitogenome, individual BLASTN searches were conducted, using all available orchid plastome sequences and flowering plant mitogenome sequences. Plastid rRNA fragments were found at a high frequency in the mitogenome. Seven plastid protein coding gene fragments (*ndh*C, *ndh*J, *ndh*K, *psa*A, *psb*F, *rpo*B, and *rps*4) were also identified in the mitogenome of *G. pubilabiata*. Phylogenetic trees using these seven plastid protein coding gene fragments suggested that horizontal gene transfer (HGT) from plastome to mitogenome occurred before losses of photosynthesis related genes, leading to the lineage of *G. pubilabiata*. Compared to species phylogeny of the lineage of orchid, it was estimated that HGT might have occurred approximately 30 million years ago.

Keywords: Orchidaceae; plastome; mitogenome; horizontal gene transfer; Mycoheterotorphy

1. Introduction

The advancement in sequencing technology has led to a surge in research on nucleotide sequences of flowering plants and their use for evolutionary and phylogenetic studies. However, these studies have mainly focused on the plastome due to cost and time constraints in sequencing. As of February 2023, 7,588 species of angiosperm plastome sequences have been deposited in NCBI whereas only 347 mitogenome sequences have been deposited in NCBI. Research using plastome sequences has been conducted on various flowering plants, such as *Amborella* [1], magnolids [2], monocots [3], and rosid [4]. Despite challenges in sequencing, nuclear genomes of some flowering plants have been decoded, including *Amborella* [5], *Oryza* [6], *Phalaenopsis* [7], and *Triticum* [8].

Both mitochondria and chloroplast are essential organelles in plant function. The plastome is the most widely used genome in plant evolutionary study because it has relatively stable genome structure, size, and rich evolutionary information. In contrast to plastome, the mitogenome vary widely in size and structure across taxa even among different populations of same species [9,10]. The smallest mitogenome is 66 kb for *Viscum* [11], while the longest mitogenome is 11.3 mb for *Silene* [12]. It has been suggested that these differences result from genomic rearrangements, repetitive sequences, and/or foreign DNA insertions [13–15]. Due to many repeats, long length, and rearrangement of the mitogenome, it is difficult to decipher the overall structure of the mitogenome in majority of plant species. Frequent gene transfers have been reported in the mitogenome of

© 2023 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 June 2023 doi:10.20944/preprints202306.2204.v1

Valeriana of photosynthetic Dipsacales [16]. Intracellular gene transfer (IGT) and horizontal gene transfer (HGT) have been reported in the parasitic plant *Aeginetia* [17]. Gene transfer has been reported during the symbiotic relationship between fungi and orchids that are closely related to saprophytic life styles of the Orchidaceae [18]. The trace of gene transfer has been studied in a few orchid mitochondria. These imported sequences might be due to past biological contact [10,18–20].

The Orchidaceae are one of the largest families of flowering plants, containing approximately 25,000 species [21]. Recent studies have divided this family into five subfamilies: Apostasioideae, Vanilloideae, Orchidoideae, Cypripedioideae, and Epidendroideae [22]. Following the first complete plastome sequencing report of *Phalaenopsis aphrodite* [23], plastome sequencing reports have been published for all major lineages within Orchidaceae, such as *Corallorhiza* [24,25], *Cymbidium* [26], *Dendrobium* [27], *Holcoglossum* [28], *Neottia* [29], and *Vanilla* [30]. To date, at least 321 orchid plastome sequences have been deposited in NCBI (as of February 2023). The growing number of reported orchid plastome sequences has led to various studies of the Orchidaceae family, focusing on phylogenetics, evolution, and plastome structure [3,30–36]. These studies often concentrate on specific lineages such as epiphytic or non-photosynthetic orchids. There are a total of 43 genera that contain non-photosynthetic mycoheterotrophs [37]. Several plastome sequences of these types of orchids have been reported, such as *Aphyllorchis* [29], *Corallorhiza* [38], *Cyrtosia* [33], *Epipogium* [39], *Gastrodia* [40], *Hexalectris* [35], *Lecanorchis* [41] and *Rhizantella* [31]. Typical photosynthetic orchid plastomes contain 83 protein coding genes [23]. However, non-photosynthetic species such as *Epipogium roseum* have significantly reduced plastomes, containing only 17 protein coding genes [39]. This contraction of plastomes results in low levels of phylogenetically informative sites. Genes retained in these plastomes often exhibit elevated evolutionary rates [42], making it difficult to construct accurate phylogenetic trees and interpret their results [3,41].

For the Orchidaceae family, the mitogenome sequence has only been reported for *Gastrodia elata* [40] and nuclear genome sequences have only been reported for *Cymbidium* [43], *Dendrobium* [44], *Gastrodia* [40], and *Phalaenopsis* [7]. Despite active research into phylogenetic and evolutionary aspects of Orchidaceae, phylogenetic trees from different organelles are often inconsistent [41,45,46]. This might be due to factors such as incomplete lineage sorting, ancient plastome capture, and/or sampling issues [41]. To address these incongruences, more mitochondrial and nuclear genome sequence data are needed.

The species *Gastrodia pubilabiata* belongs to the tribe Gastrodieae of the subfamily Epidendroideae of the family Orchidaceae [22]. It is widely distributed in Korea, Japan, and Taiwan [47,48], but it is a rare and endangered species. For its closely related species, *G. elata*, plastome sequences along with nuclear and mitogenome sequences have been reported [40,41]. The plastome of *G. elata* is greatly reduced compared to that of a typical photosynthetic orchid. Flowers and stems of *G. pubilabiata* are pale greenish in color. In a previous study on *Cymbidium macrorhizon* [49] which also has pale greenish stems and flowers, the plastome was almost intact condition, although it is known as a non-photosynthetic species. Only a few *ndh* genes were missing from its plastome. Therefore, the extreme reduction of plastome sequence in *G. elata* raises the question of what type of plastome sequence will be present in *G. pubilabiata* which has pale greenish stems and flowers.

The current state of research on the plastome and mitogenome of non-photosynthetic orchids presents several difficulties, including a lack of parsimonious informative sites in the plastome and incongruence in phylogenetic trees based on different genomes. To address these issues, this study aimed to decode both of the plastome and mitogenome of the non-photosynthetic orchid *G. pubilabiata*. Through this research, organelle genomes of non-photosynthetic orchids will be better understood and interactions between different organelle genomes will be confirmed. In addition, any potential foreign nucleotide sequences in organelle genomes will be identified. Through this study, if any evidence of externally imported nucleotide sequences is found, the source of sequences will be investigated and the relationship with the taxon will be determined. Results of this study will contribute to the further understanding of the evolutionary modes of mitogenome in non-photosynthetic orchid species.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 June 2023 doi:10.20944/preprints202306.2204.v1

2. Results

2.1. Organelle characteristics and comparative analysis

The plastome sequence of *Gastrodia pubilabiata* was a 30,698 bp in length (Table 1 and Figure 1). The GC content of the plastome was 24.9% with a depth coverage of 236.0x. Unlike other reported orchid plastome sequences such as *Cypripedium* or *Habenaria*, the plastome of *G. pubilabiata* was much shorter in length. In addition, it lacked the typical quadripartite structure of angiosperm plastomes, which typically consist of a large single copy (LSC) and a small single copy (SSC) separated by two inverted repeats (IRs). Instead, the plastome of *G. pubilabiata*, like that of *G. elata*, lacked IRs. As a non-photosynthetic, mycoheterotrophic orchid without any leaves, *G. pubilabiata* lost all photosynthetic related genes, including *ndh*, *pet*, *psa*, *psb*, and *rpo* as well as several housekeeping genes, *atp* genes, *rrn*5, *ccs*A, *cem*A, *inf*A, *mat*K, *rbc*L, *ycf*3, and *ycf*4. Finally, the plastome of *G. pubilabiata* was annotated. It had only 19 protein coding genes, three tRNA genes, and three rRNA genes.

Table 1. The general features of *Gastrodia pubilabiata*'s plastome and mitogenome.

	Plastome	**Mitogenome**
Contigs	1	44
Total Length (bp)	30,698	867,349
GC (%)	24.9	42.1
Depth Coverage (x)	236.0	150.0
Protein Coding Genes	19	38
tRNAs	3	9
rRNAs	3	3

As a result of de novo assembly, 44 mitochondrial contigs of *G. pubilabiata* were obtained (Table 1). The goal was to obtain a complete, master circle of the mitogenome of *G. pubilabiata*. However, it was proved to be challenging due to the presence of AT-rich regions, rearrangements, and the recombination. These 44 mitochondrial contigs were considered as the complete mitogenome with a total length of, 867,349 bp (Table 1 and Figure 2). The coverage depth was 150.0x, with the highest coverage being 627.2x (14,604 bp) and the lowest coverage being 40.1x (4,510 bp). The GC content of the mitogenome was 42.1%. With these 44 mitochondrial contigs, a total of, 38 protein coding genes, nine tRNA genes, and three rRNA genes were annotated. Compared to other monocot mitogenomes, three genes (*rpl*10, *sdh*3, and *sdh*4), *nad*1 intron1, and *nad*1 intron2 were missing in the mitogenome of *G. pubilabiata* (Figure 2).

