

RESEARCH

Open Access



Complete chloroplast genome structural characterization of two *Aerides* (Orchidaceae) species with a focus on phylogenetic position of *Aerides flabellata*

Kaifeng Tao^{1†}, Lei Tao^{1†}, Jialin Huang², Hanning Duan¹, Yan Luo^{3*} and Lu Li^{1*}

Abstract

Background The disputed phylogenetic position of *Aerides flabellata* Rolfe ex Downie, due to morphological overlaps with related species, was investigated based on evidence of complete chloroplast (cp) genomes. The structural characterization of complete cp genomes of *A. flabellata* and *A. rosea* Lodd. ex Lindl. & Paxton were analyzed and compared with those of six related species in “*Vanda-Aerides* alliance” to provide genomic information on taxonomy and phylogeny.

Results The cp genomes of *A. flabellata* and *A. rosea* exhibited conserved quadripartite structures, 148,145 bp and 147,925 bp in length, with similar GC content (36.7 ~ 36.8%). Gene annotations revealed 110 single-copy genes, 18 duplicated in inverted regions, and ten with introns. Comparative analysis across related species confirmed stable sequence identity and higher variation in single-copy regions. However, there are notable differences in the IR regions between two *Aerides* Lour. species and the other six related species. The phylogenetic analysis based on CDS from complete cp genomes indicated that *Aerides* species except *A. flabellata* formed a monophyletic clade nested in the subtribe Aeridinae, being a sister group to *Renanthera* Lour., consistent with previous studies. Meanwhile, a separate clade consisted of *A. flabellata* and six *Vanda* R. Br. species was formed, as a sister taxon to *Holcoglossum* Schltr.

Conclusions This research was the first report on the complete cp genomes of *A. flabellata*. The results provided insights into understanding of plastome evolution and phylogenetic relationships of *Aerides*. The phylogenetic analysis based on complete cp genomes showed that *A. flabellata* should be placed in *Vanda* rather than in *Aerides*.

Keywords *Aerides flabellata*, *Aerides rosea*, Chloroplast genome, Orchidaceae, *Vanda-Aerides* alliance, Structural characterization

[†]Kaifeng Tao and Lei Tao contributed equally to this work and should be considered as co-first authors.

*Correspondence:

Yan Luo
luoyan@xtbg.org.cn
Lu Li
lilu@swfu.edu.cn

Full list of author information is available at the end of the article



Background

Aerides Lour. (Aeridinae, Vandaeae, Epidendroideae, Orchidaceae) consists of about 29 species, which are distributed from India to Papua New Guinea [1–3]. There are five species recorded in China, including one endemic species, which occurs in Southern China [4]. The distinct fragrance emitted by *Aerides* species has made them a valuable source for the production of numerous artificial hybrids and cultivars [5].

Aerides has been a focus of taxonomic disagreement within the subtribe Aeridinae [3, 5–7]. Since *Aerides* was first described, many members previously placed in other genera have been moved into it [7]. Conversely, dozens of species once included in *Aerides* have now been removed into other related genera [7]. The intrageneric taxonomy of *Aerides* were questioned due to the transfer of several species to other genera, such as *Ornithochilus* (Lindl.) Wall. ex Heynh., *Papilionanthe* Schltr., and *Seidenfadenia* Garay [8, 9]. *Aerides* was characterized by the presence of two cleft pollinia and divided into five groups based predominantly on pollinia morphology [10, 11]. However, two cleft pollinia were observed in other related genera, including *Brachypeza* Garay, *Phalaenopsis* Bl., *Rhynchostylis* Bl., *Vanda* R. Br. and among others [7]. Then, the concept of the “*Vanda-Aerides* alliance”, comprising *Aerides*, *Ascocentrum* Schltr., *Holcoglossum* Schltr., *Neofinetia* Hu, *Papilionanthe*, *Rhynchostylis* and *Vanda*, was

proposed [12], while the intergeneric delimitation has been controversial based on nuclear DNA data [3]. It is worth mentioning that the phylogenetic position of *Aerides flabellata* Rolfe ex Downie has been a focus issue [13, 14]. It was placed in *Aerides* based on an analysis using a plastid *matK* gene [15], but moved into *Vanda* in the latter treatment supported by an analysis of combined DNA datasets (nrITS and *matK*, *trnL*, *trnL-F*) [16].

The chloroplast (cp) genome has been increasingly utilized in taxonomy and phylogeny of Orchidaceae [17–19]. The complete cp genomes of six *Aerides* species (*Aerides crassifolia* C. S. P. Parish ex Burb., *Aerides falcata* Lindl. & Paxton, *Aerides lawrenceae* Rchb.f., *Aerides odorata* Lour., *Aerides quinquevulnera* Lindl., and *Aerides rosea* Lodd. ex Lindl. & Paxton) were published [20]. The results indicated that *Aerides* should be a separate clade within Aeridinae, sister to *Renanthera* Lour [20]. However, it should be noted that the complete cp genomic data of *A. flabellata* have not been reported. In this study, the structural and genomic information of the cp genomes of *A. flabellata* and *A. rosea* was characterized in detail and compared with those of six related species in the “*Vanda-Aerides* alliance”. The objectives of this study were: (1) to characterize and compare the complete cp genome structures of *A. flabellata* and *A. rosea* in detail, (2) to reconstruct the phylogenetic tree of Aeridinae to verify the position of *A. flabellata*, and (3) to

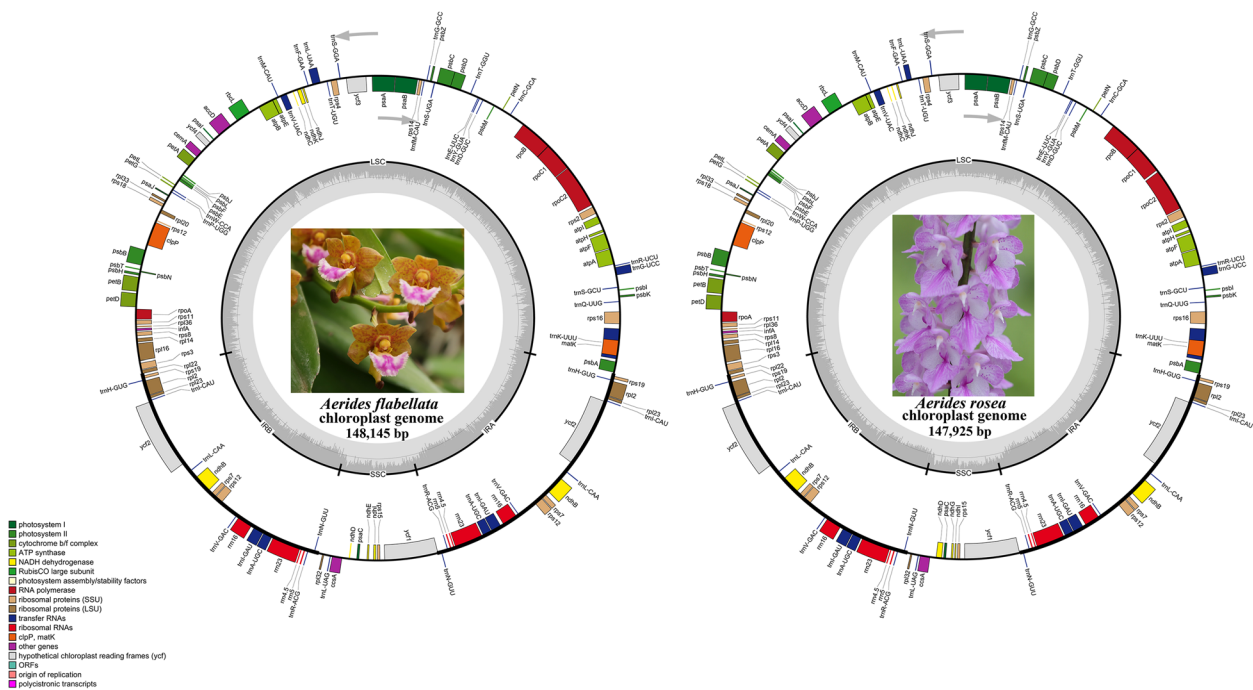


Fig. 1 The chloroplast genome maps of *Aerides flabellata* and *A. rosea*. Internal genes were clockwise transcribed, while external genes were counterclockwise transcribed. The inside circle bright and dark gray coloring indicated the genome guanine-cytosine (GC) content

