

RESEARCH

Open Access



Comparative and phylogenetic analysis of *Chiloschista* (Orchidaceae) species and DNA barcoding investigation based on plastid genomes

Ding-Kun Liu^{1,2†}, Cheng-Yuan Zhou^{2†}, Xiong-De Tu^{1,2}, Zhuang Zhao², Jin-Liao Chen², Xu-Yong Gao², Shao-Wei Xu², Meng-Yao Zeng², Liang Ma³, Sagheer Ahmad², Ming-He Li^{1,2}, Siren Lan^{1,2} and Zhong-Jian Liu^{1,2*}

Abstract

Background *Chiloschista* (Orchidaceae, Aeridinae) is an epiphytic leafless orchid that is mainly distributed in tropical or subtropical forest canopies. This rare and threatened orchid lacks molecular resources for phylogenetic and barcoding analysis. Therefore, we sequenced and assembled seven complete plastomes of *Chiloschista* to analyse the plastome characteristics and phylogenetic relationships and conduct a barcoding investigation.

Results We are the first to publish seven *Chiloschista* plastomes, which possessed the typical quadripartite structure and ranged from 143,233 bp to 145,463 bp in size. The plastomes all contained 120 genes, consisting of 74 protein-coding genes, 38 tRNA genes and eight rRNA genes. The *ndh* genes were pseudogenes or lost in the genus, and the genes *petG* and *psbF* were under positive selection. The seven *Chiloschista* plastomes displayed stable plastome structures with no large inversions or rearrangements. A total of 14 small inversions (SIs) were identified in the seven *Chiloschista* plastomes but were all similar within the genus. Six noncoding mutational hotspots (*trnN^{GUU}-rpl32* > *rpoB-trnC^{GCA}* > *psbK-psbI* > *psaC-rps15* > *trnE^{UUC}-trnT^{GGU}* > *accD-psaI*) and five coding sequences (*ycf1* > *rps15* > *matK* > *psbK* > *ccsA*) were selected as potential barcodes based on nucleotide diversity and species discrimination analysis, which suggested that the potential barcode *ycf1* was most suitable for species discrimination. A total of 47–56 SSRs and 11–14 long repeats (> 20 bp) were identified in *Chiloschista* plastomes, and they were mostly located in the large single copy intergenic region. Phylogenetic analysis indicated that *Chiloschista* was monophyletic. It was clustered with *Phalaenopsis* and formed the basic clade of the subtribe Aeridinae with a moderate support value. The results also showed that seven *Chiloschista* species were divided into three major clades with full support.

[†]Ding-Kun Liu and Cheng-Yuan Zhou contributed equally to this work.

*Correspondence:
Zhong-Jian Liu
zjliu@fafu.edu.cn

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



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Conclusion This study was the first to analyse the plastome characteristics of the genus *Chiloschista* in Orchidaceae, and the results showed that *Chiloschista* plastomes have conserved plastome structures. Based on the plastome hotspots of nucleotide diversity, several genes and noncoding regions are suitable for phylogenetic and population studies. *Chiloschista* may provide an ideal system to investigate the dynamics of plastome evolution and DNA barcoding investigation for orchid studies.

Keywords Aeridinae, Chloroplast genome, DNA barcoding, Small inversions, Phylogenetic analysis

Background

Species identification is crucial for biodiversity conservation, evolutionary analysis and molecular breeding [1–3]. DNA barcodes have been developed for species identification using DNA sequences from specific genes or intergenic regions [4, 5]. Barcoding facilitates quick and accurate identification of species with the benefits of uniformity, digitization, scalability and high reproducibility [6]. There are different DNA barcodes suitable for different taxa. The mitochondrial gene *cytochrome oxidase I* (*COI*) is the standard barcode for animals [7], but it is not suitable for plant species due to the low substitution rates [8]. Standardized barcodes have been used in plants. For example, the plastid genes *matK* and *rbcL* are used as the core barcodes for plant species identification, and the plastid intergenic sequences *trnL-F* and *trnH-psbA* serve as the spare barcodes [9]. Moreover, the internal transcribed spacer (ITS) of nuclear ribosomal DNA (nrDNA) comprises plant core barcodes, and it is also suitable for fungal species identification [10, 11]. The plant core barcodes are widely applied in most taxonomic studies [12, 13] but are limited in some taxonomically complex groups due to pseudogene amplification and radiation evolution [14–16].

An extended DNA barcode has been proposed using advanced genome sequencing technology [5, 17]. Genome skimming can concurrently obtain plastid genome (plastome), mitochondrial genome, and nrDNA sequences, which are congruent with standard DNA barcodes [17, 18]. The plastome is suitable and convenient for barcoding investigation and phylogenetic analysis due to its uniparental inheritance, moderate mutation rate and high copy number [19, 20]. The plastome also contains more informative sites than normal DNA barcodes and can be easily derived from genome skimming data by *de novo* assembly, which can work on herbarium specimens with degraded DNA [18, 21]. Therefore, all plastome sequences and nrDNA sequences are also known as ultrabarcodes or next generation barcodes [1, 22, 23]. Plastome barcodes may promote the classification and phylogenetic analysis of problematic taxonomic groups.

Orchidaceae is a taxonomically complex group that contains 28,000 species in 700 genera [24–26]. The standard barcodes (*matK*, *rbcL*, ITS) and supplementary barcodes (*psbA-trnH*, *trnL-F*, *atpI-H*) were applied in

previous studies that resolved subfamily relationships and main clade relationships in some genera or subtribes [14–16, 27]. However, their use is limited in rapidly evolving orchids. For example, in the subtribe Aeridinae and related genera, five DNA barcodes have been used in phylogenetic analysis, providing a lower resolution tree due to radiation evolution [16, 28–30].

The genus *Chiloschista* Lindl. (1832), comprising approximately 20 species [31], is mainly distributed in China and the Indian subcontinent through South-east Asia to Australia [32]. The species of *Chiloschista* are listed in the Convention on International Trade in Endangered Species (CITES). The orchids of *Chiloschista* are epiphytic with characteristic stemless and leafless bodies and are commonly known as shootless orchids [33]. The genus belongs to the subtribe Aeridinae of Orchidaceae, containing several genera (*Phalaenopsis* and *Vanda*) with high ornamental value. Previous phylogenetic studies indicated that *Chiloschista* was sister to *Dimorphorchis* and *Thrixspermum* in the subtribe Aeridinae basic clade A [32], but Zou et al. [16] recognized it as an independent secondary basic clade of Aeridinae. Therefore, phylogenetic analysis and barcoding investigation of *Chiloschista* may contribute to the classification of the Aeridinae subtribe.

The use of orchid plastomes has experienced explosive growth, especially in the past ten years, with the low-cost and widespread use of next-generation sequencing technology. A total of 474 plastomes have been released by NCBI [34], and the associated studies mainly concentrated on genome structure comparison, barcoding investigation, plastid phylogenomics and mycoheterotrophic plastome evolution [2, 3, 24, 35]. Recently, plastome studies have been performed on genome repeats, structural characteristics and mutation hotspots to investigate plastome barcodes, suggesting the effective use of plastome barcodes for species identification [3, 23, 36, 37]. Plastome data have been used to resolve the phylogenetic relationships of tribes, subtribes and genera in Orchidaceae [3, 24, 38–43], which further indicates that the plastomes are useful for phylogenetic relationship reconstruction. Moreover, plastomes have also been used to investigate the evolution of mycoheterotrophic orchid genomes suffering extreme gene loss and structural variation [43–47]. However, plastome analysis has yet to be

implemented in leafless orchids, comprehensive plastome comparison and phylogenetic analysis in *Chiloschista*.

In this study, seven complete plastomes of *Chiloschista* were assembled, with the aim of investigating the plastome evolution, potential barcodes and phylogenetic relationships of the genus. We used plastome data to address the following specific questions: (1) What are the plastome characteristics of *Chiloschista* and the differences compared with other orchids? (2) What is the phylogenetic relationship (intergeneric and intrageneric) of *Chiloschista*, and how can useful barcodes be identified?

Results

Plastome characteristics and structure

The *k*-mer coverage of the seven *Chiloschista* species plastomes (accession number OP953683–OP953689) sequenced and assembled with Illumina reads was 108.5–122.8x (Table S1). Plastome sizes ranged from 143,223 bp for *C. pusilla* and 145,463 bp for *C. yunnanensis*, which all fell within the typical angiosperm plastome size range. Each plastome possessed the quadripartite structure common to angiosperm plastomes, with comparable percentages in each region (LSC 57.7–58.3%, IR 34.5–35.1%, and SSC 7.0–7.4%) and similar G/C contents (36.8–36.9%) (Table 1).

A total of 120 genes (including repeat genes) were contained in each *Chiloschista* plastome, of which 74 were protein-coding genes, 38 were transfer RNA (tRNA) genes, and eight were ribosomal RNA (rRNA) genes (Table 1). Pseudogenization has occurred extensively in Aeridinae species (Li et al. 2019; Liu et al. 2020; Kim et al. 2020), wherein *ndh* genes were generally lost or truncated in *Chiloschista* (Fig. 1; Table 1). The *ndh* genes in the genus were all pseudogenes, ranging from four to six members in each plastome (Table 1). The plastomes of *C. guangdongensis*, *C. exuperei* and *C. yunnanensis* possessed six (*ndhB/C/E/G/I/J/K*), five (*ndhB/C/G/I/J/K*) and six (*ndhB/C/E/G/I/J/K*) pseudogenes, respectively. The other four species possessed four *ndh* pseudogenes (*ndhB/E/G/I*).

The results of IR boundary analysis indicated that JLA/B were more conserved than JSA/B (Fig. 2). The adjacent regions of the LSC and IRA (JLA) located at the *psbA* gene were similar in *Chiloschista*. The adjacent region of LSC and IRB (JLB) located at the *rpl22* gene was the same in *Phalaenopsis hygrophila* and *Chiloschista*. The adjacent region of SSC and IRA (JSA) containing the *ycf1* gene, the genes of *C. yunnanensis* and *P. hygrophila*, was complete in the SSC region with no *ycf1* fragments in the adjacent region of SSC and IRB (JSB). The SSC regions of *C. yunnanensis* and *P. hygrophila* were expanded in the above two species.

