Al-Hayat: Journal of Biology and Applied Biology
Volume 8, No 1 (2025): 59-68
DOI. 10.21580/ah.v8i1.25671.

In Silico Approach to DNA Barcoding in *Cattleya* Orchids Using *nrDNA* and *matK* Markers

Risma Wiharyanti[1*], Ixora Sartika Mercuriani[1], Evy Yulianti[1], Paramita Cahyaningrum Kuswandi[1]

[1]Department of Biology, Universitas Negeri Yogyakarta, Indonesia

Abstract

The *Cattleya* genus is among the most commercially valuable orchids. However, identifying these orchids based on morphological and phenotypic traits remains challenging. Accurate identification is essential for distinguishing between native and hybrid species, as well as determining the conservation status of certain *Cattleya* orchids. This study explored the potential of *nrDNA* and *matK* loci as DNA barcoding markers for the *Cattleya* genus using an *in silico* approach. Sequences were retrieved from NCBI and analyzed using ClustalX2 for alignment, BioEdit for format conversion, and MEGA11 for phylogenetic tree construction. The results showed that *nrDNA* exhibited higher genetic variation than *matK*. However, neither phylogenetic tree could precisely discriminate species, as some *Cattleya* species were placed closer to the outgroup. These findings may serve as a reference for the molecular identification of *Cattleya* using DNA barcoding markers.

Keywords: *Cattleya*, *matK*, molecular barcode, *nrDNA*

Introduction

Orchids represent one of the largest families of flowering plants, with approximately 6,000 species thriving across various ecosystems. Their high genetic diversity is crucial for survival and adaptation, supporting the sustainability of natural populations and enabling the development of new varieties for commercial purposes. The beauty and uniqueness of orchids give them significant aesthetic, economic, and ecological value. Additionally, several orchid species contain chemical compounds with therapeutic potential (Perwitasari et al., 2020; Sindiya et al., 2018).

The *Cattleya* genus is among the most commercially valuable orchids. These epiphytic orchids are characterized by thick pseudobulbs and fragrant, exceptionally large blooms, earning them the title "The Queen of Orchids" (Harahap et al., 2023). Due to high market demand, breeders frequently cross *Cattleya* species to produce new variants with enhanced traits and higher commercial value. However, such hybridization has made it increasingly challenging to identify *Cattleya* orchids based solely on morphological and phenotypic characteristics (Buddhachat et al., 2022). Accurate identification is essential for distinguishing native from hybrid species and for determining the conservation status of certain *Cattleya* orchids, as several species like *C. labiata*, *C. granulosa*, and *C. walkeriana* are endangered (Galetti, 2023; Rivera-Jiménez et al., 2017).

**Corresponding author: Risma Wiharyanti, email: rismawiharyanti@uny.ac.id, Universitas Negeri Yogyakarta , Jl. Colombo No.1, Karang Malang, Caturtunggal, Kec. Depok, Kabupaten Sleman, Daerah Istimewa Yogyakarta 55281*

DNA barcoding, a molecular identification technique, offers a more precise method for species identification. This approach employs standardized short DNA sequences from specific genomic regions to detect genetic variation within and among species (Mukaromah et al., 2023; Rahayu & Jannah, 2019). Molecular markers facilitate rapid, consistent, and accurate identification using small tissue samples from any part of the plant (Perwitasari et al., 2020; Sindiya et al., 2018; Sunaryo, 2015). DNA barcoding typically uses gene regions from nuclear DNA (nDNA), chloroplast DNA (cpDNA), and mitochondrial DNA (mtDNA) for species classification and phylogenetic reconstruction (Martiansyah, 2021; Rahayu & Jannah, 2019).

The maturase K (*matK*) gene and the ribulose-1,5-bisphosphate carboxylase/ oxygenase large subunit (*rbcL*) gene are two plastid genes recommended by the Consortium for the Barcode of Life (CBOL) as standard molecular markers for plant identification (CBOL Plant Working Group, 2009). However, in some plant species, the Internal Transcribed Spacer (ITS) region performs better than *matK* for species discrimination. The ITS region of the nuclear genome is easily amplified and highly variable among closely related species; ITS2, in particular, has demonstrated a success rate of up to 92.7% in species identification, making it widely used in molecular taxonomy and phylogenetics (Martiansyah, 2021). Barcoding studies on the *Cattleya* genus have successfully distinguished species such as *C. walkeriana* and *C. loddigesii* using ITS1, ITS2, and *rpoC1* markers. Therefore, further exploration of molecular markers in *Cattleya* is warranted (Rivera-Jiménez et al., 2017). The present study investigated the potential of nuclear ribosomal DNA (*nrDNA*), particularly the ITS1-5.8S-ITS2 region, along with the *matK* locus, as effective molecular markers for DNA barcoding of *Cattleya* species using an *in silico* approach.