Table 1 The general genome characteristics of the two *Aerides* species

Characteristics and Parameters	<i>Aerides flabellata</i>	<i>Aerides rosea</i>
Total cp genome size (bp)	148,145	147,925
LSC length (bp)	84,905	85,317
SSC length (bp)	11,636	11,018
IR length (bp)	25,802	25,795
Total GC content (%)	36.8	36.7
GC content for LSC (%)	34.1	34
GC content for SSC (%)	28.2	28.2
GC content for IR (%)	43.1	43.2
Total number of genes	128	128
CDS genes	80	79
rRNAs genes	8	8
tRNAs genes	38	38
Pseudogenes	2	3

provide new genomic data for a better understanding of the phylogeny of *Aerides*.

Results

General data on the chloroplast genome

The depth of the assemblies was 494.99 (*Aerides flabellata*) and 240.80 (*A. rosea*) (Fig.S1). The structures of cp genomes of the two *Aerides* species are highly similar. The total sizes of two cp genomes were 148,145 bp (*A. flabellata*) and 147,925 bp (*A. rosea*) (Fig. 1, Table 1). Same as most angiosperms, their cp genome displayed a typical quadripartite structure with a large single-copy (LSC) region (84,905 bp, 85,317 bp), a small single-copy (SSC) region (11,636 bp, 11,018 bp), and two inverted repeats (IR) regions (25,802 bp, 25,795 bp). The two cp genomes were all AT-rich, overall GC content ranged from 36.7~36.8%. The GC content in IR regions (43.1~43.2%) was higher than in LSC (34~34.1%) and SSC regions (28.82%) (Table 1). The GC content of the three codon positions of the two cp genomes was very similar. Furthermore, the third codon position was related to codon bias and mRNA stability. However, the third letter GC (36.28%) content was lower than the first (37.18%) and second (36.80%) letter GC content in *A. flabellata*. In contrast, the third letter GC content (36.53%) was lower than the second (37.18%) letter GC content,

Table 2 The GC content of the three positions of the two *Aerides* species

Species	1st letter GC	2nd letter GC	3rd letter GC
<i>Aerides flabellata</i>	37.18%	36.80%	36.28%
<i>A. rosea</i>	36.49%	37.23%	36.53%

but higher than the first letter GC (36.49%) content in *A. rosea* (Table 2). Both cp genomes contained 128 genes, including 2 (*A. flabellata*)~3 (*A. rosea*) pseudogenes, 79 (*A. rosea*)~80 (*A. flabellata*) CDS (coding sequences), eight rRNAs, and 38 tRNAs (Table 1). Among these, there were 110 unique genes in each cp genome. The LSC region contained 62 CDS genes and 21 tRNA genes in the two cp genomes. The SSC region comprised only one tRNA gene in the two cp genomes but eight CDS genes in *A. flabellata* and seven CDS genes in *A. rosea*. Six CDS genes (*rpl2*, *rpl23*, *rps7*, *rps12*, *rps19*, and *ycf2*), eight tRNA genes (*trnA-UGC*, *trnH-GUG*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, and *trnV-GAC*), and four rRNA genes (*rrn4.5*, *rrn5*, *rrn16*, and *rrn23*) were repeated in the IR regions (Table S1). There were ten genes with introns in the two cp genomes, seven genes with one intron (*rps16*, *rpoC1*, *rpl2*, *rpl16*, *petD*, *petB*, and *atpF*), and the other three genes with two introns (*clpP*, *ycf3*, *rps12*) (Table S2). However, the length of ten intron-containing genes were different in the two *Aerides* species (Table S2). Only one of the ten intron-containing genes were in the IR regions, while the other genes spread across the LSC region. In addition, *rps12* was a unique trans-splicing gene in which the first exon dispersed in the LSC region, but the second and third exons were in IR regions. Seven *ndh* (NA (D)H dehydrogenase) genes were identified in the cp genome of *A. flabellata* (*ndh B/C/D/E/I/J/K*) and *A. rosea* (*ndh B/C/D/G/I/J/K*) (Fig. 1, Table S1).

Repeat sequences analysis

The number of SSRs was analyzed to elucidate allied species or intra-species variations. There were 57 (*Aerides flabellata*) and 76 (*A. rosea*) SSRs detected in the two cp genomes, respectively consisting of 39 mononucleotides, seven dinucleotides, four trinucleotides, five tetranucleotides, one pentanucleotide and one hexanucleotide in *A. flabellata*, but of 52 mononucleotides, 12 dinucleotides, six trinucleotides, four tetranucleotides, two pentanucleotides in *A. rosea* (Table 3). Repeat units were composed mainly of A or T, and the mononucleotides were A/T type rather than G/C type in the two cp genomes. Furthermore, the C/G mononucleotide and AAAT/ATTT type tetranucleotide only existed in *A. flabellata* (Fig. S2).

Four different types of long repeats were also identified based on the complete genome sequence: complement (C), forward (F), palindromic (P), and reverse (R) (Table S3). Forty-nine large repeats were detected in the two cp genomes. In *A. flabellata*, almost all the repeats ranged from 20 to 39 bp, with the fewest in 40~49 bp. However, the number of long repeats above 40 bp in length was similar to the repeats from 20 to 39 bp in *A. rosea*. No complement repeats were detected above 40 bp

Table 3 The number of SSRs types distributed in different copy regions of the two *Aerides* species

SSRs types	<i>Aerides flabellata</i>	<i>Aerides rosea</i>
Mononucleotide	39	52
Dinucleotide	7	12
Trinucleotide	4	6
Tetranucleotide	5	4
Pentanucleotide	1	2
Hexanucleotide	1	0
Total	57	76

in length, and they were rare even in the smaller size ranges (Table S3).

Codon usage analysis

Based on coding sequences (CDS), codon usage frequency and relative synonymous codon usage (RSCU) were computed in the cp genomes of the two *Aerides* species and other six related species from “*Vanda-Aerides* alliance” (*Aerides falcata*, *A. lawrenceae*, *A. odorata*, *Vanda coerulea* Griff. ex Lindl., *V. coerulescens* Griff., and *V. subconcolor* Tang & F. T. Wang) downloaded from NCBI (<https://www.ncbi.nlm.nih.gov>) (Table S4) [21]. These CDS were composed of 48,830 to 49,803 codons, respectively, and encoded 20 amino acids in the eight cp genomes (Fig. S3, Table S4). The RSCU value of seven chloroplast genomes was similar, except *A. odorata*, which possessed the lower RSCU of leucine (Leu) and the higher RSCU of serine (Ser). Among them, leucine (Leu: 9.65~10.46%) was the amino acid that was utilized the most frequently, whereas tryptophan (Trp: 1.27~1.45%) was the least ubiquitous amino acid in the eight cp genomes (Table S5). According to the RSCU value, the eight cp genome could be divided into five groups: 28 codons (RSCU>1) and 33 codons (RSCU<1) in *A. odorata*; 29 codons (RSCU>1) and 31 codons (RSCU<1) in *A. falcata*; 30 codons (RSCU>1) and 32 codons (RSCU<1) in *A. flabellata* & *V. coerulea*; 31 codons (RSCU>1) and 30 codons (RSCU<1) in *A. lawrenceae* & *V. subconcolor*; 31 codons (RSCU>1) and 31 codons (RSCU<1) in *V. coerulescens*; 32 codons (RSCU>1) and 30 codons (RSCU<1) in *A. rosea* (Table S4). Almost all CDS in the eight species had the standard ATG start codon, but *rpl2* started with ATA/TAT. Among three stop codons, the TAA was the most common.