Repeat analysis was conducted to compare the mitogenome and plastome of *G. pubilabiata*. Three additional organelles were also included (the mitogenome of *G. elata*, the plastome of *G. elata*, and the plastome of *G. longistyla*) in this analysis. The analysis revealed a total of 437 repeats (Supplementary Table S1). Of these, 279 repeats were found in the mitogenome of *G. pubilabiata*, 79 repeats were shared between the mitogenomes of *G. pubilabiata* and *G. elata*, 44 repeats were shared among the three *Gastrodia* plastomes, 34 repeats were identified within the mitogenome of *G. elata*, and one repeat was shared between the *G. elata* plastome and mitogenome (Figure 3 and Supplementary Table S1). Additionally, 89 simple sequences repeats (SSRs) were found within the plastome of *G. pubilabiata*, including 16 mononucleotide SSRs, five dinucleotide SSRs, four trinucleotide SSRs, 13 tetranucleotide SSRs, and 51 pentanucleotide SSRs (Supplementary Table S2).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 June 2023 doi:10.20944/preprints202306.2204.v1

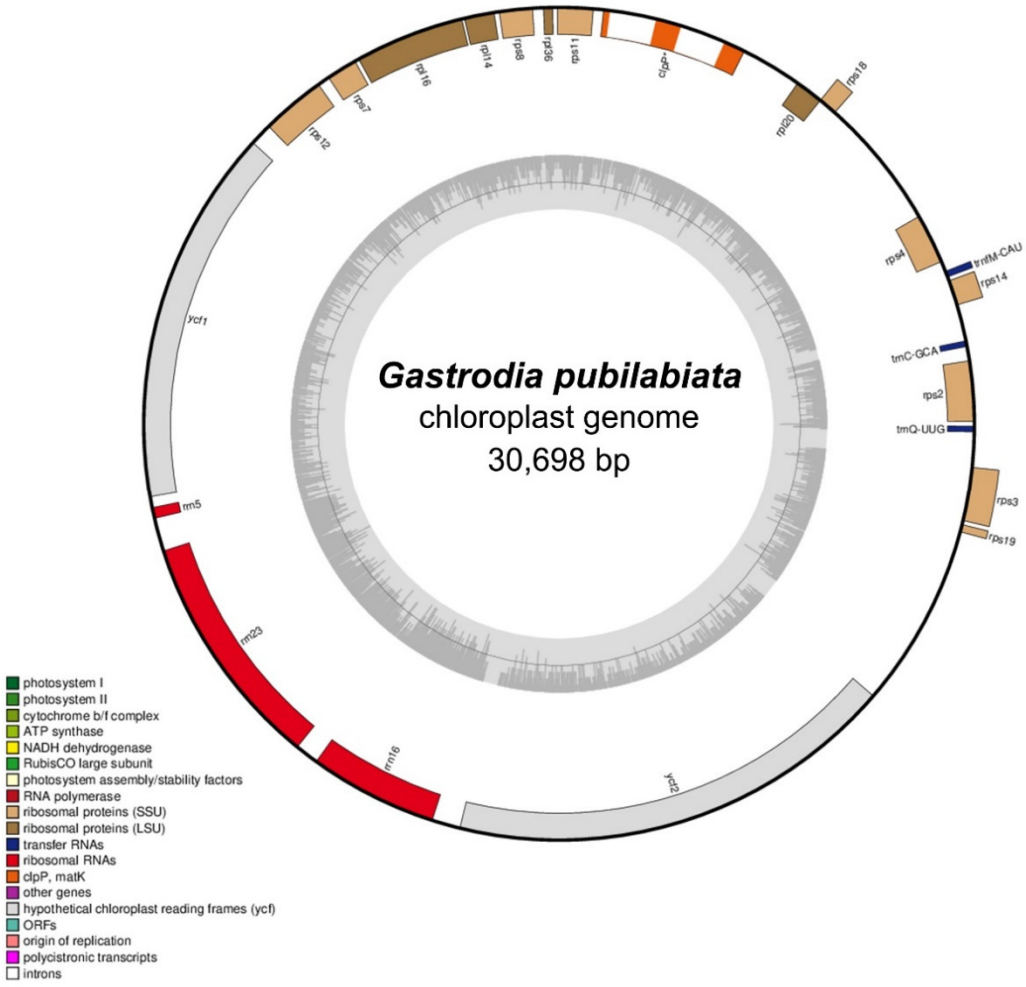

Figure 1. The complete plastome of Gastrodia pubilabiata. The plastome is 30.698 bp in length and lack IR region. It contains 19 protein coding genes, three tRNA genes, and three rRNA genes.

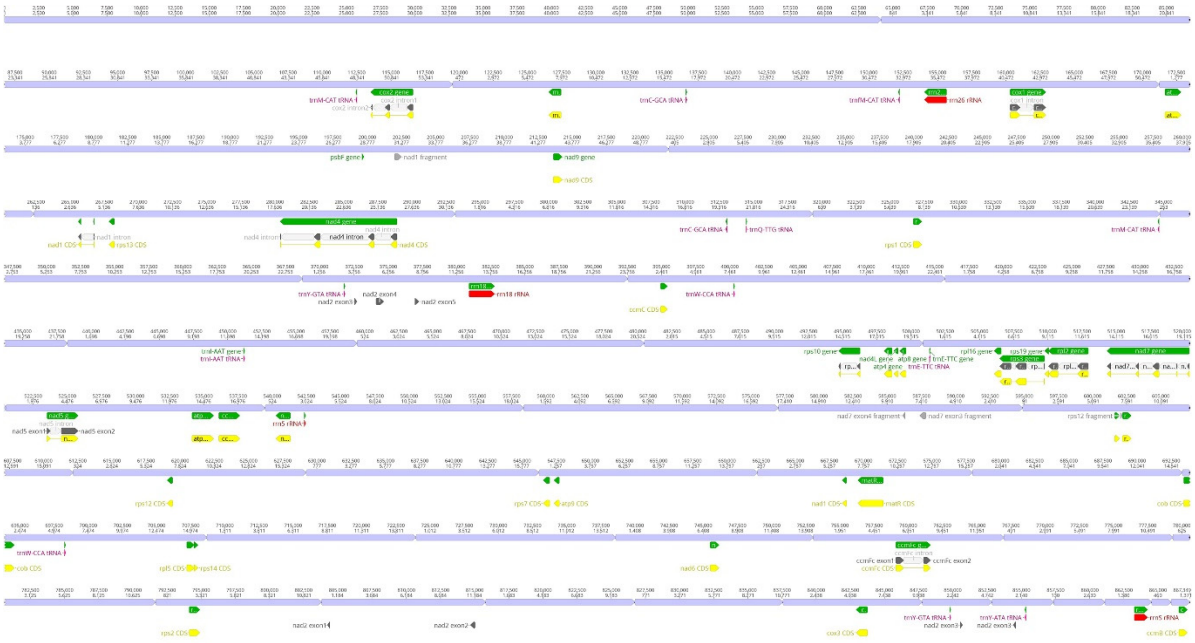

Figure 2. The complete mitogenome of *Gastrodia pubilabiata*. It is composed of 44 mitocondrial contigs. A total length of mitogenome is 867,349 bp in length and the genome include 38 protein coding genes, nine tRNA genes, and three rRNA genes.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 June 2023 doi:10.20944/preprints202306.2204.v1

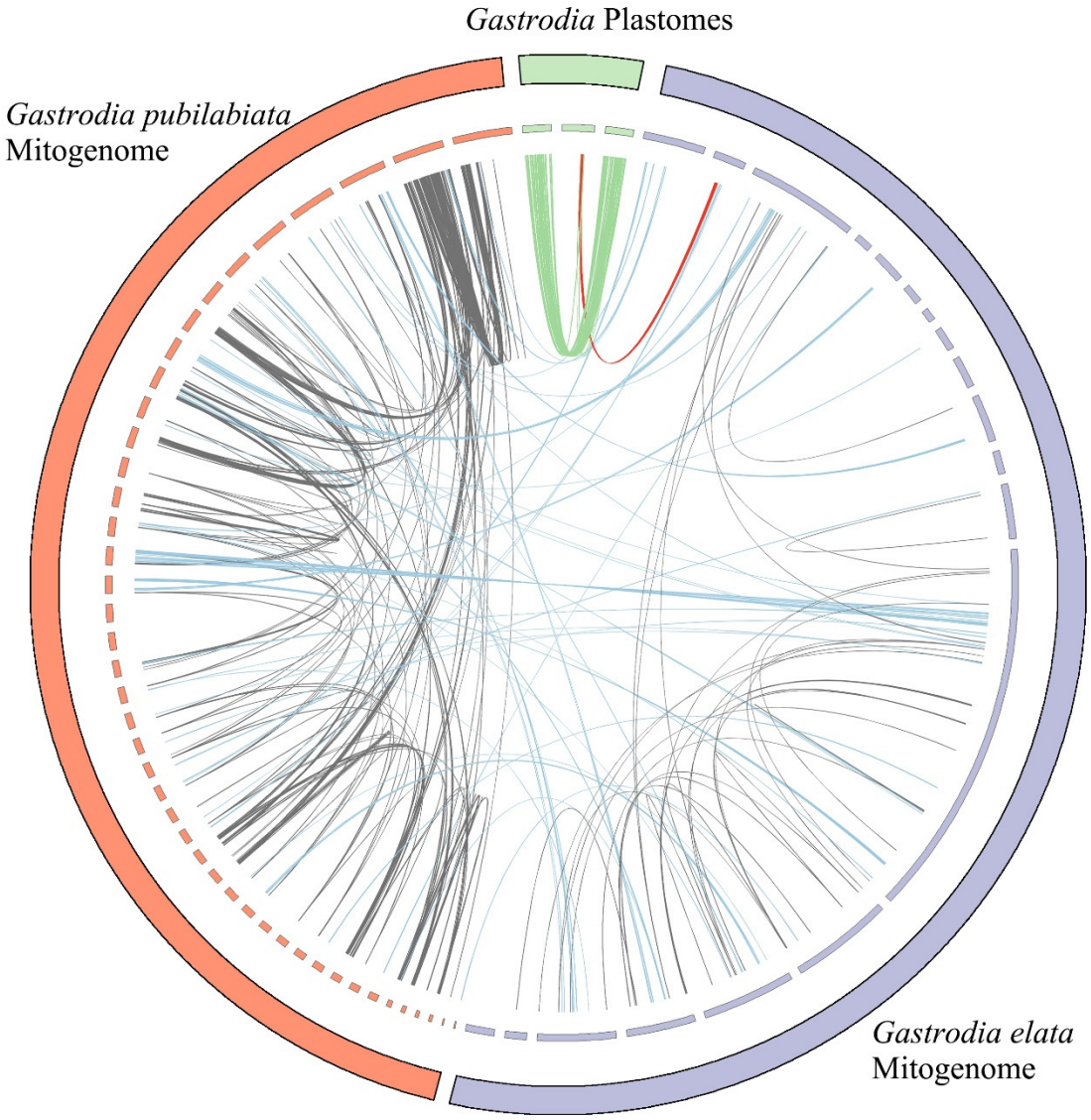

Figure 3. Results of repeat analysis between Gastrodia organelle genomes. The blue lines represent repeats between mitogenomes of different species, the green lines represent repeats between plastomes of different species, the grey lines indicate repeats between mitogenomes of the same species, and the red lines indicate repeats between plastomes and mitogenomes, respectively.

