ENHANCED KNOWLEDGE IN SCIENCES AND TECHNOLOGY
e-ISSN: 2773-6385

EKST

Vol. 5 No. 2 (2025) 526-537
https://publisher.uthm.edu.my/periodicals/index.php/ekst

DNA Barcoding of The Penang Hill Slipper Orchid (*Paphiopedilum Barbatum*)

Khairun Ayuni Afiqah Hassan[1], Rexy Prakash Chacko[1], Yap Jing Wei[1*]

[1] *Department of Technology and Natural Resources, Faculty of Applied Sciences and Technology, UTHM Kampus Cawangan Pagoh, Hab Pendidikan Tinggi Pagoh, KM 1, Jalan Panchor, 84600 Pagoh, Muar, Johor, MALAYSIA.*

*Corresponding Author: jwyap@uthm.edu.my
DOI: https://doi.org/10.30880/ekst.2025.05.02.060

Article Info

Received: 1 January 2025
Accepted: 18 January 2025
Available online: 19 December 2025

Keywords

DNA Barcoding, *Paphiopedilum*, *Paphiopedilum Barbatum*, ITS2, matK

Abstract

Paphiopedilum is a genus of the slipper orchid from the subfamily Cypripedioideae. Penang Hill, Malaysia is home to a distinct variety of Paphiopedilum barbatum (Paphiopedilum Penang Hill) which has been almost extirpated from its native range. Although there is interest in breeding and reintroducing Paphiopedilum Penang Hill from individuals in Penang Botanic Garden, it is vital to ensure the origin of these plants. In this study, DNA barcoding was used to confirm the taxonomic status of Paphiopedilum Penang Hill. Representative samples of these species were obtained by fresh field collections. PCR amplification was attempted using primers for two DNA barcoding regions: maturase K (matK) and the internal transcribed spacer II (ITS2) nuclear region. Both of the regions were successfully amplified except for one sample from Penang Botanic Garden. Maximum likelihood and maximum parsimony trees were constructed on those samples and other outgroup sequences from GenBank. The phylogenetic analysis revealed that while both regions were successfully amplified, the bootstrap values for the phylogenetic trees were low, indicating weak statistical support for genetic distinction between the two populations. The ITS2 region results further suggested a close genetic relationship between the Penang Hill and Botanic Garden samples.

1. Introduction

The family Orchidaceae, one of the largest and most diverse plant families [1], includes the genus Paphiopedilum, commonly known as slipper orchids. *Paphiopedilum barbatum*, an endangered species, especially notable for its striking appearance and ecological importance. Orchids are among the most threatened taxa globally due to increasing anthropogenic threats and specific conservation needs [2]. The species has faced severe population decline due to habitat destruction caused by urban development, deforestation, and environmental changes. Additionally, illegal collection for ornamental trade has significantly reduced its numbers in the wild. The risk of genetic contamination from cultivated individuals in botanic gardens further threatens the genetic integrity of the remaining wild populations, complicating conservation efforts. Conservation efforts are further complicated by challenges in species identification due to overlapping morphological traits.qw2

DNA barcoding, a molecular tool that uses genetic sequences for species identification of an unknown species [3], has proven effective in addressing such taxonomic ambiguities. Studies have demonstrated its utility in addressing such taxonomic ambiguities. Studies have demonstrated its utility in distinguishing closely related orchid species and understanding their genetic diversity. For instance, DNA barcoding has been successfully applied to various *Paphiopedilum* species [4], facilitating their conservation by providing reliable genetic data.

Despite its ecological significance, *P. barbatum* in Penang Hill has been nearly eradicated. A small number of individuals remain in the collection from Penang Botanic Garden, and there is growing interest in reintroducing the species to its original habitat. However, this raises a critical issue which is ensuring that plants used in reintroduction programs are genetically authentic and representative of the Penang Hill population. The potential risk of genetic contamination from non-native populations could compromise the ecological integrity and genetic purity of *P. barbatum*. This underscores the importance of ensuring accurate species identification and conserving populations in their native. The purpose of this study is to collect and barcode wild P.*barbatum* from Penang Hill. Compare the DNA barcodes of *P. barbatum* from Penang Hill with those from collections of Penang Botanic Garden.

2. Methodology

2.1 Study Site and Sample Collection

This study was carried out at Penang Hill situated on Penang Island, Malaysia, 5. 415° N latitude 100. 276° E longitude. Due to the vulnerability of the species and its habitat, the geographical coordinates of sampling sites are excluded in this study.

Paphiopedilum samples were collected from three locations on Penang Hill and from the orchid collection at Penang Botanic Gardens. Leaves from *P. barbatum* individuals were stored in airtight containers in the field before being transported back to the of Universiti Tun Hussein Onn Malaysia (UTHM).

2.2 DNA Extraction

DNA extraction was carried out with the aid of the PrimeWay Plant DNA Extraction Kit (CTAB Method). Although both fresh and dry samples could be used according to the manufacturer's protocol, fresh leaf samples of *P. barbatum* were used in this process. Firstly, 50 µl Elution buffer was preheated beforehand to 65 °C. Then, samples with a weight of less than or equal to 100 mg were homogenized and transferred into 1.5 ml tubes before 400 µl of CTAB buffer was added and vortexed to mix thoroughly. 10 µl of RNase A solution was added to the solution before it got mixed and incubated at 65 °C for about 10 minutes. The sample was centrifuged at 11,000 x g (9,919 rpm) for 5 minutes. PrimeWay Plant Filter (violet) was placed into a new Collection tube and then the supernatant was transferred into the filter and centrifuged at 11,000 x g (9,919 rpm) for 2 minutes. The flow-through was transferred into a new 1.5 ml microcentrifuge tube, and the PrimeWay Plant Filter (violet) was discarded. Then, PPC Buffer was added into the tube and mixed thoroughly by vortexed. A new PrimeWay Plant Column (Green) was placed into a new collection tube, and 700 µl of lysate was transferred into the Column. The tube was centrifuged at 11,000 x g (9,919 rpm) for I minute and the flow through was discarded. The column was placed back into the Collection tube. The steps were followed by added 400 µl Wash Buffer P1 into the column, followed by 700 µl and 200 µl of wash Buffer P2, respectively. For each wash, the tube was centrifuged at 11,000 x g, and the flow-through was discarded. The PrimeWay Plant Column was placed into a new 1.5 ml microcentrifuge tube, and 50 µl pre-heated Elution Buffer was added. Lastly, the tube was centrifuged at 11,000 x g for 1 minute to elute DNA.

