MARKER ASSESSMENT, MOLECULAR CHARACTERISATION AND PHYLOGENY OF PENINSULAR MALAYSIAN *Begonia* spp. UNDER THE SECTION *Jackia*

By

YONG CAYLEE

A project report submitted to the Department of Agricultural and Food Science

Faculty of Science

Universiti Tunku Abdul Rahman

in partial fulfilment of the requirements for the degree of

Bachelor of Science (Honours) Agricultural Science

September 2025

ABSTRACT

MARKER ASSESSMENT, MOLECULAR CHARACTERISATION AND PHYLOGENY OF PENINSULAR MALAYSIAN *Begonia* spp. UNDER THE SECTION *Jackia*

YONG CAYLEE

Begonia L. encompasses over 2,000 species categorised into 70 sections. Peninsular Malaysia harbours numerous endemic *Begonia* species, many of which belong to Section Jackia, and are at risk of extinction. Most of these species are poorly identified and documented, with limited genetic and phylogenetic data available. This study employed nuclear (ITS) and chloroplast DNA markers (*ndh*A intron, *ndh*F–*rpl*32, and *rpl*32–*trn*L) as recommended by previous *Begonia* studies. These markers, either in individual or in concatenated form, were used to assess species delimitation and phylogeny of five *Begonia* samples from Peninsular Malaysia, alongside species from GenBank and earlier studies. The concatenated dataset (ITS+*ndh*A intron+*ndh*F–*rpl*32+*rpl*32–*trn*L) successfully delineated species but was unable to resolve the phylogenetic relationships between *B. nurii* and *B. foxworthyi*. Species delimitation analyses using ASAP, ABGD, PTP and GMYC revealed at least 14

species of *Begonia* in the dataset, although the classification of specimens BGNM1 and BGNG1 varied depending on the algorithm. This was further supported by genetic distance data, which recorded an interspecific genetic divergence of 0.51–2.16%. At least two species are potentially novel, while the taxonomic status of others (i.e., *B*. cf. *nurii* and *B*. cf. *foxworthyi*) remains uncertain due to missing or incorrect genetic data of type specimens. The study provided a preliminary phylogeny of *Begonia* within Section *Jackia* and highlights the rich diversity of Begonia in Peninsular Malaysia. Furthermore, the incorporation of floral morphological and genetic data of a wild-collected *B. rajah* confirmed its identity, establishing that it is extant in the wild after nearly a century of presumed extinction. The performance of each genetic marker (i.e., ease of amplification and sequencing, phylogenetic informativeness, etc.) was assessed, laying valuable groundwork for advancing taxonomic research optimised for the fast-evolving *Begonia* species of Southeast Asia.

ACKNOWLEDGEMENTS

Above all, I would like to show my deepest appreciation to my supervisor, Dr. Tan Ji, for his dedication, thorough guidance and unwavering patience throughout the entirety of this project. Besides this, I would also like to thank Miss Hazlinda for helping me with all the issues I faced in the lab. Special thanks to Vanesa who accompanied me through all the trials and tribulations. Lastly, I extend my heartfelt gratitude to my family, friends, and my cat for their emotional support and encouragement in the completion of my final year project.

DECLARATION

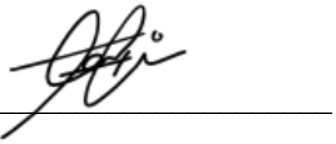
I hereby declare that this final year project report is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UTAR or other institutions.

YONG CAYLEE

APPROVAL SHEET

This final year project entitled "MARKER ASSESSMENT, MOLECULAR CHARACTERISATION AND PHYLOGENY OF PENINSULAR MALAYSIAN *Begonia* spp. UNDER THE SECTION *Jackia*" was prepared by YONG CAYLEE and submitted as partial fulfilment of the requirements for the degree of Bachelor of Science (Honours) Agricultural Science at Universiti Tunku Abdul Rahman.