2.2. Foreign sequences in the organelle

BLASTN and BLASTP searches were carried out for the plastome of *G. pubilabiata* to identify exotic regions. Results of the BLASTN showed that most regions had the highest matches with plastome sequences of green plants. However, there were also matches with algae, bacteria, and animal sequences in regions corresponding to rRNAs (*rrn*4.5, *rrn*5, *rrn*18, and *rrn*23). Results of the BLASTP for open reading frames (ORFs) in the *G. pubilabiata* plastome confirmed that most ORFs were of green plant origin. In some cases, ORFs located near or within *rpl*16, *rrn*16, *rrn*23, and *rps*12 showed a high proportion of matches with bacterial sequences in BLASTP search results.

During annotation of protein coding genes in the plastome of *G. pubilabiata*, no debris from other mitogenome or nuclear genomes was found. However, during annotation of mitogenome, sequences like plastome derived *psb*F genes were detected, which was 119 bp in length. Regarding this regions, similar regions were found in other previously reported mitogenomes. These sequences were aligned and used to construct a maximum likelihood tree (Supplementary Figure S1 and Supplementary

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 June 2023 doi:10.20944/preprints202306.2204.v1

Figure S2). In this ML tree, all mitochondrial *psb*F-like regions formed a monophyletic group, indicating that mitochondrial and plastome *psb*F genes evolved independently.

The plastome and mitogenome of *G. pubilabiata* were searched using local BLASTN against two databases of published angiosperm plastomes and mitogenomes. As a result, 23 blocks of the *G. pubilabiata* plastome were found to match with the mitogenome database. Of these 23 blocks, 17 were related to *rrn*16 and *rrn*23, while the remaining were related to *clp*P, *rpl*2, *rpl*16, and *ycf*1. In the *G. pubilabiata* mitogenome, 118 blocks were identified as having plastome-related sequences, 82 of which had a higher degree of similarity to mitogenomes than to the plastome database. Thirty-six blocks were found to be plastome-related blocks, largely located in intergenic spacers or rRNAs. Six regions were found to be coding regions in the plastome (*ndh*C, *ndh*J, *ndh*K, *psa*A, *rpo*B, and *rps*4) (Supplementary Table S3). A phylogenetic tree was constructed using these regions after performing MAFFT alignment (Figure 4).

Figure 4. A maximum likelihood (ML) phylogenetic tree of six regions, found to be plastid-origin coding regions in the mitogenome (ndhC, ndhJ, ndhK, psaA, rpoB, and rps4). The position of Gastrodia pubilabiata is highlighted in red.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 June 2023 doi:10.20944/preprints202306.2204.v1

2.3. Phylogenetic position of Gastrodia pubilabiata

The phylogenetic position of *G. pubilabiata* was determined using 79 protein coding regions and four rRNAs in its plastome. Results showed that *G. pubilabiata*, *G. elata*, and *G. longystyla* formed a monophyletic group. All nodes in the *Gastrodia* clade were supported by 100% bootstrap values (Figure 5). General phylogenetic position of the Orchidaceae was found to be Apostasioideae[Vanilloideae[Cypripedioideae[Orchidoideae, Epidendroideae]]].

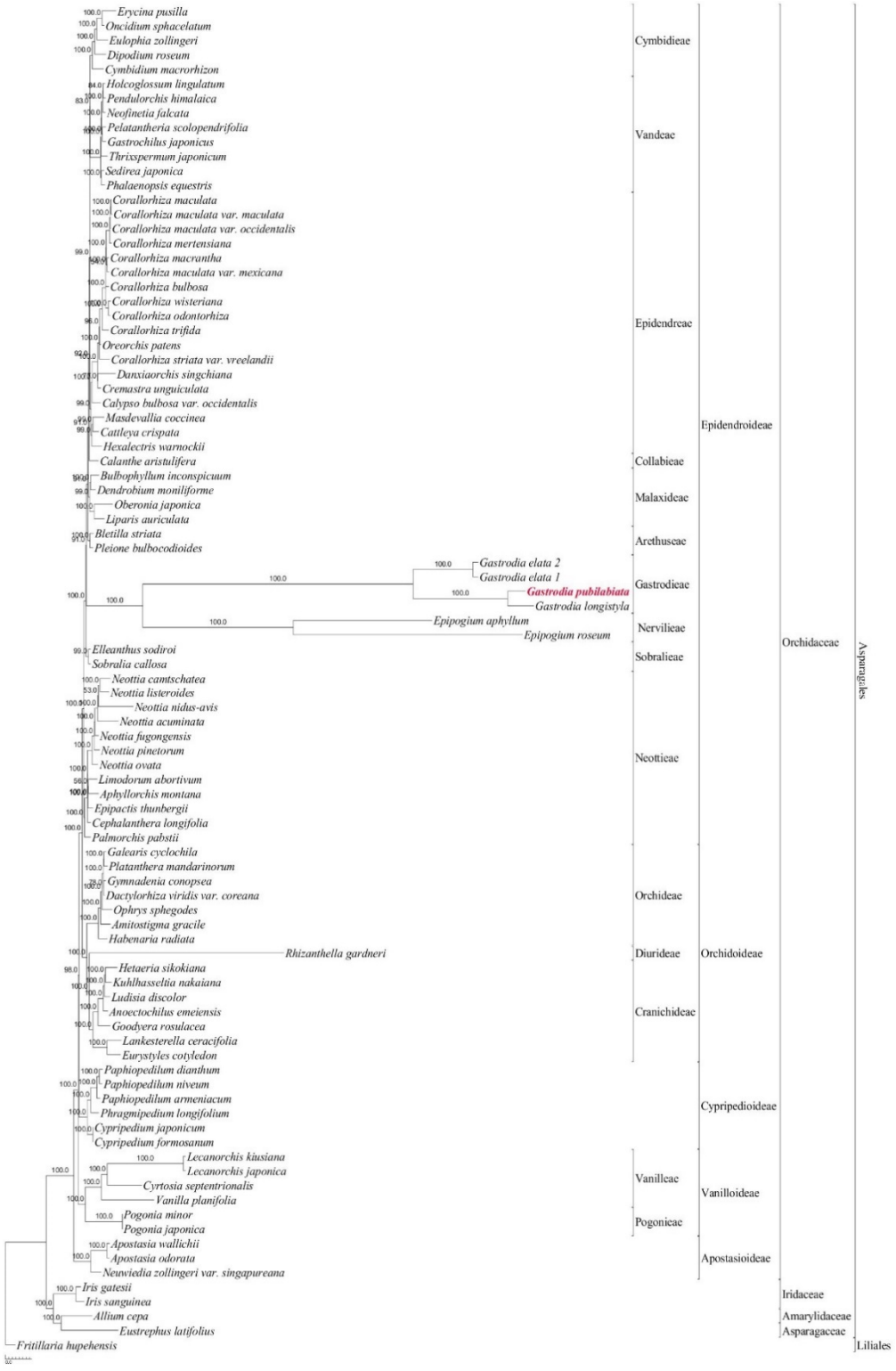

Figure 5. A maximum likelihood (ML) phylogenetic tree inferred from 92 species (87 species from Orchidaceae, four species from Asparagales, and one species from Liliales were used). The nucleotide

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 June 2023 doi:10.20944/preprints202306.2204.v1

sequences of 79 protein coding genes and four rRNA genes were aligned independently and concatenated to be a single datamatrix. The alignment was 85,749 bp and the tree was constructed by RaxML with the GTR + I + G model (-621773.060310 of ML value). The scientific name of Gastrodia pubilabiata was highlighted with red color. The numbers above or below the nodes are the bootstrap values.

As previously mentioned, studies decoding mitogenomes of orchids are limited. Thus, a phylogenetic tree was constructed using previously reported angiosperm mitogenomes. Results showed that *G. pubilabiata* and *G. elata* formed a monophyletic group and the genus *Gastrodia* formed a sister relationship with *Allium cepa* and *Asparagus officinalis*. Most species were supported by high bootstrap values (above 90%). However, species such as *Beta*, *Daucus*, *Fagus*, *Geranium*, *Luffa*, *Nelumbo*, and *Prunus* showed relatively lower bootstrap values. The phylogenetic tree based on the mitogenomes was in line with previously reported phylogenetic trees based on plastomes or nuclear genomes (Figure 6).

