MDPI

Article

Plastome Evolution and Comparative Analyses of a Recently Radiated Genus *Vanda* (Aeridinae, Orchidaceae)

Wanshun Lei [1], Peng Zhou [1], Zelong Pei [1], Yizhen Liu [1], Yan Luo [2] and Xiaoguo Xiang [1,*]

[1] Key Laboratory of Poyang Lake Environment and Resource Utilization, Ministry of Education, School of Life Sciences, Nanchang University, Nanchang 330031, China; 405600220103@email.ncu.edu.cn (W.L.); 355600230021@email.ncu.edu.cn (P.Z.); 5604120024@email.ncu.edu.cn (Z.P.); liuyizhen@ncu.edu.cn (Y.L.)

[2] Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla 666300, China; luoyan@xtbg.org.cn

[*] Correspondence: xiangxg2018@ncu.edu.cn

Abstract: *Vanda* R.Br. is an epiphytic orchid genus with significant horticultural and ornamental value. Previous molecular studies expanded *Vanda* including some members from five other genera. However, the interspecific relationships of this recently radiated genus have remained unclear based on several DNA markers until now. In this study, the complete plastome has been used to infer the phylogenetic relationships of *Vanda s.l.* The five newly obtained plastomes ranged from 146,340 bp to 149,273 bp in length, with a GC content ranging from 36.5% to 36.7%. The five plastomes contained 74 protein-coding genes (CDSs), 38 tRNAs, and 8 rRNAs, and their *ndh* genes underwent loss or pseudogenization. Comparative plastome analyses of 13 *Vanda* species revealed high conservation in terms of genome size, structure, and gene order, except for a large inversion from *trnG*^{GCC} to *ycf*3 in *V. coerulea*. Moreover, six CDSs and five non-CDSs were selected as candidate DNA barcodes. Our phylogenetic analyses demonstrated that *Vanda s.l.* is a monophyletic group with high supporting values based on five different datasets (complete plastome with one IR, 68 CDSs, LSC, five hypervariable non-CDSs, and six hypervariable CDSs), while the phylogenetic relationships among species were fully resolved based on the complete plastome with one IR dataset. Our results confirmed that the complete plastome has a great power in resolving the phylogenetic relationships of recently radiated lineages.

Keywords: *Vanda*; plastome; mutational hotspots; phylogenomics; rapid radiation lineages

1. Introduction

Vanda R.Br. (Aeridinae, Vandeae, Epidendroideae, and Orchidaceae) comprises approximately 73 species [1,2], and is found in subtropical and tropical Asia and Australia (Queensland). Over half of the species exhibit narrow endemism, with a center of species diversity in the South-East Asian archipelagos [3]. Particularly, there is great diversity in floral characters, and thus, *Vanda* has been recognized as one of the five most horticulturally important orchid genera in the world [3–5].

Vanda is considered to be one of the most taxonomically complicated groups in Aeridinae, due to its morphological similarity to *Ascocentrum*, *Euanthe*, *Trudelia*, and *Christensonia* [4]. The genus *Vanda* was established in 1820 by Robert Brown [6]. Lindley [7] redefined this genus, and postulated five sections. Gardiner [8] transferred 17 species from *Ascocentrum*, *Ascocentropsis*, *Christensonia*, *Eparmatostigma*, and *Neofinetia* into *Vanda*, and proposed to establish *Vanda s.l.* Subsequently, Gardiner et al. [3] used three plastid regions (*matK*, *psbA-trnH*, and *trnL-F*) to partly resolve the phylogenetic relationships among *Vanda s.l.*, and proposed 13 morphological sectional classifications. Chase et al. [2] adopted this broad taxonomic generic concept. Although the *Vanda s.l.* has been widely accepted, the infrageneric relationships remained controversial in phylogenetic research. For example, *V. alpina* and *V. pumila* were clustered together based on five DNA sequences (ITS, *atpI-H*,

matK, *psbA-trnH*, and *trnL-F*) [9], but this relationship was not supported when using ITS and five plastid DNA regions (*atpI-H*, *matK*, *psbA-trnH*, *trnL-F*, and *trnS-trnfM*) [10]. Moreover, many sections proposed by Gardiner et al. [3] were not supported as monophyletic groups based on ITS and three plastid DNA regions (*matK*, *trnH-psbA*, and *trnL-F*) [11]. Zhang et al. [12] investigated the diversification of Orchidaceae and inferred that *Vanda s.l.* originated 8.19 million years ago (Ma), and then rapidly diversified following the Pliocene. Briefly, several DNA markers had limited power to reveal the interspecific relationships of the radiated *Vanda s.l.* Therefore, it is crucial to explore more genetically varied loci to elucidate the phylogenetic relationships within *Vanda*.

Until now, the complete plastome has successfully clarified the phylogenetic relationships of recently radiated lineages. For instance, Li et al. [13] used plastome data to explore the relationships within *Holcoglossum*, which had divergently radiated since the late Miocene [14]. Moreover, the phylogenetic relationships of *Eriocaulon*, which diverged following the late Miocene and diversified in the Quaternary, were well resolved by the complete plastome [15]. Wang et al. [16] successfully verified the species relationships within *Lagerstroemia* (Lythraceae) based on the plastome data, which rapidly radiated following the late Miocene. Thus, it is necessary to explore the plastome character of *Vanda s.l.* and employ the complete plastome to investigate these recently rapidly diverging lineages.

In this study, the plastomes of five *Vanda* species were newly obtained, and eight published plastomes were downloaded from GenBank. In total, 13 plastomes were combined to conduct plastome comparative analyses and phylogenetic reconstruction within *Vanda s.l.* The goals of this study are: (1) to reveal the gene content, gene order, and structures of *Vanda* plastomes; (2) to detect the hypervariable regions of these plastomes as prospective DNA markers for species identification; (3) to explore the phylogenetic relationships within this genus. Our study provides valuable genetic information on the phylogeny, species identification, and conservation of *Vanda s.l.*

2. Result

2.1. The Plastome Characters of Vanda

The plastomes of *Vanda* species contained four regions: a large single copy (LSC) region, a small single copy (SSC) region, and two copies of inverted repeats (IRa/b) (Figure 1). The total length of the 13 plastomes in the genus *Vanda* ranged from 146,340 bp (*V. ampullaceum*) to 149,490 bp (*V. subconcolor*) (Table 1). The lengths of the LSC regions were 83,808–85,982 bp, those of the SSC regions were 11,424–12,002 bp, and a pair of IR regions ranged from 24,523 to 25,965 bp (Table 1). The total GC contents ranged from 36.5% to 36.7%. All of the 13 *Vanda* plastomes contained 120 genes, comprising 74 protein-coding genes (CDSs), 38 transfer RNA (tRNA) genes, and eight ribosomal RNA (rRNA) genes (Table 1).

