RESEARCH ARTICLE

Evaluation of the DNA Barcodes in *Dendrobium* (Orchidaceae) from Mainland Asia

Songzhi Xu[1], Dezhu Li[2], Jianwu Li[3], Xiaoguo Xiang[1], Weitao Jin[1], Weichang Huang[4], Xiaohua Jin[1]*, Luqi Huang[5]*

1 State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, P. R. China, 2 Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650201, China, 3 Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun Township, Mengla County, Yunnan 666303, China, 4 Shanghai Chenshan Botanical Garden, Chenhua Road 3888, Songjiang, Shanghai 201602, China, 5 National Resource Centre for Chinese Material Medica, China Academy of Chinese Medical Science, Beijing 100700, China

* xiaohuajin@ibcas.ac.cn (XJ); huangluqi01@126.com (LH)

Abstract

DNA barcoding has been proposed to be one of the most promising tools for accurate and rapid identification of taxa. However, few publications have evaluated the efficiency of DNA barcoding for the large genera of flowering plants. *Dendrobium*, one of the largest genera of flowering plants, contains many species that are important in horticulture, medicine and biodiversity conservation. Besides, *Dendrobium* is a notoriously difficult group to identify. DNA barcoding was expected to be a supplementary means for species identification, conservation and future studies in *Dendrobium*. We assessed the power of 11 candidate barcodes on the basis of 1,698 accessions of 184 *Dendrobium* species obtained primarily from mainland Asia. Our results indicated that five single barcodes, i.e., ITS, ITS2, *matK*, *rbcL* and *trnH-psbA*, can be easily amplified and sequenced with the currently established primers. Four barcodes, ITS, ITS2, ITS+*matK*, and ITS2+*matK*, have distinct barcoding gaps. ITS+*matK* was the optimal barcode based on all evaluation methods. Furthermore, the efficiency of ITS+*matK* was verified in four other large genera including *Ficus*, *Lysimachia*, *Paphiopedilum*, and *Pedicularis* in this study. Therefore, we tentatively recommend the combination of ITS+*matK* as a core DNA barcode for large flowering plant genera.

Introduction

DNA barcoding has been widely evaluated since the mitochondrial gene cytochrome *c* oxidase I (COI) was proposed as a DNA barcode for species identification[1]. Significant progress has been made in the DNA barcoding of higher plants, and the followingcore DNA barcodes have been proposed: *matK, rbcL*, ITS, or ITS2 and *matK+rbcL*[2–11]. On the other hand, although many efforts have been made to establish a universal barcode for plants, these efforts have not been very successful due to the low substitution rates of mitochondrial DNA[11] and the complicated evolutionary processes and patterns of higher plants, such as genome duplication,

hybridization, and introgression[12–15]. In addition, there are few studies that tested the capacity of DNA barcoding among the largest genera of flowering plants, especially for recently evolved genera, which may present another challenge for DNA barcoding.

Dendrobium, which includes approximately1200–1500 species, is among the largest genera of flowering plants and is primarily distributed in tropical and subtropical Asia, northeast Australia, and New Zealand[16–19]. *Dendrobium* species have important medicinal[20,21]and horticultural value. Many *Dendrobium* species are considered critically endangered or endangered (IUCN Redlist of higher plants in China, http://www.zhb.gov.cn/gkml/hbb/bgg/201309/t20130912_260061.htm) due to over-collection, loss of habitat and habitat fragmentation, and all *Dendrobium* species are included in Appendices I and II of CITES. *Dendrobium* species are notoriously difficult to identify due to their vegetative similarity, large number of species and the overlapping morphological variation within some species[19,22,23]. Furthermore, because they are important economic plants, some species were highly processed in the medicinal market and the shoots of some species were internationally traded, making the species more difficult to recognize. Recent results of molecular systematic studies have indicated that mainland Asian *Dendrobium* is a recent radiation and is divided into eight clades[24]. Given the conservation status and economic value of *Dendrobium*, the difficulties in morphological identification of Asian *Dendrobium* species, and the fact that *Dendrobium* is one of the largest genera with recent radiation, it is an excellent group for testing the effectiveness of DNA barcoding in large flowering plant genera. Moreover, there is an urgent need to develop a DNA barcoding system for conservation and future studies. However, it is difficult to sample all 1200–1500 species of this genus throughout a large geographic region. Thus, here, we focused on species mostly from mainland Asia to evaluate the effectiveness of DNA barcoding.

Recently, five studies focused on evaluating barcodes in *Dendrobium*[25–29](Table S1 in S1 File). However these results were based on sparse sampling (at most 52 species) or used limited evaluation methods (two evaluation methods), some conclusions made by these studies are inconsistent or even conflict with each other. In this study, we assessed 11 candidate barcodes by sampling 184 species of *Dendrobium* obtained mostly from mainland Asia and using various evaluation methods with the following aims: (1) propose a more practical and universal barcode for *Dendrobium* and (2)test the effectiveness of DNA barcoding in four other large plant genera.

Materials and Methods

Plant materials, DNA extraction, PCR amplification, sequencing and sequence download

We first obtained sequences generated from molecular experiments in our lab. Total DNA was isolated from leaves dried in silica-gel using a modified CTAB protocol[30]. Three plastid barcodes (the coding genes *matK* and *rbcL*, and the spacer *trnH-psbA*) and a nuclear internal transcribed spacer (ITS) were amplified and sequenced using universal primers (Table 1). The selected DNA regions were amplified by using a standard polymerase chain reaction (PCR). The PCR mixtures (25 μL) each contained approximately 10 ng (1–2 μL) of template DNA, 12.5 μL of 2×PCR mix (0.005 units/μL Taq DNA polymerase; 4 mM $MgCl_2$; and 0.4 mM dNTPs), 0.2 μL of each primer and 6.5–7.5 μL of ddH_2O. The sequencing reactions were performed using the Applied Biosystems Prism Bigdye Terminator Cycle Sequencing Kit (Foster City, CA).