Approved by:



Dr. Tan Ji
Supervisor
Department of Agricultural and Food Science
Faculty of Science
Universiti Tunku Abdul Rahman

Date: 05/09/2025

FACULTY OF SCIENCE

UNIVERSITI TUNKU ABDUL RAHMAN

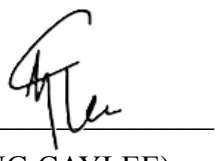
Date: 6/9/2025

PERMISSION SHEET

I, **Yong Caylee** (ID No: **23ADB00540**) hereby certify that I have completed the final year project titled "MARKER ASSESSMENT, MOLECULAR CHARACTERISATION AND PHYLOGENY OF PENINSULAR MALAYSIAN *Begonia* spp. UNDER THE SECTION *Jackia*" under the supervision of Dr. Tan Ji (Supervisor) from the Department of Agricultural and Food Science, Faculty of Science.

I understand that the University may upload the softcopy of my final year project in PDF format to the UTAR Institutional Repository, which may be made accessible to the UTAR community and public.

Yours truly,



(YONG CAYLEE)

Universiti Tunku Abdul Rahman	FORM	REF NO. : FM-LIB-501
		REVISION NO. : 0
	UTAR Institutional Repository (UTAR-IR): Permission Form	EFFECTIVE DATE : 24/06/2024
		PAGE : 1

Title of Final Work : MARKER ASSESSMENT, MOLECULAR CHARACTERISATION AND PHYLOGENY OF PENINSULAR MALAYSIAN *Begonia* spp. UNDER THE SECTION *Jackia*

Degree Awarded : Bachelor of Science (Honours) Agricultural Science

Faculty : Faculty of Science

By signing below, I acknowledge that I have fully read and understood the Open Access (OA) Policy for Scholarly Output, and agree to abide by the terms and conditions outlined in the Policy.

I certify that the version of the document(s) that I deposited in UTAR-IR is the same as that approved by the examination committee, and hereby declare that I own the copyright to this work and it does not contain any unauthorised third party copyrighted materials.

I hereby grant to Universiti Tunku Abdul Rahman (UTAR) permission to make available my thesis / project work open access in UTAR-IR, and to exercise the copyright in those materials. More specifically, I grant to UTAR a non-exclusive, irrevocable, worldwide licence to exercise any and all rights under copyright relating to my thesis / project work, in any medium, and to authorise others to do the same. I shall retain copyright in my thesis / project work.

Alternative Access Option (if applicable):	
	Restricted Access. Reason(s): _____
	Embargo for up to 24 months. Reason(s): _____ Start Date: _____End Date: _____Total Duration: _____months

Signature: _____

Name: Yong Caylee

ID: 23ADB00540

Date: 8/10/2025

Signature: _____

Supervisor's Name: Dr. Tan Ji

Date: 8/10/2025

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DNA	Deoxyribonucleic acid
ITS	Internal Transcribed Spacer
***ndh*A**	NADH dehydrogenase subunit A
***ndh*F–*rpl*32**	NADH dehydrogenase subunit F gene–ribosomal protein L32 gene
***rpl*32–*trn*L**	Ribosomal protein L32 gene–transfer RNA-Leucine gene
PCR	Polymerase Chain Reaction
MSA	Multiple sequence alignment
ML	Maximum likelihood
MP	Maximum parsimony
BI	Bayesian inference
MCMC	Markov chain Monte Carlo
ASAP	Assemble species by automatic partitioning
ABGD	Automatic barcode gap discovery
bPTP	Bayesian Poisson tree processes
GMYC	Generalised mixed Yule Coalescent
SSC	Small-single copy
TE	Tris-EDTA

CHAPTER 1

INTRODUCTION

1.1 Overview of *Begonia* Taxonomy and Conservation Status in Malaysia

Begonia L. is amongst the top 10 largest genera of flowering plants, with approximately 2,198 accepted documented species making up 70 different sections (Hughes et al., 2025). The genus comprises morphologically distinct species ranging from terrestrial herbs to shrubs that generally grow perennially and rarely annually, and they typically possess asymmetrical leaves and unisexual flowers (Dewitte et al., 2011; Rjosk, Neinhuis and Lautenschläger, 2022). *Begonia* displays a wide natural distribution and can be found in both tropical and subtropical areas, including Africa, South and Central America and Southern Asia (Neale, Goodall-Copestake, and Kidner, 2006; Emelianova et al., 2021). While some species are consumed as food and others offer valuable medicinal properties, begonias are primarily valued commercially as ornamental houseplants for their striking flowers and unique foliage patterns and colours (Buyun et al., 2019; Marchioni et al., 2022; Permata and Susandarini, 2022).