Table 1. Summary of plastome characters among 13 *Vanda* species.

| Taxa | Length (bp) | | | | GC Content (%) | Gene Numbers | | | |
	Plastome	LSC	IR	SSC		Total	CDS	tRNA	rRNA
V. alpina	149,273	85,819	25,860	11,734	36.7	120	74	38	8
V. ampullaceum	146,340	85,436	24,523	11,858	36.5	120	74	38	8
V. brunnea	149,216	85,783	25,860	11,713	36.7	120	74	38	8
V. coelestis	148,073	85,105	25,772	11,424	36.8	120	74	38	8
V. coerulea	149,268	85,982	25,787	11,712	36.6	120	74	38	8
V. coerulescens	149,410	85,954	25,965	11,526	36.7	120	74	38	8
V. concolor	149,474	85,678	25,897	12,002	36.6	120	74	38	8
V. cristata	146,993	83,927	25,772	11,522	36.7	120	74	38	8
V. falcata	146,497	83,808	25,457	11,775	36.6	120	74	38	8
V. pumila	149,273	85,819	25,860	11,734	36.7	120	74	38	8
V. richardsiana	146,498	83,809	25,457	11,775	36.6	120	74	38	8
V. subconcolor	149,490	85,691	25,912	11,975	36.6	120	74	38	8
V. xichangensis	146,681	83,920	25,505	11,751	36.6	120	74	38	8

All the *ndh* genes of *Vanda* plastomes had been lost or pseudogenized (Figure 1; Table 1). The plastomes of *V. cristata*, *V. falcata*, *V. richardsiana*, and *V. xichangensis* possessed five pseudogenes (*ndhB*, *ndhD*, *ndhE*, *ndhG*, and *ndhI*) and lost six genes (*ndhA*, *ndhC*, *ndhF*, *ndhH*, *ndhJ, and ndhK*). The other nine species possessed eight pseudogenes (*ndhB*, *ndhC*, *ndhD*, *ndhE*, *ndhG*, *ndhI*, *ndhJ*, and *ndhK*), and lost three genes (*ndhA*, *ndhF*, and *ndhH*) (Figure 1).
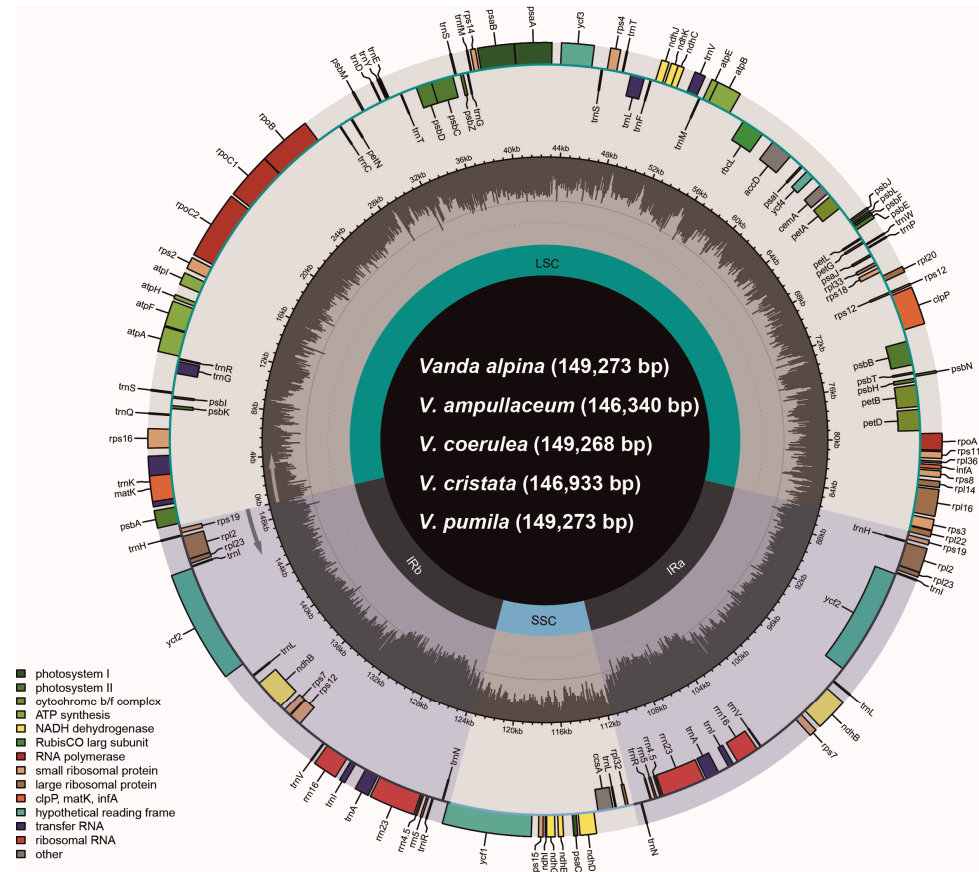

Figure 1. The gene map of five *Vanda* plastomes. The darker gray in the inner circle corresponds to the GC content. The IRa/b, LSC, and SSC are shown inside the GC content.

There was no expansion and contraction in the IR/SC boundary regions across the 13 species of *Vanda* (Figure S1). *Rpl22* (318–329 bp) genes of the LSC crossed over into IRb. The *trnN* (72 bp) and *rpl32* (174 bp) genes were near to the junction between SSC and IRb. Within the junction region between SSC and IRa, the *ycf1* genes were across the SSC-IRa boundary, and primarily located in the SSC region, with lengths ranging from 5211 bp (*V. coerulescens*) to 5328 bp (*V. concolor*). The adjacent regions of LSC and IRb were located in the *rps19* (279 bp) genes and *psbA* (1062 bp) genes. The collinearity analysis revealed that the SSC and IR regions of 13 *Vanda* plastomes were highly conserved. Additionally, conservation was observed in the LSC region of 12 *Vanda* plastomes, and a significant inversion from *trnG*^{GCC} to *ycf3* was identified in the LSC region of the *V. coerulea* plastome (Figure 2).

Figure 2. Visualization of the collinearity. Structural alignment of 13 *Vanda* plastomes using progressive Mauve. The tree topology and species order were consistent with that in Figure 5. The lines linking the collinear blocks represent homology between different genomes. Numbers on the upper *x*-axis are genome map coordinates in kilobases (kb).

