
Zhuang Zhou 1,2,†, Jinliao Chen 1,*, Fei Wang 1, Xiaopei Wu 1, Zhongjian Liu 1, Donghui Peng 1, and Siren Lan 1, *

1 Key Laboratory of National Forestry and Grassland Administration for Orchid Conservation and Utilization at Landscape Architecture and Art, Fujian Agriculture and Forestry University, Fuzhou 350002, China; zhuangchow@126.com (Z.Z.); fjchenjl@126.com (J.C.)
2 Zhejiang Institute of Subtropical Crops, Zhejiang Academy of Agricultural Sciences, Wenzhou 325005, China
* Correspondence: fpdh@fafu.edu.cn (D.P.); lkzx@fafu.edu.cn (S.L.)
† These authors contributed equally to this work.

Abstract: *Taeniophyllum* is a distinct taxon of epiphytic leafless plants in the subtribe Aeridinae of Orchidaceae. The differences in chloroplast genomes between extremely degraded epiphytic leafless orchids and other leafy orchids, as well as their origins and evolution, raise intriguing questions. Therefore, we report the chloroplast genome sequence of *Taeniophyllum complanatum*, including an extensive comparative analysis with other types of leafless orchids. The chloroplast genome of *T. complanatum* exhibited a typical quadripartite structure, and its overall structure and gene content were relatively conserved. The entire chloroplast genome was 141,174 bp in length, making it the smallest known chloroplast genome of leafless epiphytic orchids. It encoded a total of 120 genes, including repetitive genes, comprising 74 protein-coding genes, 38 transfer RNA (tRNA) genes, and 8 ribosomal RNA (rRNA) genes. A phylogenetic analysis was conducted on the chloroplast genomes of 43 species belonging to five subfamilies of Orchidaceae. The results showed that the five subfamilies were monophyly, with nearly all segments having a 100% bootstrap value. *T. complanatum* and Chiloschista were clustered together as a sister group to *Phalaenopsis* and occupied the highest position in the Epidendroideae. Phylogenetic analysis suggested that *T. complanatum* and other leafless orchids within the Orchidaceae evolved independently. This study may provide the foundation for research on phylogenetic and structural diversity in leafless epiphytic orchids, thereby enhancing the resources available for chloroplast genome studies in Orchidaceae.

Keywords: chloroplast genome; *Taeniophyllum complanatum*; codon usage; SSR and repeat analysis; phylogenetic analysis; epiphytic leafless orchid

1. Introduction

Leaflessness is relatively common among seed plants. In some plants, such as Cactaceae, Euphorbiaceae, and Casuarinaceae, the leaves have degenerated into spines or membranous structures, and photosynthesis is carried out by the stem [1–4]. In other groups, like the Orobancheaceae, Balanophoraceae, Rafflesiaceae, *Cassylth*, and *Cuscuta*, plants obtain nutrients by parasitizing host plants through specialized structures [5–8]. *Monotropa uniflora* absorb nutrients from decaying plant materials through mycorrhizal associations [9,10]. These plants are typically terrestrial, parasitic, or saprophytic in life form and are relatively diverse in nature [10,11].

The Orchidaceae family comprises approximately 736 genera and 28,000 species [12,13]. There are some unique cases in this plant family, such as *Vanilla aphylla*, which photosynthesizes primarily through its stems, and genera like *Gastrodia*, *Epipogium*, *Platanthera*, and *Neottia*, which obtain nutrients by parasitizing or forming symbiotic relationships.
with fungi [9,11,14–17]. In the group of orchids called Aeridinae, there is a special group of leafless epiphytic orchids [18]. This plant grows epiphytically, and its life history is characterized by an extended absence of leaves or the complete absence of leaves [18]. Moreover, its shoots only come out during the flowering. These are even called “shootless” plants due to their lack of underground storage organs, like tubers or bulbs and specialized nutrient-absorbing structures [18,19]. There are three genera and 209 species in this group, including Chiloschista (20 species), Taeniophyllum (185 species), and the Aphyllae section of the genus Phalaenopsis (4 species) [20]. Compared to other leafless groups, these unique orchids occupy a relatively narrow ecological niche, mainly grown epiphytically on the trunks or branches of trees in tropical and subtropical forests from Asia to Australia [20]. However, understanding of the genetic makeup and evolution of these groups and other leafy orchids is still limited.

_T. complanatum_ is a rare and endangered epiphytic leafless orchid, distributing only in southern China and southern Japan [20]. In addition, little is known about its chloroplast genome and phylogenetic relationships. In this study, we presented the chloroplast genome sequence of _T. complanatum_ and examined its distinctions from the chloroplast genomes of other leafless orchids within the Orchidaceae family. Additionally, we examined its phylogenetic position within the Orchidaceae.

