ORIGINAL ARTICLE

Accurate authentication of *Dendrobium officinale* and its closely related species by comparative analysis of complete plastomes

Shuying Zhu[a,b,†], Zhitao Niu[a,†], Qingyun Xue[a], Hui Wang[a], Xuezhu Xie[a], Xiaoyu Ding[a,*]

[a]*College of Life Sciences, Nanjing Normal University, Nanjing 210023, China*
[b]*College of Life Sciences, Huzhou University, Huzhou 313000, China*

Abstract Owing to its great medicinal and ornamental values, *Dendrobium officinale* is frequently adulterated with other *Dendrobium* species on the market. Unfortunately, the utilization of the common DNA markers ITS, ITS2, and *matK*+*rbcL* is unable to distinguish *D. officinale* from 5 closely related species of it (*D. tosaense*, *D. shixingense*, *D. flexicaule*, *D. scoriarum* and *D. aduncum*). Here, we compared 63 *Dendrobium* plastomes comprising 40 newly sequenced plastomes of the 6 species and 23 previously published plastomes. The plastomes of *D. officinale* and its closely related species were shown to have conserved genome structure and gene content. Comparative analyses revealed that small single copy region contained higher variation than large single copy and inverted repeat regions, which was mainly attributed to the loss/retention of *ndh* genes. Furthermore, the intraspecific sequence variability among different *Dendrobium* species was shown to be diversified, which necessitates a cautious evaluation of genetic markers specific for different *Dendrobium* species. By evaluating the maximum likelihood trees inferred from different datasets, we found that the complete plastome sequence dataset had the highest discriminatory power for *D. officinale* and its closely related species, indicating that complete plastome sequences can be used to accurately authenticate *Dendrobium* species.

1. Introduction

The genus *Dendrobium*, one of the largest genera in the family Orchidaceae, comprises approximately 1200–1500 species, mainly distributed in tropical and subtropical Asia and eastern Australia[1,2]. There are about 80 species of this genus in China. *Dendrobium* orchids are one of the most well known orchids in global horticultural trade due to their beautiful flowers and ideal characteristics as houseplants[3]. Moreover, many species in this genus have been extensively used as traditional herbal medicine in many Asian countries for hundreds of years[4]. *Dendrobium officinale* Kimura et Migo is a rare and endangered species endemic to China, mainly distributed in southern provinces such as Yunnan, Guizhou, Guangxi, Fujian and Zhejiang[5,6]. However, because of its high medicinal and ornamental values, *D. officinale* has often been adulterated with other *Dendrobium* species on the market[7,8].

Dendrobium species are notoriously difficult to identify due to their similar appearance and tissue structure[9–11]. Usually, their identification relies heavily on morphological methods as well as phytochemical approaches such as capillary electrophoresis (CE)[12] and high-pressure liquid chromatography (HPLC)[13]. Unfortunately, these methods are unable to distinguish *D. officinale* from several *Dendrobium* species, especially 5 closely related species of it (*D. tosaense* Makino; *D. shixingense* Z. L. Chen, S. J. Zeng et J. Duan; *D. flexicaule* Z. H. Tsi, S. C. Sun et L. G. Xu; *D. scoriarum* W. W. Smith; *D. aduncum* Lindl.)[14,15] because of their close affinities.

Recently, molecular techniques have been used in the authentication studies of *Dendrobium*, and a great number of molecular markers have been developed to identify *D. officinale* and its closely related species[3,16–18]. Moreover, a single or a combination of DNA barcodes has also been adopted to infer their species relationships[14,19–23]. However, the identification of *D. officinale* and its close relatives remains difficult because of the high similarities in their genetic backgrounds. For instance, by using 5 DNA markers (ITS, *rbcL, matK, psbA-trnH* and *trnL* intron) to construct the relationships among 109 *Dendrobium* species, the relationships among most of these *Dendrobium* species were clearly settled, yet *D. officinale* and its closely related species were still nested with each other[14]. Xu et al.[23] employed the sequence combination of ITS+*matK* to identify *Dendrobium* species, which showed high specificity for most of the species, but failed to distinguish *D. officinale* and its closely related species. Therefore, it is desirable to develop an effective method for authenticating *D. officinale* and its closely related species.

Chloroplasts, the photosynthetic plastids, are usually uniparentally inherited and have their own genomes called plastomes. In general, plastomes of seed plants have relatively small sizes, conserved gene contents, and dense coding regions as compared with nuclear and mitochondrial genomes[24]. With these unique features, plastomes can be easily assembled and compared[25]. With the advance of next-generation sequencing (NGS), the number of complete plastome sequences has increased rapidly, which provided the opportunity to identify species using the complete plastome sequences. For example, Nock et al.[26] employed complete plastome sequences for the species identification studies of grasses. Using the similar approach, Hu et al.[27] established a phylogeny for the species of *Orychophragmus* and discriminated 6 close relatives in this genus, and Zhang et al.[28] successfully distinguished 9 species

of *Echinacea*. The positive results in these studies implied that the complete plastome sequences had the potential ability to clarify the relationships of *D. officinale* and its closely related species. In addition, the mutational hotspots among *Dendrobium* species were assessed by Niu et al.[29], with 10 loci being put forth as mutational hotspots for *Dendrobium* plastomes. However, it remains to be determined whether these loci are still maintaining the highest degree of sequence variability within species.

In this study, 63 *Dendrobium* plastomes were compared, comprising 40 newly sequenced plastomes of *D. officinale* and its closely related species as well as 23 previously published plastomes. Our aims were: (1) to characterize the plastomes of *D. officinale* and 5 close relatives of it regarding genome structure, gene content, and sequence divergence; (2) to detect the intraspecific plastomic mutational hotspots that are useful for population studies and conservation genetic research; (3) to assess the species discriminatory power of the complete plastome sequences for *D. officinale* and its closely related species. Based on the comparison of different sequence datasets, our results showed that complete plastome sequences could be used for the authentication study of *D. officinale* and its closely related species.

2. Materials and methods

2.1. Plant materials and DNA extraction

A total of 40 representative individuals were collected from the main distribution areas of *D. officinale* and 5 closely related species of it: *D. tosaense*, *D. shixingense*, *D. flexicaule*, *D. scoriarum* (*D. guangxiense*), and *D. aduncum* (Table 1[30,31]). All the plant samples were identified by Prof. Xiaoyu Ding and grown in a greenhouse at Nanjing Normal University, China. Total genomic DNA of each sample was extracted from 2 g fresh leaves using Dneasy Plant Mini Kits (QIAGEN, Germany). DNA samples that met the quality requirements (A260/280 ratio=1.8–2.0, A260/230 ratio >1.7, and DNA concentration >100 ng/μL) were used for sequencing.

2.2. DNA sequencing, plastome assembly, annotation, and PCR-based validation

Paired-end sequencing of 150 bp was conducted on an Illumina Hiseq. 4000 platform, and >6 Gb of sequence data for each sample was obtained. The raw reads were trimmed with an error probability <0.05 and by removing one nucleotide at both terminal ends, and then assembled on CLC Genomics Workbench 8.5.1 (CLC Bio, Aarhus, Denmark) by using the *de novo* assembling method coupled with reference-guided assembling method as described by Niu et al[32]. The plastome of *D. officinale* NC_024019[30] was used as reference. The gaps and 4 junctions between inverted repeat (IR) regions and single copy (SC) regions were confirmed by PCR amplification and Sanger sequencing with specific primers.

The complete plastome sequences were annotated by using the online program DOGMA[33]. The tRNA genes were detected with tRNAscan-SE 1.2[34]. The exact boundaries of each gene were manually checked by comparing them with homologous genes of other plastomes in the genus of *Dendrobium*.

Table 1 Sampling information and plastome characteristics of *D. officinale* and its closely related species.