Malaysia is recognised as one of the world's biodiversity hotspots, where its tropical rainforests and limestone habitats support a rich variety of flora and

fauna, including numerous *Begonia* species (Latiff, 2018). These habitats often provide shaded areas of high yet stable humidity—conditions that allow for begonias to grow and thrive (Ardi et al., 2018; Siregar et al., 2021). To date, around 167 named species of *Begonia* are reportedly native to Malaysia (57 in P. Malaysia, 113 in East Malaysia), and many more species are yet to be described, especially in Sabah and Sarawak (Sang and Kiew, 2014; Tan, Tam and Kiew, 2018). Despite the species diversity, *Begonia* populations are usually confined to a niche environment that they have adapted to and are thus susceptible to illegal poaching as well as loss of habitat due to deforestation for agricultural activities, quarry mining and infrastructure development (Tian et al., 2018). As such, 57% of *Begonia* species in Peninsular Malaysia have been classified as threatened, where 24 taxa are considered endangered, and 21 species have only been observed in one location (Chua, Kiew and Chan, 2009).

Begonia is one of the fastest-growing genera in terms of new species discovery and classification, encompassing a morphologically diverse range of species that can make it taxonomically complex (Moonlight et al., 2018). Therefore, it is crucial for proper species delimitation, species delimitation, which is the process or method used to identify and define boundaries between species (Wiens, 2007). Before the advent of molecular systematics, species descriptions and cataloging of *Begonia* relied on morphology and geographic distribution (Doorenbos, Sosef, and Wilde, 1998). However, this approach led to inconsistencies in identifications, primarily due to phenotypic plasticity, inadequate plant distribution records, and

the incomplete collection of data when they were first described (Naive et al., 2022). Therefore, the incorporation of genetic information for species identification and phylogeny has significantly advanced the taxonomic development of *Begonia* over the past three decades (Badcock, 1998). This was achieved with the use of DNA markers (sequences of DNA of known location) such as the internal transcribed spacer (ITS), NADH dehydrogenase subunit A (*ndh*A) intron, spacer between NADH dehydrogenase subunit F gene and ribosomal protein L32 gene (*ndh*F–*rpl*32), and spacer between ribosomal protein L32 gene and transfer RNA-Leucine gene (*rpl*32–*trn*L) (Al-Samarai and Al-Kazaz, 2015; Ardi et al., 2022). This approach facilitates a more reliable and objective detection of both intra- and interspecific genetic variations. Although most DNA markers are well-amplified and sequenced using standard *Taq* DNA polymerase, recent studies suggested that DNA polymerase with proofreading ability, such as Phusion™ DNA polymerase, could aid in better sequencing accuracy (Gohl et al., 2021; McInerney, Adams and Hadi, 2014).

DNA markers that are amplified and sequenced are then used for DNA barcoding, a technique that utilises a short segment of DNA to quickly and accurately identify a species. This method offers an easy alternative to a time-consuming process, which is especially important for low-level, non-taxonomic studies (de Vere et al., 2014; Mishra et al., 2015). Despite multiple studies involving nuclear, chloroplast and mitochondrial markers, a consensus on which universal region is to be used as a DNA barcode for all begonias worldwide has not been reached

(Jiao and Shui 2013; Tseng et al., 2022; Xiong et al., 2023). This is particularly true for begonias in Southeast Asia, which evolve more rapidly and have frequently been undersampled and underrepresented in previous studies (Harrison, Harrison and Kidner, 2016; Jiao and Shui, 2013; Mahardika, 2016). Such limitations in data underscore the importance of better genetic characterisation of *Begonia* species in this geographical region.

