Journal of Tropical Biodiversity and Biotechnology
Volume 08, Issue 03 (2023): jtbb73130
DOI: 10.22146/jtbb.73130

Research Article

In Silico Approach for DNA Barcoding using Phylogenetic Analysis of *Coelogyne* spp. based on the *mat*K, *rpo*C1, *rbc*L and nrDNA Markers

Apriliana Pratiwi[1,2], Anggiresti Kinasih[1,2], Maura Indria Meidianing[1,2], Febri Yuda Kurniawan[1,3], Endang Semiarti[2]*
1)Biology Orchid Study Club (BiOSC), Faculty of Biology, Universitas Gadjah Mada, 55281 Yogyakarta, Indonesia
2)Department of Tropical Biology, Faculty of Biology, Universitas Gadjah Mada, 55281 Yogyakarta, Indonesia
3)Study Program of Biotechnology, Graduate School, Universitas Gadjah Mada, 55281 Yogyakarta, Indonesia
*Corresponding author, email: endsemi@ugm.ac.id

Keywords:
Coelogyne
conservation
DNA barcoding
in silico
phylogenetic analysis
Submitted:
18 February 2022
Accepted:
19 May 2023
Published:
02 October 2023
Editor:
Furzani Binti Pa'ee

ABSTRACT
In silico biology is considered as an effective and applicable approach to initiate various research, such as biodiversity taxonomical conservation. Phylogenetic analysis using *in silico* taxonomy method for orchid species can provide data on genetic diversity and evolutionary relationships. One particular method that can be used to evaluate specific targets of gene loci in the taxonomic study is DNA barcoding. This research was conducted to determine the specific target locus gene using *mat*K, *rbc*L, *rpo*C1, and nrDNA markers for DNA barcoding of the *Coelogyne* genus with *in silico* approach using phylogenetic analysis. All marker sequences were collected from the NCBI website and analysed using several softwares and methods, namely Clustal X for sample sequence alignment and MEGA 11 for phylogenetic tree construction and analysis. The results showed that the gene locus in *Coelogyne* recommended was the nrDNA gene locus. Phylogenetic analysis revealed that the use of the nrDNA gene locus was able to separate 17 *Coelogyne* species with two outgroup species, namely *Cymbidium* and *Vanilla*, then followed with ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (*rbc*L) while the other gene loci, namely maturase K (*mat*K) and polymerase beta' subunit (*rpo*C1) provided a visual phylogenetic tree in which the two outgroup species entered into the same clade as the *Coelogyne* species. Thus, the results of this study can be used as a reference to support the *Coelogyne* breeding and conservation program.

INTRODUCTION
Orchids (Orchidaceae) are herbaceous plants that have the potential to be used as research objects because of their large diversity. Orchid plant taxonomy is important for the plant classification, and various researchers can easily describe and identify variations and relationships between one species and another. *Coelogyne* is a genus of orchids with about 200 species spread throughout Asia (Chase et al. 2015), including India, China, Philippines, Pacific Indonesia, and Fiji Islands (Singh & Kumaria 2020). In Indonesia, most *Coelogyne* species are found in dense tropical forests, especially in Kalimantan, Sumatra, and Sulawesi, and several species have not been identified. One of Indonesia's endemic species is *Coelogyne pandurata*, or black orchid found in East Kalimantan (Hartati &

Muliawati 2020) The IUCN's ecological conservation status of *Coelogyne* is least concerned in CITES (Convention on International Trade in Endangered Species of Fauna and Flora), *Coelogyne pandurata is* classified as Appendices II. Identification of *Coelogyne* species in Indonesia based on morphological characterisation was already done in previous research using phenetic taxonomy. The study found that *C. pandurata* from East Kalimantan has high morphological similarities with *C. rumphii* from South Sulawesi, and *C. mayeriana* from Kalimantan has high similarity with *C. asperata* from West Kalimantan (Hartati et al. 2019). However, the combination of homologous and non-homologous data becomes biased and the evolutionary relationship was not given as information in phenetic taxonomy.

Advances in the field of science strongly support the development of a classification system, namely phylogenetic taxonomy (Haider 2018) Phylogenetic taxonomic studies produce classification data for species collection and provide information about the relationship between species (Rivero 2016) because the data used include DNA sequence data (Nauheimer et al. 2018; Zhang et al. 2021). The phylogenetic analysis system can also contribute to the conservation of orchid plants such as *Coelogyne*. The taxonomic status of *Coelogyne* can be clarified through phylogenetic analysis so that conservation priorities are known (Li et al. 2018). Unique evolutionary lineages of *Coelogyne* can also be identified to compare rare and widespread species. The nucleotide sequence of a standard genome region is needed as a tool for species identification to produce a phylogenetic tree with a higher level of discrimination. This process can be achieved by determining DNA barcodes. DNA barcoding method is currently known as one aspect of genetic conservation managements (Kim et al. 2014) . DNA barcoding is now widely practiced and often used because it can support various studies, such as complementing information in plant classification and increasing authentication and identification of medicinal plants (Mishra et al. 2016; Parveen et al. 2016).

Research on DNA barcoding in *Coelogyne* has been carried out previously in India. Still, this study only used the *rbc*L gene and *Coelogyne* species in India, so there is no information about *C. pandurata* (Ramudu & Khasim 2016). The previous research also conducted phylogenetic analysis but it only focused on the relationship between *C. fimbriata* and *C. ovalis* using a combination of chloroplast fragments and *mat*K. Still, *mat*K phylogenetic tree has low bootstrap results (Jiang et al. 2020). A comparison of five different loci, such as *rbc*L, *rpo*B, *rpo*C1, *mat*K, and *ITS* has been done and shown that *ITS* is the most efficacy barcode that can be used for discriminate 47 genera of Indian Orchid then, followed by *mat*K (Parveen et al. 2017). Based on genetic distance, phylogenetic tree, and similarity of 94 genera of five subfamilies of medicinal Orchidaceae from Asia, a single barcode region such as *ITS* then *mat*K has higher species discrimination capability than the combination of two barcode region (Raskoti & Ale 2021). It also has similarities with the jewel orchids in Vietnam, where multi-locus barcodes cannot improve resolution for species classification (Ho et al. 2021). The results of DNA barcoding in the genera level, infrageneric rank, and species level may have a different result, such as in *Euphrasia* where *ITS* barcoding cannot provide a clear resolution of species-level separation (Wang et al. 2018). In some arid plant, *rbc*L (88%) have a higher success amplification rate than *mat*K as a barcode (Bafeel et al. 2011). Previous study showed that *mat*K have high resolution in discriminate orchid in intergeneric level of family Orchidaceae, but not a good barcode to resolve phylogeny at intrageneric of *Dendrobium* (Chattopadhyay et al. 2017).