No.	Species	Code	Location	Plastome length (bp)	LSC length (bp)	SSC length (bp)	IR length (bp)	AT content (%)	Voucher	Accession No.
1	*D. officinale*	DoWS	Wenshan, Yunnan Province	152156	85016	14522	26309	62.52	ZSY01005	LC331062
2	*D. officinale*	DoGN	Guangnan, Yunnan Province	152018	84910	14514	26297	62.50	ZSY01011	LC348520
3	*D. officinale*	DoSP	Shiping, Yunnan Province	152246	85106	14510	26315	62.54	ZSY01017	LC348521
4	*D. officinale*	DoXY	Xingyi, Guizhou Province	152165	85025	14522	26309	62.56	ZSY01213	LC348522
5	*D. officinale*	DoSD	Sandu, Guizhou Province	152029	84919	14516	26297	62.46	ZSY01808	LC348523
6	*D. officinale*	DoTE	Tian'e, Guangxi Province	152042	84920	14522	26300	62.51	ZSY01710	LC348524
7	*D. officinale*	DoGL	Guilin, Guangxi Province	152167	85040	14509	26309	62.52	ZSY01709	LC348525
8	*D. officinale*	DoSG	Shaoguan, Guangdong Province	152224	85094	14514	26308	62.53	ZSY01622	LC348526
9	*D. officinale*	DoHS	Huoshan, Anhui Province	152163	85033	14514	26308	62.52	ZSY01521	LC348527
10	*D. officinale*	DoSY	Shaoyang, Hunan Province	152023	84931	14476	26308	62.51	ZSY01423	LC348528
11	*D. officinale*	DoLN	Longnan, Jiangxi Province	152226	85096	14514	26308	62.53	ZSY01314	LC348529
12	*D. officinale*	DoJG	Jinggangshan, Jiangxi Province	152066	84936	14514	26308	62.51	ZSY01315	LC348530
13	*D. officinale*	DoLC	Liancheng, Fujian Province	152163	85031	14514	26309	62.52	ZSY01807	LC348725
14	*D. officinale*	DoLS	Lishui, Zhejiang Province	152219	85104	14515	26300	62.61	ZSY01918	LC348531
15	*D. officinale*	DoYD	Yandang, Zhejiang Province	152221	85109	14516	26298	62.53		Luo et al.[30]
16	*D. officinale*	DoZJ	Zhejiang Province	152018	84944	14506	26284	62.43		Yang et al.[31]
17	*D. tosaense*	DtTD	Taidong, Taiwan	152253	85094	14507	26321	62.63	ZSY03221	LC348532
18	*D. tosaense*	DtHL	Hualian, Taiwan	152250	85091	14517	26321	62.53	ZSY03226	LC348720
19	*D. tosaense*	DtLN	Longnan, Jiangxi Province	152256	85099	14508	26320	62.53	ZSY03325	LC348721
20	*D. shixingense*	DsDA	Du'an, Guangxi Province	152123	85063	14502	26279	62.50	ZSY09212	LC348726
21	*D. shixingense*	DsGL	Guilin, Guangxi Province	152195	85076	14501	26309	62.56	ZSY09331	LC348722
22	*D. shixingense*	DsLZ	Liuzhou, Guangxi Province	152175	85056	14501	26309	62.52	ZSY09330	LC348723
23	*D. shixingense*	DsCZ	Chenzhou, Hunan Province	152183	85062	14503	26309	62.50	ZSY09688	LC348724
24	*D. shixingense*	DsNX	Nanxiong, Guangdong Province	152184	85063	14503	26309	62.52	ZSY09008	LC348861
25	*D. shixingense*	DsSX	Shixing, Guangdong Province	152178	85057	14503	26309	62.52	ZSY09001	LC348860
26	*D. shixingense*	DsLN	Longnan, Jiangxi Province	152178	85058	14502	26309	62.52	ZSY09335	LC348863
27	*D. shixingense*	DsQN	Quannan, Jiangxi Province	152181	85061	14502	26309	62.52	ZSY09213	LC348862
28	*D. flexicaule*	DfGZ	Ganzi, Sichuan Province	152191	85030	14483	26339	62.55	ZSY05441	LC348965
29	*D. flexicaule*	DfGL	Ganluo, Sichuan Province	152245	85045	14492	26354	62.56	ZSY05225	LC348856
30	*D. flexicaule*	DfSN	Shennongjia, Hubei Province	152252	85049	14495	26354	62.54	ZSY05002	LC348854
31	*D. flexicaule*	DfNZ	Nanzhao, Henan Province	152242	85039	14495	26354	62.53	ZSY05008	LC348855
32	*D. scoriarum*	DgDB	Debao, Guangxi Province	151982	84929	14449	26302	62.52	ZSY02011	LC348847
33	*D. scoriarum*	DgHC	Hechi, Guangxi Province	151992	84933	14455	26302	62.52	ZSY02022	LC348848
34	*D. scoriarum*	DgXL	Xilin, Guangxi Province	151998	84941	14455	26301	62.53	ZSY02035	LC348849
35	*D. scoriarum*	DgFN	Funing, Yunnan Province	151995	84936	14455	26302	62.52	ZSY02215	LC348852
36	*D. scoriarum*	DgWS	Wenshan, Yunnan Province	151994	84938	14454	26301	62.53	ZSY02205	LC348851
37	*D. scoriarum*	DgXC	Xichou, Yunnan Province	151978	84921	14455	26301	62.52	ZSY02238	LC348853
38	*D. scoriarum*	DgXY	Xingyi, Guizhou Province	151990	84934	14442	26307	62.42	ZSY02106	LC348864
39	*D. scoriarum*	DgAL	Anlong, Guizhou Province	151985	84951	14450	26292	62.53	ZSY02118	LC348850
40	*D. aduncum*	DaLF	Luofushan, Guangdong Province	152112	84952	14522	26319	62.52	ZSY06111	LC348858
41	*D. aduncum*	DaTY	Taoyuan, Hunan Province	152104	84944	14522	26319	62.46	ZSY06116	LC348859
42	*D. aduncum*	DaXY	Xingyi, Guizhou Province	152123	84961	14524	26319	62.48	ZSY06088	LC348857

2.3. The analysis of plastomic sequence divergence of D. officinale and its closely related species

In addition to 40 plastomes newly sequenced in this study, two published plastomes of *D. officinale* with the GenBank accession numbers NC_024019[30] and KJ862886[31] were also retrieved for comparative analyses. Full alignments of 42 these plastomes were performed using mVISTA program[35], with *D. officinale* (NC_024019) used as the reference. To estimate the sequence divergence of different regions of the plastomes, sequences of coding regions and non-coding regions (introns, intergenic spacer regions, and pseudogenes) were retrieved from these 42 plastomes. The syntenic loci were aligned using MUSCLE 3.8.31[36] implemented in MEGA 5.2[37]. The sequences of non-coding regions were firstly aligned with the default parameters and then realigned with the "Refining" option. The sequences of protein-coding genes were aligned with the Align Codons option using the default parameters. The gaps located at the 5′- and 3′-ends of alignments were excluded. Then, single nucleotide polymorphisms (SNPs) and insertions/deletions (Indels) were identified by DnaSP v5[38].

2.4. Estimates of sequence variability

The intraspecific-level sequence variability (SV) for each intergenic and intronic locus with the length of more than 150 bp was estimated according to the formula described by Niu et al.[39] as follows: SV (%)=(The number of SNPs+the number of Indel events) / (The number of conserved sites+the number of SNPs+the number of Indel events) × 100. For each locus, there were 120, 3, 28, 12, 28, and 3 pairwise alignments for *D. officinale*, *D. tosaense*, *D. shixingense*, *D. flexicaule*, *D. scoriarum*, and *D. aduncum*, respectively. Finally, we determined the average SV of each syntenic locus for each species.

