

Identification of Indonesian Endemic Orchidaceae using Sequence Alignment and Genetic Distance

Tutut Indria Permana¹, Navik Khusniah¹, Roimil Latifa¹, Eko Susetyarini^{1,2},
Husamah Husamah^{1,2*}, Diani Fatmawati^{1,3}

¹Department of Biology Education, Faculty of Teacher Training and Education, Universitas Muhammadiyah Malang, Jl. Raya Tlogomas 246 Malang 65144, Indonesia

²Master of Biology Education, Postgraduate Program, Universitas Muhammadiyah Malang, Jl. Raya Tlogomas 246 Malang 65144, Indonesia

³Graduate School of Biotechnology, College of Life Sciences, Kyung Hee University, 1732 Deogyong-daero, Giheung-gu, Yongin-si, Gyeonggi-Do 17104, Republic of Korea

*Corresponding Author: usya_bio@umm.ac.id

Submitted: 2025-12-16. Revised: 2026-02-20. Accepted: 2026-04-23.

Abstract. Orchid is an ornamental plant commodity with a high contribution to horticultural production in Indonesia. Efforts to fulfill market demand have caused the hybridization of many orchids to improve quality and create uniqueness in orchid horticulture. Hybridization is influenced by the genetic distance between orchid species to be crossing over. This study aimed to analyze the genetic relationship of the Indonesian endemic Orchidaceae family using sequence alignment and genetic distance. This descriptive exploratory research uses the bioinformatics method with the genetic distance as a variable. The procedure in this study begins with collecting sequence data in FASTA format, aligning sample sequence data, constructing phylogenetic trees, computing amino acid composition, computing Pairwise Distance, and ends with data analysis. The sample was 16 species of orchids that were selected using a purposive sampling technique with complete sequence data in GenBank. The genetic relationships of the 16 species consist of two major clades. Each clade has a pair of species connected by a straight vertical line indicating genetic distance <0.003 or close category. In the first clade, there are three pairs of closely related species (*Vanda pumila* – *Dendrobium trilmellatum*, *Cymbidium dayanum* – *Aerides odorata*, and *Coelogyne pandurata* – *Coleogyne asperata*). While in the second clade, only one pair of species is closely related, namely *Phalaenopsis corningiana* – *Phalaenopsis tetrapis*. This research information confirms the DNA-based identification of endemic Indonesian orchids, which has a positive impact on the accuracy of biodiversity inventory and monitoring. This research supports SDG 5. The data provides a scientific basis for terrestrial ecosystem conservation and prevents biodiversity loss through more accurate species identification.

Keywords: Genetic distance; Orchidaceae; sequence alignment

How to Cite: Permana, T. I., Khusniah, N., Latifa, R., Susetyarini, E., Husamah, H., & Fatmawati, D. (2026). Identification of Indonesian Endemic Orchidaceae using Sequence Alignment and Genetic Distance. *Biosaintifika: Journal of Biology & Biology Education*, 18(1), 174-184.

DOI: <http://dx.doi.org/10.15294/biosaintifika.v18i1.38007>

INTRODUCTION

Orchids in Indonesia with high diversity have produced a variety of genetic patterns that differ between populations, making Indonesia known as the world's mega-diversity of orchids (Sun et al., 2024). Orchid or Orchidaceae is a community with the most species compared to other flowering plant families, with vegetation ranging from beaches, mangrove forests, meadows, and mountains (Mirioba et al., 2024). The morphology of Orchidaceae consists of sepals, petals, labellum, and columns with unique characteristics that are different for each species (Takeda et al., 2025).

The Central Bureau of Statistics explains that orchid production was 11.35 million in 2021. Thus, orchids are an ornamental plant commodity that contributes to horticultural production in Indonesia. Large quantities of high-quality seeds were needed to fulfill this demand (Vilcherrez-Atoche et al., 2022). Cross-breeding (hybridization) is one way to improve the quality of orchid plants. Orchids can not only be crossed with species of the same genus as other ornamental plants but can also be crossed with species between genera so that they will have the opportunity to create uniqueness in horticulture (Han et al., 2025). According to Iiyama et al. (2024) the results of crosses known as hybrid

orchids will always have their market share in line with the growing public interest in maintaining orchids.

The efficacy of hybridization hinges upon the degree of consanguinity among the implicated species. Proximity in genetic relatedness between the parental orchid species positively correlates with the probability of a successful interspecific cross (Zhang et al., 2019). Adherence to a common ancestral lineage or monophyletic origin is imperative in hybridization endeavors. Species diverging from monophyletic criteria are predisposed to manifest rejection phenomena during the hybridization process, attributable to genetic incongruities (Prihartin et al., 2023; J. Wang et al., 2025). To surmount these plausible hurdles, a comprehensive understanding of the genetic interplay among the species earmarked for hybridization is recommended, facilitated through phylogenetic analysis (Nuryady et al., 2019; Z. Wang et al., 2025).

In phylogenetics, species or individuals that have many similar characters will be categorized as closely related from one ancestor so that their descendants will form a monophyletic group (Zander, 2025). In phylogenetic analysis, outgroup groups will form apomorphic and plesiomorphic characters (Post et al., 2023). Apomorphic characters are characters whose changes are inherited from the ingroup, while plesiomorphic characters are characters with primitive genetics from the outgroup. Synapomorphic characters are characters found in monophyletic groups (Mavrodiev, 2020). In the identification of kinship relationships and phylogenetic analysis of plants, various genes have been extensively employed, among which the *rbcL* gene stands out as prominent.

The *rbcL* gene encodes the large subunit of the ribulose-1,5-bisphosphate carboxylase/oxygenase enzyme, also known as Rubisco. Rubisco plays a role in the initial part of the Calvin cycle as a catalyst for the CO₂-fixing reaction. This enzyme converts carbon dioxide and water into carbohydrates, which play a crucial role in photosynthesis. The *rbcL* gene is typically found in the chloroplast DNA of plants and algae (Fang et al., 2019). This gene is used as a DNA barcode to identify and confirm species within a taxonomic framework. The *rbcL* gene has been widely used in phylogenetic studies to analyze genetic diversity and relationships between different plant species. It has a low mutation rate but is effective and stable in plants, making it quite suitable for analyzing relationships among plants within a

single genus. *RbcL*'s low mutation rate compared to other chloroplast genes, such as *trnH* and *ORF3*, thus retaining ancestral gene information (W. Zhang et al., 2023). Furthermore, this results in protein sequences with high homology across species within the same genus.