In this study, a phylogenetic analysis is conducted on intrageneric level in *Coelogyne* genera using the *in silico* method, through DNA barcode of *mat*K, *rbc*L and *rpo*C. In this research, *ITS* region was not included because the data was not available in the Gene Bank. Instead, nuclear ribosomal DNA (nrDNA) that consists of *ITS*1-5,85S-*ITS*2 region was used. The use of DNA barcodes is based on various aspects, such as the type of living thing itself. The Plant Working Group Consortium for the Barcode of Life (CBOL) recommends the use of *rbc*L + *mat*K as the core barcode in plant phylogenetic analysis (CBOL Plant Working Group 2009). In addition to the use of these two types of barcodes, it is highly recommended to carry out additional analysis using other types of barcodes (Hollingsworth et al. 2011). In various studies in the genetics of orchids and plants, *rpo*C1 is one of the chloroplast genes often used (Parveen et al. 2016; Kim et al. 2020) which belongs to the *Coelogyne* genus (Singh & Kumaria 2020). In this research, *rpo*C1 was present as an additional barcode to enhance the carried-out analysis and to compare the exact resolution between three loci that had better results. This research aim was to determine the specific target locus gene using *mat*K, *rbc*L, and *rpo*C1 markers for DNA barcoding of each species in *Coelogyne* genus with *in silico* approach using phylogenetic analysis.

MATERIALS AND METHODS

Materials

The research was carried out by accessing the DNA sequences of the *Coelogyne* genus from the *mat*K, *rbc*L, *rpo*C1 and nrDNA gene loci as the sample. Currently, research has found out that *Coelogyne* species with the *mat*K gene locus have DNA sequence lengths up to 500-900 base pairs, *rpo*C1 gene locus has up to 400-550 DNA base pairs, *rbc*L gene locus has ±600 DNA base pairs, and nrDNA has 600-800 base pair. All DNA sequences belonging to Asian *Coelogyne* species were accessed via the nucleotide database at The European Bioinformatics Institute (EBI) (https://www.ebi.ac.uk/) and National Center for Biotechnology Information (NCBI) GenBank (www.ncbi.nlm.nih.org/nucleotide) (Vu et al. 2018). Based on NCBI, there were 21 sequences and 19 sequences collected from EBI which had *mat*K, *rbc*L, *rpo*C1 and nrDNA gene loci. Gene loci with unverified names were not used and removed from the data. Each sequence was selected into a separate gene locus and was selected based on the specifications of the nucleotide base length of each locus. The final 19 loci were used for analysis in this study (Table 1).

Methods

The sequences of each species that have been found were then collected in FASTA format. The alignment of the collected sequences was performed using the CLUSTAL X software program, and the .aln output format was predefined. The .aln output of the CLUSTAL X software program was processed with MEGA 11 and then saved in MEGA format for phylogenetic tree analysis (Hall 2013). The best model for constructing a phylogenetic tree for each data set is determined in advance in the program. The phylogenetic tree construction method used in this study is Neighbor-Joining with 1000 bootstrap based on the best model tamura-3-parameter (Ho & Nguyen 2020).

Data analysis was performed from the most representative gene locus for phylogenetic analysis using MEGA 11 and DNAsp. Genetic variation was observed to learn about the polymorphic data that represent the mutation of each nucleotide. The genetic distance matrix was conducted to determine the divergence between each species in the *Coelo-*

Table 1. The *mat*K, *rpo*C1, *rbc*L and nrDNA sequences of 19 species of *Coelogyne* and 2 outgroup species of *Cymbiduum* and *Vanilla* used in the study.

Orchid Species	Accession Number			
	matK	***rbcL***	***rpoC1***	**nrDNA**
Coelogyne asperata	KU877844	KU877824	KX037361	AF281128
Coelogyne mayeriana	MN400412	MN400420	-	AF281129
Coelogyne pandurata	KU877841	MN416671	KU219954	AF281130
Coelogyne trinervis	KF974497	JN005393	KP662087	AF302744
Coelogyne rochusseni	MK398201	MN416673	KP662089	MK356175
Coelogyne cumingii	MK398200	MN400414	KP662086	MK356172
Coelogyne verrucosa	AY003884	-	-	AF281131
Coelogyne ovalis	MN416677	MN416668	MT067929	KY966509
Coelogyne fimbriata	KR905392	KU219968	KP662093	EU441205
Coelogyne nitida	JN004370	MK155298	KP662088	HQ130496
Coelogyne schilleriana	KU877839	KU219970	KU219952	-
Coelogyne pachystachya	KU877838	KU219969	KU219951	-
Coelogyne barbata	KX298581	KU219960	KP662085	AF302755
Coelogyne velutina	KU877840	MN416675	KU219953	AF302753
Coelogyne xyrekes	MK398225	KU219966	KP662092	MK356198
Coelogyne pulverula	KU877846	KU877826	KX037358	MK356157
Coelogyne viscosa	KX298597	KU219955	KP662080	MK356152
Coelogyne fuscescens	KF974501	KU219959	KP662084	KF866234
Coelogyne eberhardtii	MN400408	MN400416	KP662091	AF302754
Cymbidium aloifolium	KX298600	JN005425	HM053600	JF729014
Vanilla planifolia	MF349972	JN005701	JN005354	AF030049

gyne genus. Haplotype was determined to learn about the location of inherited allele groups from a single parent using DNA sp, and a haplotype map was constructed using Network. GC content and nucleotide diversity were determined to ensure the data is valid for *Coelogyne* genus.

RESULTS AND DISCUSSION

This research was conducted using the NCBI website to search for *Coelogyne* DNA barcodes consisting of three 4 gene loci, namely *mat*K, *rpo*C1, *rbc*L, and nrDNA. Based on result given in the Table 2, *mat*K locus have the lowest G+C content and nrDNA has the highest G+C content. Moderate amount of G+C content about 29.48% for *mat*K, 44.12% for *rpo*C1, 42.18% for *rbc*L and 57.27% for nrDNA . Based on (Wu et al. 2020) the overall G+C content of *Coelogyne* is 43.3%, so the most representative data for G+C content was *rpo*C1 and *rbc*L. The result is similar with previous study that stated *rbc*L and *rpo*C1 has good quality of sequence compared to other locus (Parveen et al. 2016; Hosein et al. 2017; El-Sherif & Ibrahim 2020).