Research Method

DNA sequences were obtained from the GenBank database (https://www.ncbi.nlm.nih.gov/) using the search feature in the nucleotide section by entering the species name and the desired DNA target. The targeted sequences were *matK* and *nrDNA* containing the ITS1-5.8S-ITS2 region, which were used as molecular markers. The collected sequence information included the accession number and nucleotide length (whether the sequences were partial or complete). This information was stored in a Microsoft Excel file, and the nucleotide sequences were saved in FASTA format for further analysis.

All nucleotide data from each locus were processed by performing sequence alignment using the CLUSTALX2 software. The purpose of the alignment was to identify differences and similarities among the sequences, as well as to explore sequences that could potentially serve as DNA barcodes. The *.aln output format from CLUSTALX2 was further processed in BioEdit and saved as *.FAS files, which were used to construct a phylogenetic tree with MEGA11 software. In constructing the phylogenetic tree, *Dendrobium officinale* and *Vanilla planifolia* were used as outgroup species, as they are distantly related to *Cattleya*. This study employed the Neighbor-Joining (NJ) method with 1000 bootstrap replications and the Tamura 3-parameter model (Pratiwi et al., 2023). The NJ method constructs phylogenetic trees based on genetic distance, enabling the determination of the level of relatedness among *Cattleya* species.

Research Results and Discussion

Identifying species based solely on morphological characteristics has proven inadequate (Mukaromah et al., 2023; Mursyidin et al., 2021). In the case of *Cattleya*, morphological identification frequently leads to misclassification. Therefore, molecular identification through a genomic approach, specifically DNA barcoding, is necessary. This method utilizes short DNA sequences as molecular markers to efficiently identify species. Among the most commonly used markers for *Cattleya* are *nrDNA* and *matK*. Given their respective advantages, the present study compared these two markers to assess their potential for identifying *Cattleya* species.

A total of 36 *nrDNA* and 37 *matK Cattleya* sequences were retrieved from the NCBI database (see **Table 1**), originating from studies conducted in several countries. Most *matK* sequences were derived from Brazilian studies, with a few from Canada (2) and the USA (2). In

contrast, *nrDNA* sequences were predominantly from the UK, with additional sequences from Poland (1) and Brazil (3). Sequence homology is indicated by the number of asterisks (*) in the alignment results, denoting identical nucleotides and conserved regions among species. Gaps represent insertions, deletions, or rearrangements of genetic material during evolution (Sindiya et al., 2018). A greater number of asterisks corresponds to a higher degree of sequence homology.

