

Article

Complete Chloroplast Genome of *Paphiopedilum delenatii* and Phylogenetic Relationships among Orchidaceae

Huyen-Trang Vu ^{1,2,†}, Ngan Tran ^{2,†} , Thanh-Diem Nguyen ¹, Quoc-Luan Vu ³, My-Huyen Bui ¹, Minh-Tri Le ² and Ly Le ^{2,4,*}

¹ Faculty of Biotechnology, Nguyen Tat Thanh University, District 4, Hochiminh City 72820, Vietnam; vthtrang@ntt.edu.vn (H.-T.V.); nthanhdziem1708@gmail.com (T.-D.N.); huyen99xx@gmail.com (M.-H.B.)

² Faculty of Biotechnology, International University-Vietnam National University, Linh Trung Ward, Thu Duc District, Hochiminh City 7000000, Vietnam; kimngan.tran@hdr.qut.edu.au (N.T.); lequangminhtri@gmail.com (M.-T.L.)

³ Tay Nguyen Institute for Scientific Research, Vietnam Academy of Science and Technology, Dalat 670000, Vietnam; vuquocluan07@yahoo.com.vn

⁴ Vingroup Big Data Institute, Hai Ba Trung District, Hanoi 100000, Vietnam

* Correspondence: ly.le@hcmu.edu.vn; Tel.: +84-906578836

† These authors contributed equally to this work.

Received: 23 November 2019; Accepted: 25 December 2019; Published: 2 January 2020



Abstract: *Paphiopedilum delenatii* is a native orchid of Vietnam with highly attractive floral traits. Unfortunately, it is now listed as a critically endangered species with a few hundred individuals remaining in nature. In this study, we performed next-generation sequencing of *P. delenatii* and assembled its complete chloroplast genome. The whole chloroplast genome of *P. delenatii* was 160,955 bp in size, 35.6% of which was GC content, and exhibited typical quadripartite structure of plastid genomes with four distinct regions, including the large and small single-copy regions and a pair of inverted repeat regions. There were, in total, 130 genes annotated in the genome: 77 coding genes, 39 tRNA genes, 8 rRNA genes, and 6 pseudogenes. The loss of *ndh* genes and variation in inverted repeat (IR) boundaries as well as data of simple sequence repeats (SSRs) and divergent hotspots provided useful information for identification applications and phylogenetic studies of *Paphiopedilum* species. Whole chloroplast genomes could be used as an effective super barcode for species identification or for developing other identification markers, which subsequently serves the conservation of *Paphiopedilum* species.

Keywords: *Paphiopedilum*; chloroplast genome; comparative genomics; conservation genomics

1. Introduction

The sequence of chloroplast (cp) genomes can provide information for studying genetic relationships, gene transfer, cloning, and domestication of species [1]. Much research has shown the role of plastomes in practice. The project of chloroplast genome sequencing and barcoding *Fritillaria* [2] helped identify *Fritillaria* species for medical use. The sequencing of the chloroplast genome of the rice variety Nagina-22 [3] enriched genetic resources to support the breeding and crossbreeding of next-generation rice varieties. Na Tian et al. (2018) sequenced and analyzed the chloroplast genome of *Epipremum aureum* that contributed to the propagation and support of gene transfer of this medicinal plant [4]. Shuai Guo et al. (2018) successfully sequenced the chloroplast genome of *Paeonia ostii* which enhanced the productivity of this medicinal herb [5]. In orchids, the sequencing and analysis of cp genomes helped explain phylogenetic relationships and the evolutionary path of Orchidaceae [6,7].

Currently in Vietnam, only *Panax vietnamesis* (Ngoc Linh Ginseng), a valuable endemic Vietnamese ginseng, has had the chloroplast genome sequenced based on four species (two of *Panax vietnamesis*, one of *P. bipinnatifidus*, and one of *P. stipuleanatus*). From then, the study analyzed phylogenetic species and identified four potential indicators for molecular barcode for classification of this target group [8].

Paphiopedilum are the favorite ornamental plants in the world including in Vietnam. Due to its natural and specific beauty, *Paphiopedilum* is hunted and traded in large numbers, leading to danger of extinction. Vietnam is the country with the largest number of *Paphiopedilum* species in the world. However, according to IUCN (International Union for Conservation of Nature) statistics, of nearly 24 species of Vietnamese *Paphiopedilum*, 23 species are on the list of threatened with extinction [9]. *Paphiopedilum delenatii* (Cypripedioideae, Orchidaceae) is usually found at the elevations of 300–750 m and is distributed mostly in the southern region of Vietnam. *P. delenatii* has important ornamental value and is being subjected to overexploitation and habitat destruction. According to the IUCN, *P. delenatii* is now a critically endangered (CR) species with approximately 200 mature individuals left. Controlling over-exploitation and illegal trade is difficult, in which official protectors need to have basic knowledge of identification techniques to distinguish valuable and common species. *Paphiopedilum* species are easily recognized by their specific flower morphology. However, most illegal trades are with immature, non-flowered plants, which leads to species misidentification. The situation of uncontrolled exploitation and smuggling subsequently leads to the destruction of more *Paphiopedilum* species. Timely identification and control helps to limit illegal collection and smuggling, reducing the risk of extinction.

Identifying species using molecular techniques is considered to be the most effective because they give high accuracy results and help to quickly and accurately identify species using a very small number of samples from plant parts, such as roots, stems, leaves. Because of this, an understanding of genomics plays a crucial role. Divergent and conserved regions in the genome provide useful information to establish DNA-based [10–14] as well as PCR-based [13,15–20] identification markers, supporting the protection and management of species. However, in Orchidaceae, most of the studies are of the Epidendroideae subfamily. There were up to 99 complete chloroplast genomes of Epidendroideae species in the RefSeq database (accessed on 8 August 2019). In contrast, there are 3 of Vanilloideae, 3 of Apostasioideae, 9 of Orchidoideae, and 7 of Cypripedioideae. Up to now, 4 cp genomes of *Paphiopedilum* (belong to Cypripedioideae) have been sequenced and analyzed, i.e., *P. tranlienianum* [21], *P. dianthum* [22], *P. armeniacum*, and *P. niveum* [23]. Hence, this study contributes to data resource of chloroplast genomes of Cypripedioideae, in particular, and of Orchidaceae, in general.

For the above reasons, we carried out next-generation sequencing of *P. delenatii* (Figure 1), an endemic species of Vietnam, and assembled its complete chloroplast genome. Our end goal was to extend the genetic resources for the endangered *P. delenatii* and *Paphiopedilum* species, in general.

2. Results and Discussion

2.1. Chloroplast Genome of *Paphiopedilum delenatii*

The whole genome of *P. delenatii* was sequenced from total DNA and resulted in 11.6 million high-quality paired-end reads with high read coverage of 800×. The complete chloroplast genome was then assembled separately following the procedure of our previous study [24]. To verify the plastid genome sequence, we independently compared it to four barcoded markers, i.e., *matK*, *trnL-UAA*, *rpoB*, and *rpoC1* (GenBank accessions MK792631, MK787353, MK876160, and MK792704, respectively), belonging to the chloroplast genome of the same *P. delenatii* sample that was sequenced by the Sanger method in our previous study [25]. The pairwise-alignment results, which showed 100% similarity between single regions of the assembled genome, proved that the whole plastid genome sequence was of *P. delenatii*. This complete plastome sequence was deposited in GenBank under accession MK463585.

The assembled chloroplast genome is 160,955 base pairs (bp) in length (Figure 1). It exhibits a typical quadripartite structure of the large single-copy (LSC, 89,869 bp) and small single-copy (SSC,

2694 bp) regions, separated by a pair of inverted repeat regions (IRs, 34,196 bp each). There are 107 unique genes that were annotated (Table 1), including 68 protein-coding genes, 30 transfer RNA genes, 4 ribosomal RNA genes, and 5 pseudogenes. In particular, there are 23 genes in double copies (Table 1).