Phylogenetic research on orchid relationships has been conducted by Liu et al. (2019) through a phylogenetic analysis of *Dendrobium* orchids using ITS rDNA. The results of this study resulted in the reconstruction of a phylogenetic tree of *Dendrobium*, divided into two main clades: the first clade consisting of eight orchid species from the genus *Dendrobium*, and the second clade consisting of two orchid species from the genus *Dendrobium*. Another study conducted by Chi et al. (2021) reconstructed a phylogenetic tree for the genus *Thrixspermum*. Both studies produced phylogenetic trees similar to this study. The novelty of this study lies in the use of 16 orchid species samples from eight genera distributed throughout Indonesia.

This study also included genetic distance analysis for each species found in Indonesia. Genetic distance analysis is a method used to infer genetic relationships between individuals or groups based on phenotypic data from skeletal remains or genomic coding sequences. It is commonly used in bioarchaeological research when DNA preservation is poor or destructive sampling is not possible. Genetic distance measures the degree of genetic divergence between species or populations and is used to analyze genetic differentiation between populations. Different measures of genetic distance have unique properties and can lead to different evolutionary relationships (P. Liu et al., 2025). In the context of population genetics, genetic distance is often used to estimate population structure and assess genetic relationships between populations (Ragauskas et al., 2025).

A study related to genetic distance using molecular data was conducted on *Zanclea* sp. and the *Zanclea* clade based on the mitochondrial COI gene (Maggioni et al., 2022). These data indicate that both study samples are congruent and similar to each other. The absence of research related to genetic distance analysis on orchids, especially those spread across Indonesia, makes this a recent research issue. This research utilized eight orchid genera endemic to Indonesia, including *Aerides*, *Coelogyne*, *Cymbidium*, *Dendrobium*, *Paphiopedilum*, *Phalaenopsis*, *Oncidium*, *Vanda*.

This study aimed to analyze the genetic

relationship of the Indonesian endemic Orchidaceae family using sequence alignment and genetic distance. This research makes a significant contribution to science and society by providing accurate information on the identification of endemic Indonesian orchids based on DNA data, particularly sequence alignment and genetic distance. This can certainly strengthen biodiversity inventories, monitor closely related species, and inform orchid conservation decisions.

METHODS

This research was carried out using bioinformatics methods using research objects in the form of Orchidaceae family sequence datasets and *rbcL* marker genes obtained through the NCBI (National Center for Biotechnology Information) GenBank on the website <https://www.ncbi.nlm.nih.gov> and the MEGA X application. The variable in this study is a single variable, namely genetic distance.

The *rbcL* gene has 1430 base pairs and functions to code for rubisco (ribulose 1,5-bisphosphate carboxylase/oxygenase). Rubisco is an enzyme that is responsible for CO₂ fixation in the Calvin cycle in many varieties of photosynthetic organisms. This is because RuBisCo produced by the *rbcL* gene is very

abundant in plant cells, namely up to 15% of total plant protein and more than 50% of total protein in chloroplasts. The *rbcL* gene is a chloroplast gene coding region that has a high amplification success rate in various flowering plant species, gymnosperms and cryptogamous species. DNA barcoding of the *rbcL* region typically analyzes partial sequences from either the 5' *rbcLa* region or the 3' *rbcL* region as seen in Figure 1. Multiple sequence alignment of this barcode dataset with sequences from the GenBank database shows very little sequence variation, with only four characteristic single nucleotide polymorphic (SNP) positions were observed at Positions 172, 216, 392 and 431 (Masiero et al., 2020).

The samples in this study used 16 orchid species, namely *Aerides odorata* Lour, *Coelogyne asperata* Lindl., *Coelogyne pandurata* J.J. Sm. & Schltr, *Cymbidium dayanum* Rchb.f, *Dendrobium anosmum* Lindl., *Dendrobium spectabile* (Blume) Miq., *Dendrobium trillamellatum* J.J. Sm, *Paphiopedilum glaucophyllum* J.J. Sm., *Phalaenopsis amboinensis* J.J. Sm., *Oncidium* sp., *Phalaenopsis corningiana* Rchb.f., *Phalaenopsis gigantea* J.J. Sm., *Phalaenopsis pantherina* Rchb.f., *Phalaenopsis tetraspis* Rchb.f., *Phalaenopsis modesta* J.J. Sm., and *Vanda pumila*. Table 1 shows the detail sequence length of *rbcL* gene in the sample.

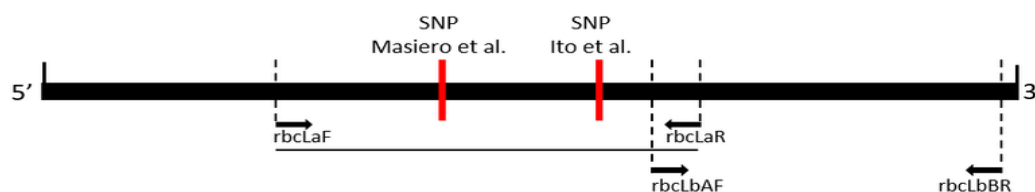


Figure 1. Diagram of the *rbcL* marker gene

Table 1. The research samples

No.	Species	GenBank accession number	Sequence length (bp)
1.	<i>Vanda pumila</i>	JQ180387.1	486
2.	<i>Phalaenopsis pantherina</i> Rchb.f.	AY389429.1	669
3.	<i>Phalaenopsis modesta</i> J.J. Sm.	AY389398.1	681
4.	<i>Phalaenopsis gigantea</i> J.J. Sm.	MT519356.1	553
5.	<i>Phalaenopsis corningiana</i> Rchb.f.	FJ460417.1	681
6.	<i>Phalaenopsis tetraspis</i> Rchb.f.	AY389419.1	681
7.	<i>Paphiopedilum glaucophyllum</i> J.J. Sm.	MT519316.1	553
8.	<i>Oncidium</i> sp.	MW497538.1	553
9.	<i>Dendrobium trillamellatum</i> J.J. Sm	KF496773.1	464
10.	<i>Phalaenopsis amboinensis</i> J.J. Sm.	MT519345.1	546
11.	<i>Dendrobium spectabile</i> (Blume) Miq.	MT519104.1	553
12.	<i>Cymbidium dayanum</i> Rchb.f	MF861081.1	1.293
13.	<i>Dendrobium anosmum</i> Lindl.	MZ223423.1	582
14.	<i>Coelogyne pandurata</i> J.J. Sm. & Schltr	OL537595.1	553
15.	<i>Coelogyne asperata</i> Lindl.	MN877294.1	516
16.	<i>Aerides odorata</i> Lour	JX068006.1	237

Sequence Data Collection

The sequences of the sixteen samples from the Orchidaceae family were obtained through the NCBI (National Center for Biotechnology Information) GenBank database by entering the keywords species name and rbcL marker gene. The sequences taken have different base pair lengths, then saved in FASTA format, where the sequences will then be aligned.