For purity and concentration of DNA samples, each DNA samples were assessed by nucleic acid spectrophotometry using Implen NanoPhotometer on the samples to determine the purity and concentration of DNA. The spectrophotometer was 'blanked' by using distilled water. 1 µl of distilled water was placed onto pedestal. The concentration and purity were recorded for each sample.

2.3 PCR

Two chosen barcode regions were used in this DNA barcoding: ITS2 (Internal transcribed spacer 2) and matK (Maturase K) [5].

2.3.1 PCR Amplification of ITS2

The PCR amplification of the ITS2 region was performed according to the manufacturer's instructions. The reagents were thawed on ice and 25 micro liter reaction mixtures were made in 0.2 micro litter PCR tubes. The master mix consisted of 5.0 µl 5× Green GoTaq Flexi Buffer, 2.0 µl MgCl2 (25 mM), 2.0 µl each of forward and reverse primers (Table 1), 0.5 µl dNTPs (10 mM), 2.0 µl DNA template, 0.126 µl DNA polymerase, and 11.374 µl nuclease-free water. The PCR amplification cycle for the ITS2 region is shown in Table 2. This consists of initial denaturation at 94˚C for about 5 minutes, 40 amplification cycles at 94˚C for 30 seconds, annealing at 56˚C at 30 seconds and extension at 72˚C at 45 seconds. Lastly, the final extension at 72˚C for 10 minutes.

Table 1 Primer used for amplification and sequencing of ITS2

Primers	5' sequence 3'

ITS2-S2F	ATG CGA TAC TTG GTG TGA AT
ITS2-S3R	GAC GCT TCT CCA GAC TAC AAT

Table 2 PCR profile for ITS2

Parameter	Temperature
Initial Denaturation	94˚C, 5 min
Denaturation	94˚C, 30 sec
Annealing	56˚C, 30 sec
Extension	72˚C, 45 sec
Final Extension	72˚C, 10 min

2.3.2 PCR Amplification of matK

Following the instruction of the PCR kit manufacturer, PCR amplification of the matK region was performed. Reagents were thawed on ice and a 25 µl reaction was set up in 0.2 mL PCR tubes also on ice. The master mix included 5.0 µl 5× Green GoTaq Flexi Buffer, 2.0 µl MgCl2 (25 mM), 2.0 µl each of forward and reverse primers (Table3), 0.5 µl dNTPs (10 mM), 2.0 µl DNA template, 0.126 µl DNA polymerase and 11.374 µl nuclease-free water. This consists of initial denaturation at 94˚C for about 1 minute, 40 amplification cycles at 94˚C for 1 minute annealing at 48˚C at 30 seconds, and extension at 72˚C at 1 minute. Lastly, the final extension at 72˚C for 7 minutes.

Table 3 Primer used for amplification and sequencing of matK

Primers	5' sequence 3'
MatK390F	CGA TCT ATT CAT TCA ATA TTT
MatK1326R	TCT AGC ACA CGA AAG TCG AAG T

Table 4 PCR profile for matK

Parameter	Temperature
Initial Denaturation	94˚C, 1 min
Denaturation	94˚C, 1 min
Annealing	48˚C, 30 sec
Extension	72˚C, 1 min
Final Extension	72˚C, 7 min

2.4 Sequencing and Data Analysis

For purification sequencing, all of the PCR samples were delivered to Apical Scientific Sdn. Bhd. The unedited sequence was displayed, and then the sequence was edited, assembled, aligned, and analysed using Geneious v9.8. However, one of the samples, B.Garden 3 ITS2 displayed significant levels of noise and unclear which resulting the sample could not be incorporated into the alignment. Then, the Nucleotide Basic Local Alignment Search Tool (Nucleotide BLAST) web-based program from GenBank, National Centre for Biotechnology Information (NCBI) was used to check the sequence and compare it with the other sequence in the GenBank database.

Multiple forward reaction sequences for both region: *ITS2* and *matK* were aligned using MEGA X v10.2.6. Moreover, two type of tree models were used which are Maximum Likelihood, and Maximum Parsimony model model, and constructed using MEGA X v10.2.6. A sequence of *P. niveum* was used as the outgroup species. The outgroup species was chosen as it is genetically less close to the ingroup. A model test was conducted in MEGA X prior to Maximum Likelihood analysis.

Table 5 List of ITS2 sequences acquired from GenBank for phylogenetic analysis

No.	Species Identity	GenBank accession no.

1.	*Paphiopedilum iowii*	JQ660867.1
2.	*Paphiopedilum* niveum	JQ660879.1
3.	*Paphiopedilum barbatum*	KC6992104.1
4.	*Paphiopedilum barbatum*	JQ660872.1
5.	*Paphiopedilum barbatum*	KC692106.1
6.	*Paphiopedilum bullenianum*	KC692110.1
7.	*Paphiopedilum callosum*	KX931029.1
8.	*Paphiopedilum javanium*	GUI20212.1

Table 6 List of matK sequences acquired from GenBank for phylogenetic analysis

No.	Species Identity	GenBank accession no.
1.	*Paphiopedilum iowii*	KC692135.1
2.	*Paphiopedilum* niveum	KC692139.1
3.	*Paphiopedilum barbatum*	MG522898.1
4.	*Paphiopedilum barbatum*	MG522900.1
5.	*Paphiopedilum barbatum*	MG522901.1
6.	*Paphiopedilum bullenianum*	KC692127.1
7.	*Paphiopedilum callosum*	KX886258.1
8.	*Paphiopedilum javanium*	GUI20222.1

3. Result and Discussion

3.1 DNA Extraction

In this study, DNA was collected from fresh leaf samples of each P. *barbatum* by using the ionic detergent cetyltrimethylammonium bromide (CTAB) method and extracted with PrimeWay Plant DNA Extraction Kit. The DNA quality of these samples was assessed by Nanodrop, gel electrophoresis, and PCR and compared against those DNA samples from different places.

Table 7 Concentration of DNA of each sample Concentration of DNA of each sample

Location	Sample Name	Concentration of DNA (ng/µl)	A260/A280	A260/A230
Penang Botanic Garden	P1	17.850	1.794	1.092
	P2	16.500	1.774	1.823
	P3	6.100	1.627	1.052
	P4	9.7500	1.696	1.444
Penang Hill	P5	10.800	1.701	1.367
	P6	4.3500	1.642	1.582
	P7	9.650	1.708	1.270

Based on Table 7, DNA extraction was successful; however, multiple attempts were required to obtain a good number of results. Although all samples were fresh, DNA extractions were conducted on different days, which may have contributed to slight variations in DNA quality. The freshness of the samples could have

Enhanced Knowledge in Sciences and Technology Vol. 5 No. 2 (2025) p. 526-537

declined over time, potentially affecting extraction yields. Additionally, handling time differences during the extraction process may have further influenced the DNA quality across samples.