Second, we downloaded all sequences (ITS, *matK, rbcL*, and *trnH-psbA*) in *Dendrobium* from NCBI. The downloaded sequences from NCBI were filtered according to the following three criteria: i) length less than 300 bp; ii) lacking of voucher specimens; iii)vouchers without specific names (such as *Dendrobium* sp. and *Dendrobium* cff.).

Table 1. A list of primers used for PCR and sequence in this study.

region	primer	Sequence (5′-3′)	Reference
rbcL	1F	ATG TCA CCA CAA ACA GAA AC	[52]
	1360R	CTT CAC AAG CAG CAG CTA GTT C	
matK	390F	CGA TCT ATT CAT TCA ATA TTT C	[53]
	1326R	TCT AGC ACA CGA AAG TCG AAG T	
ITS	17SE	ACG AAT TCA TGG TCC GGT GAA GTG TTC G	[54]
	26SE	TAG AAT TCC CCG GTT CGC TCG CCG TTA C	
trnH-psbA	trnH	CGC GCA TGG TGG ATT CAC AAT CC	[55]
	psbA	GTT ATG CAT GAACGT AAT GCT C	

doi:10.1371/journal.pone.0115168.t001

Although we tried to include at least five individuals for each species, some species had less than five individuals in NCBI and sometimes it was difficult to obtain five individuals in the field. Meanwhile, some species had many individuals. To save computational time, the representatives of each species were limited to fifteen. The taxa, voucher specimens and GenBank accession numbers used in this study are shown in Table S2 in S1 File.

Data analysis

Sequences for each region were aligned with Clustal X v1.8.7[31] and adjusted manually in BioEdit v7.1.3.0[32]. As for ITS, after aligning by Clustal X, we adjusted the regions (ITS1 and ITS2) in two ends of 5.8S rDNA based on parsimony principle. The sequence character-based method were performed for the aligned matrices of each barcode using the 'polymorphic sites' function of the DnaSP5 program[33]. Genetic pairwise distances was computed with the K2P model[34] in MEGA5[35].Differences between intra- and inter-specific distances for each pair of five single barcodes were compared using IBM SPSS Statistics v19.0[36] with Wilcoxon signed-rank tests[37]. Barcoding gaps comparing the distributions of the pairwise intra- and inter-specific distances for each candidate barcode with 0.005 distance intervals were estimated in TaxonDNA with a 'pairwise summary' function[38]. To test the accuracy of the barcode regions for species identification, the proportion of correct identifications were calculated using TaxonDNA with 'Best match', 'Best close match' and 'All species barcodes' functions. To further evaluate the effectiveness of candidate barcodes, we evaluated whether species were considered monophyletic for each barcode by conducting a tree-based analysis. The phylogenetic trees were estimated using the neighbor joining (NJ) feature of MEGA5, and node support was assessed by a bootstrap test[39] with 1000 pseudo-replicates of NJ run with the K2P distance options. *Liparis kumokiri* was used as outgroup for the tree-based analysis following the procedure described by Xiang et al. (2013).

Singh et al. [29]indicatedthat species identification success rate changed with the number of samples. In order to predict the relationships between the number of species sampled and the species identification success rate more accurately, gradient evaluation was used. Gradient evaluation is a method by using different gradient of species in sampling and then evaluating the corresponding efficiency of species identification success of each gradient of ITS+*matK* with the tree-method (NJ).Based on the sampling size of previous studies (Table S3 in S1 File), we here chose 8 species gradients, i.e., 5, 17, 36, 52, 60, 70, 80, and 91species.

Our primary results indicated that ITS+*matK* had the highest species identification success rate. To test the universality of ITS+*matK* as a DNA barcode for species identification in large flowering plant genera, we searched for recent literatures about DNA barcoding in Google Scholar and Web of Science. Four large plant genera, including *Paphiopedilum* (approximately

80 species)[40], *Ficus* (approximately 500 species)[41], *Pedicularis* (approximately 600 species) [42] and *Lysimachia* (approximately 200 species)[43], were found(Table S4-S7 in S1 File). We evaluated the effectiveness of ITS+*matK* for species identification in these genera by calculating genetic distance, constructing NJ trees and conducting analyses using the TaxonDNA program and then compared with the core barcode proposed by the previous study.

Results

PCR amplification and sequencing

The success rates of the amplification of the four loci (ITS, *matK, rbcL,* and *trnH-psbA*)were 100% using the universal primers proposed by CBOL(Table 1). Sequencing success rates were 96.77% (ITS), 97.42% (*matK*), 100% (*rbcL*) and 49.68% (*trnH-psbA*). For *trnH-psbA*, the success rate was relatively low due topoly(T) at about 100bp in the forwards direction when sequencing. The present study submitted 221 new sequences to NCBI, which included 97 sequences of ITS from 37 species, 39 sequences of *matK* from 17 species, 43 sequences of *rbcL* from 18 species and 42 sequences of *trnH-psbA* from 18 species. After screening according to three criteria (see methods), we obtained 1477 sequences from NCBI, including 567, 392, 330 and 188 sequences of ITS, *matK, rbcL* and *trnH-psbA*, respectively. In total, 664 accessions of ITS from 166 species, 431 accessions of *matK* from 105 species, 373 accessions of *rbcL* from 108 species and 230 accessions of *trnH-psbA* from 86 species were collected (Table S2 in S1 File).

Intra- and inter-specific diversity and barcoding gap

The aligned sequence lengths ranged from 1460 bp for *trnH-psbA* to 312 bp for ITS2 (Table 2). ITS had the most variable sites and parsimony-informative characters, followed by *matK* (Table 2). The pairwise intra-specific distances in the eleven barcodes ranged from a minimum of 0.0% to a maximum of 8.29% (Table 3). The mean intraspecific distances were the minimum for *matK*+*rbcL* (0.06%) and the maximum for ITS2 (0.82%). The pairwise interspecific distances in the eleven barcodes ranged from a minimum of 0% to a maximum of 61% (Table 3). The mean interspecific distances were minimum for *trnH-psbA* and *matK*+*trnH-psbA*(0.8%) and maximum for ITS2 (21.6%). In summary, ITS2 exhibited the highest mean intra- and

Table 2. Evaluation of six DNA markers and combinations of the markers.