An example of the valuable application of molecular data would be the erection of the section *Jackia* M. Hughes (previously sect. *Reichenhemia*) by Moonlight et al. (2018), to house *Begonia* spp. found throughout Malesia (Malaysia, Singapore, Indonesia, Brunei, the Philippines, Timor-Leste and Papua New Guinea). Members of this section are rhizomatous and have inflorescence with stamen columns which are protandrous as well as trilocular ovaries with continuous placentation (Ardi et al., 2018). Currently, *Begonia* sect. *Jackia* is comprised of 76 species, 11 of which could be found in Peninsular Malaysia, namely *B. corneri, B. forbesii, B. foxworthyi, B. ignorata, B. lengguanii, B. nurii, B. rajah, B. reginula, B. tigrina, B. yappii,* and *B. yenyeniae* (Hughes et al., 2025). Due to the endemicity of most of these species as well as the lack of funds and manpower, they are poorly documented both morphologically and molecularly (Chua, Kiew and Chan, 2009; Kiew, 2005). A popular species from this list is *Begonia rajah*, desired by many collectors for its bronzy variegation between its green veins. After over a century since it was first found in 1892, *B. rajah* is still

thought to be extinct in the wild, although recent DNA data suggested otherwise (Ho, 2024; Tan, Tam and Kiew, 2018).

1.2 Problem Statement

Although begonias are diverse in Peninsular Malaysia, they are poorly documented, endemic and thus susceptible to habitat destruction and extinction. These issues highlight the need to characterise *Begonia* species of the section *Jackia*, to provide better insights into the genetic and phylogenetic data of samples, primarily those from Kelantan and Terengganu, two poorly sampled states of Malaysia.

This can be achieved using gross morphological data supplemented by genetic data based on four most commonly-used *Begonia* DNA markers, i.e., ITS, *ndh*A intron, *ndh*F–*rpl*32 spacer, and *rpl*32–*trn*L spacer. The use of high-fidelity DNA polymerases offers valuable insights into overall sequencing quality and helps determine whether their use is truly necessary. Analysis of these DNA markers will contribute valuable genetic data on *Begonia* sect. *Jackia* to the currently limited database, offering insights into its phylogeny, and also identifying the most effective DNA markers for barcoding. This taxonomic framework is essential in facilitating species discovery, future conservation and bioprospecting efforts.

1.3 Objectives

The objectives of this study are:

i. To infer the phylogenetic relationships between members of *Begonia* sect. *Jackia* from Peninsular Malaysia using one nuclear DNA marker and three chloroplast DNA markers.

ii. To delimitate *Begonia* species under sect. *Jackia* based on distance matrixes and ultrametric trees.

iii. To assess individual marker performance and identify which one is best suited for the DNA barcoding of *Begonias* within Peninsular Malaysia.

iv. To complete the morphological and genetic characterisation of wild *B. rajah* collected from Terengganu in 2024 and confirm its 'extinct-in-the-wild' status.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview of *Begonia*

Begonia L. is a genus of angiosperms that houses an approximate of 2,198 accepted species, which are categorised into 70 sections (taxonomic rank between genus and species to categorise large genera; Hughes et al., 2025). It was first named by Charles Plumier in 1695, which was later adopted by Charles Linnaeus in 1753 (Doorenbos, Sosef, and Wilde, 1998). Given the vast number of species within the genus, there is great diversity to be seen throughout the morphology, making them either very similar or distinct from each other. Despite this, they share common characteristics such as asymmetrical leaves and monoecious flowers (Rjosk, Neinhuis and Lautenschläger, 2022). Most of the species grow perennially and rarely annually in the form of terrestrial herbs (Dewitte et al., 2011).

2.1.1 Economic Roles of *Begonia*

Begonias are popular as ornamental plants due to their dainty flowers but mainly their unique foliar morphology (Buyun et al., 2019). There is a large market for *Begonia* hybrids, especially in Europe and the United States (Neale, Goodall-Copestake and Kidner, 2006). However, they also have cultural significance and are traditional potherbs for countries such as Japan, Indonesia, India, and Nepal (Rajbhandary, 2015). For culinary purposes, some species of *Begonia*, i.e., *Begonia* × *tuberhybrida*, *B. cucullata*, *B. nelumbifolia* etc., are eaten raw in salads or as a snack (Laferrière, 1992), even though most *Begonia* are toxic to humans. The acidity of edible *Begonia* is also useful in the curdling and manufacturing of cheese. In addition to this, many begonias are used as traditional medication which can aid illnesses such as colds, fever, gastrointestinal diseases, diabetes, jaundice, etc. (Lalawmpuii and Tlau, 2021). This is because some species, like *Begonia roxenburgii,* are chemically rich and possess antioxidative and antimicrobial properties.