The typical *Chiloschista* plastomes exhibited no inversions or rearrangements, but one inversion was identified compared with the *P. hygrophila* reference plastome (Fig. 3). Fourteen small inversions (SIs) were identified in the seven *Chiloschista* plastomes (Fig. 4; Table S2). The locations of all 14 SIs were in stem- and loop-forming regions, with 12 SIs being detected in the intergenic region and two in the intron (Fig. 4). The LSC region contained 12 SIs, and the IR and SSC regions contained one SI (Fig. 1). Those of the 14 primary SI types that have base substitutions in the stem or loop were catalogued as subtypes. The catalogue of subtypes in the stem or loop may easily identify the trait distribution states during plastome evolutionary processes. The dG value for each hairpin is displayed in Table S2, with high stability between different species. dG represents the quantity of energy needed to fully break a secondary DNA structure. The dG values also indicate the stability of each hairpin.

To clarify allied species or intragenus variations, the number of repeat sequences and regions of distribution were examined. First, four types of repeats (complement, forward, palindrome and reverse) were examined in *Chiloschista* plastomes (Fig. 5A, Table S3). The majority of the repeat sequences were in the 20–29 bp range, followed by 30–39 bp, then over 40 bp, with the fewest in the over 40 bp range. There were no C and R repeats found over 40 bp in length, and they were infrequent even in the smaller size ranges. In the 30–39 bp group, there were no C repeats, and R repeats were detected

Table 1 Characteristics of the complete plastomes of the *Chiloschista* lineages

Species name	Size (bp)	GC content (%)	LSC size in bp (%)	IR size in bp (%)	SSC size in bp (%)	Total number of gene	Protein-coding gene	tRNA gene	rRNA gene	<i>ycf1</i> fragment (bp)	Number of <i>ndh</i> fragment
<i>Chiloschista exuperei</i>	144,909	36.9	84,191 (58.1)	25,297 (34.9)	10,124 (7.0)	120	74	38	8	69	5
<i>C. guangdongensis</i>	145,322	36.9	84,713 (58.3)	25,231 (34.7)	10,147 (7.0)	120	74	38	8	59	6
<i>C. lunifera</i>	145,004	36.8	83,684 (57.7)	25,390 (35.0)	10,540 (7.3)	120	74	38	8	148	4
<i>C. pusilla</i>	143,223	36.8	83,088 (58.0)	24,835 (34.7)	10,465 (7.3)	120	74	38	8	179	4
<i>C. sp. 128</i>	144,689	36.9	83,508 (57.7)	25,357 (35.1)	10,467 (7.2)	120	74	38	8	148	4
<i>C. viridiflava</i>	143,233	36.8	83,084 (58.0)	24,842 (34.7)	10,465 (7.3)	120	74	38	8	179	4
<i>C. yunnanensis</i>	145,463	36.9	84,497 (58.1)	25,107 (34.5)	10,752 (7.4)	120	74	38	8	0	6

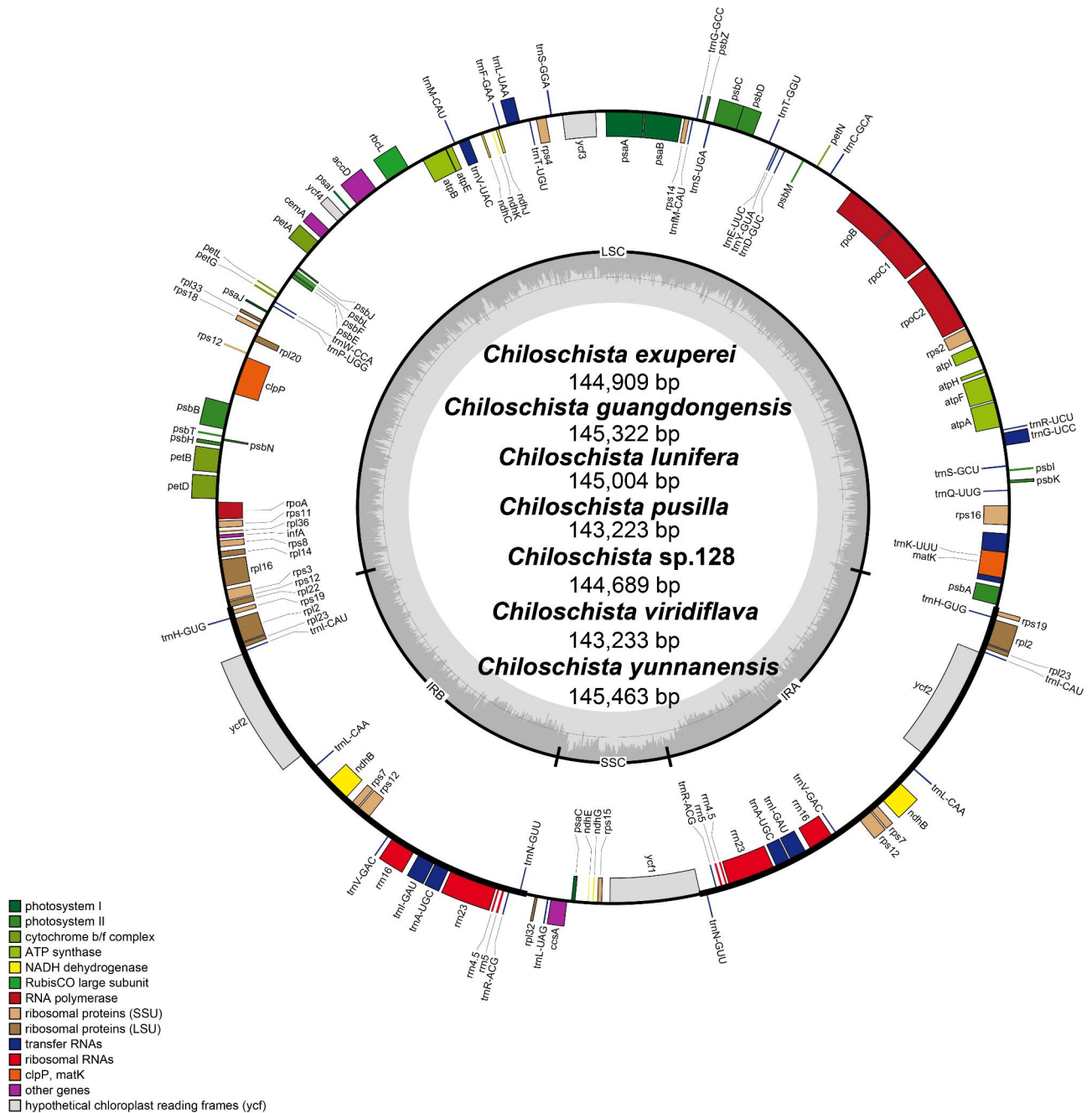


Fig. 1 The annotation map of seven *Chiloschista* plastomes. The darker gray in the inner circle corresponds to the GC content. The IRA and IRB (two inverted repeating regions); LSC (large single-copy region); and SSC (Small single-copy region) are indicated outside of the GC content

in most species (Fig. 5A). Additionally, we investigated regions in which six types (mono-, di-, tri-, tetra-, penta-, and hexa-) of SSRs were analysed. A total of 47 (*C. exuperei*) – 56 (*C. pusilla*) SSRs were found in *Chiloschista* (Fig. 5B, Table S4). The majority of SSRs were found in the LSC region, while few SSRs were found in the SSC region (Fig. 5B).

Plastome sequence divergence, evolution and barcoding investigation

We used mVISTA to find regions with high variations between conserved regions to further describe the differences between seven newly assembled *Chiloschista* species plastomes and that of *P. hygrochila* (outgroup species). High variation was identified in the intergenic and intragenic regions of plastomes in *Chiloschista* and *P. hygrochila* (Fig. 6), particularly in the LSC (from *rpoB* to

Inverted Repeats

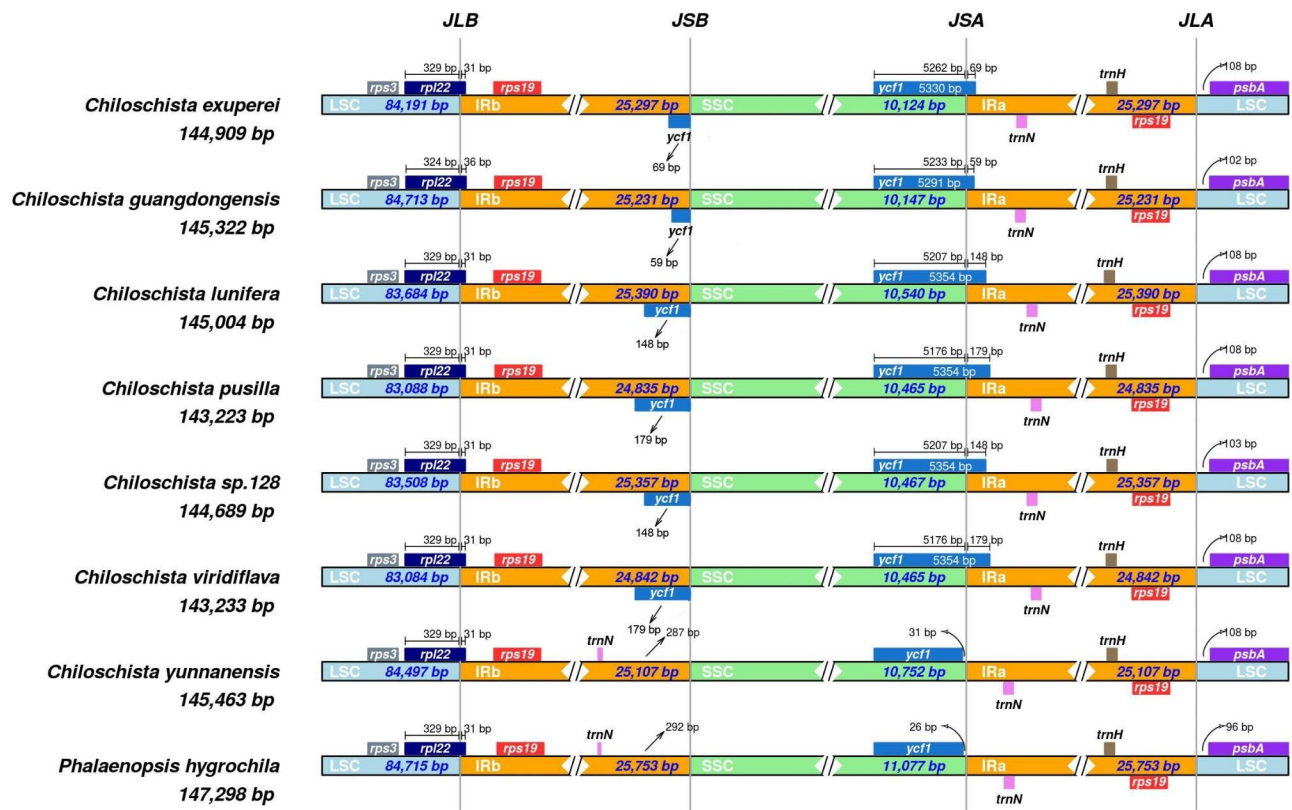


Fig. 2 Comparison of junctions between the LSC, SSC, and IR regions among seven newly assembled *Chiloschista* plastomes and *Phalaenopsis hygrochila*

psbD) and SSC regions (from *rpl32* to *ycf1*). As a result of these findings, multiple intergenic and intragenic regions may be suitable for DNA barcode investigations that can easily distinguish between *Chiloschista* species.