### 2. Materials and Methods

#### 2.1. Plant Materials, DNA Extraction and Sequencing

_T. complanatum_ was introduced and grown at Fujian Agriculture and Forestry University in Fujian Province, China. Voucher information for the specimens can be found in Table S1. Total DNA was extracted from fresh roots according to the manufacturer’s instructions and clustered using the HiSeq 4000 PE Cluster Kit (Illumina, Wuhan, China) [21,22]. The library preparations were sequenced using the Illumina HiSeq 4000 platform (Illumina, Wuhan, China) and 150 bp paired-end reads were produced. Illumina data were filtered by script in the cluster with the default parameter [22,23]. In this study, the paired reads were excluded from further analysis when the proportion of low-quality (Q ≤ 5) bases exceeded 50% of the read’s base number, and the percentage of ‘N’ bases in the reads exceeded 10% of the total bases [24,25]. A minimum of 10 gigabases (Gb) of high-quality sequencing data were obtained for each species.

#### 2.2. Chloroplast Genome Assembly and Annotation

The chloroplast genome assembly and annotation were conducted in accordance with the previously established procedures [26]. In summary, paired-end reads were assembled using the GetOrganelle pipeline (https://github.com/Kinggerm/GetOrganelle (accessed on 10 April 2024)), and then the filtered reads were assembled using SPAdes version 3.10 [27]. For the chloroplast genome assembly, the published sequence of the chloroplast genome of _Phalaenopsis wilsonii_ (MW218959) was used as a reference. Gene annotation was performed using DOGMA [28] with default parameters and validated through Geneious Prime v2021.1.1 [29]. Circular maps were generated using the online tool Chloroplot (https://irscope.shinyapps.io/chloroplot/ (accessed on 10 April 2024)) [30].

#### 2.3. Genome Comparison and Analysis, IR Border and Divergence Analyses

The chloroplast sequences from the ten orchid species of different life forms were aligned using the LAGAN alignment program within mVISTA (http://genome.lbl.gov/vista/mvista/submit.shtml (accessed on 10 April 2024)), with the sequence of _Apostasia odorata_ serving as the reference [31]. The chloroplast rearrangements were identified and visualized by employing Mauve [32]. To examine the boundaries between the inverted repeat regions (IRs), small single-copy (SSC), and large single-copy (LSC) regions of the chloroplasts, the online tool program IRscope was used (https://irscope.shinyapps.io/irapp (accessed on 10 April 2024)) [33].
For the identification of mutational hotspot regions and genes, the chloroplast sequences were aligned using MAFFT v7 [34]. Subsequently, the nucleotide diversity (Pi) for the four Aeridinae chloroplast genomes was calculated through DnaSP v6.12.03 (DNA Sequences Polymorphism) [35]. Highly mutable regions were pinpointed using a sliding window strategy with a 100 bp window length and a 25 bp step size.

2.4. Codon Usage Frequency and Amino Acid Abundance

A total of 68 protein-encoding genes sourced from the chloroplast genome of *T. complanatum* were isolated using PhyloSuite v1.2.2 [36]. Subsequently, these protein-coding genes were combined with PhyloSuite v1.2.2 and scrutinized for the relative synonymous codon usage (RSCU) in *T. complanatum* species, employing DAMBE [37]. Finally, the RStudio v0.1.2 software package “ggpubr” was employed to visualize the frequency of codon usage [38].

2.5. Repeat Sequence Analysis

The online tool REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer (accessed on 10 April 2024)) was utilized for the identification of repeat sequences, including forward, palindrome, reverse, and complement long repeats. The specified criteria for repeat size were set to a minimum of 20 bp and a maximum of 50 bp, with a hamming distance set of 3 [39]. To detect simple sequence repeats (SSRs), MISA-web (http://webblast.ipk-gatersleben.de/misa/ (accessed on 10 April 2024)) was employed with the following parameters: a threshold for mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide SSRs, along with minimal repeat numbers of 10, 5, 4, 3, 3, and 3, respectively [40].

2.6. Reconstruction of Phylogenetic Relationship

The phylogenetic analysis of 43 Orchidaceae species was conducted using complete chloroplast sequences, with two species from *Iris* as the outgroups. Among these 45 species, *T. complanatum* is a species we recently sequenced, while the remaining 44 species were sourced from publicly available complete chloroplast data on NCBI. A detailed list of the taxa analyzed, along with voucher information and GenBank accessions, can be found in Table S1. The chloroplast sequences were aligned using Geneious Prime v2021.1.1 [29], and a total of 68 protein-coding genes were aligned using PhyloSuite v1.2.2 [36]. The phylogenetic relationships were investigated through maximum likelihood (ML) analysis, which was conducted on the CIPRES Science Gateway website [41]. The GTRCAT model was applied to all datasets, and a total of 1000 repeated self-expanding analyses were carried out [42].