IR expansion and contraction

The cp genomes of the two *Aerides* species were highly conserved structurally, as well as those of the six species selected from “*Vanda-Aerides* alliance”. There were four boundaries (LSC/IRb, IRb/SSC, SSC/IRa, IRa/LSC) with structural variations (Fig. 2). The *rpl22* gene was

expanded from LSC to the IRb region. The *rpl32* gene was present in the SSC region in the eight species. The *trnN* gene was observed in the IRa and IRb region in the eight species. Notably, the *ycf1* gene was expanded from SSC to the IRa region in *A. flabellata* and three *Vanda* species, while it was only located in the SSC region in the other four *Aerides* species. In addition, the *ycf1* gene was also present in the IRb region of *V. coerulea* and *V. coerulescens*, and it expanded from IRb to the SSC region in *V. subconcolor*, but it is absent in *A. flabellata* and *A. rosea*.

Structural comparison and divergence hotspot identification analysis

Using *Aerides flabellata* as the reference, the cp genome sequences were compared by mVISTA (Fig. 3). The IR regions were more stable than the LSC and the SSC regions, and the rRNA genes were highly conserved. Meanwhile, the non-coding regions (CNS) were more diverse than the coding regions. The exons of *ycf1* and *ycf2* gene exhibited the highest polymorphism.

It was shown that the Pi value of LSC and SSC regions was greater than those of the IR regions based on the examination of CDS DNA polymorphism, demonstrating that the former were more varied than the latter. Three out of 62 CDS possessed the highest Pi values: *psbT* (0.01753), *ycf1* (0.01970) and *rps12* (0.03228) (Fig. 4A, Table S5). There were two locations with high Pi value (>0.05) for the IGS (intergenic spacer), including *psbB_psbT* (0.05291) and *psbE_petL* (0.08433) (Fig. 4B, Table S6). The Pi value of IGS locations (0.00~0.07, average 0.01965) was greater than that of CDS (0.00~0.024, average 0.00505) (Fig. 4, Table S5, S6).

Positive selection analysis

The Bayes Empirical Bayes (BEB) method identified 53 genes under positive selection, with *rpl22*, *rps4*, *rps8*, *rps14*, *rps16*, *rps18*, *rpl32*, *ycf1*, and *ycf2* genes having two or more significant positive selection sites. Other genes had just one substantial positive selection site aside. The number of positive selections of genes in LSC was higher than in SSC and IR regions (Table 4, Table S7).

Phylogenetic analysis

A Maximum-likelihood (ML) phylogenetic tree was reconstructed based on 62 single-copy CDS sequences of the two *Aerides* species and 45 representatives from *Aeridinae*, with six *Polystachya* species as outgroups, to shed a light on the phylogeny of *Aerides*, as well as the position of *A. flabellata* (Fig. 5, Table S8). *A. flabellata* and six *Vanda* species were formed as a stable clade with strong support (UFBoot: 100%), which was sister to *Holcoglossum* in the “*Vanda-Aerides* alliance”. It was shown that *A.*

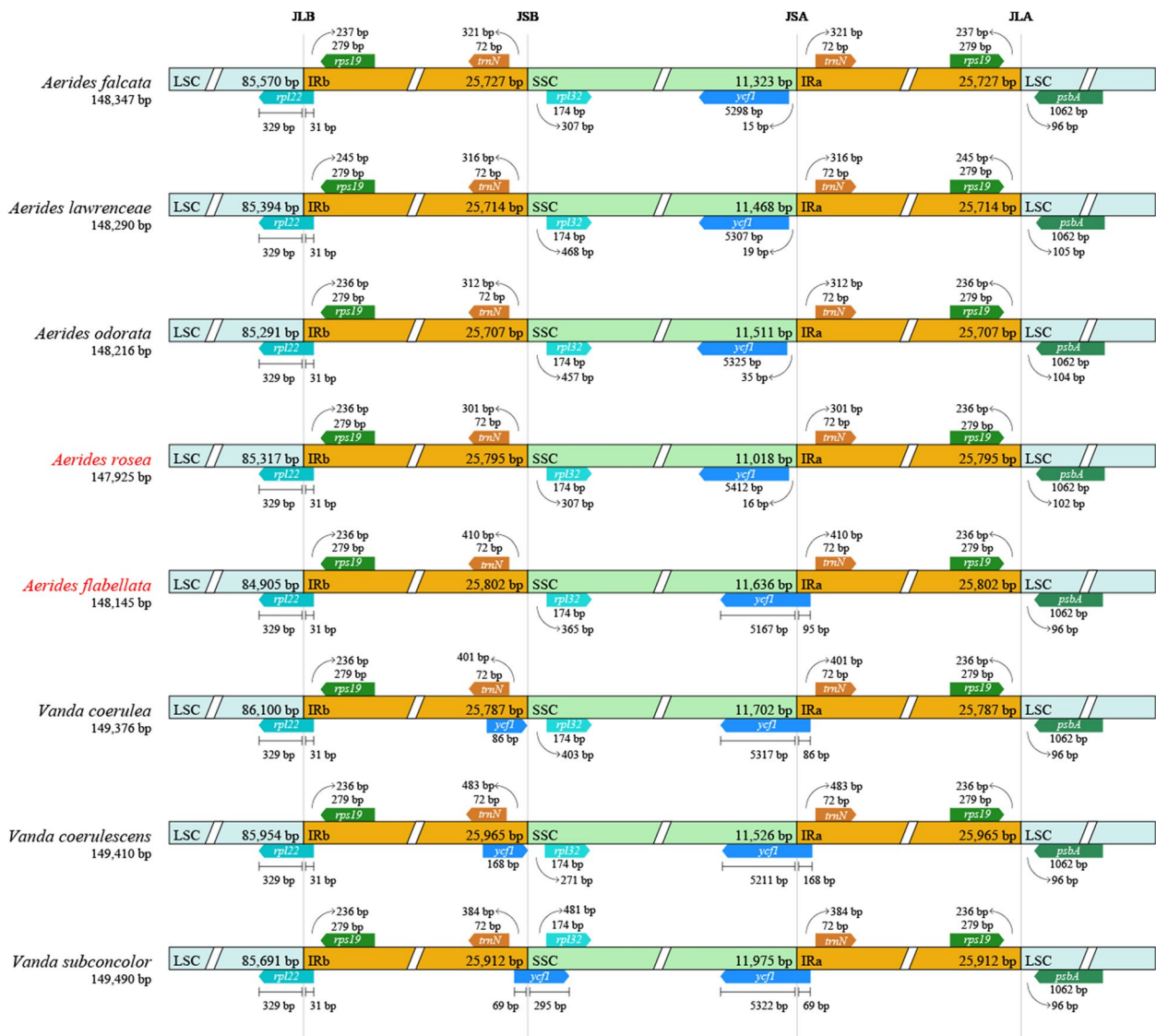


Fig. 2 Comparison of the boundaries of LSC, SSC and IR regions among chloroplast genomes of the two *Aerides* species and six species selected from “*Vanda-Aerides* alliance”. The arrow indicated the number of bp representing genes that were distant from a particular region of the cp genome. JLB (LSC/IRb), JSB (IRb/SSC), JSA (SSC/IRa), and JLA (IRa/LSC) denoted the junction sites between each corresponding two regions on the cp genome

flabellata should be placed in *Vanda*, which was sister to *V. coerulea* with strong support (UFBoot: 98%). Meanwhile, six *Aerides* species formed a monophyletic clade, with *A. rosea* as the sister taxon to the other five species. This monophyletic clade of *Aerides* was also found to be sister to *Renanthera*. All the branch nodes in the clade of *Aerides* were strongly supported by the ML analysis.