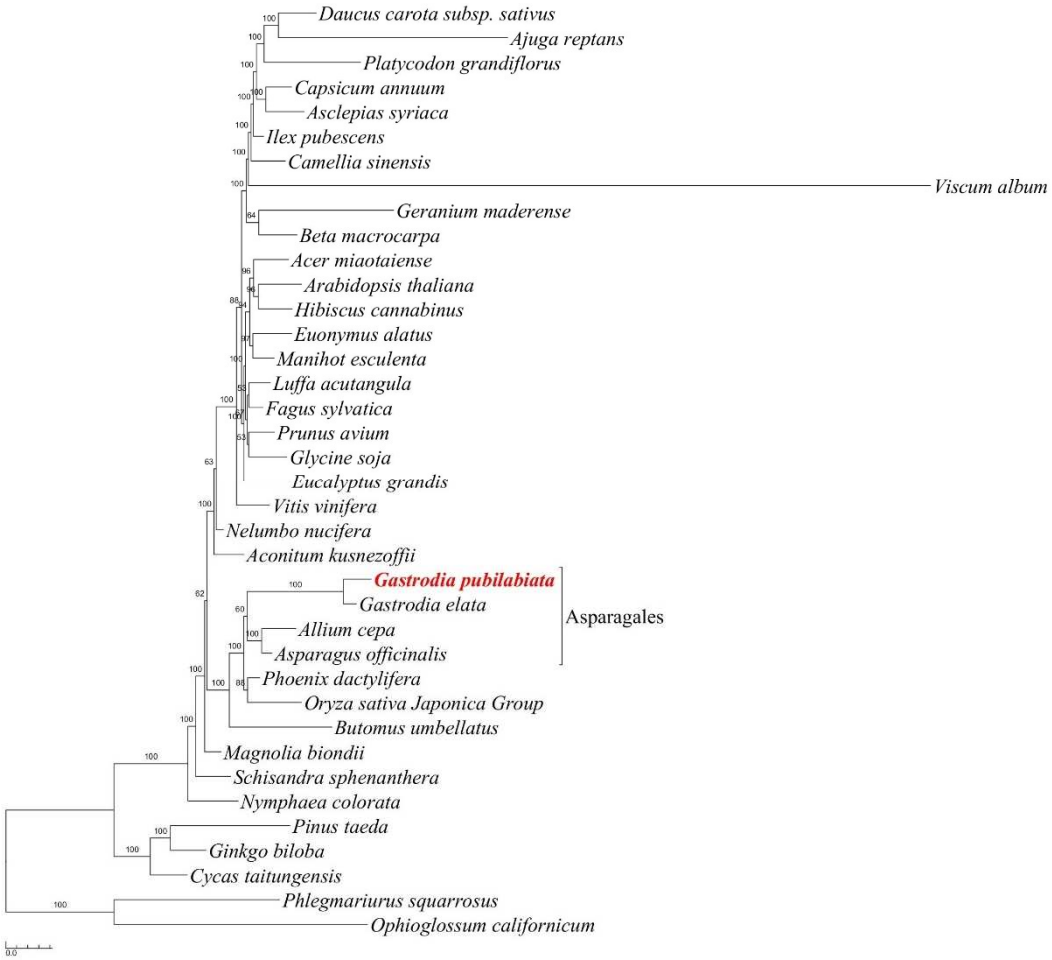

Figure 6. A maximum likelihood (ML) phylogenetic tree inferred from 38 speices. The nucleotide sequences of 35 protein coding genes and three rRNA genes were aligned independently and concatenated to be a single datamatrix. The alignment was 51,862 bp and the tree was constructed by RaxML with the GTR + I + G model (-318584.844612 of ML value). The scientific name of Gastrodia pubilabiata was highlighted with red color. The numbers above or below the nodes are the bootstrap values.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 June 2023 doi:10.20944/preprints202306.2204.v1

3. Discussion

3.1. Organelle genome evolution

The plastome of *Gastrodia pubilabiata* had a distinct feature in that it has extremely reduced genome size and consist of only a single copy region. This same plastome structure has previously been also reported in the same genus, *G. elata* [40]. Other taxa without IR include Orobanchaceae [50], Cactaceae [51], Geraniaceae [52,53], Fabaceae [54], and Gymnosperms [55,56]. In Fabaceae, a representative group of taxa lacking IR, deletions of genes or introns, rearrangements of the plastome structure, gene transfer, and hypermutations have been reported [57]. The lack of IR in the legume plastome might be a result of plastome rearrangement mediated by repetitive DNA [58]. These repeats are thought to have facilitated rearrangements by enabling recombination [53].

The plastome of *G. pubilabiata* contains simple sequence repeats (SSRs) located around *rps*2-*rps*14, *rps*2-*rps*3, and *ycf*2 regions (Supplementary Table S2). The gene *trn*Q-UUG is located between *rps*2 and *rps*14 and the *trn*C-GCA gene is situated near *rps*2 and *rps*3. Recombination of repeats and tRNAs might be the primary mechanism behind genomic rearrangements in *Trachelium* [59]. In a typical orchid plastome, *rps*2 is found near the LSC-IRa junction and *rps*3 is located near the LSC-IRb junction. *Rps*2 and *rps*14 are 20 kb apart. The change in position of *rps*2, *rps*3, and *rps*14 in *G. pubilabiata* is believed to be due to the presence of repeats or tRNAs in the intergenic spacer (IGS) near the relevant region. However, it is uncertain whether gene reduction, in which protein coding regions between corresponding genes are lost, precedes genome rearrangement and reduction [60]. Some studies have suggested that repetitive sequences contribute to genome rearrangement in the Geraniaceae [53]. Accumulation of SSRs and repeats in the plastome of *G. pubilabiata* might have also contributed to gene loss or plastome size reduction.

Several changes such as *ndh* gene contents, IR boundary shifts, and IR absence have been reported in plastomes of various orchid taxa [3,30,40,41,60]. For non-photosynthetic taxa, changes have also been reported in gene groups other than just *ndh* genes [41]. This level of change is consistent with previous reports for non-photosynthetic flowering plants [61]. In the case of *G. pubilabiata*, most genes have disappeared. The plastome of *G. pubilabiata* belongs to the fourth stage of plastome degradation pattern (Supplementary Figure S3). Other orchid taxa that belongs to stages 4 of plastome degradation, including *Epipogium* [39], *Gastrodia* [40], *Rhizanthella* [31], and *Lecanorchis* [41], are also non-photosynthetic mycoheterotrophs with greatly altered plastomes. These taxa have only retained some housekeeping genes.

Flowers and stems of *G. pubilabiata* have a green color, which sets them apart from the other fourth stages of orchid with red, white, or yellow flowers and stems. In contrast to the other green stemmed mycoheterotrophs *C. macrorhizon*, which only lacks *ndh* genes in its plastome [49], all photosynthesis related genes have disappeared from the confirmed plastome of *G. pubilabiata* (Figure 1 and Supplementary Figure S3). Despite this, low level photosynthesis has been observed in flowers and stems of *C. macrorhizon* due to residual chlorophyll [62–64]. Given this and the presence of residual plastids in other non-photosynthetic taxa, it is hypothesized that there might be an alternative pathway for photosynthesis in *G. pubilabiata*. Further research such as RNA-seq is needed to confirm this.

Mitogenomes of flowering plants are kwown to be variable in structure due to the presence of repeats, rearrangements, and submolecules [65]. Mitogenomes of monocots can vary greatly in length and genetic composition (Table 2 and Supplementary Figure S4). In the genus *Gastrodia*, multiple mitogenomes have been reported [40]. The mitogenome of *G. pubilabiata* was found to have 44 mitogenomes (Table 1 and Figure 2). Unlike the plastome, which displays a significant difference in genetic composition between photosynthetic and non-photosynthetic taxa, it is believed that gene composition of the mitogenome is not significantly impacted by photosynthesis or the lack thereof.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 June 2023 doi:10.20944/preprints202306.2204.v1

Table 2. The general features of reported mitogenomes of monocotyledon.

Scientific Name	Accesions	Contigs	Total Length (bp)	Protein Coding Genes	tRNAs	rRNAs
Gastrodia pubilabiata	OR004100-OR004143	44	867,349	38	9	3
Gastrodia elata	MF070084~MF070102	19	1,339,825	38	14	3
Allium cepa	NC030100	1	316,363	24	5	3
Asparagus officinalis	NC053642	1	492,062	36	14	3
Butomus umbellatus	NC021399	1	450,826	28	12	4
Oryza sativa	NC011033	1	490,520	31	15	3
Phoenix dactylifera	NC016740	1	715,001	37	16	3

3.2. Putative gene transfer

BLAST search was carried out to detect any foreign sequences in the plastome and mitogenome of *G. pubilabiata*. The search uncovered regions with high plastome frequency having mitochondrial features and regions with high mitochondrial frequency having plastome characteristics (Figure 7). Regions with high frequency of mitogenome even in the plastome were mostly made up of rRNA regions (Supplementary Table S3). Other protein coding regions in the plastome (*acc*D, *clp*P, *rpl*16, and *rpl*2) did show BLASTN results for the mitogenome. However, it was confirmed that there were many more search results for the plastome (Supplementary Table S3). The plastome might have infiltrated to the mitogenome, or there was a coincidental similarity of nucleotide sequences, rather than the presence of corresponding regions derived from the mitogenome. Reports of the presence of mitochondrial plastome sequences (MTPTs) have been documented in flowering plants [13]. A recent study has also reported the presence of plastome originating sequences in the mitogenome of *Mangifera* species [66].