3. Results

3.1. Characteristics of the Chloroplast

The chloroplast genome of *T. complanatum* had a length of 141,174 bp and exhibited the typical quadripartite structure, consisting of a large single-copy region (LSC, 81,791 bp), a small single-copy region (SSC, 9766 bp), and a pair of inverted repeat regions (IR, 25,123 bp) (Figure 1). The overall GC (guanine-cytosine) content of the entire chloroplast genome was 36.3%. The chloroplast genome encoded a total of 120 genes, including duplicated genes, comprising 74 protein-coding genes, 38 transfer RNA (tRNA) genes, and 8 ribosomal RNA (rRNA) genes (Table 1). Most of the genes were present as single copies in either the LSC or SSC regions, with 19 genes duplicated in the IR regions. There were two genes that contained one intron each (rpl2, rpl16, rps16, trnA\[UGC\], trnC\[UCC\], trnF\[SAU\], trnL\[UAA\], trnV\[GAC\], trnK\[UGU\], atpF, petB, petD), while three genes had two introns each (ycf3, clpP, and rps12) (Table 2). In terms of their functions, 68 genes were involved in the self-replication process, and 36 genes played a role in photosynthesis. These genes encoded five genes for photosystem I, fourteen genes for photosystem II, two genes for NADH dehydrogenase subunits, six genes for the cytochrome b6/f complex, six genes for ATP
synthase subunits, and one gene for RUBISCO. Notably, functional ndh genes were found to be lost or pseudogenes in *T. complanatum*.

**Figure 1.** Chloroplast genome structure of *T. complanatum*.

**Table 1.** Characteristics of the complete chloroplast genomes of *T. complanatum* and three related species in Aeridinae.

<table>
<thead>
<tr>
<th>Species</th>
<th>Size (bp)</th>
<th>LSC (bp)</th>
<th>SSC (bp)</th>
<th>IRs (bp)</th>
<th>Number of Genes</th>
<th>Protein-Coding Genes</th>
<th>tRNA Genes</th>
<th>rRNA Genes</th>
<th>Total GC (%)</th>
<th>LSC GC (%)</th>
<th>SSC GC (%)</th>
<th>IR GC (%)</th>
<th>Number of ndh Gene Loss/Pseudogenization</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>T. complanatum</em></td>
<td>141,174</td>
<td>81,791</td>
<td>9766</td>
<td>25,123</td>
<td>120</td>
<td>74</td>
<td>38</td>
<td>8</td>
<td>36.3</td>
<td>33.6</td>
<td>26.9</td>
<td>42.2</td>
<td>9 (3)</td>
</tr>
<tr>
<td><em>Aerides rosea</em></td>
<td>148,110</td>
<td>85,373</td>
<td>11,033</td>
<td>25,852</td>
<td>120</td>
<td>74</td>
<td>38</td>
<td>8</td>
<td>36.7</td>
<td>34.0</td>
<td>28.3</td>
<td>43.1</td>
<td>7 (5)</td>
</tr>
<tr>
<td><em>Chiloschista yunnanensis</em></td>
<td>144,370</td>
<td>83,084</td>
<td>10,568</td>
<td>25,359</td>
<td>119</td>
<td>73</td>
<td>38</td>
<td>8</td>
<td>36.9</td>
<td>33.7</td>
<td>27.6</td>
<td>43.2</td>
<td>7 (5)</td>
</tr>
<tr>
<td><em>Phalaenopsis zelionii</em></td>
<td>145,408</td>
<td>85,039</td>
<td>10,473</td>
<td>24,948</td>
<td>120</td>
<td>74</td>
<td>38</td>
<td>8</td>
<td>36.8</td>
<td>34.1</td>
<td>28.2</td>
<td>43.3</td>
<td>6 (6)</td>
</tr>
</tbody>
</table>
Table 2. Gene content and its functions in *T. complanatum* chloroplast genome.

