Vietnam Journal of Biotechnology **24**(1): 109-120, 2026. DOI: 10.15625/vjbt-23173

Research Article

GENETIC EVALUATION AND CLASSIFICATION OF HARD-LEAF *Cymbidium* ORCHIDS BASED ON DNA BARCODING MARKERS

Truong Giang Nguyen[1,2], Phuong Thao Pham[1], Van Biet Huynh[3] and Huu Duc Huynh[1,✉]

[1]*Biotechnology Center of Ho Chi Minh City, 2374 Do Muoi, Quarter 10, Trung My Tay, Ho Chi Minh City, Vietnam*

[2]*Faculty of Biological Sciences, Nong Lam University – Ho Chi Minh City, Quarter 33, Linh Xuan, Ho Chi Minh City, Vietnam*

[3]*Research institute for Biotechnology and Environment, Nong Lam University – Ho Chi Minh City, Quarter 33, Linh Xuan, Ho Chi Minh City, Vietnam*

✉To whom correspondence should be addressed. Email: huuduchuynh82@gmail.com

Received: 18 July 2025; Accepted: 15 March 2026; Published online: 25 March 2026

ABSTRACT

The hard-leaf *Cymbidium* orchid group comprises several species within its genus, including *Cymbidium finlaysonianum, Cymbidium aloifolium, Cymbidium bicolor*, and *Cymbidium atropurpureum*, which are adapted to tropical and subtropical climates. These native Vietnamese orchids have high economic value and are currently overexploited in the wild. To conserve and utilize these genetic resources, 19 samples of hard-leaf *Cymbidium* orchids from various regions were analyzed and genetically evaluated using DNA barcoding markers. The analysis of two chloroplast gene regions (*rbc*L and *mat*K) and one nuclear gene region (ITS2) revealed the sizes of each gene region: *rbc*L (680-715 bp), *mat*K (760-762 bp), and ITS2 (453 bp). Comparing the DNA sequences of these regions among the samples showed both conservation and diversity within and between species, with conserved sites (678/715, 744/762, 228/453), variable sites (37/715, 18/762, 225/453), Pi (36/715, 11/762, 191/453), and InDels (35/715, 2/762, 0/453). Phylogenetic analysis indicated that the *mat*K gene could clearly distinguish four species; however, the *rbc*L gene could not differentiate between *C. aloifolium* and *C. atropurpureum.* The ITS2 region displayed high intraspecific variation but was insufficient for species discrimination. Combining *rbc*L and *mat*K regions provided both intra-species diversity insights and effective species discrimination. These findings demonstrate that DNA barcoding markers are useful for species identification and assessing genetic diversity among the four hard-leaf *Cymbidium* orchids. They also highlight some limitations of using the ITS2 region for species identification in some instances.

Keywords: *Cymbidium,* DNA barcoding markers, genetic diversity, genetic resources, maturase K (*mat*K), sequencing.

INTRODUCTION

Cymbidium is one of the most critically cultivated orchids due to its ease of growth, diversity, beautiful flowers, and long-lasting nature. They are highly adapted to vast temperature ranges in nature and horticulture, with many hybrid varieties created. At the same time, the diversity of flower colors and shapes resulting from hybridization brings new varieties to the market. Many hybrids with attractive flower characteristics, vibrant colors, and large flowers have been created, and they hold the most significant commercial value in the orchid cultivation industry today. However, hybrids often have complex origins involving many different species. Many small-flowered hybrids, adapted to pot-growing conditions, are produced through hybridization between *Cymbidium devonianum, Cymbidium floribundum, Cymbidium tigrinum, Cymbidium ensifolium*, and *Cymbidium madidum* with large-flowered varieties. This can confuse the taxonomy of this genus (Puy and Cribb, 2007). *Cymbidium* has approximately 50 species and numerous hybrids that are commonly found in the tropics and subtropics, primarily in Asia and northern Australia. In Vietnam, about 21 species are present, including four hard-leaf *Cymbidium* orchids: *C. aloifolium, C. finlaysonianum, C. bicolor*, and *C. atropurpureum* (Tran, 2000; Nayak *et al.*, 2006; Yang *et al.*, 2013). They exhibit a wide range of colors and forms within the same species. Notably, *C. finlaysonianum* is a native hard-leaf *Cymbidium* orchid in Vietnam that is currently well accepted and favored in the market. At the same time, very few studies have been conducted on the genetic resources and genetic diversity of hard-leaf *Cymbidium* orchids in Vietnam.