2.2. Repeated Analysis

Different types of the simple sequence repeats (SSRs) of *Vanda* were detected (Figure 3A,B). The number of identified SSRs ranged from 52 (*V. coerulea*) to 69 (*V. subconcolor*). There were 775 SSRs detected in the 13 *Vanda* plastomes. Six types of SSRs (mono-nucleotide, di-nucleotide, tri-nucleotide, tetra-nucleotide, penta-nucleotide, and hexa-nucleotide) were identified, with 514 SSRs (66.32%) being mono-nucleotide type, particularly A and T repeat motifs. In di-nucleotide repeat type (115), the number of AT/TA repeat motifs (89) exceeded that of TC/GA (26). Hexa-nucleotide repeats (3) occurred with the lowest frequency, which was only detected in *V. coerulescens*, *V. concolor*, and *V. subconcolor*. Additionally, the LSC region had a higher content of SSRs (530) than the SSC region (131) and the IR regions (114). There were 637 long repeats in the *Vanda* plastomes, of which 213 (33.44%) were forward repeats, 324 (50.86%) were palindromic repeats, 96 (15.07%) were reverse repeats, and four (0.63%) were complement repeats. The lengths of the long repeats mainly ranged from 20 bp to 39 bp.
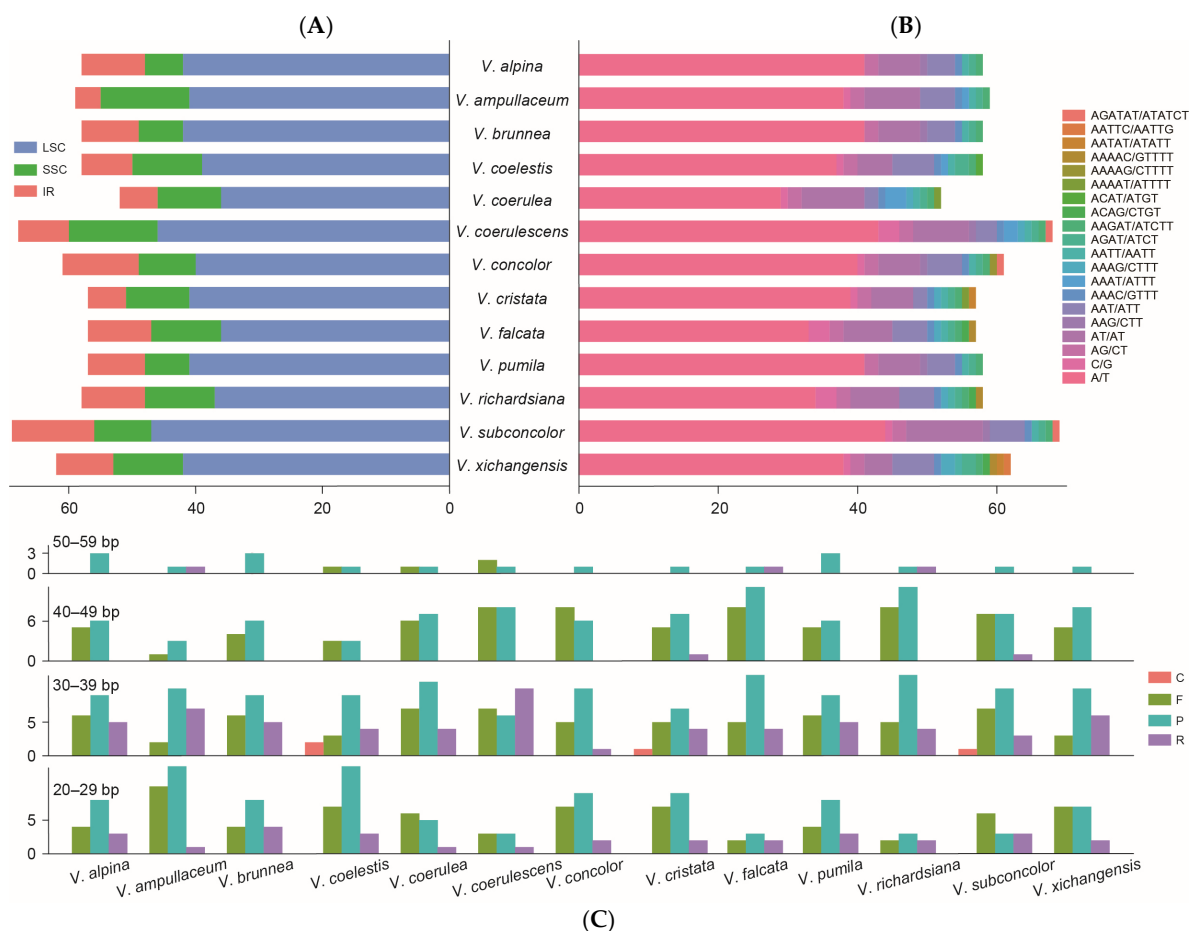
Figure 3. Summary of the simple sequence repeats (SSR) and long repeat sequences of 13 *Vanda* plastomes. (**A**) SSR distributions in the LSC, SSC, and IR regions. (**B**) Number of different types of SSRs identified in plastomes. (**C**) Number of long repeats by sequence length (F: forward repeats; R: reverse repeats; P: palindromic repeats; C: complementary repeats).

2.3. Codon Usage Bias

The codon usage frequency of *Vanda* plastomes was calculated based on concatenated sequences of 68 CDSs. The heatmap showed that codon usage bias was highly conserved in *Vanda* (Figure S2). The relative synonymous codon usage (RSCU) analysis demonstrated that codons containing A/T at 3′ exhibited high RSCU values (RSCU ≥ 1), and codons ending with C/G at 3′ showed low RSCU values (RSCU < 1). The RSCU values of *Vanda* plastomes ranged from 0.29 to 1.89 (Table S1). The AGA codon had the highest RSCU values (1.86–1.89), while AGC exhibited the lowest RSCU values (0.29–0.30). The UGA codon had higher RSCU values (0.66–0.70) than the other two termination codons (UAA and UAG). In addition, leucine (Leu) had the highest count of codons (2062–2086), while cysteine (Cys) exhibited the lowest (226–229).