2.1.2 Distribution and Habitat of *Begonia*

Begonias have a pantropical distribution which spans across Africa, Central America, and Asia, with the exception of Australia (Neale, Goodall-Copestake and Kidner, 2006). It is theorised that the diversification of *Begonia* lineages originated in Africa, followed by the other two regions (Goodall-Copestake et al., 2010). This is supported by how African *Begonia* species are seemingly older, as

they have the greatest morphological diversity despite having the least number of species (Neale, Goodall-Copestake and Kidner, 2006).

Generally, wild begonias thrive in shaded locations with high and stable humidity such as alongside rivers and creeks within primary and secondary as well as limestone karsts (Chung et al., 2014; Satyanti, 2012). They tend to grow on the forest floor, lithophytically and occasionally epiphytically, in areas which have little leaf litter coverage as it limits their growth and dispersal (Thomas, 2010). Their sensitivity to fluctuating light and humidity causes them to be highly endemic, making them more vulnerable to disturbances.

2.1.3 Conservation Status of *Begonia*

Under the IUCN Red List, there are 79 threatened species of *Begonia* as well as one extinct species, *Begonia eiromischa* (IUCN, 2025). Despite being internationally credible, the list does not accurately reflect the conservatory status of *Begonia* due to the lack of species listings within the database, as they predominantly focus on commercially important species (Edgar, 2025). As begonias are highly endemic, they face a lot of threats to their population (Radbouchoom et al., 2024). Since they are highly desirable and valued as ornamental plants, they are subjected to illegal poaching. Besides this, they are vulnerable to habitat loss as a result of global warming and human development.

Deforestation and illegal logging disrupt the habitats of begonias causing it to stray from their optimal growing conditions, affecting the distribution of begonias (Satyanti, 2012). Karst formations rarely receive the conservatory attention required, especially in tropical regions, making them susceptible to destruction through human activities such as quarry mining (Radbouchoom et al., 2025; Tian et al., 2018). Due to this, 57% of *Begonia* species in Peninsular Malaysia have been classified as threatened, where 24 taxa are considered endangered, and 21 species have only been observed in one location (Chua, Kiew and Chan, 2009).

2.1.4 *Begonia* in Southeast Asia

Southeast Asia is a hotspot for biodiversity as its tropical climate is an ideal habitat for many plants to flourish in; therefore, it is no surprise that there is a large diversity of *Begonia* found (Dias, 2019). There are over 1,000 species of *Begonia* within it, which make up 20 sections (Hughes et al., 2025), with over 650 endemic to Malesia (Malaysia, Singapore, Indonesia, Brunei, the Philippines, Timor-Leste and Papua New Guinea). Many new species have been discovered in the region during the past few decades, often attributed to faster evolutionary rates due to high endemism and genetic bottlenecks (Ardi et al. 2022; Mahardika, 2016). Despite this, many areas within Southeast Asia are left unexplored due to the lack of resources, interest and manpower, leaving many species undiscovered (Ardi et al., 2018). Conservation is also a challenge in the region, with the onset of deforestation and quarrying being a threat to many microendemic *Begonia*

species, which are restricted to single river valleys or limestone cliffs (Radbouchoom et al., 2025).

2.2 Taxonomy of *Begonia*

2.2.1 Conventional Taxonomy

Taxonomy is the study of the classification and description of organisms (Padial et al., 2010). It bears great importance as it allows for easier and clearer communication of biological information. Traditional means of classification were based on morphology in order to differentiate between species (Haider, 2018). Augustin Pyramus de Candolle is presumed to be the first to integrate taxonomy within plant science based on disparities and similarities in morphology. This created a taxonomic classification for vascular plants, which became an important foundation for plant taxonomy (Briggs, 1991). In this system, which is used even today, plants were sorted into a hierarchical series of groups based on their shared morphological characteristics, which signify they share a common ancestry (Haider, 2018). The genus level is where plants start to be generally recognisable to be part of the same group, which is later further given their individuality at the species level. For large genera such as *Begonia*, sectional segregation is utilised for better organisation and management above the species level, as it is challenging to monograph genera with over 500 species,

especially by one person in their lifetime (Atkins et al., 2021). Historically, botanical illustrations served as visual documentation of plants (Hickman, Yates and Hopper, 2017). Despite the introduction of photography, they remain useful in highlighting distinctive structural characteristics with high attention to detail.