However, based on examination using DNAsp (Table 3), *rbc*L and *rpo*C1 has lowest parsimony, also *rbc*L shown lowest polymorphic, number of mutation, nucleotide diversity, and gaps compared to another locus. The *rbc*L gene encodes the formation of rubisco enzyme to fixation of carbon dioxide in light independent photosynthetic reactions, therefore, *rbc*L is important to maintain, so mutations are rare (Rajaram et al. 2019).

High number of mutation and haplotype diversity is shown in nrDNA that consists of internal transcriber spacer that located between small-subunit of ribosomal DNA. The pattern of high haplotype diversity (Hd:1.00) corresponds to relatively low nucleotide diversity (Pi:0.096)

Table 2. GC content of *Coelogyne* species based *mat*K, *rpo*C1, *rbc*L, nrDNA gene locus.

Species	Guanin and Cytosin (G + C) content (%)			
	*mat*K	*rpoC1*	*rbc*L	nrDNA
Coelogyne pulverula	28.04	44.27	42.19	55.64
Coelogyne pandurata	27.10	44.27	41.69	55.90
Coelogyne rochusseni	32.50	44.27	42.02	56.47
Coelogyne asperata	27.10	45.29	41.86	56.75
Coelogyne xyrekes	32.14	44.27	42.02	56.81
Coelogyne viscosa	28.97	44.27	42.19	56.99
Coelogyne velutina	28.50	44.52	42.19	57.08
Coelogyne nitida	28.50	44.27	42.69	57.30
Coelogyne fuscescens	28.97	44.27	42.52	57.39
Coelogyne fimbriata	27.10	44.27	42.19	57.41
Coelogyne ovalis	32.14	41.33	42.35	57.50
Coelogyne eberhardtii	31.79	44.27	42.35	57.69
Coelogyne cumingii	32.14	44.27	42.35	57.75
Coelogyne trinervis	28.97	44.02	42.19	57.96
Coelogyne mayeriana	32.50	-	41.86	58.02
Coelogyne verrucosa	27.10	-	-	58.06
Coelogyne barbata	27.57	44.02	42.19	58.81
Cymbidium aloifolium	32.24	42.96	42.02	68.27
Vanilla planifolia	30.48	42.74	43.02	53.87

indicating that the *Coelogyne* population has experienced rapid growth and expansion over time (Yun et al. 2020). The evolutionary processes can occur in an organism because of genetic mutations and or recombinant processes which then form new species (Dharmayanti 2011). The results in this study were a phylogenetic tree of the species collected and constructed using the MEGA 11 program. The formation of the clade represents the genetic relationship between the *Coelogyne* species. In the phylogenetic tree, *Cymbidium* and *Vanilla* used as a correction factor, because both were very distantly and far related from other *Coelogyne* species as can be seen in the values.

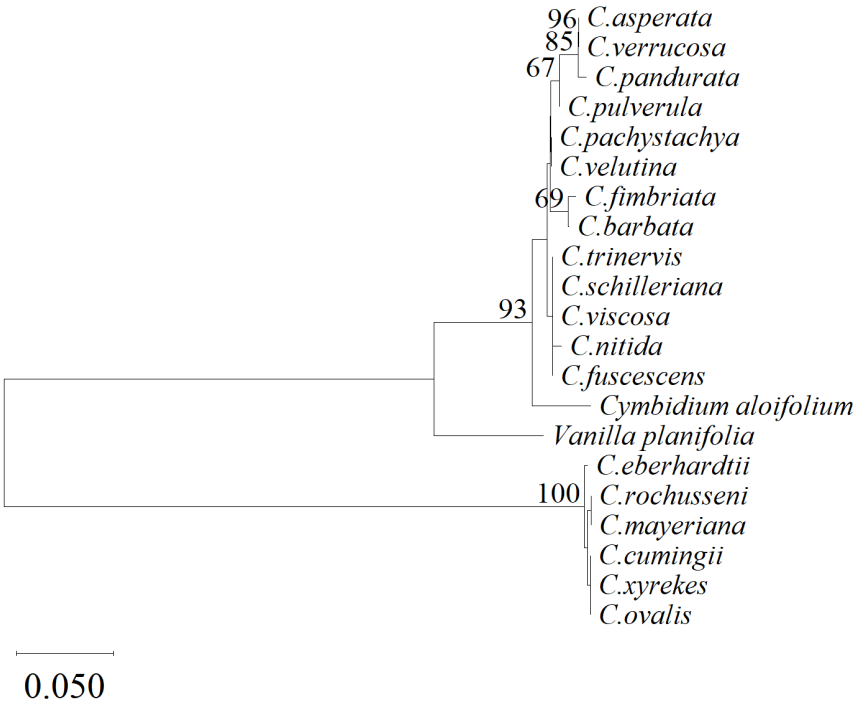

Figure 1. *Coelogyne* phylogenetic tree using the maturase K barcode (*mat*K).

Table 3. Examination using DNAsp.

Data	*mat*K	*rpoC1*	*rbc*L	nrDNA
Number of sites	290	1040	602	363
Gaps/missing data	86	651	0	70
Parsimony informative site	85	7	7	52
Variable polymorphic site	95	161	43	135
Total number of mutation	102	165	44	175
Nucleotide diversity (Pi)	0.18490	0.04627	0.0097	0.09602
Haplotype diversity (Hd)	0.914	0.784	0.9316	1.000
Number of haplotype (h)	12	10	13	19

Utilisation of the *mat*K gene locus in the phylogenetic tree showed that *Coelogyne* was divided into 2 clades, where *Cymbidium* and *Vanilla planifolia* join the first clade of *Coelogyne* species (Figure 1). From the first clade, *C. pandurata*, *C. asperata* and *C. verrucosa* have a close evolutionary relationship. The second clade consists of *C. eberhardtii*, *C. rochusseni*, *C. cumingii*, *C. mayeriana*, *C. ovalis* and *C. xyrekes* (Figure 1). This indicated that *mat*K gene locus showed inconsistent results due to the presence of *Cymbidium* and *Vanilla* in the same clade as several *Coelogyne* genera and *C. mayeriana* that located far away from another Indonesian orchid. This result is similar with previous research, barcode of *mat*K showed the putative incongruence, therefore, it is non-functional for phylogenetic analysis of Orchidaceae . Many closely related species of Indian orchid cannot be discriminated by *mat*K based on genetic distance, blast, and tree building method (Srivastava & Manjunath 2020).