2.5. Species authentication analyses

To determine whether the complete plastome sequences or commonly used DNA markers have higher discriminatory power for *D. officinale* and its closely related species, a total of 10 DNA regions (ITS, ITS2, *matK*, *rbcL*, *psbA-trnH*, *trnT-trnL*, *rpl32-trnL*, *clpP-psbB*, *trnL* intron, and *rps16-trnQ*) and the complete plastome sequences of 72 samples were employed for the authentication studies (Supplementary Information Table S1). The sequences of DNA markers were aligned with MEGA 5.2[37]. The complete plastome sequences were aligned with MAFFT v7.221[40] and then adjusted manually in MEGA 5.2[37]. All of the gaps and ambiguous sites were removed. Subsequently, these sequences were categorized into 9 datasets: (1) ITS, (2) ITS2, (3) ITS+*matK*, (4) ITS+*psbA-trnH*, (5) ITS2+*rbcL*, (6) *matK*+*rbcL*, (7) ITS+*matK*+*rbcL*, (8) *trnT-trnL*+*rpl32-trnL*+*clpP-psbB*+*trnL* intron+*rps16-trnQ*, and (9) the complete plastome sequences. Maximum likelihood (ML) trees for the 9 datasets were reconstructed using RAxML 8.0.2[41] with the GTRGAMMA model. A thousand bootstrap replicates were executed to estimate the robustness of the ML trees. Moreover, given that different regions of plastomes vary in molecular evolutionary rates[42], we also reconstructed ML trees using the following datasets: (1) the large single copy (LSC) regions, (2) the small single copy (SSC) regions, (3) the IRs, (4) protein-coding genes, and (5) non-coding regions.

2.6. Statistical analyses

Statistical analyses were performed by using SPSS Statistics 22.0.

3. Results

3.1. Plastome features

A total of 40 complete plastomes from *D. officinale* and 5 closely related species of it were sequenced, which came from 14, 3, 8, 4, 8, and 3 individual plants of *D. officinale*, *D. tosaense*, *D.shixingense*, *D. flexicaule*, *D. scoriarum*, and *D. aduncum*, respectively (Table 1[30,31]). Additionally, two published plastomes of *D. officinale* (NC_024019[30] and KJ862886[31]) were also included in comparative analyses. The plastome length of *D. officinale* ranged from 152,018 to 152,246 bp. All of the 16 plastomes displayed the typical quadripartite structure comprising a pair of IRs (26,284–26,315 bp) separated by the LSC (84,910–85,109 bp) and SSC (14,476–14,522 bp) regions. The overall AT content ranged from 62.43% to 62.61%. Each of these plastomes contained 103 unique genes (Fig. 1) consisting of 69 protein-coding genes, 30 tRNA genes, and 4 rRNA genes. Nine pseudogenes were detected, comprising 7 *ndh* remnants (*ψndhA, D, E, F, G, H* and *J*) and 2 incompletely duplicated genes at the IR/SC boundaries (*ψrpl22* and *ψycf1*). The coding regions occupied 53.88–53.97% of the complete plastome. Non-coding regions that were composed of pseudogenes, introns, and intergenic spacers occupied 3.33–3.36%, 11.70–11.74%, and 30.95–31.04% of the plastome sequences, respectively. On the other hand, the plastomes of other 5 species had the lengths of 151,978–152,256 bp, and also possessed highly conserved structure, gene content and order.

The sequences flanking IR/SC junctions were compared between *D. officinale* and its closely related species (Fig. 2). The boundaries of IR/LSC rarely changed among the 6 species. The junctions of IR$_a$/SSC were located in the 5′ end of *ycf1*, which resulted in a duplicated *ψycf1* in the IR$_b$ regions. The length of *ψycf1* was 309 bp consistently in *D. officinale*, *D. tosaense*, and *D. shixingense*, while varing slightly among *D. flexicaule*, *D. scoriarum*, and *D. aduncum* (being 342, 326, and 318 bp, respectively). The IR$_b$/SSC junctions of *D. officinale*, *D. tosaense*, and *D. shixingense* were located upstream of *ψndhF* by 3 bp; whereas the junctions of *D. flexicaule, D. scoriarum*, and *D. aduncum* were expanded into 3′ end of *ψndhF*, resulting in an overlap of *ψycf1* and *ψndhF* by approximately 10 bp. These results revealed that the expansion/contraction of IRs was conversed among *D. officinale* and its closely related species.

3.2. Plastomic sequence divergence

The sequence divergence of 42 plastomes of *D. officinale* and its closely related species were estimated. As expected, the non-coding regions exhibited higher divergence levels than the coding regions (Fig. 3). Based on the plastome-wide investigation, totally 1168 SNPs and 452 Indels were detected among these plastomes (Table 2 and Supplementary information Table S2), with the average densities of 7.6 SNPs per kb and 3.0 Indels per kb across the complete plastome. Most of the variants (585 SNPs and 355 Indels) were located in intergenic spacers. In addition, the SNP and Indel distribution were also compared among the LSC, SSC, and IR regions. The remarkably higher SNP and Indel densities in SSC region
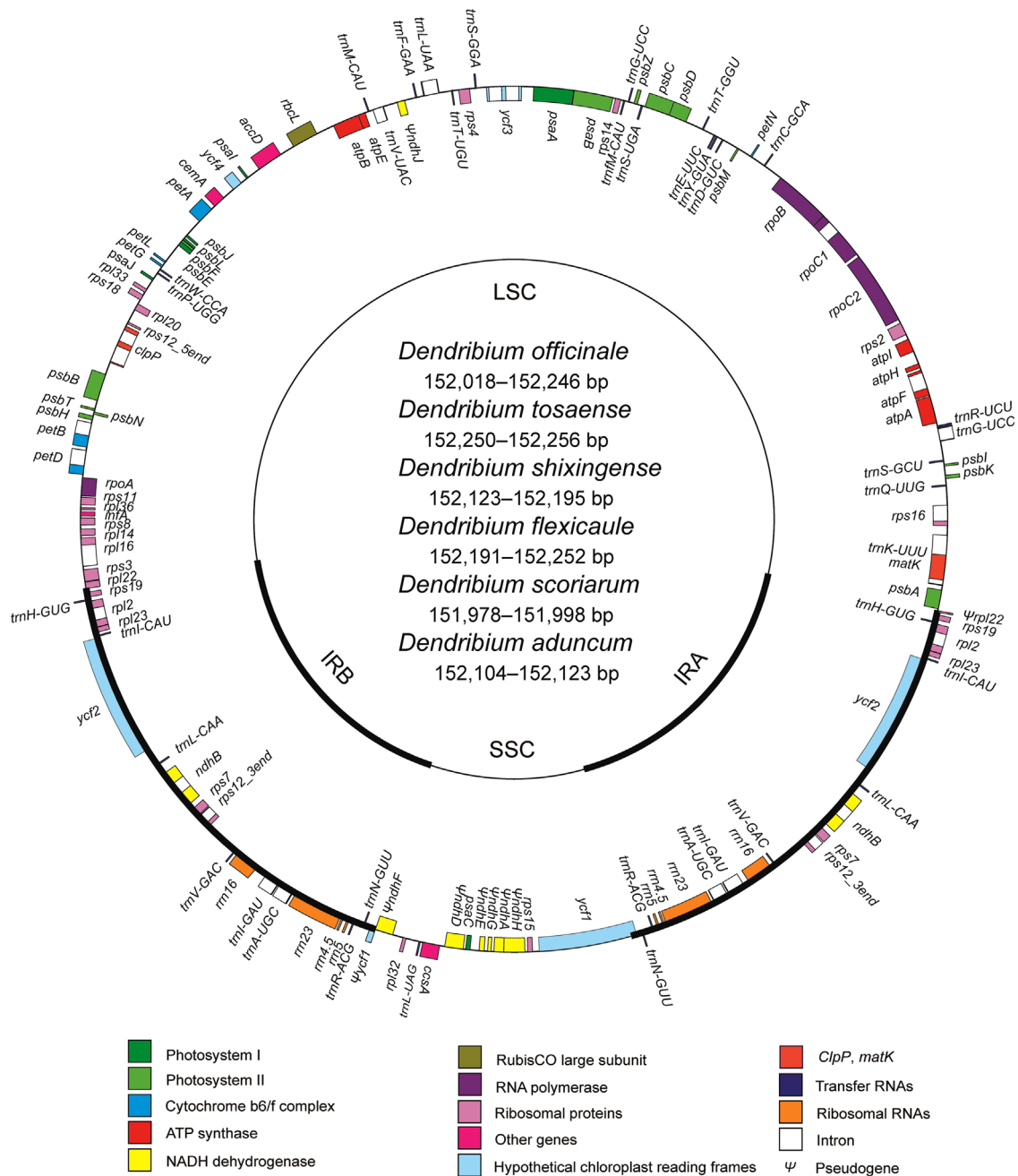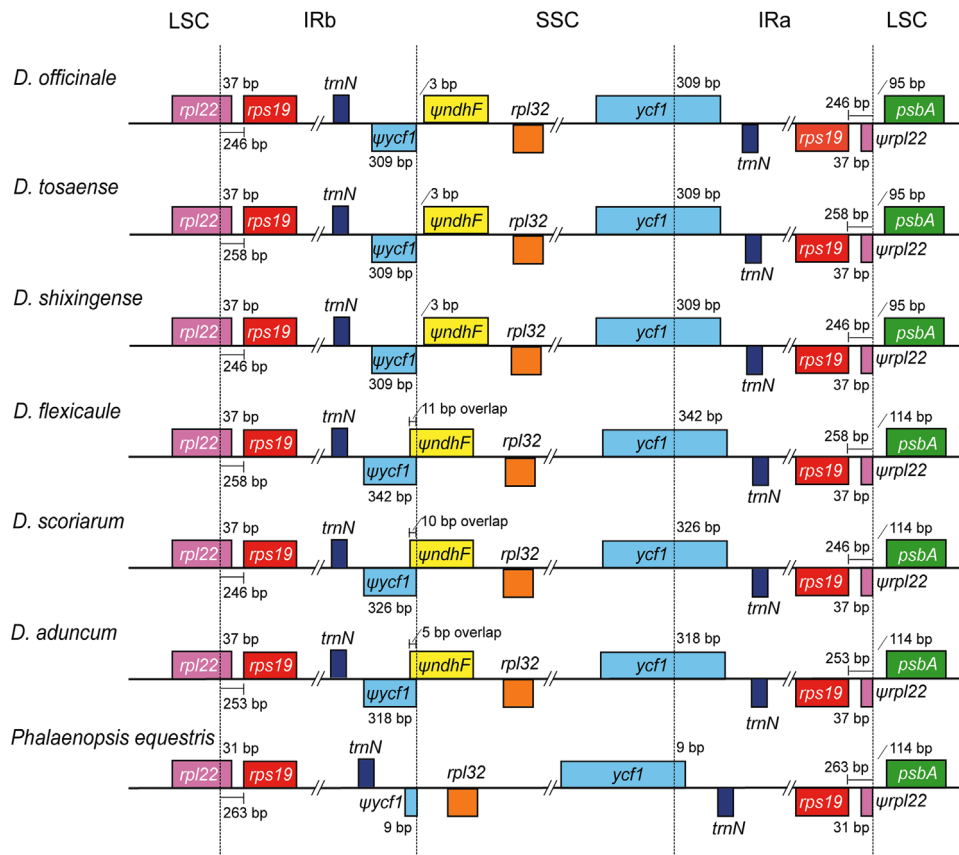
Figure 1 Plastome map of *D. officinale* and its closely related species. The genes outside and inside the circle are transcribed clockwise and counterclockwise, respectively. The length of the plastomes among the six species ranges from 151,978 to 152,256 bp.