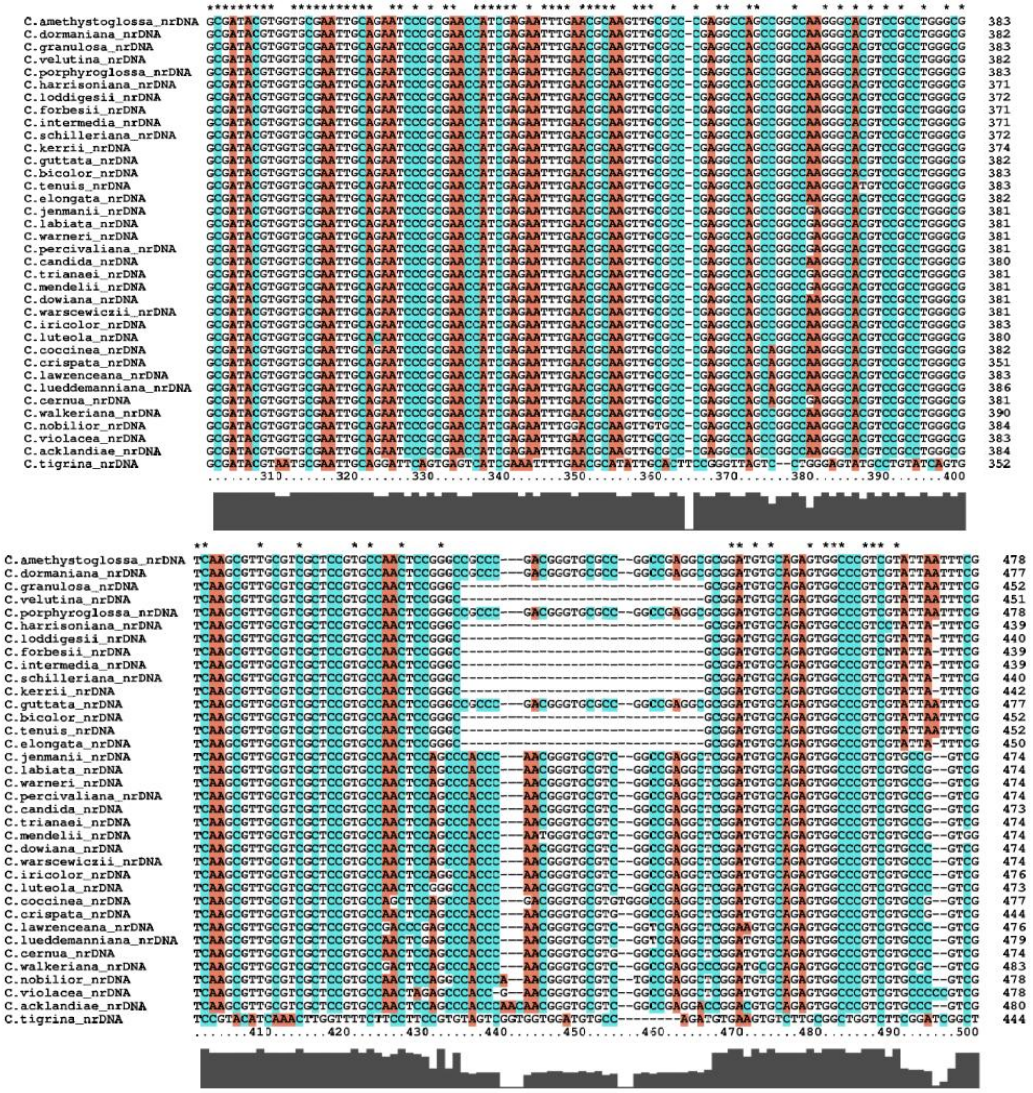
The alignment results (see **Figure 1**) show that *nrDNA* markers exhibit a lower degree of homology, indicating a high mutation rate and greater interspecies variation. This finding is consistent with previous studies, which reported that *nrDNA* regions have high mutation rates and haplotype diversity in orchids (Pratiwi et al., 2023).

Table 1. *nrDNA* and *matK* Sequence Data from NCBI

Species	*matK*		*nrDNA*	
	Accession	Length (bp)	Accession	Length (bp)
Cattleya bicolor	EU139961.1	797	AY008625.1	615
Cattleya walkeriana	EU139990.1	792	KY006872.1	646
Cattleya trianae	EU139987.1	802	AY008602.1	632
Cattleya coccinea	EU140001.1	788	AF260201.1	635
Cattleya labiata	EU139973.1	747	AF260214.1	632
Cattleya forbesii	EU139965.1	752	AY429394.1	604
Cattleya intermedia	EU139969.1	792	AF260204.1	605
Cattleya loddigesii	EU139975.1	747	KY006869.1	607
Cattleya nobilior	GQ248092.1	789	AY008607.1	641
Cattleya aclandiae	GQ248091.1	794	AF260207.1	637
Cattleya dowiana	EU139958.1	802	AY008593.1	632
Cattleya cernua	EU140000.1	797	AY429395.1	634
Cattleya porphyroglossa	EU139980.1	800	AY008612.1	644
Cattleya guttata	EU139967.1	785	AY008609.1	641
Cattleya luteola	EU139977.1	801	AY008605.1	632
Cattleya percivaliana	MT518350.1	823	AY008599.1	632
Cattleya tigrina	EU139986.1	790	OR644503.1	604
Cattleya granulosa	EU139966.1	790	AY008621.1	615
Cattleya crispata	EU140003.1	793	AY008665.1	606
Cattleya warscewiczii	EU139992.1	798	AY008603.1	632
Cattleya warneri	EU139991.1	792	AY008598.1	632
Cattleya violacea	EU139989.1	800	AF20206.1	641
Cattleya velutina	EU139988.1	798	AY008618.1	614
Cattleya tenuis	EU139985.1	803	AY008622.1	519
Cattleya schilleriana	EU139982.1	792	AY008614.1	606
Cattleya rex	EU139981.1	749	N/A	—
Cattleya mendelii	EU139978.1	727	AY008597.1	632
Cattleya lueddemanniana	EU139976.1	802	AF266744.1	639
Cattleya lawrenceana	EU139974.1	764	AF260208.1	638
Cattleya kerrii	EU139972.1	800	AY008613.1	605
Cattleya jenmanii	EU139971.1	792	AY008604.1	632
Cattleya iricolor	EU139970.1	802	AY008606.1	634
Cattleya harrisoniana	EU139968.1	788	AY008615.1	605
Cattleya elongata	EU139964.1	802	AY008619.1	614
Cattleya dormaniana	MT518339.1	798	AY008608.1	640
Cattleya candida	EU139962.1	805	AY008600.1	631