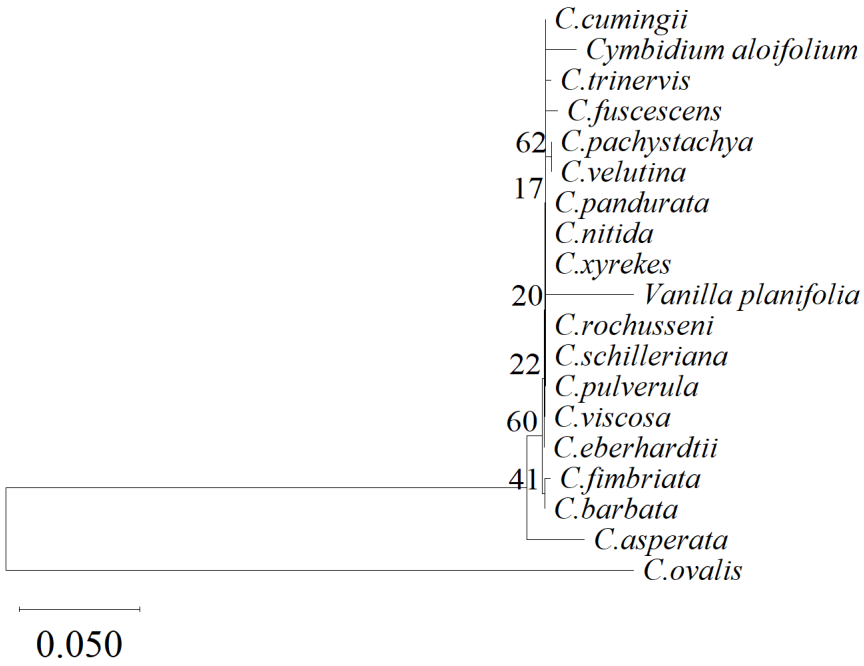

Figure 2. *Coelogyne* phylogenetic tree using RNA barcode of polymerase beta' subunit (*rpo*C1).

The result of phylogenetic tree construction using the *rpo*C1 gene locus (Figure 2) cannot discriminate species in different genus and has low bootstrap value. *Cymbidium* and *Vanilla* should not be in the same clade as *Coelogyne* because it has a different genus. Species *Coelogyne* from Indonesia such as *C. pandurata*, *C. mayeriana* and *C. asperata* also located far apart. This indicated that the use of the *rpo*C1 and *mat*K gene locus was not effective as a basis or material for phylogenetic analysis in *Coelogyne*. The *rpo*C1 locus is not recommended to be used in DNA barcoding

because it exhibits low polymorphism data compared to *mat*K locus (Hosein et al. 2017). Locus gene of *rpo*C1 showed low power in distinguishing genetic variability between some species compared to *rbc*L (El-Sherif & Ibrahim 2020). Species discrimination rates of *rpo*C1 is the lowest compared to *rbc*L, *mat*K, and *ITS* based on genetic distance, phylogenetic tree, and blast method after calculated using Kimura-2-parameter model (Parveen et al. 2017).

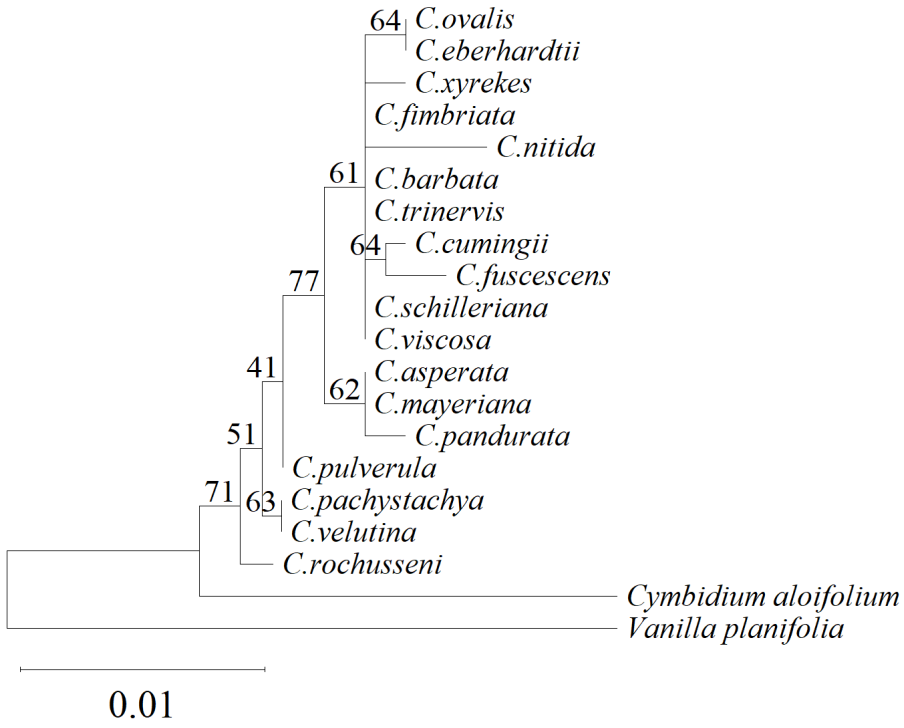

Figure 3. *Coelogyne* phylogenetic tree using the barcode of ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (*rbc*L).

The use of the *rbc*L gene locus in the phylogenetic tree (Figure 3) gave representative results as the basis for *Coelogyne* phylogenetic analysis because *Cymbidium* was in a separate clade with another 19 *Coelogyne* species. This result was also supported by (Ho et al. 2021), *rbc*L shown the best result as a DNA barcoding marker than *mat*K for distinguishing some jewel orchid species. Gene locus of *rbc*L is the best for phylogenetic analysis compared to another plastid chloroplast regions. Discriminatory power of *rbc*L is higher than *mat*K because it has good sequence quality, recoverability, and universality (Maloukh et al. 2017). The *rbc*L locus is a plastid gene in the chloroplast genome that encodes rubisco, considered sufficient and suitable for discrimination of Orchidaceae at the generic and species level (Ramudu & Khasim 2016). The visualisation of the phylogenetic tree shows that *Vanilla* lies on a different evolutionary path from other species. This showed that there is a close relationship between *Coelogyne* and *Cymbidium*, but *Coelogyne* has a distant evolutionary relationship with *Vanilla*.