	ITS	ITS2	matK	rbcL	trnH-psbA	ITS +matK	ITS2 +matK	matK +rbcL	ITS +trnH-psbA	matK +trnH-psbA	ITS+matK +trnH-psbA
Universality of primers	Yes	Yes	Yes	Yes	Yes	-	-	-		-	-
Percentage PCR success (%)	100	100	100	100	100	-	-	-		-	-
Percentage sequencing success (%)	96.77	96.77	97.42	100	49.68	-	-	-		-	-
Length of aligned sequence (bp)	857	312	833	1297	1460	1690	1145	2130	2267	2243	3050
No. of parsimony informative sites/variable sites	129/146	49/50	55/79	0/0	34/47	283/312	147/171	81/117	254/287	82/99	341/388
No. of species samples (individuals)	166 (664)	166 (664)	105 (431)	108 (373)	86(230)	91(406)	91(406)	100 (354)	80(222)	68(185)	67(183)
Ability to discriminate (NJ)	31.93%	22.29%	10.48%	5.56%	8.14%	76.92%	64.84%	24%	60%	25%	73.13%

doi:10.1371/journal.pone.0115168.t002

Table 3. Summary of the pairwise intraspecific and interspecific distances in the barcode loci of *Dendrobium* species.

Barcode locus	Intraspecific distances (%)			Interspecific distances (%)		
	Minimum	Maximum	Mean	Minimum	Maximum	Mean
ITS	0	4.91	0.52	0	24.5	10.4
ITS2	0	8.29	0.85	0	61	21.6
matK	0	1.14	0.08	0	10.1	1.4
rbcL	0	0.92	0.08	0	4.9	1.1
trnH-psbA	0	1.14	0.17	0	3.2	0.8
ITS+matK	0	2.09	0.30	0	17.2	8.3
ITS2+matK	0	3.42	0.31	0	17.4	7
matK+rbcL	0	0.56	0.06	0	5.3	1.1
ITS+trnH-psbA	0	1.57	0.32	0	11.6	6.2
matK+trnH-psbA	0	0.84	0.15	0	4.4	0.8
ITS+matK+trnH-psbA	0	1.55	0.31	0	10.5	5.5

doi:10.1371/journal.pone.0115168.t003

inter-specific distance and the results were supported by using Wilcoxon signed-rank tests (Table S2 in S1 File).

Four barcodes, i.e., ITS (Fig. 1A), ITS2 (Fig. 1B), ITS+*matK*(Fig. 1F) and ITS2+*matK*(Fig. 1G), had relatively clear barcoding gaps. All remaining barcodes had overlaps between their intra- and inter-specific distances without distinct barcoding gaps (Fig. 1C, Fig. 1D, Fig. 1E, Fig. 1H, Fig. 1I, Fig. 1J, Fig. 1K).

Species discrimination

For the analysis using TaxonDNA, ITS+*matK* had the highest success rate for the correct identification of species (Best match: 91.62%;Best close match: 91.62%; All species barcodes: 72.16%) followed by ITS2+*matK*, ITS+*matK+trnH-psbA*, ITS+*trnH-psbA* (Table 4) and *rbcL* had the lowest discrimination success rate (Best match:17.69%; and Best close match: 17.69%). For the tree-based analysis, the performance of eleven candidate barcodes at discriminating species were summarized in Table 2 and Fig. S1-S11 in S1 File. All single-locus barcodes had very low levels of species discrimination, varying from 5.56% (*rbcL*) to 31.93% (ITS). The core barcode *matK+rbcL* proposed by CBOL had the lowest species resolution (24%) among six multi-locus barcodes. ITS+*matK* had the highest success rate (76.92%, Fig. 2) followed by ITS+*matK+trnH-psbA* (73.13%).For these two methods, species discrimination was higher when ITS was included among the six combinations (Table 2, Table 4).

Effectiveness of ITS+matK in gradient evaluation

The species identification success rate decreased as the number of species increased from 5 to 52. However, when the number of species reached the range of 52～91, the success rate of identifying species was stable at approximately 80% (Fig. 3, Table S4 in S1 File).