<table>
<thead>
<tr>
<th>Category</th>
<th>Group of Genes</th>
<th>Name of Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-replication</td>
<td>Proteins of large ribosomal subunit</td>
<td>rpl2(×2) a, rpl14, rpl16 a, rpl20, rpl22, rpl23(×2), rpl32, rpl33, rpl36</td>
</tr>
<tr>
<td></td>
<td>Proteins of small ribosomal subunit</td>
<td>rps2, rps3, rps4, rps7(×2), rps8, rps11, rps12(×2) b, rps14, rps15, rps16 a, rps18, rps19(×2)</td>
</tr>
<tr>
<td></td>
<td>Subunits of RNA polymerase</td>
<td>rpoA, rpoB, rpoC1 b, rpoC2</td>
</tr>
<tr>
<td>Ribosomal RNA genes</td>
<td></td>
<td>rrr4.5(×2), rrm5(×2), rrm6(×2), rrm23(×2)</td>
</tr>
<tr>
<td></td>
<td>Subunits of photosystem I</td>
<td>trnA-UGC(×2) a, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnG-UCC a, trnI-GUG(×2), trnL-CAU(×2), trnL-GAU(×2) a, trnL-UGC, trnL-UAA a, trnL-UAG, trnM-CAU, trnN-GLU(×2), trnP-UGG, trnQ-UGU, trnR-ACG(×2), trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGC, trnT-GGU, trnT-GGU, trnV-GAC(×2), trnV-UC(×2), trnW-CCA, trnY-GUA, trnF-M-CAU</td>
</tr>
<tr>
<td>Transfer RNA genes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genes for photosynthesis</td>
<td>Subunits of photosystem II</td>
<td>psaA, psaB, psaC, psaI, psaF</td>
</tr>
<tr>
<td></td>
<td>Subunits of NADH dehydrogenase</td>
<td>atpA, atpB, atpE, atpF a, atpH, atpl</td>
</tr>
<tr>
<td></td>
<td>Subunits of cytochrome b/f complex</td>
<td>ndhB(×2), ndhD</td>
</tr>
<tr>
<td></td>
<td>Subunits of ATP synthase</td>
<td>petA, petB a, petD a, petG, petL, petN</td>
</tr>
<tr>
<td></td>
<td>Large subunit of rubisco</td>
<td>rbcL</td>
</tr>
<tr>
<td>Other genes</td>
<td>Protofease</td>
<td>cpl2 b</td>
</tr>
<tr>
<td></td>
<td>Maturase</td>
<td>matK</td>
</tr>
<tr>
<td></td>
<td>Envelope membrane protein</td>
<td>ccmA</td>
</tr>
<tr>
<td></td>
<td>Translation initiation factor</td>
<td>infA</td>
</tr>
<tr>
<td></td>
<td>C-type cytochrome synthesis gene</td>
<td>ccsA</td>
</tr>
<tr>
<td></td>
<td>Subunit of acetyl-CoA-carboxylase</td>
<td>accD</td>
</tr>
<tr>
<td>Genes of unknown function</td>
<td>Conserved hypothetical chloroplast</td>
<td>ycf1, ycf(×2), ycf3 b, ycf4</td>
</tr>
</tbody>
</table>

Note: a Gene with one intron; b gene with two introns; (×2) gene with two copies.

3.2. Comparative Genome Analysis

The IRscope online tool was used to analyze the expansion and contraction of the IR regions in the chloroplast genomes of ten Orchidaceae plants, including *T. complanatum* and five leafless orchids in the vegetative phase, as well as four leafy orchids (Figure 2). The gene arrangement of these species’ chloroplast genomes was generally conserved, but there was considerable variation in the length of the IR regions, ranging from 24,948 to 30,320 bp, with some differences in expansion and contraction. These chloroplast genomes shared similar LSC/IR boundaries, with the rpl22 gene of the LSC extending into the IRb by a length of 17 to 66 bp at the junction of LSC and IRb (JLB). In the adjacent region of LSC and IRA (JLA), the positions of the rps19 and psbA genes were similar. However, these chloroplast genomes differed from each other at the junction of the IR/SSC regions. Except for *Vanilla aphylla*, which had ccsA at the IRb/SSC junction, *T. complanatum*, similarly to other leafless orchids, had the rpl32 gene near JSB, while *Phalaenopsis wilsonii* had the ndhF gene, *Apostasia odorata* had the ndhF and ycf1 genes, and *Paphiopedilum glanduliferum* had the trnL gene. At the junction of IRA and SSC (JSA), only *P. glanduliferum* had the ndhD gene, *Vanilla aphylla* had the ccsA and rpl32 genes, and the other eight Orchidaceae species had the ycf1 gene (Figure 2).

The gene order and structure of *T. complanatum* were compared using MAUVE against the chloroplast genomes of *Aerides rosea* and other orchids from different subfamilies (Figures 3 and S1). There was a high degree of similarity in gene content and order between the chloroplast genome of *A. rosea* and other members of the Aeridinae subtribe (Figure 3). Except for *Apostasia odorata* and *Gastrodia elata*, other Orchidaceae plants were found to have two and three segmental inversions and generally consistent gene sequences, with variations primarily in gene content (Figure S1).
LSC and IRA (JLA), the positions of the rps19 and psbA genes were similar. However, these chloroplast genomes differed from each other at the junction of the IR/SSC regions. Except for *Vanilla aphylla*, which had the ccsA gene at the IRb/SSC junction, *T. complanatum*, similarly to other leafless orchids, had the rpl32 gene near JSB, while *Phalaenopsis wilsonii* had the ndhF gene, *Apostasia odorata* had the ndhF and ycf1 genes, and *Paphiopedilum glanduliferum* had the trnL gene. At the junction of IRa and SSC (JSA), only *P. glanduliferum* had the ndhD gene, *Vanilla aphylla* had the ccsA and rpl32 genes, and the other eight Orchidaceae species had the ycf1 gene (Figure 2).