indicated that the variation of SSC was higher than that of LSC and IR regions. The drastic variation of SSC region might be related to the independent loss/retention of *ndh* genes. Therefore, in order to determine the effects of the loss/retention of *ndh* genes on the variation of SSC, we divided the SSC sequences into 3 parts: *ndh* pseudogenes, intergenic spacer regions adjacent to *ndh* pseudogenes, and the other regions (Table 3). Our comparison results showed a higher density and number of SNPs and Indels in *ndh* pseudogenes and their adjacent intergenic spacer regions, suggesting that the loss/retention of *ndh* genes played an important role in causing the high variation of SSC.

3.3. Diversified evolution of plastomic sequence variability among Dendrobium species

Recent studies showed diversified evolution of the plastome sequence among different orchid genera[39]. In this study, to determine whether the evolution of SV was conserved within the genus of *Dendrobium*, we estimated the pairwised intraspecific SV of 90 syntenic non-coding loci including intergenic spacers and introns of *D. officinale* and its closely related species (Supplementary Information Table S3). Moreover, we also performed correlation tests for the 90 non-coding loci between the 5 species. Although they are closely related to each

Figure 2 Comparison of the regions flanking IR/SC junctions among *D. officinale* and its closely related species. The plastome of *Phalaenopsis equestris* is used as the reference.

other, our results showed that the SV values were statistically uncorrelated ($P > 0.05$) or weakly correlated (Spearman's $r = 0.208$ to 0.458, $P < 0.05$) between the 5 species. These results suggested that the evolution of SV among different *Dendrobium* species was variable.

The top-12 mutational hotspots that contain the highest intraspecific variability (SV > 0.1%) and the lowest interspecific variability (SV < 3%) of *D. officinale* were shown in Fig. 4[29]. However, with the comparison of the SV values of other 4 species, our results have shown that: (1) none of top-12 hotspots is common to all the 5 species; (2) only 3 of them (*rps2-rpoC2*, *accD-psaI*, and *matK-5′trnK*) are present in more than 3 species. This pointed to the fact that the plastomic mutational hotspots for the intraspecific-level studies of *Dendrobium* species were diversified.

3.4. Species authentication analyses

DNA barcodes or plastome mutational hotspots, such as ITS[43] ITS2[21], ITS+*matK*[23], and *trnT-trnL*+*rpl32-trnL*+*clpP-psbB*+*trnL* intron+*rps16-trnQ*[29], were demonstrated to have high discriminatory power for most *Dendrobium* species. However, as shown in Fig. 5 and Supplementary Information Fig. S1, they all failed to distinguish *D. officinale* and its closely related species. For example, in the ML tree inferred from ITS2 dataset, *D. officinale*, *D. tosaense*, *D. shixingense*, and *D. scoriarum* were nested with each other. Though the species of *D. flexicaule* formed a monophyletic group in the ML trees based on the ITS and ITS2+*rbcL* datasets, the support values were below 50%.

By contrast, our analyses using the complete plastome dataset achieved satisfactory results as follows: (1) the phylogenetic tree yielded high resolution (bootstrap value (BS) > 85%) for all tree nodes with few exceptions within some species; (2) the individuals of each of the 6 species were resolved as a monophyletic group. Furthermore, the complete plastome dataset also showed higher resolution than other plastome-scale datasets, *i.e.*, LSC, SSC, IRs, protein-coding genes, and non-coding sequences (Supplementary Information Fig. S2). These results indicated that the complete plastome sequences could be used to identify *D. officinale* and its closely related species.

4. Discussion

4.1. Variations in the SSC regions of D. officinale and its closely related species were mainly contributed by the loss/retention of ndh genes