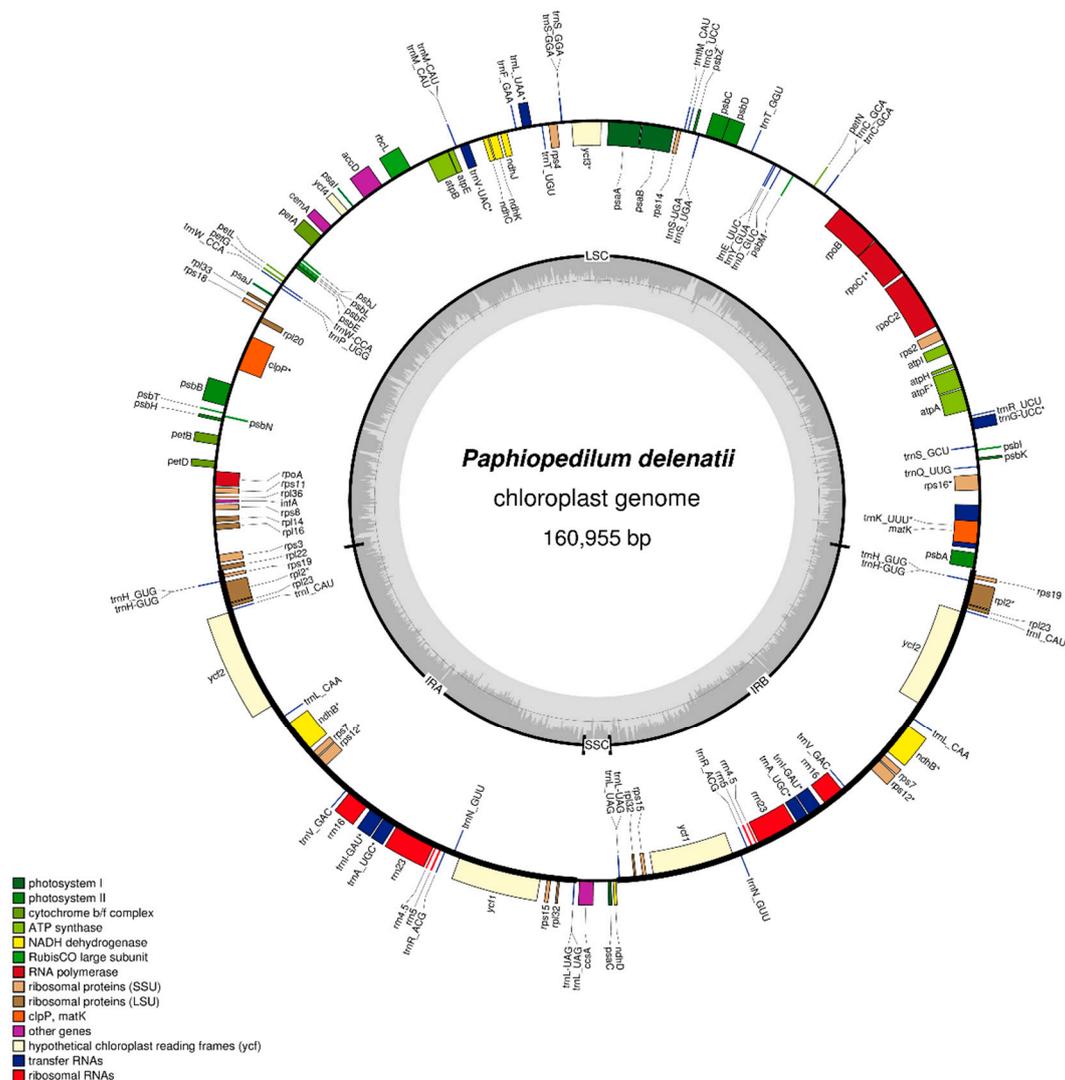


Figure 1. Gene map of the *Paphiopedilum delenatii* chloroplast genome. Annotated genes are colored according to functional categories. Genes lying outside of the outer circle were transcribed clockwise, while those inside the circle were transcribed counterclockwise. The innermost, darker gray corresponds to GC content, while the lighter gray corresponds to AT content. IR, inverted repeat; LSC, large single-copy region; SSC, small single-copy region; SSU, small subunit; LSU, large subunit; RNA, ribonucleic acid; NADH, nicotinamide adenine dinucleotide (NAD) + hydrogen (H); ATP, Adenosine Triphosphate; RubisCO, Ribulose-1,5-bisphosphate carboxylase/oxygenase.

The overall GC content of *P. delenatii* is 35.6%. The GC contents of the LSC, SSC, and IR regions are 33.0%, 28.5%, and 39.3%, respectively (Table 2). The GC content in the IR region was higher than both LSC and SSC regions in all examined *Paphiopedilum* plastomes. This result agreed with previous studies [6,7,26,27]. The presence of four ribosomal RNA (rRNA) genes is considered to be the reason for high GC contents in the IR regions [6,26]. However, another hypothesis proposed that the higher GC content evolution in IR regions does not relate to natural selection but to GC-biased gene conversion (gBGC) [7,27]. Accordingly, GC to AT mutations were unstable while AT to GC ones were fixed after gBGC, and hence the gBGC process prefers repairing DNA mismatches in

recombining DNA over evolutionary time [28]. Meanwhile, IRs are considered to be recombination hotspots due to their identical inverted repeat structures, which increase the frequency of intraplasmidic recombination [29,30] and hence increase gBGC.

Table 1. List of annotated genes in the *Paphiopedilum delenatii* chloroplast genome.

| Classification of Genes | | Name of Genes | Number |
|--|---|---|--------|
| RNA genes | Ribosomal RNAs | <i>rrn4.5</i> (× 2), <i>rrn5</i> (× 2), <i>rrn16</i> (× 2), <i>rrn23</i> (× 2) | 8 |
| | Transfer RNAs | <i>trnA</i> _UGC(× 2), <i>trnC</i> _GCA, <i>trnD</i> _GUC, <i>trnE</i> _UUC, <i>trnF</i> _GAA, <i>trnM</i> _CAU, <i>trnG</i> _GCC, <i>trnG</i> _UCC, <i>trnH</i> _GUG(× 2), <i>trnI</i> _CAU(× 2), <i>trnI</i> _GAU(× 2), <i>trnK</i> _UUU, <i>trnL</i> _CAA(× 2), <i>trnL</i> _UAA, <i>trnL</i> _UAG(× 2), <i>trnM</i> _CAU, <i>trnN</i> _GUU(× 2), <i>trnP</i> _UGG, <i>trnQ</i> _UUG, <i>trnR</i> _ACG(× 2), <i>trnR</i> _UCU, <i>trnS</i> _GCU, <i>trnS</i> _GGA, <i>trnS</i> _UGA, <i>trnT</i> _GGU, <i>trnT</i> _UGU, <i>trnV</i> _GAC(× 2), <i>trnV</i> _UAC, <i>trnW</i> _CCA, <i>trnY</i> _GUA | 39 |
| Protein-coding genes | Photosystem I | <i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i> | 5 |
| | Photosystem II | <i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbH</i> , <i>psbI</i> , <i>psbJ</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbN</i> , <i>psbT</i> , <i>psbZ</i> | 15 |
| | Cytochrome | <i>petA</i> , <i>petB</i> , <i>petD</i> , <i>petG</i> , <i>petL</i> , <i>petN</i> | 6 |
| | ATP synthase | <i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> , <i>atpH</i> , <i>atpI</i> | 6 |
| | Rubisco | <i>rbcL</i> | 1 |
| | Ribosomal proteins-small units | <i>rps11</i> , <i>rps12</i> (× 2), <i>rps14</i> , <i>rps15</i> (× 2), <i>rps16</i> , <i>rps18</i> , <i>rps19</i> (× 2), <i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> (× 2), <i>rps8</i> | 16 |
| | Ribosomal proteins-large units | <i>rpl14</i> , <i>rpl16</i> , <i>rpl2</i> (× 2), <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> (× 2), <i>rpl32</i> (× 2), <i>rpl33</i> , <i>rpl36</i> | 12 |
| | RNA polymerase | <i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> , <i>rpoC2</i> | 4 |
| | Miscellaneous | <i>accD</i> , <i>ccsA</i> , <i>cemA</i> , <i>clpP</i> , <i>infA</i> , <i>matK</i> | 6 |
| Hypothetical chloroplast reading frames (<i>ycf</i>) | <i>ycf1</i> (× 2), <i>ycf2</i> (× 2), <i>ycf3</i> , <i>ycf4</i> | 6 | |
| Pseudogenes | NADH dehydrogenase | <i>ndhB</i> (× 2), <i>ndhC</i> , <i>ndhD</i> , <i>ndhJ</i> , <i>ndhK</i> | 6 |
| Total | | | 130 |

(× 2) refers to genes in double copies.

Table 2. Basic features of four *Paphiopedilum* chloroplast genomes.

| Species | <i>Paphiopedilum delenatii</i> | <i>Paphiopedilum armeniacum</i> | <i>Paphiopedilum niveum</i> | <i>Paphiopedilum dianthum</i> |
|------------------------------|--------------------------------|---------------------------------|-----------------------------|-------------------------------|
| Total length (bp) | 160,955 | 162,682 | 159,108 | 154,699 |
| IR length (bp) | 34,196 | 33,641 | 31,978 | 32,711 |
| LSC length (bp) | 89,869 | 91,734 | 89,958 | 86,861 |
| SSC length (bp) | 2694 | 3666 | 5194 | 2416 |
| Total gene number | 130 | 129 | 126 | 130 |
| Coding sequence (CDS) number | 77 | 77 | 74 | 79 |
| rRNA number | 8 | 8 | 8 | 8 |
| tRNA number | 39 | 38 | 38 | 38 |
| Pseudogene number | 6 | 6 | 6 | 5 |
| Overall GC content (%) | 35.6 | 35.4 | 35.0 | 35.0 |
| GC content of IR (%) | 39.3 | 39.0 | 40.0 | 39.0 |
| GC content of LSC (%) | 33 | 32.6 | 32.0 | 33.0 |
| GC content of SSC (%) | 28.5 | 31.0 | 29.0 | 29.0 |
| GenBank accession | MK463585 | NC_026779.1 | NC_026776.1 | NC_036958.1 |

GC content has been reported to be different not only between different regions of a genome, but also between different genomes of different species [31,32]. Overall GC content of some orchid plastomes were recorded, e.g., *Dendrobium moniliforme* (37.54%), *Goodyera schlechtendaliana* (37.07%), *Vanilla aphylla* (35.02%) [33], *Cremastra appendiculata* (37.2%), *Calanthe davidii* (36.9%), *Epipactis mairei* (37.2%), and *Platanthera japonica* (37%) [6], with their plastome lengths ranging from 148,778 to 162,835 bp. GC content of the *Paphiopedilum* species in our study range from 35% to 35.6%, with genome lengths of 154.699 to 162.682 bp, similar to other orchids. Hence, *Paphiopedilum* and *Vanilla* contain lower a GC percent, which has been suggested to mean that the sequence variability was higher toward the enrichment of AT nucleotides [33,34]. The difference in GC content of the nuclear genome has been proposed as a useful value for identification of species [31]. Karimi et al. (2018) introduced the GCSpeciesSorter tool for accurately and quickly determining GC content and, hence, classifying species in a mix of DNA relationships for metagenomic studies [35]. Hence more studies on the GC content of plant plastomes might provide a useful measurement for the identification of species.