3.2 PCR Results of ITS2

The amplification of the ITS2 region was performed on DNA samples from P. *barbatum*. The agarose gel electrophoresis results are shown in the figure below. PCR products were successfully amplified for most of the samples, producing bands of approximately 500bp, consistent with the expected size of the ITS2 region.

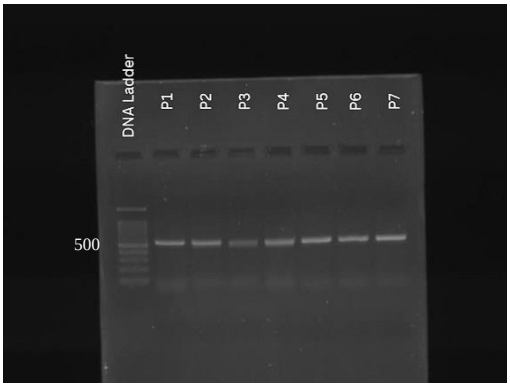

Fig. 1 The results of PCR of P. *barbatum* using ITS2

Clear and distinct bands were observed for the majority of the samples, indicating successful amplification. However, variability in band intensity was observed across some lanes, suggesting differences in DNA concentration or quality among the samples [6]. In certain instances, the amplification results required multiple attempts to achieve clear bands. Adjustments such as increasing the DNA template volume, refining the master mix composition, and optimizing the annealing temperature were necessary to improve amplification efficiency.

Sample P3 exhibited faint than the other, likely due to DNA degradation over time as fresh samples were used instead of dried samples. The decline in DNA quality over storage duration may have contributed to reduced amplification efficiency [6]. Overall, the ITS2 marker proved to be a reliable region for DNA barcoding of P.*barbatum* with successful amplification in most samples.

3.3 PCR Results of *matK*

All samples were used as templates for PCR amplification of the *matK* region to evaluate the usefulness of DNA barcoding and verify the origin of the DNA samples. The agarose gel electrophorosis results of seven samples are shown in the figure below. PCR products were successfully amplified. All of the bands were approximately 800 bp in length, confirming the accuracy of the primers used.

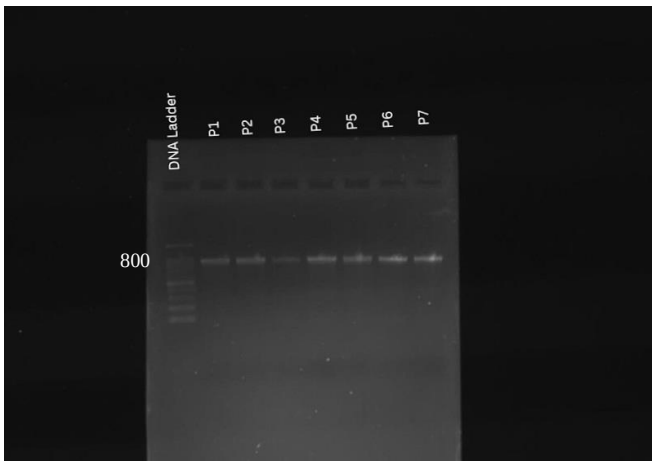

Fig. 2 The results of PCR of P. barbatum using matK

Multiple PCR attempts were required to obtain clear bands. Adjustments to the master mix included increasing $MgCl_2$ concentration to 2.5 mM, and optimizing primer concentrations to 0.5 µM. The thermocycling program was also refined, with the annealing temperature adjusted from 55 °C to 60 °C to reduce non-specific amplification. These optimizations significantly improved the amplification efficiency and clarity of the bands.

The variability in band intensity was observed, likely due to differences in DNA quality among samples. Fresh samples yielded stronger bands compared to those stored for longer durations. DNA freshness might be one of the reasons for the weak bands in earlier attempts highlight the importance of keeping the DNA fresh and quality for consistent results.

The findings highlight the effectiveness of *matK* as a dependable barcoding region for P. *barbatum*, with successful amplification across samples demonstrating strong primer accuracy and suitability for phylogenetic studies.

3.4 Phylogenetic Analysis of ITS2 Region

The phylogenetic relationships of *Paphiopedilum* species and related samples were analyzed using the *ITS2* region. Two approaches which are Maximum Parsimony, and Maximum Likelihood were used to construct the phylogenetic trees. The results of the model test for ITS2 are shown in Appendix A. These trees aim to resolve the relationships among samples and assess the genetic distinction between Penang Hill *P. barbatum* (B.Bendera1 ITS2, B.Bendera2 ITS2, B.Bendera3) ITS2 and samples collected from the Penang Botanic Garden (B.Garden1 ITS2, B.Garden2 ITS2, B.Garden4 ITS2).

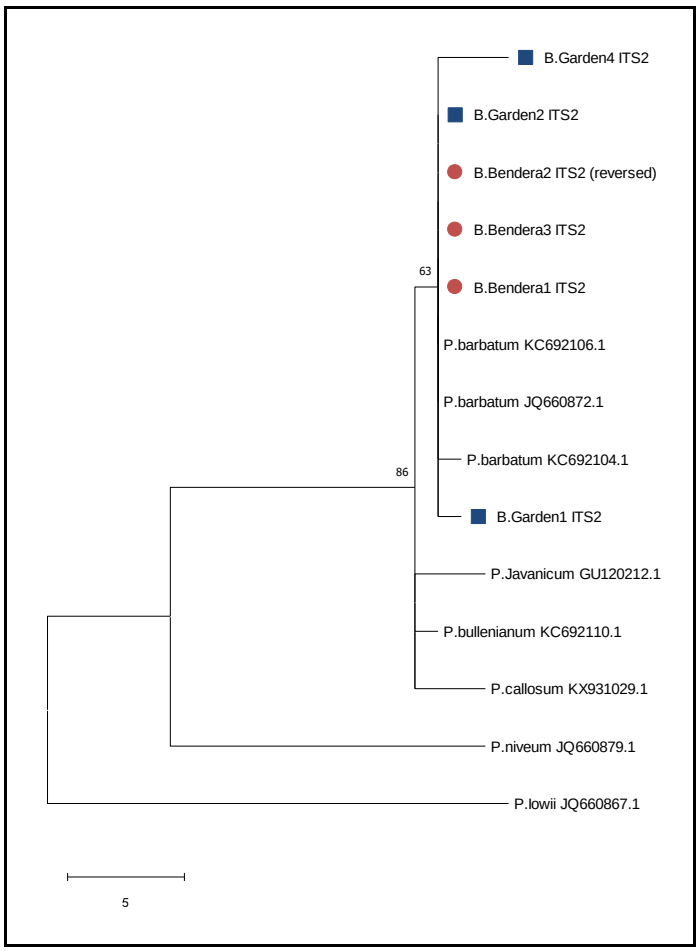

Fig. 3 Original tree of Maximum Parsimony method.