Cattleya amethystoglossa	EU139957.1	791	AY008610.1	641
Dendrobium officinale	MG760737.1	650	GU339109.1	636
Vanilla planifolia	MF349972	814	AF030049.1	622

Figure 1. Alignment of *Cattleya nrDNA* sequences. The asterisk (*) indicates homology, while gaps represent insertions or deletions. Each color denotes a specific nitrogenous base.



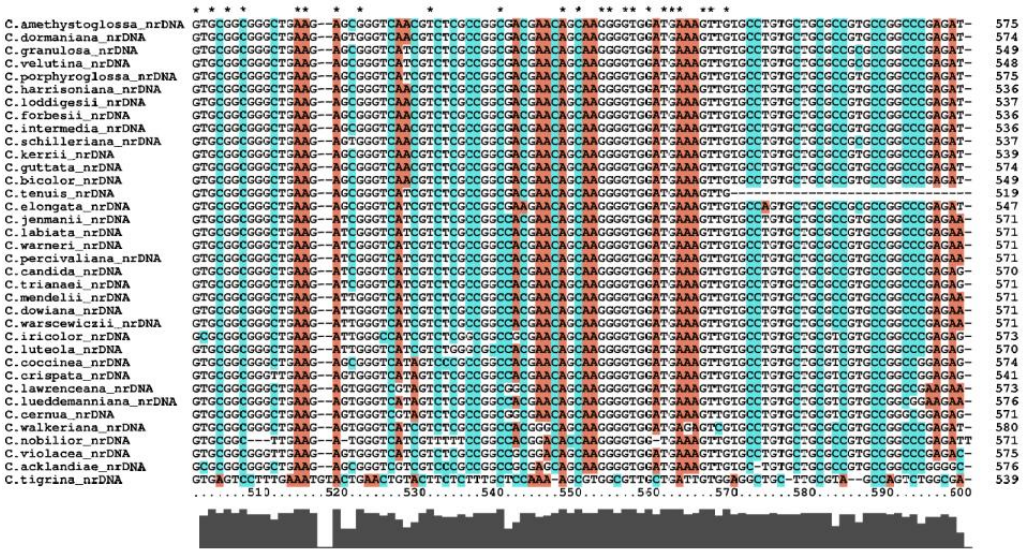
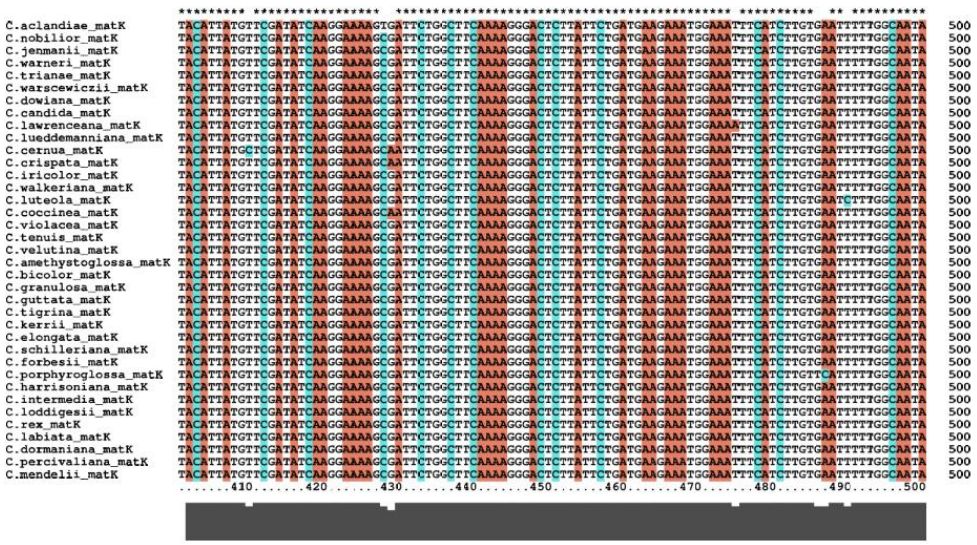