Multiple issues arise from the sole use of morphology for plant taxonomy with the main one being phenotypic plasticity. The phenotypic expression of plants is highly unstable as it is heavily reliant on environmental factors (Schlichting, 1986). Although this ability allows plants to better adapt to their environment, this creates taxonomic confusion and misidentifications, as different plant species may end up looking similar and vice versa. Convergent evolution is another factor that complicates morphological taxonomy as separate lineages that do not share a common ancestor could evolve similar structure or traits independently from each other, driven by similarity in environmental pressures and natural selection, causing parallel evolution where similar genetic mutation is observed in independent lineages (Stayton, 2015; Stern, 2013). Additionally, cryptic species, i.e., species that are indistinguishable by their phenotypic traits but belong to different species, can also occur and have been reported in plants and arthropods (Jörger and Schrödl, 2013).

2.2.1.1 Historical Taxonomy of *Begonia*

The history of *Begonia* taxonomy started off with the description of what would be a plant belonging to the genus in 1651 by Francisco Hernandez in Mexico (Doorenbos, Sosef, and Wilde, 1998). The genus name "*Begonia*" was officially coined by Plumier along with the description of six new species. This was then condensed into one species, *Begonia obliqua*, by Carl Linnaeus. Following this, the genus grew steadily with the introduction of more species. However, the genus was then split into three different genera, *Mezierea, Casparya*, and *Begonia*, due to the foresight that the original genus would be too extensive (Doorenbos, Sosef, and Wilde, 1998). Because of the vast number of species categorised under *Begonia*, the genus is segmented into sections. The framework that inspired the taxonomic structure used today originated in 1864 when Alphonse de Candolle segmented the genus *Begonia* into 61 sections (Moonlight et al., 2018). The separation into sections assists in the organisation of large genus such as *Begonia* as it resides between the genera and species level. Following this, a revision of the characteristics of the sections was published by Doorenbos, Sosef, and Wilde (1998) where 63 sections were recognised with their characteristics being mostly followed presently.

2.2.2 Molecular Taxonomy

With the advent of DNA sequencing, it is possible to compare genetic sequences to be able to distinguish between species on a molecular level (Alexander, Levett

and Turenne, 2015). The concept is simple where individuals of the same species would have highly similar or identical DNA, while distantly related individuals would be genetically different. Genetic data often offers a more objective stance in species delimitation, which is the process or method used to identify and define boundaries between species (Wiens, 2007). This is because DNA is intrinsic and less susceptible to environmental influences, e.g., the same *Begonia* species grown under different environmental conditions will retain the same DNA, even though its morphology may vary significantly due to plasticity.

Additionally, the genetic sequences of an organism will change through phenomena like mutation and hybridisation over long periods of time, which acts as a foundation for evolutionary divergence (Schleifer and Ludwig, 1994), which then allows scientists to identify how different species evolved by means of a phylogenetic tree. Furthermore, the inclusion of fossil or carbon dating information, if available, will generate chronograms, which allow scientists to estimate when species evolved (Bromham and Penny, 2003). However, the use of DNA data is not without drawbacks, which may include higher cost and expertise, DNA extraction difficulties in certain plant species, and contamination (Hollingsworth, Graham and Little, 2011; Lauri and Mariani, 2009). Today, DNA data is commonly used alongside morphological traits as complementary evidence to achieve a more comprehensive and robust taxonomy, which has since revolutionised plant systematics (Simpson, 2019).