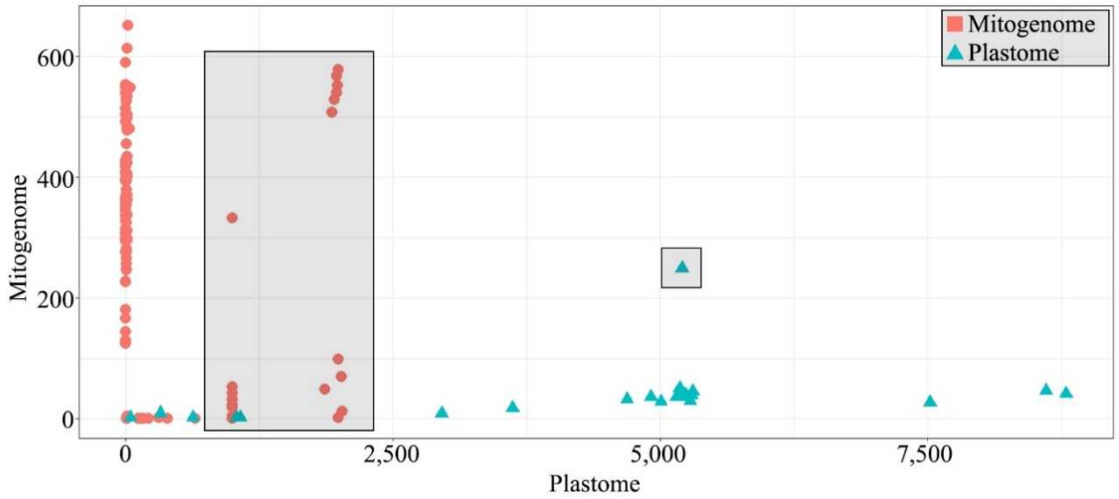

Figure 7. A dot-plot based on local BLASTN search results. Blue triangular dots represent plastome-origin sequences. Red circular dots indicate mitogenome-origin sequences. Irregular search results are highlighted with gray boxes.

BLASTN results for the mitogenome were like those for the plastome. Many results showing the high frequency for the plastome contained regions having rRNA, such as 16s rRNA, and 18s rRNA, suggesting that there might have been an exchange between rRNAs of the mitogenome and the plastome. However, for some regions, BLASTN results were found for plastome regions (*ndh*C, *ndh*J, *ndh*K, *psa*A, *psb*F, *rpo*B, and *rps*4), but not for mitogenome references. This suggests that gene transfer from the plastome to the mitogenome might have occurred. The mitogenome contains many

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 June 2023 doi:10.20944/preprints202306.2204.v1

photosynthesis-related genes from the plastome (*ndh*, *psb*, *psa, rpo*), even though *G. pubilabiata* is a non-phtoosynthetic plant and all photosynthesis related genes have been deleted from the current plastome (Figure 1 and Supplementary Figure S3). This suggests that transfer of photosynthesis related genes to the mitogenome might have occurred before the loss of photosynthetic function in plastome. Therefore, this gene transfer might have occurred a longer time ago than previously expected. Except for a few obligate mycoheterotrophic taxa, a previous study have shown that the divergence of mycoheterotrophs might have occurred earlier than 30 mya [41]. This suggests that photosynthesis-related genes might have disappeared after that time. Gene transfer might have occurred a longer than 30 mya. However, as pointed out in that previous study, research results on the plastome and mitogenome of related obligated mycoheterotrophic taxa will be needed to verify that hypothesis.

A phylogenetic tree was constructed to determine the origin of gene fragments (*ndh*, *psa*, *psb*, and *rpo*) in the mitogenome of *G. pubilabiata* with other major flowering plants. Results are shown in Figure 4. In the case of *psb*F, *G. pubilabiata* formed a monophyletic relationship with other mitochondrial *psb*F. It tended to be distinguished between the plastome and the mitogenome. This suggests that these gene fragments might have been transferred from the plastome to the mitogenome in the distant past. In the case of *psa*A, the mitogenome derived *psa*A displayed a long branch, which was believed to result from the absence of other mitogenome derived sequences with the same cause as the *psb*F phylogenetic tree. The phylogenetic tree using *rpo*B region was distinct from trees using *psa*A and *psb*F gene fragments. It did not show a long branch and *rpo*B sequences of *G. pubilabiata* from a monophyletic branch with *Euonymus* of Celastrales, which was supported by a bootstrap value of 100. This suggests that *G. pubilabiata* has a close relationship with Celastraceae plants in its vicinity. Similar relationships have been reported for parasitic plants such as *Aeginetia* with nearby Poaceae plants [17]. Although it is difficult to directly compare mycoheterotrophic orchid to parasitic plants, this orchid might be related to surrounding plants. It was cornfirmed that *Celastrus* and *Euonymus* lived some distance away, not nearby.

Unlike photosynthesis-related genes found in the mitogenome, the *rps*4 gene is classified as a housekeeping gene. The presence of this gene in both the plastome and mitogenome of *G. pubilabiata* was confirmed and a phylogenetic tree was constructed to examine its origin. The phylogenetic tree showed that both plastome and mitogenome derived sequences of the *rps*4 gene formed a monophyly with other orchids. However, the plastome derived sequence showed a long branch and formed a monophyly, while the mitogenome derived sequence was at the base of the clade. This difference in evolutionary rates between the two sequences was likely due to relaxed selection in degraded plastomes of mycoheterotrophic plants, which tended to drive the acceleration of gene evolution [67]. Moreover, mitochondrial datasets shows much slower evolutionary rates than the plastome datasets in the previous study [68]. Thus, it could be concluded that the *rps*4 gene in the plastome evolved at a faster evolutionary rate than the gene in the mitogenome.

It is known that *ndh* genes are typically the first to be lost in the plastome degradation process of non-photosynthetic species, followed by other photosynthesis related genes [50]. This study found that *ndh*C, J, and K genes in the mitogenome of *G. pubilabiata* formed a monophyly with other orchids, unlike photosynthesis-related gene *rpo*, *psa*, or *psb*, which showed relationships with other taxa of flowering plants (Figure 4). This suggested that the degradation of *G. pubilabiata*'s plastome was more recent or that alternative pathways were involved. Reference guided assembly analysis only found a few base sequences for *ndh*C, J, K in the mitogenome. No base sequence was found to support the possibility of transfer to the nuclear genome. Hence, it was assumed that *ndh*C, J, K genes disappeared more recently from the plastome or that the degradation of the plastome was more recent. While it was believed that non-photosynthesis in Orchidaceae developed within the past 30 mya [41], further experiments such as transcriptome analysis are needed to explore the possibility of alternative pathways.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 June 2023 doi:10.20944/preprints202306.2204.v1

3.3. Phylogenetic implications

Phylogenetic relationships within the Orchidaceae are presented as Apostasioideae[Vanilloideae[Cypripedioideae[Orchidoideae, Epidendroideae]]] based on the analysis of protein coding regions and rRNAs of orchid plastomes. This finding aligns with a previous study [3]. Three species of the genus *Gastrodia*, *G. elata*, *G. longyphyla*, and *G. pubilabiata*, formed a monophyletic group and the tribe Nervilieae was the sister group to Gastrodieae (Figure 5). In the case of Gastrodieae, Nervilieae, Diurideae, and Vanilleae, containing non-photosynthetic orchids, a long branch was formed, which might be due to reduced number of genes used in the analysis or elevated mutation rates in mycoheterotrophic species. A phylogenetic tree based on mitochondrial genes indicated that *G. pubilabiata* was part of a monophyletic group along with other Asparagales species (Figure 6). Unlike the plastome phylogenetic tree, non-photosynthetic species exhibited limited levels of long branches.

4. Materials and Methods

4.1. Plant material and DNA extraction

Plant samples were collected from Jeju Island, South Korea. To protect plant habitat, only a few flowers were taken from individuals of *Gastrodia pubilabiata*. Fresh samples were ground into powder using liquid nitrogen in a mortar. Ground samples were then used to extract genomic DNA using a G-spin II Genomic DNA extraction Kit (Intron, Seoul, Korea) following the manufacturer's manual. The quality of the extracted DNA was checked with a UV/VIS spectrophotometer. DNA was then stored in Plant DNA bank in Korea (PDBK2020-0086). Due to limitations of sample collection, the voucher specimen was replaced with photographs of *G. pubilabiata*'s in its habitat. These photographs were deposited in Korea University Herbarium (KUS) under the voucher-specimen number of KUS2020-0086 (Supplementary Table S4).

4.2. Sequencing, Assembly and Annotations

Extracted genomic DNA of *G. pubilabiata* was used for sequencing using two different NGS platforms: Illumina NovaSeq (Illumina, CA, US) and Pacbio Sequel (Pacbio, CA, US). A total of 224,926,562 NovaSeq reads, and 746,492 Pacbio Sequel reads were produced. The quality of sequencing reads was examined using FastQC v.0.11.9 [69]. Raw reads from Illumina NovaSeq were about 50 gb. They were trimmed with BBduk 37.64, as implemented in Geneious 11.1.5 (length: 27 kmer) [70]. BBNorm 37.64 was used to normalize trimmed reads (target coverage level: 30; minimum depth: 12). Raw reads from Pacbio Sequel were about 5 gb. They were corrected and polished using trimmed NOVAseq reads with CANU v.1.9 [71].

Error corrected Pacbio reads were used for de-novo assembly using FALCON v.0.3.0 [72,73] and CANU v.1.9. Mitogenome of *Allium cepa* (NC030100) and 18 mitochondrial sequences of *G. elata* (MF070084 ~ MF070102) were used as references to verify putative mitochondrial contigs among de novo assembly results obtained using FALCON and CANU. Trimmed NovaSeq reads and selected contigs from FALCON and CANU results were used for hybrid de novo assembly using SPAdes v.3.14.1 [74]. Mitochondrial contigs were selected from SPAdes results by performing local BLASTN [75] and using BANDAGE v.0.8.0. [76]. The final selection of mitogenome was based on their coverage, which was obtained using hybridSPAdes [77]. The depth-coverage of mitogenomes of *G. pubilabiata* is represented in Supplementary Table S5.