Tree #1 out of 10 most parsimonious trees (length = 62) is shown. The consistency index is (0.916667), the retention index is (0.928571), and the composite index is 0.913594 (0.851190) for all sites and parsimony-informative sites (in parentheses). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (10000 replicates) are shown next to the branches [7]. The MP tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm (pg. 126 in ref. [13]) with search level 1 in which the initial trees were obtained by the random addition of sequences (10 replicates). This analysis involved 14 nucleotide sequences. There were a total of 1020 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [8]

The phylogenetic tree constructed using the Maximum Parsimony method (see Fig. 3) reveals clustering patterns within *P. barbatum* samples from Penang Hill (B.Bendera) and Penang Botanic Garden (B.Garden). However, the low bootstrap values (<70%) indicate weak statistical support for some of these groupings,

Enhanced Knowledge in Sciences and Technology Vol. 5 No. 2 (2025) p. 526-537

suggesting that ITS2 alone may not provide sufficient resolution to distinguish between closely related populations.

B.Garden1 and B.Garden2 do not cluster closely, indicating possible genetic variation within the Penang Botanic Garden population. B.Garden1 appears more closely related to the *P. barbatum* reference sequences, while B.Garden2 forms a separate branch. Meanwhile, B.Bendera samples (B.Bendera1, B.Bendera2, B.Bendera3) form a distinct clade with a moderate bootstrap value of 63%, suggesting some degree of genetic divergence from *P. barbatum* and B.Garden samples. The placement of outgroup species (*P. lowii* and *P. niveum*) at earlier branches confirms their evolutionary distinction from *P. barbatum* and its related samples. This further supports the idea that B.Bendera and B.Garden samples are more closely related to *P. barbatum* than to these outgroup species.

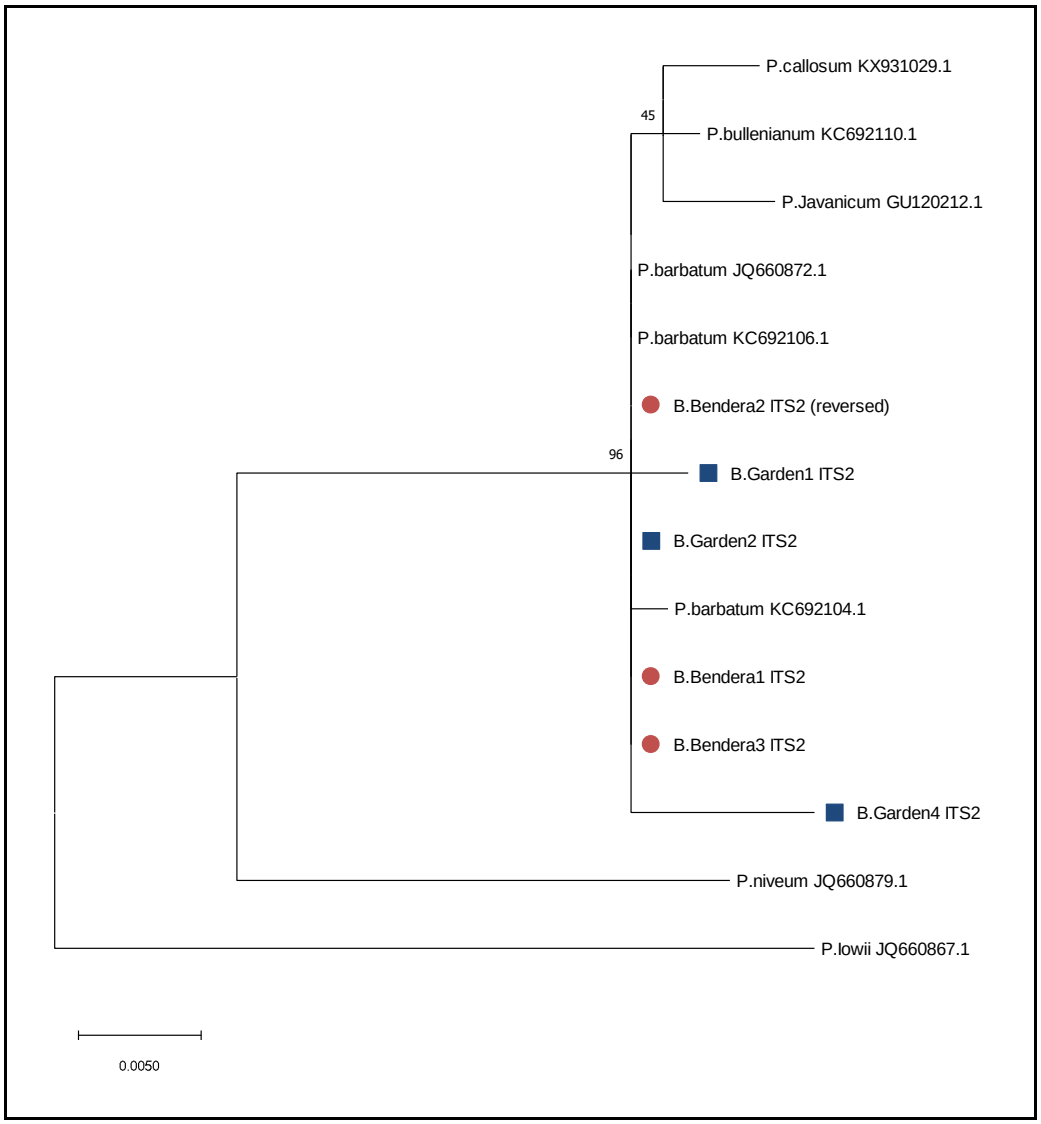

Fig. 4 Original tree of Maximum Likelihood method.