Sequence Alignment

Sequences that have been saved in FASTA format will be analyzed using MEGAX software. The sequence is input, followed by an alignment process using the Clustal W menu, then the sequence is cut to remove gaps so that the results obtained are conservative. Results are saved in fasta format.

Measuring Genetic Distance

The results of sequence analysis from the 16 samples were converted into a genetic distance table using the compute pairwise distance model. The reconstruction results are saved in Excel format.

Reconstruction the Phylogenetic Tree

The sequence alignment results are used to reconstruct a phylogenetic tree in the MEGA application option with a maximum composite likelihood model with a bootstrap value of 1000. The reconstruction results are saved in PNG format.

RESULTS AND DISCUSSION

Sequence Alignment

Sequence alignment is carried out to align and determine the level of homology between sequences so that they can be compared. The sequence alignment display shows that bases that have the same code will have the same color as seen in Table 2. Differences in the sequence of nucleotide bases in the sequence alignment results prove that changes have occurred in genes from species in the same genus. This can be seen in the *Phalaenopsis* genus where the *Phalaenopsis gigantea* species has a different base sequence from species from other *Phalaenopsis* genera.

Table 2. Sequence alignment results of the samples

No.	Species	DNA Sequences
1.	<i>Vanda pumila</i>	TGCGAGCTCTACGTCTGGAAGATCTGCGAATTCCTTCTTCTTCCAAAACCTTTCCAAGG TCCGCCTCATGGC
2.	<i>Phalaenopsis pantherina</i>	TAGGAATAGAAATTTTCAAGAAGATTGGGTTGCGCCATATATATCAAAGAGTCTAAAATA ATGATGTATTTGGT
3.	<i>Phalaenopsis modesta</i>	TAGGAATAGAAATTTTCAAGAAGATTGGGTTGCGCCATATATATCAAAGAGTATAAAATA ATGATGTATTTGGT
4.	<i>Phalaenopsis gigantea</i>	TACGTGCACTACGACTAGAAGATTTACGTATTCCTCCAGCATATGTAAAAACATTCCAAG GTGCTCCACATGGT
5.	<i>Phalaenopsis corningiana</i>	TAGGAATAGAAATTTTCAAGAAGATTGGGTTGCGCCATATATATCAAAGAGTATAAAATA ATGATGTATTTGGT
6.	<i>Phalaenopsis tetraspis</i>	TAGGAATAGAAATTTTCAAGAAGATTGGGTTGCGCCATATATATCAAAGAGTATAAAATA ATGATGTATTTGGT
7.	<i>Paphiopedilum glaucophyllum</i>	TGCGAGCTCTACGTCTGGAAGATCTACGAATTCCTTCTTATTCCAAAACCTTTCCAAGG TCCACCTCATGGC
8.	<i>Oncidium sp.</i>	TGCGAGCTCTACGTCTGGAAGATCTGCGAATTCCTTCTTATTCCAAAACCTTTCCAAGG TCCACCTCATGGT
9.	<i>Dendrobium trilamellatum</i>	TGCGAGCTCTACGTCTGGAAGATCTGCGAATTCCTTCTTATTCCAAAACCTTTCCAAGG TCCGCCTCATGGC
10.	<i>Phalaenopsis amboinensis</i>	TACGTGCACTACGACTAGAAGATTTACGTATTCCTCCAGCATATGTAAAAACATTCCAAG GTGCTCCACATGGT
11.	<i>Dendrobium spectabile</i>	TGCGAGCTCTACGTCTGGAAGATCTGCGAATTCCTTCTTATTCCAAAACCTTTCCAAGG TCCGCCTCATGGC
12.	<i>Cymbidium dayanum</i>	TTCGAGCTTTACGTCTGGAAGATCTGCGAATTCCTTCTTATTCCAAAACCTTTCCAAGG TCCGCCTCATGGC
13.	<i>Dendrobium anosmum</i>	TGCGAGCTCTACGTCTGGAAGATCTGCGAATTCCTTCTTATTCCAAAACCTTTCCAAGG TCCGCCTCATGGC
14.	<i>Coelogyne pandurata</i>	TGCGAGCTCTACGTCTGGAAGATCTACGAATTCCTTCTTATTCCAAAACCTTTCCAAGG TCCGCCTCATGGC
15.	<i>Coelogyne asperata</i>	TGCGAGCTCTACGTCTGGAAGATCTACGAATTCCTTCTTATTCCAAAACCTTTCCAAGG TCCGCCTCATGGC
16.	<i>Aerides odorata</i>	TGCGAGCTCTAAGTCTGGAAGATCTGGGAATTCCTTCTTATTCCAAAACCTTTCAAAGT TCCGCCTCATGGC

Note: T: Timin, C: Sitosin, G: Guanin, A: Adenin

Differences in nucleotide base sequences prove that changes have occurred in genes from species in the same genus. The lower the difference in nucleotide bases of each species, the closer the relationship between species will be. Different nitrogen base sequences cause differences in amino acid sequences. The results of the sequence alignment are followed by data on the probability of substitution from one nucleotide base to another based on the pattern and rate of substitution based on the 1993 Tamura-Nei model as shown in Table 3.

Table 3. Probability of nucleotide base substitution

	A	T/U	C	G
A	-	8.65	7.05	8.28
T/U	7.96	-	9.62	5.52
C	7.96	11.81	-	5.52
G	11.94	8.65	7.05	-

The different levels of substitution are shown in the data display in bold (see Table 3). In the sequence alignment results, it was found that the frequency of the code bases Adenine= 27.28%, Thymine/Uracil= 29.65%, Cytosine= 24.16%, and Guanine= 18.92%. To estimate the probability value in the table, calculations were carried out using the maximum Log, namely -361.454 involving 16 nucleotide sequences. Table 3 shows that in the sequences of the 16 samples, no synonymous substitutions occurred in the form of pyrimidine bases becoming purine or purine bases becoming pyrimidines, and only non-synonymous

substitutions occurred in the form of purine bases becoming pyrimidines. The occurrence of substitution mutations can result in changes in the amino acids formed, thereby affecting gene expression (Oelschlaeger, 2024).

Gene expression refers to the process by which genetic information in DNA is translated into functional proteins. It begins with transcription, where RNA is synthesized from DNA, followed by translation, where ribosomes translate the mRNA sequence into the amino acid sequence in the protein. Amino acids in proteins determine the structure and function of the protein, so that proteins differ from each other according to the sequence of amino acids. The differences in amino acids in each sample for *rbcl* gene can be seen in Table 4.