Effectiveness of ITS+matK in four tested large plant genera

Parveen *et al.*[40] proposed *matK* as the core barcode in slipper orchid *Paphiopedilum*, our results indicated ITS+*matK* (2.1%) has larger mean interspecific distance than *matK*(0.8%) in *Paphiopedilum*. Li *et al.*[41]suggested ITS as the core barcode in *Ficus*, our results indicated ITS+*matK*(62.71%) performed better than ITS (59.32%) based on NJ tree method. Yu *et al.*[42] found ITS was most effective as a core barcode in *Pedicularis*, our results showed that the
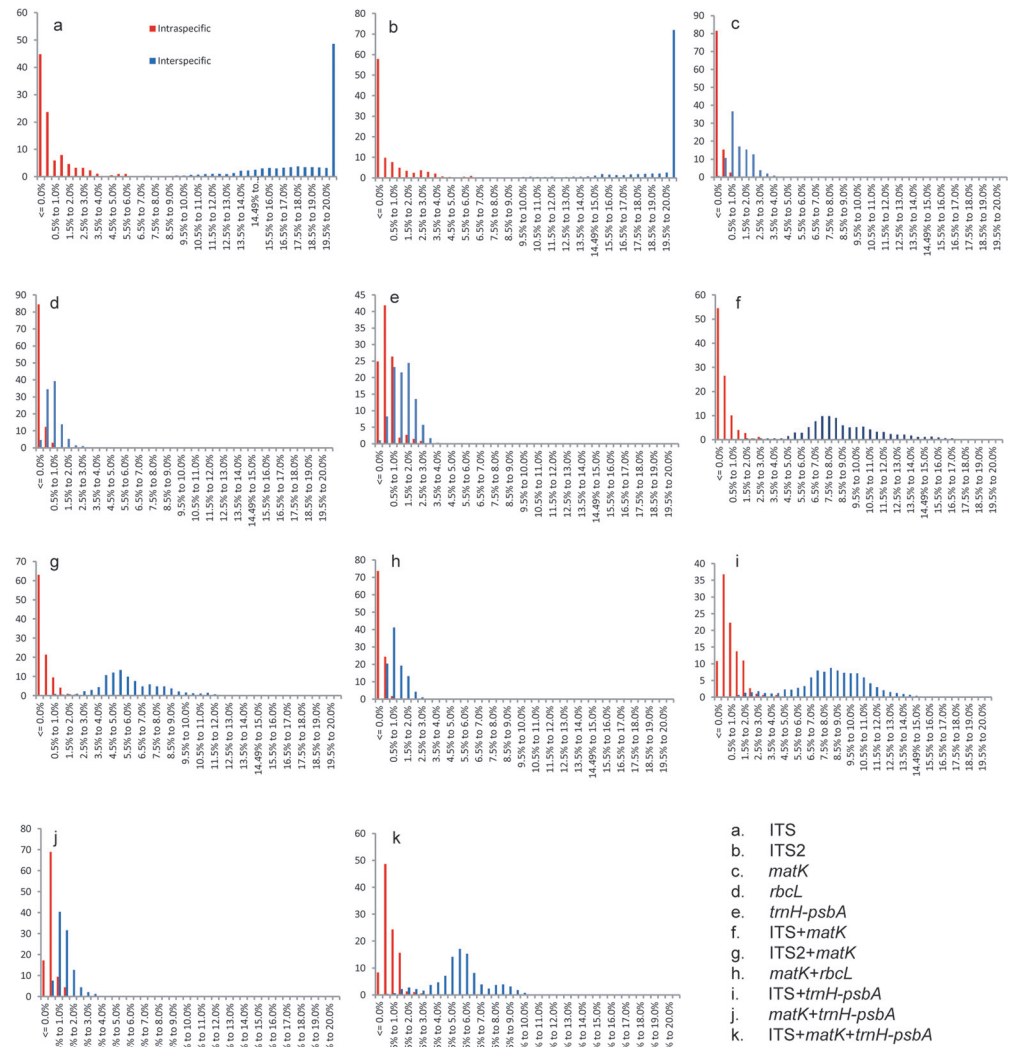
Figure 1. Distribution of intra- and inter-specific Kimura 2-parameter (K2P) distances among all samples for the five candidate loci and their combinations.

success rate of identifying species of ITS+*matK*(76.74%) was larger than ITS (70.93%) based NJ tree method. Zhang *et al.*[43] suggested ITS+*matK*+*rbcL* as a core barcode in *Lysimachia*, our results demonstrated that ITS+*matK*(6.2%) has larger interspecific divergence than ITS+*matK*+*rbcL* (4.4%).Therefore, our results suggested that ITS+*matK* is better than the core barcodes proposed by previous results for these four generastudied here(Tables S5-S8 in S1 File).

Discussion

Evaluation of the DNA barcodes in Dendrobium

Many efforts have been made to discover the core barcodes for different land plant taxa; however, a consensus has not been reached[6,44,45].According to our results, ITS and ITS2have more parsimony informative sites and better discriminatory power among the five proposed loci, i.e., ITS, ITS2, *matK, rbcL,* and *trnH-psbA*, which is consistent with the results of many previous studies[3,7,45].The distance analysis demonstrated that ITS2 had the highest intra-

Table 4. Identification success based on the 'best match', 'best close match' and 'all species barcodes' function of the program TaxonDNA.

| Region	Best match			Best close match			All species barcodes		
	Correct (%)	Ambiguous (%)	Incorrect (%)	Correct (%)	Ambiguous (%)	Incorrect (%)	Correct (%)	Ambiguous (%)	Incorrect (%)
ITS	77.71	6.02	16.26	77.1	5.87	8.73	56.47	33.13	2.1
ITS2	72.28	13.25	14.45	71.53	12.5	7.68	56.62	33.28	1.8
matK	49.18	42.45	8.35	49.18	42.45	8.35	51.97	44.54	3.47
rbcL	17.69	74.53	7.77	17.69	74.53	7.77	47.18	47.98	4.82
trnH-psbA	43.47	29.13	27.39	43.47	29.13	27.39	8.26	89.56	2.17
ITS+*matK*	91.62	1.23	7.14	91.62	1.23	4.67	72.16	22.66	2.7
ITS2+*matK*	90.88	2.46	6.65	90.88	2.21	4.92	69.7	25.86	2.46
matK+rbcL	71.46	17.23	11.29	71.46	17.23	11.29	46.04	50.28	3.67
ITS+*trnH-psbA*	82.88	1.35	15.76	82.43	1.35	9.0	28.64	41.44	2.7
matK+trnH-psbA	68.1	7.56	24.32	68.1	7.56	24.32	22.16	74.05	3.78
ITS+*matK+trnH-psbA*	86.33	0.54	13.11	86.33	0.54	3.27	45.9	46.44	4.37

doi:10.1371/journal.pone.0115168.t004

and inter-specific sequence divergence (Table 3). However, according to the NJ tree, ITS/ITS2 had low species discrimination rates for *Dendrobium* (less than 35%, Table 2), even though ITS has long been used to infer the phylogenies of plants[24,46–48].