The evolutionary history was inferred by using the Maximum Likelihood method and Kimura 2-parameter model [9]. The tree with the highest log likelihood (-1612.74) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. This analysis involved 14 nucleotide sequences. There were a total of 1020 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [8].

The Maximum Likelihood analysis for the ITS2 region was performed using the Kimura 2-parameter model. The resulting phylogenetic tree reveals that *Paphiopedilum barbatum* specimens form a single clade with high bootstrap support (96%), confirming their genetic similarity. This suggests that the ITS2 region does not

provide sufficient resolution to distinguish between different populations within *P. barbatum*. However, the tree successfully separates *P. barbatum* from other *Paphiopedilum* species, forming distinct clades with measurable genetic divergence. The high bootstrap value of 96% further supports the reliability of these evolutionary relationships, indicating strong confidence in the clustering pattern. The separation of *P. barbatum* from other species, such as *P. callosum, P. bulleniianum, and P. javanicum*, highlights its genetic distinction.

The presence of high bootstrap values across multiple nodes in the tree suggests that the inferred phylogenetic relationships are well-supported and statistically robust. These results confirm that while ITS2 is a useful marker for species delimitation, it may not be suitable for resolving intraspecific variation within P. barbatum. Despite these findings, the ITS2 region placed all P. barbatum specimens, including both garden and wild populations, within a single major clade. This suggests that ITS2 does not provide enough resolution to distinguish between closely related populations within the same species. Future studies should consider using additional genetic markers, such as rbcL or trnH-psbA, to improve phylogenetic resolution.

3.5 Phylogenetic Analysis of matK

The evolutionary relationships among *Paphiopedilum* species and related samples were examined using the matK region. Two methods, Maximum Parsimony and Maximum Likelihood, were used to generate the phylogenetic trees. The results of the model test for matK are shown in Appendix B. For the Maximum Likelihood analysis of the matK region, the Tamura-3-parameter model was applied. These trees aim to clarify the relationships among samples and evaluate the genetic differentiation between Penang Hill P. *barbatum* (B.Bendera1 *matK*, B.Bendera2 *matK*, B.Bendera3 *matK*) and specimens obtained from the Penang Botanic Garden (B.Garden1 *matK*, B.Garden2 *matK*, B.Garden3 *matK*, B.Garden4 *matK*).

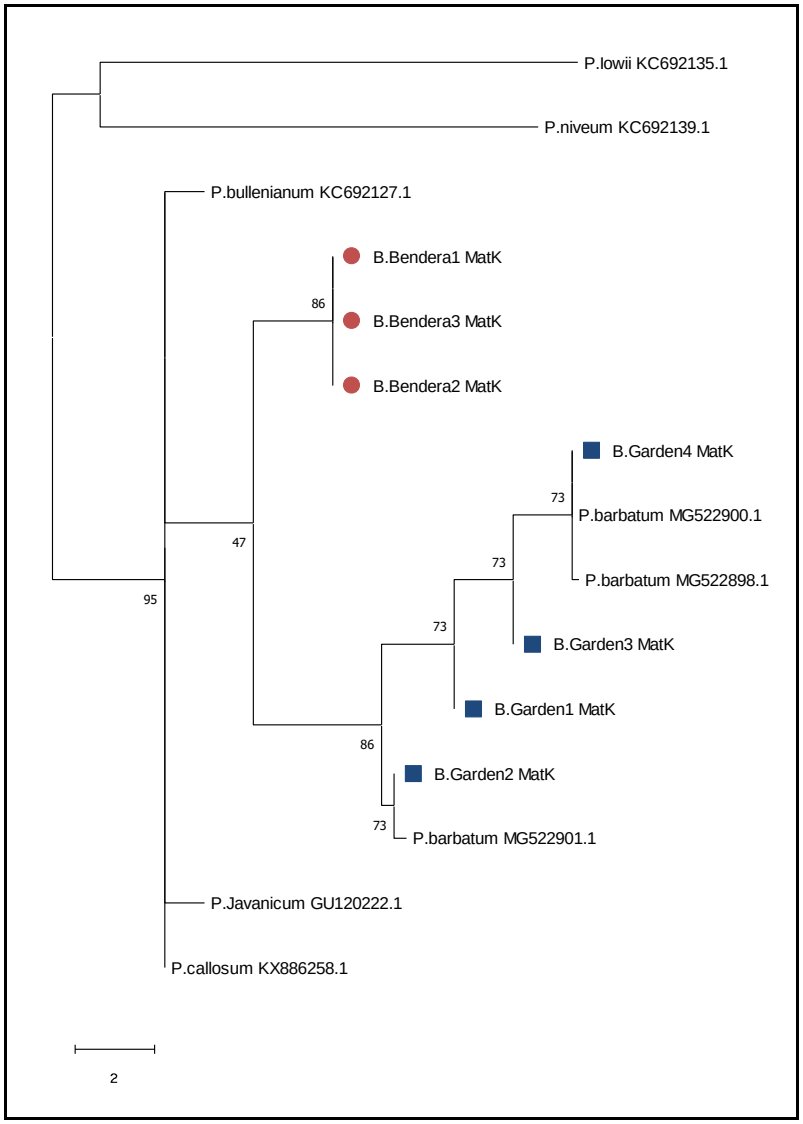

Fig. 5 Original tree of Maximum Parsimony method.

Enhanced Knowledge in Sciences and Technology Vol. 5 No. 2 (2025) p. 526-537

The evolutionary history was inferred using the Maximum Parsimony method. Tree #1 out of 9 most parsimonious trees (length = 42) is shown. The consistency index is (1.000000), the retention index is (1.000000), and the composite index is 1.000000 (1.000000) for all sites and parsimony-informative sites (in parentheses). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (10000 replicates) are shown next to the branches [8]. The MP tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm (pg. 126 in ref. [13]) with search level 1 in which the initial trees were obtained by the random addition of sequences (10 replicates). This analysis involved 15 nucleotide sequences. There were a total of 1506 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [8].

Samples from Penang Hill (B.Bendera1, B.Bendera2, B.Bendera3) form a distinct cluster separate from Penang Botanic Garden samples (B.Garden1, B.Garden2, B.Garden3, B.Garden4). This indicates a closer relationship among the B.Bendera samples. Meanwhile, the B.Garden samples are grouped, suggesting a high level of similarity among them. However, the node separating B.Bendera from B.Garden samples has a low bootstrap value (47) indicating low confidence in this separation. The grouping of other species (*P. iowii, P. niveum, P. calcosum*) at earlier nodes shows high bootstrap support (95), suggesting evolutionary divergence between the Penang Hill samples and Penang Botanic Garden samples. This shows that Penang Hill samples are genetically distinct from Botanic Garden samples.