Slight differences in nucleotide bases cause several species to be in the same clade in phylogenetic tree reconstruction. In a phylogenetic tree, species sequences connected by straight vertical lines indicate that the two species are identical (Luo et al., 2025). As for the first clade, the species connected by straight vertical lines are *Vanda pumila* with *Dendrobium trilamellatum*, *Cymbidium dayanum* with *Aerides odorata*, *Coelogyne pandurata* with *Coelogyne asperata*, and *Phalaenopsis gigantea* with *Phalaenopsis amboinensis*, which are species of the same genus. In the second class, the species connected by a straight vertical line are only *Phalaenopsis corningiana* and *Phalaenopsis tetrapis*.

Table 4. The amino acids composition of samples

Species	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr	Total
1	0.00	8.33	0.00	8.33	4.17	0.00	0.00	4.17	12.50	25.00	4.17	0.00	8.33	0.00	4.17	4.17	0.00	8.33	0.00	4.17	24
2	4.55	4.55	0.00	9.09	4.55	4.55	0.00	13.64	13.64	9.09	9.09	0.00	0.00	0.00	9.09	9.09	0.00	0.00	0.00	9.09	22
3	4.55	4.55	0.00	9.09	4.55	4.55	0.00	13.64	13.64	9.09	9.09	0.00	0.00	0.00	9.09	4.55	0.00	0.00	0.00	13.64	22
4	0.00	0.00	4.55	0.00	4.55	0.00	18.18	4.55	13.64	9.09	9.09	0.00	0.00	4.55	0.00	4.55	0.00	13.64	0.00	13.64	22
5	4.55	4.55	0.00	9.09	4.55	4.55	0.00	13.64	13.64	9.09	9.09	0.00	0.00	0.00	9.09	4.55	0.00	0.00	0.00	13.64	22
6	4.55	4.55	0.00	9.09	4.55	4.55	0.00	13.64	13.64	9.09	9.09	0.00	0.00	0.00	9.09	4.55	0.00	0.00	0.00	13.64	22
7	0.00	4.17	0.00	8.33	4.17	0.00	4.17	8.33	12.50	20.83	4.17	0.00	8.33	0.00	0.00	4.17	0.00	8.33	4.17	8.33	24
8	0.00	8.33	0.00	8.33	4.17	0.00	4.17	8.33	12.50	20.83	4.17	0.00	8.33	0.00	0.00	4.17	0.00	8.33	4.17	4.17	24
9	4.17	8.33	0.00	8.33	4.17	0.00	0.00	8.33	12.50	20.83	4.17	0.00	8.33	0.00	4.17	4.17	0.00	4.17	4.17	4.17	24
10	0.00	0.00	4.55	0.00	4.55	0.00	18.18	4.55	13.64	9.09	9.09	0.00	0.00	4.55	0.00	4.55	0.00	13.64	0.00	13.64	22
11	0.00	8.33	0.00	8.33	4.17	0.00	0.00	8.33	12.50	20.83	4.17	0.00	8.33	0.00	4.17	4.17	0.00	8.33	4.17	4.17	22
12	0.00	4.17	0.00	8.33	8.33	0.00	0.00	8.33	12.50	20.83	4.17	0.00	8.33	0.00	4.17	4.17	0.00	8.33	4.17	4.17	22
13	0.00	8.33	0.00	8.33	4.17	0.00	0.00	8.33	12.50	25.00	4.17	0.00	4.17	0.00	4.17	4.17	0.00	8.33	4.17	4.17	22
14	0.00	4.17	0.00	8.33	4.17	0.00	0.00	8.33	12.50	20.83	4.17	0.00	8.33	0.00	4.17	4.17	0.00	8.33	4.17	8.33	22
15	0.00	4.17	0.00	8.33	4.17	0.00	0.00	8.33	12.50	20.83	4.17	0.00	8.33	0.00	4.17	4.17	0.00	8.33	4.17	8.33	22
16	0.00	4.35	0.00	8.70	8.70	0.00	0.00	8.70	13.04	21.24	4.35	0.00	8.70	0.00	4.35	4.35	0.00	4.35	8.70	0.00	23

Description:

- | | |
|---------------------------------------|-------------------------------------|
| 1. <i>Vanda pumila</i> | 9. <i>Dendrobium trilamellatum</i> |
| 2. <i>Phalaenopsis pantherine</i> | 10. <i>Phalaenopsis amboinensis</i> |
| 3. <i>Phalaenopsis modesta</i> | 11. <i>Dendrobium spectabile</i> |
| 4. <i>Phalaenopsis gigantea</i> | 12. <i>Cymbidium dayanum</i> |
| 5. <i>Phalaenopsis corningiana</i> | 13. <i>Dendrobium anosmum</i> |
| 6. <i>Phalaenopsis tetrapis</i> | 14. <i>Coelogyne pandurata</i> |
| 7. <i>Paphiopedilum glaucophyllum</i> | 15. <i>Coelogyne asperata</i> |
| 8. <i>Oncidium</i> sp. | 16. <i>Aerides odorata</i> |

The identical species connected by a straight vertical line indicate the similarity of the nucleotide base sequence, which influences the amino acid composition in a sequence. *Vanda pumila* and *Dendrobium trilamellatum*, which are in one vertical line, have the same content of several types of amino acids, namely, cysteine 8.33, aspartic acid 0.00, glutamic acid 8.33, phenylalanine 4.17, glycine 0.00, histidine 0.00, lysine 12.50, methionine 4.17, asparagine 0.00, proline, 8.33, glutamine 0.00, arginine 4.17, serine 4.17, threonine 0.00, and tyrosine 4.17 (see Table 4). Meanwhile, for other types of amino acids, there are differences in composition between the two, where *Vanda pumila* has 0.00 alanine, while *Dendrobium trilamellatum* has 4.17, *Vanda pumila* has 4.17 isoleucine, while *Dendrobium trilamellatum* has 8.33, *Vanda pumila* has 25.00 leucine, while *Dendrobium trilamellatum* has 20.83, *Vanda pumila* has 0.00 tryptophan, while *Dendrobium trilamellatum* has 4.17. Other species that are connected in a vertical line are *Phalaenopsis corningiana* and *Phalaenopsis tetrapis*, where these two species of the same genus have identical and exactly the same amino acid composition. *Phalaenopsis corningiana* and *Phalaenopsis tetrapis* have alanine 4.55, cysteine 4.55, and aspartic acid 0.00, glutamic acid 9.09, phenylalanine 4.55, glycine 4.55, histidine 0.00, isoleucine 13.64, lysine 13.64, leucine 9.09, methionine 9.09, asparagine 0.00, proline, 8.33, glutamine 0.00, arginine 9.09, serine, 4.55, threonine 0.00, valine 0.00, tryptophan 0.00, and tyrosine 13.64.