On the other hand, we made several new findings regarding the candidate barcodes.Several combinations of two or three barcodes have been proposed as core barcodes, including *matK+rbcL*[11], ITS+*trnH-psbA*[49], ITS+*matK+rbcL*[43]and ITS2+*rbcL*[42], but a consensus regarding the utility of these barcodes has not been reached. The combination of *matK+rbcL* proposed by CBOL as a universal barcode for all land plants has the lowest species resolution (24%) among all six combinations because of the low substitution rates of these coding genes. In contrast, the combination of ITS+*matK* has the highest percent of species identification compared to the other single candidates or combinations (Table 2, Table 4) and has well-defined gaps (Fig. 1F). In agreement with previous results, the combination of ITS+*matK+trnH-psbA* did not provide a higher species identification success rate in comparison with ITS+*matK* [6,50,51].

According to the results of the gradient evaluation for *Dendrobium*, we can predict that ITS+*matK* probably still shows a high success rate of species identification (at approximately 80%)when the number of species exceeds 91.However,there are about 1200–1500 species in *Dendrobium*, and only 184 species (one tenth of the diversity of *Dendrobium*)were included in our analyses. It seems that success rate of species identification will decrease if more species (e.g. 900 species) is included in the analysis. One potential solution for the application of DNA barcoding of large genus as *Dendrobium* is to know the geographical information of specimens, which has been illustrated by some recent results of DNA barcoding[40,41].The relationship between sampling size and success rate of species identification remain to be further tested.

There are three criteria to filter the downloaded sequences from Genbank, however, it is impossible to eliminate the downloaded sequences from misidentified samples or mixed-up materials. Our analyses indicated that these sequences have three possible effects on the results of DNA barcoding. First, these sequences will increase mean intraspecific distances of some taxa and the pairwise interspecific distance between taxa; second, these sequences will overlap between their intra- and inter-specific distances without distinct barcoding gaps; third, these sequence will lower the rate for the correct identification of species and the effectiveness of barcodes. Therefore, it seems that the rate of correct identification of species of ITS+*matK* may
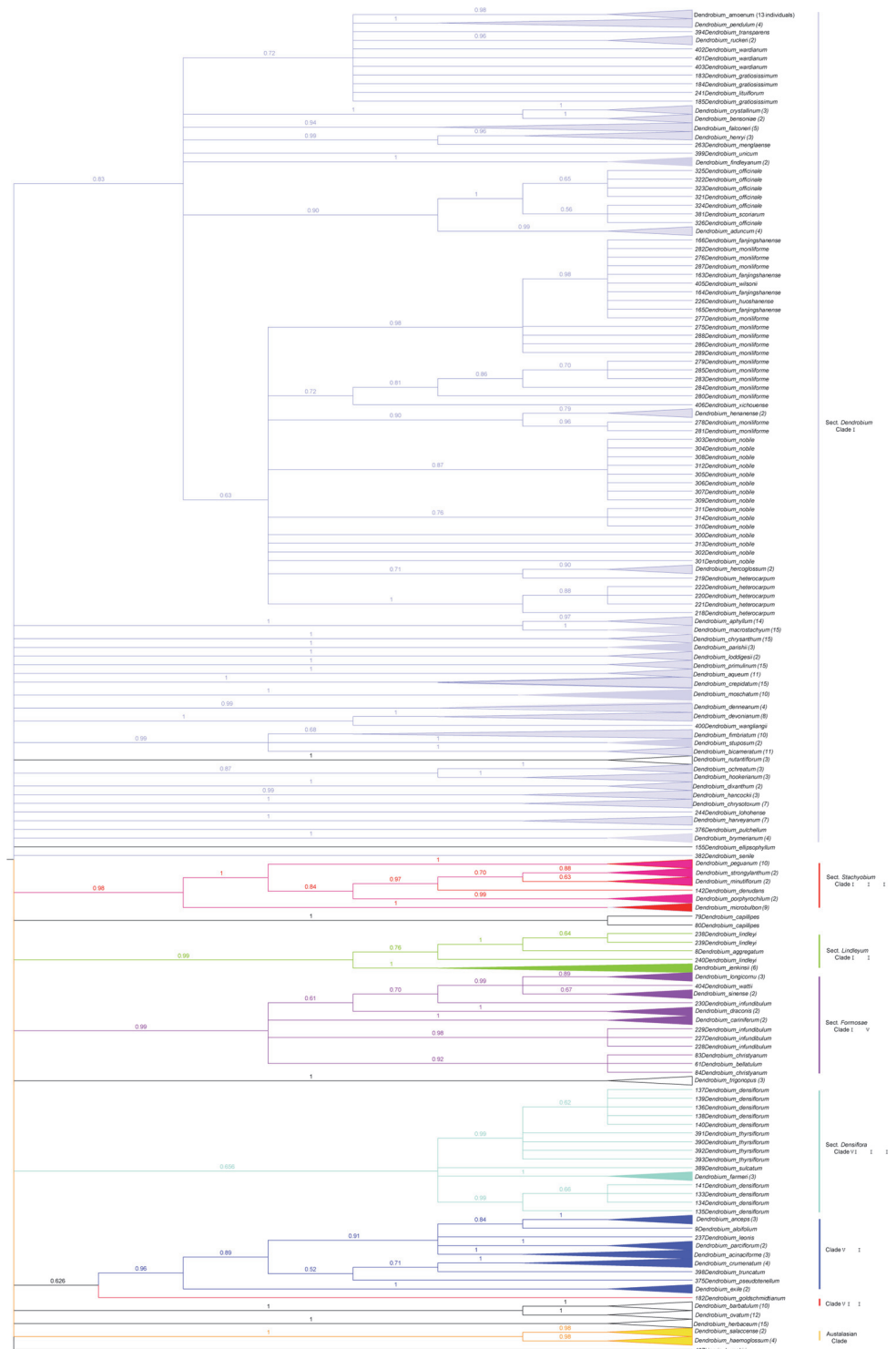
Figure 2. Neighbor joining (NJ) tree generated using ITS+*matK* sequences of *Dendrobium*. Bootstrap values (>50%) are shown above the relevant branches. Corresponding clades are color-coded. Unresolved species according to recent phylogeny research are highlighted in black. More details are presented in Figure S6 in S1 File.