The comparison of four *Paphiopedilum* plastomes showed that *P. delenatii* inherits a similar conserved plastome structure to its *Paphiopedilum* sisters (Table 2). Some differences are that *P. niveum* has the least total gene number and coding sequence (CDS) number and has no protein-coding gene *infA*. *P. dianthum* contains four genes, *ycf68*($\times 2$) and *orf42*($\times 2$), that are not observed in the other three plastomes. *P. delenatii* is distinguished by double copies of *trnL_UAG* and *rpl32* (Table S1).

In *P. delenatii*, there are six *ndh* genes, all of which are pseudogenes: *ndhB*($\times 2$), *ndhC*, *ndhD*, *ndhJ*, and *ndhK*. *ndhA*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, and *ndhI* are entirely absent from the *P. delenatii* cp genomes. The same applies to *P. armeniacum* and *P. niveum* chloroplast genomes. As for *P. dianthum*, the RefSeq annotation showed two pseudogenes: *ndhB* and *ndhD*. This result was consistent with the report of Guo et al. (2012) [36]. However, we aligned the *ndhC*, *ndhJ*, and *ndhK* sequences of *P. delenatii* to the three chloroplast genomes of *Paphiopedilum* and found a version of *ndhJ* present in the *P. dianthum* cp genome (Figure S1).

Ndh genes code for the enzyme NADH dehydrogenase [37,38] that is responsible for electron transport of chloroplasts. The loss and variation of *ndh* genes in IR boundaries in orchid species have been the focus of multiple investigations [23,39,40]. Here, we compared the boundaries between two inverted repeat regions, i.e. IRa and IRb, and LSC and SSC regions of *P. delenatii* to those of 14 other orchid species from 5 subfamilies of Orchidaceae (Figure 2). All Cyprripedioideae species with high-resolution plastid genome annotations were included. *P. delenatii* exhibited a highly-conserved pattern of IR boundaries to those of *P. dianthum*, *P. armeniacum*, as well as *Vanilla pompona* (Vanilloideae): *rpl22* at the IRb/LSC border, 2 copies of the *ycf1* gene in IRs, and a pseudogenized *rpl22* next to the border of IRa/LSC. These 4 species possessed rather small SSCs (2037–3666 bp) compared to the remaining 11 species (13,066–21,921 bp). Interestingly, in the other four Cyprripedioideae species, which were *Phragmipedium longifolium*, *Cypripedium formosanum*, *Cypripedium japonicum*, and *Cypripedium macranthos*, *ycf1* genes shifted to the borders of IRs and SSC, leaving one functional *ycf1* at the SSC/IRa junction and one pseudogene in IRb, next to the IRb/SSC. The three *Cypripedium* species had the *ndhF* gene in the SSC region, which was absent in all *Paphiopedilum*. These Cyprripedioideae had very similar features in IR junctions to those of Orchidoideae and Apostasioideae. In Epidendroideae, IR boundaries of the 4 examined species exhibited higher heterogeneity: *ndhF* is present in *Calanthe davidii* and *Neottia ovata*, but not in *Eulophia zollingeri* and *Cattleya crispata*; *ycf1* is mostly in the SSC region but at the IRb/SSC border in *Neottia ovata*. In the case of *P. niveum*, the IR regions, which contain six palindromic pairs instead of one (Table S2), were not determined. Hence, this species was not included in Figure 2 and its plastome structure should be interpreted further.

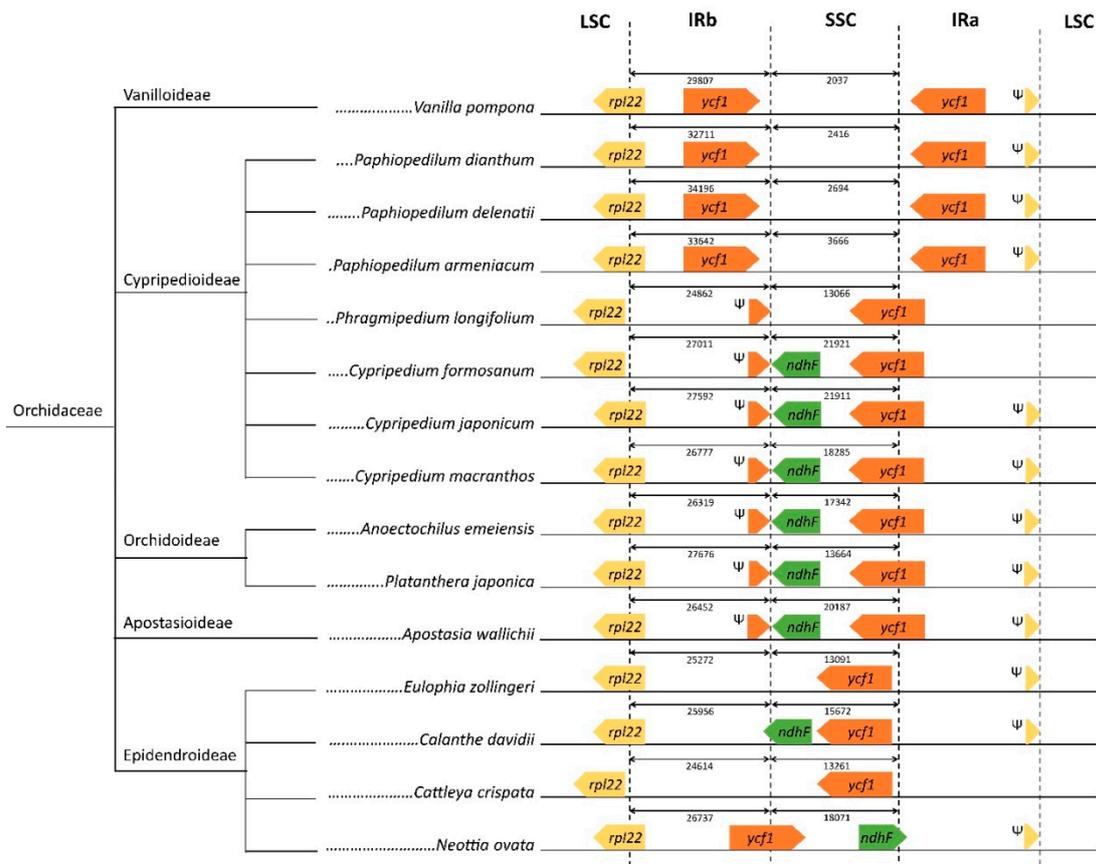


Figure 2. Schematic representation of the borders between inverted repeats (IRs) and large single-copy (LSC) and small single-copy (SSC) of *P. delenatii* and other 15 species from all five subfamilies of Orchidaceae. Ψ indicates a pseudogene.

The IR region is highly conserved and stable in the chloroplast genome. The expansion and contraction of the IR region is a common characteristic of the chloroplast genome. Luo et al. (2014) proposed four types of IR/SSC junction when they examined seven orchids [41]. In our study, the IR pattern of *Cypripedium* and Orchidoideae and Apostasioideae species matched with type I; *Phragmipedium longifolium* matched with type III; *Eulophia zollingeri* and *Cattleya crispata* matched with type IV. None of our examined species matched with Type II, which was the overlap of Ψ*ycf1* and *ndhF* genes [41]. Instead, *Vanilla pompona*, *Calanthe davidii*, and *Neottia ovata* expressed entirely new IR patterns (Figure 2). Together with the shift of the *ycf1* gene, the presence or absence of the *ndhF* gene was one of the factors observed in different IR/SSC patterns. According to Guo et al. [36], among the Cypripedioideae, *Cypripedium* species inherited the *ndhF* gene from their ancestors; this gene has been lost from the other genera. *ndhF* gene loss was previously proposed to be correlated with the instability of IR/SSC boundary [40]. More studies are required to better understand IR evolution. However, this study emphasized the diversity of IR/SSC boundaries in orchids.

2.2. Repeat and Microsatellite Analysis

For *Paphiopedilum delenatii*, we found 645 repeats with lengths from 30 to 58 bp (Table S3). The number of forward repeats were the most common (176/645), followed by palindromic repeats (168/645), reverse repeats (167/645), and complement repeats (133/645). Most repeats were inside intergenic spacers. Notably, 23 repeat sequences were entirely located in the *ycf2* gene. There were 87 simple sequence repeats (SSRs) or microsatellite sequences identified in the *P. delenatii* chloroplast

genome (Table S4). Fifty SSR loci were in intergenic regions, while 21 were in the coding areas (12 of which were in *ycf1* gene). Most SSR loci were mononucleotide repeats with AT motifs (57/87).

Repeats are units of DNA that are similar in the genome. There are short repeats and longer repeats. SSR (simple sequence repeat), or microsatellite, is a type of low-complexity, short repeat with 1–6 nucleotides. Generally, microsatellite SSRs are widely distributed throughout the genome and have a great effect on recombination and rearrangement of the genome [42,43]. In our study, there were no tetranucleotide or longer repeats in the cp genome. This result was consistent with previous reports that most SSRs include mono- and di-nucleotide repeats while tri-, tetra-, penta-, and hexa-nucleotide repeat sequences were detected at much lower frequencies [6]. The longer repeat, known as a minisatellite, contains 10–100 nucleotides. In term of direction and complementary, these repeats are divided into four types: forward (direct) repeats, reverse repeats (also known as inverted repeat –IR), complement repeats, and palindromic repeats (reverse complement repeat) [44]. Both microsatellite and minisatellite repetitive sequences play significant roles in species identification. SSRs were used as DNA barcodes to clear identify 5 genotypes of *Solanum melongena* L. by Chinnappareddy et al. [45]. In Orchidaceae, SSR markers were developed and utilized as identification tools in various studies [18,46–48] due to their high reproducibility and variability [45,49,50]. In particular, SSRs in the chloroplast genome were reported to have a high level of polymorphism among species and loci [51]. Sets of cp SSRs were also isolated and developed for recognizing valuable plants, serving conservation genetics, investigation of chloroplast genetic structure, adaptive evolution, and population [52–57]. Even single loci, i.e., *trnL* and *trnL-F*, in chloroplast were used to develop specific primers for amplifying SSR sequences for a population genetic study [58]. The longer repeat minisatellites were supposed to have specific mechanisms of evolution and function, forming common, as well as unique, repeat patterns [44,59], hence they are important tools for taxonomic and phylogenetic studies. Eight out of 13 species in the *Phoenix* genus (Arecaceae) were unambiguously distinguished using minisatellites developed from a 700 bp region in the chloroplast spacer *trnG* (GCC)-*trnF*M (CAU) [60]. Our study provided primary data of SSRs and minisatellite repeats for further research on identifying applications.