Plastome sequences of *G. elata* (NC037409) and *Habenaria radiata* (NC035834) were used as references to identify plastome contigs from results of FALCON and CANU. Selected contigs were assembled using Geneious assembler in Geneious 11.1.5 to obtain complete plastome sequence. Trimmed NovaSeq reads were then mapped to the complete plastome sequence of *G. pubilabiata*'s using Geneious assembler, and the mapped NovaSeq reads were extracted. Extracted NovaSeq reads were used to perform de novo assembly using Geneious assembler to validate *G. pubilabiata*'s complete plastome sequence.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 June 2023 doi:10.20944/preprints202306.2204.v1

The complete plastome sequence of *G. pubilabiata* was annotated using BLASTN, tRNAscan-SE 2.0 [78], ORF finder, and the find annotation function in Geneious 11.1.5. Plastome sequences of *G. elata* and *H. radiata* were used as references. Alternative start codons (ACG and TTG) were also considered in ORF finder. Pseudogene judgement was performed based on the criteria described previously [41]. ORFs that were larger than 150 bp were translated into protein sequences. These sequences were used to perform a psi-blast search [79] using several reported Orchidaceae plastome sequences as references. A circular plastome map was generated using the OGdraw web server [80].

4.3. Phylogenetic analysis

Ninety-two plastome sequences were downloaded from NCBI for phylogenetic analysis (Supplementary Table S6), including 87 sequences for the Orchidaceae, four for the Asparagales, and one for Liliales. Seventy-nine protein coding genes and four rRNA genes were extracted from these plastome sequences. Each extracted gene was aligned using MAFFT [81]. Alignments were manually checked. All alignments were concatenated to a length of 85,749 bp and subjected to jModeltest in the CIPRES Sience Gateway to determine the best fit model [82,83]. The best fit model was determined to GTR + I + G. Missing genes were treated as missing data. A maximum likelihood (ML) tree was constructed using RaxML-HPC2 on XSEDE in CIPRES Science Gateway with 100 bootstrap replicates [84]. The result ML tree was visualized using Treegraph2 [85].

Thirty-eight mitogenome sequences were downloaded from NCBI (Supplementary Table S6). Thirty-five coding genes and three rRNA genes were extracted from these mitogenomes. Each gene was aligned using MAFFT. All alignments were manually checked. Alignments were concatenated to a total length of 51.862 bp and then subjected to jModeltest in CIPRES Science Gateway to determine the best model. An ML tree was constructed using RaxML-HPC2 on XSEDE in CIPRES Science Gateway with a GTR + I + G model and 100 bootstrap replicates. The obtained ML tree was graphically represented using Treegraph2.

Eighteen mitochondrial sequences of *G. elata* (MF070084 ~ MF070102) and two plastome sequences of *G. elata* (NC037409) and *G. longistyla* (MW879162) were used to compare organelles among the genus *Gastrodia*. Simple sequence repeats (SSRs) were identified using Phobos 3.3.12 [86]. REPuter was used to distinguish palindromic repeats among organelles with default options [87]. The repeat finder in Geneious was used to find repeats larger than 50 bp with a perfect match option. Repeat information were summarized and presented graphically using Circos v.0.69-9 [88].

A total of 5,199 reported angiosperm plastome sequences and 219 mitogenome sequences were downloaded from NCBI and used as a local BLAST database. Plastome and mitogenome sequences of *G. pubilabiata* were divided into 150 bp fragments, which were then used to perform local BLASTN to identify potential instances of IGT or HGT. Results were summarized in terms of chloroplast counts and mitochondria counts based on the number of results generated. Summarized data were visualized using ggplot2 in R [89]. Based on these data, mitochondrial plastome (MTPT) sequences were extracted to construct a phylogenetic tree. Fifty-three plastome sequences of the main angiosperm lineage were used as references for this phylogenetic tree. For *psa*A and *rpo*B genes, a total of 301 Malpighiales and Celastrales plastome sequences were used as references to construct the ML tree. All obtained ML trees were visualized using Treegraph2.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org. Table S1: The repeats analysis results among plastomes and mitogenomes of the genus *Gastrodia*, Table S2: The list of simple sequences repeats of the *Gastrodia pubilabiata*'s plastome, Table S3: Summarized BLASTN results of putative gene transfers, Table S4: The NGS results of *Gastrodia pubilabiata* with two different method (NovaSeq and Pacbio Sequel), Table S5: The depth-coverage of mitogenomes of *Gastrodia pubilabiata*, Table S6: Used sequences information in this study, Figure S1: The alignment of mitochondrial *psb*F fragments, Figure S2: A maximum likelihood (ML) phylogenetic tree of *psb*F fragments with plastome origin *psb*F and mitochondrial *psb*F fragments, Figure S3: The plastome gene contents heatmap of Orchidaceae. Non-photosynthetic orchids were highlighted with red font color, Figure S4: The mitogenome gene contents heatmap of reported monocotyledon mitogenomes including two *Gastrodia* species.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 June 2023 doi:10.20944/preprints202306.2204.v1

Author Contributions: Conceptualization, K.-J.K.; Methodology, Y.-K.K.; Software, Y.-K.K.; Validation, Y.-K.K. and K.-J.K.; Formal analysis, Y.-K.K., S. Jo, S.-H.C. and J.-R.H.; Investigation, Y.-K.K.; Resources, K.-J.K.; Data curation, Y.-K.K., S. Jo, S.-H.C. and J.-R.H.; Writing—original draft, Y.-K.K.; Writing—review & editing, K.-J.K.; Supervision, K.-J.K.; Project administration, K.-J.K.; Funding acquisition, Y.-K.K. and K.-J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by National Research Foundation of Korea, NRF-2021R1A2C1013731 to K.-J.K. and NRF-2020R1A6A3A01100512 to Y.-K.K.

Data Availability Statement: The newly sequenced plastome and mitogenome of *Gastrodia pubilabiata* were submitted to GenBank and the accession number are OR004100-OR004143 for mitogenome and OR031839 for plastome, respectively.

Acknowledgments: We thanks to two anonymous reviewers to their valuable comment for the manuscript improvement. We also thanks to PDBK (Plant DNA Bank of Korea) for DNA curation and KUS (Korea University Herbarium) for voucher information curation, respectively.

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

Abbreviations

LSC	**Large Single Copy**
SSC	Small Single Copy
IR	Inverted Repeat
ML	Maximum Likelihood
HGT	Horizontal Gene Transfer
IGT	Intracellular Gene Transfer
SSR	Simple Sequence Repeat
MTPT	Mitochondrial Plastome sequences

References

1. Goremykin, V. V.; Hirsch-Ernst, K.I.; Wölfl, S.; Hellwig, F.H. Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm. *Mol. Biol. Evol.* **2003**, *20*, 1499–1505, doi:10.1093/molbev/msg159.
2. Cai, Z.; Penaflor, C.; Kuehl, J. V.; Leebens-Mack, J.; Carlson, J.E.; DePamphilis, C.W.; Boore, J.L.; Jansen, R.K. Complete plastid genome sequences of *Drimys*, *Liriodendron*, and *Piper*: Implications for the phylogenetic relationships of magnoliids. *BMC Evol. Biol.* **2006**, *6*, 1–20, doi:10.1186/1471-2148-6-77.
3. 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 Biol. Sci.* **2015**, *282*, 20151553, doi:10.1098/rspb.2015.1553.
4. Sun, M.; Naeem, R.; Su, J.X.; Cao, Z.Y.; Burleigh, J.G.; Soltis, P.S.; Soltis, D.E.; Chen, Z.D. Phylogeny of the Rosidae: A dense taxon sampling analysis. *J. Syst. Evol.* **2016**, *54*, 363–391, doi:10.1111/jse.12211.
5. DePamphilis, C.W.; Palmer, J.D.; Rounsley, S.; Sankoff, D.; Schuster, S.C.; Ammiraju, J.S.S.; Barbazuk, W.B.; Chamala, S.; Chanderbali, A.S.; Determann, R.; et al. The *Amborella* genome and the evolution of flowering plants. *Science (80-.).* **2013**, *342*, doi:10.1126/science.1241089.
6. Chen, J.; Huang, Q.; Gao, D.; Wang, J.; Lang, Y.; Liu, T.; Li, B.; Bai, Z.; Luis Goicoechea, J.; Liang, C.; et al. Whole-genome sequencing of *Oryza brachyantha* reveals mechanisms underlying *Oryza* genome evolution. *Nat. Commun.* **2013**, *4*, doi:10.1038/ncomms2596.
7. Liu, Z.J. The genome sequence of the orchid *Phalaenopsis equestris*. *Nat. Genet.* **2015**, *47*, 65–72, doi:10.1038/ng.3149.
8. Mayer, K.F.X.; Marcussen, T.; Sandve, S.R.; Heier, L.; Pfeifer, M.; Kugler, K.G.; Zhan, B.; Spannagl, M.; Pfeifer, M.; Jakobsen, K.S.; et al. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome Ancient hybridizations among the ancestral genomes of bread wheat Genome interplay in the grain transcriptome of hexaploid bread wheat Structural and functional pa. *Science* **2014**, *345*, 1250092.
9. Wu, Z.; Sloan, D.B. Recombination and intraspecific polymorphism for the presence and absence of entire chromosomes in mitochondrial genomes. *Heredity (Edinb).* **2019**, *122*, 647–659, doi:10.1038/s41437-018-0153-3.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 June 2023 doi:10.20944/preprints202306.2204.v1