2.4. Plastome Sequence Divergence and DNA Marker Investigation

The structural characteristics of 13 *Vanda* plastomes were investigated to evaluate the sequence of the plastomes (Figure S3). The coding regions exhibited higher conservation than the non-coding regions, and the LSC and SSC regions displayed greater variability than the IR regions (Figure S3). The nucleotide diversity (Pi) was calculated to identify the highly mutated hotspots of *Vanda* plastomes (Figure 4). The results showed that the Pi values of the LSC region, SSC region, and IR regions were 0.0075, 0.0064, and 0.0016, respectively. Based on the ranking of the Pi values, five hypervariable regions, including *clpP-psbB*, *rps15-ycf1*, *trnN*^{GUU}*-rpl32*, *trnS*^{GCU}*-trnG*^{GCC}, and *trnT*^{UGU}*-trnL*^{UAA} were identified in the

complete plastomes. As for 68 CDSs, six hypervariable regions were identified, including *ccsA*, *psbI*, *rpl22*, *rpoA*, *rps15*, and *ycf1*.
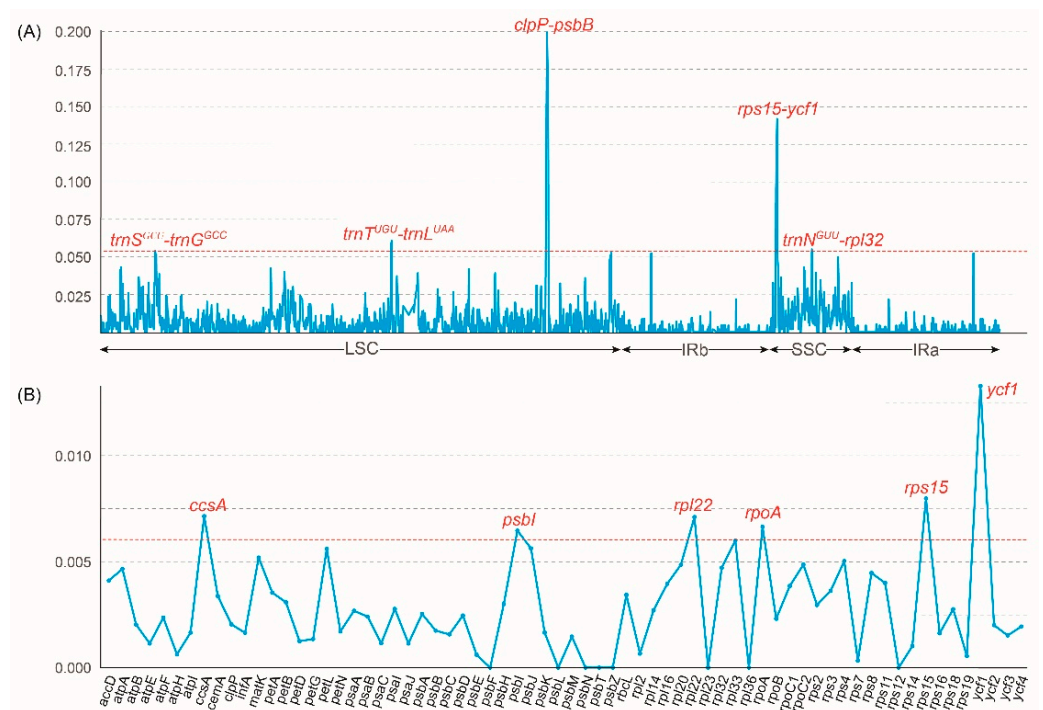

Figure 4. Sliding window test of nucleotide diversity (Pi) in the *Vanda* plastomes. (**A**) The nucleotide diversity of the complete plastomes showing five mutation hotspot regions (dashed line marked the Pi > 0.54). The *x*-axis represented the position of the midpoint of a window, while the *y*-axis represents the Pi value of each window. (**B**) The nucleotide diversity of 68 CDSs showing six mutation hotspot regions (dashed line marked the Pi > 0.06).

2.5. Phylogenetic Analyses and Character Reconstruction

Maximum likelihood (ML) and Bayesian inference (BI) analyses yielded nearly identical topologies based on the complete plastome with one IR, 68 CDSs, LSC, five hypervariable non-CDSs and six hypervariable CDSs, respectively, except the uncertain systematic position of *V. ampullaceum* (Figure 5 and Figure S4). The monophyly of *Vanda s.l.* have confirmed by the five datasets with high Bootstrap percentages (BS values) from the ML analyses and Posterior Probability (PP values) from the Bayesian analyses, respectively. The phylogenetic relationships among species were fully resolved based on the complete plastome (excluding one IR) matrix, and the phylogenetic tree was illustrated here. *Vanda s.l.* was well supported as a monophyletic group (PP-BI = 1.00, BS-ML = 100%) (Figure 5). Three species from sect. *Neofinetia* formed a clade and diverged first (PP-BI = 1.00, BS-ML = 100%). The other sections formed another clade with high to moderate supporting values (PP-BI = 1.00, BS-ML = 71%). The three species from sect. *Cristatae* and the two species from sect. *Longicalcarata* did not form a clade.

The reconstruction of *ndh* gene status showed that the most recent common ancestor (MRCA) of *Vanda* compared to its relatives involved the *ndhI* pseudogene. The *ndhC*, *ndhJ*, and *ndhK* were lost in sect. *Neofinetia* and *V. cristata* (Figure 5).
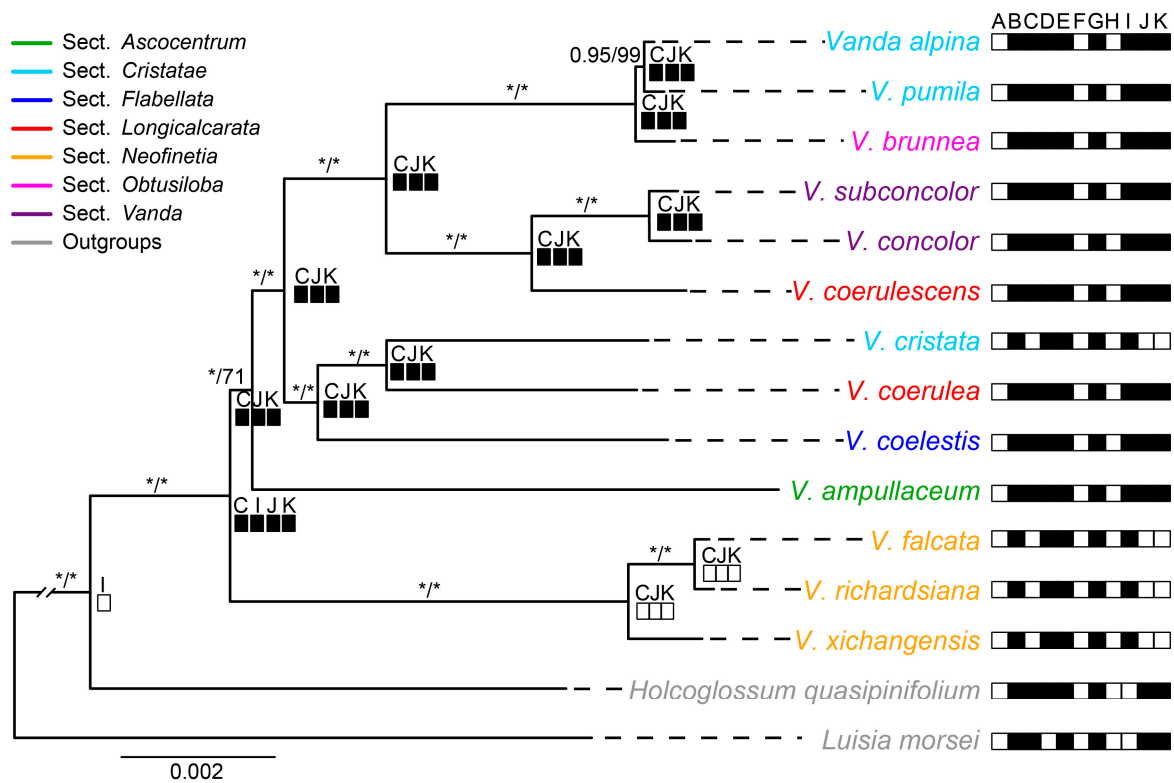
Figure 5. Maximum likelihood phylogenetic tree of *Vanda* based on whole plastomes with one IR and reconstruction of *ndh* gene loss and pseudogenization. The numbers above represent the supporting values of Bayesian inference and maximum likelihood analyses, respectively, and the star reflects that the supporting value is equal to 100%. The black box represents the pseudogene of the *ndh* gene, and the blank box represents the absence of the *ndh* gene.