To investigate the evolutionary characteristics of the *Chiloschista* plastomes, we conducted a substitution rate comparison analysis of 68 protein-coding genes. Among the genes in *Chiloschista* plastomes, the estimated non-synonymous (dN)/synonymous (dS) substitution rates were 0.00010–0.62310 under purifying selection, except for *petG* (2.13591) and *psbF* (3.87157) genes under positive selection (Fig. S1, Table S5).

To further analyse the mutation hotspots in the *Chiloschista* plastomes, we used DnaSP6 to assess the nucleotide diversity (Pi) of the alignment of the complete genome. The results showed high divergence of the SSC region and conservation of the IR region (Fig. 7, Tables S6, S7). We selected six mutational hotspots (Pi value: *trnN^{GUU}-rpl32* (0.191) > *rpoB-trnC^{GCA}* (0.097) > *psbK-psbI* (0.065) > *psaC-rps15* (0.063) > *trnE^{UUC}-trnT^{GGU}* (0.061) > *accD-psaI* (0.059)) for candidate barcodes. The protein-coding genes were also used for nucleotide diversity analysis. The results showed five coding sequences (Pi

value: *ycf1* (0.050) > *rps15* (0.040) > *matK* (0.034) > *psbK* (0.032) > *ccsA* (0.031)) with high nucleotide diversity that were appropriate for phylogeny.

Eleven datasets were obtained to perform species discrimination analysis using the tree-building method (Table 2). The results showed that the lengths of the five coding sequence potential barcode (*ycf1*, *matK*, *ccsA*, *rps15*, *psbK*) matrices were 6,105, 1,595, 996, 276, and 192, respectively. The informative sites of the above five potential barcodes contained 784, 147, 90, 27, and 16 informative sites. The species discrimination rate (95.45%, 68.18%, 59.09%, 45.45%, 13.64%) corresponded with the number of informative sites in coding sequences. The results showed that the lengths of six noncoding sequence potential barcodes (*psbK-psbI*, *rpoB-trnC^{GCA}*, *psaC-rps15*, *trnE^{UUC}-trnT^{GGU}*, *accD-psaI*, *trnN^{GUU}-rpl32*) were 688, 2,048, 1,017, 1,846, 1,230, and 5,158, respectively. The informative sites of six non-coding potential barcodes contained 72, 278, 137, 259, 128, and 356 informative sites, which did not correspond to the species discrimination rates (76.19%, 71.43%, 68.75%, 66.67%, 40.91%, and 35.00%, respectively). The

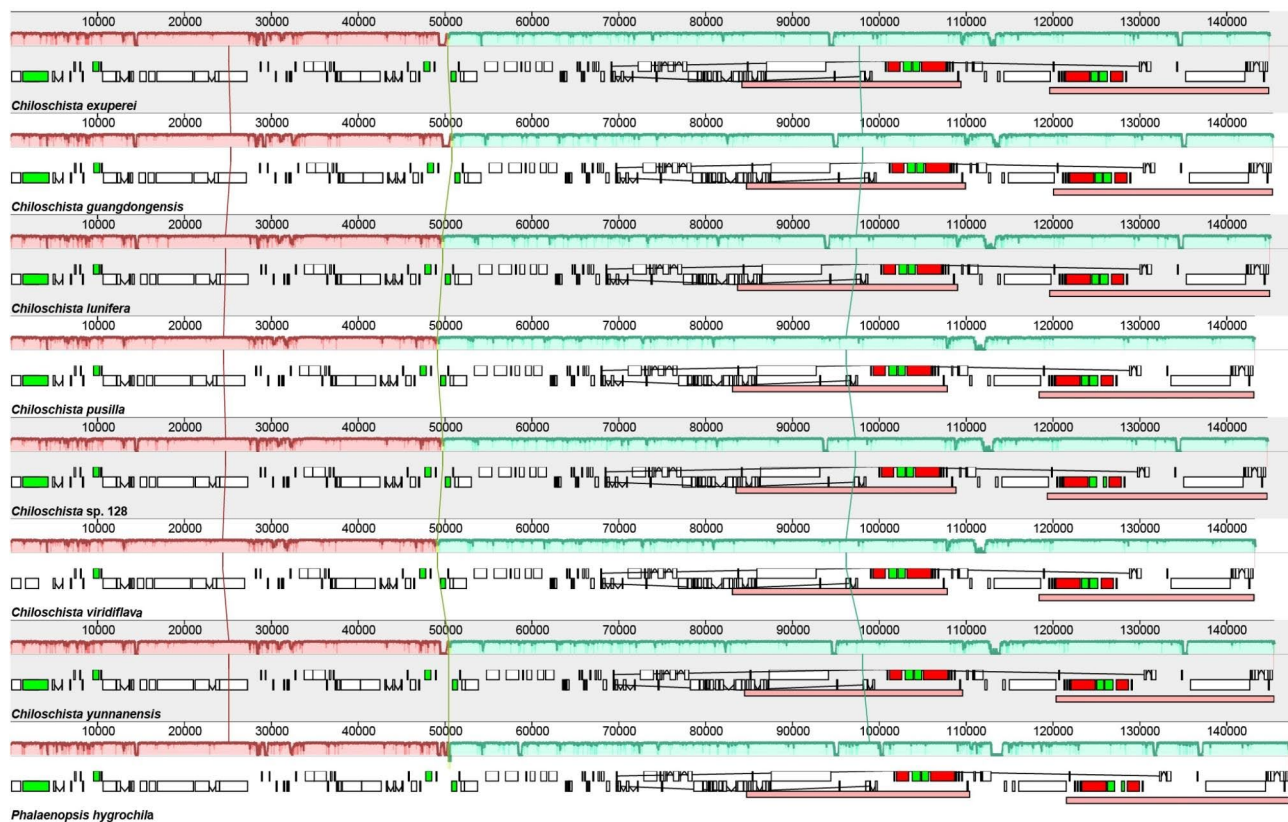


Fig. 3 Plastome comparison of seven species of *Chiloschista* and *Phalaenopsis hygrochila* using a progressive MAUVE algorithm

results showed that *ycf1* and *psbK-psbI*, with high species discrimination, were appropriate for use as barcodes.

Phylogenetic analysis

The phylogenetic relationship analysis was conducted by three methods (ML, MP and BI) based on the whole plastome, and 68 protein-coding sequences resulted in a similar topology (Fig. 8; Fig. S2). Seven species of *Chiloschista* formed a monophyletic genus, which was clustered with *Phalaenopsis* at the basic clade of subtribe Aeridinae with moderate support values. The seven *Chiloschista* species were classified into three major clades. *C. guangdongensis* and *C. exuperei* formed the first clade of *Chiloschista*, *C. yunnanensis* was the second clade, and *C. viridiflava*, *C. pusilla*, *C. sp. 128* and *C. lunifera* formed the third clade.

We also used informative noncoding sequences and coding sequences to reconstruct the phylogenetic relationship for useful barcode investigation. The informative noncoding sequence *trnN^{GUU}-rpl32*, coding sequence *ycf1*, concatenation of six noncoding sequences, and concatenation of five coding sequences were used for phylogenetic analysis (Fig. S3). The phylogenetic analysis of *Chiloschista* based on *trnN^{GUU}-rpl32* resulted in high support values but different relationships among *C. guangdongensis*, *C. exuperei* and *C. yunnanensis*. Three

phylogenetic trees presented same topologies, and the tree based on five coding sequences had high support values.

Discussion

Plastome characteristics and structural evolution

Plastome sizes in Orchidaceae are diverse due to variable lifeforms, ranging from 19,047 bp (*Epipogium roseum*) to 212,688 bp (*Cypripedium tibeticum*) [48, 49]. The plastome size of *Chiloschista* was conserved, ranging from 143,233 bp to 145,643 bp, which is in accordance with the ranges of previously reported orchid plastomes. The GC content (36.8–36.9%) also fell within the range of 23.1% (*Gastrodia flexistyla*) to 37.8% (*Cypripedium macranthos*) [49, 50]. Plastome size in orchids is closely related to gene loss and IR boundary shifts [41, 51]. We found that the *ndh* genes of *Chiloschista* were pseudogenes (Fig. S4, Table 1), which is consistent with the plastome study of the subtribe Aeridinae [2, 38]. Previous studies suggested that the IR boundary shift was also related to plastome size; the *Pelargonium* plastomes ranged from 165,508 bp to 242,575 bp, and IR expansion or contraction was the main contribution [52–54]. The present study revealed that the IR boundary JLA/B was conserved compared with JSA/B (Fig. 2), while compared with other orchids, the *Chiloschista* IR boundary was