Although the plastomes of land plants are conserved in terms of genome structure and gene content, some structure changes (*e.g.*, inversions, rearrangements, and IR expansion/contraction) and the sequence variations have been detected in complete plastomes[24]. In this study, while the plastomes of *D. officinale* and its closely related species were found to be highly conserved in plastome structure, and gene content and order, sequence variations (SNPs and Indels) were also observed in these plastomes. SNP loci and Indel events are very useful resources for phylogenetic analysis and species identification[25]; SNPs have
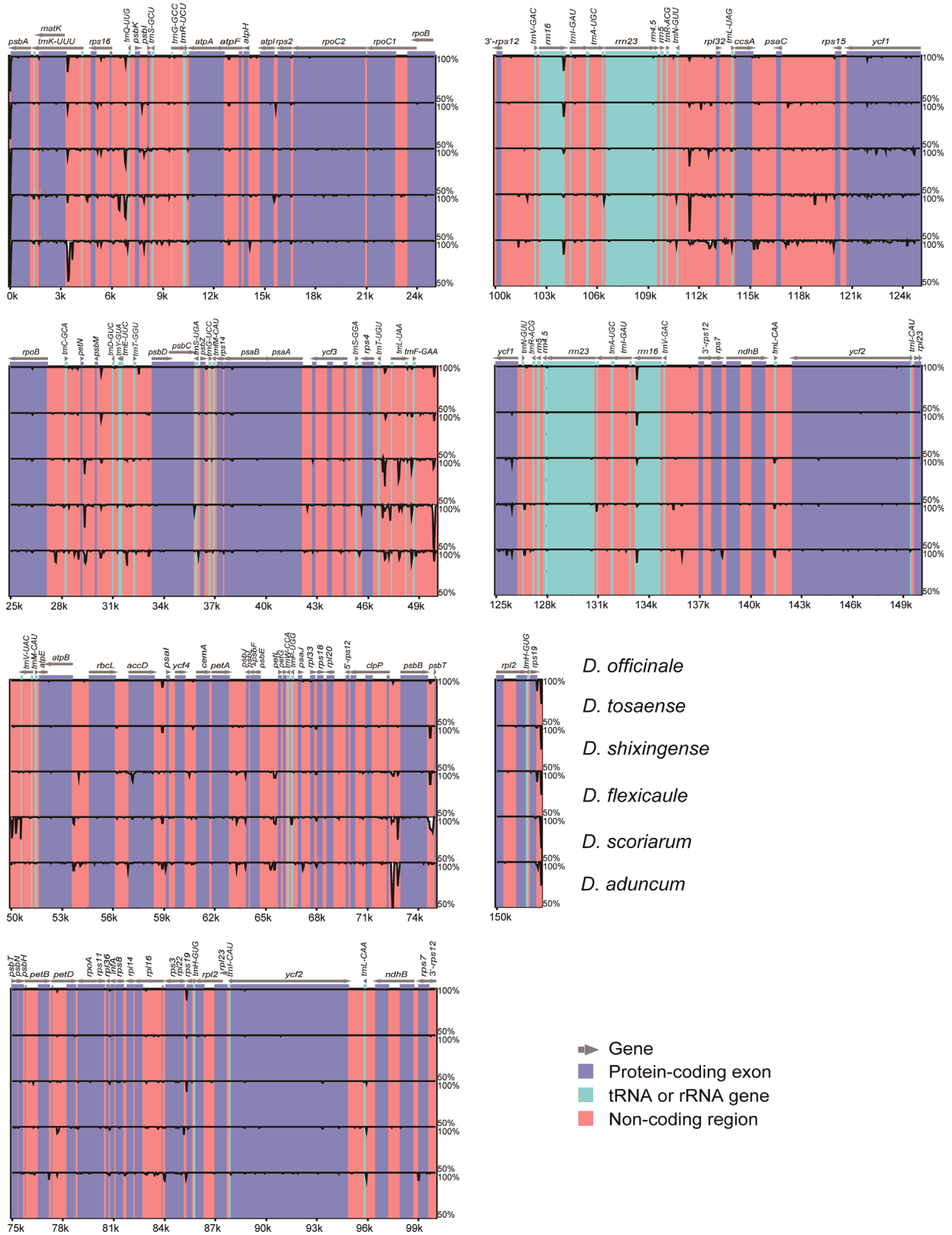
Figure 3 Sequence identity plots among the plastomes of *D. officinale* and its closely related species with *D. officinale* (NC_024019) sequence as the reference by using mVISTA. Each species is represented by one accession.

Table 2   The SNPs and Indels in 42 complete plastomes of *D. officinale* and its closely related species.

Region	Coding region		Intergenic spacer		Intron		Pseudogene		Summary	
	SNP	Indel	SNP	Indel	SNP	Indel	SNP	Indel	SNP	Indel
LSC	174 (4.0)	3 (0.07)	425 (13.6)	280 (9.0)	112 (10.6)	42 (4.0)	7 (15.9)	2 (4.6)	718 (8.4)	327 (3.8)
SSC	105 (15.1)	6 (0.9)	104 (31.7)	33 (10.1)	–	–	91 (20.9)	24 (5.6)	300 (20.6)	63 (4.3)
IRs	66 (2.1)	10 (0.3)	56 (4.2)	42 (3.1)	26 (3.5)	10 (1.4)	2 (5.8)	0	150 (2.8)	62 (1.2)
Total	345 (4.2)	19 (0.2)	585 (12.2)	355 (7.4)	138 (7.7)	52 (2.9)	100 (19.5)	26 (5.1)	1168 (7.6)	452 (3.0)

The figures in the bracket indicate the density (per kb) of SNPs or Indels in a target region.

Table 3   The SNPs and Indels in SSC regions of *D. officinale* and its closely related species.

Region	SNP	Indel
Ndh pseudogene	91 (21.1)	24 (5.6)
Adjacent regions of *ndh* pseudogene	77 (38.3)	20 (10.0)
Other regions	132 (16.0)	19 (2.3)
Ndh pseudogene + adjacent regions of *ndh* pseudogene	168 (26.6)	44 (7.0)

The figures in the bracket indicate the density (per kb) of SNPs or Indels in a target region.

been successfully used to infer the phylogenetic relationships within the genus of *Citrus*[44], and Indels have been employed to design specific primers for authenticating buckwheat species[45]. Recently, Niu et al.[29] revealed a nonrandom location of Indels in *Dendrobium* plastomes. In line with their findings, the present study demonstrated higher densities of SNPs and Indels in SSC than in LSC and IR regions, pointing to the fact that both Indels and SNPs were nonrandomly distributed in plastomes of *D. officinale* and its closely related species.

The remarkably higher SNP and Indel densities in SSC might be caused by 2 factors: (1) the drastic expansion/contraction of IRs and (2) the independent loss/retention of *ndh* genes. The expansion/contraction of IRs is known to be largely responsible for the variation of plastomes in different orchids[46,47] and many other species[48–50]. However, the current research showed that only slight changes occurred in the sequences flanking IR/SC boundaries among the plastomes of *D. officinale* and its closely related species, suggesting that the effect of expansion/contraction of IRs on the variation of SSC was negligible. This signified that the loss/retention of *ndh* genes was a potential determinant for the variation of SSC. Recent studies revealed an independent loss/retention of *ndh* genes in SSC region of *Dendrobium* plastomes[29], which led us to expect a higher variation in their located regions. Indeed, we found that *ndh* pseudogenes and their adjacent intergenic spacer regions possessed a higher density and a larger number of variants (SNPs and Indels) than the other regions in SSC. Therefore, we infer that the loss/retention of *ndh* genes mainly accounted for the variation of SSC.

4.2. Plastome-wide comparison is required for studies of orchid species at different levels

Mutational hotspots of plastome sequences are the most useful tools for the systematic and phylogeographic studies at the different taxonomic levels. However, Shaw et al.[51] showed that the hotspot regions were diverse among different plant lineages. For orchid species, different mutational hotspots have been proposed for different orchid genera, *i.e.*, the loci of *rpl32-trnL*, *trnH-psbA*, *trnE-trnT*, *trnK-rps16*, and *trnT-trnL* were employed for the species identification study of *Cymbidium*[52]; and the loci of *trnS-trnG*, *psaC-ndhE*, *clpP-psbB*, *rpl16* intron, *rpoB-trnC*, *trnT-psbD*, *rbcL-accD*, *rpl32-trnL*, *ccsA-ndhD*, and *ndhC-trnV* were listed as the top-10 mutational hotspots for the genus of *Phalaenopsis*[51]. Recently, on the basis of comprehensive plastome-wide comparison, Niu et al.[39] found that the mutational hotspots for orchid species were genus specific; they also proposed the most valuable hotspot combination for the interspecific-level studies of *Dendrobium* genus[29]. However, the mutational hotspots for the intraspecific-level studies of *Dendrobium* species were still uncertain. For example, the loci of *trnC-petN* and *trnE-trnT* were employed to investigate the phylogeographic relationship among different population of *D. moniliforme*[53], while the loci of *accD-psaI*, *trnC-petN*, and *rps15-ycf1* were used to infer the phylogeographic history of *D. officinale*[15]. Thus, aiming to identify the mutational hotspots or combinations applicable for the intraspecific-level studies of *Dendrobium* species, we estimated the pairwised intraspecific SV of 90 syntenic non-coding loci of *D. officinale* and its closely related species. Nevertheless, our careful examination failed to select a common mutational hotspot for the 5 species *D. officinale*, *D.shixingense*, *D. flexicaule*, *D. scoriarum*, and *D. aduncum*, although they are closely related to each other. Our results indicated that the mutational hotspots for the intraspecific-level studies of *Dendrobium* species were diversified. Therefore, it is necessary to make a cautious evaluation of genetic markers specific for different *Dendrobium* species. Considering diversified mutational hotspots among different orchid genera and different species, we proposed that the plastome-wide comparison is required for studies of orchid species at different levels.