**Figure 2.** Comparison of the borders of large single-copy (LSC), small single-copy (SSC), and inverted repeat (IR) regions among ten chloroplast genomes.

**Figure 3.** Chloroplast genomes comparison of *T. complanatum* and three related species using a progressive MAUVE algorithm. Protein coding genes, rRNA genes, tRNA genes and intron containing tRNA genes are marked with block in white, red, black and green colors, respectively.
3.3. Repeated Analysis

A total of 57 SSRs were identified in *T. complanatum* (Figure 4A, Table S2). Mononucleotide repeats were the most abundant, with 41–49 nucleotide repeats. This was followed by eight tetranucleotide repeats, five dinucleotide repeats, two trinucleotide repeats, and one pentanucleotide repeat. Furthermore, there were no hexanucleotide repeats. Most of the SSRs were located in the LSC region (Figure 4B). The majority of single-nucleotide repeats were composed of A/T, while dinucleotide repeats were mostly AT/AT (Figure 4C).

![Figure 4. Analysis of simple sequence repeats (SSRs) and repeated sequences in the chloroplast genomes of *T. complanatum*.](image)

*Figure 4.* Analysis of simple sequence repeats (SSRs) and repeated sequences in the chloroplast genomes of *T. complanatum*. (A) Type and frequency of each identified SSR; (B) number of SSRs for *T. complanatum* by location in IR, LSC, and SSC; (C) number of different SSR units in the chloroplast genome of *T. complanatum*; (D) variation in repeat abundance and type; (E) distribution of repetitive regions; (F) distribution of repeat in IR, LSC, and SSC.

*T. complanatum* had 49 long repeat sequences in the chloroplast genome, which nearly fell within the range of 20–29 bp, and a few were found in the 30–39 bp range. Among these, there were 28 palindrome repeat sequences (P), accounting for 57.14%; 20 forward repeat sequences (F), representing 40.82%; and only 1 reverse repeat sequence. There was no complementary repeat sequence. These repeats were predominantly found in the LSC and IR regions, and none were detected in the SSC region (Figure 4D–F, Table S3).

3.4. Codon Usage Frequency and Aminoacid Abundance

In the chloroplast genome of *T. complanatum*, a total of 68 protein-coding genes were analyzed, excluding the lost and pseudogenized *ndh* genes. These genes were encoded by 19,269 codons (Figure 5A, Table S4). The codon usage patterns, as shown in Table 2, revealed a highly conserved codon usage bias (CUB). The results of relative synonymous codon usage (RSCU) analysis indicated that UUA, AGA, and GCU had the highest CUB values, with the average values of 1.9327, 1.9098, and 1.9007, respectively. The CGC, AGC, and GAC had the lowest CUB values, with average values of 0.3018, 0.3111, and 0.3191, respectively. Among the three stop codons, UAA showed the highest frequency, accounting for 48.52%. Furthermore, 30 codons exhibited strong bias (RSCU > 1), 32 codons had weak
bias (RSCU < 1), and the start codons AUG and UGG, encoding methionine and tryptophan, respectively, had no bias (RSCU = 1) (Figure 5A, Table S4). The most abundant amino acids used in the protein-coding genes were Leu, Ile, and Ser, accounting for 10.18%, 8.44%, and 7.61% of the total, respectively. Cys was the least frequent amino acid in the entire chloroplast genome, accounting for only 1.16% (Figure 5B, Table S4).

**Figure 5.** (A) Codon usage distribution of each amino acid in the *T. complanatum* chloroplast genome. The *y*-axis represents the codon usage frequency in percentage, whereas the *x*-axis represents the codons. (B) The amino acid distribution of *T. complanatum* chloroplast genome. The *x*-axis shows the abundance of each amino acid in percentage.
3.5. Sequence Divergence and Barcoding Investigation

To uncover genomic variations within the chloroplast genome of *T. complanatum*, an online tool mVISTA was utilized, wherein *Apostasia odorata* was used as a reference (Figure S2). The results from mVISTA revealed that, except for *T. complanatum*, most species’ chloroplast genomes were well conserved, exhibiting a high level of nucleotide variation across different Orchidaceae plants. To further examine these mutation hotspots within the chloroplast genomes of orchids with leaves, *T. complanatum*, *Aerides rosea*, *Phalaenopsis wilsonii*, and *Chiloschista yunnanensis* were compared for nucleotide diversity (Pi) using DNASP6 (Figure 6, Table S5). The nucleotide diversity (Pi) values for the four orchids from the subtribe Aeridinae ranged from 0 to 0.3333. At a threshold of Pi $\geq 0.2$, five mutation hotspots (trn$^{\text{UUC}}$-trn$^T$ GGU $> \text{trn}^S$ GCU-trn$^G$ UCC $> \text{trn}^L$ UAA $> \text{psaC-rps15} > \text{clpP-psbB}$) were selected (Figure 6).