2.2.2.1 DNA Markers and Amplification

A DNA marker (also known as a molecular or genetic marker) is a sequence of DNA where its location on the chromosome is known (Al-Samarai and Al-Kazaz, 2015). DNA markers can be found in the nucleus, mitochondrion or chloroplast and may or may not be protein-encoding (i.e., a gene). A DNA marker has different arrangements of alleles in different individuals or species that can be defined and be distinguished from one another. This makes it useful to analyse and compute evolutionary pathways as well as for DNA barcoding, a method for rapidly and accurately identifying species by analysing a short, standardised segment of DNA. Basically, the DNA of a sample is checked against a repository of DNA data to determine its identity based on similarity scores. DNA barcoding is especially valuable for low-cost studies requiring the identification of biological samples without involving taxonomic experts. Species identification using DNA barcoding only requires the use of one DNA marker, however, the use of multiple markers is more beneficial to infer on phylogenetic relationships due to the different evolutionary rates of different markers (Bidyananda et al., 2024). Additionally, as the identification is based on DNA, an incomplete sample (e.g., one insect leg) can also be used for species identification.

DNA barcoding is done using DNA barcodes, which are basically a DNA marker that is globally recognised and accepted (Kress and Erickson, 2008). A good plant DNA barcode must fulfil several criteria: (i) it must be present in all plants; (ii) it

must display sufficient genetic variation to confidently differentiate species, (iii) it should be easy to amplify using Polymerase Chain Reaction (PCR), (iv) it should be easy to DNA sequence, and (v) it should ideally form a consistent multiple sequence alignment block, i.e., DNA length the same across species (Hebert et al. 2003; Jinbo et al. 2006). A universal plant DNA barcode is hard to find and establish due to the slow mutation rate of plants, which makes it difficult to identify a sequence that can differentiate every plant species (Letsiou, 2024). The two most common DNA barcodes for plants are *mat*K and *rbc*L. However, these DNA barcodes, despite being able to identify a large portion of global plant species, were unable to differentiate closely related plant species such as those found in Southeast Asia (Hollingsworth, Graham and Little, 2011).

Genetic material is first extracted from the plants, where the process consists of breaking down the plant material, physically and enzymatically, followed by its precipitation and cleaning (Wang et al., 2021). Following this, PCR is done to amplify the desired sequences using DNA polymerase. The process has three different phases: denaturation, annealing, and extension (Khehra, Padda and Swift, 2023). In the first phase, DNA is heated up to split the double-stranded molecules by breaking the hydrogen bonds between the complementary base pairs. The annealing of DNA is the rapid cooling of the DNA for the specific primers, a short segment of single-stranded nucleic acid, which starts DNA synthesis of a specific region, to bind to their complementary sites. Lastly, the temperature is raised to the optimal temperature for the polymerase to bind and

elongate the DNA strand. DNA polymerase is the enzyme responsible for the duplication of DNA as it binds to the primers and extends the DNA sequence by adding in the complementary nucleotides (Garcia-Diaz and Bebenek, 2007). *Taq* DNA polymerase is the most-well known polymerase due to its high heat tolerance (Ishino and Ishino, 2014). However, minor errors in amplification, such as slip-strand mispairing, can be observed occasionally (McInerney, Adams and Hadi, 2014). In contrast to this, high-fidelity DNA polymerases, like Phusion™ DNA polymerase, have proofreading abilities, resulting in the amplification of higher quality DNA sequences (i.e., very few mistakes)usually used for cloning (Dolgova and Stukolova, 2017).

The success of amplification can be assessed using gel electrophoresis before the amplicons are sent for sequencing. Agarose gels are made with wells on one end to load the amplicons after being submerged in a buffer solution (Lee et al., 2012). An electric current is then sent through to make use of the negative charge of the DNA phosphate backbone, allowing the amplicons to migrate towards the positive anode. Within the 3D, complex network of the gel, the DNA is separated, where the rate of migration depends on the size of the amplicon, allowing for the estimation of its size (Lee et al., 2012). With the addition of dyes, e.g., GelRed® nucleic acid gel stain, the amplicons can be visualised under ultraviolet light, where a correctly amplified DNA marker should result in a single DNA band of expected size in gel electrophoresis (Huang, Baum and Fu, 2010).