4.3. Complete plastome sequences can be used for the authentication of D. officinale and its closely related species

The accurate identification of medicinal plants is essential to their safe utilization and genetic resource conservation.
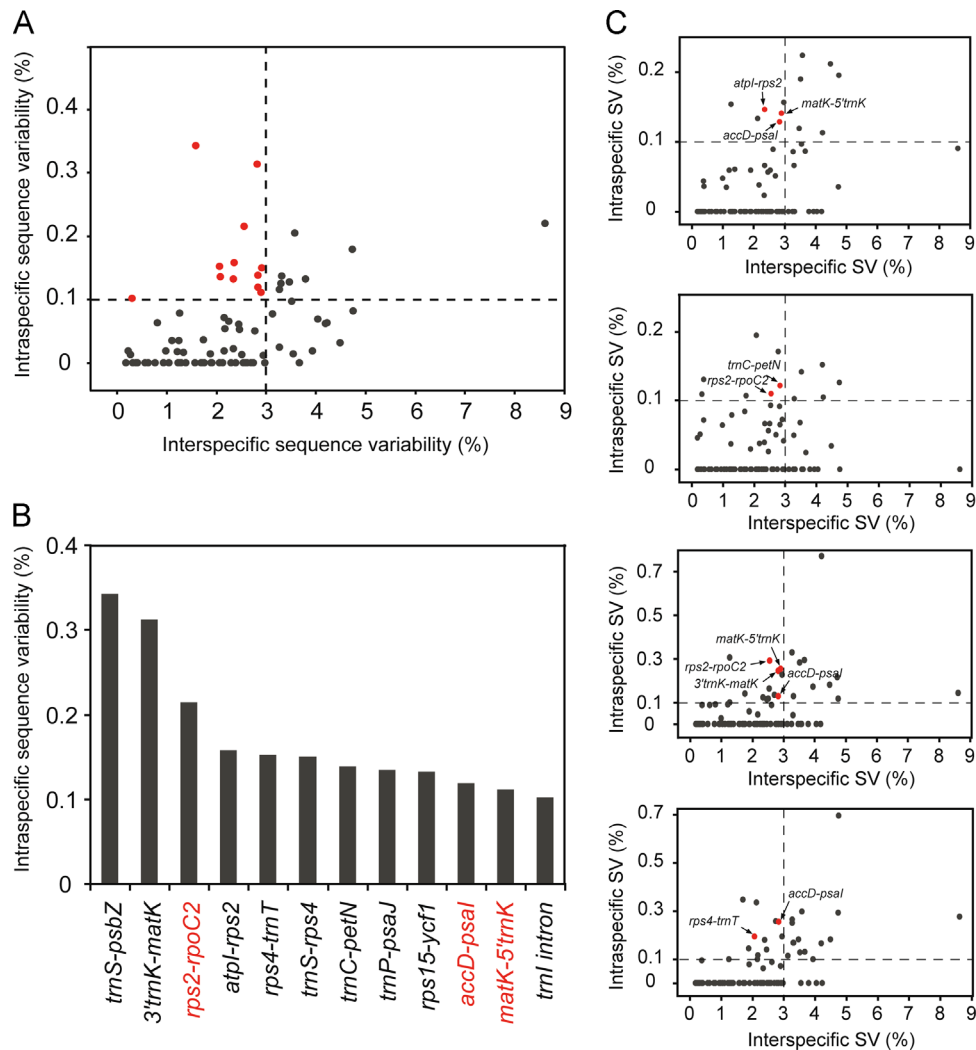
Figure 4 The pairwised intraspecific sequence variability (SV) of 90 syntenic non-coding loci among *D. officinale* and its closely related species. (A) The comparison of interspecific and intraspecific SV values for the loci in the plastome of *D. officinale*. The interspecific SV values of these loci were obtained from Niu et al[29]. (B) The twelve intergenic and intronic loci with the highest SV values in the plastome of *D. officinale*. (C) The comparison of interspecific and intraspecific SV values for the loci in the plastome of *D. shixingense*, *D. flexicaule*, *D. scoriarum* and *D. aduncum*, respectively. The intraspecific SV values of *D. tosaense* were excluded because of the limited variable sites. Only three loci of *rps2-rpoC2*, *accD-psaI* and *matK-5′trnK* (in red) are present in more than three species.

However, *Dendrobium* species have been documented to be difficult for discrimination analysis due to their similar appearance and tissue structure[9–11]. Among them, *D. officinale* and its closely related species are a particularly difficult group for identification studies and phylogenetic analyses due to the overlapping morphological variations[14,54,55], complex evolutionary histories[15], and the lack of effective molecular markers. Previous molecular identification and phylogenetic studies about this group were generally focused on one or a few DNA regions, but none of them successfully resolved relationships among the 6 species, which was mainly due to the inadequate variations provided by a limited number of DNA loci[14,19–23]. Recently, complete plastome sequences containing massive variable sites have been successfully applied to authenticate species and resettle the phylogenetic relationships in taxonomically difficult groups, such as *Orychophragmus*[27], *Cymbidium*[52], and *Schima*[56]. Thus, to overcome the disadvantage of lacking variations for DNA

regions, the complete plastome was exploited to authenticate *D. officinale* and its closely related species and to resolve their relationships.

In comparison with previous studies, we proposed that the method of discriminating *D. officinale* and its closely related species by using the complete plastome sequences was highly reliable, accurate and feasible, as demonstrated by 3 facts as follows. Firstly, the relationships among *D. officinale* and its closely related species were resolved with high support values (BS > 85%). As mentioned above, the complete plastome could provide sufficient informative sites, which can help to infer robust relationships for intractable groups at low taxonomic levels[52,56,57]. Secondly, the monophyletic groups of *D. officinale*, *D. tosaense*, *D. shixingense*, *D. flexicaule*, *D. scoriarum*, and *D. aduncum* resolved in this study were based on a comprehensive taxon sampling, which included representative individuals of documented main distribution areas for the 6 species. In previous studies, the sampling of these species was relatively limited. Moreover,
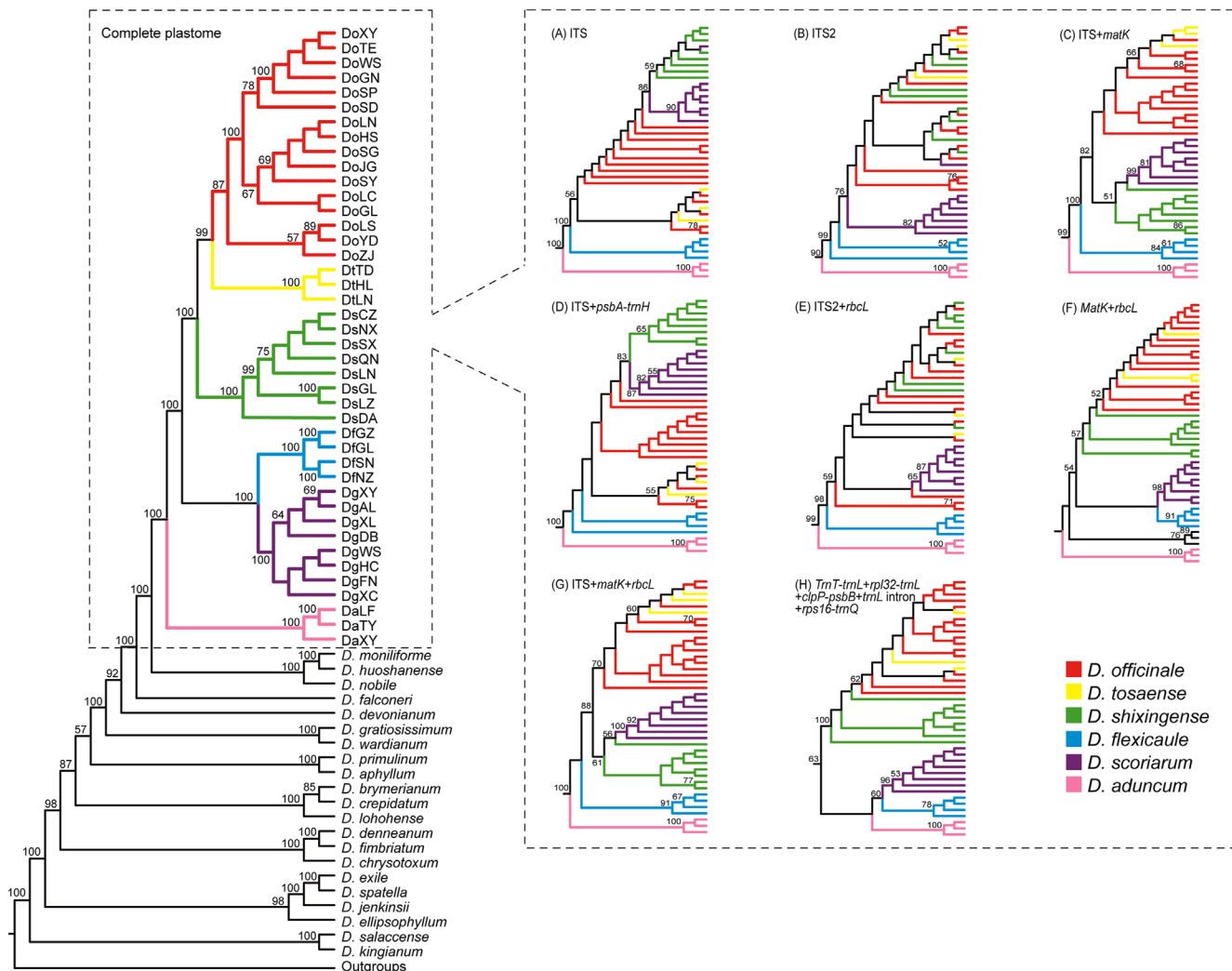
Figure 5 ML trees of *D. officinale* and its closely related species inferred from the complete plastome (left) and commonly used DNA markers (right). Each of the six species is color coded. Numbers near the nodes are bootstrap support values (only values >50% are shown).