DNA molecular markers have been utilized in genetic research to assess genetic diversity, identify genetic resources, analyze target genes, investigate genetic variation, and facilitate species identification, all of which are effective across various fields. In particular, the DNA barcode is a modern technique that has been extensively researched and applied worldwide for classification, evaluating genetic diversity, and conserving genetic resources. This method relies on the variation and evolution of DNA sequences in the nucleus and chloroplasts, which helps determine genetic relationships among organisms and partially assists with species identification within the taxonomy system (Group *et al.*, 2009; Fazekas *et al.*, 2012). Many previous studies successfully used DNA barcoding markers for genetic diversity assessment and species identification in various orchid resources, such as those from Australia (Farrington *et al.*, 2009), Thailand (Siripiyasing *et al.*, 2012a), China (Zhu *et al.*, 2022), Indonesia (Su'udi and Setyati, 2022), Malaysia (Rajaram *et al.*, 2019), and others, targeting gene regions such as *rbc*L, *mat*K, *trn*H-*psb*A, *rpo*C, *rpo*B, and *ycf*1. These studies highlight the effectiveness of different gene regions in evaluating genetic diversity and identifying species across various orchid genera, including *Dendrobium, Phalaenopsis, Vanda, Oncidium,* and *Cymbidium*. Especially for *Cymbidium*, numerous studies have been conducted on genetic resource diversity and species identification. However, these mainly focus on soft-leaf *Cymbidium* orchids adapted for cold climates, such as *C. floribundum, C. tigrinum*, and *C. ensifolium*, whereas species adapted to tropical and subtropical climates, such as *C. finlaysonianum, C. aloifolium, C. bicolor,* and *C. atropurpureum* (Puy and Cribb, 2007; Sharma *et al.*, 2012;

Siripiyasing *et al.*, 2012b). Therefore, evaluating and identifying species within the *C. finlaysonianum, C. bicolor, C. aloifolium,* and *C. atropurpureum* groups based on DNA barcoding markers is essential for conservation and the sustainable use of these genetic resources. In this study, 19 orchid samples representing four hard-leaf *Cymbidium* species collected from different regions were analyzed and genetically assessed using three gene regions (*rbc*L, *mat*K, and ITS) to develop a DNA-based database for species identification and the preservation and conservation of these genetic resources.

MATERIALS AND METHODS

Plant materials

A total of 19 hard-leaf *Cymbidium* orchid samples (10 *C. finlaysonianum* samples, 2 *C. bicolor* samples, 5 *C. aloifolium* samples, and 2 *C. atropurpureum* samples) were collected from different geographical regions and grown under net house conditions in Ho Chi Minh City for classification and evaluation of genetic relationships (Table 1).

Table 1. Source of hard-leaf *Cymbidium* orchid samples included in the study.

Symbol	Scientific name	Collection site
CF01	*C. finlaysonianum*	Kon Ka Kinh National Park (Gia Lai)
CF02	*C. finlaysonianum*	Chu Yang Sin National Park (Dak Lak)
CF03	*C. finlaysonianum*	Ta Dung National Park (Dak Nong)
CF04	*C. finlaysonianum*	Bidoup Nui Ba National Park (Lam Dong)
CF05	*C. finlaysonianum*	Yok Don National Park (Dak Lak)
CF06	*C. finlaysonianum*	Bu Gia Map National Park (Binh Phuoc)
CF07	*C. finlaysonianum*	Chu Mom Ray National Park (Kon Tum)
CF08	*C. finlaysonianum*	Dak Lak
CF09	*C. finlaysonianum*	Kon Tum
CF10	*C. finlaysonianum*	Lam Dong
CB01	*C. bicolor*	Yok Don National Park (Dak Lak)
CB02	*C. bicolor*	Ta Dung National Park (Dak Nong)
CA01	*C. aloifolium*	Chu Yang Sin National Park (Dak Lak)
CA02	*C. aloifolium*	Yok Don National Park (Dak Lak)
CA03	*C. aloifolium*	Kon Tum
CA04	*C. aloifolium*	Gia Lai
CA05	*C. aloifolium*	Kon Tum
CT01	*C. atropurpureum*	Dak Lak
CT02	C. atropurpureum	Quang Nam

DNA extraction

Total DNA was extracted from leaf samples using the GeneJET Plant Genomic DNA Purification Kit (Thermo Fisher Scientific, USA). The purity and concentration of the DNA extract were verified by 0.8% agarose gel electrophoresis and measured using a NanoDrop spectrophotometer (Thermo Fisher Scientific, USA).

Amplification

Polymerase Chain Reaction (PCR) amplification of the *rbc*L, *mat*K, and ITS2 gene regions with primer pair *rbc*L1-F: 5'-ATG TCA CCA CAA ACA GAA AC-3', *rbc*L724-R: 5'-TCG CAT GTA CCT GCA GTA GC-3' (Newmaster *et al.*, 2013); *mat*K472-F: 5'-CCC RTY CAT CTG GAA ATC TTG GTT C-3', *mat*K1248-R: 5'-GCT

RTR ATA ATG AGA AAG ATT TCT GC-3' (Asahina *et al*., 2010); and ITS-p3 (5.8S): 5'-YGA CTC TCG GCA ACG GAT A-3', ITS-p4 (26S): 5'-CCG CTT AKT GAT ATG CTT AAA-3' (Cheng *et al*., 2016). The composition for a 20 µL PCR reaction contained 10 µL Dream Taq Green PCR Master Mix (Thermo Fisher Scientific, USA), 0.4 µL forward primer (20 µM), 0.4 µL reverse primer (20 µM) (synthesized at Integrated DNA Technologies (IDT)), 8.2 µL double-distilled sterilized water and 1 µL template DNA (50 ng/µL). The PCR cycle included pre-denaturation at 95°C for 1 minute, followed by 30 cycles (denaturation at 95°C for 30 seconds, annealing at 52°C, 53°C or 54°C for 30 seconds, and extension at 72°C for 40 seconds). The final extension was performed at 72°C for 10 minutes. The PCR products were analyzed and evaluated on a 1.2% agarose gel and photographed using a GelDoc-It 2315 imager (UVP, USA).