doi:10.1371/journal.pone.0115168.g002
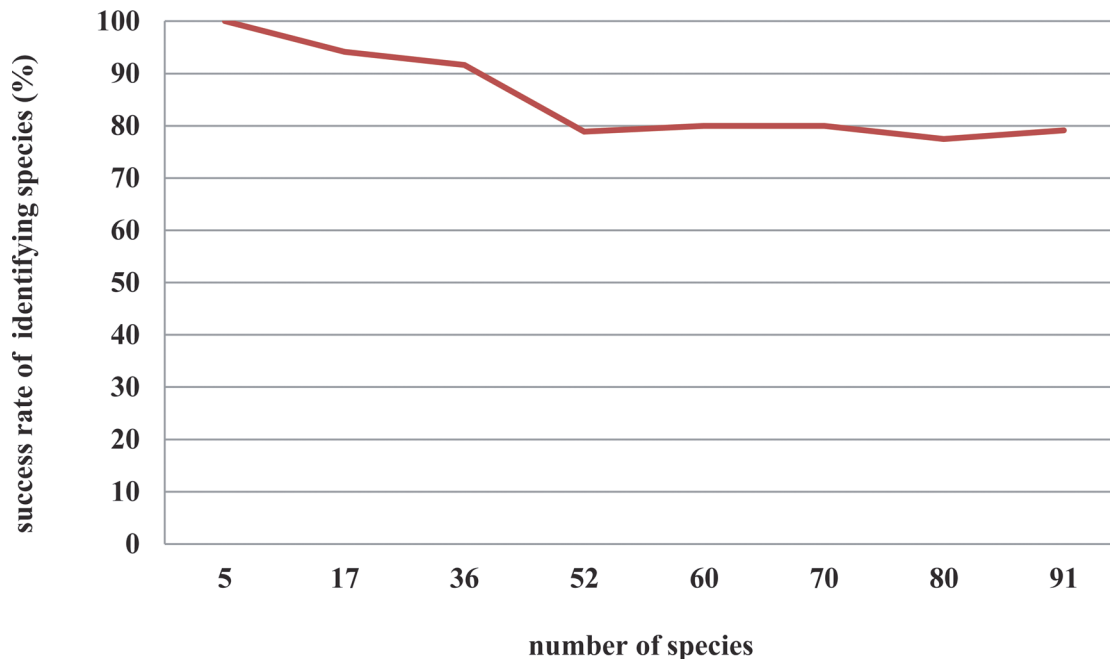
Figure 3. The relationship between number of samples and species identification success rate based on neighbor joining (NJ) tree using ITS+*matK* sequences of *Dendrobium*.

increase if sequences from the misidentified samples or mixed-up materials could be excluded from analyses.

The evaluation of ITS+*matK* in four other large plant genera indicated that this combination showed a higher species discrimination success rate compared with the barcodes proposed in previous publications. Therefore, we tentatively propose ITS+*matK* as a core barcode for large flowering plant genera. This result needs to be further validated in more large flowering plant genera.

Supporting Information

S1 File. Table S1 in S1 File Main information in prior study about DNA barcoding in *Dendrobium*. Table S2 in S1 File Samples and voucher information for the *Dendrobium* species used in this study (the accession numbers in red represent sequences which were newly submitted). Table S3 in S1 File Gradient evaluation of ITS+*matK* in *Dendrobium*. Table S4 in S1 File Summary of species identification success rate based on distance method, NJ tree and the programe TaxonDNA in *Paphiopedilum*. Table S5 in S1 File Summary of species identification success rate based on distance method, NJ tree and the programe TaxonDNA in *Ficus*. Table S6 in S1 File Summary of species identification success rate based on distance method, NJ tree and the programe TaxonDNA in *Pedicularis*. Table S7 in S1 File Summary of species identification success rate based on distance method, NJ tree and the programe TaxonDNA in *Lysimachia*. Table S8 in S1 File Wilcoxon signed-rank tests of intra- and inter-specific divergence among five single loci. Figure S1 in S1 File 50% consensus NJ tree based on ITSfor *Dendrobium* species. Numbers on branches represent NJ support values. Figure S2 in S1 File 50% consensus NJ tree based on ITS2 for *Dendrobium* species. Numbers on branches represent NJ support values. Figure S3 in S1 File 50% consensus NJ tree based on *matK* for *Dendrobium* species. Numbers on branches represent NJ support values. Figure S4 in S1 File 50% consensus NJ tree based on *rbcL*

for *Dendrobium* species. Numbers on branches represent NJ support values. Figure S5 in S1 File 50% consensus NJ tree based on *trnH-psbA* for *Dendrobium* species. Numbers on branches represent NJ support values. Figure S6 in S1 File 50% consensus NJ tree based on ITS+*matK* for *Dendrobium* species. Numbers on branches represent NJ support values. Figure S7 in S1 File 50% consensus NJ tree based on ITS2+*matK* for *Dendrobium* species. Numbers on branches represent NJ support values. Figure S8 in S1 File 50% consensus NJ tree based on *matK+rbcL* for *Dendrobium* species. Numbers on branches represent NJ support values. Figure S9 in S1 File 50% consensus NJ tree based on ITS+*trnH-psbA* for *Dendrobium* species. Numbers on branches represent NJ support values. Figure S10 in S1 File 50% consensus NJ tree based on *matK+trnH-psbA* for *Dendrobium* species. Numbers on branches represent NJ support values. Figure S11 in S1 File 50% consensus NJ tree based on ITS+*matK+trnH-psbA* for *Dendrobium* species. Numbers on branches represent NJ support values.
(PDF)

Author Contributions

Conceived and designed the experiments: XJ XX SX LH. Performed the experiments: XX. Analyzed the data: SX. Contributed reagents/materials/analysis tools: XX XJ JL WJ SX. Wrote the paper: SX. Revised the draft: XJ XX DL JL WH.