![Figure 6](image)

Figure 6. Sliding window test of nucleotide diversity (Pi) in the *T. complanatum* and three related species chloroplast genomes. Window length: 100 bp; step size: 25 bp. X-axis: the position of the midpoint of each window. Y-axis: nucleotide diversity of each window.

3.6. Phylogenetic Reconstruction

We conducted a systematic phylogenetic analysis of 43 Orchidaceae whole-genome genes using Maximum Likelihood (ML) in the past tense, with two *Iris* species as our outgroup, and almost all branching nodes received strong support (BS $\geq 95$%) (Figure 7). The ML tree revealed that the five subfamilies of orchids were each monophyletic. *T. complanatum* and *Chiloschista* were sister taxa, positioned at the top branch of the Epidendroideae subfamily. Among the five subfamilies of Orchidaceae, neither the subfamily Orchidoidae nor the subfamily Vanilloideae had any leafless taxa reported. *Vanilla aphylla* showed a closer phylogenetic relationship with species in Orchidoidae and Vanilloideae, while it had a more distant relationship with epiphytic leafless taxa within the subfamily Vanilloideae. This suggests that different types of leafless orchids, in the evolution of orchids, have evolved independently multiple times.
Figure 7. Phylogenetic tree of *T. complanatum* and 42 other Orchidaceae species based on the complete chloroplast genome data, with two *Iris* species as outgroup. Numbers near the nodes are bootstrap percentages; (red) indicates leafless orchid.
4. Discussion

4.1. Chloroplast Genome Organization and Gene Content of T. complanatum and Comparison among Other Orchidaceae

The size of the chloroplast genome varies significantly among different life forms of orchids. Current studies have reported that the chloroplast genome size in the subtribe Aeridinae ranges from 142,859 bp (Schoenorchis seidenfadenii) [22] to 149,689 bp (Thrixspermum tsii) [43]. In this study, we obtained the chloroplast genome sequence of T. complanatum using next-generation sequencing technology. The genome size was 141,174 bp, making it the smallest chloroplast genome species within Aeridinae known to date. Its genome structure follows the typical quadripartite structure consistent with previously reported Aeridinae chloroplasts [22,25,44–47].

There are 11 ndh genes in the plant chloroplast genome [22]. The ndhA, F, and H genes were completely lost in the Aeridinae subtribe, and the remaining ndh genes became pseudogenes. In T. complanatum, only two ndh genes (ndhB and ndhD) were detected, which were pseudogenes. Consistent with the results from the subtribe Aeridinae chloroplast genomes, the other nine ndh genes (ndhA/C/E/F/G/H/I/J/K) were completely absent [25,44,46,47]. However, it is noteworthy that the number of ndh gene losses in T. complanatum was higher than in most other Aeridinae species, which could be one reason for its smaller chloroplast genome compared to the other known Aeridinae species. Previous studies have also found that loss and pseudogenization of ndh genes have been observed across all five subfamilies of the Orchidoideae [48]. Furthermore, epiphytic orchids experience a higher frequency of these events compared to terrestrial orchids [24,49]. It has been hypothesized that this may be related to epiphytic habitats [49]. T. complanatum is typically epiphytic, growing on tree trunks or rocks. Therefore, the results provide support to some extent for the close association between the loss or pseudogenization of ndh genes and epiphytic habitats.

In most angiosperms, plastid genomes have shown limited recombination in the past, exhibiting a highly conserved structure [38]. The majority of angiosperm plastids have a typical quadripartite structure consisting of a large single-copy region, a small single-copy region, and a pair of inverted repeat regions [50]. The inverted repeat regions play a role in the structural stability of plastids [51]. Previous studies have shown that contractions and expansions of the inverted repeat regions are common events in the evolutionary process [52], which is a major cause of variation in the plastid genome length [53]. For example, Gastrodia elata lost an inverted repeat region, which results in a plastid genome with only one copy [24]. These variations can occur at the borders of the inverted repeats (IRs) and single-copy regions (LSC and SSC), allowing certain genes to enter the IR or SC regions. This is also a major factor contributing to plastid genome length differences [53].