The translation of amino acids synthesizes proteins, each characterized by a distinct sequence and number of amino acids, crucial for determining the unique three-dimensional structure and biological functions inherent in the morphology of every species (Wahyudiati, 2017). Species in the first clade have similar morphological characters, which are dominated by the sympodial stem type and live at low to high altitudes. Meanwhile, species in the second clade have a monopodial stem type and live at low altitudes.

Furthermore, the sequence alignment result of the 16 samples was used as material for reconstructing a phylogenetic tree to describe kinship relationships, which then produced two large clades with bootstrap values displayed at the base of each branch. The first clade consists of 12 species, including *Vanda pumila*, *Dendrobium trilamellatum* J.J. Sm., *Dendrobium spectabile*

(Blume) Miq., and *Cymbidium dayanum* Rchb.f., *Dendrobium anosmum* Lindl., *Aerides odorata* Lour., *Coelogyne pandurata* J.J. Sm. & Schltr., *Coelogyne asperata* Lindl., *Oncidium* sp., *Paphiopedilum glaucophyllum* J.J. Sm., *Phalaenopsis gigantea* J.J. Sm., *Phalaenopsis amboinensis* J.J. Sm. The second clade consists of four species, including *Phalaenopsis pantherina* Rchb.f., *Phalaenopsis modesta* J.J. Sm., *Phalaenopsis corningiana* Rchb.f., and *Phalaenopsis tetrapis* Rchb.f. (see Figure 2). The 16 samples, which were divided into two clades, showed that species in one clade were similar in the arrangement of nucleotide bases. If the genetic difference value is 1–3%, this indicates that the individuals being compared have similar nucleotide base sequences and are thought to be identical or similar individuals. On the other hand, if the genetic difference value is more than 3%, then the individuals being compared do not have similar nucleotide base sequences and are different individuals or different species (Kulkarni et al., 2025).

The phylogenetic tree view shows the different lengths of branch lines. Different line lengths indicate the evolutionary level of each species (Koshkarov & Tahiri, 2024). A longer line shows a further evolutionary distance, while a shorter line shows a closer evolutionary distance of a species. The longer the line, the further the evolutionary distance, and it can be said that the species has more advanced characters than its ancestral species (Allard et al., 2025). Meanwhile, the shorter the line, indicates that a species has more primitive the characters or the more similar it is to its ancestor species. The difference in line lengths in the phylogenetic tree is clearly described in Figure 3.

The phylogenetic tree in Figure 3 shows that the 12 species in the first clade have differences in line length, where *Phalaenopsis gigantea* and *Phalaenopsis amboinensis* have shorter lines compared to the other 10 species, which indicates that the characters of these two species are more primitive. The *Phalaenopsis* genus in the second clade, which consists of *Phalaenopsis corningiana*, *Phalaenopsis tetrapis*, *Phalaenopsis modesta*, and *Phalaenopsis pantherina*, has the shortest line in the results of the phylogenetic tree reconstruction, which indicates that of the 16 species used in this study, these 4 species have the best characters, primitive and considered closest to ancestral characters.

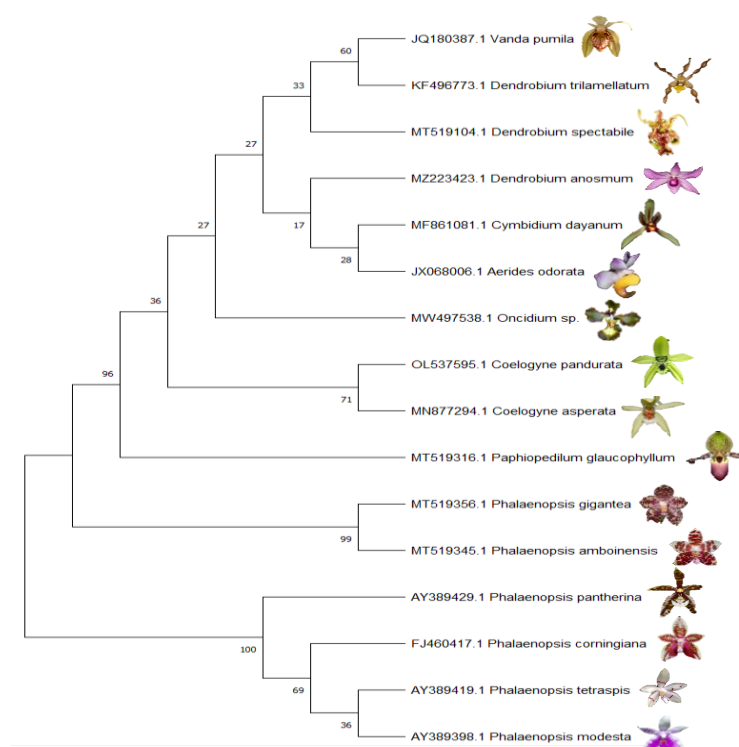


Figure 2. Phylogenetic tree of 16 orchid samples

Genetic Distance

The genetic distance of an organism is calculated based on nucleotide and amino acid sequence data, basically interpreting the amount of nitrogen base substitution in a nucleotide sequence. These substitutions can be synonymous (purine base to purine; pyrimidine to pyrimidine)

or non-synonymous (purine to pyrimidine and vice versa). The result of this substitution is polymorphism in the nucleotide sequence (more than one haplotype). The results of genetic distance analysis using the compute pairwise distance model display a table for analyzing genetic distance between species, as in Table 5.

Table 5. The result of the genetic distance analysis of the samples

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1																
2	1.61															
3	1.61	0.01														
4	1.61	0.00	0.01													
5	0.36	1.13	1.20	1.13												
6	1.61	0.00	0.01	0.00	1.13											
7	0.36	1.13	1.05	1.01	0.00	1.01										
8	0.04	1.50	1.26	1.26	0.32	1.26	0.32									
9	0.06	1.51	1.26	1.27	0.28	1.27	0.34	0.04								
10	0.03	1.61	1.34	1.34	0.30	1.34	0.36	0.03	0.04							
11	0.03	1.61	1.27	1.28	0.30	1.28	0.36	0.02	0.02	0.02						
12	0.04	1.73	1.34	1.34	0.28	1.34	0.34	0.03	0.02	0.03	0.01					
13	0.06	1.65	1.31	1.32	0.31	1.32	0.39	0.04	0.03	0.04	0.02	0.02				
14	0.04	1.61	1.30	1.31	0.28	1.31	0.34	0.01	0.02	0.03	0.01	0.01	0.02			
15	0.04	1.61	1.30	1.31	0.28	1.31	0.34	0.01	0.02	0.03	0.01	0.01	0.02	0.00		
16	0.09	1.35	1.21	1.22	0.44	1.22	0.46	0.12	0.11	0.12	0.09	0.10	0.11	0.10	0.07	

Description:

1. *Vanda pumila*
2. *Phalaenopsis tetraspis*
3. *Phalaenopsis pantherina*
4. *Phalaenopsis modesta*
5. *Phalaenopsis gigantea*
6. *Phalaenopsis corningiana*
7. *Phalaenopsis amboinensis*
8. *Paphiopedilum glaucophyllum*

9. *Oncidium sp.*
10. *Dendrobium trilamellatum*
11. *Dendrobium spectabile*
12. *Dendrobium anosmum*
13. *Cymbidium dayanum*
14. *Coelogyne pandurata*
15. *Coelogyne asperata*
16. *Aerides odorata*

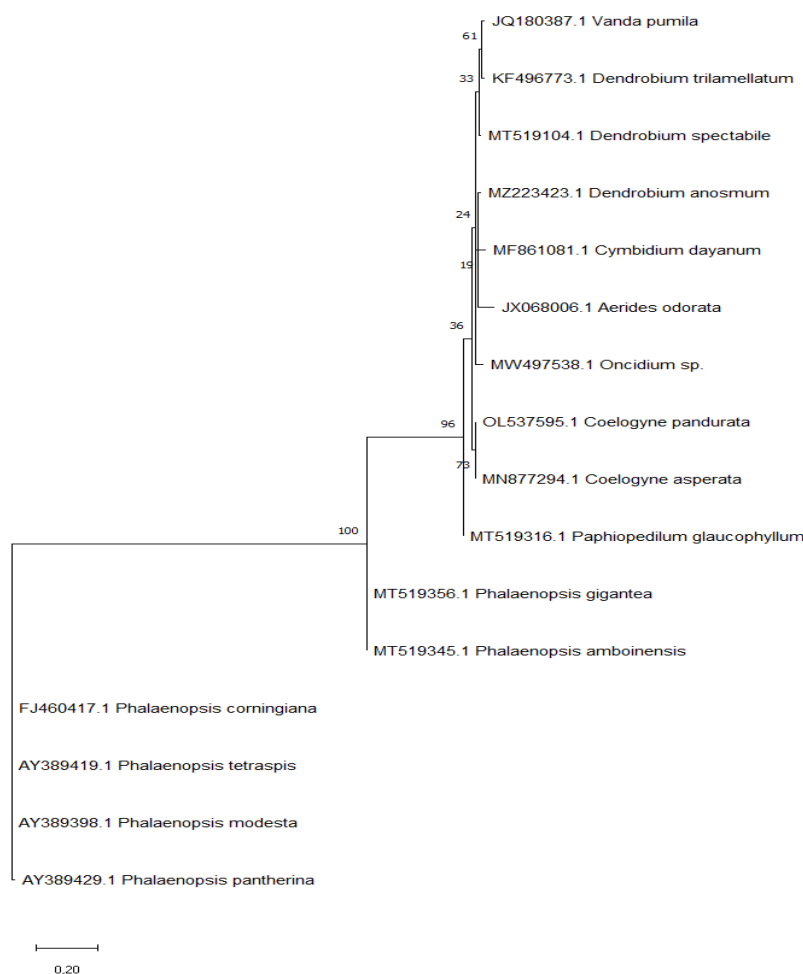


Figure 3. Phylogenetic tree with toggle display node stats

This genetic distance analysis result is used to assess the relationship between different species in interbreeding. This helps determine the proximity of genetic traits and the potential for successful hybridization. In a study, Van Oost et al. used genetic relationships and genome size to predict cross-compatibility among different lavender species, aiding targeted breeding of better cultivars (Oost et al., 2021). Transgressive phenotypes, which exceed the range of traits of the parent species, are more likely to occur in hybrids between more distant species, suggesting a relationship between genetic distance and transgression frequency (Gao et al., 2025). Jordan et al. observed limited gene flow between outcrossing and self-fingering *Geum* species, suggesting that genetic exchange is influenced by genetic distance. A hump-shaped relationship between genetic similarity and offspring performance in *Ranunculus reptans* highlights the importance of genetic distance in mating systems and mate choice.

In the genetic distance results presented in

Table 5, it can be seen that species in one genus have a close genetic distance with a distance of <0.03 . This indicates the similarity of the nucleotide base sequence, which will have an impact on the similarity of amino acid composition, so that phenotypic gene expression also tends to be the same. Long genetic distance or >0.03 is also directly proportional to the differences between species and genera, such as the *Oncidium* genus having a terrestrial way of life, while the *Phalaenopsis* genus has an epiphytic way of life. Meanwhile, in terms of morphology, the *Oncidium* genus has wavy petals and sepals with a bright yellow or brownish color, accompanied by a fragrant fragrance, and has elongated and thin leaves, while the *Phalaenopsis* genus has star-shaped flowers like *Phalaenopsis pantherina*, while other species have rounded-type flowers on the sepals and dorsal. The genetic distance used as a reference to distinguish between species is 3% (Iiyama et al., 2024), if the results of the genetic distance calculation exceed 3%, or 0.03, then the species is declared different from

other species, and can be stated to have a long genetic distance.

The novelty of this research lies in the DNA-based identification framework for endemic Indonesian Orchidaceae, specifically by combining sequence alignment and genetic distance. Through this effort, the genetic relationships of the 16 species studied were mapped. This mapping resulted in two large clades and a very close species pair (genetic distance <0.003). Both can be identified with greater precision.

This research provides tangible benefits in strengthening terrestrial biodiversity inventory and monitoring activities and improving the accuracy of the scientific basis for setting conservation priorities and habitat management. This research supports SDG 15, especially targets 15.1 and 15.5. Molecular-level identification is a prerequisite for conserving and sustainably managing terrestrial ecosystems and their biodiversity (Target 15.1) and for taking urgent action to halt biodiversity loss and prevent species extinction by reducing misidentification and improving monitoring of endemic lineages (Target 15.5) (Hariyanto et al., 2019; Novinovrita & Kencanawati, 2025; Rahmawati et al., 2026; Siregar et al., 2025; Umami et al., 2025).

CONCLUSION

The kinship relationships among the 16 species form two major clades, with each clade featuring pairs of species linked by straight vertical lines, denoting a genetic distance of <0.003 or falling within the close category. Within the first clade, *Vanda pumila* is linked with *Dendrobium trilamellatum*, *Cymbidium dayanum* with *Aerides odorata*, and *Coelogyne pandurata* with *Coelogyne asperata*. Conversely, the second clade contains only one pair of species connected by a straight vertical line: *Phalaenopsis corningiana* and *Phalaenopsis tetrapis*. Notably, species within the first clade exhibit varying line lengths, with *Phalaenopsis gigantea* and *Phalaenopsis amboinensis* displaying shorter lines compared to the remaining 10 species, indicative of their more primitive characteristics. In contrast, the *Phalaenopsis* genus within the second clade, comprising *Phalaenopsis corningiana*, *Phalaenopsis tetrapis*, *Phalaenopsis modesta*, and *Phalaenopsis pantherina*, demonstrates the shortest line lengths in the phylogenetic tree reconstruction results, suggesting that among the 16 species examined, these four possess the most

primitive traits and are considered closest to ancestral characteristics.