Discussion

In this study, the complete cp genomes of *Aerides flabellata* and *A. rosea* were sequenced and compared with those of other six related species within “*Vanda-Aerides*

alliance” to learn more about the cp genomic information and the molecular phylogeny of *Aerides*.

The cp genomes of *Aerides flabellata* and *A. rosea* were highly similar. Both cp genomes showed a typical quadripartite circular structure with the LSC and SSC regions partitioned by the IR regions, which were similar to the other orchids and most of the angiosperms with no significant differences [19, 22]. Notably, the genome size differed from previous research, with 79~80 CDS were annotated in these two cp genomes, as opposed to the 74 CDS reported previously [20]. The annotation of the *ndh* CDS caused this difference. *A. flabellata* and *A. rosea*

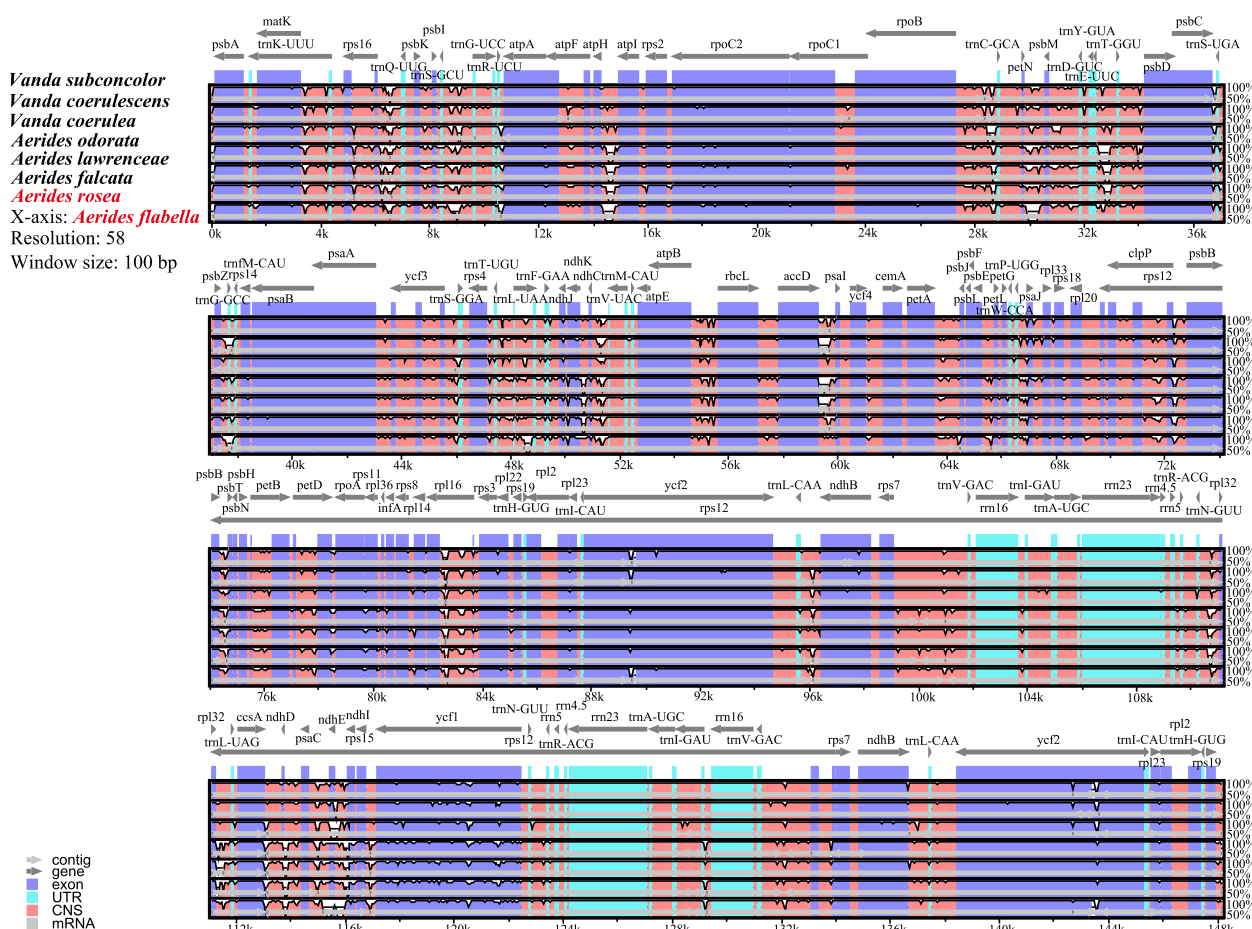


Fig. 3 Sequence alignment of chloroplast genomes of the two *Aerides* species and six species selected from “*Vanda-Aerides* alliance” using mVISTA. The vertical scale indicates the percentage of identity, ranging from 50 to 100%. The horizontal axis indicated the coordinates within the cp genome. Genome regions were color coded as exon, intron, and conserved non-coding sequences (CNS) and mRNA

contained seven *ndh* genes with five~six *ndh* CDS. In contrast, other *Aerides* species lacked some *ndh* genes or *ndh* CDS [20]. Eleven *ndh* genes in cp genomes encode the NAD(p)H dehydrogenase [23]. Previous research delineated Apostasioideae as *ndh*-complete, Vanilloideae as *ndh*-deleted, Cypripedioideae, Orchidoideae, and Epidendroideae as both *ndh*-complete and *ndh*-deleted. These findings suggested the presence of a complete functioning set of *ndh* genes in the common ancestor of orchids [24]. In certain photoautotrophic plants, the NDH complex is deemed unnecessary [24, 25]. Additionally, the GC content of the IR regions was much higher than that of the LSC and SSC regions, and these characteristics were also observed in *Cardamine* species [26]. This phenomenon is caused by the presence of rRNA and tRNA genes in the IR regions, which is the same as in other Orchidaceae cp genomes [18, 19].

Simple sequence repeats (SSRs), also known as microsatellites, represent shorter tandem repeats consisting

of 1~6 bp repeat units dispersed widely across the cp genome, and could be used for phylogenetic analysis [18, 27–29]. A total of 57 SSRs were identified in *Aerides flabellata*, while 76 were detected in *A. rosea*. Notably, the count of SSRs in *A. flabellata* diverged from recent research on *Aerides*, which reported a total of 71~77 SSRs [20]. Mononucleotide repeats emerged as the most prevalent SSRs within the cp genomes of both *A. flabellata* and *A. rosea*. Similar to six *Polystachya* species and three *Bulbophyllum* species, cp SSRs are predominantly comprised of short poly-A or poly-T repeats, and the mononucleotide repeats are the most commonly encountered forms [18, 30]. Repeated sequences play a pivotal role in species evolution, as well as in the inheritance and variation of genes within species [31, 32]. These repetitive sequences were widely used in the studies on genetic diversity, population structure, and the identification of closely related species [20, 33, 34]. In this study, 49 long repeats were identified from the two *Aerides* cp genomes,

Table 4 The positive selection analysis of two *Aerides* species and six species selected from “*Vanda-Aerides* alliance”