2.3. Phylogenetic and Species Resolution Analyses

We included chloroplast genomes of *P. niveum* and all the Orchidaceae species examined in the above IR boundaries section to understand the phylogenetic relationship between *P. delenatii* and other species across subfamilies (Figure 3). *Artemisia argyi* was set as the outgroup. Firstly, a whole-plastome tree was constructed. As expected, *P. delenatii* along with other *Paphiopedilum* again showed a high similarity and close phylogenetic relationship to each other. The 6 species of Cyripedioideae were clustered together into one clade, excluding *C. macranthos* and *C. formosanum*. Although the IR boundary analysis showed similar patterns between *V. pompona* and those of *Paphiopedilum* accessions, in term of the whole plastome, *V. pompona* was more similar to Epidendroideae and Orchidoideae. Next, we constructed phylogenetic trees based on two popular short barcodes: *matK* and *rbcL*. All three phylogenetic trees presented a full separation of 16 species. This might be because the analysis consisted of distinct species. In contrast to the plastome tree, *Apostasia wallichii* was now segregated in an independent monophyletic branch, as expected, since it belongs to a different subfamily. Furthermore, *C. formosanum* was grouped with the members of its genus. A dot plot similar analysis on the whole-genome alignment of *P. delenatii* to other species (Figure S2) found that *C. formosanum* and *A. wallichii* both had an inverse similarity fragments located in the LSC region, at the position of 10,000–80,000 bp. This led to greater genetic similarities between *C. formosanum* and *A. wallichii*, which grouped them into one clade in the whole-plastome tree. *C. macranthos* was closer to its sisters of Cyripedioideae in terms of nucleotide polymorphism (Figures S3 and S4), but differed in terms of amino acid variations, which explains the separation of this species in the phylogenetic trees.

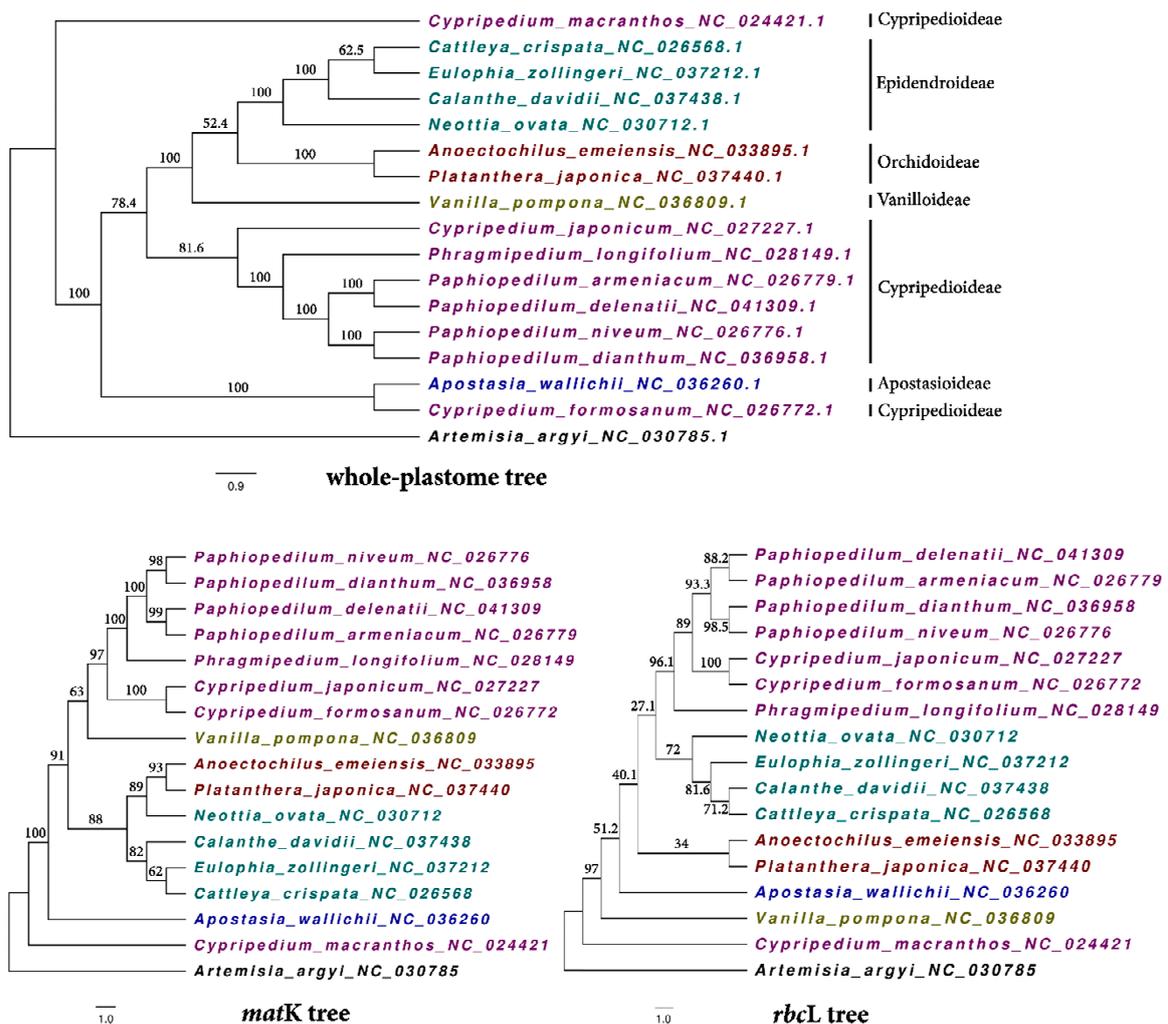


Figure 3. Cladogram of the phylogenetic relationship between *P. delenatii* and other Orchidaceae species based on maximum likelihood analysis. The number on the branches are bootstrap values.

The chloroplast genome was proposed to be used as meta-barcode by some previous studies [61–64]. Hence, two universal mini-barcodes, *matK* and *rbcL*, were established in order to compare the effects of species resolution under the tree-based method. From the above analyses, using the whole genome for phylogenetic relationship analysis might not be practical due to the unexpected problem of inversion. However, for barcoding, the species separation, but not relationship, is of first concern. In this respect, using the whole plastome as a super-barcode is effective. Besides, we also analyzed genetic distance matrices among species using plastome, *matK* and *rbcL*, data independently (Table S5). The average, minimum, and maximum of distances were calculated. Except for *rbcL* with a low average value (0.034), the average genetic distance by *matK* (0.122) and by plastome (0.112) were all high. Although the average values were not much different between plastomic and *matK* data, the minimum value of *matK* (0.005) was much lower than that of the plastome (0.012). While some previous studies showed that *matK* was not able to identify all species in some close taxonomic groups [65,66], our result suggests that the entire cp genome could do better than *matK* in identifying closely related species.

2.4. Divergence of Hotspot Regions

We calculated the nucleotide variability in the chloroplast genomes of the four analyzed species: *Paphiopedilum delenatii*, *Paphiopedilum armeniacum*, *Paphiopedilum niveum*, and *Paphiopedilum dianthum*. The *P. delenatii* plastome was highly conserved compared to other cp genomes of *Paphiopedilum* species.

The number of single nucleotide polymorphism (SNP) was 24,211 out of 170,423 bp of the alignment. The values of nucleotide diversity (π) ranged from 0 to 0.34889 and the diversity threshold was 0.079. However, the diversity threshold decreased at 0.0377 when the SSC regions were excluded from the analysis. At this threshold, 11 highly-variable regions were suggested as potential markers in species identification and phylogeny study of the *Paphiopedilum* genus (Figure 4). These 11 highly-variable regions included 1 protein-coding gene (*clpP*) and 8 intergenic spacers (*matK-rps16*, *trnR_UCU-atpA*, *psbM-trnD_GUC*, *trnE_UUC-trnT_GGU*, *accD-psaI*, *psbE-petL*, *trnP_UGG-psaJ*, and *rpl23-trnI_CAU*) in the LSC region; 1 intergenic spacer (*ycf1-rps15*) in the repeat region IRb; and 1 intergenic spacer (*ccsA-psaC*) in the SSC region (Table S6).