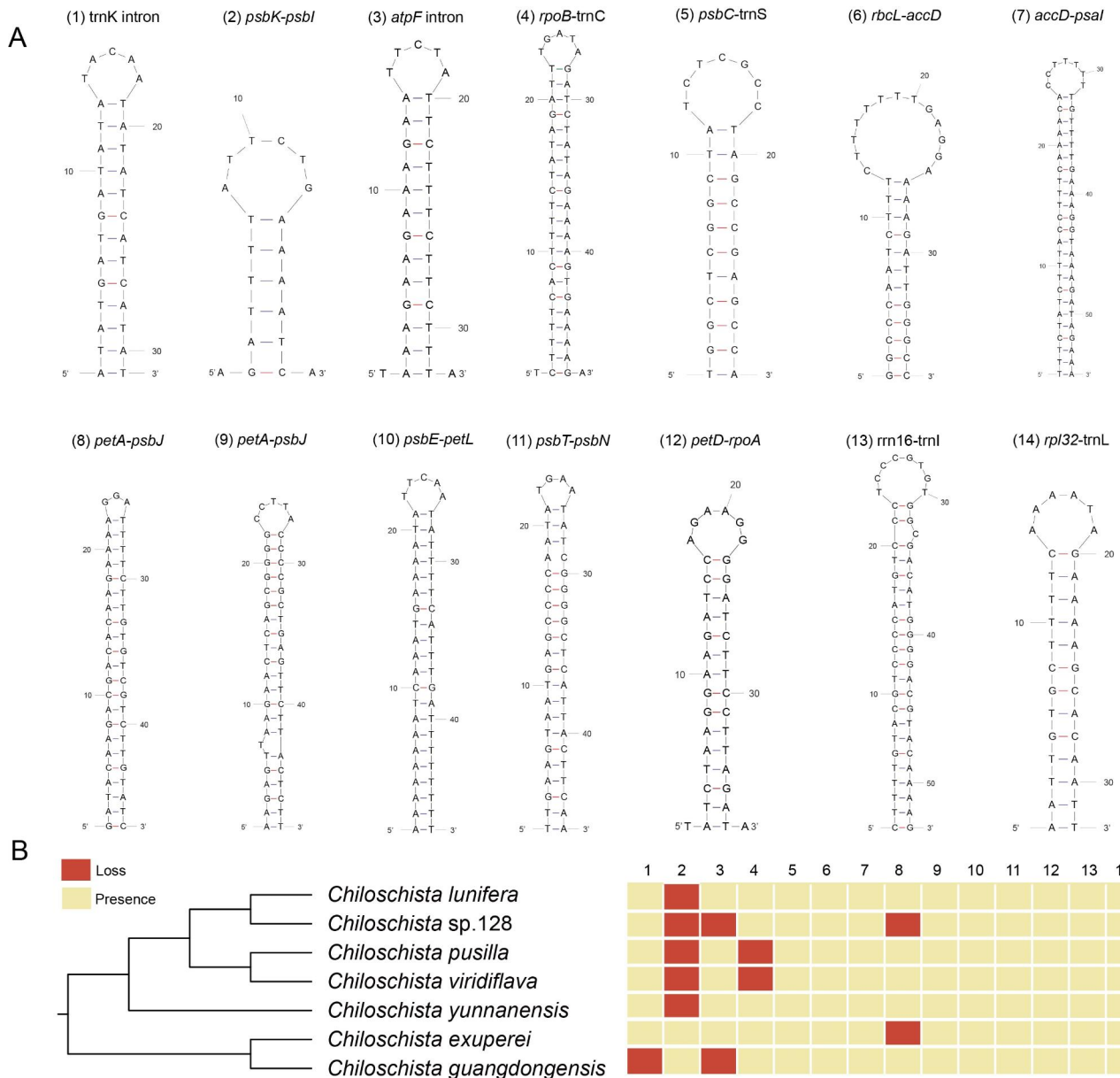


Fig. 4 Stem-loop structure of fourteen small inversions across seven *Chiloschista* species. Major types are represented. Details of free energy, sequences, loop length, and subtypes are described in Table S3

conserved [41, 49]. The plastome size of *Chiloschista* was relatively conserved, and the variation could be contributed by *ndh* loss instead of the IR boundary shift.

Plastome protein-coding gene loss is usually found in orchids, and mycoheterotrophic orchids obtain energy from fungi, leading to plastome degradation [44, 45, 48, 55]. However, autotrophic orchids also suffer gene loss on some occasions [2, 38, 56, 57]. We annotated 120 genes (containing repeat genes) in *Chiloschista*, and the gene number was less than that of other autotrophic orchids due to *ndh* gene loss. The *ndh* gene loss was general in epiphytic orchids, and previous studies suggested that the *ndh* loss did not correlate to taxonomic or evolutionary

relationships [36, 56–58]. We also analysed *ndh* gene loss/pseudogenization in the evolutionary process. The results of Aeridinae did not exhibit an obvious *ndh* deletion pattern, as in previous studies [2, 38]. However, we found that *ndh* loss in *Chiloschista* can be divided into two parts: *ndhC/K* was absent in four species in a clade, and three species in basic clades were pseudogenes (Fig. S4). Our results suggested that *ndh* gene loss or pseudogenization might be accompanied by *Chiloschista* evolution. We also conducted substitution rate analysis to compare the evolution of genes. The results showed that the *petG* and *psbF* genes were under positive selection, and the other genes were under purifying selection.

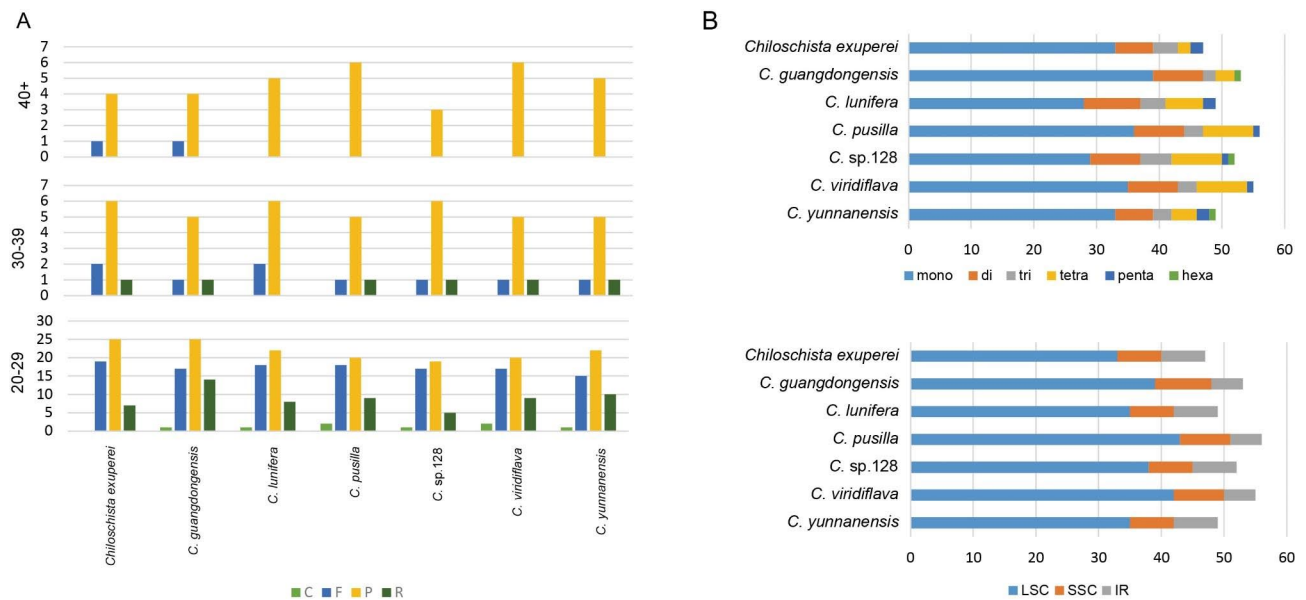


Fig. 5 Summary of simple sequence repeats (SSR) across the *Chiloschista* species. **(A)** Variation in repeat abundance and type in seven plastomes. **(B)** Number of SSRs for each *Chiloschista* species by SSR unit size, and number of SSRs for each Aeridinae species by location in IR, LSC, and SSC

A previous study suggested that genes under positive selection may affect photosynthetic efficiency [59], and whether they can benefit *Chiloschista* adaptation needs more study in the future.

Plastome inversions are widespread in flowering plants [60–62], and recent studies have also found that they are universal in orchids [2, 41, 43]. Distinct inversions were not detected in *Chiloschista* in this study, but compared with the reference plastome of *Phalaenopsis hygrophila*, there was an inversion of approximately 500 bp in the intergenic region of *rps4-atpE* (Fig. 3); this inversion could be unique to the genus. We also performed small inversion/hairpin inversion (SI) analysis. SIs are ubiquitous in the angiosperm plastome and are always flanked by inverted repeats of approximately 8 to 50 bp [63]. There were 14 SIs detected in *Chiloschista* plastomes (Fig. 4, Table S2), which was consistent with the study of Aeridinae [38]. In each species of *Chiloschista*, there ranged from 11 to 13 SIs, and no phylogenetic signal was observed (Fig. 4). Eight SIs (3, 6, 7, 9, 10, 11, 12, 13) were different between *Chiloschista* plastomes (Table S2). A previous study indicated that intergeneric SI was easily distinguished [38], and based on our results, SIs may not be easily distinguished intrageneric relationships.

Phylogenetic analysis and barcoding investigation

Our results revealed the phylogenetic position of *Chiloschista* and the intrageneric relationships. Few studies have been conducted on the molecular phylogenetics of *Chiloschista*. Carlswald et al. [33] indicated that *Chiloschista* was sister to the clade of four species (*Amesiphilippinensis*, *Tuberolabium kotoense*, *Dyakia*

hendersoniana and *Tuberolabium brevirachis*), and the intrageneric relationship was ((*C. lunifera*, *C. pusilla*) *C. viridiflava*) *C. parishii*). Topik et al. [64] suggested that *C. viridiflava* was sister to *Ornithochilus difformis* (accepted name *Phalaenopsis difformis*) and belonged to the *Pelatantheria* alliance based on ITS and *matK*. Pridgeon et al. [32] suggested that *Chiloschista* was embedded in *Dimorphorchis* and sister to *Thrixspermum* formed the basic clade of Aeridinae. According to Zou et al. [16], *Chiloschista* was an independent clade of Aeridinae, and the intrageneric relationship was (*C. sp. 4516* (*C. yunnanensis* (*C. lunifera* (*C. pusilla*, *C. parishii*))). Previous studies documented that the relationships of *Chiloschista* usually exhibited unstable topology and low support values. Our results suggested that *Chiloschista* was an independent clade of Aeridinae, and it was clustered with *Phalaenopsis* at the basic clade with moderate support values (55/-/0.91) (Fig. 8). The relationship was not consistent with previous studies [16, 32], and the moderate support value implied that the *Chiloschista* could have different topology. We showed the full support values of the intrageneric relationship of *Chiloschista*, which provided new insight into relationship clarification.