M8							
Region	Gene name	Positive sites	Pr($\omega > 1$)	Region	Gene name	Positive sites	Pr($\omega > 1$)
LSC	<i>atpA</i>	508	0.988*	LSC	<i>rpl22</i>	55 K	0.982*
LSC	<i>atpE</i>	135	1.000**	LSC		58 Y	0.982*
LSC	<i>atpH</i>	82	1.000**	LSC		87 I	0.982*
LSC	<i>atpI</i>	248	1.000**	LSC		88 V	0.982*
LSC	<i>cemA</i>	230	1.000**	LSC		110 N	0.982*
LSC	<i>infA</i>	78	1.000**	LSC		120	1.000**
LSC	<i>petA</i>	321	1.000**	LSC	<i>rps4</i>	166 I	0.965*
LSC	<i>petB</i>	216	1.000**	LSC		202	1.000**
LSC	<i>petD</i>	164	1.000**	LSC	<i>rps8</i>	3 R	0.967*
LSC	<i>petG</i>	38	1.000**	LSC		132	1.000**
LSC	<i>petL</i>	32	1.000**	LSC	<i>rps11</i>	139	1.000**
LSC	<i>petN</i>	30	1.000**	LSC	<i>rps14</i>	22 F	0.960*
LSC	<i>psaB</i>	735	0.991**	LSC		101	1.000**
LSC	<i>psaI</i>	37	1.000**	LSC	<i>rps16</i>	34 Q	0.977*
LSC	<i>psaJ</i>	45	1.000**	LSC		38 F	0.977*
LSC	<i>psbA</i>	354	1.000**	LSC		86 K	0.977*
LSC	<i>psbB</i>	509	1.000**	LSC	<i>rps18</i>	4 F	0.964*
LSC	<i>psbD</i>	354	1.000**	LSC		102	1.000**
LSC	<i>psbE</i>	84	1.000**	LSC	<i>ycf4</i>	185	1.000**
LSC	<i>psbF</i>	40	1.000**	SSC	<i>ccsA</i>	322	1.000**
LSC	<i>psbH</i>	74	1.000**	SSC	<i>psaC</i>	82	1.000**
LSC	<i>psbI</i>	37	1.000**	SSC	<i>rpl32</i>	52 K	0.957*
LSC	<i>psbJ</i>	41	1.000**	SSC		58	1.000**
LSC	<i>psbK</i>	62	1.000**	SSC	<i>ycf1</i>	829 E	0.971*
LSC	<i>psbL</i>	39	0.999**	SSC		1020 L	0.976*
LSC	<i>psbM</i>	35	1.000**	IR	<i>rpl2</i>	272	1.000**
LSC	<i>psbN</i>	44	1.000**	IR	<i>rpl23</i>	94	1.000**
LSC	<i>psbT</i>	36	1.000**	IR	<i>rps7</i>	156	1.000**
LSC	<i>psbZ</i>	63	1.000**	IR	<i>rps19</i>	93	1.000**
LSC	<i>rpl14</i>	123	1.000**	IR	<i>ycf2</i>	562 S	0.994**
LSC	<i>rpl16</i>	136	1.000**	IR		563 G	0.985*
LSC	<i>rpl20</i>	118	1.000**	IR		564 C	0.989*
LSC	<i>rpl33</i>	67	1.000**	IR		771 M	0.971*
LSC	<i>rpl36</i>	38	1.000**	IR		1562 D	0.972*
LSC	<i>rpoA</i>	338	1.000**	IR		1782 N	0.969*
LSC	<i>rps3</i>	219	1.000**				

* $p > 95\%$; ** $p > 99\%$

indicating that the *Aerides* cp genome retained abundant genetic information. The above findings can provide a data basis for further studies on population genetics.

The formation of codons is a critical process in translating genetic information from mRNA to protein [35], which is influenced by codon bias, particularly the third base usage pattern [36]. It has been empirically established that the GC composition exerts an influence on the utilization of codons and amino acids, and the GC

content of the third codon base (GC3) is deemed to most closely reflect codon usage trends [37]. Regarding *Aerides* species, the GC content observed in this study aligns with previous research [20]. Based on the RSCU analysis, six codons encoded arginine, leucine and serine. However, only one codon encoded methionine and tryptophan, which was also reported in other orchid species [19, 38].

The IR region is the most conservative section within the cp genome. However, its boundaries have

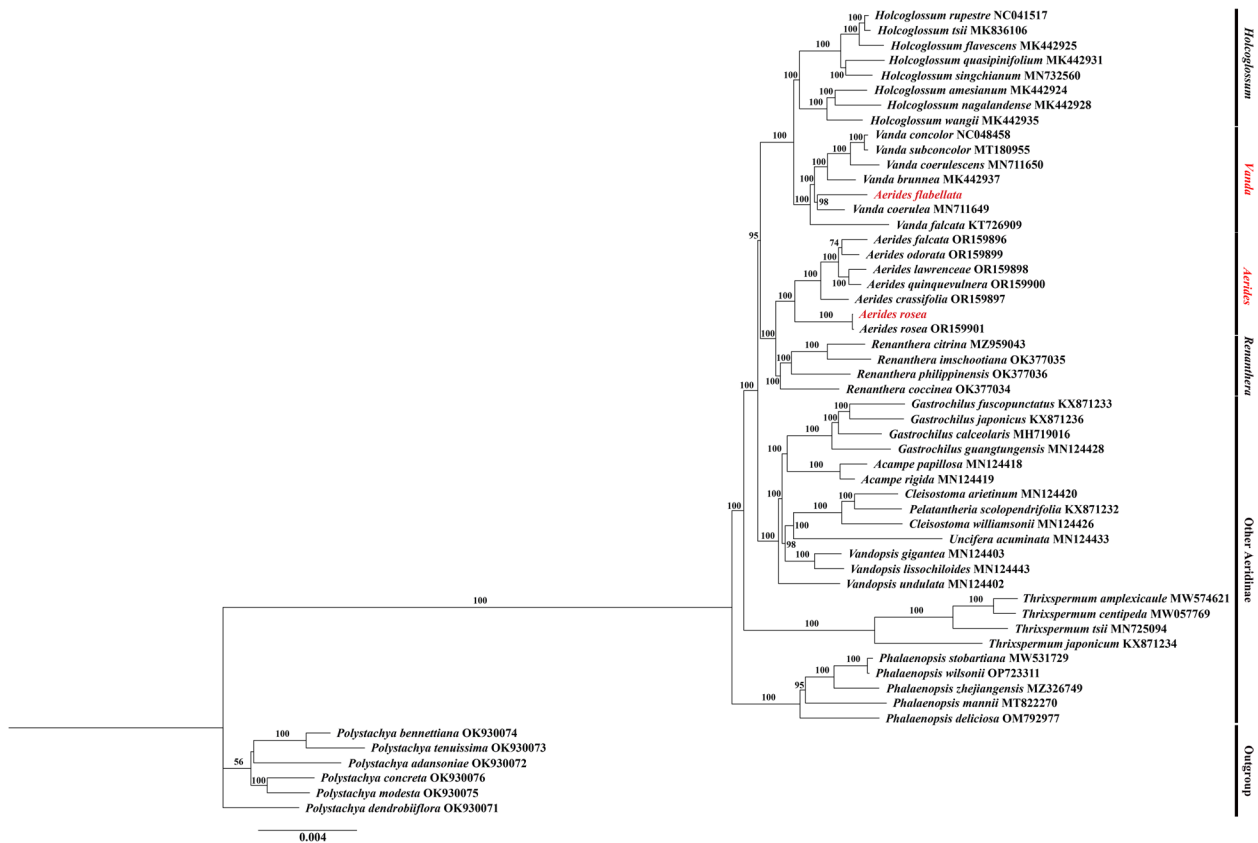


Fig. 5 Phylogenetic tree reconstructed of Aeridinae using Maximum-likelihood (ML) method based on 62 single-copy CDS sequences of 47 Aeridinae species, with six *Polystachya* species as outgroups

demonstrated frequent contractions and expansions, associated with the evolution of the cp genome, representing the primary driver for variations in cp genome length [39, 40]. Unlike basal angiosperms and eudicots, most monocots typically harbor *trnH-rps19* clusters in each IR region [41]. In this study, the *trnH-rps19* clusters were also located in each IR region, which was consistent with other five *Aerides* species [20], *Paphiopedilum henryanum* Braem [42], *Phalaenopsis stobartiana* Rchb.f., *P. wilsonii* Rolfe [19], and *Platanthera ussuriensis* (Regel) Maxim [17]. The presence of the *trnH-rps19* gene cluster in the IR of most monocots has been suggested as evidence of a duplication event predating the divergence of monocot lineages. Contractions and expansions in the IR borders have also been proposed to implicate taxonomic relationships among angiosperms [27, 41]. Additionally, *Aerides crassifolia*, *A. quinquevulnera*, *A. lawrenceae*, *A. odorata*, and *A. falcata* were consistent with *A. rosea* [20], wherein the *ycf1* gene was exclusively located in the SSC region. In contrast, the *ycf1* gene spanned the SSC and IRa regions in *A. flabellata*, aligning with observations in *Vanda subconcolor*.