In line with this research, future researchers can orient their analysis by increasing the number of samples per species and using diverse location backgrounds (islands/altitudes). This effort can test the stability of the two-clade pattern and ensure the closeness of species pairs (genetic distance <0.003). Researchers can also use multi-locus or genomic approaches (e.g., SNP/ddRAD). Furthermore, these steps can be integrated with morphological-ecological data to provide more certainty and robustness in species boundaries for "close category" pairs.

ACKNOWLEDGEMENT

We would like to express our gratitude to the Universitas Muhammadiyah Malang for funding this research through the FTTE-UMM Block Grant scheme for the 2023–2024 period (No. E.6.1/1188a/BAA-AIK/UMM/X/2023).

AUTHOR CONTRIBUTION STATEMENT

TIP conceptualization, methodology, data interpretation, supervision, funding acquisition, writing—original draft, writing—review & editing, and final interpretation of results. NK data curation, investigation, and software/computational implementation support. RL formal analysis, methodology, validation, and writing—review & editing. ES formal analysis, methodology, validation, and writing—review & editing. HH formal analysis, methodology, data interpretation, visualization, validation, writing—original draft, writing—review & editing, and final interpretation of results. DF formal analysis, methodology, validation, and writing—review & editing. All authors read and approved the final manuscript and agreed to be accountable for all aspects of the work.

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict of interest regarding the publication of this paper.

USE OF ARTIFICIAL INTELLIGENCE (AI)-ASSISTED TECHNOLOGY

The authors declare that no artificial intelligence (AI) tools were used in the generation, analysis, or writing of this manuscript. All aspects of the research, including data collection,

interpretation, and manuscript preparation, were carried out entirely by the authors without the assistance of AI-based technologies.

REFERENCES

- Allard, J. B., Sharma, S., Patel, R., Sanderford, M., Tamura, K., Vucetic, S., Gerhard, G. S., & Kumar, S. (2025). Evolutionary sparse learning reveals the shared genetic basis of convergent traits. *Nature Communications*, 16(1), 3217. <https://doi.org/10.1038/s41467-025-58428-8>
- Chi, M.-W., Liu, D.-K., Zhou, C.-Y., Li, M.-H., & Lan, S.-R. (2021). The complete plastid genome of *Thrixspermum centipeda* (Orchidaceae, Aeridinae). *Mitochondrial DNA. Part B, Resources*, 6(3), 1245–1246. <https://doi.org/10.1080/23802359.2021.1906171>
- Fang, T., Liao, S., Chen, X., Zhao, Y., Zhu, Q., Cao, Y., Wang, Q., Zhang, S., Gao, Z., Yang, Y., Wang, Y., & Zhang, J. (2019). Forensic drowning site inference employing mixed pyrosequencing profile of DNA barcode gene (*rbcL*). *International Journal of Legal Medicine*, 133(5), 1351–1360. <https://doi.org/10.1007/S00414-019-02075-4>
- Gao, Z., Xiao, Y., Chen, J., & Shi, T. (2025). eQTL Analysis Elucidates the Genetic Basis Underlying Transgressive Gene Expression in *Nelumbo* Hybrids. *Molecular Ecology*, 34(24), e70203. <https://doi.org/10.1111/mec.70203>
- Han, C., Dong, F., Qi, Y., Wang, Y., Zhu, J., Li, B., Zhang, L., Lv, X., & Wang, J. (2025). The Breeding, Cultivation, and Potential Applications of Ornamental Orchids with a Focus on *Phalaenopsis*—A Brief Review. *Plants*, 14(11), 1689. <https://doi.org/10.3390/plants14111689>
- Hariyanto, S., Fahmi, A. K., Soedarti, T., & Suwarni, E. E. (2019). Vegetation and Community Structure of Mangrove in Bama Resort Baluran National Park Situbondo East Java. *Biosaintifika*, 11(1), 132–138.
- Iiyama, C. M., Vilcherrez-Atoche, J. A., Germanà, M. A., Vendrame, W. A., & Cardoso, J. C. (2024). Breeding of ornamental orchids with focus on *Phalaenopsis*: current approaches, tools, and challenges for this century. *Heredity*, 132(4), 163–178. <https://doi.org/10.1038/s41437-024-00671-8>
- Koshkarov, A., & Tahiri, N. (2024). Novel Algorithm for comparing phylogenetic trees with different but overlapping taxa. *Symmetry*, 16(7), 790. <https://doi.org/10.3390/sym16070790>
- Kulkarni, P., Porter, L., Chou, T.-F., Chong, S., Chiti, F., Schafer, J. W., Mohanty, A., Ramisetty, S., Onuchic, J. N., Tuite, M., Uversky, V. N., Weninger, K. R., Koonin, E. V., Orban, J., & Salgia, R. (2025). Evolving concepts of the protein universe. *IScience*, 28(3), 112012. <https://doi.org/10.1016/j.isci.2025.112012>
- Liu, H., Fang, C., Zhang, T., Guo, L., & Ye, Q. (2019). Molecular authentication and differentiation of *Dendrobium* species by rDNA ITS region sequence analysis. *AMB Express*, 9(1), 53. <https://doi.org/10.1186/s13568-019-0767-8>
- Liu, P., Jiang, X., Guo, H., Jia, T., Wu, S., Ou, F., Tian, W., Liu, L., Chang, Y., Ding, J., & Zhang, W. (2025). The Influence of Parent Pairs with Different Genetic Distances on the Genetic Diversity of Offspring in *Strongylocentrotus intermedius*. *Biology*, 14(7). <https://doi.org/10.3390/biology14070745>
- Luo, Y., Luo, W., Zhao, T., Yang, J., Yuan, L., Zhang, P., Gong, Z., Li, H., Sima, Y., & Xu, T. (2025). The Complete Chloroplast Genomes of Three *Manglietia* Species and Phylogenetic Insight into the Genus *Manglietia* Blume. *Current Issues in Molecular Biology*, 47(9), 737. <https://doi.org/10.3390/cimb47090737>
- Maggioni, D., Arrigoni, R., Seveso, D., Galli, P., Berumen, M. L., Denis, V., Hoeksema, B. W., Huang, D., Manca, F., Pica, D., Puce, S., Reimer, J. D., & Montano, S. (2022). Evolution and biogeography of the *Zanclaea-Scleractinia* symbiosis. *Coral Reefs*, 41(3), 779–795. <https://doi.org/10.1007/s00338-020-02010-9>
- Masiero, E., Banik, D., Abson, J., Greene, P., Slater, A., & Sgamma, T. (2020). Molecular verification of the uk national collection of cultivated *liriope* and *ophiopogon* plants. *Plants*, 9(5), 1–12. <https://doi.org/10.3390/plants9050558>
- Mavrodiev, E. V. (2020). Synapomorphies Behind Shared Derived Characters: Examples from the Great Apes' Genomic Data. *Acta Biotheoretica*, 68(3), 357–365. <https://doi.org/10.1007/s10441-019-09368-6>
- Mirioba, J., Emitaro, W., Obwanga, B., Gaya, H., Leley, N., Otuoma, J., M. Maina, J., & Kawaka, F. (2024). Orchid species diversity across a forest disturbance gradient in west Mau forest, Kenya. *PLOS One*, 19(8), e0307887. <https://doi.org/10.1371/journal.pone.0307887>
- Novinovrita, M., & Kencanawati, I. (2025). Ethnobotanical study of *zingiberaceae* to reinforce plant taxonomy concepts through field