Moreover, we performed nucleotide diversity analysis of the complete plastome and coding genes to investigate useful DNA barcodes for phylogenetic analysis. Plastomic mutational hotspots are convenient and practical regions for DNA barcoding development, as suggested by previous studies in orchids [3, 23, 37, 65–67]. The complete plastome nucleotide diversity analysis suggested that the SSC region was more variable than the LSC and IR regions (Figs. 6 and 7). Six intergenic regions

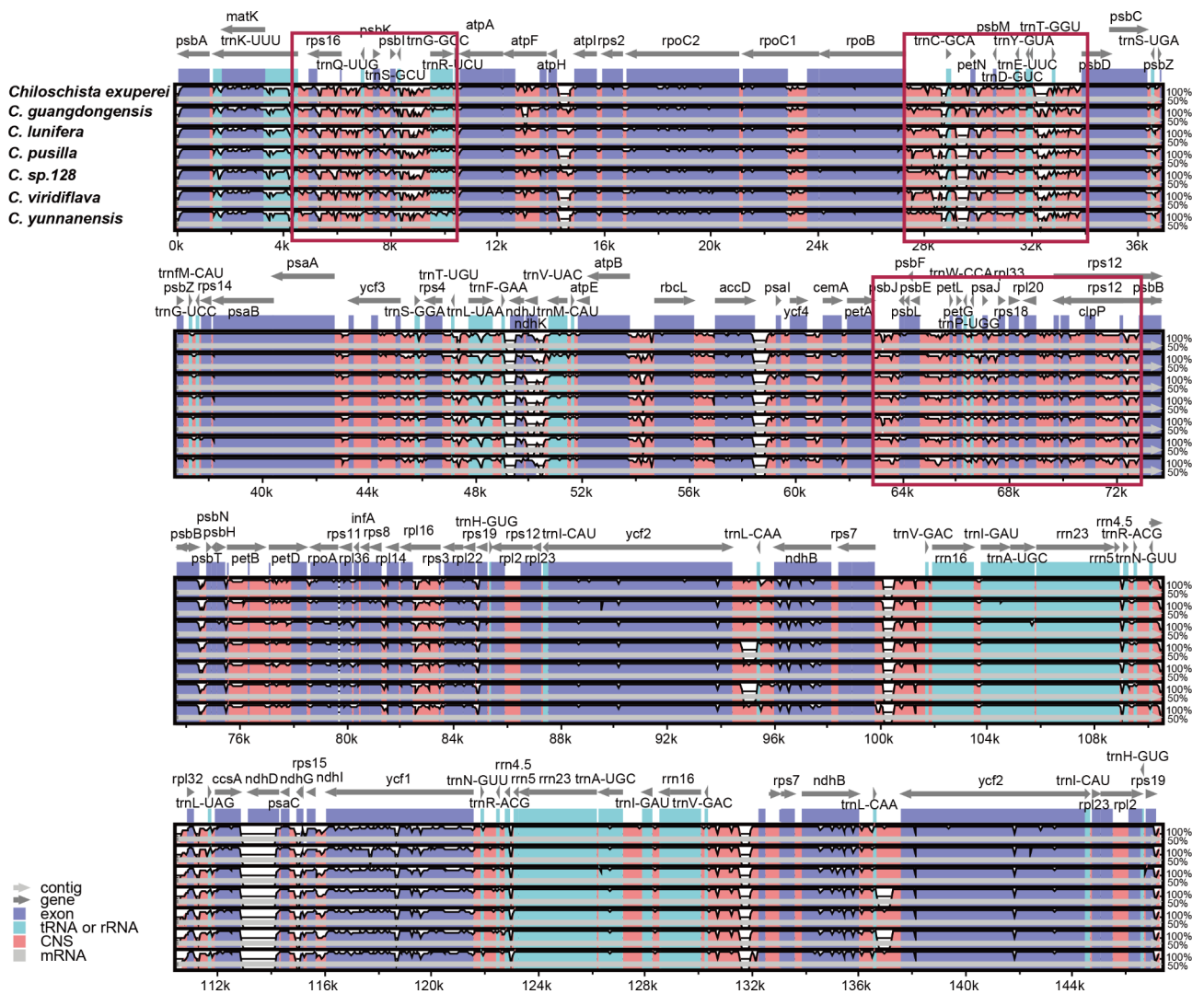


Fig. 6 Global alignment of seven *Chiloschista* plastomes using mVISTA with *Phalaenopsis hygrochila* as reference. The y-axis shows the coordinates between the plastomes. The red boxes mean high variation regions in plastome sequence

(*trnN^{GLU}-rpl32*, *rpoB-trnC^{GCA}*, *psbK-psbI*, *psaC-rps15*, *trnE^{LUC}-trnT^{GGU}*, *accD-psaI*, $P_i > 0.06$) were selected for phylogenetic analysis (Fig. 2, Table S6). The results suggested that the potential barcodes possessed high diversity but might not be suitable for phylogenetic analysis due to the unstable topology. Based on the nucleotide diversity of protein-coding genes, five coding genes (*ycf1*, *rps15*, *matK*, *psbK*, *ccsA*, $P_i > 0.03$) were selected for phylogenetic analysis (Fig. S2, Table S7). The results based on five genes showed relatively high diversity and suitability for phylogenetic analysis. We also evaluated the potential barcodes of five coding sequences and six noncoding sequences for species discrimination (Table 2). The results indicated that the potential candidate barcode *ycf1* may be suitable for *Aeridinae* species discrimination. The applicability of this potential barcode to this group

will be further evaluated through the identification of similar species in subsequent studies.

High mutational regions also contain SSRs (simple sequence repeats) and large repeats, which are widely used in studies of genetic diversity, population structure and species identification [68–70]. A total of 47–56 SSRs and 11–14 long repeats (>30 bp) were identified in *Chiloschista* plastomes, which were mostly located in the intergenic region of LSC (Fig. 5, Tables S3, S4). The repeats in the coding regions were mainly located in the exons of *accD*, *rpoC2*, *ycf1* and *ycf2*. Most of the SSR types were mononucleotide repeats in the seven *Chiloschista* species. The mVISTA percent identity plot and sliding window analysis showed that the most divergent regions were located in regions of the *Chiloschista* plastomes. The results provide a data basis for future population genetics studies.

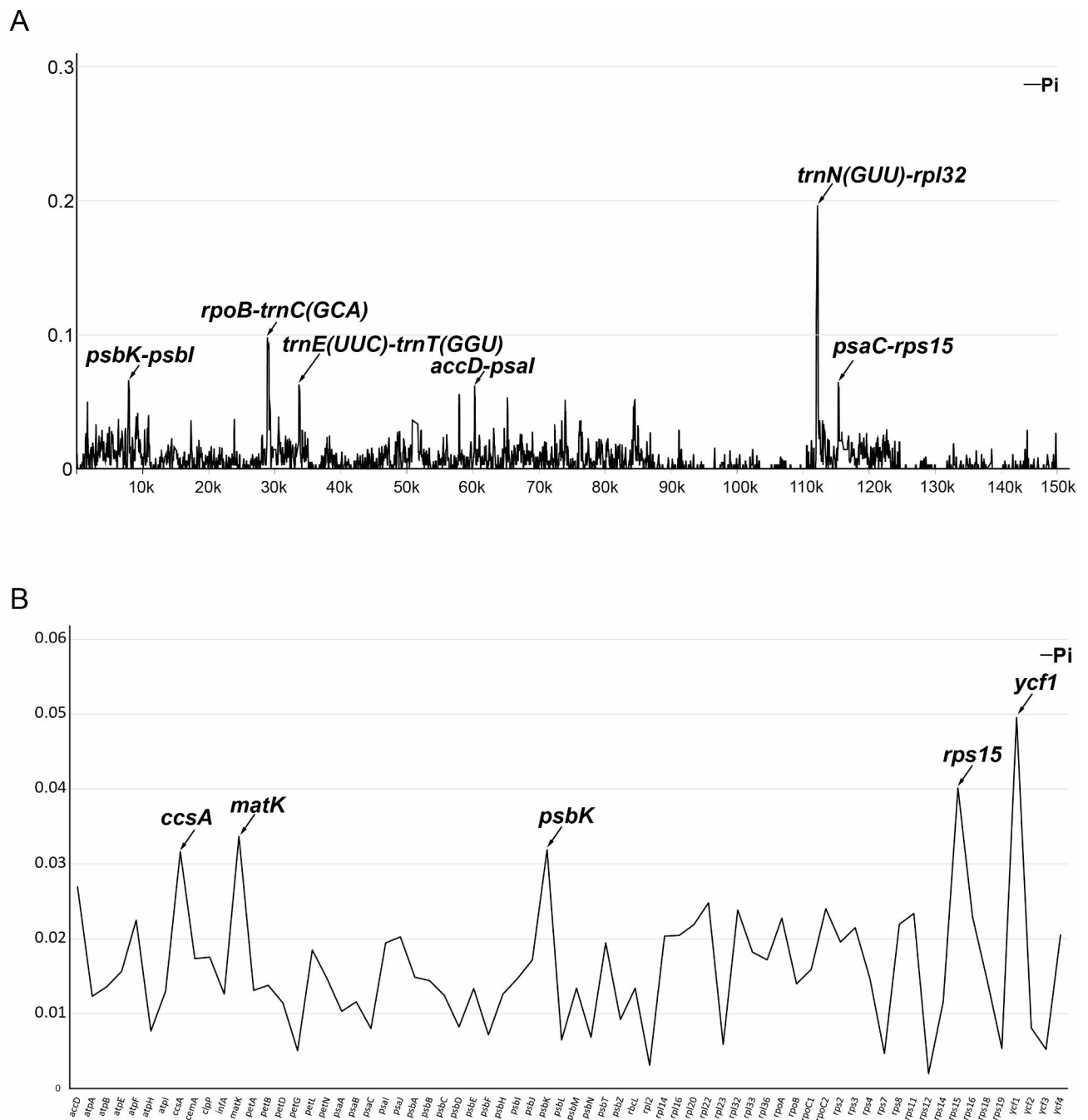


Fig. 7 Sliding window test of nucleotide diversity (π) in the *Chiloschista* Plastomes. **(A)** The nucleotide diversity of complete plastome, six mutation hotspot regions ($\pi > 0.06$) were annotated. **(B)** The nucleotide diversity of 68 protein coding sequence, five mutation hotspot region ($\pi > 0.03$). The window size was set to 100 bp and the sliding windows size was 25 bp. X-axis, position of the midpoint of a window; Y-axis, π values of each window

Conclusions

In this study, we obtained the complete plastomes of seven *Chiloschista* species (*C. exuperei*, *C. guangdongensis*, *C. lunifera*, *C. pusilla*, *C. sp. 128*, *C. viridiflava* and *C. yunnanensis*) and found that the plastomes of the seven species possessed a generally preserved overall structure and gene content. Genome sizes, GC contents, gene contents, repeats and IR boundary variations showed little

variance. It is important to note that all *ndh* genes in the plastomes of *Chiloschista* were deleted or truncated, as was also found in other species of the subtribe Aeridinae. The genes *petG* and *psbF* were under positive selection. We offer a resource for creating DNA barcodes to advance research on *Chiloschista* species, which suggests that the potential barcode *ycf1* is most suitable for species discrimination. Based on the available data, phylogenetic

Table 2 Evaluation of five coding genes and six noncoding sequences and ability to discriminate species