In this study, the boundary region between the inverted repeat (IR) and small single-copy (SSC) regions of the plastid genome in epiphytic orchids (including T. complanatum) was slightly different from that of other orchids. For example, similarly to other orchids, the boundary between IRb and SSC corresponds to the rpl32 gene in T. complanatum. However, it is outside ccsA in Vanilla aphylla, it is the trnN and rpl32 genes in Eulophia zollingeri and Cymbidium macrorhizon, it is the ndhF gene in Platanthera japonica, it is the ndhF and ycf1 genes in Apostasia odorata, and it is the trnL gene in Paphiopedilum glanduliferum. Our study suggests that, compared to other orchids, the size of the plastid genome and the IR region of T. complanatum are conserved (Figure 2). Therefore, the differences in the plastid genome size in T. complanatum may be due to the presence of indels in the intergenic regions and the absence of ndh genes [54,55].

4.2. Comparative Analysis of Sequence Comparison and Gene Order

We conducted MAUVE alignments to determine the unique structure and homology of chloroplast genomes among leafless orchid plants with different lifestyles. These genome-specific structural variations are highly informative and supportive, since they reveal evolutionary relationships [38]. We did not observe significant rearrangements in the chloroplast genomes of T. complanatum and other leafless orchid plants (Figure S1), and the
gene content and homology of the *T. complanatum* chloroplast genome remained the same as other subtribes (Figure 3).

Comparative genomics research provides important information for observing rearrangements or mutations (such as insertions or deletions at the genomic level) [56]. There was a higher frequency of non-coding-region variations in *T. complanatum*, similar to other Orchidaceae and epiphytic leafless orchid species (Figure S2) [25,47]. Therefore, coding regions are more conserved than non-coding regions. Moreover, there is a high level of nucleotide variation among different orchid plants, indicating that chloroplast genomes can serve as important markers for Orchidaceae phylogenetic studies.

### 4.3. SSRS and Repetitive Region Analysis

Microsatellites or Simple Sequence Repeats (SSRs) are defined as repetitive nucleotide units consisting of 1–6 repeats [38]. These repetitive sequences are scattered throughout the chloroplast genome and play a crucial role in the evolution of species and genetic variation within species [57]. Due to their high polymorphism, these repeat units have strong potential as molecular markers and have been widely used in studies of genetic diversity, population structure, and close species identification [58]. The dominance of A/T bases in microsatellites has been frequently reported, indicating a higher content of thymine (T) and adenine (A) repeats in the chloroplast genome (Table S2) [47,59]. Higher levels of A and T bases were detected in the repetitive regions of the *T. complanatum* chloroplast genome, mainly located in the intergenic regions of the LSC, and enriched in the non-coding regions. These results are consistent with most orchid species [25,47].

There was a high ratio of palindromic repeat sequences (P) and forward repeats in the chloroplast genome of *T. complanatum* (Figure 3). In the chloroplast genomes of the *Aerides*, *Phalaenopsis*, *Chiloschista*, *Trichoglottis*, and *Paraphalaenopsis*, within the subtribe Aeridinae (Supplementary Table S3) [25,44,46,47], the most common types of repeat regions were palindromic repeat sequences (P) and forward repeats (F), while the least common types were R and C repeat sequences [25,44–47]. This phenomenon was also observed in other life forms of leafless orchids, such as *Gastrodia elata*, *Eulophia zollingeri*, *Cymbidium macrorhizoz*, and *Vanilla aphylla* [24,60,61]. In general, chloroplast genomes contain significantly diverse SSRs and repeat regions, making them valuable for detecting polymorphism at the population, intraspecific, and interspecific levels (Supplementary Table S3) [38,62].

### 4.4. Codon Usage Frequency Analysis and Amino Acid Abundance

Codon usage preference signifies that each gene in each different species has its own preferred amino acid codon [58]. The relative synonymous codon usage (RSCU) is thought to be an important indicator for calculating the preference of synonymous codon usage. If the value of RSCU is higher than one, then this indicates preferential selection of that codon; if it is lower than one, then this suggests no preferential selection [25]. In the chloroplast of *T. complanatum*, 30 codons showed strong bias (RSCU > 1) and 32 codons exhibited weak bias (RSCU < 1). Comparatively, *Paphiopedilum glanduliferum* and *Vanilla aphylla* exhibited weak bias in 31 codons (RSCU < 1), while the least common types were K and C repeat sequences [25,44–47]. This phenomenon was also observed in other life forms of leafless orchids, such as *Gastrodia elata*, *Eulophia zollingeri*, *Cymbidium macrorhizoz*, and *Vanilla aphylla* [24,60,61]. In general, chloroplast genomes contain significantly diverse SSRs and repeat regions, making them valuable for detecting polymorphism at the population, intraspecific, and interspecific levels (Supplementary Table S3) [38,62].