Sequencing and genetic classification

PCR products of the *rbc*L gene region, *mat*K gene region, and ITS2 were sequenced using the Sanger method at 1st Base Company (Malaysia). DNA sequences were edited with ATGC ver. 7.1 software and checked for discrepancies by removing about 20-30 nucleotides from the beginning and end, comparing the forward strand and reverse strand DNA sequences, and collecting the corresponding DNA sequence. These sequences were used to assess Percent Identity and Query Coverage compared to those available in the GenBank database using the BLAST tool (Basic Local Alignment Search Tool). Phylogenetic trees were constructed using MEGA 7.0 software (Molecular Evolution Genetics Analysis) with the Construct/Test Maximum Likelihood Tree algorithm, and branch support was assessed with 1000 bootstrap replicates.

RESULTS AND DISCUSSION

The total DNA obtained from the leaf samples was quantified, and its purity was assessed. The results showed that the total DNA had high purity, with an OD 260/280 ratio ranging from 1.77 to 2.13, and a DNA concentration between 151.13 ng/µL and 249.09 ng/µL. The DNA was diluted to a concentration of 50 ng/µL for PCR analysis in the next section.

Amplification and DNA sequence analysis

PCR amplification results of the *rbc*L gene region, *mat*K gene region, and ITS2 using primer pairs *rbc*L1-F and *rbc*L724-R; *mat*K472-F and *mat*K1248-R; ITS-p3 (5.8S) and ITS-p4 (26S) were obtained from 19 hard-leaf *Cymbidium* orchid samples, with a success rate of 100%. The PCR product sizes were approximately 700-750 bp for the *rbc*L gene region, 800-850 bp for the *mat*K gene region, and 450-500 bp for the ITS2 gene region. The PCR products of the *rbc*L gene, *mat*K gene, and ITS2 were sequenced using the Sanger method at First Base Company (Malaysia), achieving a 100% success rate. Raw DNA sequences were processed with ATGC ver 7.1 software, revealing sizes of 680-715 bp for *rbc*L, 760-762 bp for *mat*K, and 453 bp for ITS2. These sequences were used to analyze variable sites, conserved sites, Pi index, and insertion/deletion sites (InDels) in the 19 samples, using MEGA 7.0 software. For the *rbc*L gene region, conserved sites were at 678/715, 744/762, and 228/453 sites; variable sites at 37/715, 18/762, and 225/453; Pi at 36/715, 11/762, and 191/453; and InDel at 35/715 and 2/762 sites. When analyzing across four species, the *rbc*L region had variant positions at

37/715 sites. Still, no variants were observed among *C. aloifolium, C. bicolor,* and *C. atropurpureum*. In contrast, only one or two variants appeared in *C. finlaysonianum* samples, such as at position 73 of sample CF09 (T, others C) and position 340 of CF010 (T, others C). InDel positions only appeared when comparing the four hard-leaf *Cymbidium* species, except for sample CF07, which had one InDel at position 658 (G) among *C. finlaysonianum* samples (Table 2). For the *mat*K gene region within *C. finlaysonianum*, six of 18 sites were variable, with one insertion-deletion (InDel) at two positions: specifically, at position 686, where CF10 had an A nucleotide, and other samples exhibited a deletion at this site. No InDels were observed among the other three species (Table 3). The ITS2 gene region showed more differences, with 225/453 variable sites across the four species, and 114/453 among the hard-leaf *Cymbidium* samples, indicating greater variability than in the chloroplast gene regions. InDel positions appeared when comparing *C. finlaysonianum* with the other species based on the *rbc*L gene region. For the *mat*K region, InDel positions were only noted in two samples compared to the others. Overall, the ITS2 gene region exhibited more variation among samples, followed by the *mat*K gene region. Morphologically, focusing on flowers, samples with different DNA sequences compared to others displayed notable differences in flower coloration and shape.

Analysis of percent identity and query coverage of hard-leaf *Cymbidium* orchid samples compared to data on GenBank.

Gene region DNA sequences of *rbc*L, *mat*K, and ITS2 from 19 collected *Cymbidium* orchids were compared with the GenBank database using the BLAST tool. The percent identity ranged from 91.7% to 100%, 98.30% to 99.87%, and 75.6% to 98.9%, respectively, with coverage of 93% to 100%, 94% to 100%, and 78% to 97%. Among these, *C. finlaysonianum* samples showed similarity and coverage based on three gene regions, with data on GenBank reaching from *rbc*L (91.7-92.46%; 93-100% and similar to OR413862.1-*C. huoshanense*), *mat*K (99.35-99.61%; 99-100% and similar to NC_079579.1-*C. finlaysonianum*), and ITS2 (83.52-96%; 91-95% and identical to AF470514.1-*C. finlaysonianum*). In *C. bicolor* samples, percent identity and coverage were *rbc*L (97%, 100% and similar to MN654912.1-*C. bicolor*), *mat*K (99.34-99.87%, 94-99% and similar to OQ405312.1-*C. mannii=C. bicolor*), and ITS2 (97.35-98.9%; 85-95% and similar to AF470514.1-*C. finlaysonianum*, MF86-1139.1-*C. aloifolium*). In *C. aloifolium* samples, the percent identity with OQ405325.1-*C. aloifolium* was 100% with a coverage of 100%. ITS2 showed similarity levels of 75.6-91.86% and 78-97.85%, and was similar to MF861151.1-*C. paucifolium*, HQ263142.1-*C. serratum*, and AJ300273.1-*C. ensifolium* (Table 4). The *C. atropurpureum* sample showed similarity with OQ405302.1-*C. atropurpureum*. These results demonstrate that the studied hard-leaf *Cymbidium* orchid samples share some similarities and differences with existing sequences on GenBank. This indicates that gene regions *rbc*L and *mat*K are effective for classifying four species of hard-leaf *Cymbidium* orchids and assessing genetic diversity within these species.