	<i>ycf1</i>	<i>matK</i>	<i>ccsA</i>	<i>rps15</i>	<i>psbK</i>	<i>psbK-psbI</i>	<i>rpoB-trnC^{GCA}</i>	<i>psaC-rps15</i>	<i>trnE^{UUC}-trnT^{GGU}</i>	<i>accD-psal</i>	<i>trnN^{GUU}-rpl32</i>
Length of aligned sequence (bp)	6105	1595	996	276	192	688	2048	1017	1846	1230	5158
No. of variable sites (bp)	1605	295	160	48	27	184	663	314	624	289	1034
No. of informative sites (bp)	784	147	90	27	16	72	278	137	259	128	356
No. of species samples	22	22	22	22	22	21	21	16	21	22	20
Species discriminate	95.45%	68.18%	59.09%	45.45%	13.64%	76.19%	71.43%	68.75%	66.67%	40.91%	35.00%

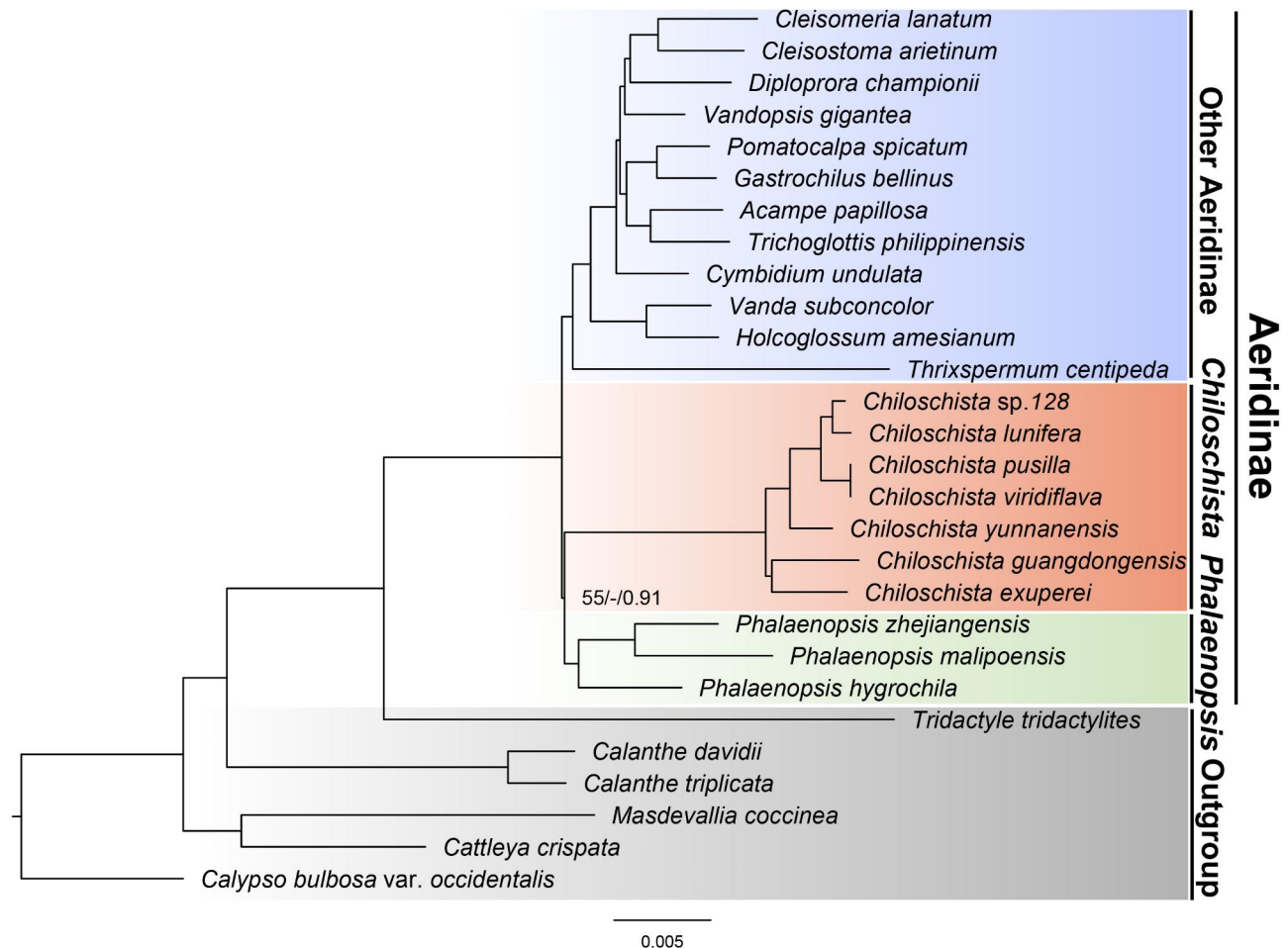


Fig. 8 Phylogenetic tree of *Chiloschista* and other 15 Aeridinae species based on the complete plastome data. Numbers near the nodes are bootstrap percentages and Bayesian posterior probabilities (BS_{ML} left, BS_{MP} middle, and PP right). The nodes without values are 100% bootstrap or 1.00 posterior probability

analysis was performed to identify the genus *Chiloschista* in the subtribe Aeridinae and, to a considerable extent, to establish the phylogenetic relationships of the major taxonomic groups in the subtribe and above subtribe of Orchidaceae. These discoveries help us better comprehend the characteristics and evolution of *Chiloschista* plastomes, which further our knowledge of phylogenetic relationships and DNA barcoding for Aeridinae species conservation and even extend to the Orchidaceae family.

Methods

Taxon sampling and sequencing

Plant materials of seven *Chiloschista* species were collected from the Forest Orchid Garden greenhouse at Fujian Agriculture and Forestry University (Fuzhou, Fujian Province, China). The formal identification of plant material was conducted by Dr. Ming-He Li and Prof. Zhong-Jian Liu. The voucher specimen of materials was deposited at the Herbarium of Fujian Agriculture and Forestry University, and deposition numbers and

GenBank accessions of species are listed in Table S8. A total of 28 species in 19 genera were analysed in combination with publicly accessible plastome data, including six species from five genera (*Calanthe*, *Calypso*, *Cattleya*, *Masdevallia* and *Tridactyle*) serving as the outgroup. According to the manufacturer's protocol, total DNA was extracted from fresh leaves using the Plant Mini Kit (Qiagen, CA, USA), and DNA degradation and contamination were examined by 1% agarose gel electrophoresis. According to the manufacturer's instructions, the HiSeq 4000 PE Cluster Kit (Illumina) was used to cluster the index-coded sample data on a cBot Cluster Generation System. The library preparations were sequenced on an Illumina HiSeq 4000 platform following cluster creation, and 150-bp paired-end reads were produced [71]. Scripts were used to filter the Illumina data in the cluster (default parameter: -L 5, -p 0.5, -N 0.1). When the low-quality ($Q \leq 5$) base number in sequencing reads surpassed 50% of the read base number and the N content in reads exceeded 10% of the read base number, paired reads were eliminated from the analysis.

Plastome assembly and annotation

To obtain plastid-like reads, the paired-end reads were filtered using the GetOrganelle pipeline (<https://github.com/Kinggerm/GetOrganelle>) [72]. The filtered reads were assembled using SPAdes version 3.10 [73]. To acquire pure plastid contigs, the final "fastg" files were filtered by the GetOrganelle script. The filtered De Bruijn graphs were then examined and corrected by Bandage [74]. The circular plastome was obtained through the above steps.

GeSeq [75] was used to annotate the newly assembled plastomes, while tRNAscan-SE v2.0.3 [76] was used to further verify tRNA genes. The start and stop codons in protein-coding genes found by GeSeq were manually visualized and corrected by alignment with the plastomes of related species in Geneious R11.1.5 [77]. The translation of each protein-coding gene was also validated by Geneious R11.1.5 [77]. A gene containing one or several internal stop codons compared to homologous genes was determined to be a pseudogene or partial copy. The plastome annotation file was generated using GB2Sequin [78], which was submitted to GenBank at the National Center for Biotechnology Information (NCBI) with unique accession numbers. The circle diagrams of annotated plastomes were drawn by OGDRAW [79]. The genes with $\geq 50\%$ loss of the complete CDSs or similarity $\leq 50\%$ were considered lost genes [38].

Plastome structure analysis

REPuter [80] was used to identify the long repeats of seven *Chiloschista* plastomes with default parameters, and four repeat types (F, forward, P, palindrome, R,

reverse, and C, complement) were identified. The Perl script of MISA [81] was used to identify simple sequence repeats (SSRs) with minimal thresholds of 10, 5, 4, 3, 3, and 3 repeat units for mono-, di-, tri-, tetra-, penta-, and hexa-motif microsatellites, respectively. The collinearity and rearrangements of plastomes were analysed and drawn by Mauve [82]. The boundary expansion or contraction analysis between the inverted repeat (IR) and single-copy (SC) regions was compared using Geneious 11.1.5 and IRscope [76, 83].

Sequence divergence and barcoding investigation

The online tool mVISTA [84] was used to analyse the plastome sequence diversity through the comparison of seven *Chiloschista* plastomes by the Shuffle-LAGAN alignment program [85]. The *P. hygrochila* plastome was used as a reference. Complete chloroplast genomes and 68 coding sequences of seven *Chiloschista* species alignment files were used to analyse the nucleotide diversities (P_i) with a window length of 100 sites and a step size of 25 sites by DnaSP 6 [86]. According to the P_i value, five protein-coding genes and six noncoding sequences were selected for species discrimination analysis. The tree-building method was adopted to analyse the eleven datasets. For the tree-building method, all datasets were aligned by MAFFT software [87], and the maximum likelihood (ML) tree was constructed by IQTREE [88, 89].

Phylogenetic analysis

The phylogenetic analysis of *Chiloschista* and other Aroidae species was conducted based on the whole plastome and 68 protein-coding sequences. The whole plastome sequences were aligned by MAFFT [87]. The protein-coding sequences (*ndh* genes were widely lost or truncated in Aroidae species) were aligned using MEGA 7.0 [90]. The alignment of the whole plastome was trimmed using trimAl v1.2 [91] with a heuristic approach (-automated1) to choose the best-automated method to decrease the systematic errors produced from poor quality. The online tool CIPRES Science Gateway (RaxML-HPC2 on XSEDE 8.2.12, PAUP on XSEDE 4.a168 and MrBayes on XSEDE 3.2.7) was used to perform phylogenetic analysis with three methods, including maximum likelihood (ML), maximum parsimony (MP) and Bayesian inference (BI) [92]. For MP analysis, 1000 tree-bisection-reconnection (TBR) searches with MAXTR-EES set to increase without limit were performed on the combined dataset by PAUP [93]. A heuristic search using 1000 random addition sequence repeats and TBR branch switching was conducted with all characters being equally weighted and unordered. For ML analysis, 1000 repeated self-expanding analyses using the GTRCAT model were carried out on all datasets [94]. For Bayesian analysis, the GTR+I+ Γ substitution model was

performed by MrBayes v. 3.2.7 [95]. The Markov chain Monte Carlo (MCMC) algorithm was run for 10,000,000 generations, with one tree sampled every 100 generations. To construct majority-rule consensus trees and estimate posterior probability (PP), the first 25% of trees were eliminated as burn-in.