Figure 2. Alignment of *Cattleya matK* sequences. The asterisk (*) indicates homology, while gaps represent insertions or deletions. Each color denotes a specific nitrogenous base.



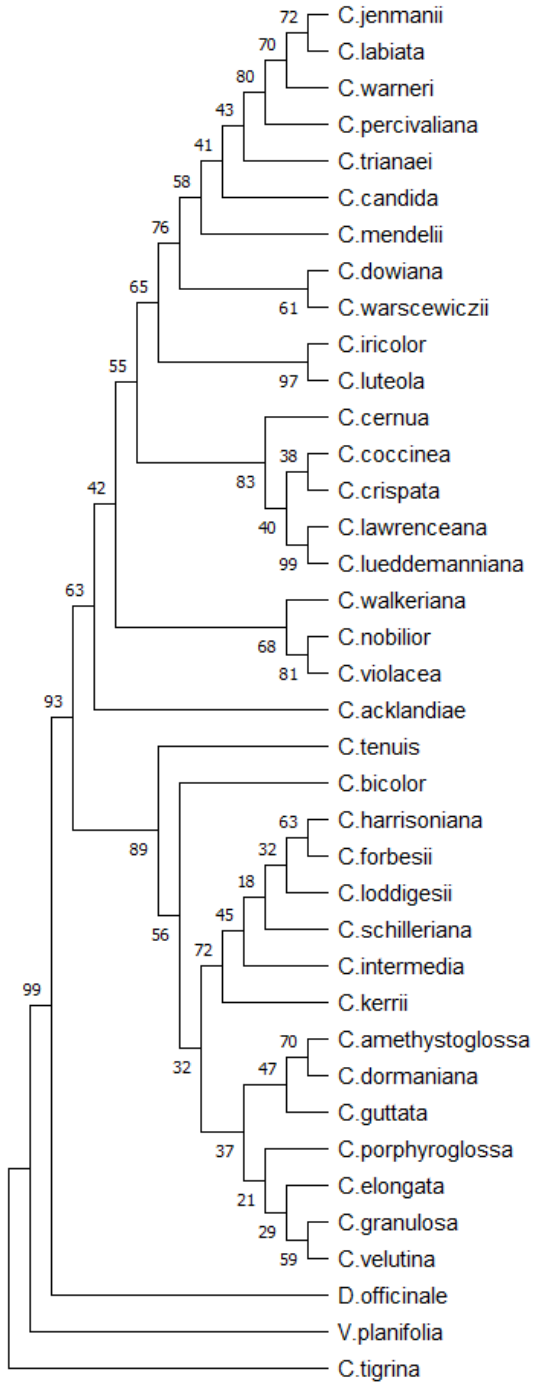
The *nrDNA* sequence includes the ITS, located between the small-subunit ribosomal RNA genes. The 5.8S region of *nrDNA* is highly conserved, whereas the ITS regions are nuclear loci that undergo frequent recombination, leading to increased genetic variation among closely related taxa (Pratiwi et al., 2023; Wathon et al., 2023). ITS1 and ITS2 exhibit higher mutation rates, making them promising markers for species-level identification.