References

1. Hebert PDN, Cywinska A, Ball SL (2003) Biological identifications through DNA barcodes. Proceedings of the Royal Society of London Series B: Biological Sciences 270: 313–321. doi: 10.1098/rspb.2002.2218 PMID: 12614582

2. Chase MW, Cowan RS, Hollingsworth PM, Van Den Berg C, Madriñán S, et al. (2007) A proposal for a standardised protocol to barcode all land plants. Taxon 56: 295–299.

3. Chen SL, Yao H, Han J, Liu C, Song J, et al. (2010) Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PloS ONE 5: e8613. doi: 10.1371/journal.pone.0008613 PMID: 20062805

4. Ford CS, Ayres KL, Toomey N, Haider N, Van Alphen Stahl J, et al. (2009) Selection of candidate coding DNA barcoding regions for use on land plants. Botanical Journal of the Linnean Society 159: 1–11.

5. Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. Proceedings of the National Academy of Sciences of the United States of America 102: 8369–8374. doi: 10.1073/pnas.0503123102 PMID: 15928076

6. Lahaye R, Van der Bank M, Bogarin D, Warner J, Pupulin F, et al. (2008) DNA barcoding the floras of biodiversity hotspots. Proceedings of the National Academy of Sciences of the United States of America 105: 2923–2928. doi: 10.1073/pnas.0709936105 PMID: 18258745

7. Li DZ, Gao LM, Li HT, Wang H, Ge XJ, et al. (2011) Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. Proceedings of the National Academy of Sciences of the United States of America 108: 19641–19646. doi: 10.1073/pnas.1104551108 PMID: 22100737

8. Pennisi E (2007) Taxonomy—Wanted: A barcode for plants. Science 318: 190–191. doi: 10.1126/science.318.5848.190 PMID: 17932267

9. Hollingsworth PM, Graham SW, Little DP (2011) Choosing and Using a Plant DNA Barcode. PloS ONE 6: e19254. doi: 10.1371/journal.pone.0019254 PMID: 21637336

10. Hollingsworth PM (2011) Refining the DNA barcode for land plants. Proceedings of the National Academy of Sciences of the United States of America 108: 19451–19452. doi: 10.1073/pnas.1116812108 PMID: 22109553

11. CBOL Plant Working Group (2009) A DNA barcode for land plants. Proceedings of the National Academy of Sciences of the United States of America 106: 12794–12797. doi: 10.1073/pnas.0905845106 PMID: 19666622

12. Bodt S, Maere S, Van de Peer Y (2005) Genome duplication and the origin of angiosperms. Trends in Ecology & Evolution 20: 591–597. doi: 10.1016/j.tree.2005.07.008 PMID: 16701441

13. Fazekas AJ, Kesanakurti PR, Burgess KS, Percy DM, Graham SW, et al. (2009) Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? Molecular Ecology Resources 9: 130–139. doi: 10.1111/j.1755-0998.2009.02652.x PMID: 21564972

14. Rhymer JM, Simberloff D (1996) Extinction by hybridization and introgression. Annual Review of Ecology and Systematics 27: 83–109.

15. Rieseberg LH, Wendel JF (1993) Introgression and its consequences in plants. In: HR G., editor. Hybrid Zones and the Evolutionary Process: Oxford University Press. pp. 70–109.

16. Cribb P, Govaerts R, Prat D (2005) Just how many orchids are there? In: Roguenant A, Raynal-Roques A, editors. Proceedings of the 18th World Orchid Conference, Dijon, France, 11–20 March, 2005: Naturalia Publications. pp. 161–172.

17. Wood HP (2006) The Dendrobiums. Ruggell: A.R.G.Gantner Verlag.

18. Rubinoff D, Cameron S, Will K (2006) Are plant DNA barcodes a search for the Holy Grail? Trends in Ecology & Evolution 21: 1–2. doi: 10.1016/j.tree.2005.10.019 PMID: 16701459

19. Adams P (2011) Systematics of Dendrobiinae (Orchidaceae), with special reference to Australian taxa. Botanical Journal of the Linnean Society 166: 105–126.

20. Hu SY (1970) *Dendrobium* in Chinese medicine. Economic Botany 24: 165–174.

21. State Pharmacopoeia Committee (2010) Pharmacopoeia of the People's Republic of China. Beijing: People's Medical Publishing House.

22. Morris MW, Stern WL, Judd WS (1996) Vegetative anatomy and systematics of subtribe Dendrobiinae (Orchidaceae). Botanical Journal of the Linnean Society 120: 89–144.

23. Yukawa T, Uehara K (1996) Vegetative diversification and radiation in subtribeDendrobiinae (Orchidaceae): Evidence from chloroplast DNA phylogeny and anatomical characters. Plant Systematics and Evolution 201: 1–14.

24. Xiang XG, Schuiteman A, Li DZ, Huang WC, Chung SW, et al. (2013) Molecular systematics of *Dendrobium* (Orchidaceae, Dendrobieae) from mainland Asia based on plastid and nuclear sequences. Molecular Phylogenetics and Evolution 69: 950–960. doi: 10.1016/j.ympev.2013.06.009 PMID: 23811435

25. Asahina H, Shinozaki J, Masuda K, Morimitsu Y, Satake M (2010) Identification of medicinal *Dendrobium* species by phylogenetic analyses using *matK* and *rbcL* sequences. Journal of Natural Medicines 64: 133–138. doi: 10.1007/s11418-009-0379-8 PMID: 20140532

26. Yao H, Song JY, Ma XY, Liu C, Li Y, et al. (2009) Identification of *Dendrobium* species by a candidate DNA barcode sequence: the chloroplast psbA-trnH intergenic region. Planta Medica 75: 667–669. doi: 10.1055/s-0029-1185385 PMID: 19235685

27. Xu H, Wang ZT, Ding XY, Zhou KY, Xu LS (2005) Differentiation of *Dendrobium* species used as Huangcao shihu by rDNA sequence analysis. Planta Medica 71: 1–3.