Nucleotide substitution rate analysis

The sequences of 68 protein-coding genes were retrieved from the *Chiloschista* plastomes to investigate the nucleotide substitution rate (Table S5). The branch model was selected for nucleotide substitution rate analysis. The *Chiloschista* clade was the foreground clade, and the others were background clades. The values of dN/dS (LRT: M0 vs. M2) in plastid protein-coding genes were estimated by EasyCodeML v1.0 [96].

Abbreviations

BI	Bayesian inference
IR	Invert repeat
ITS	Internal transcribed spacer
LSC	Large single copy
ML	Maximum likelihood
MP	Maximum parsimony
nrDNA	Nuclear ribosomal DNA
SI	Small inversion
SSC	Small single copy
SSR	Simple sequence repeats

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09847-8>.

Supplementary Material 1: Fig. S1. Comparison of non-synonymous (dN) / synonymous (dS) substitution rates among *Chiloschista* plastid protein-coding genes. **Fig. S2.** Phylogenetic analysis of 22 Aerialinae species based on 68 protein coding genes. **Fig. S3.** Phylogenetic analysis of 22 Aerialinae species based on six noncoding barcodes and five coding barcodes. **Fig. S4.** The *ndh* genes loss across the subtribe Aerialinae. **Table S1.** The statistics of raw data and plastome assembly results. **Table S2.** The description small inversion sequences of seven *Chiloschista*. **Table S3.** The details information of long repeats. **Table S8.** A list of the taxa analysed with voucher information and GenBank accessions

Supplementary Material 2: Table S4. The details information Small simple repeats. **Table S5.** Comparison of non-synonymous (dN) / synonymous (dS) substitution rates among *Chiloschista* plastid protein-coding genes. **Table S6.** The nucleotide diversity of seven *Chiloschista* plastome. **Table S7.** The nucleotide diversity of 68 protein coding genes in *Chiloschista*

Acknowledgements

Not applicable.

Authors' contributions

S.L. and Z.-J. L. and conceived and designed the study. D.-K. L., X.-D. T., M.-Y. Z. and L.M. collected samples, Z.Z., X.-Y. G. and S.-W. X. conducted the experiments. D.-K. L. and C.-Y. Z. analyzed the data, D.-K. L., M.-H. L. and A.S. wrote the paper. All authors have read and approved the final manuscript.

Funding

This work was supported by the Forestry Peak Discipline Construction Project of Fujian Agriculture and Forestry University (72202200205) to ZJL and National Key Research and Development Program of China (2019YFD1000400) to SL.

Data Availability

The plastome sequences of *Chiloschista* are openly available in NCBI at the GenBank database with accession number OP953683–OP953689, the raw data has been submitted to the SRA database (BioSample: SAMN37279563–SAMN37279569; BioProject: PRJNA1012891; SRA: SRR25905744–SRR25905750).

Declarations

Ethics approval and consent to participate

Plant materials for this investigation were obtained with authorization from local Forest Department authorities. Voucher specimens (Please see Table S8) have been placed in herbaria that are open to the public (Herbarium of Fujian Agriculture and Forestry University). All procedures were carried out in conformity with the applicable norms and regulations.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹College of Forestry, Fujian Colleges and Universities Engineering Research Institute of Conservation and Utilization of Natural Bioresources, Fujian Agriculture and Forestry University, Fuzhou 350002, China

²Key Laboratory of Orchid Conservation and Utilization of National Forestry and Grassland Administration, College of Landscape Architecture and Art, Fujian Agriculture and Forestry University, Fuzhou 350002, China

³Fujian Health College, Fuzhou 350101, Fujian, China

Received: 6 January 2023 / Accepted: 27 November 2023

Published online: 06 December 2023

References

- Ji Y, Liu C, Yang Z, Yang L, He Z, Wang H, et al. Testing and using complete plastomes and ribosomal DNA sequences as the next generation DNA barcodes in *Panax* (Araliaceae). *Mol Ecol Resour*. 2019;19:1333–45.
- Liu DK, Tu XD, Zhao Z, Zeng MY, Zhang S, Ma L. Plastid phylogenomic data yield new and robust insights into the phylogeny of *cleisostoma-gastrochilus* clades (Orchidaceae, Aerialinae). *Mol Phylogenet Evol*. 2020;145:106729.
- Zhang L, Huang Y, Huang J, Ya J, Zhe M, Zeng C, et al. DNA barcoding of *Cymbidium* by genome skimming: call for next-generation nuclear barcodes. *Mol Ecol Resour*. 2022;23:1–16.
- Hebert PDN, Cywinska A, Ball SL, DeWaard JR. Biological identifications through DNA barcodes. *Proc R Soc London Ser B Biol Sci*. 2003;270:313–21.
- Coissac E, Hollingsworth PM, Lavergne S, Taberlet P. From barcodes to genomes: extending the concept of DNA barcoding. *Mol Ecol*. 2016;25:1423–8.
- Xu SZ, Li ZY, Jin XH. DNA barcoding of invasive plants in China: a resource for identifying invasive plants. *Mol Ecol Resour*. 2018;18:128–36.
- Hebert PDN, Ratnasingham S, DeWaard JR. Barcoding animal life: *cytochrome c oxidase subunit 1* divergences among closely related species. *Proc R Soc B Biol Sci*. 2003;270:96–9.
- Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. Use of DNA barcodes to identify flowering plants. *Proc Natl Acad Sci*. 2005;102:8369–74.
- Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, van der Bank M, et al. A DNA barcode for land plants. *Proc Natl Acad Sci*. 2009;106:12794–7.
- Li DZ, Gao LM, Li HT, Wang H, Ge XJ, Liu JQ, et al. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proc Natl Acad Sci*. 2011;108:19641–6.
- Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, et al. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. *Proc Natl Acad Sci*. 2012;109:6241–6.
- Chatrou LW, Pirie MD, Erkens RHJJ, Couvreur TLPP, Neubig KM, Abbott JR, et al. A new subfamilial and tribal classification of the pantropical flowering plant