10. Lin, Y.; Li, P.; Zhang, Y.; Akhter, D.; Pan, R.; Fu, Z.; Huang, M.; Li, X.; Feng, Y. Unprecedented organelle genomic variations in morning glories reveal independent evolutionary scenarios of parasitic plants and the diversification of plant mitochondrial complexes. *BMC Biol.* **2022**, *20*, 1–16, doi:10.1186/s12915-022-01250-1.
11. Skippingtona, E.; Barkmanb, T.J.; Ricea, D.W.; Palmera, J.D. Miniaturized mitogenome of the parasitic plant viscum scurruloideum is extremely divergent and dynamic and has lost all nad genes. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E3515–E3524, doi:10.1073/pnas.1504491112.
12. 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*, doi:10.1371/journal.pbio.1001241.
13. Sloan, D.B.; Wu, Z. History of plastid DNA insertions reveals weak deletion and AT mutation biases in angiosperm mitochondrial genomes. *Genome Biol. Evol.* **2014**, *6*, 3210–3221, doi:10.1093/gbe/evu253.
14. Wynn, E.L.; Christensen, A.C. Repeats of unusual size in plant mitochondrial genomes: Identification, incidence and evolution. *G3 Genes, Genomes, Genet.* **2019**, *9*, 549–559, doi:10.1534/g3.118.200948.
15. Gandini, C.L.; Garcia, L.E.; Abbona, C.C.; Sanchez-Puerta, M.V. The complete organelle genomes of *Physochlaina orientalis*: Insights into short sequence repeats across seed plant mitochondrial genomes. *Mol. Phylogenet. Evol.* **2019**, *137*, 274–284, doi:10.1016/j.ympev.2019.05.012.
16. Kim, H.T.; Kim, J.S. Structural mutations in the organellar genomes of *Valeriana sambucifolia* f. *dageletiana* (Nakai. ex Maekawa) hara show dynamic gene transfer. *Int. J. Mol. Sci.* **2021**, *22*, doi:10.3390/ijms22073770.
17. Choi, K.S.; Park, S. Complete plastid and mitochondrial genomes of *Aeginetia indica* reveal intracellular gene transfer (IGT), horizontal gene transfer (HGT), and cytoplasmic male sterility (cms). *Int. J. Mol. Sci.* **2021**, *22*, doi:10.3390/ijms22116143.
18. Sinn, B.T.; Barrett, C.F. Ancient Mitochondrial Gene Transfer between Fungi and the Orchids. *Mol. Biol. Evol.* **2020**, *37*, 44–57, doi:10.1093/molbev/msz198.
19. 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 (80-.).* **2013**, *342*, 1468–1473, doi:10.1126/science.1246275.
20. Sanchez-Puerta, M.V.; Edera, A.; Gandini, C.L.; Williams, A. V.; Howell, K.A.; Nevill, P.G.; Small, I. Genome-scale transfer of mitochondrial DNA from legume hosts to the holoparasite *Lophophytum mirabile* (Balanophoraceae). *Mol. Phylogenet. Evol.* **2019**, *132*, 243–250, doi:10.1016/j.ympev.2018.12.006.
21. Christenhusz, M.J.M.; Byng, J.W. The number of known plants species in the world and its annual increase. *Phytotaxa* **2016**, *261*, 201–217.
22. 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, doi:10.1111/boj.12234.
23. Chang, C.C.; Lin, H.C.; Lin, I.P.; Chow, T.Y.; Chen, H.H.; Chen, W.H.; Cheng, C.H.; Lin, C.Y.; Liu, S.M.; Chang, C.C.; et al. The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): Comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. *Mol. Biol. Evol.* **2006**, *23*, 279–291, doi:10.1093/molbev/msj029.
24. 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, doi:10.3732/ajb.1200256.
25. Barrett, C.F.; Wicke, S.; Sass, C. Dense infraspecific sampling reveals rapid and independent trajectories of plastome degradation in a heterotrophic orchid complex. *New Phytol.* **2018**, doi:10.1111/nph.15072.
26. 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, doi:10.1186/1471-2148-13-84.
27. Niu, Z.; Xue, Q.; Zhu, S.; Sun, J.; Liu, W.; Ding, X. The Complete Plastome Sequences of Four Orchid Species: Insights into the Evolution of the Orchidaceae and the Utility of Plastomic Mutational Hotspots. *Front. Plant Sci.* **2017**, *8*, 1–11, doi:10.3389/fpls.2017.00715.
28. Li, Z.H.; Ma, X.; Wang, D.Y.; Li, Y.X.; Wang, C.W.; Jin, X.H. Evolution of plastid genomes of *Holcoglossum* (Orchidaceae) with recent radiation. *BMC Evol. Biol.* **2019**, *19*, 1–10, doi:10.1186/s12862-019-1384-5.
29. Feng, Y.L.; Wicke, S.; Li, J.W.; Han, Y.; Lin, C.S.; Li, D.Z.; Zhou, T.T.; Huang, W.C.; Huang, L.Q.; Jin, X.H. Lineage-specific reductions of plastid genomes in an orchid tribe with partially and fully mycoheterotrophic species. *Genome Biol. Evol.* **2016**, *8*, 2164–2175, doi:10.1093/gbe/evw144.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 June 2023 doi:10.20944/preprints202306.2204.v1

30. Lin, C.S.; Chen, J.J.W.; 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*, 1–10, doi:10.1038/srep09040.
31. Delannoy, E.; Fujii, S.; Colas Des Francs-Small, C.; Brundrett, M.; Small, I. Rampant Gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes. *Mol. Biol. Evol.* **2011**, *28*, 2077–2086, doi:10.1093/molbev/msr028.
32. Kim, J.S.; Kim, H.T.; Kim, J.-H. The largest plastid genome of monocots: a novel genome type containing AT residue repeats in the slipper orchid *Cypripedium japonicum*. *Plant Mol. Biol. Report.* **2015**, *33*, 1210–1220.
33. Kim, Y.K.; Jo, S.; Cheon, S.H.; Joo, M.J.; Hong, J.R.; Kwak, M.H.; Kim, K.J. Extensive losses of photosynthesis genes in the plastome of a mycoheterotrophic orchid, *Cyrtosia septentrionalis* (Vanilloideae: Orchidaceae). *Genome Biol. Evol.* **2019**, *11*, 565–571, doi:10.1093/gbe/evz024.
34. Freudenstein, J. V.; Chase, M.W.M.; Lanfear, R.; Synergy, B.; Phytol, N.; Page, V.I.; De, A.A.P.; Wu, G.A.; Terol, J.; Ibanez, V.; et al. Phylogenetic relationships in Epidendroideae (Orchidaceae), one of the great flowering plant radiations: Progressive specialization and diversification. *PLoS One* **2017**, *7*, 1–12, doi:10.1093/aob/mcu253.
35. Barrett, C.F.; Kennedy, A.H. Plastid genome degradation in the endangered, mycoheterotrophic, North American orchid *Hexalectris warnockii*. *Genome Biol. Evol.* **2018**, *10*, 1657–1662, doi:10.1093/gbe/evy107.
36. Guo, Y.Y.; Yang, J.X.; Bai, M.Z.; Zhang, G.Q.; Liu, Z.J. The chloroplast genome evolution of Venus slipper (*Paphiopedilum*): IR expansion, SSC contraction, and highly rearranged SSC regions. *BMC Plant Biol.* **2021**, *21*, 1–14, doi:10.1186/s12870-021-03053-y.
37. Merckx, V.S.F.T.; Freudenstein, J. V; Kissling, J.; Christenhusz, M.J.M.; Stotler, R.E.; Crandall-Stotler, B.; Wickett, N.; Rudall, P.J.; Maas-van de Kamer, H.; Maas, P.J.M. Taxonomy and Classification BT - Mycoheterotrophy: The Biology of Plants Living on Fungi. In; Merckx, V., Ed.; Springer New York: New York, NY, 2013; pp. 19–101 ISBN 978-1-4614-5209-6.
38. Barrett, C.F.; Freudenstein, J. V.; Li, J.; Mayfield-Jones, D.R.; Perez, L.; Pires, J.C.; Santos, C. Investigating the path of plastid genome degradation in an early-transitional clade of heterotrophic orchids, and implications for heterotrophic angiosperms. *Mol. Biol. Evol.* **2014**, *31*, 3095–3112, doi:10.1093/molbev/msu252.
39. Schelkunov, M.I.; Shtratnikova, V.Y.; Nuraliev, M.S.; Selosse, M.A.; Penin, A.A.; Logacheva, M.D. Exploring the limits for reduction of plastid genomes: A case study of the mycoheterotrophic orchids *Epipogium aphyllum* and *Epipogium roseum*. *Genome Biol. Evol.* **2015**, *7*, 1179–1191, doi:10.1093/gbe/evv019.
40. 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*, doi:10.1038/s41467-018-03423-5.
41. Kim, Y.K.; Jo, S.; Cheon, S.H.; Joo, M.J.; Hong, J.R.; Kwak, M.; Kim, K.J. Plastome Evolution and Phylogeny of Orchidaceae, With 24 New Sequences. *Front. Plant Sci.* **2020**, *11*, 1–27, doi:10.3389/fpls.2020.00022.
42. Lam, V.K.Y.; Darby, H.; Merckx, V.S.F.T.; Lim, G.; Yukawa, T.; Neubig, K.M.; Abbott, J.R.; Beatty, G.E.; Provan, J.; Soto Gomez, M.; et al. Phylogenomic inference in extremis: A case study with mycoheterotroph plastomes. *Am. J. Bot.* **2018**, *105*, 480–494, doi:10.1002/ajb2.1070.
43. Yang, F.X.; Gao, J.; Wei, Y.L.; Ren, R.; Zhang, G.Q.; Lu, C.Q.; Jin, J.P.; Ai, Y.; Wang, Y.Q.; Chen, L.J.; et al. The genome of *Cymbidium sinense* revealed the evolution of orchid traits. *Plant Biotechnol. J.* **2021**, *19*, 2501–2516, doi:10.1111/pbi.13676.
44. Zhang, G.-Q.; Xu, Q.; Bian, C.; Tsai, W.-C.; Yeh, C.-M.; Liu, K.-W.; Yoshida, K.; Zhang, L.-S.; Chang, S.-B.; Chen, F. The *Dendrobium catenatum* Lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution. *Sci. Rep.* **2016**, *6*, 19029.
45. Zhang, G.Q.; Chen, G.Z.; Chen, L.J.; Zhai, J.W.; Huang, J.; Wu, X.Y.; Li, M.H.; Peng, D.H.; Rao, W.H.; Liu, Z.J.; et al. Phylogenetic incongruence in *Cymbidium* orchids. *Plant Divers.* **2021**, *43*, 452–461, doi:10.1016/j.pld.2021.08.002.
46. Li, Y.X.; Li, Z.H.; Schuitman, A.; Chase, M.W.; Li, J.W.; Huang, W.C.; Hidayat, A.; Wu, S.S.; Jin, X.H. Phylogenomics of Orchidaceae based on plastid and mitochondrial genomes. *Mol. Phylogenet. Evol.* **2019**, *139*, 106540, doi:10.1016/j.ympev.2019.106540.
47. Committee, F. of K.E.; Park, C.-W. *The genera of vascular plants of Korea*; Academy Publ., 2007; ISBN 897616380X.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 June 2023 doi:10.20944/preprints202306.2204.v1