The phylogenetic tree (Figure 3) also showed that all *Coelogyne* from Indonesia such as *C. mayeriana* from Kalimantan, *C. asperata* from West Kalimantan have close evolutionary relationship then followed with *C. pandurata* from East Kalimantan that classified in the same subclade. This result is in agreement with phenetic taxonomy that has been done previosly in morphology comparison of each species from Indonesia (Hartati et al. 2019; Hartati & Muliawati 2020). However, the tree are not representative for the separation of orchids originating from Asia, such as *C. fimbriata* and *C. ovalis* which should have close relationship

(Jiang et al. 2020) but located far apart in the phylogenetic tree constructed with *rbc*L. The ability of *rbc*L to discriminate each genus within the same family was higher than *mat*K, but both w ere less effective at differentiating species within the same genus. The discrimination power of *rbc*L in the intrageneric group is low compared to other locus (Chattopadhyay et al. 2017). Although having high quality sequence, *rbc*L species resolution was inferior, when using BLASTn and Neighboor -joining method tree, *rbc*L was unable to discriminate species in the same *Phaphiopedilum* genus (Rajaram et al. 2019). Discrimination rate and resolution of *rbc*L used to distinguish each *Coelogyne* species from India was 36.36% based distance method, 72.72% based on phylogenetic tree Kimura-2-parameter, 44.44% based on cluster and all of that considered as low (Ramudu & Khasim 2016).

In contrast, phylogenetic tree constructed with nrDNA locus (Figure 4) have high separation and discrimination power, this is related to *Cymbidium* and *Vanilla* that are already located as outgroup. Some of species already resolved into some different subclades and the evolutionary relationship can be distinguished. High bootstrap values are also shown in the branches which represent the close evolutionary relationship between *C. fimbriata* and *C. ovalis*. Orchid species from Indonesia, such as *C. asperata*, *C. pandurata*, *C. mayeriana*, also showed a close evolutionary relationship. *C. asperata* was closer to *C. verrucosa* compared to other species, it was relevant with previous research (Jiang et al. 2020). The phylogenetic tree constructed by nrDNA demonstrated high resolution and promising discrimination between intrageneric group of *Dendrobium* groups compared to *mat*K and *rbc*L (Chattopadhyay et al. 2017). The reason nrDNA shows the best phylogenetic results is because 5,8S comprises conserved sequence and *ITS* generally carries variation among closely related genera. The phylogenetic tree constructed by by maximum-likelihood method using *ITS* region showed high resolution and discrimination compared to *mat*K and *rbc*L on 7 genera Indian endemic orchid (Srivastava & Manjunath 2020). Higher mutation rate obviously represent in *ITS* and *ITS2*, so it has great potential in systematic study and species identification (Duan et al. 2019).

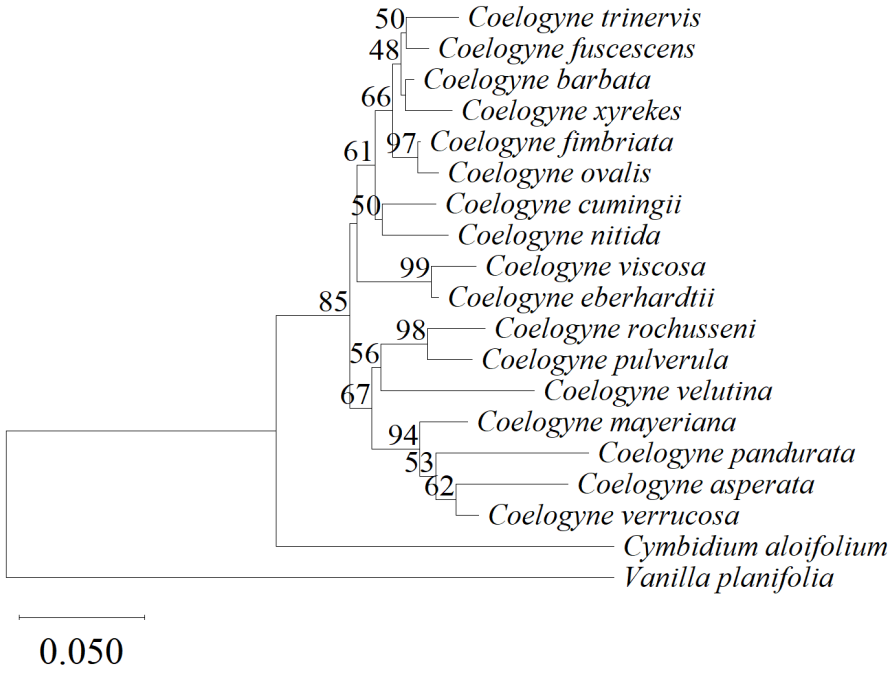

Figure 4. *Coelogyne* phylogenetic tree using the barcode of nrDNA (*ITS1*+5,8S+*ITS2*).

Total genetic distance shown in Table 4, for *rbc*L after generate using Tamura-3-parameter model shown the low value at 0.002-0.012. Some of species like *C.mayeriana*, *C.asperata*, *C.viscosa*, *C.schilleriana*, *C.trinervis*, *C.barbata*, *C.fimbriata* have the same pattern of genetic distance at 0.007. This result is in accordance with previous research that generate average interspecific distance 0.007 using Kimura-2-parameter model in some of *Coelogyne* species (Ramudu & Khasim 2016). This proves that *rbc*L is the most conserved one because it displays the lowest genetic sequence divergence value (Raskoti & Ale 2021).

Total genetic distance of nrDNA that consist of *ITS* region has highest value compared to *rbc*L. This result shown in Table 5, related to previous research that stated intraspecific genetic distance and interspecific variation in *ITS* ranged as highest followed by *ITS2* then the lowest is *rbc*L (Raskoti & Ale 2021). The genetic distance pattern of *C. mayeriana*, *C. pandurata*, and *C. asperata* which are considered as endemic orchid from Indonesia was distinct from other species due to differences in the origin of geographical population growth and evolutionary sources based on previous research (Yun et al. 2020).