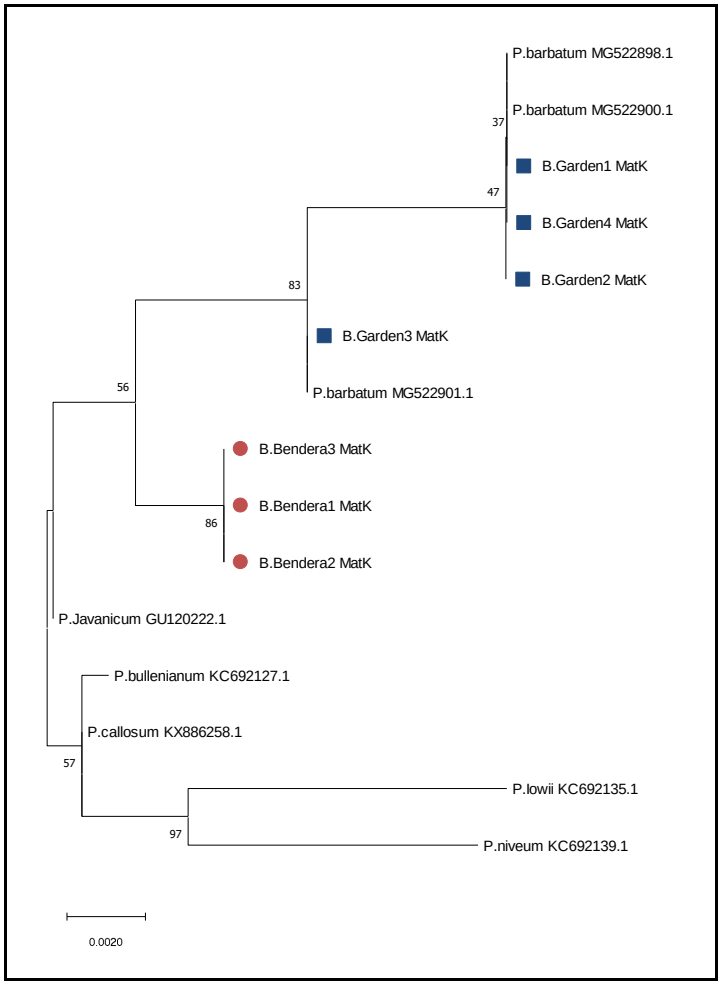

Fig. 6 Original tree of Maximum Likelihood method.

The evolutionary history was inferred by using the Maximum Likelihood method and Tamura 3-parameter model [10]. The tree with the highest log likelihood (-2280.80) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Tamura 3 parameter model and then selecting the topology with superior log likelihood value. This analysis involved 15 nucleotide sequences. There were a total of 1506 positions in the final dataset. Evolutionary analyses were conducted in MEGA X [8].

Based on Appendix A, the Maximum Likelihood (ML) analysis for matK uses an evolutionary models chosen based on the best fit model selection. For the matK region, the Tamura 3-parameter model was applied. Similar

to the Maximum Parsimony tree, B.Bendera samples cluster distinctly, indicating a close genetic relationship among them. Meanwhile, the B.Garden samples formed a seperate branch, distinct from the B.Bendera group.

The matK phylogenetic trees demonstrated a clearer distinction between the Penang Hill and Penang Botanic Garden samples compared to ITS2. While some Penang Botanic Garden samples grouped closely with known P. barbatum sequences, others displayed a slight divergence, suggesting potential genetic variation within the population. Despite this, the bootstrap values remained relatively low, indicating that the genetic differentiation is not strongly supported. This could be due to limitations in the matK region's variability or possible gene flow between populations. Overall, the results suggest that while matK provides better phylogenetic resolution than ITS2, additional markers are needed to confirm the genetic distinction of Penang Hill P. barbatum. Future research should look at the entire DNA of the plant or use other DNA regions that show more difference. This will help get clearer results and better confirm how the Penang Hill and Botanic Garden samples are related.

However, there are significant concerns regarding the genetic integrity of the Botanic Garden samples. If they are not genetically identical to the original Penang Hill population, reintroduction could lead to genetic contamination through hybridization, potentially reducing the adaptive potential of the wild population. Additionally, plants that have been in controlled environments may lack the necessary adaptations to survive in their natural habitat, leading to a low survival rate after reintroduction. Careful genetic verification is crucial before proceeding with reintroduction efforts [11]. The findings of this study suggest uncertainty in the genetic distinction between the Penang Hill and Botanic Garden samples due to the low bootstrap values in the phylogenetic analysis. To ensure successful and ecologically responsible repopulation, further genetic studies, such as whole-genome sequencing or population genetic analysis, should be conducted to confirm the suitability of the Botanic Garden samples for reintroduction without compromising the native population.

4. Conclusion

This study was able to obtain and barcode several wild *P. barbatum* specimens originating from Penang Hill and used the DNA barcoding technique with the matK and ITS2 regions. While the phylogenetic analysis provided some insights into genetic relationships, the bootstrap values for both regions were relatively low, indicating weak statistical support for genetic distinction between the two populations. The ITS2 results, in particular, suggest a close genetic similarity between the Penang Hill and Botanic Garden samples, making it inconclusive to determine their exact origin based solely on these markers. Future studies incorporating additional molecular markers or whole-genome sequencing would be beneficial in resolving taxonomic ambiguities and supporting conservation and reintroduction efforts.

Acknowledgement

The authors would like to thank the Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, for its support. Also, the Penang Botanic Garden for their collaboration throughout this project.