48. Chen, X.Q.; Liu, Z.J.; Zhu, G.H.; Lang, K.Y.; Ji, Z.H.; Luo, Y.B.; Jin, X.H.; Cribb, P.J.; Wood, J.J.; Gale, S.W. Flora of China: Orchidaceae. *Chen XQ, Wood JJ* **2009**, *175*, 477.
49. 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, doi:10.1007/s00606-017-1472-1.
50. Wicke, S.; Muller, K.F.; de Pamphilis, C.W.; Quandt, D.; Wickett, N.J.; Zhang, Y.; Renner, S.S.; Schneeweiss, G.M. Mechanisms of Functional and Physical Genome Reduction in Photosynthetic and Nonphotosynthetic Parasitic Plants of the Broomrape Family. *Plant Cell* **2013**, *25*, 3711–3725, doi:10.1105/tpc.113.113373.
51. Sanderson, M.J.; Copetti, D.; Burquez, A.; Bustamante, E.; Charboneau, J.L.M.; Eguiarte, L.E.; Kumar, S.; Lee, H.O.; Lee, J.; McMahon, M.; et al. Exceptional reduction of the plastid genome of saguaro cactus (*Carnegiea gigantea*): loss of the *ndh* gene suite and inverted repeat. *Am. J. Bot.* **2015**, *102*, 1115–1127, doi:10.3732/ajb.1500184.
52. Guisinger, M.M.; Kuehl, J. V; Boore, J.L.; Jansen, R.K. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. *Mol. Biol. Evol.* **2011**, *28*, 583–600.
53. Blazier, J.C.; Jansen, R.K.; Mower, J.P.; Govindu, M.; Zhang, J.; Weng, M.L.; Ruhlman, T.A. Variable presence of the inverted repeat and plastome stability in Erodium. *Ann. Bot.* **2016**, *117*, 1209–1220, doi:10.1093/aob/mcw065.
54. Palmer, J.D.; Thompson, W.F. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. *Cell* **1982**, *29*, 537–550, doi:10.1016/0092-8674(82)90170-2.
55. 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.
56. Raubeson, L.A.; Jansen, R.K. A rare chloroplast-DNA structural mutation is shared by all conifers. *Biochem. Syst. Ecol.* **1992**, *20*, 17–24.
57. Moghaddam, M.; Ohta, A.; Shimizu, M.; Terauchi, R.; Kazempour-Osaloo, S. The complete chloroplast genome of *Onobrychis gaubae* (Fabaceae-Papilionoideae): comparative analysis with related IR-lacking clade species. *BMC Plant Biol.* **2022**, *22*, 75.
58. Wu, S.; Chen, J.; Li, Y.; Liu, A.; Li, A.; Yin, M.; Shrestha, N.; Liu, J.; Ren, G. Extensive genomic rearrangements mediated by repetitive sequences in plastomes of *Medicago* and its relatives. *BMC Plant Biol.* **2021**, *21*, 1–16, doi:10.1186/s12870-021-03202-3.
59. Haberle, R.C.; Fourcade, H.M.; Boore, J.L.; Jansen, R.K. Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. *J. Mol. Evol.* **2008**, *66*, 350–361, doi:10.1007/s00239-008-9086-4.
60. 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 associated instability of the inverted repeat/small single-copy region boundaries. *PLoS One* **2015**, *10*, doi:10.1371/journal.pone.0142215.
61. Wicke, S.; Müller, K.F.; dePamphilis, C.W.; Quandt, D.; Bellot, S.; Schneeweiss, G.M. Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. *Proc. Natl. Acad. Sci.* **2016**, *113*, 9045–9050, doi:10.1073/pnas.1607576113.
62. Suetsugu, K.; Ohta, T.; Tayasu, I. Partial mycoheterotrophy in the leafless orchid *Cymbidium macrorhizon*. *Am. J. Bot.* **2018**, *105*, 1595–1600, doi:10.1002/ajb2.1142.
63. Blumenfeld, H. Beitrage zur Physiologie des Wurzepilzes von Limodorum abortivum (L.) Sw; Universitat Basel., 1935;
64. Girlanda, M.; Selosse, M.A.; Cafasso, D.; Brilli, F.; Delfine, S.; Fabbian, R.; Ghignone, S.; Pinelli, P.; Segreto, R.; Loreto, F.; et al. Inefficient photosynthesis in the Mediterranean orchid *Limodorum abortivum* is mirrored by specific association to ectomycorrhizal Russulaceae. *Mol. Ecol.* **2006**, *15*, 491–504, doi:10.1111/j.1365-294X.2005.02770.x.
65. 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, doi:10.1111/nph.12395.
66. Niu, Y.; Gao, C.; Liu, J. Complete mitochondrial genomes of three *Mangifera* species, their genomic structure and gene transfer from chloroplast genomes. *BMC Genomics* **2022**, *23*, 1–8, doi:10.1186/s12864-022-08383-1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 June 2023 doi:10.20944/preprints202306.2204.v1

67. Kim, Y.; Cheon, S.; Hong, J.; Kim, K. Evolutionary Patterns of the Chloroplast Genome in Vanilloid Orchids (Vanilloideae , Orchidaceae). **2023**.
68. Fonseca, L.H.M.; Lohmann, L.G. Exploring the potential of nuclear and mitochondrial sequencing data generated through genome-skimming for plant phylogenetics: A case study from a clade of neotropical lianas. *J. Syst. Evol.* **2020**, *58*, 18–32, doi:10.1111/jse.12533.
69. Andrews, S. FastQC: a quality control tool for high throughput sequence data 2010.
70. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649.
71. 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.
72. Chin, C.-S.; Peluso, P.; Sedlazeck, F.J.; Nattestad, M.; Concepcion, G.T.; Clum, A.; Dunn, C.; O'Malley, R.; Figueroa-Balderas, R.; Morales-Cruz, A. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **2016**, *13*, 1050–1054.
73. Chin, C.-S.; Alexander, D.H.; Marks, P.; Klammer, A.A.; Drake, J.; Heiner, C.; Clum, A.; Copeland, A.; Huddleston, J.; Eichler, E.E. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **2013**, *10*, 563–569.
74. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477.
75. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410.
76. Wick, R.R.; Schultz, M.B.; Zobel, J.; Holt, K.E. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* **2015**, *31*, 3350–3352.
77. Antipov, D.; Korobeynikov, A.; McLean, J.S.; Pevzner, P.A. HybridSPAdes: An algorithm for hybrid assembly of short and long reads. *Bioinformatics* **2016**, *32*, 1009–1015, doi:10.1093/bioinformatics/btv688.
78. 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, doi:10.1093/nar/gkw413.
79. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402.
80. Lohse, M.; Drechsel, O.; Bock, R. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* **2007**, *52*, 267–274.
81. Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066, doi:10.1093/nar/gkf436.
82. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **2012**, *9*, 772.
83. Miller, M.A.; Pfeiffer, W.; Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In Proceedings of the 2010 gateway computing environments workshop (GCE); Ieee, 2010; pp. 1–8.
84. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313.
85. Stöver, B.C.; Müller, K.F. TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics* **2010**, *11*, 1–9, doi:10.1186/1471-2105-11-7.
86. Mayer, C. Phobos Version 3.3. 12. *A tandem repeat search Progr.* **2010**, *20*.
87. Kurtz, S. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **2001**, *29*, 4633–4642, doi:10.1093/nar/29.22.4633.
88. Krzywinski, M.; Schein, J.; Birol, I.; Connors, J.; Gascoyne, R.; Horsman, D.; Jones, S.J.; Marra, M.A. Circos: an information aesthetic for comparative genomics. *Genome Res.* **2009**, *19*, 1639–1645.
89. Wickham, H.; Chang, W.; Wickham, M.H. Package 'ggplot2.' *Creat. Elegant Data Vis. Using Gramm. Graph. Version* **2016**, *2*, 1–189.

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.