Table 2. Variant position and InDel of 19 hard-leaf *Cymbidium* orchid samples based on the *rbc*L gene region and 4 reference DNA sequences on GenBank.

Sample	Variant positions																																			
	07	28	40	47	64	73	100	114	143	151	166	325	326	340	341	362	386	420	434	468	473	477	483	510	521	551	584	605	620	643	648	659	660	664	665	668	691
OR413862.1	A	G	G	G	G	C	T	C	T	G	T	A	G	T	G	A	C	C	C	G	G	C	C	G	G	G	G	C	C	C	T	A	A	A	G	A	G
CF01-CF07	A	G	G	G	G	C	T	C	T	G	T	A	G	C	G	A	C	C	C	G	G	C	C	G	G	G	G	C	C	C	T	A	A	A	G	A	G
CF08	A	G	G	G	G	C	T	C	T	G	T	A	G	C	G	A	C	C	C	G	G	C	C	G	G	G	G	C	C	C	T	A	A	A	G	A	G
CF09	A	G	G	G	G	T	T	C	T	G	T	A	G	T	G	A	C	C	C	G	G	C	C	G	G	G	G	C	C	C	T	A	A	A	G	A	G
CF10	A	G	G	G	G	C	T	C	T	G	T	A	G	T	G	A	C	C	C	G	G	C	C	G	G	G	G	C	C	C	T	A	A	A	G	A	G
MN641752.1	A	T	T	T	T	T	G	A	G	C	A	C	T	T	T	T	T	A	G	T	T	C	T	T	T	A	T	A	A	A	G	C	T	T	A	G	T
CA01-CA05	A	T	T	T	T	T	G	A	G	C	A	C	T	T	T	T	T	A	G	T	T	C	T	T	T	A	T	A	A	A	G	C	T	T	A	G	T
NC060583.1	A	T	T	T	T	T	G	A	G	C	A	C	T	T	T	T	T	A	G	T	T	C	T	T	T	A	T	A	A	A	G	C	T	T	A	G	T
CB01-CB02	A	T	T	T	T	T	G	A	G	C	A	C	T	T	T	T	T	A	G	T	T	T	C	T	T	A	T	A	A	A	G	C	T	T	A	G	T
OQ405302.1	A	T	T	T	T	T	G	A	G	C	A	C	T	T	T	T	T	A	G	T	T	C	T	T	T	A	T	A	A	A	G	C	T	T	A	G	T
CT01	T	T	T	T	T	T	G	A	G	C	A	C	T	T	T	T	T	A	G	T	T	C	T	T	T	A	T	A	A	A	G	C	T	T	A	G	T
CT02	A	T	T	T	T	T	G	A	G	C	A	C	T	T	T	T	T	A	G	T	T	C	T	T	T	A	T	A	A	A	G	C	T	T	A	G	T

Sample	InDel positions																																	
	123	124	125	126	127	235	236	237	238	239	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	552	553	554	555	556	557	658
OR413862.1	-	-	-	-	-	G	G	A	C	G	C	A	T	C	G	A	G	G	C	-	-	-	-	-	-	-	-	-	G	T	T	T	A	T	G
CF01-CF06	A	C	C	G	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CF07	A	C	C	G	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G
CF08-CF10	A	C	C	G	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MN641752.1	-	-	-	-	-	G	G	A	C	G	C	A	T	C	G	A	G	G	C	T	G	T	T	G	T	T	G	G	G	T	T	T	A	T	G
CA01-CA05	-	-	-	-	-	G	G	A	C	G	C	A	T	C	G	A	G	G	C	T	G	T	T	G	T	T	G	G	G	T	T	T	A	T	G
NC060583.1	-	-	-	-	-	G	G	A	C	G	C	A	T	C	G	A	G	G	C	T	G	T	T	G	T	T	G	G	G	T	T	T	A	T	G
CB01-CB02	-	-	-	-	-	G	G	A	C	G	C	A	T	C	G	A	G	G	C	T	G	T	T	G	T	T	G	G	G	T	T	T	A	T	G
OQ405302.1	-	-	-	-	-	G	G	A	C	G	C	A	T	C	G	A	G	G	C	T	G	T	T	G	T	T	G	G	G	T	T	T	A	T	G
CT01-CT02	-	-	-	-	-	G	G	A	C	G	C	A	T	C	G	A	G	G	C	T	G	T	T	G	T	T	G	G	G	T	T	T	A	T	G

Note: OR413862.1-D. huoshanense, NC_060583.1-C. bicolor, MN641752.1-C. aloifolium, OQ405302.1-C. atropurpureum, "-": Positions with no nucleotide

Table 3. Variant position, Indel of 19 hard-leaf *Cymbidium* orchid samples based on the *mat*K gene region and 4 reference DNA sequence on GenBank.