Conflict of Interest

The authors confirm that there are no conflicts of interest related to the publication of this paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Khairun Ayuni Afiqah Hassan; Yap Jing Wei;* ***sampling and data collection:*** *Khairun Ayuni Afiqah Hassan, Rexy Prakash Chako;* ***analysis and interpretation of results:*** *Khairun Ayuni Afiqah Hassan, Yap Jing Wei;* ***draft manuscript preparation:*** *Khairun Ayuni Afiqah Hassan, Yap Jing Wei. All authors reviewed the results and approved the final version of the manuscript.*

Appendix A: The Substitution Model (Maximum Likelihood) for ITS2

Table. Maximum Likelihood fits of 24 different nucleotide substitution models

Model	Parameters	BIC	AICc	lnL	(+I)	(+G)	R	f(A)	f(T)	f(C)	f(G)	r(AT)	r(AC)	r(AG)	r(TA)	r(TC)	r(TG)	r(CA)	r(CT)	r(CG)	r(GA)	r(GT)	r(GC)
K2	26	3460.225	3277.745	-1612.788	n/a	n/a	1.78	0.250	0.250	0.250	0.250	0.045	0.045	0.160	0.045	0.160	0.045	0.045	0.160	0.045	0.160	0.045	0.045
K2+G	27	3469.061	3279.569	-1612.693	n/a	2.13	1.80	0.250	0.250	0.250	0.250	0.045	0.045	0.161	0.045	0.161	0.045	0.045	0.161	0.045	0.161	0.045	0.045
K2+I	27	3469.250	3279.758	-1612.788	0.00	n/a	1.78	0.250	0.250	0.250	0.250	0.045	0.045	0.160	0.045	0.160	0.045	0.045	0.160	0.045	0.160	0.045	0.045
T92	27	3469.580	3280.089	-1612.953	n/a	n/a	1.78	0.247	0.247	0.253	0.253	0.044	0.046	0.162	0.044	0.162	0.046	0.044	0.158	0.046	0.158	0.044	0.046
JC	25	3475.488	3300.021	-1624.932	n/a	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
K2+G+I	28	3478.085	3281.582	-1612.693	0.00	2.13	1.80	0.250	0.250	0.250	0.250	0.045	0.045	0.161	0.045	0.161	0.045	0.045	0.161	0.045	0.161	0.045	0.045
T92+G	28	3478.420	3281.917	-1612.860	n/a	2.16	1.80	0.247	0.247	0.253	0.253	0.044	0.045	0.163	0.044	0.163	0.045	0.044	0.158	0.045	0.158	0.044	0.045
T92+I	28	3478.608	3282.105	-1612.954	0.00	n/a	1.78	0.247	0.247	0.253	0.253	0.044	0.046	0.162	0.044	0.162	0.046	0.044	0.158	0.046	0.158	0.044	0.046
JC+G	26	3484.450	3301.970	-1624.900	n/a	4.19	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC+I	26	3484.513	3302.033	-1624.932	0.00	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
HKY	29	3484.705	3281.191	-1611.490	n/a	n/a	1.78	0.236	0.257	0.238	0.269	0.046	0.043	0.172	0.042	0.153	0.048	0.042	0.165	0.048	0.151	0.046	0.043
T92+G+I	29	3487.445	3283.931	-1612.860	0.00	2.16	1.80	0.247	0.247	0.253	0.253	0.044	0.045	0.163	0.044	0.163	0.045	0.044	0.158	0.045	0.158	0.044	0.045
JC+G+I	27	3493.475	3303.983	-1624.900	0.00	4.19	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
HKY+G	30	3493.542	3283.018	-1611.396	n/a	2.15	1.80	0.236	0.257	0.238	0.269	0.046	0.042	0.173	0.042	0.153	0.048	0.042	0.166	0.048	0.152	0.046	0.042
HKY+I	30	3493.730	3283.205	-1611.490	0.00	n/a	1.78	0.236	0.257	0.238	0.269	0.046	0.043	0.172	0.042	0.153	0.048	0.042	0.165	0.048	0.151	0.046	0.043
TN93	30	3493.759	3283.235	-1611.505	n/a	n/a	1.78	0.236	0.257	0.238	0.269	0.046	0.043	0.172	0.042	0.153	0.048	0.042	0.166	0.048	0.150	0.046	0.043
TN93+G	31	3502.566	3285.032	-1611.396	n/a	2.15	1.80	0.236	0.257	0.238	0.269	0.046	0.042	0.172	0.042	0.154	0.048	0.042	0.167	0.048	0.151	0.046	0.042
HKY+G+I	31	3502.569	3285.034	-1611.397	0.00	2.15	1.80	0.236	0.257	0.238	0.269	0.046	0.042	0.173	0.042	0.153	0.048	0.042	0.166	0.048	0.152	0.046	0.042
TN93+I	31	3502.757	3285.222	-1611.491	0.00	n/a	1.78	0.236	0.257	0.238	0.269	0.046	0.043	0.172	0.042	0.153	0.048	0.042	0.166	0.048	0.150	0.046	0.043
TN93+G+I	32	3511.590	3287.046	-1611.396	0.00	2.15	1.80	0.236	0.257	0.238	0.269	0.046	0.042	0.172	0.042	0.154	0.048	0.042	0.167	0.048	0.151	0.046	0.042
GTR	33	3516.537	3284.984	-1609.356	n/a	n/a	1.55	0.236	0.257	0.238	0.269	0.066	0.025	0.154	0.060	0.154	0.038	0.025	0.166	0.075	0.135	0.036	0.066
GTR+G	34	3524.271	3285.710	-1608.711	n/a	2.91	1.79	0.236	0.257	0.238	0.269	0.065	0.025	0.172	0.060	0.153	0.022	0.024	0.166	0.075	0.150	0.021	0.066
GTR+I	34	3525.447	3286.886	-1609.299	0.00	n/a	1.56	0.236	0.257	0.238	0.269	0.066	0.025	0.155	0.060	0.154	0.037	0.025	0.166	0.075	0.136	0.036	0.066
GTR+G+I	35	3533.299	3287.729	-1608.712	0.00	2.91	1.79	0.236	0.257	0.238	0.269	0.065	0.025	0.172	0.060	0.153	0.022	0.024	0.166	0.075	0.150	0.021	0.066