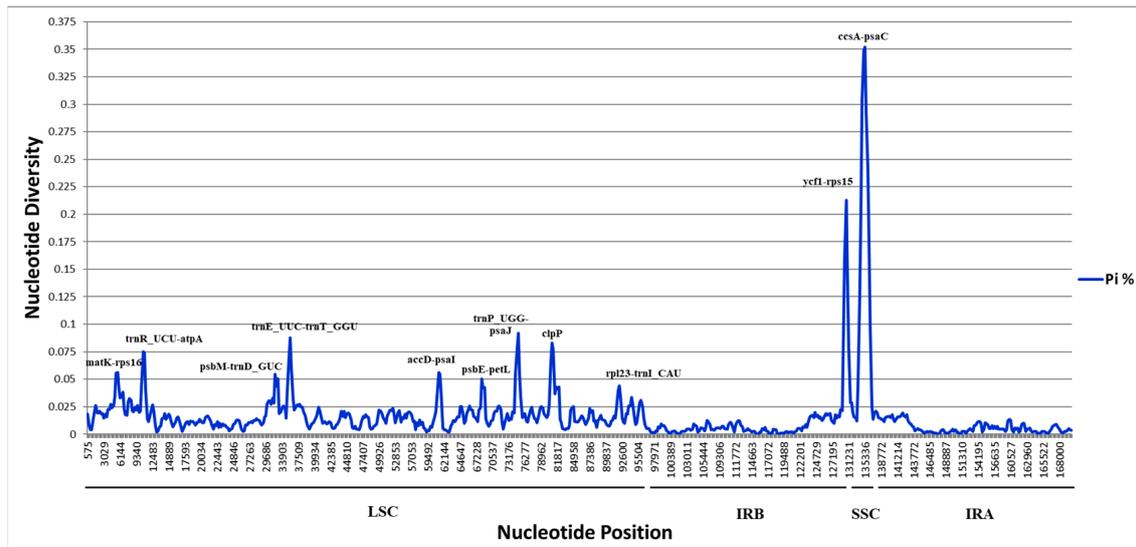


Figure 4. Sliding window of nucleotide diversity from the alignment of four *Paphiopedilum* plastomes. (π) Parsimony rate; (IR) inverted repeat regions; (LSC) large single-copy region; (SSC) small single-copy region; red line: diversity threshold.

In the study of Bi et al. (2018), nucleotide diversity of eight *Fritillaria* species ranged from 0 to 0.02583 [67]. The π value was from 0 to 0.05872 for the nine cp genomes of *Eragrostis* species in the study of Somaratne et al. (2019) [68]. In comparison with those studies, the π value (0–0.34889) in our study was much higher. The reason was that there were two fragments with extremely high divergence within the SSC region that were not present in the mentioned studies. Therefore, we tried another analysis in which the SSC region was removed. The threshold was significantly decreased, at 0.0377 instead of 0.079, for the whole cp genomes. The high divergence of SSC was also reported recently by Cui et al. (2019). A comparison of 10 ginger species showed an average value of nucleotide variability of 0.0187, while it was only 0.0075 when comparing 4 species of the same family [69]. Therefore, SSC seemed to rapidly evolve compared to LSC. The two inverted regions, IRa and IRb, were quite conserved, with low nucleotide diversity. In addition, the narrow endemism is also a prominent feature of *Paphiopedilum* species. Of the known species, 72% are narrowly endemic with very limited distribution [70]. The local distribution and ecological separation might be the reason for considerable genetic differences between species in the same genus and might clarify the high nucleotide diversity of *Paphiopedilum* species in our study.

Using short sequences for the identification of a species is still the current universal method due to its simple and time-saving traits. Indeed, the average divergence of *matK* (0.122) was even higher than that of the plastome (0.112) in our study. The reason for this is that the plastome contains both diverse and conserved regions while *matK* is a selected, high-variable locus [71]. This result proved the species separation capability of this mini-barcode in comparison with the super one. Common markers, such as the internal transcribed spacer (ITS) in the nucleus and cp loci *rbcL*, *matK*, *ycf1*, *trnL*, *trnL-F*,

atpF-atpH, *trnH-psbA*, etc., were the most used. However, as no single locus can resolve the entire plant species [72,73], new variable sites are still being sought [6,10,12,74,75]. In the chloroplast genome, SNPs were not random but clustered as “hotspots” [76], which were defined as highly-variable loci [77]. A series of new findings for species-specific barcodes was developed recently. Protein-coding gene *clpP* and intergenic spacer *rps15-ycf1* from our proposal were also recommended previously [76,78–80]. The hot region of *trnE-trnT* was matched with the study of Zhao et al. (2018) [77]. Although *rps16* and its intergenic spacer with *trnQ* were suggested in much research [10,11,26,74,81], our study introduced a new intergenic spacer of this gene, i.e., *rps16-matK*. However, searching for new variable sites as candidate barcodes is just the first step. Primer designing and amplification success are also required for barcoding effects [10,25,74].

Besides barcoding applications, hot region information is also used for developing other PCR-based identification techniques. rDNA-ITS (internal transcribed spacer) sequences were used to design species-specific SCAR (sequence characterized amplified regions) markers, in which 3 primer pairs—SCAR-600armF/Pap-ITS2R, SCAR-300delF/Pap-ITS2R, and SCAR-700micF/Pap-ITS2R—were effectively used to amplify and recognize three *Paphiopedilum* species (*P. armeniacum*, *P. micranthum*, and *P. delenatii*) and their hybrids [17]. Similarly, two DNA sequences, *rpoC2* and *atpF-atpH*, were reported to contain species-specific SNPs, insertions, and deletions. This information was utilized to develop 3 species-specific primers for quickly identifying species: *Cypripedium guttatum* var. *koreanum*, *C. japonicum*, and *C. formosanum* of the Korean *Cypripedium* genus. This ARMS (amplification refractory mutation system) method was also based on electrophoresis technique without a sequencing step [15]. In another study (2014), the divergent nucleotide sequence of the ITS region and 3 cpDNA fragments were amplified and subsequently cut with several restriction enzymes to create species-specific types of DNA patterns. The PCR-restriction fragment length polymorphism (PCR-RFLP) approach was successful in the identification of 25 native *Dendrobium* species in Thailand [16].

3. Materials and Methods

3.1. Plant Material, DNA Extraction, and Sequencing

The sample of *Paphiopedilum delenatii* plant (Figure 5) was identified by the shape of the flowering plant and stored at $-80\text{ }^{\circ}\text{C}$ at Tay Nguyen Biological Institute, Vietnam. First, 0.2 g of leaf was ground with 5 μL proteinase K, 3mL of a mixture of beta-mer and extract buffer at $65\text{ }^{\circ}\text{C}$, then incubated for 30 min at $65\text{ }^{\circ}\text{C}$. The sample was had 600 μL P:C:I added and was centrifuged for 10 min at 10,000 rpm. After adding 5 μL of RNase and incubating at $37\text{ }^{\circ}\text{C}$, the sample had 600 μL C:I added. DNA was precipitated by isopropanol and incubated overnight at $-20\text{ }^{\circ}\text{C}$. The pellet, obtained by centrifugation, was washed with 70%, 80%, and 90% ethanol. DNA was suspended in 25 μL TE and stored at $-20\text{ }^{\circ}\text{C}$. The library construction and whole-genome sequencing of *P. delenatii* was performed by GENEWIZ (South Plainfield, NJ, USA). Sequencing was carried out on an Illumina HiSeq (Genewiz, South Plainfield, USA) using a 2×150 paired-end (PE) configuration.



Figure 5. The flower of endemic species *Paphiopedilum delenatii*.

3.2. Read Data Processing and Chloroplast Genome Assembly

Demultiplexing was performed by bcl2fastq 2.17. Raw data was filtered as follows: (1) discard pair-end reads with adapter, (2) discard pair-end reads when the content of N bases is more than 10% in either read, and (3) discard pair-end reads when the ration of bases of low quality ($Q < 20$) is more than 0.5 in either read. The chloroplast genome of *P. delenatii* was reconstructed using NOVOPlasty 2.7.2 [82], with the complete chloroplast genome of *P. armeniacum* (RefSeq: NC_026779.1) as the reference genome and the *rbcl* from the same plastid genome of *P. armeniacum* as the seed sequence. The annotation was done by GeSeq [83] and further manually curated by comparison to the annotations of *P. armeniacum*, *P. dianthum*, and *P. niveum* in GenBank. The genome map was drawn by OGDRAW [84].

3.3. Repeat Sequence and Microsatellite Identification

REPuter [85] was used to calculate DNA repeats, including forward, reverse, complement, and palindromic kinds of repeat sequences. The repeats were identified with a hamming distance of 3 and minimum repeat size of 30 [6]. MISA [86] was used to identify microsatellite sequences with default parameters.

3.4. Examination of IR Junctions

We manually examined the IR junctions of all included orchid species. Annotations of IRs, SSC, LSC, and genes were based on their respective annotations in the RefSeq database. For genomes without IR annotations, we used REPuter to identify their pairs of inverted repeats.

3.5. Phylogenetic Analysis

The phylogenetic analysis was based on the complete genome sequences of 16 orchid species under the maximum likelihood criterion and the GTR + I + G nucleotide substitution model using R package phangorn [87]. Node was calculated from 1000 bootstrap replicates. Figtree [88] was used to visualize the resulting tree. The multiple alignment data of 16 plastomes was used to calculate variable sites and genetic distance matrices using MEGA [89]. MAFFT [90] was used to pairwise align and construct dot-plot graph of these 16 plastomes.

3.6. Nucleotide Variability Calculation

DnaSP v6.1 was used to extract the parsimony variable site density over the plastid genome alignment of four analyzed *Paphiopedilum* species with a sliding window (window length ≤ 600 and step size = 200). Nucleotide diversity was calculated by the ratio of Pi and window length. The diversity threshold was 0.079, calculated by the sum of the average and double the standard deviation [67]. Regions with diversity higher than the threshold were recommended as highly variable regions.