Sample	Variant positions																		InDel positions	
	27	53	69	82	109	134	155	179	207	232	240	265	286	318	397	418	426	473	682	686
NC079579.1	A	T	G	C	G	C	C	T	C	T	T	G	G	A	T	A	A	T	-	-
CF01-CF06	A	T	G	A	G	C	C	T	C	T	T	G	G	C	T	A	A	C	-	-
CF07	A	T	G	A	G	C	C	T	C	T	T	T	G	C	T	A	A	C	-	-
CF08	A	T	G	C	G	C	A	C	C	T	T	G	G	C	T	A	A	C	-	-
CF09	T	T	G	A	G	C	C	T	C	T	T	G	G	C	T	A	A	C	-	-
CF10	A	T	G	C	G	C	C	T	C	G	T	G	G	C	T	A	A	C	-	A
MN641752.1	A	G	G	C	G	C	C	T	C	T	C	G	G	C	T	G	G	T	-	-
CA01-CA05	A	G	G	C	G	C	C	T	C	T	C	G	G	C	T	G	G	C	-	-
OQ405312.1	A	G	C	C	A	C	C	T	C	T	T	G	A	C	T	G	G	T	-	-
CB01-CB02	A	G	G	C	A	C	C	T	T	T	T	G	A	C	T	G	G	C	-	-
OQ405302.1	A	G	G	C	G	C	C	T	C	T	T	G	G	C	G	G	G	T	-	-
CT01-CT02	A	G	G	C	G	A	C	T	C	T	T	G	G	C	G	G	G	C	A	-

Note: NC_079579.1-C. finlaysonianum, OQ405312.1-C. mannii=C. bicolor, MN641752.1-C. aloifolium, OQ405302.1-C. atropurpureum, "-": Positions with no nucleotide.

Table 4. Comparison of percent identity and query coverage of gene region DNA sequences (*rbc*L, *mat*K, ITS2) of the studied hard-leaf *Cymbidium* orchid samples versus available sequences on GenBank.

Symbol	Gene region	Accession		Percent identity (%)	Query coverage (%)
CF01-CF10	*rbcL*	OR413862.1	*D. huoshanense*	91.7-92.46	93-100
CB01-CB02		MN654912.1	*C. bicolor*	97.2	100
CA01-CA05		OQ405325.1	*C. aloifolium*	100	100
CT01-CT02		OQ405302.1	*C. atropurpureum*	99.72	99
CF01-CF10	*matK*	NC_079579.1	*C. finlaysonianum*	99.35-99.61	99-100
CB01		OQ405312.1	*C. mannii*	99.60	99
		MN654912.1	*C. bicolor*	99.34	99
CB02		MN654912.1	*C. bicolor*	99.87	94
CA01-CA05		MN641752.1	*C. aloifolium*	98.30-99.87	99-100
CT01-CT02		OQ405302.1	*C. atropurpureum*	99.61	99
CF01-CF10	*ITS2*	AF470514.1	*C. finlaysonianum*	83.52-96	91-95
CB01		AF470514.1	*C. finlaysonianum*	97.35	85
		M861139.1	*C. aloifolium*	98.9	95
CA01		MF861151.1	*C. paucifolium*	98.30-99.87	99-100
CA02-CA04		HQ263142.1	*C.serratum*	75.6-97.85	78-91
CA05		AJ300273.1	*C. ensifolium*	79.49	91
CT01-CT02		AJ300286.1	*C. aspidistrifolium*	81.49-81.69	87-88

Classification and genetic evaluation of four hard-leaf *Cymbidium* orchid species

The phylogenetic tree of 19 *Cymbidium* orchid samples (including 10 *C. finlaysonianum*, 2 *C. bicolor*, 5 *C. aloifolium*, and 2 *C. atropurpureum*) based on DNA sequences from each gene region (*rbc*L, *mat*K, and ITS2) shown in GenBank, indicates different levels of phylogenetic classification depending on whether chloroplast or nuclear gene regions are analyzed. For the chloroplast genome, results based on the *rbc*L gene show that *C. finlaysonianum* is classified into one group (Group I) and is separated from the remaining group (Group II), which includes *C. aloifolium, C. bicolor,* and *C. atropurpureum* (Figure 2A). Within this, the *C. bicolor* group is distinct from the *C. atropurpureum* and *C. aloifolium* groups. Analyses of the *mat*K region reveal that *C. bicolor* is separated from *C. finlaysonianum*, *C. atropurpureum*, and *C. aloifolium* (Figure 2B). These three groups are also clearly distinct from each other. These findings indicate genetic variation both within species and between different species, with some differences in phylogenetic classification depending on the gene region. Additionally, the *mat*K region exhibits a higher level of diversity than *rbc*L within the *C. finlaysonianum* group, whereas samples from other groups display slight intraspecific variation, despite differences in morphology and flower color (results of morphology and flower coloration analysis are pending publication). Furthermore, the combined analysis of the *rbc*L and *mat*K regions (Figure 2C) provides a clear phylogenetic distinction among four *Cymbidium* species and demonstrates genetic stability within some samples of *C. finlaysonianum* and *C. aloifolium*. The *C. finlaysonianum* group is separated from the others, while *C. aloifolium* is closely related to *C. atropurpureum*. The species identification results based on these two gene regions are consistent with previous morphological classifications of the genus *Cymbidium* (Puy and Crib, 2007). Moreover, the species identification ability of the *rbc*L and *mat*K regions has been explored in *Cymbidium*, with the *mat*K region capable of distinguishing more species than *rbc*L, depending on the section (Worthy *et al*., 2022; Zang *et al*., 2023). The use of DNA barcoding markers in plant taxonomy has been common in past studies, supporting morphological classification and analyzing genetic relationships within the orchid family using gene regions such as *rbc*L, *psa*B, *atp*B, and *mat*K (Cameron *et al*., 1999; Cameron, 2004; Cameron, 2006; Freudenstein and Chase, 2015). These studies found a correlation between morphological and molecular methods. Additionally, some research indicates that the *mat*K gene region evolves approximately three times faster than the *rbcL* and *atpB* genes, making it an effective tool for distinguishing species in angiosperms (Gao *et al*., 2011). Furthermore, for the ITS2 nuclear gene region, the analysis results showed that these four *Cymbidium* orchid species were not clearly separated from each other, and the samples within the same species also differed (Figure 2C). The results also showed that some samples, although from different species, were grouped closely together, such as CB01, which belongs to the same group as *C. finlaysonianum*, and CF07, which is in the same group as *C. aloifolium*. These findings indicate that the ITS region is less effective for species identification in these studied samples. The above results are also consistent with studies on the phylogenetic classification of species within