Appendix B: The Substitution Model (Maximum Likelihood) for matK

Table. Maximum Likelihood fits of 24 different nucleotide substitution models

Model	Parameters	BIC	AICc	lnL	(+I)	(+G)	R	f(A)	f(T)	f(C)	f(G)	r(AT)	r(AC)	r(AG)	r(TA)	r(TC)	r(TG)	r(CA)	r(CT)	r(CG)	r(GA)	r(GT)
T92	29	4850.441	4628.247	-2285.068	n/a	n/a	0.87	0.343	0.343	0.157	0.157	0.085	0.039	0.079	0.085	0.079	0.039	0.085	0.173	0.039	0.173	0.085
T92+G	30	4857.644	4627.792	-2283.837	n/a	0.18	0.88	0.343	0.343	0.157	0.157	0.085	0.039	0.079	0.085	0.079	0.039	0.085	0.173	0.039	0.173	0.085
T92+I	30	4858.388	4628.536	-2284.209	0.49	n/a	0.87	0.343	0.343	0.157	0.157	0.085	0.039	0.079	0.085	0.079	0.039	0.085	0.173	0.039	0.173	0.085
HKY	31	4861.372	4623.862	-2280.868	n/a	n/a	0.87	0.307	0.379	0.164	0.150	0.095	0.041	0.075	0.077	0.082	0.037	0.077	0.189	0.037	0.154	0.095
T92+G+I	31	4867.320	4629.810	-2283.842	0.47	0.71	0.88	0.343	0.343	0.157	0.157	0.085	0.039	0.079	0.085	0.079	0.039	0.085	0.173	0.039	0.173	0.085
HKY+G	32	4868.526	4623.358	-2279.612	n/a	0.17	0.88	0.307	0.379	0.164	0.150	0.094	0.041	0.075	0.077	0.082	0.037	0.077	0.190	0.037	0.154	0.094
TN93	32	4868.589	4623.421	-2279.644	n/a	n/a	0.87	0.307	0.379	0.164	0.150	0.094	0.041	0.105	0.076	0.057	0.037	0.076	0.130	0.037	0.214	0.094
HKY+I	32	4869.302	4624.135	-2280.000	0.49	n/a	0.87	0.307	0.379	0.164	0.150	0.095	0.041	0.075	0.077	0.082	0.037	0.077	0.190	0.037	0.154	0.095
GTR	35	4873.303	4605.164	-2267.502	n/a	n/a	0.75	0.307	0.379	0.164	0.150	0.015	0.048	0.084	0.012	0.053	0.099	0.091	0.121	0.025	0.173	0.251
TN93+G	33	4875.785	4622.961	-2278.409	n/a	0.18	0.88	0.307	0.379	0.164	0.150	0.093	0.040	0.105	0.076	0.056	0.037	0.076	0.129	0.037	0.216	0.093
TN93+I	33	4876.533	4623.709	-2278.783	0.49	n/a	0.87	0.307	0.379	0.164	0.150	0.093	0.041	0.105	0.076	0.056	0.037	0.076	0.130	0.037	0.215	0.093
GTR+G	36	4877.098	4601.303	-2264.567	n/a	0.21	0.89	0.307	0.379	0.164	0.150	0.013	0.049	0.100	0.011	0.052	0.101	0.091	0.121	0.000	0.206	0.255
GTR+I	36	4877.739	4601.944	-2264.887	0.49	n/a	0.88	0.307	0.379	0.164	0.150	0.015	0.049	0.099	0.012	0.053	0.101	0.091	0.122	0.000	0.204	0.254
HKY+G+I	33	4878.201	4625.377	-2279.617	0.47	0.69	0.88	0.307	0.379	0.164	0.150	0.094	0.041	0.075	0.077	0.082	0.037	0.077	0.190	0.037	0.154	0.094
TN93+G+I	34	4885.455	4624.973	-2278.411	0.45	0.66	0.88	0.307	0.379	0.164	0.150	0.093	0.040	0.105	0.076	0.056	0.037	0.076	0.129	0.037	0.216	0.093
GTR+G+I	37	4889.841	4606.390	-2266.106	0.39	0.64	0.77	0.307	0.379	0.164	0.150	0.013	0.048	0.087	0.011	0.052	0.100	0.090	0.120	0.022	0.179	0.253
JC	27	5055.790	4848.913	-2397.408	n/a	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
K2	28	5062.242	4847.707	-2395.802	n/a	n/a	0.87	0.250	0.250	0.250	0.250	0.067	0.067	0.117	0.067	0.117	0.067	0.067	0.117	0.067	0.117	0.067
JC+G	28	5062.597	4848.062	-2395.979	n/a	0.11	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC+I	28	5063.602	4849.066	-2396.481	0.49	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
K2+G	29	5068.932	4846.738	-2394.314	n/a	0.10	0.88	0.250	0.250	0.250	0.250	0.066	0.066	0.117	0.066	0.117	0.066	0.066	0.117	0.066	0.117	0.066
K2+I	29	5070.035	4847.841	-2394.865	0.49	n/a	0.88	0.250	0.250	0.250	0.250	0.067	0.067	0.117	0.067	0.117	0.067	0.067	0.117	0.067	0.117	0.067
JC+G+I	29	5072.215	4850.021	-2395.955	0.00	0.11	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
K2+G+I	30	5078.597	4848.745	-2394.314	0.00	0.10	0.88	0.250	0.250	0.250	0.250	0.066	0.066	0.117	0.066	0.117	0.066	0.066	0.117	0.066	0.117	0.066

References

[1] Swarts, N. D., & Dixon, K. W. (2009). Terrestrial orchid conservation in the age of extinction. Annals of Botany, 104(3), 543-556.

[2] Wraith, J., Norman, P., & Pickering, C. (2020). Orchid conservation and research: An analysis of gaps and priorities for globally Red Listed species. Ambio, 49, 1601 - 1611. https://doi.org/10.1007/s13280-019-01306-7.

[3] Hollingsworth, P. M., Li, D.-Z., van der Bank, M., & Twyford, A. D. (2016). Telling plant species apart with DNA: From barcodes to genomes. Philosophical Transactions of the Royal Society B: Biological Sciences, 371(1702), 20150338. https://doi.org/10.1098/rstb.2015.0338

[4] Rajaram, M. C., Yong, C. S. Y., Gansau, J. A., & Go, R. (2019). Strong DNA barcoding of endangered Paphiopedilum species (Orchidaceae) of Peninsular Malaysia. Phytotaxa, 387(2), 94–104. https://doi.org/10.11646/phytotaxa.387.2.2

[5] CBOL Plant Working Group. (2009). A DNA barcode for land plants. Proceedings of the National Academy of Sciences, 106(31), 12794-12797. https://doi.org/10.1073/pnas.0905845106

[6] Cankar, K., Štebih, D., Dreo, T., Žel, J., & Gruden, K. (2006). BMC Biotechnology, 6(1), 37. https://doi.org/10.1186/1472-6750-6-37

[7] Felsenstein J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783-791.

[8] Kumar S., Stecher G., Li M., Knyaz C., and Tamura K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution 35:1547-1549.

[9] Kimura M. (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16:111-120.
[10] Tamura K. (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C-content biases. Molecular Biology and Evolution 9:678-687
[11] St. Clair, A., Dunwiddie, P., Fant, J., Kaye, T., & Kramer, A. (2020). Mixing source populations increases genetic diversity of restored rare plant populations. Restoration Ecology, 28. https://doi.org/10.1111/rec.13131.