Divergent regions, serving as valuable sources of data for DNA barcoding and phylogenetic research, were frequently employed as molecular markers in studies focused on phylogenetic reconstruction [43]. In this study, the nucleotide sequence of non-coding regions was more varied than the coding regions, which was generally consistent with other Orchidaceae cp genomes [18, 19]. Furthermore, the analysis of coding sequence regions revealed that the genes *rps12*, *psbT* and *ycf1* had significantly higher Pi values. Notably, *ycf1*, akin to *matK*, has been utilized as a DNA marker for phylogenetic studies [43]. In this research, *psbB-psbT* and *psbE-petL* also possessed the higher degree of variability. Simultaneously, sequences such as *trnS-trnG*, *psaC-ndhE*, *clpP-psbB*, and others exhibited the highest degree of variability in *Phalaenopsis* [19], while *rpl32-trnL*, *trnE-trnT*, and others showed the highest degree of variability in *Cymbidium* Sw. [44]. These indicated a diversity array of highly variable sequences in the Orchidaceae cp genome.

The utilization of the substitution rate ratio at synonymous and nonsynonymous sites (dN/dS, ω) has been pivotal in discerning adaptive signals among species and inferring evolutionary processes [45, 46]. Additionally, it

could suggest that environmental factors impacted the evolution of cp genomes, representing a primary cause for the divergence of numerous genes within the cp genome [47]. In this study, 53 genes were significantly identified under positive selection. Among them, the *atpH*, *petL*, and *rps4* genes have also been observed in other orchids [19, 48]. Furthermore, these genes could be used for orchid identification and phylogenetic research.

Aerides flabellata (synonym: *Vanda flabellata*) has been a focus of considerable taxonomic disagreement [6, 49]. Some taxonomists placed it within *Aerides* on account of features such as a long column foot and motile lip [10], while others assigned it to *Vanda*, emphasizing the species' short spur and broad lip [3, 5, 8, 50]. The species *Christensonia vietnamica* Haager, exhibiting morphological resemblances to both *Vanda* and *Rhynchostylis* [13], has been affiliated with *A. flabellata*, being described as 'almost a yellow *Aerides flabellata*' [13]. Therefore, *A. flabellata* and *C. vietnamica* were placed into *Vanda* based on combined DNA datasets (nrITS and *matK*, *trnL*, *trnL-F*) [3, 6, 15, 51].

The structural features of the cp genome have been utilized in constructing the phylogeny of Orchidaceae [17–19], because protein-coding regions and conserved sequences were informative for taxonomy [52]. In this study, based on CDS data from complete cp genomes, it was showed that *Aerides flabellata* was embedded within the clade of *Vanda*, while other six *Aerides* were grouped into a stable monophyletic clade. Therefore, it was supported that *A. flabellata* should be moved into *Vanda* from *Aerides* based on the comparative and the phylogenetic analyses.

Conclusion

The complete cp genomes of *Aerides flabellata* and *A. rosea* were sequenced and analyzed to unveil their genomic intricacies. This investigation encompassed a holistic exploration of various facets, including the general genome structure, codon usage, repeat sequences, boundaries within the inverted repeats, DNA polymorphism, and phylogenetic position. These cp genomic datasets were compared with the other six related species from the “*Vanda-Aerides* alliance”. It was confirmed that the cp genomic features of the “*Vanda-Aerides* alliance” was almost congruent and highly conserved, which could be used to understand the plastome evolution and evolutionary relationships of the “*Vanda-Aerides* alliance”. In addition, it was supported that *A. flabellata* should be removed into *Vanda* from *Aerides* based on cp genomic data.

Materials and methods

Ethical statement

No specific permits were required for the collection of specimens for this study. This research was carried out in compliance with the relevant laws of China.

Plant materials and chloroplast genome sequencing

Leaf samples of *Aerides flabellata* and *A. rosea* were cultivated and obtained from the Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Yunnan. The specimen was deposited in the Herbarium of Southwest Forestry University (HSFU, Lilu20180015, lilu@swfu.edu.cn). Genomic DNA of each sample was extracted from the silica gel-dried leaf tissues using the modified CTAB method with the TiangenDNA kit (TIANGEN, China) [53]. Paired-end libraries with an average insert size of approximately 400 bp were prepared using a TruSeq DNA Sample Prep Kit (Illumina, Inc., San Diego, CA, USA) according to the manufacturer's instructions. The libraries were sequenced on the Illumina HiSeq 2500 platform at Personalbio (two times 150 bp; Illumina, Shanghai, China). Raw data were filtered using Fastp v0.23.1 to obtain high-quality reads by the sliding window method to drop the low-quality bases of each read's head and tail [54].

Chloroplast genome assembly and annotation

The two complete cp genomes from the clean reads were assembled by the GetOrganelle version 1.7.7.0 [55] and annotated the new sequences using the Geneious Prime version 2020.0.4 [56]. The complete cp genomes sequences of *Aerides flabellata* and *A. rosea* were submitted to GenBank (Accession number: PP003956 and PP003955). The circular genome maps were drawn by the OGDRAW program (<https://chlorobox.mpimp-golm.mpg.de/OGDraw>) [44].

Sequence analysis and statistics

The repetitive structures, repeat sizes, and locations of forward match (F), reverse match (R), palindromic match (P), and complementary match (C) nucleotide repeat sequences were identified by REPuter v2.74 (<https://bibiserv.cebitec.uni-bielefeld.de/reputer/>) [57], with maximal repeat size set to 50 bp, minimal repeat size set to 20 bp, and hamming distance set to 3 [20]. By setting the minimum number of repeats to 10, 5, 4, 3, and 3 for mononucleotide (mono-), dinucleotide (din-), trinucleotide (tri-), tetranucleotide (tetra-), pentanucleotide (penta-), and hexanucleotide (hexan-), respectively, simple sequence repeats (SSR), a tract of repetitive DNA that typically ranges in length from 1 to 6 nucleotides, were detected via MISA (<https://webblast.ipk-gatersleben.de/misa/index.php?action=1>) [58, 59]. Condon usage was analyzed by MEGA11 software [60], and the relative synonymous codon usage (RSCU) and amino acid frequencies were calculated with default settings [61]. Finally, the RSCU figure was drawn by PhyloSuite version 1.2.2 [62, 63]. In addition, the GC content of the three position was analyzed by CUSP on EMBOSS program (<http://emboss.toulouse.inra.fr/cgi-bin/emboss/cusp>) [64].

Sequence divergence and genome comparison

The pairwise alignments and sequence divergence of *Aerides flabellata* and *A. rosea* with other six related species from “*Vanda-Aerides* alliance” (Table S9) were performed by the mVISTA with Shuffle-LAGAN mode (https://genome.lbl.gov/cgi-bin/VistaInput?num_seqs=2) [65]. Using an online application CPJSDraw v1.0.0 (<http://112.86.217.82:9929/#/tool/alltool/detail/335>), the contraction and extension of the IR borders between the four major areas (LSC/IRa/SSC/IRb) of the eight cp genome sequences were performed [66].

Positive selection analysis

The CDS sequences of *Aerides flabellata* and *A. rosea* with other six related species from “*Vanda-Aerides* alliance” (Table S9) were extracted by PhyloSuite version 1.2.2 [62, 63], and the single-copy CDS sequences were aligned by MAFFT version 7 [67]. The phylogenetic tree based on CDS was platformed by MEGA 11 with Neighbor-Joining (NJ) methods [60]. The non-synonymous (dN) and synonymous (dS) substitution rates were calculated by the CodeML algorithm implemented in EasyCodeML [68] and selected the M8 mode for selection suites to detect the protein-coding genes under selection in the two *Aerides* species and six related species.