The amino acid frequency for leucine (Leu) was the highest, while the frequency for cysteine (Cys) was the lowest in the encoded amino acids of the *T. complanatum* chloroplast. Our results matched well with previous studies on codon preferences in the Orchidaceae [25,63]. These findings also suggest a high similarity in codon usage frequency in orchid plants, further demonstrating the high conservation of the *T. complanatum* chloroplast genome (Table S4).

### 4.5. Divergent Hot Spot Analysis

Nucleotide diversity (Pi) is the amount of variation in nucleic acid sequences among different species, and variable regions can be used as molecular markers for population
Generally, in nucleotide diversity analysis for land plants, Pi values can vary depending on the number of species analyzed or the genomic region under study [38]. To further analyze the mutational hotspots of *T. complanatum* and three related species’ chloroplast genomes, DnaSP6 was used to analyze the nucleotide diversity (Pi) for the alignment of the complete genome. The Pi values for the four chloroplast genomes ranged between 0 and 0.3333. There were, primarily, five distinct hotspots with higher Pi values (Pi > 0.2): four were located in the LSC region, and one was found in the SSC region.

Previous research has identified some regions, such as *trnS*~GCU~–*trnG*~UCC~ and *trnL*~UAA~, as having the maximum nucleotide differences in *Trichoglottis* [25]. However, in *Aerides*, regions like *rpl20–rpl33*, *psbM*, *petB*, *rpoB–*trn*G*~GCA~, and *psbM–*trn*G*~UCC~, have been recognized to have significant differences [46]. Similarly, for the genus *Paraphalaenopsis*, regions such as *psbB*, *ccsA*, *trnK*~UUA~, *trnS*~GCU~–*trnG*~UCC~, and *psbM–*trn*G*~UCC~ are known for their maximum differences [47]. Furthermore, regions like *trnC*~GCA~–pet*N*, *trnD*~GUC~–*trnY*~GUA~, *trnS*~GCU~–*trnG*~UCC~, and *rpl32–*trn*G*~UAG~ are typically used as marker genes to detect the evolution relationships within Aeridinae [66]. These highly variable regions can serve as molecular markers for plant species identification, which may provide valuable insights into the phylogenetic relationships among *Taeniophyllum* and other Aeridinae species.

### 4.6. Phylogenetic Analysis

Analyzing the entire chloroplast genome could effectively address various questions in molecular evolution and the systematics of the same genus or family, thereby enhancing our understanding of molecular evolution [23]. The role of chloroplast genome data in reconstructing tribal, subtribal, and generic phylogenetic relationships in the Orchidaceae has also been established [22,24,61]. In order to gain insight into the evolutionary relationships among diverse groups of leafless orchids, we utilized chloroplast whole-genome sequences from 45 distinct species and constructed a phylogenetic tree based on maximum likelihood methods (Figure 7). According to the phylogenetic tree, the monophyly of five subfamilies in the Orchidaceae was strongly supported. Nearly all of the branches displayed 100% bootstrap values, indicating high support for these clades. This is the same as previous findings regarding Orchidaceae phylogenetics, based on different molecular data sources [24,61]. *T. complanatum* and *Chiloschista* are grouped together as sisters of *Phalaenopsis* at the tip of the tree, orchid subfamily branch, confirming that various types of leafless orchid plants have independently evolved multiple times.

### 5. Conclusions

As a rare and endangered epiphytic leafless orchid, *T. complanatum* has received widespread attention, yet research on its chloroplast genome information and phylogenetic relationships is still unclear. Our study indicated that, as a rare and endangered epiphytic leafless orchid, the overall structure and gene content of the chloroplast genome of *T. complanatum* were relatively conserved compared to other orchid plants. Only certain differences were found in terms of genome size, gene content, GC content, and repetitive sequences. All of the *ndh* genes were either lost or methylated in the chloroplasts of orchids that have leaves. This research provides a reference for further investigation into the DNA barcoding of species in the Aeridinae subtribe. Phylogenetic analysis based on available data confirmed that epiphytic leafless orchids such as *T. complanatum* had independently evolved, similar to other leafless orchid plants. It is possible to determine the phylogenetic relationships of most taxonomic units at the subtribal and higher subtribal levels within the Orchidaceae by using these findings.

**Supplementary Materials:** The following supporting information can be downloaded at [https://www.mdpi.com/article/10.3390/horticulturae10060660/s1](https://www.mdpi.com/article/10.3390/horticulturae10060660/s1). Table S1: Source and voucher information for this study. Voucher specimens were deposited in the herbariums of Forestry College of Fujian Agriculture and Forestry University (FFJC) and National Center for Biotechnology Information


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