- practice. *Al Jahiz: Journal of Biology Education Research*, 6(1), 115-132.
- Nuryady, M. M., Widayanti, R., Nurcahyo, R. W., Fadrinatha, B., & Fahrurrozi Z S, A. (2019). Characterization and phylogenetic analysis of multidrug-resistant protein - encoding genes in *Trypanosoma evansi* isolated from buffaloes in Ngawi district, Indonesia. *Veterinary World*, 12(10), 1573–1577. <https://doi.org/10.14202/vetworld.2019.1573-1577>
- Oelschlaeger, P. (2024). Molecular Mechanisms and the Significance of Synonymous Mutations. *Biomolecules*, 14(1). <https://doi.org/10.3390/biom14010132>
- Oost, E. Van, Leus, L., Rybel, B. De, & Laere, K. Van. (2021). Determination of Genetic Distance, Genome Size and Chromosome Numbers to Support Breeding in Ornamental *Lavandula* Species. *Agronomy*, 11(11), 2173. <https://doi.org/10.3390/AGRONOMY11112173>
- Post, N. W., Gilbert, C. C., Pugh, K. D., & Mongle, C. S. (2023). Implications of outgroup selection in the phylogenetic inference of hominoids and fossil hominins. *Journal of Human Evolution*, 184, 103437. <https://doi.org/10.1016/j.jhevol.2023.103437>
- Prihartin, I., Ismail, A. S., Sukorini, H., Nursandi, F., Zakia, A., & Farahdina, F. A. R. (2023). Identification of bacterial isolates of Tumpang and Bumiasri (East Java, Indonesia) using 16S rRNA gene sequencing and screening of their active compounds as a biofertilizer. *Biodiversitas*, 24(6), 3338–3343. <https://doi.org/10.13057/biodiv/d240629>
- Ragauskas, A., Maziliauskaitė, E., Prakas, P., & Butkauskas, D. (2025). Population Genetic Structure: Where, What, and Why? *Diversity*, 17(8), 584. <https://doi.org/10.3390/d17080584>
- Rahmawati, D., Aurora, W. I. D., Styarsih, L., & Sulistiawan, A. (2026). Local food-based mediterranean-style nutrition education improves dietary adherence, sleep quality, and psychological wellbeing among third-trimester pregnant women. (2026). *Al Jahiz: Journal of Biology Education Research*, 7(1), 122-139.
- Siregarar, N. F., Amprasto, A., & Sriyati, S. (2025). Biodiversity literacy and tradition: Development of student worksheets based on the local potential of the indahan tupporobu dish for biology learning. *Al Jahiz: Journal of Biology Education Research*, 6(2), 242-258.
- Sun, J., Liu, B., Rustiami, H., Xiao, H., Shen, X., & Ma, K. (2024). Mapping Asia Plants: Plant Diversity and a Checklist of Vascular Plants in Indonesia. *Plants*, 13(16), 2281. <https://doi.org/10.3390/plants13162281>
- Takeda, S., Nishikawa, Y., Tachibana, T., Higaki, T., Sakamoto, T., & Kimura, S. (2025). Morphological and Transcriptome Analysis of the Near-Threatened Orchid *Habenaria radiata* with Petals Shaped Like a Flying White Bird. *Plants*, 14(3), 393. <https://doi.org/10.3390/plants14030393>
- Umami, M., Narakusumo, R. P., Encilia, E., Nada, S., Amrulloh, I., Agung, A., & Sekarwangi, D. (2025). Insect-Assisted Biodegradation of Waste Plastics for Enhancing Sustainable Development Goals. *Biosaintifika*, 17(1), 68–79.
- Vilcherrez-Atoche, J. A., Iiyama, C. M., & Cardoso, J. C. (2022). Polyploidization in Orchids: From Cellular Changes to Breeding Applications. *Plants (Basel, Switzerland)*, 11(4). <https://doi.org/10.3390/plants11040469>
- Wang, J., Mao, X., Liu, B., Wang, Z., Ma, Y., Li, Q., & Liu, J. (2025). Recurrent Hybridisations During Diversification of One *Rhododendron* Species Complex. *Molecular Ecology*, 34(21), e70046. <https://doi.org/10.1111/mec.70046>
- Wang, Z., Sun, Y., Liao, T., Zhong, H., Gu, Q., & Luo, K. (2025). Molecular Phylogenetics of Seven Cyprinidae Distant Hybrid Lineages: Genetic Variation, 2nNCRC Convergent Evolution, and Germplasm Implications. *Biology*, 14(11), 1527. <https://doi.org/10.3390/biology14111527>
- Zander, R. H. (2025). Geometric Models of Speciation in Minimally Monophyletic Genera Using High-Resolution Phylogenetics. *Plants*, 14(4). <https://doi.org/10.3390/plants14040530>
- Zhang, W., Zhang, Z., Liu, B., Chen, J., Zhao, Y., & Huang, Y. (2023). Comparative analysis of 17 complete chloroplast genomes reveals intraspecific variation and relationships among *Pseudostellaria heterophylla* (Miq.) Pax populations. *Frontiers in Plant Science*, 14, 1163325. <https://doi.org/10.3389/fpls.2023.1163325>
- Zhang, Z., Gale, S. W., Li, J.-H., Fischer, G. A., Ren, M.-X., & Song, X.-Q. (2019). Pollen-mediated gene flow ensures connectivity among spatially discrete sub-populations of *Phalaenopsis pulcherrima*, a tropical food-deceptive orchid. *BMC Plant Biology*, 19(1), 597. <https://doi.org/10.1186/s12870-019-2179-y>