There are 13 different haplotypes in total were observed in *rbc*L (Figure 5a). Based on *rbc*L haplotype distribution (Table 6), *C.viscosa*, *C.barbata*, *C.schilleriana*,*C.fimbriata*, and *C.trinervis* shared the same haplotype at H4. Species *C.pachystachya* and *C.velutina* also share the same haplotype at H9. *C.ovalis* and *C.eberhardtii* share same haplotype at H1. The frequency of H4 is 5 meanwhile frequency of H1, H2, and H7 is 2. Based on *rbc*L haplotype map (Figure 5a), H4 individu is the source of evolutionary formation of H1, H9, H6, H11 with low mutation line, and H7 with more mutation line. This is in accordance to each species such as *C. fimbriata* with H4 that has close evolutionary relationship with *C. ovalis* with H1. Indonesian orchid species, *C. asperata*, and *C.mayeriana* share the same haplotype in H2 and evolutionarily develop into H3 which consist of *C. pandurata.* Meanwhile, the form haplotype map that builds using nrDNA (Figure 5b), *C. pandurata* has H2 that develop from *C. mayeriana* with H16, then both evolved from *C. verrucosa* with H19, then all of them evolved from *C. asperata* with H17. Meanwhile *C. fimbriata* (H1) and *C. ovalis* (H18) join in the same median vectors, *Cymbidium* (H14) shows many mutation lines that developed from *C. fimbriata.*

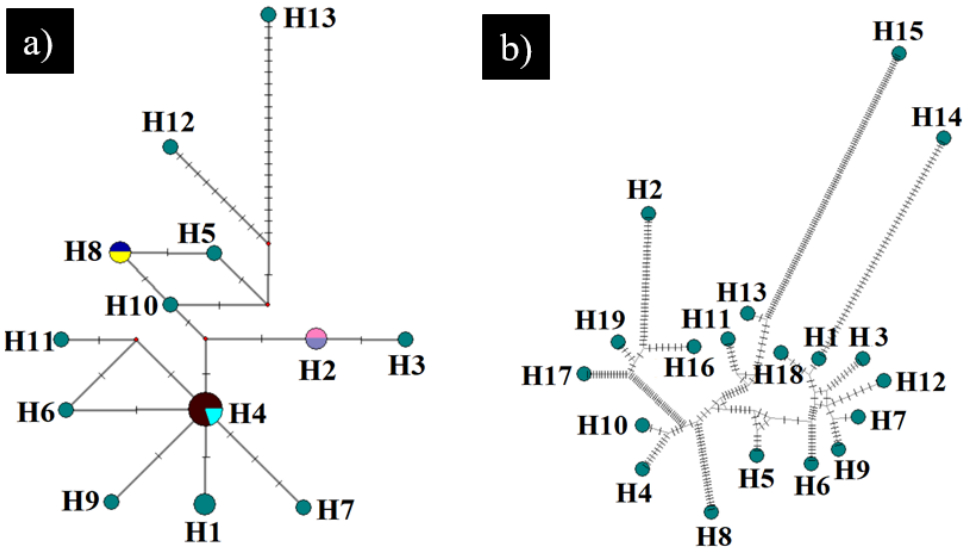

Figure 5. Haplotype map of a) *rbc*L loci gene and b) nrDNA loci gene of *Coelogyne*

Table 4. Total genetic distance between *Coelogyne* species using *rbc*L

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
C.rochusseni	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C.velutina	0.002	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C.pachystachya	0.002	0.000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C.pulverula	0.003	0.002	0.002	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C.mayeriana	0.007	0.005	0.005	0.003	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C.asperata	0.007	0.005	0.005	0.003	0.000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C.viscosa	0.007	0.005	0.005	0.003	0.003	0.003	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C.schilleriana	0.007	0.005	0.005	0.003	0.003	0.003	0.000	-	-	-	-	-	-	-	-	-	-	-	-	-
C.trinervis	0.007	0.005	0.005	0.003	0.003	0.003	0.000	0.000	-	-	-	-	-	-	-	-	-	-	-	-
C.barbata	0.007	0.005	0.005	0.003	0.003	0.003	0.000	0.000	0.000	-	-	-	-	-	-	-	-	-	-	-
C.fimbriata	0.007	0.005	0.005	0.003	0.003	0.003	0.000	0.000	0.000	0.000	-	-	-	-	-	-	-	-	-	-
C.pandurata	0.008	0.007	0.007	0.005	0.002	0.002	0.005	0.005	0.005	0.005	0.005	-	-	-	-	-	-	-	-	-
C.eberhardtii	0.008	0.007	0.007	0.005	0.005	0.005	0.002	0.002	0.002	0.002	0.002	0.007	-	-	-	-	-	-	-	-
C.ovalis	0.008	0.007	0.007	0.005	0.005	0.005	0.002	0.002	0.002	0.002	0.002	0.007	0.000	-	-	-	-	-	-	-
C.cumingii	0.008	0.007	0.007	0.005	0.005	0.005	0.002	0.002	0.002	0.002	0.002	0.007	0.003	0.003	-	-	-	-	-	-
C.xyrekes	0.008	0.007	0.007	0.005	0.005	0.005	0.002	0.002	0.002	0.002	0.002	0.007	0.003	0.003	0.003	-	-	-	-	-
C.fuscescens	0.010	0.008	0.008	0.007	0.007	0.007	0.003	0.003	0.003	0.003	0.003	0.008	0.005	0.005	0.003	0.005	-	-	-	-
C.nitida	0.012	0.010	0.010	0.008	0.008	0.008	0.005	0.005	0.005	0.005	0.005	0.010	0.007	0.007	0.007	0.007	0.008	-	-	-
Cymbidium	0.020	0.022	0.022	0.020	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.025	0.025	0.025	0.025	0.025	0.027	0.029	-	-
Vanilla	0.036	0.038	0.038	0.036	0.039	0.039	0.039	0.039	0.039	0.039	0.039	0.041	0.041	0.041	0.041	0.041	0.043	0.045	0.050	-