3. Discussion

3.1. The Plastome Characteristics and Structural Evolution

In this study, the complete plastome sequences of five *Vanda* species were newly obtained. The *Vanda* plastomes contained one LSC, one SSC, and two IR regions (Figure 1), which is similar to most angiosperm [13,15,16]. The genome sizes ranged from 146,340 bp to 149,490 bp and the GC contents of the plastomes range from 36.5% to 36.7% and fall within the range of other orchids, such as *Aerides* and *Angraecum* [17,18]. A total of 120 genes were identified in the *Vanda* plastomes, consisting of 74 CDSs, 38 tRNA genes, and 8 rRNA genes (Table 1). Further, all *ndh* genes were pseudogenized or lost (Figure 5). The *ndh* genes have been involved in photosynthesis, the photosynthetic response and stress acclimation, and have been hypothesized to be closely related to plant transition to terrestrial habitats and are usually absent in epiphytic plants [19,20]. Previous studies showed that all 11 *ndh* genes were lost or pseudogenized in nearly all photosynthetic and non-photosynthetic orchids, which suggests that the *ndh* genes have been independently lost in various orchid lineages [21]. As such, all of the *ndh* genes were truncated or pseudogenized in the epiphytic orchid genera *Holcoglossum* [13], *Paraphalaenopsis* [22], and *Trichoglottis* [23].

Our results reveal that the gene arrangement of the IR/SC boundary in *Vanda* plastomes was extremely conserved (Figure S1), and the gene content of this genus was identical. This finding implied that the length variation of *Vanda* plastomes did not attribute to the IR/SC boundary shift, although the IR/SC boundary shift has been considered to be the main factor for the variations in plastome length and gene content [24].

Previous studies have revealed that gene structure and gene order are highly conserved in orchid plastomes [25,26]. Our collinearity analysis revealed that the structures of 13 *Vanda* plastomes were conserved, while there was a large inversion from *trnG^GCC* to *ycf3* existing

in the *V. coerulea* plastome (Figure 2). Plastome rearrangement was also reported for the plastome of Orchidaceae. For example, *Cypripedium formosanum* has an inversion from *atpA* to *petG* [27], and there is a large inversion from *trnS*^{GCU} to *trnS*^{GGA} in the *Apostasia wallichii* plastome [28].

Our results showed that among 13 *Vanda* plastomes, the codon usage bias was extremely conserved (Figure S2). Leucine (Leu) had the highest amino acid frequency, while cysteine (Cys) had the lowest frequency. The high leucine frequency may be related to its critical function in photosynthesis-related metabolism [29]. The measure of codon usage bias could be helpful to explore the evolutionary patterns of lineages [30].

3.2. The Barcoding Investigation and Phylogenetic Analyses

SSRs have been recognized as a crucial molecular marker in population genetics and conservation biology [31,32]. A total of 775 SSRs were identified in the *Vanda* plastome, with 66.32% of them being mono-nucleotide repeats, which are mainly located in the intergeneric regions of the LSC (Figure 3). The majority of the nucleotides consisted of A/T motifs rather than C/G motifs, potentially attributed to the A/T-rich plastome structure, which was consistent with the results of previous studies [33,34].

In this study, five non-CDSs regions (*clpP-psbB*, *rps15-ycf1*, *trnN*^{GUU}-*rpl32*, *trnS*^{GCU}-*trnG*^{GCC}, and *trnT*^{UGU}-*trnL*^{UAA}) and six CDSs (*ccsA*, *psbI*, *rpl22*, *rpoA*, *rps15*, and *ycf1*) were selected as candidate barcodes (Figure 4). Similarly, numerous plastome comparative studies within orchids also detected some hypervariable hotspots for the species diversity groups [35,36]. Our findings can serve as specific DNA markers for species identification in *Vanda s.l.*

Plastomes provide abundant genetic information to reveal the phylogenetic relationships within a genus [37]. Previous research about the phylogenetic relationships of *Vanda* partly resolved the interspecific relationships based on several DNA markers [9–11,38]. In this study, our phylogenetic results based on complete plastomes (excluding one IR) and four other datasets strongly supported that *Vanda s.l.* was a monophyletic group, which is consistent with previous studies [9–11]. In addition, the complete plastome perfectly resolved the interspecific and infrageneric relationships of *Vanda s.l.* Among the seven sampled sections of 13 morphological sections proposed in Gardiner et al. [3], our phylogenetic results indicate that five sections were recognized, and two sections (sect. *Longicalcarata* and sect. *Cristatae*) were not monophyletic. The stem of *Vanda s.l.* diverged following the late Miocene, and rapidly diversified following the Pliocene [16]. It is acknowledged that the phylogenetic relationships among the radiated groups were difficult to resolve with several markers. For instance, compared with the results of a few DNA markers, the complete plastome, except in the case of IR, was successfully used to clarify the interspecific relationships of *Holcoglossum*, which diverged and diversified following the latest Miocene [12,13]. The plastomes also showed their power in reconstructing phylogenetic relationships of other recently diverged non-orchid lineages, such as *Eriocaulon* [14] and *Lagerstroemia* [15]. Therefore, we posit that the plastome could be valuable for exploring the phylogenetic relationships of *Vanda s.l.*, and may also be useful for other lineages that have recently undergone rapid radiation.