D. shixingense (a recently reported species)[55] or *D. tosaense* was always not included. In addition, other 21 *Dendrobium* species, which were often used as the adulterants of *D. officinale* on the marker, were also sampled in this study. Thirdly, it is becoming simple and relatively inexpensive to obtain the complete plastome. With the development of NGS technologies, the sequencing cost has fallen sharply. Furthermore, the approach of assembling plastome from genomic sequencing data has led to a convenient way to generate complete plastome sequences[26]. Therefore, the use of complete plastome is a promising way for authenticating *Dendrobium* species.

Dendrobium species are famous for their great medical values, which has led to many adulterants sold as the Material Medica from *Dendrobium*. Unfortunately, the use of many effective DNA markers, such as *rbcL*, *matK* and even the sequences of ITS or ITS2, could not effectively identify the Material Medica from *Dendrobium*, especially for *D. officinale* and its closely related species because of their close genetic relationships. In this study, we successfully authenticated *D. officinale* and its closely related species by using their complete plastome sequences, indicating that the complete plastome could be considered an efficient super-barcode[58,59]

for the authentication of Material Medica from *Dendrobium*. Compared to conventional barcoding approaches, using the complete plastome as super-barcode has many advantages, including higher accuracy in taxonomically difficult groups, and universality, as it does not require the use of taxon-specific primers[60,61]. On the other hand, the main challenges of super-barcoding are the establishment of a rich plastome database and the improvement of processing power and analytical capacity for big data. Nevertheless, with the rapid advancement of molecular technologies and methodologies, we believe that it will soon be practical to apply the complete plastomes to the authentication studies of most land plants.

5. Conclusions

In conclusion, this is the first study to authenticate the taxonomically difficult group of *D. officinale* and its closely related species. Firstly, we investigated the relationship between plastome variations and loss/retention of *ndh* genes, and found that the variations in the SSC regions of *D. officinale* and its closely related species were mainly contributed by the loss/retention of *ndh* genes. Then,

based on plastome-wide comparison, our analysis revealed the diversified evolution of SV among these plastomes, signifying that the plastome-wide comparison is essential for studies of *Dendrobium* species at different levels. Most importantly, after having carefully examined the ML trees inferred from different sequence datasets, we proposed that *D. officinale* and its closely related species could be unequivocally distinguished using the complete plastome sequences.

Acknowledgments

Appendix A.   Supporting information

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.apsb.2018.05.009.

References

1. Wood HP. *The Dendrobiums*. Ruggell: ARG Gantner Verlag; 2006.
2. Zhu GH, Ji ZH, Wood JJ, Wood HP. *Flora of China*. Beijing: Scientific Press; 2009.
3. Teixeira da Silva JA, Jin X, Dobránszki J, Lu J, Wang H, Zotz G, et al. Advances in *Dendrobium* molecular research: applications in genetic variation, identification and breeding. *Mol Phylogenet Evol* 2016;**95**:196–216.
4. Bao XS, Shun QS, Chen LZ. *The medicinal plants of Dendrobium (Shi-hu) in China*. Shanghai: Shanghai Medicinal University Press and Fudan University Press; 2001.
5. Tsi ZH. *Flora Reipublicae Popularis Sinicae*. Beijing: Science Press; 1999.
6. Li X, Ding X, Chu B, Zhou Q, Ding G, Gu S. Genetic diversity analysis and conservation of the endangered Chinese endemic herb *Dendrobium officinale* Kimura et Migo (Orchidaceae) based on AFLP. *Genetica* 2008;**133**:159–66.
7. Xu H, Hou B, Zhang J, Min T, Yuan Y, Niu Z, et al. Detecting adulteration of *Dendrobium officinale* by real-time PCR coupled with ARMS. *Int J Food Sci Technol* 2012;**47**:1695–700.
8. Ding G, Zhang D, Feng Z, Fan W, Ding X, SNP Li X. ARMS and SSH authentication of medicinal *Dendrobium officinale* KIMURA et MIGO and application for identification of Fengdou Drugs. *Biol Pharm Bull* 2008;**31**:553–7.
9. Zhang YB, Pui-Hay BP, Wang ZT, Shaw PC. Current approaches for the authentication of medicinal *Dendrobium* species and its products. *Plant Genet Resour* 2005;**3**:144–8.
10. Niu Z, Pan J, Xue Q, Zhu S, Liu W, Ding X. Plastome-wide comparison reveals new SNV resources for the authentication of *Dendrobium huoshanense* and its corresponding medicinal slice (Huoshan Fengdou). *Acta Pharm Sin B* 2018;**8**:466–77.
11. Yukawa T, Uehara K. Vegetative diversification and radiation in subtribe Dendrobiinae (Orchidaceae): evidence from chloroplast DNA phylogeny and anatomical characters. *Plant Syst Evol* 1996;**201**:1–14.
12. Zha XQ, Luo JP, Wei P. Identification and classification of *Dendrobium candidum* species by fingerprint technology with capillary electrophoresis. *S Afr J Bot* 2009;**75**:276–82.
13. Yang L, Wang Z, Xu L. Simultaneous determination of phenols (bibenzyl, phenanthrene, and fluorenone). *Dendrobium* species by high-performance liquid chromatography with diode array detection. *J Chromatogr A*. 2006;**1104**:230–7.
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. Hou BW, Luo J, Zhang YS, Niu ZT, Xue QY, Ding XY. Iteration expansion and regional evolution: phylogeography of *Dendrobium officinale* and four related taxa in southern China. *Sci Rep* 2017;**7**:43525.
16. Shen J, Ding X, Liu D, Ding G, He J, Li X, et al. Intersimple sequence repeats (ISSR) molecular fingerprinting markers for authenticating populations of *Dendrobium officinale* Kimura et Migo. *Biol Pharm Bull* 2006;**29**:420–2.
17. Feng SG, Lu JJ, Gao L, Liu JJ, Wang HZ. Molecular phylogeny analysis and species identification of *Dendrobium* (Orchidaceae) in China. *Biochem Genet* 2014;**52**:127–36.
18. Kang JY, Lu JJ, Qiu S, Chen Z, Liu JJ, Wang HZ. *Dendrobium* SSR markers play a good role in genetic diversity and phylogenetic analysis of Orchidaceae species. *Sci Hort* 2015;**183**:160–6.
19. Ding XY, Wang ZT, Hong Xu, Xu LS, Zhou KY. Database establishment of the whole rDNA ITS region of *Dendrobium* species of "Fengdou" and authentication by analysis of their sequences. *Acta Pharm Sin* 2002;**37**:567–73.
20. Takamiya T, Wongsawad P, Sathapattayanon A, Tajima N, Suzuki S, Kitamura S, et al. Molecular phylogenetics and character evolution of morphologically diverse groups, *Dendrobium* section *Dendrobium* and allies. *AoB Plants* 2014;**6**:plu045.
21. Feng S, Jiang Y, Wang S, Jiang M, Chen Z, Ying Q, et al. Molecular identification of *Dendrobium* species (Orchidaceae) based on the DNA barcode ITS2 region and its application for phylogenetic study. *Int J Mol Sci* 2015;**16**:21975–88.
22. Xiang XG, Mi XC, Zhou HL, Li JW, Chung SW, Li DZ, et al. Biogeographical diversification of mainland Asian *Dendrobium* (Orchidaceae) and its implications for the historical dynamics of evergreen broad-leaved forests. *J Biogeogr* 2016;**43**:1310–23.
23. Xu S, Li D, Li J, Xiang X, Jin W, Huang W, et al. Evaluation of the DNA barcodes in *Dendrobium* (Orchidaceae) from mainland Asia. *PLoS One* 2015;**10**:e0115168.
24. Raubeson LA, Jansen RK. *Chloroplast genomes of plants*. In: Henry RJ, editor. *Plant diversity and evolution: genotypic and phenotypic variation in higher plants*. London: CAB International; 2005. p. 45–68.
25. Tonti-Filippini J, Nevill PG, Dixon K, Small I. What can we do with 1000 plastid genomes?. *Plant J* 2017;**90**:808–18.
26. Nock CJ, Waters DL, Edwards MA, Bowen SG, Rice N, Cordeiro GM, et al. Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnol J* 2011;**9**:328–33.
27. Hu H, Hu Q, Al-Shehbaz IA, Luo X, Zeng T, Guo X, et al. Species delimitation and interspecific relationships of the genus *Orychophragmus* (Brassicaceae) inferred from whole chloroplast genomes. *Front Plant Sci* 2016;**7**:1826.
28. Zhang N, Erickson DL, Ramachandran P, Ottesen AR, Timme RE, Funk VA, et al. An analysis of *Echinacea* chloroplast genomes: implications for future botanical identification. *Sci Rep* 2017;**7**:216.
29. Niu Z, Zhu S, Pan J, Li L, Sun J, Ding X. Comparative analysis of *Dendrobium* plastomes and utility of plastomic mutational hotspots. *Sci Rep* 2017;**7**:2073.
30. Luo J, Hou BW, Niu ZT, Liu W, Xue QY, Ding XY. Comparative chloroplast genomes of photosynthetic orchids: insights into evolution of the Orchidaceae and development of molecular markers for phylogenetic applications. *PLoS One* 2014;**9**:e99016.
31. Yang P, Zhou H, Qian J, Xu H, Shao Q, Li Y, et al. The complete chloroplast genome sequence of *Dendrobium officinale*. *Mitochondrial DNA A DNA Mapp Seq Anal* 2016;**27**:1262–4.
32. Niu Z, Xue Q, Wang H, Xie X, Zhu S, Liu W, et al. Mutational biases and GC-biased gene conversion affect GC content in the plastomes of *Dendrobium* genus. *Int J Mol Sci* 2017;**18**:2307.
33. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 2004;**20**:3252–5.