As shown in Figure 1, the variable region at positions 434–468 of the nrDNA alignment may serve as a candidate barcode for distinguishing Cattleya species. Studies have demonstrated that the ITS region is more effective in differentiating species, with higher pairwise genetic distances than other markers. For instance, ITS outperformed rbcL, rpoB, rpoC1, and matK in identifying Indian orchids (Parveen et al., 2017; Rivera-Jiménez et al., 2017; Srivastava & Manjunath, 2020). Similar results were observed in Paphiopedilum and Coelogyne spp., where ITS was the most suitable barcode compared to plastid regions (Pratiwi et al., 2023; Sindiya et al., 2018). ITS also successfully distinguished between Dendrobium discolor, C. walkeriana, and C. loddigesii (Perwitasari et al., 2020; Rivera-Jiménez et al., 2017).

The matK marker showed a higher degree of homology among Cattleya species than nrDNA (see Figure 2), suggesting a lower mutation rate and reduced interspecies variation. As maternally inherited chloroplast DNA, matK has stable and conserved characteristics, resulting in limited genetic variability (Sindiya et al., 2018). While matK is effective at distinguishing species at the intergeneric level within Orchidaceae, it is less effective for resolving phylogenies within the genus Dendrobium (Chattopadhyay et., 2017). Additionally, matK showed lower specificity as a barcode for Paphiopedilum (Sindiya et al., 2018).

The phylogenetic trees (see **Figure 3** and **Figure 4**) illustrate the genetic relationships among *Cattleya* species. The tree based on *nrDNA* successfully separated *D. officinale* and *V. planifolia* as outgroups. However, it failed to correctly place *C. tigrina*, classifying it as an outgroup rather than within *Cattleya*. According to Van Den Berg (2014), *Cattleya* has been reclassified into four subgenera: *Cattleya*, *Cattleyella*, *Intermedia*, and *Maximae*. The subgenus *Cattleya* includes three sections: *Cattleya*, *Crispae*, and *Lawrenceanae*.

Figure 3. Phylogenetic tree constructed using *nrDNA* sequences



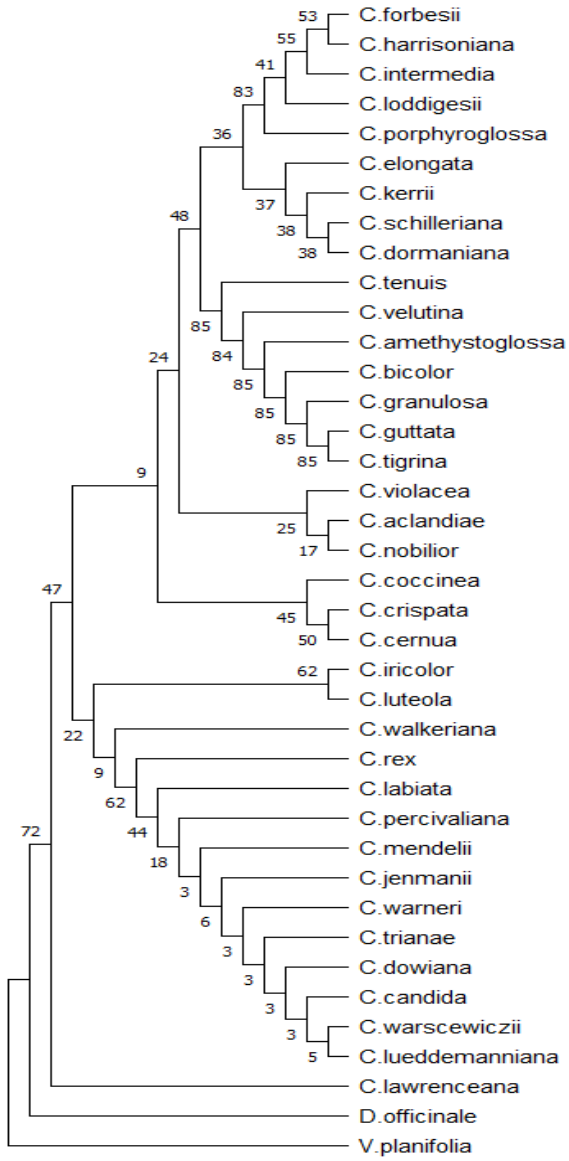
The *nrDNA*-based phylogeny (see **Figure 3**) clustered the subgenus *Cattleya*, section *Cattleya* (e.g., *C. jenmanii*, *C. labiata*, *C. warneri*, *C. percivaliana*, *C. trianaei*, *C. mendelii*, *C. dowiana*, *C. warscewiczii*, *C. iricolor*, and *C. luteola*) into a single clade. Similarly, section *Crispae* (e.g., *C. coccinea*, *C. crispata*, and *C. cernua*) also formed a distinct group. However, the *nrDNA* tree positioned some members of subgenus *Intermedia* (e.g., *C. walkeriana*, *C. nobilior*, *C. violacea*, *C. aclandiae*) closer to subgenus *Cattleya*, although they still formed a separate cluster.