4. Conclusions

In this study, we aimed to expand the genetic resource of the endangered species *P. delenatii* by next-generation sequencing and chloroplast genome assembly. *P. delenatii* chloroplast genome exhibited a quadripartite structure of 160,955 bp and a total of 130 genes, which were highly conserved compared to other *Paphiopedilum* species. All six *ndh* genes in the *P. delenatii* cp genome were pseudogenes. The presence or absence of the *ndhF* gene and the shift of *ycf1* and *rpl22* genes on the boundaries between IRs and LSC and SSC regions resulted in different IR/SSC patterns that can be useful in inferring species relationships. A reference of 87 SSRs/minisatellite repeats in *P. delenatii* was proposed for further research on identification applications. Eight highly-variable regions were suggested as the potential markers in barcoding and phylogeny studies of *Paphiopedilum* genus. Hence, the sequence data of *P. delenatii* complete chloroplast genome could be used directly in the identification of *Paphiopedilum* species or for the development of other identification markers, such as SSR, barcoding, or species-specific PCR-based techniques. Although it is still costly and time-consuming compared to short DNA sequencing, genome sequencing costs have decreased significantly in recent years. Along with the development of other whole genome sequencing without amplification steps, e.g., Nanopore technique, we might hope that sequencing of the whole genome would be easier and convenient in the future.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2223-7747/9/1/61/s1>, Figure S1: Alignment of *ndhC*, *ndhJ*, and *ndhK* genes on 4 *Paphiopedilum* using SimpleSynteny; Figure S2: Pairwise dot plot cladograms of *P. delenatii* and other Orchids; Figure S3: Nucleotide polymorphism of Orchidaceae species based on *rbcL* region; Figure S4: Amino acid polymorphism and number of variable sites between *C. macranthos* and other Orchidaceae species using *rbcL* sequence; Table S1: List of annotated genes in *Paphiopedilum* plastomes; Table S2: Positions of some palindromic repeat pairs in *P. niveum*; Table S3: Full list of DNA repeats in *P. delenatii* cp genome; Table S4: Full list of simple sequence repeats (SSRs) in *P. delenatii* cp genome; Table S5: Genetic distance matrix of analysis data; Table S6. Variable sites from plastomes of 4 *Paphiopedilum* species.

Author Contributions: Conceptualization, L.L.; Data curation, H.-T.V. and M.-H.B.; Formal analysis, H.-T.V., N.T., T.-D.N., and M.-T.L.; Funding acquisition, L.L.; Investigation, H.-T.V., N.T., Q.-L.V., T.-D.N., and M.-H.B.; Methodology, N.T. and M.-T.L.; Project administration, H.-T.V. and L.L.; Resources, Q.-L.V.; Supervision, L.L.; Writing—original draft, H.-T.V., N.T., and T.-D.N.; Writing—review & editing, L.L., H.-T.V., and N.T. All authors have read and agreed to the published version of the manuscript.

Funding: The study was supported by Air Force Office of Scientific Research and Asian Office of Aerospace Research and Development (grant number FA23861514119). All authors declared no other competing financial interests.

Acknowledgments: This study was supported by the Science and Technology Development Fund of Nguyen Tat Thanh University. We thank to Tay Nguyen Institute for Scientific Research for providing *Paphiopedilum* samples for this study. We are also very grateful to the Computational Biology Center of International University, Vietnam National University for providing the computer resources.

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

Abbreviation

| | |
|------|--|
| IUCN | International Union for Conservation of Nature |
| Cp | chloroplast |
| CR | Critically endangered |
| IR | Inverted repeat region |
| SSC | Small single-copy region |
| LSC | Large single-copy region |

| | |
|------|--------------------------------|
| gBGC | GC-biased gene conversion |
| SSR | Simple sequence repeat |
| SNP | Single nucleotide polymorphism |

References

- Daniell, H.; Lin, C.S.; Yu, M.; Chang, W.J. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biol.* **2016**, *17*, 134. [[CrossRef](#)] [[PubMed](#)]
- Xiang, L.; Su, Y.; Li, X.; Xue, G.; Wang, Q.; Shi, J.; Wang, L.; Chen, S. Identification of Fritillariae bulbosus from adulterants using ITS2 regions. *Plant Gene* **2016**, *7*, 42–49. [[CrossRef](#)]
- Yeisoo, Y.; Hyun Oh, L.; Joong Hyoun, C.; Han Yong, P.; Soo-Cheul, Y. The complete chloroplast genome sequence of *Oryza sativa* aus-type variety Nagina-22 (Poaceae). *Mitochondrial DNA Part B* **2017**, *2*, 819–820.
- Tian, N.; Han, L.; Chen, C.; Wang, Z. The complete chloroplast genome sequence of *Epipremnum aureum* and its comparative analysis among eight Araceae species. *PLoS ONE* **2018**, *13*, e0192956. [[CrossRef](#)] [[PubMed](#)]
- Guo, S.; Guo, L.; Zhao, W.; Xu, J.; Li, Y.; Zhang, X.; Shen, X.; Wu, M.; Hou, X. Complete chloroplast genome sequence and phylogenetic analysis of *Paeonia ostii*. *Molecules* **2018**, *23*, 246. [[CrossRef](#)] [[PubMed](#)]
- Dong, W.-L.; Wang, R.-N.; Zhang, N.-Y.; Fan, W.-B.; Fang, M.-F.; Li, Z.-H. Molecular Evolution of Chloroplast Genomes of Orchid Species: Insights into Phylogenetic Relationship and Adaptive Evolution. *Int. J. Mol. Sci.* **2018**, *19*, 716. [[CrossRef](#)]
- Niu, Z.; Xue, Q.; Wang, H.; Xie, X.; Zhu, S.; Liu, W.; Ding, X. Mutational Biases and GC-Biased Gene Conversion Affect GC Content in the Plastomes of *Dendrobium* Genus. *Int. J. Mol. Sci.* **2017**, *18*, 2307.
- Manzanilla, V.; Kool, A.; Nguyen Nhat, L.; Nong Van, H.; Le Thi Thu, H.; de Boer, H.J. Phylogenomics and barcoding of *Panax*: Toward the identification of ginseng species. *BMC Evol. Biol.* **2018**, *18*, 44. [[CrossRef](#)]
- International Union for Conservation of Nature. *The IUCN Red List of Threatened Species*; Version 2019-2; International Union for Conservation of Nature: Gland, Switzerland, 2019.
- Zhou, Y.; Nie, J.; Xiao, L.; Hu, Z.; Wang, B. Comparative Chloroplast Genome Analysis of Rhubarb Botanical Origins and the Development of Specific Identification Markers. *Molecules* **2018**, *23*, 2811. [[CrossRef](#)]
- Li, X.; Tan, W.; Sun, J.; Du, J.; Zheng, C.; Tian, X.; Zheng, M.; Xiang, B.; Wang, Y. Comparison of Four Complete Chloroplast Genomes of Medicinal and Ornamental *Meconopsis* Species: Genome Organization and Species Discrimination. *Sci. Rep.* **2019**, *9*, 10567.
- Park, I.; Yang, S.; Kim, W.J.; Song, J.-H.; Lee, H.-S.; Lee, H.O.; Lee, J.-H.; Ahn, S.-N.; Moon, B.C. Sequencing and Comparative Analysis of the Chloroplast Genome of *Angelica polymorpha* and the Development of a Novel Indel Marker for Species Identification. *Molecules* **2019**, *24*, 1038. [[CrossRef](#)] [[PubMed](#)]
- Zhou, T.; Wang, J.; Jia, Y.; Li, W.; Xu, F.; Wang, X. Comparative Chloroplast Genome Analyses of Species in *Gentiana* section *Cruciata* (Gentianaceae) and the Development of Authentication Markers. *Int. J. Mol. Sci.* **2018**, *19*, 1962. [[CrossRef](#)] [[PubMed](#)]
- Jheng, C.F.; Chen, T.C.; Lin, J.Y.; Chen, T.C.; Wu, W.L.; Chang, C.C. The comparative chloroplast genomic analysis of photosynthetic orchids and developing DNA markers to distinguish *Phalaenopsis* orchids. *Plant Sci.* **2012**, *190*, 62–73. [[CrossRef](#)] [[PubMed](#)]
- Kim, J.S.; Kim, H.T.; Son, S.-W.; Kim, J.-H. Molecular identification of endangered Korean lady's slipper orchids (*Cypripedium*, Orchidaceae) and related taxa. *Botany* **2015**, *93*, 603–610. [[CrossRef](#)]
- Peyachoknagul, S.; Mongkolsiriwatana, C.; Wannapinpong, S.; Huehne, P.S.; Srikulnath, K. Identification of native *Dendrobium* species in Thailand by PCR-RFLP of rDNA-ITS and chloroplast DNA. *Sci. Asia* **2014**, *40*, 113–120. [[CrossRef](#)]
- Sun, Y.W.; Liao, Y.J.; Hung, Y.S.; Chang, J.C.; Sung, J.M. Development of ITS sequence based SCAR markers for discrimination of *Paphiopedilum armeniacum*, *Paphiopedilum micranthum*, *Paphiopedilum delenatii* and their hybrids. *Sci. Hortic.* **2011**, *127*, 405–410. [[CrossRef](#)]
- Fattmah, D.; Sukma, D. Development of Sequence-Based Microsatellite Marker for *Phalaenopsis* Orchid. *HAYATI J. Biosci.* **2011**, *18*, 71–76. [[CrossRef](#)]
- Lin, J.-Y.; Lin, B.-Y.; Chang, C.-D.; Liao, S.-C.; Liu, Y.-C.; Wu, W.-L.; Chang, C.-C. Evaluation of chloroplast DNA markers for distinguishing *Phalaenopsis* species. *Sci. Hortic.* **2015**, *192*, 302–310. [[CrossRef](#)]
- Yu, X.Q.; Drew, B.T.; Yang, J.B.; Gao, L.M.; Li, D.Z. Comparative chloroplast genomes of eleven *Schima* (Theaceae) species: Insights into DNA barcoding and phylogeny. *PLoS ONE* **2017**, *12*, e0178026. [[CrossRef](#)]