34. Schattner P, Brooks AN, Lowe TM. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res* 2005;**33**:W686–9.

35. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res* 2004;**32**:W273–9.

36. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004;**32**:1792–7.

37. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011;**28**:2731–9.

38. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 2009;**25**:1451–2.

39. Niu Z, Xue Q, Zhu S, Sun J, Liu W, Ding X. The complete plastome sequences of four orchid species: insights into the evolution of the Orchidaceae and the utility of plastomic mutational hotspots. *Front Plant Sci* 2017;**8**:715.

40. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;**30**:772–80.

41. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;**30**:1312–3.

42. Palmer JD. Comparative organization of chloroplast genomes. *Annu Rev Genet* 1985;**19**:325–54.

43. Chattopadhyay P, Banerjee G, Banerjee N. Distinguishing orchid species by DNA barcoding: increasing the resolution of population studies in plant biology. *OMICS* 2017;**21**:711–20.

44. Carbonell-Caballero J, Alonso R, Ibañez V, Terol J, Talon M, Dopazo J. A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus *Citrus*. *Mol Biol Evol* 2015;**32**:2015–35.

45. Cho KS, Yun BK, Yoon YH, Hong SY, Mekapogu M, Kim KH, et al. Complete chloroplast genome sequence of tartary buckwheat (*Fagopyrum tataricum*) and comparative analysis with common buckwheat (*F. esculentum*). *PLoS One* 2015;**10**:e0125332.

46. Kim HT, Kim JS, Moore MJ, Neubig KM, Williams NH, Whitten WM, et al. Seven new complete plastome sequences reveal rampant independent loss of the *ndh* gene family across orchids and associated instability of the inverted repeat/small single-copy region boundaries. *PLoS One* 2015;**10**:e0142215.

47. 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 singapureana* (Apostasioideae) reveals different evolutionary dynamics of IR/SSC boundary among photosynthetic orchids. *Front Plant Sci* 2017;**8**:1713.

48. Wu CS, Lai YT, Lin CP, Wang YN, Chaw SM. Evolution of reduced and compact chloroplast genomes (cpDNAs) in gnetophytes: selection toward a lower-cost strategy. *Mol Phylogenet Evol* 2009;**52**:115–24.

49. Wu CS, Wang YN, Hsu CY, Lin CP, Chaw SM. Loss of different inverted repeat copies from the chloroplast genomes of Pinaceae and Cupressophytes and influence of heterotachy on the evaluation of gymnosperm phylogeny. *Genome Biol Evol* 2011;**3**:1284–95.

50. Sanderson MJ, Copetti D, Búrquez A, Bustamante E, Charboneau JL, Eguiarte LE, et al. Exceptional reduction of the plastid genome of saguaro cactus (*Carnegiea gigantea*): loss of the *ndh* gene suite and inverted repeat. *Am J Bot* 2015;**102**:1115–27.

51. 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.

52. 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.

53. Ye M, Liu W, Xue Q, Hou B, Luo J, Ding X. Phylogeography of the endangered orchid *Dendrobium moniliforme* in East Asia inferred from chloroplast DNA sequences. *Mitochondrial DNA A DNA Mapp Seq Anal* 2017;**28**:880–91.

54. Jin X, Chen S, Luo Y. Taxonomic revision of *Dendrobium moniliforme* complex (Orchidaceae). *Sci Hort* 2009;**120**:143–5.

55. Chen ZL, Zeng SJ, Wu KL, Duan J. *Dendrobium shixingense* sp. nov. (Orchidaceae) from Guangdong, China. *Nord J Bot* 2010;**28**:723–7.

56. Yu XQ, Drew BT, Yang JB, Gao LM, Li DZ. Comparative chloroplast genomes of eleven *Schima* (Theaceae) species: insights into DNA barcoding and phylogeny. *PLoS One* 2017;**12**:e0178026.

57. Li P, Zhang S, Li F, Zhang S, Zhang H, Wang X, et al. A phylogenetic analysis of chloroplast genomes elucidates the relationships of the six economically important *Brassica* species comprising the triangle of U. *Front Plant Sci* 2017;**8**:111.

58. 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.

59. Hollingsworth PM, Li DZ, van der Bank M, Twyford AD. Telling plant species apart with DNA: from barcodes to genomes. *Philos Trans R Soc Lond B Biol Sci* 2016;**371**:20150338.

60. 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.

61. Coissac E, Hollingsworth PM, Lavergne S, Taberlet P. From barcodes to genomes: extending the concept of DNA barcoding. *Mol Ecol* 2016;**25**:1423–8.