Prior studies confirmed that *nrDNA* could yield strong phylogenetic resolution due to the conserved nature of the 5.8S region and the variability of ITS (Pratiwi et al., 2023). A higher-resolution tree was obtained using ITS for seven endemic Indian orchid genera compared to other markers (Srivastava & Manjunath, 2020), highlighting ITS's potential for species-level systematics.

In contrast, the phylogenetic tree based on *matK* (see **Figure 4**) grouped subgenus *Intermedia* into a single clade derived from a common ancestor, although *C. walkeriana*, a member of subgenus *Cattleya*, was also included in this clade. Subgenus *Intermedia* comprised *C. forbesii*, *C. harrisoniana*, *C. intermedia*, *C. loddigesii*, *C. porphyroglossa*, *C. elongata*, *C. kerrii*, *C. schilleriana*, *C. dormaniana*, *C. tenuis*, *C. velutina*, *C. amethystoglossa*, *C. bicolor*, *C. granulosa*, *C. guttata*, *C. tigrina*, *C. violacea*, *C. aclandiae*, and *C. nobilior*. This marker also clustered section *Cattleya* into a single clade

.

Figure 4. Phylogenetic tree constructed using *matK* sequences



Like the *nrDNA* tree, the *matK*-based phylogeny separated *D. officinale* and *V. planifolia* as outgroups. However, *C. lawrenceana*, which should belong to section *Lawrenceanae* along with *C. lueddemanniana*, appeared closely related to the outgroup. Additionally, section

Crispae (*C. coccinea*, *C. crispata*, *C. cernua*) appeared more closely related to subgenus *Intermedia* than to section *Cattleya* within subgenus *Cattleya*.

Conclusion

The *nrDNA* sequences containing ITS1-5.8S-ITS2, used as molecular markers for species identification in *Cattleya*, exhibited higher genetic variation than *matK*. This greater variation arose because ITS was located in the nuclear genome and underwent frequent recombination. Additionally, ITS sequences had a higher mutation rate than *matK* sequences, contributing to their genetic diversity. In contrast, *matK* sequences were part of the chloroplast DNA, which was maternally inherited and more conserved, leading to limited genetic variation. Both *nrDNA* and *matK* markers have proven effective in distinguishing *Cattleya* species. However, neither phylogenetic tree could accurately resolve all species. For instance, *C. tigrina* was positioned closer to the outgroup in the *nrDNA*-based phylogenetic tree, and *C. lawrenceana* appeared similarly misplaced when using *matK* as the marker. Expanding the search for additional DNA barcode markers in the *Cattleya* genus may enable more comprehensive phylogenetic analysis and yield more accurate results through better species representation.

Acknowledgment

The authors express their gratitude to Universitas Negeri Yogyakarta for providing financial support through the Research Group Grant (contract number B/86/UN34.13/PM.01.01/2024).

References

Buddhachat, K., Sripairoj, N., Punjansing, T., Kongbangkerd, A., Inthima, P., Tanming, W., & Kosavititkul, P. (2022). Species discrimination and hybrid detection in terrestrial orchids using Bar-HRM: a case of the Calanthe group. *Plant Gene*. https://doi.org/https://doi.org/10.1016/j.plgene.2021.100349

CBOL Plant Working Group. (2009). *A DNA barcode for land plants*. *106*(31).