21. Sui, C.; Liu, S.; Liu, H.; Liu, H. The complete chloroplast genome of *Paphiopedilum tranlimianum* (Orchidaceae). *Mitochondrial DNA Part B* **2018**, *3*, 820–822. [[CrossRef](#)]
22. Hou, N.; Wang, G.; Zhu, Y.; Wang, L.; Xu, J. The complete chloroplast genome of the rare and endangered herb *Paphiopedilum dianthum* (Asparagales: Orchidaceae). *Conserv. Genet. Resour.* **2017**, *10*, 709–712. [[CrossRef](#)]
23. 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*, 9040. [[CrossRef](#)] [[PubMed](#)]
24. Vu, H.-T.; Nguyen, T.-D.; Ly, L.; Nguyen, T.-A.; Vu, Q.-L.; Nguyen, T.-C.; Tran, H.-D. Construction of complete chloroplast genome of the endemic *Paphiopedilum delenatii* species of Vietnam. *Vietnam J. Biotechnol.* **2019**. accepted.
25. Vu, H.-T.; Nguyen, T.-D.; Tran, H.-D.; Vu, Q.-L.; Tran, N.; Nguyen, T.-C.; Luu, P.-N.; Tran, D.-D.; Nguyen, T.-K.; Le, L. Identification of Vietnamese *Paphiopedilum* species using DNA sequences. *Biology* **2019**. accepted. [[CrossRef](#)]
26. Li, B.; Zheng, Y. Dynamic evolution and phylogenomic analysis of the chloroplast genome in Schisandraceae. *Sci. Rep.* **2018**, *8*, 9285. [[CrossRef](#)]
27. Wu, C.S.; Chaw, S.M. Evolutionary Stasis in Cycad Plastomes and the First Case of Plastome GC-Biased Gene Conversion. *Genome Biol. Evol.* **2015**, *7*, 2000–2009. [[CrossRef](#)]
28. Duret, L.; Galtier, N. Biased Gene Conversion and the Evolution of Mammalian Genomic Landscapes. *Annu. Rev. Genom. Hum. Genet.* **2009**, *10*, 285–311. [[CrossRef](#)]
29. Glemin, S.; Arndt, P.F.; Messer, P.W.; Petrov, D.; Galtier, N.; Duret, L. Quantification of GC-biased gene conversion in the human genome. *Genome Res.* **2015**, *25*, 1215–1228. [[CrossRef](#)]
30. Palmer, J.D. Chloroplast DNA exists in two orientations. *Nature* **1983**, *301*, 92–93. [[CrossRef](#)]
31. Li, X.-Q.; Du, D. Variation, evolution, and correlation analysis of C+G content and genome or chromosome size in different kingdoms and phyla. *PLoS ONE* **2014**, *9*, e88339. [[CrossRef](#)]
32. Singh, R.; Ming, R.R.; Yu, Q. Comparative Analysis of GC Content Variations in Plant Genomes. *Trop. Plant Biol.* **2016**, *9*, 136–149. [[CrossRef](#)]
33. 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*, 715. [[CrossRef](#)] [[PubMed](#)]
34. Niu, Z.; Shuying, Z.; Jiajia, P.; Ludan, L.; Jing, S.; Xiaoyu, D. Comparative analysis of *Dendrobium* plastomes and utility of plastomic mutational hotspots. *Sci. Rep.* **2017**, *7*, 2073.
35. Karimi, K.; Wuitchik, D.M.; Oldach, M.J.; Vize, P.D. Distinguishing Species Using GC Contents in Mixed DNA or RNA Sequences. *Evol. Bioinform. Online* **2018**, *14*, 1176934318788866. [[CrossRef](#)]
36. Guo, Y.-Y.; Luo, Y.-B.; Liu, Z.-J.; Wang, X.-Q. Evolution and Biogeography of the Slipper Orchids: Eocene Vicariance of the Conduplicate Genera in the Old and New World Tropics. *PLoS ONE* **2012**, *7*, e38788. [[CrossRef](#)]
37. Ifuku, K.; Endo, T.; Shikanai, T.; Aro, E.M. Structure of the chloroplast NADH dehydrogenase-like complex: Nomenclature for nuclear-encoded subunits. *Plant Cell Physiol.* **2011**, *52*, 1560–1568. [[CrossRef](#)]
38. Nelson, N.F.; Yocum, C. Structure and Function of Photosystem I and II. *Annu. Rev. Plant Biol.* **2006**, *57*, 521–565. [[CrossRef](#)]
39. Niu, Z.; Pan, J.; Zhu, S.; Li, L.; Xue, Q.; Liu, W.; Ding, X. Comparative analysis of the complete plastomes of *Apostasia wallichii* and *Neuwiedia singaporeana* (*Apostasioideae*) reveals different evolutionary dynamics of IR/SSC boundary among photosynthetic orchids. *Front. Plant Sci.* **2017**, *8*, 1713. [[CrossRef](#)]
40. 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*, e0142215. [[CrossRef](#)]
41. Luo, J.; Hou, B.-W.; Niu, Z.-T.; Liu, W.; Xue, Q.-Y.; Ding, X.-Y. Comparative Chloroplast Genomes of Photosynthetic Orchids: Insights into Evolution of the Orchidaceae and Development of Molecular Markers for Phylogenetic Applications. *PLoS ONE* **2014**, *9*, e99016. [[CrossRef](#)]
42. Cavalier-Smith, T. Chloroplast evolution: Secondary symbiogenesis and multiple losses. *Curr. Biol.* **2002**, *12*, R62–R64. [[CrossRef](#)]

43. Vieira, M.L.C.; Santini, L.; Diniz, A.L.; Munhoz, C.d.F. Microsatellite markers: What they mean and why they are so useful. *Genet. Mol. Biol.* **2016**, *39*, 312–328. [[CrossRef](#)] [[PubMed](#)]
44. Chang, Y.C.; Chang, C.H. Common repeat sequences in bacterial genomes. *J. Med. Biol. Eng.* **2003**, *23*, 65–72.
45. Chinnappareddy, L.R.D.; Khandagale, K.; Srinivas Reddy, S.H.; Kanupriya, C.; Chennareddy, A.; Singh, T.H. SSR-Based DNA Barcodes as a Tool for Identification of Eggplant Genotypes. *Int. J. Veg. Sci.* **2012**, *18*, 260–271. [[CrossRef](#)]
46. Tsai, C.-C.; Wu, P.-Y.; Kuo, C.-C.; Huang, M.-C.; Yu, S.-K.; Hsu, T.-W.; Chiang, T.-Y.; Chiang, Y.-C. Analysis of microsatellites in the vulnerable orchid *Gastrodia flavilabella*: The development of microsatellite markers, and cross-species amplification in *Gastrodia*. *Bot. Stud.* **2014**, *55*, 72. [[CrossRef](#)]
47. Phuekvilai, P.; Pongtongkam, P.; Peyachoknagul, S. Development of Microsatellite Markers for Vanda Orchid. *Kasetsart J. Nat. Sci.* **2009**, *43*, 497–506.
48. Liu, Y.T.; Chen, R.K.; Lin, S.J.; Chen, Y.C.; Chin, S.W.; Chen, F.C.; Lee, C.Y. Analysis of sequence diversity through internal transcribed spacers and simple sequence repeats to identify *Dendrobium* species. *Genet. Mol. Res.* **2014**, *13*, 2709–2717. [[CrossRef](#)]
49. Jonah, P.M.; Bello, L.L.; Lucky, O.; Midau, A.; Moruppa, S. Review: The importance of molecular markers in plant breeding programs. *Glob. J. Sci. Front. Res.* **2011**, *11*, 4–12.
50. Kalia, R.; Rai, M.; Kalia, S.; Singh, R.; Dhawan, A. Microsatellite markers: An overview of the recent progress in plants. *Euphytica* **2011**, *177*, 309–334. [[CrossRef](#)]
51. Nadeem, M.A.; Nawaz, M.A.; Shahid, M.Q.; Doğan, Y.; Comertpay, G.; Yıldız, M.; Hatipoğlu, R.; Ahmad, F.; Alsaleh, A.; Labhane, N.; et al. DNA molecular markers in plant breeding: Current status and recent advancements in genomic selection and genome editing. *Biotechnol. Biotechnol. Equip.* **2018**, *32*, 261–285. [[CrossRef](#)]
52. Echt, C.S.; Deverno, L.L.; Anzidei, M.; Vendramin, G.G. Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait. *Mol. Ecol.* **1998**, *7*, 307–316. [[CrossRef](#)]
53. Huang, J.; Yang, X.; Zhang, C.; Yin, X.; Liu, S.; Li, X. Development of Chloroplast Microsatellite Markers and Analysis of Chloroplast Diversity in Chinese Jujube (*Ziziphus jujuba* Mill.) and Wild Jujube (*Ziziphus acidojujuba* Mill.). *PLoS ONE* **2015**, *10*, e0134519. [[CrossRef](#)] [[PubMed](#)]
54. Deng, Q.; Zhang, H.; He, Y.; Wang, T.; Su, Y. Chloroplast microsatellite markers for *Pseudotsaxus chienii* developed from the whole chloroplast genome of *Taxus chinensis* var. *mairei* (Taxaceae). *Appl. Plant Sci.* **2017**, *5*, 1600153. [[CrossRef](#)] [[PubMed](#)]
55. Park, H.; Kim, C.; Lee, Y.-M.; Kim, J.-H. Development of chloroplast microsatellite markers for the endangered *Maianthemum bicolor* (Asparagaceae s.l.). *Appl. Plant Sci.* **2016**, *4*, 1600032. [[CrossRef](#)]
56. Pinheiro, F.; Palma-Silva, C.; Barros, F.; Pessoa Felix, L.; Lexer, C.; Cozzolino, S.; Fay, M. Chloroplast microsatellite markers for the Neotropical orchid genus *Epidendrum*, and cross-amplification in other *Laeliinae* species (Orchidaceae). *Conserv. Genet. Resour.* **2009**, *1*, 505–511. [[CrossRef](#)]
57. Ebert, D.; Peakall, R.O.D. Chloroplast simple sequence repeats (cpSSRs): Technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. *Mol. Ecol. Resour.* **2009**, *9*, 673–690. [[CrossRef](#)]
58. Nepal, M.; Piya, S. Characterization of Nuclear and Chloroplast Microsatellite Markers for *Falcaria vulgaris* (Apiaceae). *Am. J. Plant Sci.* **2013**, *4*, 590–595.
59. Mehrotra, S.; Goyal, V. Repetitive sequences in plant nuclear DNA: Types, distribution, evolution and function. *Genom. Proteom. Bioinform.* **2014**, *12*, 164–171. [[CrossRef](#)]
60. Ballardini, M.; Mercuri, A.; Littardi, C.; Abbas, S.; Couderc, M.; Ludeña, B.; Pintaud, J.-C. The chloroplast DNA locus *psbZ-trnFM* as a potential barcode marker in *Phoenix* L. (Arecaceae). *ZooKeys* **2013**, *365*, 71–82. [[CrossRef](#)]
61. Singh, H.K.; Parveen, I.; Raghuvanshi, S.; Babbar, S.B. The loci recommended as universal barcodes for plants on the basis of floristic studies may not work with congeneric species as exemplified by DNA barcoding of *Dendrobium* species. *BMC Res. Notes* **2012**, *5*, 42. [[CrossRef](#)]
62. Krawczyk, K.; Nobis, M.; Myszczynski, K.; Klichowska, E.; Sawicki, J. Plastid super-barcodes as a tool for species discrimination in feather grasses (Poaceae: Stipa). *Sci. Rep.* **2018**, *8*, 1924. [[CrossRef](#)] [[PubMed](#)]
63. Chen, X.; Zhou, J.; Cui, Y.; Wang, Y.; Duan, B.; Yao, H. Identification of *Ligularia* Herbs Using the Complete Chloroplast Genome as a Super-Barcode. *Front. Pharmacol.* **2018**, *9*, 695. [[CrossRef](#)] [[PubMed](#)]