4. Materials and Methods

4.1. Sampling and Sequencing

In this study, five *Vanda* species, representing three sections (sect. *Ascocentrum*, sect. *Cristatae* and sect. *Longicalcarata*), were collected from Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences. Eight published plastomes of *Vanda* were downloaded from GenBank and the annotations were updated. In total, 13 species were sampled to represent seven sections of *Vanda* as accurately as possible. Additionally, two closely related species (*Holcoglossum quasipinifolium* and *Luisia morsei*) were selected as outgroups based on Zhang et al. [16]. The detailed sampling information was listed in Table S2.

Using the modified CTAB method [39], we extracted the total DNA of five sampled *Vanda* species from silica gel-dried leaves, respectively. Then, following the manuals of NEB Next® Ultra DNA Library Prep Kit (NEB, Ipswich, MA, USA), the libraries for paired-end 150 bp sequencing were constructed using an Illumina HiSeq 2000 platform. Finally, around 4 Gb of raw data was generated for each sampled species.

4.2. Assembly and Annotation

After the assessment of the quality of raw sequence reads using FastQC v0.11.9 [40], we filtered the low-quality reads by Trimmomatic v0.39 [41], and assembled the clean reads using GetOrganelle v1.7.3.2 [42]. Then, to evaluate the completeness of the assembled plastomes, we checked and visualized them in Bandage v0.7.1 [43]. Finally, annotation was conducted using PGA software [44] and manually adjusted by Geneious v9.05 [45]. The plastome of *Vanda brunnea* (NC_041522) was selected as reference. The gene map was plotted by OGDRAW (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html, accessed on 15 July 2024).

4.3. Sequence and Structure Divergence Analyses

The online program mVISTA [46] (https://genome.lbl.gov/vista/index.shtml, accessed on 15 June 2024) was used to visualize the sequence divergence of *Vanda* with the plastome of *Holcoglossum quasipinifolium* (NC_041516) as a reference. We analyzed the expansion and contraction of the IR boundary using IRscope v3.1 [47]. Possible structure divergence (rearrangements and inversions) of plastomes were detected using the Mauve v1.1.3 [48] plugin in Geneious v9.05 [45]. Moreover, the IRa region was removed to avoid the overrepresentation of the inverted repeats. To further detect hypervariable regions within *Vanda*, the Pi values of complete plastomes and 68 CDSs of 13 *Vanda* species were estimated using DnaSP v6 [49]. The window length was set to 100 sites with the step size of 25 sites.

4.4. Repetitive Sequence and Codon Usage Analyses

The SSRs were detected using the online program MISA [50] (https://webblast.ipk-gatersleben.de/misa/, accessed on 17 June 2024), and the parameters for SSR motifs were set to 10, 5, 4, 3, 3, and 3 nucleotide repeats for mono-nucleotide, di-nucleotide, tri-nucleotide, tetra-nucleotide, penta-nucleotide, and hexa-nucleotide, respectively. Four long repeat types in 13 plastomes of *Vanda* were determined using the online program REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer, accessed on 17 June 2024), including F (forward), P (palindrome), R (reverse), and C (complement) repeats. The maximum and minimum repeat sizes of long repeats in the REPuter program were set to 50 bp and 20 bp, respectively, with the Hamming distance as 3. In addition, RSCU values and amino acid frequencies were calculated using MEGA v11 [51]. Protein-coding regions in the IRa region were also eliminated from the codon usage analyses to avoid overrepresentation.

4.5. Phylogenetic Analyses and Measurement of Divergence Variables

The phylogenetic relationships of *Vanda* were reconstructed based on the following five matrices: (1) the complete plastome sequences (excluding IRa), (2) concatenation of 68 CDSs, (3) LSC region, (4) hypervariable regions of the whole plastomes, and (5) hypervariable regions of 68 CDSs. We removed the IRa region from the complete plastome and CDSs to avoid the overrepresentation of the inverted repeats. Five matrices with 13 *Vanda* species and two outgroups were aligned using MAFFT v7 [52] and manually adjusted in BioEdit v7.0 [53]. After the selection of the best nucleotide substitution model for each matrix using jModelTest2 [54], respectively, ML analyses were carried out in RAxML v8.2.12 [55] with the best-fit model. Moreover, BI analyses were conducted in MrBayes v3.2.6 [56] for 10,000,000 generations and sampled every 1000 generations. After the assessment of the convergence in Tracer v1.7 [57], we obtained the majority rule consensus tree with the first

25% samples as "burn-in". Finally, the phylogenetic trees were plotted using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/Figtree/, accessed on 22 June 2024).

4.6. Character Reconstruction of Ndh Gene Status in Vanda

For all 13 *Vanda* species and two outgroups in our dataset, all 11 *ndh* genes were scored as lost and pseudogene. The character state reconstruction was conducted in Mesquite v3.51 (http://www.mesquiteproject.org, accessed on 12 August 2024) using the parsimony method. The best ML topology was selected as the input tree.

5. Conclusions

In the present study, we reported five *Vanda* plastomes (*V. ampullaceum*, *V. alpina*, *V. coerulea*, *V. cristata*, and *V. pumila*). The plastome characteristics and results of the comparative analyses indicate that the gene content and genomic structure of sampled *Vanda* plastomes were highly conserved, with the only difference was a large inversion from *trnG^{GCC}* to *ycf3* in the *V. coerulea* plastome. All *ndh* genes were found to be lost or pseudogenized. According to the ranking of Pi values, five non-CDSs (*clpP-psbB*, *rps15-ycf1*, *trnN^{GUU}-rpl32*, *trnS^{GCU}-trnG^{GCC}*, and *trnT^{UGU}-trnL^{UAA}*) and six CDSs (*ccsA*, *psbI*, *rpl22*, *rpoA*, *rps15*, and *ycf1*) were selected as candidate barcodes. The phylogenetic topologies based on complete plastome sequences (excluding one IR) showed the highest power in resolving the interspecific relationships of *Vanda*. Our results found that plastome can offer valuable insights into the phylogenetic relationships of recently diverged lineages.