Phylogenetic analysis

For phylogenetic analysis, the cp genomes of 53 species were selected (Table S9). The ingroup contains the genomes of 47 Aeridinae species, which 45 species were downloaded from the NCBI database. As Polystachyinae was sister to Aeridinae [18], six species from Polystachyinae were selected as outgroups. The single-CDS sequences (Table S8) from cp genomes were used for the phylogenetic analysis. These single-CDS sequences were extracted by PhyloSuite version 1.2.2 [62, 63], aligned by MAFFT version 7 [67], trimmed by Gblocks [69], and concatenated by plugins in PhyloSuite version 1.2.2 [62, 63]. The Maximum-Likelihood (ML) tree was performed in GTR+F+R2 mode based on CDS sequences by IQ-TREE 2 with 5000 ultrafast bootstrap (UFBoot) [70–72].

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-024-10458-0>.

Supplementary Material 1.
Supplementary Material 2.
Supplementary Material 3.
Supplementary Material 4.

Acknowledgements

We thank Dr. Fei Zhao for suggestions and for revising the article and Associate Professor Yuxiao Zhang for providing the computer server.

Authors' contributions

K.T. and L.T. collaborated on the analysis and writing of this manuscript. Y.L. provided the material. J.H. and H.D. collected the material. LL undertook the formal identification of the plant material. L.L. and Y.L. contributed to the design and editing of this manuscript. All authors reviewed and approved the final manuscript.

Funding

This study was supported by the National Nature Science Foundation of China (NSFC 32060049).

Availability of data and materials

The datasets generated or analyzed during the current study are available in the NCBI BioProject (PRJNA994440 and PRJNA995179, SRA: SRR25256624 and SRR25293872).

Declarations

Ethics approval and consent to participate

The study was conducted the plant material that complies with relevant institutional, national, and international guidelines and legislation. *Aerides flabellata* and *A. rosea* were cultivated in Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹College of Forestry, Southwest Forestry University, Kunming, Yunnan 650224, China. ²School of Chemistry, Biology and Environment, Yuxi Normal University, Yuxi, Yunnan 653100, China. ³Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences & Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, Yunnan, China.

Received: 3 January 2024 Accepted: 27 May 2024

Published online: 03 June 2024

References

- Chase MW, Cameron KM, Freudenstein JV, Pridgeon AM, Salazar G, van den Berg C, et al. An updated classification of Orchidaceae. *Bot J Linn Soc*. 2015;177:151–74.
- Dressler R. *Phylogeny and Classification of Orchid Family*. Cambridge: Cambridge University Press; 1993.
- Kocyan A, de Vogel EF, 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;48:422–43.
- Chen XQ, Wood JJ. *Aerides* Lour. In: *Flora of China: Orchidaceae*. Vol. 25. Beijing: Science Press; 2009. p. 485–6.
- Christenson EA. Nomenclatural Changes in the Orchidaceae Subtribe Sarcanthinae. *Selbyana*. 1986;9:167–70.
- Fan J, Qin H-N, Li D-Z, Jin X-H. Molecular phylogeny and biogeography of *Holcoglossum* (Orchidaceae: Aeridinae) based on nuclear ITS, and chloroplast *trnL-F* and *matK*. *Taxon*. 2009;58:849–61.
- Pridgeon AM, Cribb PJ, Chase MW, Rasmussen FN. *Genera Orchidacearum Volume 6: Epidendroideae (Part 3)*. Oxford, New York: Oxford University Press; 2014. p. 133–7.
- Garay LA. On the Systematics of the Monopodial Orchids I. *Bot Mus Leaf Harv Univ*. 1972;23:149–212.
- Garay LA. On the Systematics of the Monopodial Orchids II. *Bot Mus Leaf Harv Univ*. 1974;23:369–75.
- Seidenfaden G. *Orchid Genera in Thailand XIV: Fifty-nine Vandoid Genera*. Copenhagen: Council for Nordic Publications in Botany; 1988.
- Senghas K. 50. Subtribus: Aeridinae (‘Sarcanthinae’). In: *Die Orchideen*, 3rd edition, Vol. I/B. Berlin: Blackwell; 1996. p. 1131–422.