64. Nock, C.; Waters, D.; Shepherd, M.; Bundock, P.; Henry, R. Plant DNA Barcoding using chloroplast genome sequences. In Proceedings of the IBC XVII International Botanical Congress, Melbourne, Australia, 23–30 July 2011.
65. Vu, H.-T.; Huynh, P.; Tran, H.-D.; Le, L. In Silico Study on Molecular Sequences for Identification of Paphiopedilum Species. *Evol. Bioinform.* **2018**, *14*, 117693431877454. [[CrossRef](#)] [[PubMed](#)]
66. Hosein, F.N.; Austin, N.; Maharaj, S.; Johnson, W.; Rostant, L.; Ramdass, A.C.; Rampersad, S.N. Utility of DNA barcoding to identify rare endemic vascular plant species in Trinidad. *Ecol. Evol.* **2017**, *7*, 7311–7333. [[CrossRef](#)] [[PubMed](#)]
67. Bi, Y.; Zhang, M.-F.; Xue, J.; Dong, R.; Du, Y.-P.; Zhang, X.-H. Chloroplast genomic resources for phylogeny and DNA barcoding: A case study on Fritillaria. *Sci. Rep.* **2018**, *8*, 1184. [[CrossRef](#)] [[PubMed](#)]
68. Somaratne, Y.; Guan, D.-L.; Abbood, N.N.; Zhao, L.; Wang, W.-Q.; Xu, S.-Q. Comparison of the Complete Eragrostis pilosa Chloroplast Genome with Its Relatives in Eragrostideae (Chloridoideae; Poaceae). *Plants* **2019**, *8*, 485. [[CrossRef](#)]
69. Cui, Y.; Nie, L.; Sun, W.; Xu, Z.; Wang, Y.; Yu, J.; Song, J.; Yao, H. Comparative and Phylogenetic Analyses of Ginger (Zingiber officinale) in the Family Zingiberaceae Based on the Complete Chloroplast Genome. *Plants* **2019**, *8*, 283. [[CrossRef](#)]
70. Averyanov, L.; Cribb, P.; Phan, K.L.; Nguyen, T.H. *Slipper Orchids of Vietnam*; (Vietnamese edition); Giao Thong Van Tai Publishing House: Ho Chi Minh City, Vietnam, 2004; p. 308.
71. Kress, W.J.; Erickson, D.L. DNA barcodes: Genes, genomics, and bioinformatics. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 2761–2762. [[CrossRef](#)]
72. Kress, W.J.; Erickson, D.L. A two-locus global DNA barcode for land plants: The coding rbcL gene complements the non-coding trnH-psbA spacer region. *PLoS ONE* **2007**, *2*, e508. [[CrossRef](#)]
73. Fazekas, A.J.; Burgess, K.S.; Kesanakurti, P.R.; Graham, S.W.; Newmaster, S.G.; Husband, B.C.; Percy, D.M.; Hajibabaei, M.; Barrett, S.C.H. Multiple Multilocus DNA Barcodes from the Plastid Genome Discriminate Plant Species Equally Well. *PLoS ONE* **2008**, *3*, e2802. [[CrossRef](#)]
74. Dong, W.; Liu, H.; Xu, C.; Zuo, Y.; Chen, Z.; Zhou, S. A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: A case study on ginsengs. *BMC Genet.* **2014**, *15*, 138. [[CrossRef](#)] [[PubMed](#)]
75. Jayakodi, M.; Choi, B.S.; Lee, S.C.; Kim, N.H.; Park, J.Y.; Jang, W.; Lakshmanan, M.; Mohan, S.V.G.; Lee, D.Y.; Yang, T.J. Ginseng Genome Database: An open-access platform for genomics of Panax ginseng. *BMC Plant Biol.* **2018**, *18*, 62. [[CrossRef](#)] [[PubMed](#)]
76. Scarcelli, N.; Barnaud, A.; Eiserhardt, W.; Treier, U.A.; Seveno, M.; d’Anfray, A.; Vigouroux, Y.; Pintaud, J.-C. A Set of 100 Chloroplast DNA Primer Pairs to Study Population Genetics and Phylogeny in Monocotyledons. *PLoS ONE* **2011**, *6*, e19954. [[CrossRef](#)] [[PubMed](#)]
77. Zhao, Z.; Wang, X.; Yu, Y.; Yuan, S.; Jiang, D.; Zhang, Y.; Zhang, T.; Zhong, W.; Yuan, Q.; Huang, L. Complete chloroplast genome sequences of Dioscorea: Characterization, genomic resources, and phylogenetic analyses. *PeerJ* **2018**, *6*, e6032. [[CrossRef](#)]
78. Yaradua, S.S.; Alzahrani, D.A.; Albokhary, E.J.; Abba, A.; Bello, A. Complete Chloroplast Genome Sequence of Justicia flava: Genome Comparative Analysis and Phylogenetic Relationships among Acanthaceae. *BioMed Res. Int.* **2019**, *2019*, 4370258. [[CrossRef](#)] [[PubMed](#)]
79. Yamane, K.; Yasui, Y.; Ohnishi, O. Intraspecific cpDNA variations of diploid and tetraploid perennial buckwheat, Fagopyrum cymosum (Polygonaceae). *Am. J. Bot.* **2003**, *90*, 339–346. [[CrossRef](#)]
80. Cho, K.S.; Yun, B.K.; Yoon, Y.H.; Hong, S.Y.; Mekapogu, M.; Kim, K.H.; Yang, T.J. Complete Chloroplast Genome Sequence of Tartary Buckwheat (Fagopyrum tataricum) and Comparative Analysis with Common Buckwheat (F. esculentum). *PLoS ONE* **2015**, *10*, e0125332. [[CrossRef](#)]
81. Song, Y.; Chen, Y.; Lv, J.; Xu, J.; Zhu, S.; Li, M. Comparative Chloroplast Genomes of Sorghum Species: Sequence Divergence and Phylogenetic Relationships. *BioMed Res. Int.* **2019**, *2019*, 5046958. [[CrossRef](#)]
82. Dierckxsens, N.; Mardulyn, P.; Smits, G. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* **2016**, *45*, e18.
83. Tillich, M.; Lehwark, P.; Pellizzer, T.; Ulbricht-Jones, E.S.; Fischer, A.; Bock, R.; Greiner, S. GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **2017**, *45*, W6–W11. [[CrossRef](#)]
84. 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. [[CrossRef](#)] [[PubMed](#)]

85. Kurtz, S.; Choudhuri, J.V.; Ohlebusch, E.; Schleiermacher, C.; Stoye, J.; Giegerich, R. REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **2001**, *29*, 4633–4642. [[CrossRef](#)] [[PubMed](#)]
86. Beier, S.; Thiel, T.; Münch, T.; Scholz, U.; Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [[CrossRef](#)] [[PubMed](#)]
87. Schliep, K.P. Phangorn: Phylogenetic analysis in R. *Bioinformatics* **2010**, *27*, 592–593. [[CrossRef](#)]
88. Rambaut, A. FigTree: Tree Figure Drawing Tool Version 1.2.2. 2009. Available online: <http://tree.bio.ed.ac.uk/software/figtree/> (accessed on 11 January 2019).
89. Kumar, S.; Nei, M.; Dudley, J.; Tamura, K. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief. Bioinform.* **2008**, *9*, 299–306. [[CrossRef](#)]
90. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [[CrossRef](#)]



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).