the genus *Cymbidium* using nuclear and chloroplast gene regions, which still show differences in genetic groupings among some species when applying a phylogenetic tree based on different gene regions (Berg *et al.*, 2002; Yukawa *et al.*, 2002).

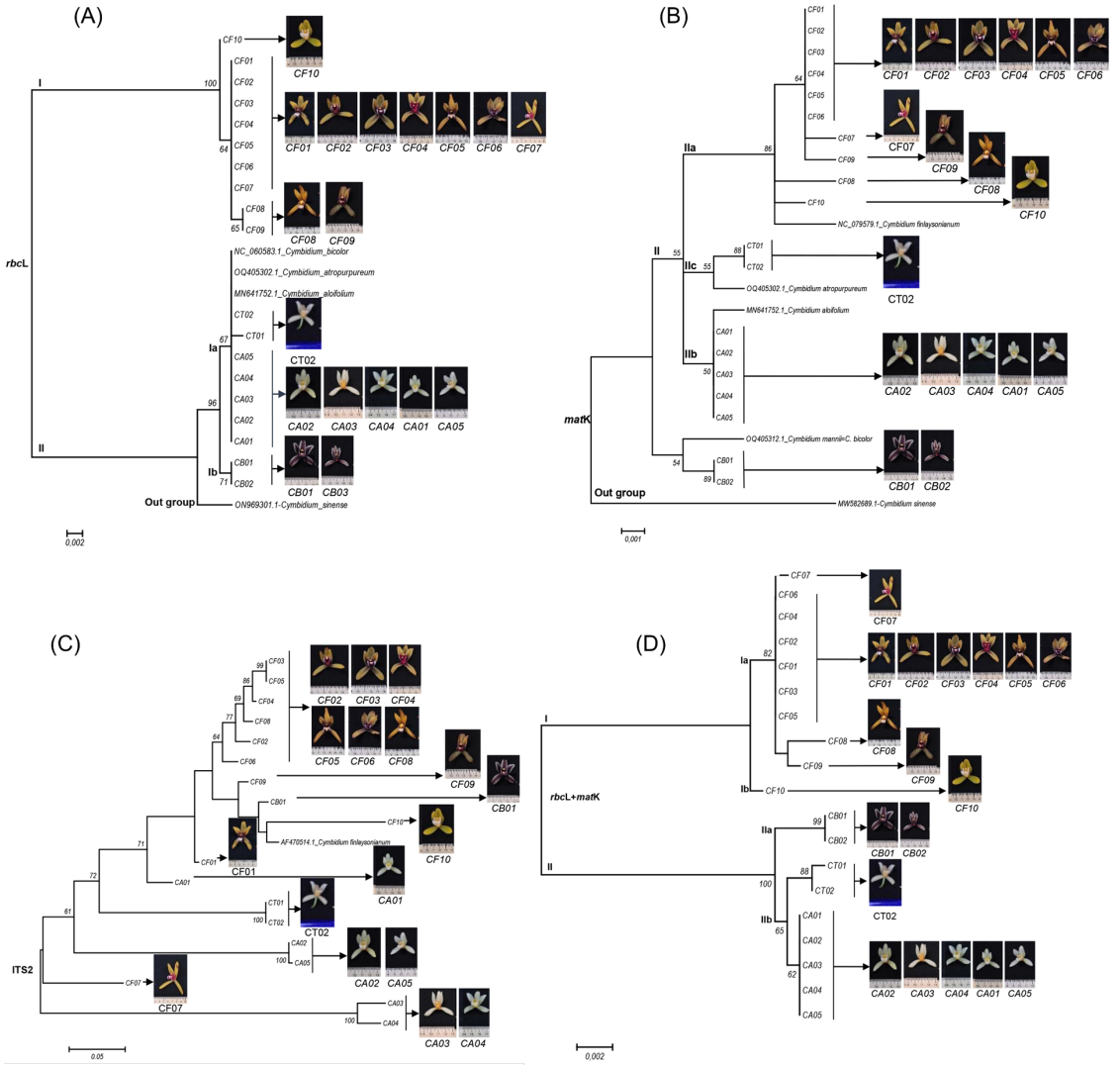

Figure 2. Phylogenetic tree based on DNA sequences of three gene regions (*rbc*L, *mat*K, ITS2, *rbc*L+*mat*K) of hard-leaf *Cymbidium* orchid samples using MEGA 7.0 software with the Maximum Likelihood tree algorithm and a bootstrap value of 1000.