Table 5. Total genetic distance between *Coelogyne* species using nrDNA

Species	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
C.viscosa	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C.eberhardtii	0.02	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C.fimbriata	0.07	0.06	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C.ovalis	0.07	0.06	0.01	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C.cumingii	0.07	0.06	0.05	0.05	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C.nitida	0.08	0.06	0.05	0.05	0.04	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C.barbata	0.08	0.06	0.02	0.03	0.04	0.04	-	-	-	-	-	-	-	-	-	-	-	-	-
C.trinervis	0.08	0.07	0.04	0.05	0.06	0.06	0.03	-	-	-	-	-	-	-	-	-	-	-	-
C.fuscescens	0.08	0.07	0.02	0.03	0.05	0.05	0.02	0.03	-	-	-	-	-	-	-	-	-	-	-
C.xyrekes	0.09	0.08	0.04	0.04	0.05	0.05	0.02	0.04	0.03	-	-	-	-	-	-	-	-	-	-
C.pulverula	0.10	0.09	0.08	0.08	0.07	0.08	0.08	0.09	0.09	0.09	-	-	-	-	-	-	-	-	-
C.mayeriana	0.10	0.08	0.08	0.08	0.08	0.09	0.07	0.08	0.08	0.08	0.08	-	-	-	-	-	-	-	-
C.rochusseni	0.10	0.09	0.08	0.08	0.08	0.09	0.09	0.10	0.09	0.10	0.04	0.09	-	-	-	-	-	-	-
C.verrucosa	0.11	0.10	0.08	0.08	0.09	0.09	0.08	0.09	0.07	0.09	0.08	0.04	0.09	-	-	-	-	-	-
C.velutina	0.13	0.10	0.10	0.11	0.11	0.11	0.10	0.11	0.10	0.11	0.10	0.10	0.11	0.12	-	-	-	-	-
C.pandurata	0.14	0.13	0.13	0.14	0.13	0.13	0.12	0.13	0.13	0.13	0.11	0.09	0.13	0.09	0.15	-	-	-	-
C.asperata	0.15	0.14	0.12	0.12	0.12	0.12	0.11	0.13	0.13	0.12	0.10	0.07	0.12	0.06	0.16	0.10	-	-	-
Cymbidium	0.21	0.19	0.19	0.20	0.21	0.22	0.19	0.20	0.19	0.20	0.23	0.20	0.20	0.18	0.25	0.26	0.27	-	-
Vanilla	0.43	0.42	0.40	0.41	0.41	0.43	0.40	0.40	0.41	0.41	0.44	0.43	0.45	0.44	0.46	0.46	0.47	0.49	-

Table 6. Haplotype distribution of *Coelogyne* and outgroup species using *rbc*L and nrDNA loci gene

Hap	Distribution of *Coelogyne* species based on loci gene	
	***rbc*L**	**nrDNA**
H1	*C.ovalis, C.eberhardtii*	*C fimbriata*
H2	*C.asperata, C.mayeriana*	*C.pandurata*
H3	*C.pandurata*	*C.trinervis*
H4	*C.trinervis, C.fimbriata, C.schilleriana C.barbata, C.viscosa*	*C.rochusseni*
H5	*C.rochusseni*	*C.cumingii*
H6	*C.rochusseni*	*C.nitida*
H7	*C.nitida*	*C.barbata*
H8	*C.pachystachya, C.velutina*	*C.velutina*
H9	*C.xyrekes*	*C.xyrekes*
H10	*C.pulverula*	*C.pulverula*
H11	*C.fuscescens*	*C.viscosa*
H12	*Cymbidium aloifolium*	*C.fuscescens*
H13	*Vanilla planifolia*	*C.eberhardtii*
H14	ND	*Cymbidium aloifolium*
H15	ND	*Vanilla planifolia*
H16	ND	*C.mayeriana*
H17	ND	*C.asperata*
H18	ND	*C.ovalis*
H19	ND	*C.verrucosa*

Abbreviation : Hap = haplotype, H = number of haplotype, ND = not detected

CONCLUSIONS

The research shows that the use of the nrDNA which consist of *ITS* gene region is more recommended in phylogenetic analysis among species in the *Coelogyne* genus then followed by *rbc*L. Based on phylogenetic tree and haplotype map constructed with nrDNA, all *Coelogyne* species from Indonesia have close evolutionary relationship, *C. pandurata* (H2) evolved from *C. mayeriana* (H16), then both evolved from *C. verrucosa* (H19), then all of them evolved from *C. asperata* (H17). In addition, it is necessary to improve the DNA barcode data from the *Coelogyne* genus thus phylogenetic analysis carried out can represent a more significant number of species so that the research results obtained will be more accurate.

AUTHORS CONTRIBUTION

AP as first author did phylogenetic analysis, full paper writing, reviewing and editing, AK did nucleotide data analysis using DNAsp, haplotype map construction, full paper writing, reviewing and editing, MIM did full paper writing and editing, FYK reviewing and editing, ES reviewing and editing.

ACKNOWLEDGMENTS

The authors would like to thank the Biology Orchid Study Club (BiOSC) members who have helped in the collection and data collection of orchid samples.

CONFLICT OF INTEREST

There is no any conflict of interest regarding the research or the funding.

REFERENCES

Bafeel, S.O. et al., 2011. Comparative evaluation of PCR success with universal primers of maturase K (matK) and ribulose-1, 5-bisphosphate carboxylase oxygenase large subunit (rbcL) for barcoding of some arid plants. *Plant Omics*, 4, pp.195-198.

CBOL Plant Working Group, 2009. A DNA barcode for land plants. In *Proceedings of the National Academy of Sciences*, 106. pp. 12794–12797.

Chase, M.W. et al., 2015. An updated classification of Orchidaceae. *Botanical Journal of the Linnean Society*, 177(2), pp.151–174. doi: 10.1111/boj.12234.

Chattopadhyay, P., Banerjee, G. & Banerjee, N., 2017. Distinguishing Orchid Species by DNA Barcoding: Increasing the Resolution of Population Studies in Plant Biology. *OMICS: A Journal of Integrative Biology*, 21(12), pp.711–720. doi: 10.1089/omi.2017.0131.

Duan, H. et al., 2019. The screening and identification of DNA barcode sequences for Rehmannia. *Scientific Reports*, 9, 17295. doi: 10.1038/s41598-019-53752-8.

El-Sherif, N. & Ibrahim, M., 2020. Implications of rbcL and rpoC1 DNA Barcoding in Phylogenetic Relationships of some Egyptian Medicago sativa L. Cultivars. *Egyptian Journal of Botany*, 60(12), pp.451–460. doi: 10.21608/ejbo.2020.20028.1399.

Haider, N., 2018. A Brief Review on Plant Taxonomy and its Components. *Jour Pl Sci Res*, 34(2), pp.275–290.

Hall, B.G., 2013. Building Phylogenetic Trees from Molecular Data with MEGA. *Molecular Biology and Evolution*, 30(5), pp.1229–1235. doi: 10.1093/molbev/mst012.

Hartati, S. et al., 2019. Morphological characterization of Coelogyne spp for germplasm conservation of orchids. *Revista Ceres*, 66(4), pp.265–270. doi: 10.1590/0034-737x201966040004.