References

1. Gardiner, L.M.; Cribb, P.C. *Renziana, Vol. 3: Vanda*; The Swiss Orchid Foundation: Basel, Switzerland, 2013.
2. Chase, M.W.; Cameron, K.M.; Freudenstein, J.V.; Pridgeon, A.M.; Salazar, G.; van den Berg, C.; Schuiteman, A. An updated classification of Orchidaceae. *Bot. J. Linn. Soc.* **2015**, *177*, 151–174. [CrossRef]
3. Gardiner, L.M.; Kocyan, A.; Motes, M.; Roberts, D.L.; Emerson, B.C. Molecular phylogenetics of *Vanda* and related genera (Orchidaceae). *Bot. J. Linn. Soc.* **2013**, *173*, 549–572. [CrossRef]
4. Motes, M. *Vandas: Their Botany, History, and Culture*; Timber Press: Portland, OR, USA, 1997.
5. Pridgeon, A.M.; Cribb, P.J.; Chase, M.W.; Rasmussen, F.N. *Genera Orchidacearum Vol. 6 Epidendroideae*; Oxford University Press: New York, NY, USA, 2014.
6. Brown, R. *Vanda Roxburghii, a Chequer-Flowered Vanda*; Botanical Register: Cambridgeshire, UK, 1820; Volume 6, p. 506.
7. Lindley, J. *Folia Orchidacea*; Matthews: London, UK, 1853.
8. Gardiner, L.M. New combinations in the genus *Vanda* (Orchidaceae). *Phytotaxa* **2012**, *61*, 47–54. [CrossRef]
9. Zou, L.H.; Huang, J.X.; Zhang, G.Q.; Liu, Z.J.; Zhuang, X.Y. A molecular phylogeny of Aeridinae (Orchidaceae: Epidendroideae) inferred from multiple nuclear and chloroplast regions. *Mol. Phylogenet. Evol.* **2015**, *85*, 247–254. [CrossRef] [PubMed]

10. Zhang, G.Q.; Liu, K.W.; Chen, L.J.; Xiao, X.J.; Zhai, J.W.; Li, L.Q.; Cai, J.; Hsiao, Y.Y.; Rao, W.H.; Huang, J.; et al. A new molecular phylogeny and a new genus, *Pendulorchis*, of the *Aerides-Vanda* alliance (Orchidaceae: Epidendroideae). *PLoS ONE* **2013**, *8*, e60097. [CrossRef]

11. Zou, L.H.; Wu, X.Y.; Lin, M.; Chen, L.J.; Liu, Z.J. *Vanda funingensis*, a new species of Orchidaceae (Epidedroideae; Vandeae; Aeridinae) from China: Evidence from morphology and DNA. *Phytotaxa* **2016**, *260*, 1–13. [CrossRef]

12. Zhang, G.J.; Hu, Y.; Huang, M.Z.; Huang, W.C.; Liu, D.K.; Zhang, D.Y.; Hu, H.H.; Downing, J.L.; Liu, Z.J.; Ma, H. Comprehensive phylogenetic analyses of Orchidaceae using nuclear genes and evolutionary insights into epiphytism. *J. Integr. Plant Biol.* **2023**, *65*, 1204–1225. [CrossRef]

13. 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*, 63. [CrossRef]

14. Zhao, J.H.; Zhou, P.; Li, X.Q.; Zhang, L.G.; Jin, X.H.; Xiang, X.G. Temporal and spatial pattern of *Holcoglossum* Schltr. (Orchidaceae), an East Asian endemic genus, based on nuclear and chloroplast genes. *Front. Ecol. Evol.* **2020**, *8*, 245. [CrossRef]

15. Li, E.Z.; Liu, K.J.; Deng, R.Y.; Gao, Y.W.; Liu, X.Y.; Dong, W.P.; Zhang, Z.X. Insights into the phylogeny and chloroplast genome evolution of *Eriocaulon* (Eriocaulaceae). *BMC Plant Biol.* **2023**, *23*, 32. [CrossRef]

16. Wang, J.; He, W.C.; Liao, X.Z.; Ma, J.; Gao, W.; Wang, H.Q.; Wu, D.L.; Tembrock, R.; Wu, Z.Q.; Gu, C.H. Phylogeny, molecular evolution, and dating of divergences in *Lagerstroemia* using plastome sequences. *Hortic. Plant J.* **2023**, *9*, 345–355. [CrossRef]

17. Chen, J.L.; Wang, F.; Zhou, C.Y.; Ahmad, S.; Zhou, Y.Z.; Li, M.H.; Liu, Z.J.; Peng, D.H. Comparative phylogenetic analysis for *Aerides* (Aeridinae, Orchidaceae) based on six complete plastid genomes. *Int. J. Mol. Sci.* **2023**, *24*, 12473. [CrossRef]

18. Zhou, C.Y.; Lin, W.J.; Li, R.; Wu, Y.; Liu, Z.J.; Li, M.H. Characterization of *Angraecum* (Angraecinae, Orchidaceae) plastomes and utility of sequence variability hotspots. *Int. J. Mol. Sci.* **2023**, *25*, 184. [CrossRef]

19. Martín, M.; Sabater, B. Plastid *ndh* genes in plant evolution. *Plant Physiol. Biochem.* **2010**, *48*, 636–645. [CrossRef]

20. Silva, S.R.; Michael, T.P.; Meer, E.J.; Pinheiro, D.G.; Varani, A.M.; Miranda, V.F.O. Comparative genomic analysis of *Genlisea* (corkscrew plants–Lentibulariaceae) chloroplast genomes reveals an increasing loss of the *ndh* genes. *PLoS ONE* **2018**, *13*, e0190321. [CrossRef]

21. 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]

22. Chen, J.L.; Wang, F.; Zhao, Z.; Li, M.H.; Liu, Z.J.; Peng, D.H. Complete chloroplast genomes and comparative analyses of three *Paraphalaenopsis* (Aeridinae, Orchidaceae) species. *Int. J. Mol. Sci.* **2023**, *24*, 11167. [CrossRef] [PubMed]

23. Zhou, C.Y.; Zeng, M.Y.; Gao, X.; Zhao, Z.; Li, R.; Wu, Y.; Liu, Z.J.; Zhang, D.; Li, M.H. Characteristics and comparative analysis of seven complete plastomes of *Trichoglottis s.l.* (Aeridinae, Orchidaceae). *Int. J. Mol. Sci.* **2023**, *24*, 14544. [CrossRef] [PubMed]

24. Dugas, D.V.; Hernandez, D.; Koenen, E.J.; Schwarz, E.; Straub, S.; Hughes, C.E.; Jansen, R.K.; Nageswara-Rao, M.; Staats, M.; Trujillo, J.T.; et al. Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions, and accelerated rate of evolution in *clpP*. *Sci. Rep.* **2015**, *5*, 16958. [CrossRef] [PubMed]

25. Xue, Q.Q.; Yang, J.P.; Yu, W.H.; Wang, H.M.; Hou, Z.Y.; Li, C.; Xue, Q.Y.; Liu, W.; Ding, X.Y.; Niu, Z.T. The climate changes promoted the chloroplast genomic evolution of *Dendrobium* orchids among multiple photosynthetic pathways. *BMC Plant Biol.* **2023**, *23*, 189. [CrossRef]

26. Zhao, Z.; Zeng, M.Y.; Wu, Y.W.; Li, J.W.; Zhou, Z.; Liu, Z.J.; Li, M.H. Characterization and comparative analysis of the complete plastomes of five *Epidendrum* (Epidendreae, Orchidaceae) species. *Int. J. Mol. Sci.* **2023**, *24*, 14437. [CrossRef] [PubMed]