Galetti, P. M. (2023). Conservation Genetics in the Neotropics. In *Conservation Genetics in the Neotropics* (Vol. 3). Springer International Publishing. https://doi.org/10.1007/978-3-031-34854-9

Harahap, F., Hariyadi, I., Silitonga, M., Suryani, C., Edi, S., & Ningsih, A. P. (2023). In vitro Growth of Cattleya sp Orchid from Leaf Explants with Growth Regulators. *Jurnal Pembelajaran Dan Biologi Nukleus*, *9*(1), 192–200. https://doi.org/10.36987/jpbn.v9i1.3945

Martiansyah, I. (2021). *Mini Review: Pendekatan Molekuler DNA Barcoding: Studi Kasus Identifikasi dan Analisis Filogenetik Syzygium (Myrtaceae)*. http://journal.uin-alauddin.ac.id/index.php/psb

Mukaromah, A. S., Ulfah, M., Rachmah, A. N., Arfan, M. R., Kusumarini, N., & Febriana, A. (2023). Authentication of Wax Apple (Syzygium samarangense (Blume) Merr. & L.M Perry) Delima and Citra Cultivars by Morphological and Molecular Approach. *Al-Hayat: Journal of Biology and Applied Biology*, *6*(1), 77–86. https://doi.org/10.21580/ah.v6i1.18713

Mursyidin, D. H., Ahyar, G. M. Z., Saputra, A. W., & Hidayat, A. (2021). Genetic Diversity and Relationships of Phalaenopsis Based on the rbcL and trnL-F Markers: In Silico Approach. *Biosaintifika*, *13*(2). https://doi.org/10.15294/biosaintifika.v13i2.29904

Parveen, I., Singh, H. K., Malik, S., & Raghuvanshi, S. (2017). Evaluating five different loci (rbcL, rpoB, rpoC1, matK, and ITS) for DNA barcoding of Indian orchids. *Genome*. https://doi.org/10.1139/gen-2016-0215

Perwitasari, D. A. G., Rohimah, S., Ratnasari, T., Sugiharto, B., & Su'udi, M. (2020). DNA Barcoding of Medicinal Orchid Dendrobium discolor Lindl. Tanimbar Using rbcL and ITS genes. *Buletin Penelitian Tanaman Rempah Dan Obat*, *31*(1), 8. https://doi.org/10.21082/bullittro.v31n1.2020.8-20

Pratiwi, A., Kinasih, A., Meidianing, M. I., Kurniawan, F. Y., & Semiarti, E. (2023). In Silico Approach for DNA Barcoding using Phylogenetic Analysis of Coelogyne spp. based on the matK, rpoC1, rbcL and nrDNA Markers. *Journal of Tropical*

Biodiversity and Biotechnology, *8*(3), 1–14. https://doi.org/10.22146/jtbb.73130

Rahayu, D. A., & Jannah, M. (2019). *Dna Barcode Hewan Dan Tumbuhan Indonesia*. 9–25.

Rivera-Jiménez, H., Rossini, B. C., Tambarussi, E. V., Veasey, E. A., Ibanes, B., & Marino, C. L. (2017). DNA barcode regions for differentiating Cattleya walkeriana and C. loddigesii. *Acta Scientiarum - Biological Sciences*, *39*(1), 45–52. https://doi.org/10.4025/actascibiolsci.v39i1.33024

Sindiya, V., Mukarramah, L., Rohimah, S., Al GhifariPerwitasari, D., & Su'udi, M. (2018). Studi In Silico Potensi DNA Barcode pada Anggrek Langka Paphiopedilum. *BIOSFER : Jurnal Biologi Dan Pendidikan Biologi*, *3*(1). https://doi.org/10.23969/biosfer.v3i1.1250

Srivastava, D., & Manjunath, K. (2020). DNA barcoding of endemic and endangered orchids of India: A molecular method of species identification. *Pharmacognosy Magazine*. https://phcog.com/article/view/2020/16/70/290-299

Sunaryo, W. (2015, September 1). *Aplikasi DNA Barcoding untuk analisis keragaman genetik lai-durian (Durio zibethinus x kutejensis) asal Kalimantan Timur*. https://doi.org/10.13057/psnmbi/m010602

Van Den Berg, C. (2014). Reaching a compromise between conflicting nuclear and plastid phylogenetic trees: A new classification for the genus Cattleya (Epidendreae; Epidendroideae; Orchidaceae). *Phytotaxa*, *186*(2), 075–086. https://doi.org/10.11646/phytotaxa.186.2.2

Wathon, S., Astikaningrum, D., Paramitha, N., Ardyah, C., Oktarianti, R., & Senjarini, K. (2023). In silico exploration of the potential barcode DNA in Anopheles sp., a malarian vector from Indonesia. *Jurnal Biolokus: Jurnal Penelitian Pendidikan Biologi Dan Biologi*, *6*(1).