For the ITS region, some studies have shown it can support species identification in some instances. (Jalil *et al.*, 2015; Kress *et al.*, 2015; Paula, 2013). However, other research suggests that it is not suitable and depends on the subject and the number of samples collected (Berg *et al.*, 2002; Yukawa *et al.*, 2002; Puy and Crib, 2007). In this study, using a large number of *C. finlaysonianum* samples (10 samples) from different regions demonstrated the stability of the *rbc*L and *mat*K regions for distinguishing species such as *C. bicolor, C. aloifolium,* and *C. atropurpureum*, whereas the ITS region exhibited significant intraspecific diversity in *C. finlaysonianum* and *C. aloifolium.* This

result aligns with previous morphological studies of the group (unpublished), which showed diversity in morphology and flower coloration in *C. finlaysonianum* and *C. aloifolium*. From these findings, it is evident that the *rbc*L and *mat*K gene regions are suitable for species identification among the four *Cymbidium* species and are more effective when used in combination. Although the ITS region exhibits intraspecific diversity in *C. finlaysonianum* and *C. aloifolium*, its capacity for species identification is limited, potentially leading to confusion in some cases. Its ability for species identification is restricted and may be confusing in some cases.

ACKNOWLEDGMENTS

The authors would like to sincerely thank the Department of Science and Technology of Ho Chi Minh City for providing funding under contract number 07/2023/HĐ-QKHCN dated March 09, 2023, between the Science and Technology Development Fund of Ho Chi Minh City and the Biotechnology Center of Ho Chi Minh City for the implementation of the scientific and technological task "Research on conservation and evaluation of the gene pool of hard-leaf *Cymbidium* (*Cymbidium* spp.) in Vietnam.". The Biotechnology Center of Ho Chi Minh City provided the equipment and facilities necessary to conduct this research.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

Asahina H., Shinozaki J., Masuda K., Morimitsu Y., and Satake M. (2010). Identification of medicinal *Dendrobium* species by phylogenetic analyses using *mat*K and *rbc*L sequences. *Journal of Natural Medicines*, *64*(2), 133-138. http://doi.org/10.1007/s11418-009-0379-8

Berg C. V. D., Ryan A., Cribb P. J., and Chase M. W. (2002). Molecular phylogenetics of *Cymbidium* (Orchidaceae: Maxillarieae): sequence data from internal transcribed spacers (ITS) of nuclear ribosomal DNA and plastid *mat*K. *Lindleyana*, *17*(2), 101-111. http://www.cabidigitallibrary.org/doi/full/10.5555/20023134595

Cameron K. M. (2004). Utility of plastid *psa*B gene sequences for investigating intrafamilial relationships within Orchidaceae. *Molecular Phylogenetics and Evolution*, *31*(3), 1157-1180. http://doi.org/10.1016/j.ympev.2003.10.010

Cameron K. M. (2006). A comparison and combination of plastid *atp*B and *rbc*L gene sequences for inferring phylogenetic relationships within Orchidaceae. *Aliso*, *22*, 447-464. http://doi.org/10.5642/aliso.20062201.36

Cameron K. M., Chase M. W., Whitten W. M., Kores P. J., Jarrell D. C., Albert V. A., *et al.* (1999). A phylogenetic analysis of the Orchidaceae: evidence from *rbc*L nucleotide sequences. *American Journal of Botany*, *86*(2), 208-224. http://doi.org/10.2307/2656938

Cheng T., Xu C., Lei L., Li C., Zhang Y., and Zhou S. (2016). Barcoding the kingdom Plantae: new PCR primers for ITS regions of plants with improved universality and specificity. *Molecular Ecology Resources*, *16*(1), 138-149. http://doi.org/10.1111/1755-0998.12438

Farrington L., MacGillivray P., Faast R., and Austin A. (2009). Investigating DNA barcoding options for the identification of *Caladenia* (Orchidaceae) species. *Australian Journal of Botany*, *57*(4), 276-286. http://doi.org/10.1071/BT08138

Fazekas A. J., Kuzmina M. L., Newmaster S. G., and Hollingsworth P. M. (2012). DNA barcode

methods for land plants. *In DNA barcodes: Methods and Protocols.* Springer. http://doi.org/10.1007/978-1-61779-591-6_11

Freudenstein J. V., and Chase M. W. (2015). Phylogenetic relationships in *Epidendroideae* (Orchidaceae), one of the great flowering plant radiations: progressive specialization and diversification. *Annals of Botany*, *115*(4), 665-681. http://doi.org/10.1093/aob/mcu253