27. Lin, C.S.; Chen, J.J.; Huang, Y.T.; Chan, M.T.; Daniell, H.; Chang, W.J.; Hsu, C.T.; Liao, D.C.; Wu, F.H.; Lin, S.Y.; et al. The location and translocation of *ndh* genes of chloroplast origin in the Orchidaceae family. *Sci. Rep.* **2015**, *5*, 9040. [CrossRef] [PubMed]

28. Niu, Z.T.; Pan, J.J.; Zhu, S.Y.; Li, L.D.; Xue, Q.Y.; Liu, W.; Ding, X.Y. Comparative analysis of the complete plastomes of *Apostasia wallichii* and *Neuwiedia singapureana* (Apostasioideae) reveals different evolutionary dynamics of IR/SSC boundary among photosynthetic orchids. *Front. Plant Sci.* **2017**, *8*, 1713. [CrossRef] [PubMed]

29. Knill, T.; Reichelt, M.; Paetz, C.; Gershenzon, J.; Binder, S. *Arabidopsis thaliana* encodes a bacterial-type heterodimeric isopropyl-malate isomerase involved in both Leu biosynthesis and the Met chain elongation pathway of glucosinolate formation. *Plant Mol. Biol.* **2009**, *71*, 227–239. [CrossRef]

30. Iriarte, A.; Lamolle, G.; Musto, H. Codon usage bias: An endless tale. *J. Mol. Evol.* **2021**, *89*, 589–593. [CrossRef]

31. Cavalier-Smith, T. Chloroplast evolution: Secondary symbiogenesis and multiple losses. *Curr. Biol.* **2002**, *12*, R62–R64. [CrossRef]

32. Cui, Y.F.; Zhou, P.; Xiang, K.L.; Zhang, Q.; Yan, H.; Zhang, L.G.; Pan, B.; Huang, Y.S.; Guo, Z.Y.; Li, Z.Y.; et al. Plastome evolution and phylogenomics of Trichosporeae (Gesneriaceae) with its morphological characters appraisal. *Front. Plant Sci.* **2023**, *14*, 1160535. [CrossRef]

33. Tao, K.F.; Tao, L.; Huang, J.L.; Duan, H.N.; Luo, Y.; Li, L. Complete chloroplast genome structural characterization of two *Aerides* (Orchidaceae) species with a focus on phylogenetic position of *Aerides flabellata*. *BMC Genom.* **2024**, *25*, 552. [CrossRef]

34. Abdullah; Mehmood, F.; Rahim, A.; Heidari, P.; Ahmed, I.; Poczai, P. Comparative plastome analysis of *Blumea*, with implications for genome evolution and phylogeny of Asteroideae. *Ecol. Evol.* **2021**, *11*, 7810–7826. [CrossRef]

35. 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*, 248. [CrossRef]

36. Li, L.; Wu, Q.P.; Fang, L.; Wu, K.L.; Li, M.Z.; Zeng, S.J. Comparative chloroplast genomics and phylogenetic analysis of *Thuniopsis* and closely related genera within Coelogyninae (Orchidaceae). *Front. Genet.* **2022**, *13*, 850201. [CrossRef]

37. Dong, W.P.; Sun, J.H.; Liu, Y.L.; Xu, C.; Wang, Y.H.; Suo, Z.L.; Zhou, S.L.; Zhang, Z.X.; Wen, J. Phylogenomic relationships and species identification of the olive genus *Olea* (Oleaceae). *J. Syst. Evol.* **2022**, *60*, 1263–1280. [CrossRef]

38. Kocyan, A.; Vogel, E.F.; Conti, E.; Gravendeel, B. Molecular phylogeny of *Aerides* (Orchidaceae) based on one nuclear and two plastid markers: A step forward in understanding the evolution of the Aeridinae. *Mol. Phylogenet. Evol.* **2008**, *48*, 422–443. [CrossRef] [PubMed]

39. Doyle, J.J.; Doyle, J.L. A rapid DNA isolation procedure for small amounts of fresh leaf tissue. *Phytochem. Bull.* **1987**, *19*, 11–15.

40. Brown, J.; Pirrung, M.; McCue, L.A. FQC Dashboard: Integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics* **2017**, *33*, 3137–3139. [CrossRef]

41. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]

42. Jin, J.J.; Yu, W.B.; Yang, J.B.; Song, Y.; dePamphilis, C.W.; Yi, T.S.; Li, D.Z. GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* **2020**, *21*, 241. [CrossRef]

43. Wick, R.R.; Schultz, M.B.; Zobel, J.; Holt, K.E. Bandage: Interactive visualization of de novo genome assemblies. *Bioinformatics* **2015**, *31*, 3350–3352. [CrossRef]

44. Qu, X.J.; Moore, M.J.; Li, D.Z.; Yi, T.S. PGA: A software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant. Methods* **2019**, *15*, 50. [CrossRef]

45. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [CrossRef]

46. Brudno, M.; Malde, S.; Poliakov, A.; Do, C.B.; Couronne, O.; Dubchak, I.; Batzoglou, S. Glocal alignment: Finding rearrangements during alignment. *Bioinformatics* **2003**, *19*, i54–i62. [CrossRef] [PubMed]

47. Amiryousefi, A.; Hyvönen, J.; Poczai, P. IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* **2018**, *34*, 3030–3031. [CrossRef]

48. Darling, A.E.; Mau, B.; Perna, N.T. Progressive Mauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* **2010**, *5*, e11147. [CrossRef] [PubMed]

49. Rozas, J.; Ferrer-Mata, A.; Sánchez-DelBarrio, J.C.; Guirao-Rico, S.; Librado, P.; Ramos-Onsins, S.E.; Sánchez-Gracia, A. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **2017**, *34*, 3299–3302. [CrossRef] [PubMed]

50. Beier, S.; Thiel, T.; Münch, T.; Scholz, U.; Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [CrossRef]

51. 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]

52. 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]

53. Hall, T.A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nuclc. Acids Symp. Ser.* **1999**, *41*, 95–98.

54. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. jModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods* **2012**, *9*, 772. [CrossRef]

55. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [CrossRef]

56. Ronquist, F.; Teslenko, M.; van der Mark, P.; Ayres, D.L.; Darling, A.; Höhna, S.; Larget, B.; Liu, L.; Suchard, M.A.; Huelsenbeck, J.P. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **2012**, *61*, 539–542. [CrossRef] [PubMed]

57. Rambaut, A.; Drummond, A.J.; Xie, D.; Baele, G.; Suchard, M.A. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **2018**, *67*, 901–904. [CrossRef]