Hartati, S. & Muliawati, E.S., 2020. Short Communication: Genetic variation of Coelogyne pandurate, C. rumphii and their hybrids based on RAPD markers. *Biodiversitas Journal of Biological Diversity*, 21(10), pp.4709-4713. doi: 10.13057/biodiv/d211033.

Ho, V.T. et al., 2021. Comparison of matK and rbcL DNA barcodes for genetic classification of jewel orchid accessions in Vietnam. *Journal of Genetic Engineering and Biotechnology*, 19(1), 93. doi: 10.1186/s43141-021-00188-1.

Ho, V.T. & Nguyen, M.P., 2020. An in silico approach for evaluation of rbcL and matK loci for DNA barcoding of Cucurbitaceae family. *Biodiversitas Journal of Biological Diversity*, 21(8), pp.3879-3885. doi: 10.13057/biodiv/d210858.

Hollingsworth, P.M., Graham, S.W. & Little, D.P., 2011. Choosing and Using a Plant DNA Barcode. *PLoS ONE*, 6(5), e19254. doi: 10.1371/journal.pone.0019254.

Hosein, F.N. et al., 2017. Utility of DNA barcoding to identify rare endemic vascular plant species in Trinidad. Ecology and Evolution. *Ecology and Evolution*, 7(18), pp.7311–7333. doi: 10.1002/ece3.3220.

Dharmayanti, N.L.P.I., 2011. Molecular phylogenetic: organism taxonomy method based on evolution history. *Wartazoa*, 21(1), pp.1–5.

Jiang, K. et al., 2020. Chloroplast Genome Analysis of Two Medicinal Coelogyne spp. (Orchidaceae) Shed Light on the Genetic Information, Comparative Genomics, and Species Identification. *Plants*, 9 (10), 1332. doi: 10.3390/plants9101332.

Kim, H.M. et al., 2014. DNA barcoding of Orchidaceae in Korea. *Molecular Ecology Resources*, 14(3), pp.499–507. doi: 10.1111/1755-0998.12207.

Kim, Y.-K. et al., 2020. Plastome Evolution and Phylogeny of Orchidaceae, With 24 New Sequences. *Frontiers in Plant Science*, 11, 22. doi: 10.3389/fpls.2020.00022.

Li, J. et al., 2018. Prioritizing the orchids of a biodiversity hotspot for conservation based on phylogenetic history and extinction risk. *Botanical Journal of the Linnean Society*, 186(4), pp.473–497. doi: 10.1093/botlinnean/box084.

Maloukh, L. et al., 2017. Discriminatory power of rbcL barcode locus for authentication of some of United Arab Emirates (UAE) native plants. *3 Biotech*, 7(2), 144. doi: 10.1007/s13205-017-0746-1.

Mishra, P. et al., 2016. DNA barcoding: an efficient tool to overcome authentication challenges in the herbal market. *Plant Biotechnology Journal*, 14(1), pp.8–21. doi: 10.1111/pbi.12419.

Nauheimer, L. et al., 2018. Australasian orchid biogeography at continental scale: Molecular phylogenetic insights from the Sun Orchids (Thelymitra, Orchidaceae). *Molecular Phylogenetics and Evolution*, 127, pp.304–319. doi: 10.1016/j.ympev.2018.05.031.

Parveen, I. et al., 2016. DNA Barcoding for the Identification of Botanicals in Herbal Medicine and Dietary Supplements: Strengths and Limitations. *Planta Medica*, 82(14), pp.1225–1235. doi: 10.1055/s-0042-111208.

Parveen, I. et al., 2017. Evaluating five different loci (rbcL, rpoB, rpoC1, matK, and ITS) for DNA barcoding of Indian orchids. *Genome*, 60 (8), pp.665–671. doi: 10.1139/gen-2016-0215.

Rajaram, M.C. et al., 2019. DNA Barcoding of Endangered Paphiopedilum species (Orchidaceae) of Peninsular Malaysia. *Phytotaxa*, 387 (2), pp.94–104. doi: 10.11646/phytotaxa.387.2.2.

Ramudu, J. & Khasim, S.M., 2016. DNA Barcoding of some Indian Coelogyne (Epidendroideae, Orchidaceae). *J. Orchid Soc. India*, 30, pp.65–73.

Raskoti, B.B. & Ale, R., 2021. DNA barcoding of medicinal orchids in Asia. *Scientific Reports*, 11(1), 23651. doi: 10.1038/s41598-021-03025-0.

Rivero, D.G., 2016. Darwinian Archaeology and Cultural Phylogenetics. *In Cultural Phylogenetics Springer*, pp.43–72. doi: 10.1007/978-3-319-25928-4_3.

Singh, N. & Kumaria, S., 2020. A Combinational Phytomolecular-Mediated Assessment in Micropropagated Plantlets of Coelogyne ovalis Lindl.: A Horticultural and Medicinal Orchid. *Proceedings of the National Academy of Sciences, India Section B: Biological Sciences*, 90(2), pp.455–466. doi: 10.1007/s40011-019-01118-5.

Srivastava, D. & Manjunath, K., 2020. DNA barcoding of endemic and endangered orchids of India: A molecular method of species identification. *Pharmacognosy Magazine*, 16(70), pp.290-299. doi: 10.4103/pm.pm_574_19.

Vu, H.-T. et al., 2018. In Silico Study on Molecular Sequences for Identification of *Paphiopedilum* Species. *Evolutionary Bioinformatics*, 14, 1176934318774542. doi: 10.1177/1176934318774542.

Wang, X. et al., 2018. DNA barcoding a taxonomically complex hemiparasitic genus reveals deep divergence between ploidy levels but lack of species-level resolution. *AoB PLANTS*, 10(3). doi: 10.1093/aobpla/ply026.

Wu, Q.-P. et al., 2020. Characterization of the complete chloroplast genome of *Coelogyne fimbriata* (Orchidaceae). *Mitochondrial DNA Part B*, 5(3), pp.3507–3509. doi: 10.1080/23802359.2020.1827058.

Yun, S.A. et al., 2020. Genetic diversity and population structure of the endangered orchid Pelatantheria scolopendrifolia (Orchidaceae) in Korea. *PLOS ONE*, 15(8), e0237546. doi: 10.1371/journal.pone.0237546.

Zhang, G.-Q. et al., 2021. Phylogenetic incongruence in Cymbidium orchids. *Plant Diversity*, 43(6), pp.452–461. doi: 10.1016/j.pld.2021.08.002.