28. Lau DTW, Shaw PC, Wang J, But PPH (2001) Authentication of medicinal *Dendrobium* species by the internal transcribed spacer of ribosomal DNA. Planta Medica 67: 456–460. doi: 10.1055/s-2001-15818 PMID: 11488462

29. Singh HK, Parveen I, Raghuvanshi S, Babbar SB (2012) The loci recommended as universal barcodes for plants on the basis of floristic studies may not work with congeneric species as exemplified by DNA barcoding of Dendrobium species. BMC Research Notes 5: 42. doi: 10.1186/1756-0500-5-42 PMID: 22260646

30. Doyle JJ, Doyle JL (1987) A rapid procedure for DNA purification from small quantities of fresh leaf tissue. Phytochemical Bulletin 19: 11–15.

31. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research 25: 4876–4882. doi: 10.1093/nar/25.24.4876 PMID: 9396791

32. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41: 95–98.

33. Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19: 2496–2497. doi: 10.1093/bioinformatics/btg359 PMID: 14668244

34. Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16: 111–120. doi: 10.1007/BF01731581 PMID: 7463489

35. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28: 2731–2739. doi: 10.1093/molbev/msr121 PMID: 21546353

36. IBM Corp. (2010) IBM SPSS Statistics for windows, Version 19.0. Armonk, NY: IBM Corp.

37. Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. PLoS Biology 3: e422. doi: 10.1371/journal.pbio.0030422 PMID: 16336051

38. Meier R, Shiyang K, Vaidya G, Ng PK (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. Systematic Biology 55: 715–728. doi: 10.1080/10635150600969864 PMID: 17060194

39. Felsenstein J (1988) Phylogenies from molecular sequences: inference and reliability. Annual Review of Genetics 22: 521–565. doi: 10.1146/annurev.ge.22.120188.002513 PMID: 3071258

40. Parveen I, Singh HK, Raghuvanshi S, Pradhan UC, Babbar SB (2012) DNA barcoding of endangered Indian *Paphiopedilum* species. Molecular Ecology Resources 12: 82–90. doi: 10.1111/j.1755-0998.2011.03071.x PMID: 21951639

41. Li HQ, Chen JY, Wang S, Xiong SZ (2012) Evaluation of six candidate DNA barcoding loci in *Ficus* (Moraceae) of China. Molecular Ecology Resources 12: 783–790. doi: 10.1111/j.1755-0998.2012.03147.x PMID: 22537273

42. Yu WB, Huang PH, Ree RH, Liu ML, Li DZ, et al. (2011) DNA barcoding of *Pedicularis* L.(Orobanchaceae): Evaluating four universal barcode loci in a large and hemiparasitic genus. Journal of Systematics and Evolution 49: 425–437.

43. Zhang C, Wang FY, Yan HF, Hao G, Hu CM, et al. (2012) Testing DNA barcoding in closely related groups of *Lysimachia* L.(Myrsinaceae). Molecular Ecology Resources 12: 98–108. doi: 10.1111/j.1755-0998.2011.03076.x PMID: 21967641

44. Farrington L, MacGillivray P, Faast R, Austin A (2009) Investigating DNA barcoding options for the identification of *Caladenia* (Orchidaceae) species. Australian Journal of Botany 57: 276–286.

45. Yao H, Song JY, Liu C, Luo K, Han JP, et al. (2010) Use of ITS2 region as the universal DNA barcode for plants and animals. PloS ONE 5: e13102. doi: 10.1371/journal.pone.0013102 PMID: 20957043

46. Burke J, Bayly M, Adams P, Ladiges P (2008) Molecular phylogenetic analysis of *Dendrobium* (Orchidaceae), with emphasis on the Australian section Dendrocoryne, and implications for generic classification. Australian Systematic Botany 21: 1–14.

47. Clements MA (2003) Molecular phylogenetic systematics in the Dendrobiinae (Orchidaceae), with emphasis on *Dendrobium* section *Pedilonum*. Telopea 10: 247–298.

48. Schuiteman A (2011) *Dendrobium* (Orchidaceae): to split or not split. Gardens Bulletin-Singapore 63: 245–257.

49. Yang JB, Wang YP, Moeller M, Gao LM, Wu D (2012) Applying plant DNA barcodes to identify species of *Parnassia* (Parnassiaceae). Molecular Ecology Resources 12: 267–275. doi: 10.1111/j.1755-0998.2011.03095.x PMID: 22136257

50. Ashfaq M, Asif M, Anjum Z, Zafar Y (2013) Evaluating the capacity of plant DNA barcodes to discriminate species of cotton (*Gossypium*: Malvaceae). Molecular Ecology Resources 13: 573–582. doi: 10.1111/1755-0998.12089 PMID: 23480447

51. Xiang XG, Hu H, Wang W, Jin XH (2011) DNA barcoding of the recently evolved genus *Holcoglossum* (Orchidaceae: Aeridinae): a test of DNA barcode candidates. Molecular Ecology Resources 11: 1012–1021. doi: 10.1111/j.1755-0998.2011.03044.x PMID: 21722327

52. Goldman DH, Freudenstein JV, Kores PJ, Molvray M, Jarrell DC, et al. (2001) Phylogenetics of Arethuseae (Orchidaceae) based on plastid *matK* and *rbcL* sequences. Systematic Botany 26: 670–695.

53. Cuénoud P, Savolainen V, Chatrou LW, Powell M, Grayer RJ, et al. (2002) Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbcL, atpB*, and *matK* DNA sequences. American Journal of Botany 89: 132–144. doi: 10.3732/ajb.89.1.132 PMID: 21669721

54. Sun Y, Skinner DZ, Liang GH, Hulbert SH (1994) Phylogenetic analysis of *Sorghum* and related taxa using internal transcribed spacers of nuclear ribosomal DNA. Theoretical and Applied Genetics 89: 26–32. doi: 10.1007/BF00226978 PMID: 24177765

55. Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, et al. (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. American Journal of Botany 92: 142–166. doi: 10.3732/ajb.92.1.142 PMID: 21652394