Gao B., Yu J. Z., Li S. X., Ding X., He Q. F., and Wang X. M. (2011). Roadside and rooftop measurements of polycyclic aromatic hydrocarbons in PM 2.5 in urban Guangzhou: evaluation of vehicular and regional combustion source contributions. *Atmospheric Environment*, *45*(39), 7184-7191. http://doi.org/10.1016/j.atmosenv.2011.09.005

Group C. P. W., Hollingsworth P. M., Forrest L. L., Spouge J. L., Hajibabaei M., Ratnasingham, S.*, et al.* (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences*, *106*(31), 12794-12797. http://doi.org/10.1073/pnas.0905845106

Jalil M., Annuar M. S. M., Tan B. C., and Khalid N. (2015). Effects of selected physicochemical parameters on zerumbone production of *Zingiber zerumbet* Smith cell suspension culture. *Evidence-Based Complementary and Alternative Medicine*, *2015*(1), 1-7. http://doi.org/10.1155/2015/757514

Kress W. J., García R. C., Uriarte M., and Erickson D. L. (2015). DNA barcodes for ecology, evolution, and conservation. *Trends in Ecology & Evolution*, *30*(1), 25-35. http://doi.org/10.1016/j.tree.2014.10.008

Nayak N. R., Tanaka M., Teixeira, D., and Silva J. (2006). *Biotechnology of Cymbidium-An overview of recent progress and future opportunities* (1nd ed.). Global Science Books.

Newmaster S. G., Grguric M., Shanmughanandhan D., Ramalingam S., and Ragupathy S. (2013). DNA barcoding detects contamination and substitution in North American herbal products. *BMC Medicine*, *11*(1), 222. http://doi.org/10.1186/1741-7015-11-222

Paula S. (2013). *A comparison of the efficiency of DNA barcode regions in a small and a large genus*. University of the Free State.

Puy D. D., and Cribb P. (2007). *The genus Cymbidium*. Royal Botanic Gardens, Kew.

Rajaram M. C., Yong C. S., Gansau J. A., and Go R. (2019). DNA barcoding of endangered *Paphiopedilum* species (Orchidaceae) of Peninsular Malaysia. *Phytotaxa*, *387*(2), 94–104. http://doi.org/10.11646/phytotaxa.387.2.2

Sharma S. K., Dkhar J., Kumaria S., Tandon P., and Rao S. R. (2012). Assessment of phylogenetic inter-relationships in the genus *Cymbidium* (Orchidaceae) based on internal transcribed spacer region of rDNA. *Gene*, *495*(1), 10-15. http://doi.org/10.1016/j.gene.2011.12.052

Siripiyasing P., Kaenratana, K., Mokkamul P., Tanee T., Sudmoon R., and Chaveerach A. (2012a). DNA barcoding of the *Cymbidium* species (Orchidaceae) in Thailand. *African Journal of Agricultural Research*, *7*(3), 393-404. http://doi.org/10.5897/AJAR11.1434

Siripiyasing P., Kaenratana K., Mokkamul P., Tanee T., Sudmoon R., and Chaveerach A. (2012b). DNA barcoding of the *Cymbidium* species (Orchidaceae) in Thailand. *African Journal of Agricultural Research*, *7*(13), 393-404. http://doi.org/10.5897/AJAR11.1434

Su'udi M., and Setyati D. (2022). DNA Barcoding of *Vanda tricolor Lindl.* Based on matK, rbcL and ITS2 Sequences. *4th International Conference on Life Sciences and Biotechnology (ICOLIB 2021). Atlantis Press*, 596-604. http://doi.org/10.2991/978-94-6463-062-6_61

Tran H. (2000). *The Orchids of Vietnam*. Ho Chi Minh City, Viet Nam: Agricultural Publishing House.

Yang J. B., Tang M., Li H. T., Zhang Z. R., and Li D. Z. (2013). Complete chloroplast genome of the genus *Cymbidium*: lights into the species identification, phylogenetic implications and population genetic analyses. *BMC Evolutionary Biology*, *13*(1), 84-95. http://doi.org/10.1186/1471-2148-13-84

Yukawa T., Miyoshi K., and Yokoyama J. (2002). Molecular phylogeny and character evolution of *Cymbidium* (Orchidaceae). *Bulletin of the National Science Museum, Tokyo, 28*, 129-139.

Zhang L., Huang Y. W., Huang J. L., Ya J. D., Zhe M. Q., Zeng C. X.*, et al.* (2023). DNA barcoding of *Cymbidium* by genome skimming: Call for next-generation nuclear barcodes. *Molecular Ecology Resources*, *23*(2), 424-439. http://doi.org/10.1111/1755-0998.137 19

Zhu S., Liu Q., Qiu S., Dai J., and Gao X. (2022). DNA barcoding: an efficient technology to authenticate plant species of traditional Chinese medicine and recent advances. *Chinese Medicine*, *17*(1), 1-17. http://doi.org/10.1186/s13020-022-00655-y

Worthy S. J., Bucalo K., Perry E., Reynolds A., Cruse-Sanders J., Pérez Á. J.*, et al.* (2022). Ability of *rbc*L and *mat*K DNA barcodes to discriminate between montane forest orchids. *Plant Systematics and Evolution*, *308*(3), 1-9. http://doi.org/10.1007/s00606-022-01809-z