- family annonaceae informed by molecular phylogenetics. Bot J Linn Soc. 2012;169:5–40.
13. Xu LS, Herrando-Moraira S, Susanna A, Galbany-Casals M, Chen YS. Phylogeny, origin and dispersal of *Saussurea* (Asteraceae) based on chloroplast genome data. Mol Phylogenet Evol. 2019;141:106613.
 14. Xiang XG, Schuiteman A, Li DZ, Huang WC, Chung SW, Li JW, et al. Molecular systematics of *Dendrobium* (Orchidaceae, Dendrobieae) from mainland Asia based on plastid and nuclear sequences. Mol Phylogenet Evol. 2013;69:950–60.
 15. Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S. Plant DNA barcoding: from gene to genome. Biol Rev. 2015;90:157–66.
 16. Zou LH, Huang JX, Zhang GQ, Liu ZJ, Zhuang XY. A molecular phylogeny of Aeriinae (Orchidaceae: Epidendroideae) inferred from multiple nuclear and chloroplast regions. Mol Phylogenet Evol. 2015;85:247–54.
 17. Hollingsworth PM, Li D-Z, van der Bank M, Twyford AD. Telling plant species apart with DNA: from barcodes to genomes. Philos Trans R Soc B Biol Sci. 2016;371:20150338.
 18. Dodsworth S. Genome skimming for next-generation biodiversity analysis. Trends Plant Sci. 2015;20:525–7.
 19. Ohyama K. Chloroplast and mitochondrial genomes from a liverwort, *Marchantia polymorpha*. Biosci Biotechnol Biochem. 1996;60:16–24.
 20. Zhang YJ, Ma PF, Li DZ. High-throughput sequencing of six bamboo chloroplast genomes: phylogenetic implications for temperate woody bamboos (Poaceae: Bambusoideae). PLoS ONE. 2011;6:e20596.
 21. Straub SCKK, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A. Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. Am J Bot. 2012;99:349–64.
 22. Kane N, Sveinsson S, Dempewolf H, Yang JY, Zhang D, Engels JMM, et al. Ultra-barcoding in cacao (*Theobroma* spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. Am J Bot. 2012;99:320–9.
 23. Yang JB, Tang M, Li HT, Zhang ZR, Li DZ. Complete chloroplast genome of the genus *Cymbidium*: lights into the species identification, phylogenetic implications and population genetic analyses. BMC Evol Biol. 2013;13:84.
 24. Givnish TJ, Spalink D, Ames M, Lyon SP, Hunter SJ, Zuluaga A, et al. Orchid phylogenomics and multiple drivers of their extraordinary diversification. Proc R Soc B Biol Sci. 2015;282:20151553.
 25. 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.
 26. Christenhusz MJM, Byng JW. The number of known plants species in the world and its annual increase. Phytotaxa. 2016;261:201.
 27. Guo YY, Luo YB, Liu ZJ, Wang XQ. Reticulate evolution and sea-level fluctuations together drove species diversification of slipper orchids (*Paphiopedilum*) in South-East Asia. Mol Ecol. 2015;24:2838–55.
 28. 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.
 29. Liu Q, Song Y, Jin XH, Gao JY. Phylogenetic relationships of *Gastrochilus* (Orchidaceae) based on nuclear and plastid DNA data. Bot J Linn Soc. 2019;189:228–43.
 30. Zhang GQ, Chen GZ, Chen LJ, Zhai JW, Huang J, Wu XY, et al. Phylogenetic incongruence in *Cymbidium* orchids. Plant Divers. 2021;43:452–61.
 31. Pham PD, Averyanov LV, Nguyen DH, Maisak T, Nguyen VC, Dang MQ, et al. A new species of *Chiloschista* (Orchidaceae, Aeriinae) from northern Vietnam. Phytotaxa. 2022;556:94–8.
 32. Pridgeon AM, Cribb PJ, Chase MW. Genera Orchidacearum volume 6: Epidendroideae (Part 3). OUP Oxford; 2014.
 33. 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.
 34. Zhang D, Zhao XW, Li YY, Ke SJ, Yin WL, Lan S, et al. Advances and prospects of orchid research and industrialization. Horticult Res. 2022;9:uhac220.
 35. Barrett CF, Wicke S, Sassi C. Dense intraspecific sampling reveals rapid and independent trajectories of plastome degradation in a heterotrophic orchid complex. New Phytol. 2018;218:1192–204.
 36. Chen Y, Zhong H, Zhu Y, Huang Y, Wu S, Liu Z, et al. Plastome structure and adaptive evolution of *Calanthe* s.l. species. PeerJ. 2020;8:e10051.
 37. 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;22:1–21.
 38. Kim YK, Jo S, Cheon SH, Kwak M, Kim YD, Kim KJ. Plastome evolution and phylogeny of subtribe Aeriinae (Vandaeae, Orchidaceae). Mol Phylogenet Evol. 2020;144:106721.
 39. Li YX, Li ZH, Schuiteman A, Chase MW, Li JW, Huang WC, et al. Phylogenomics of Orchidaceae based on plastid and mitochondrial genomes. Mol Phylogenet Evol. 2019;139:106540.
 40. Li ZH, Ma X, Wang DY, Li YX, Wang CW, Jin XH. Evolution of plastid genomes of *Holcoglossum* (Orchidaceae) with recent radiation. BMC Evol Biol. 2019;19:63.
 41. Guo YY, Yang JX, Bai MZ, Zhang GQ, Liu ZJ. The chloroplast genome evolution of Venus slipper (*Paphiopedilum*): IR expansion, SSC contraction, and highly rearranged SSC regions. BMC Plant Biol. 2021;21:248.
 42. Tang H, Tang L, Shao S, Peng Y, Li L, Luo Y. Chloroplast genomic diversity in *Bulbophyllum* section *Macrocaulia* (Bl.) Aver. (Orchidaceae, Epidendroideae, Malaxideae): insights into species divergence and adaptive evolution. Plant Divers. 2021;43:350–361.
 43. Tu X, Liu DK, Xu S, Zhou CY, Gao X, Zeng MY, et al. Plastid phylogenomics improves resolution of phylogenetic relationship in the *Cheirostylis* and *Goodyera* clades of Goodyerinae (Orchidoideae, Orchidaceae). Mol Phylogenet Evol. 2021;164:107269.
 44. Delannoy E, Fujii S, Colas des Francs-Small C, Brundrett M, Small I. Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes. Mol Biol Evol. 2011;28:2077–86.
 45. Barrett CF, Davis JI. The plastid genome of the mycoheterotrophic *Corallorhiza striata* (Orchidaceae) is in the relatively early stages of degradation. Am J Bot. 2012;99:1513–23.
 46. Feng YL, Wicke S, Li JW, Han Y, Lin CS, Li DZ, et al. Lineage-specific reductions of plastid genomes in an orchid tribe with partially and fully mycoheterotrophic species. Genome Biol Evol. 2016;8:2164–75.
 47. Barrett CF, Sinn BT, Kennedy AH. Unprecedented parallel photosynthetic losses in a heterotrophic orchid genus. Mol Biol Evol. 2019;36:1884–901.
 48. Schelkunov MI, Shtratnikova VY, Nuraliev MS, Selosse MA, Penin AA, Logacheva MD. Exploring the limits for reduction of plastid genomes: a case study of the mycoheterotrophic orchids *Epipogium Aphyllum* and *Epipogium Roseum*. Genome Biol Evol. 2015;7:1179–91.
 49. Guo YY, Yang JX, Li HK, Zhao HS. Chloroplast genomes of two species of *Cypripedium*: expanded genome size and proliferation of AT-Biased repeat sequences. Front Plant Sci. 2021;12:1–12.
 50. Wen Y, Qin Y, Shao B, Li J, Ma C, Liu Y, et al. The extremely reduced, diverged and reconfigured plastomes of the largest mycoheterotrophic orchid lineage. BMC Plant Biol. 2022;22:1–11.
 51. Kim YK, Jo S, Cheon SH, Joo MJ, Hong JR, Kwak MH, et al. Extensive losses of photosynthesis genes in the plastome of a mycoheterotrophic orchid, *Cyrtosia Septentrionalis* (Vanilloideae: Orchidaceae). Genome Biol Evol. 2019;11:565–71.
 52. Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, et al. The complete chloroplast genome sequence of *Pelargonium x hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol Biol Evol. 2006;23:2175–90.
 53. 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.
 54. Weng ML, Ruhlman TA, Jansen RK. Expansion of inverted repeat does not decrease substitution rates in *Pelargonium* plastid genomes. New Phytol. 2017;214:842–51.
 55. Graham SW, Lam VKY, Merckx VSFT. Plastomes on the edge: the evolutionary breakdown of mycoheterotroph plastid genomes. New Phytol. 2017;214:48–55.
 56. Lin CS, Chen JJW, Huang YT, Chan MT, Daniel H, Chang WJ, et al. The location and translocation of *ndh* genes of chloroplast origin in the Orchidaceae family. Sci Rep. 2015;5:9040.
 57. Lin CS, Chen JJW, Chiu CC, Hsiao HCW, Yang CJ, Jin XH, et al. Concomitant loss of *NDH* complex-related genes within chloroplast and nuclear genomes in some orchids. Plant J. 2017;90:994–1006.
 58. Kim HT, Chase MW. Independent degradation in genes of the plastid *ndh* gene family in species of the orchid genus *Cymbidium* (Orchidaceae: Epidendroideae). PLoS ONE. 2017;12:e0187318.
 59. Niu Z, Pan J, Zhu S, Li L, Xue Q, Liu W, et al. Comparative analysis of the complete plastomes of *Apostasia Wallichii* and *Neuwiedia Singaporeana* (Apostasioideae) reveals different evolutionary dynamics of IR/SSC boundary among photosynthetic orchids. Front Plant Sci. 2017;8:1713.

60. Guisinger MM, Chumley TW, Kuehl JV, Boore JL, Jansen RK. Implications of the plastid genome sequence of *Typha* (Typhaceae, Poales) for understanding genome evolution in Poaceae. *J Mol Evol*. 2010;70:149–66.
61. Walker JF, Zanis MJ, Emery NC. Comparative analysis of complete chloroplast genome sequence and inversion variation in *Lasthenia burkei* (Madieae, Asteraceae). *Am J Bot*. 2014;101:722–9.
62. Yan M, Fritsch PW, Moore MJ, Feng T, Meng A, Yang J, et al. Plastid phylogenomics resolves infrafamilial relationships of the Styracaceae and sheds light on the backbone relationships of the Ericales. *Mol Phylogenet Evol*. 2018;121:198–211.
63. Catalano SA, Saidman BO, Vilardi JC. Evolution of small inversions in chloroplast genome: a case study from a recurrent inversion in angiosperms. *Cladistics*. 2009;25:93–104.
64. Topik H, Weston PH, Yukawa T, Ito M, Rice R. Phylogeny of subtribe Aeridinae (Orchidaceae) inferred from DNA sequences data: advanced analyses including Australasian genera. *J Teknol*. 2012;59:87–95.
65. Li Y, Tong Y, Xing F. DNA barcoding evaluation and its taxonomic implications in the recently evolved Genus *Oberonia* Lindl. (Orchidaceae) in China. *Front Plant Sci*. 2016;7:1–9.
66. Niu Z, Shuying Z, Jiajia P, Ludan L, Jing S, Xiaoyu D. Comparative analysis of *Dendrobium* plastomes and utility of plastomic mutational hotspots. *Sci Rep*. 2017;7:1–11.
67. Smidt EDC, Pérez MZ, Vieira LDN, Viruel J, De Baura VA, Balsanelli E, et al. Characterization of sequence variability hotspots in Cranichideae plastomes (Orchidaceae, Orchidoideae). *PLoS ONE*. 2020;15:e0227991.
68. 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.
69. 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 December 2019:144800.
70. Hong Z, He W, Liu X, Tembrock LR, Wu Z, Xu D, et al. Comparative analyses of 35 complete chloroplast genomes from the genus *Dalbergia* (Fabaceae) and the identification of DNA barcodes for tracking illegal logging and counterfeit rosewood. *Forests*. 2022;13:626.
71. Yan L, Yang M, Guo H, Yang L, Wu J, Li R et al. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat Struct Mol Biol* 2013 209. 2013;20:1131.
72. Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, et al. GetOrganelle: a fast and versatile toolkit for accurate *de novo* assembly of organelle genomes. *Genome Biol*. 2020;21:241.
73. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19:455–77.
74. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of *de novo* genome assemblies. *Bioinformatics*. 2015;31:3350–2.
75. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq-versatile and accurate annotation of organelle genomes. *Nucleic Acids Res*. 2017;45:W6–11.
76. Lowe TM, Chan PP. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res*. 2016;44:W54–7.
77. Kearse 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.
78. Lehwark P, Greiner S. GB2sequin—a file converter preparing custom GenBank files for database submission. *Genomics*. 2019;111:759–61.
79. Lohse M, Drechsel O, Kahlau S, Bock R. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res*. 2013;41:W575–81.
80. Kurtz S. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res*. 2001;29:4633–42.
81. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. *Bioinformatics*. 2017;33:2583–5.
82. Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT. Reordering contigs of draft genomes using the Mauve Aligner. *Bioinformatics*. 2009;25:2071–3.
83. Amiryousefi A, Hyvönen J, Poccai P. IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics*. 2018;34:3030–1.
84. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res*. 2004;32:273–9. Web Server:W.
85. Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, Program NCS, et al. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res*. 2003;13:721–31.
86. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol*. 2017;34:3299–302.
87. Katoh K, Standley DM. MAFFT multiple sequence Alignment Software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30:772–80.
88. Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol*. 2013;30:1188–95.
89. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32:268–74.
90. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Evol Genet Anal*. 2016;33:1870–4.
91. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25:1972–3.
92. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES science gateway for inference of large phylogenetic trees. *Gatew Comput Environ Work GCE*; 2010.
93. Swofford DL. PAUP*. Phylogenetic analysis using parsimony *and other methods. Version 4. Sunderland, MA: Sinauer Associates; 2003.
94. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML web servers. *Syst Biol*. 2008;57:758–71.
95. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 2012;61:539–42.
96. Gao F, Chen C, Arab DA, Du Z, He Y, Ho SY. EasyCodeML: a visual tool for analysis of selection using CodeML. *Ecol Evol*. 2019;9:3891–8.

Publisher's Note

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