12. Christenson EA, Saito K, Tanaka R. In: Proceedings of the 12th World Orchid Conference 1987. In: The taxonomy of *Aerides* and related genera. 1st ed edition. Tokyo: 12th World Orchid Conference Organizing Committee; 1987. p. 35–40.
13. Christenson EA. Taxonomy of the Aeridinae with an infrageneric classification of *Vanda* Jones ex R. Br. In: Proceedings of the 14th World Orchid Conference. Edinburgh: HMSO Publications; 1994. p. 206–16.
14. Gardiner LM, Kocyan A, Motes M, Roberts DL, Emerson BC. Molecular phylogenetics of *Vanda* and related genera (Orchidaceae). *Bot J Linn Soc.* 2013;173:549–72.
15. Topik H, Yukawa T, Ito M. Molecular phylogenetics of subtribe Aeridinae (Orchidaceae): insights from plastid *matK* and nuclear ribosomal ITS sequences. *J Plant Res.* 2005;118:271–84.
16. 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;85:247–54.
17. Han C, Ding R, Zong X, Zhang L, Chen X, Qu B. Structural characterization of *Platanthera ussuriensis* chloroplast genome and comparative analyses with other species of Orchidaceae. *BMC Genomics.* 2022;23:84.
18. Jiang H, Tian J, Yang J, Dong X, Zhong Z, Mwachala G, et al. Comparative and phylogenetic analyses of six Kenya *Polystachya* (Orchidaceae) species based on the complete chloroplast genome sequences. *BMC Plant Biol.* 2022;22:177.
19. Tao L, Duan H, Tao K, Luo Y, Li Q, Li L. Complete chloroplast genome structural characterization of two *Phalaenopsis* (Orchidaceae) species and comparative analysis with their alliance. *BMC Genomics.* 2023;24:359.
20. Chen J, Wang F, Zhou C, Ahmad S, Zhou Y, Li M, et al. Comparative Phylogenetic Analysis for *Aerides* (Aeridinae, Orchidaceae) Based on Six Complete Plastid Genomes. *Int J Mol Sci.* 2023;24:12473.
21. National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – [cited 2024 Apr 09]. Available from: <https://www.ncbi.nlm.nih.gov/>.
22. Biju VC, P.R. S, Vijayan S, Rajan VS, Sasi A, Janardhanan A, et al. The Complete Chloroplast Genome of *Trichopus zeylanicus*, And Phylogenetic Analysis with *Dioscoreales*. *The Plant Genome.* 2019;12:190032.
23. Lin C-S, Chen JJW, Chiu C-C, Hsiao HCW, Yang C-J, Jin X-H, et al. Concomitant loss of NDH complex-related genes within chloroplast and nuclear genomes in some orchids. *Plant J.* 2017;90:994–1006.
24. Lin C-S, Chen JJW, Huang Y-T, Chan M-T, Daniell H, Chang W-J, et al. The location and translocation of *ndh* genes of chloroplast origin in the Orchidaceae family. *Sci Rep.* 2015;5:9040.
25. Liu D-K, Tu X-D, Zhao Z, Zeng M-Y, Zhang S, Ma L, et al. Plastid phylogenomic data yield new and robust insights into the phylogeny of *Cleisostoma-Gastrochilus* clades (Orchidaceae, Aeridinae). *Mol Phylogenet Evol.* 2020;145: 106729.
26. Hu S, Sablok G, Wang B, Qu D, Barbaro E, Viola R, et al. Plastome organization and evolution of chloroplast genes in *Cardamine* species adapted to contrasting habitats. *BMC Genomics.* 2015;16:306.
27. Agrama HA, Tuinstra MR. Phylogenetic diversity and relationship sorghum accessions using SSRs and RAPDs. *Afr J Biotech.* 2003;2:334–40.
28. Li X, Zhao Y, Tu X, Li C, Zhu Y, Zhong H, et al. Comparative analysis of plastomes in Oxalidaceae: Phylogenetic relationships and potential molecular markers. *Plant Diversity.* 2021;43:281–91.
29. Madhumati B. Potential and application of molecular markers techniques for plant genome analysis. *International Journal of Pure & Applied Bioscience.* 2014;2:169–88.
30. Yang J, Zhu Z, Fan Y, Zhu F, Chen Y, Niu Z, et al. Comparative plastomic analysis of three *Bulbophyllum* medicinal plants and its significance in species identification. *Acta Pharmaceutica Sinica.* 2020;55:2736–45.
31. Chen Y, Hu N, Wu H. Analyzing and Characterizing the Chloroplast Genome of *Salix wilsonii*. *Biomed Res Int.* 2019;2019:5190425.
32. Khan A, Asaf S, Khan AL, Al-Harrasi A, Al-Sudairy O, AbdulKareem NM, et al. First complete chloroplast genomics and comparative phylogenetic analysis of *Commiphora gileadensis* and *C. foliacea*: Myrrh producing trees. *PLOS ONE.* 2019;14:e0208511.
33. Singh RB, Mahenderakar MD, Jugran AK, Singh RK, Srivastava RK. Assessing genetic diversity and population structure of sugarcane cultivars, progenitor species and genera using microsatellite (SSR) markers. *Gene.* 2020;753: 144800.
34. Yu J, Dossa K, Wang L, Zhang Y, Wei X, Liao B, et al. PMDBase: a database for studying microsatellite DNA and marker development in plants. *Nucleic Acids Res.* 2017;45:D1046–53.
35. Qiu S, Zeng K, Slotte T, Wright S, Charlesworth D. Reduced Efficacy of Natural Selection on Codon Usage Bias in Selfing *Arabidopsis* and *Capsella* Species. *Genome Biol Evol.* 2011;3:868–80.
36. Shang M, Liu F, Hua J, Wang K. Analysis on codon usage of chloroplast genome of *Gossypium hirsutum*. *Scientia Agricultura Sinica.* 2011;44:245–53.
37. Chen L, Liu T, Yang D, Nong X, Xie Y, Fu Y, et al. Analysis of codon usage patterns in *Taenia pisiformis* through annotated transcriptome data. *Biochem Biophys Res Commun.* 2013;430:1344–8.
38. Alzahrani DA, Yaradua SS, Albokhari EJ, Abba A. Complete chloroplast genome sequence of *Barleria prionitis*, comparative chloroplast genomics and phylogenetic relationships among Acanthoideae. *BMC Genomics.* 2020;21:393.
39. Dugas DV, Hernandez D, Koenen EJM, Schwarz E, Straub S, Hughes CE, et al. Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions and accelerated rate of evolution in *clpP*. *Sci Rep.* 2015;5:16958.
40. Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boore JL, et al. Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics.* 2007;8:174.
41. Wang R-J, Cheng C-L, Chang C-C, Wu C-L, Su T-M, Chaw S-M. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evol Biol.* 2008;8:36.
42. Liu H, Ye H, Zhang N, Ma J, Wang J, Hu G, et al. Comparative Analyses of Chloroplast Genomes Provide Comprehensive Insights into the Adaptive Evolution of *Paphiopedilum* (Orchidaceae). *Horticulturae.* 2022;8:391.
43. Menezes APA, Resende-Moreira LC, Buzatti RSO, Nazareno AG, Carlsen M, Lobo FP, et al. Chloroplast genomes of *Byrsonima* species (Malpighiaceae): comparative analysis and screening of high divergence sequences. *Sci Rep.* 2018;8:2210.
44. Shaw J, Shafer HL, Leonard OR, Kovach MJ, Schorr M, Morris AB. Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: The tortoise and the hare IV. *Am J Bot.* 2014;101:1987–2004.
45. Kryazhimskiy S, Plotkin JB. The population genetics of dN/dS. *PLoS Genet.* 2008;4: e1000304.
46. Williams MJ, Zapata L, Werner B, Barnes CP, Sottoriva A, Graham TA. Measuring the distribution of fitness effects in somatic evolution by combining clonal dynamics with dN/dS ratios. *Elife.* 2020;9: e48714.
47. Zuo L-H, Shang A-Q, Zhang S, Yu X-Y, Ren Y-C, Yang M-S, et al. The first complete chloroplast genome sequences of *Ulmus* species by *de novo* sequencing: Genome comparative and taxonomic position analysis. *PLoS ONE.* 2017;12: e0171264.
48. Tang H, Tang L, Shao S, Peng Y, Li L, Luo Y. Chloroplast genomic diversity in *Bulbophyllum* section *Macrocaulia* (Orchidaceae, Epidendroideae, Malvaceae): Insights into species divergence and adaptive evolution. *Plant Divers.* 2021;43:350–61.
49. Zhang G-Q, Liu K-W, Chen L-J, Xiao X-J, Zhai J-W, Li L-Q, et al. A New Molecular Phylogeny and a New Genus, *Pendulorchis*, of the *Aerides-Vanda* Alliance (Orchidaceae: Epidendroideae). *PLoS ONE.* 2013;8: e60097.
50. Motes MR. *Vandas: their botany, history, and culture.* Portland, Or: Timber Press; 1997.
51. Carlswald BS, Whitten WM, Williams NH, Bytebier B. Molecular phylogenetics of Vandaeae (Orchidaceae) and the evolution of leaflessness. *Am J Bot.* 2006;93:770–86.
52. Bobik K, Burch-Smith TM. Chloroplast signaling within, between and beyond cells. *Front Plant Sci.* 2015;6:781.
53. Healey A, Furtado A, Cooper T, Henry RJ. Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods.* 2014;10:21.
54. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018;34:i884–90.
55. Jin J-J, Yu W-B, Yang J-B, Song Y, dePamphilis CW, Yi T-S, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* 2020;21:241.

56. Kearsse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012;28:1647–9.
57. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res*. 2001;29:4633–42.
58. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. *Bioinformatics*. 2017;33:2583–5.
59. Thiel T, Michalek W, Varshney R, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet*. 2003;106:411–22.
60. 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.
61. Byliah S, Shedole S, Suresh KP, Gowda L, Patil SS, Indrabalan UB. Analysis of Codon Usage Bias in *Cya*, *Lef*, and *Pag* Genes Exists in px01 Plasmid of *Bacillus Anthracis*. In: Fong S, Dey N, Joshi A, editors. *ICT Analysis and Applications*. Singapore: Springer Nature; 2022. p. 1–9.
62. Xiang C-Y, Gao F, Jakovlić I, Lei H-P, Hu Y, Zhang H, et al. Using PhyloSuite for molecular phylogeny and tree-based analyses. *iMeta*. 2023;2:e87.
63. Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, et al. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol Ecol Resour*. 2020;20:348–55.
64. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet*. 2000;16:276–7.
65. Brudno M, Malde S, Poliakov A, Do CB, Couronne O, Dubchak I, et al. Global alignment: finding rearrangements during alignment. *Bioinformatics*. 2003;19(Suppl 1):i54–62.
66. Li H, Guo Q, Xu L, Gao H, Liu L, Zhou X. CPJSDraw: analysis and visualization of junction sites of chloroplast genomes. *PeerJ*. 2023;11: e15326.
67. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol*. 2013;30:772–80.
68. Gao F, Chen C, Arab DA, Du Z, He Y, Ho SYW. EasyCodeML: A visual tool for analysis of selection using CodeML. *Ecol Evol*. 2019;9:3891–8.
69. Talavera G, Castresana J. Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. *Syst Biol*. 2007;56:564–77.
70. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol*. 2018;35:518–22.
71. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14:587–9.
72. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol*. 2020;37:1530–4.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.