2.2.2.2 Bioinformatics

Bioinformatics is the use of computer software that uses mathematical formulas to analyse and understand biological data, including DNA, RNA, protein, etc. (Luscombe, Greenbaum and Gerstein, 2001). Using genetic data, phylogenetic trees can be generated, which are then used for species identification and to infer on the relationship, also known as phylogeny or phylogenetic relationships, between the species and if they share a common ancestor (Hall, 2013). This is generally performed by aligning all DNA sequences together into a multiple sequence alignment (MSA) block, and scoring algorithms applied based on the similarity of one species to another (Bawono et al., 2016). During this process, the software also predicts insertion or deletion mutations (indels) to best estimate the evolutionary pathway of all the samples tested. Different models that take into account known codon mutation rates, usually for DNA markers that are genes, can often be applied for a more precise analysis (Bidyananda et al., 2024). Different algorithms can be used to estimate evolutionary pathways, including the maximum likelihood (ML) method, which identifies the tree that is most probable given the data, and the maximum parsimony (MP) method, which seeks the simplest tree requiring the fewest evolutionary changes (Zou et al., 2024). ML is a statistical model in which tree topology formed has the highest probability of producing the genetic data (Guindon et al., 2009). Bayesian inference (BI) is similar to ML in which it uses Markov chain Monte Carlo (MCMC) to estimate the posterior probability of the phylogenetic trees (Huelsenbeck and Ronquist,

2001). The thing that differs them is that priori information can be input into BI to generate phylogenetic trees which are better informed.

Ultimately, these software models allow the construction of phylogenetic trees based on similarity scores generated from multiple sequence alignment blocks (Bawono et al., 2016). These trees offer valuable information in phylogenetic inference, i.e., how different species evolve and how they are related to each other (Hall, 2013). Notably, phylogenetic trees can vary depending on the choice of DNA markers and the phylogenetic algorithms employed (Choi et al., 2019). Therefore, careful consideration and critical evaluation are essential during data analysis, and the resulting trees provide only an inference of evolutionary relationships rather than definitive conclusions. This further underscores the aforementioned importance of integrating genetic data with morphological observations to achieve more robust and reliable taxonomic conclusions.

Species delimitation can also be performed via bioinformatics, although it is traditionally done based on morphology or morphometrics, a method that is susceptible to errors and subjectivity due to morphological plasticity, convergent evolution and cryptic species (as discussed in Section 2.2.1). To address this, species delimitation based on genetic data offers a more objective approach, and can be done using different models, such as assemble species by automatic partitioning (ASAP), automatic barcode gap discovery (ABGD), Bayesian

Poisson tree processes (bPTP), and generalised mixed Yule Coalescent (GMYC). ASAP uses the pairwise distances to be input into a hierarchical clustering algorithm to produce species partitions, which are then scored (Puillandre, Brouillet and Achaz, 2020). ABGD utilises the pairwise distance between the species to identify the barcoding gap between intra- and interspecific distances to act as a threshold for partitioning (Puillandre et al., 2012). Both bPTP and GMYC require the input of a phylogenetic tree, while ASAP and ABGD only require aligned species from a single locus. bPTP predicts the number of species based on the number of substitutions on the branches of the Bayesian phylogenetic tree (Zhang et al., 2013). GMYC models branching events between and within species of an ultrametric tree using a Yule process and a neutral coalescent process, respectively (Pons et al., 2006). In summary, these four species delimitation models use different algorithms to identify "species" based on genetic data, and when all four models produce consistent delimitation patterns, the DNA data (or DNA marker) can be considered more reliable.

2.2.2.3 Molecular Taxonomic Framework of *Begonia*

Challenges in morphological identification and the limited phylogenetic resolution of the standard plant DNA barcodes *mat*K and *rbc*L for the highly diverse southeastern *Begonia* have driven the development of alternative DNA markers (Ardi et al., 2022). In early *Begonia* studies, the chloroplast markers *trnC-trnD* and *trnL* introns were used; however, this led to partially resolved

phylogenies (Badcock, 1998; Plana, 2003). Thomas et al. (2011) were able to use three chloroplast markers (*ndh*A intron, *ndh*F–*rpl*32 and *rpl*32–*trn*L) to revise a relatively resolved and strongly supported phylogenetic framework for a selection of Asian *Begonia*. The markers, designed by Shaw et al. (2007), demonstrated high levels of variability and phylogenetic utility, making them the best choice for low-level molecular studies. These are non-coding spacers found on the small-single copy (SSC) region of chloroplast DNA (Figure 2.1).

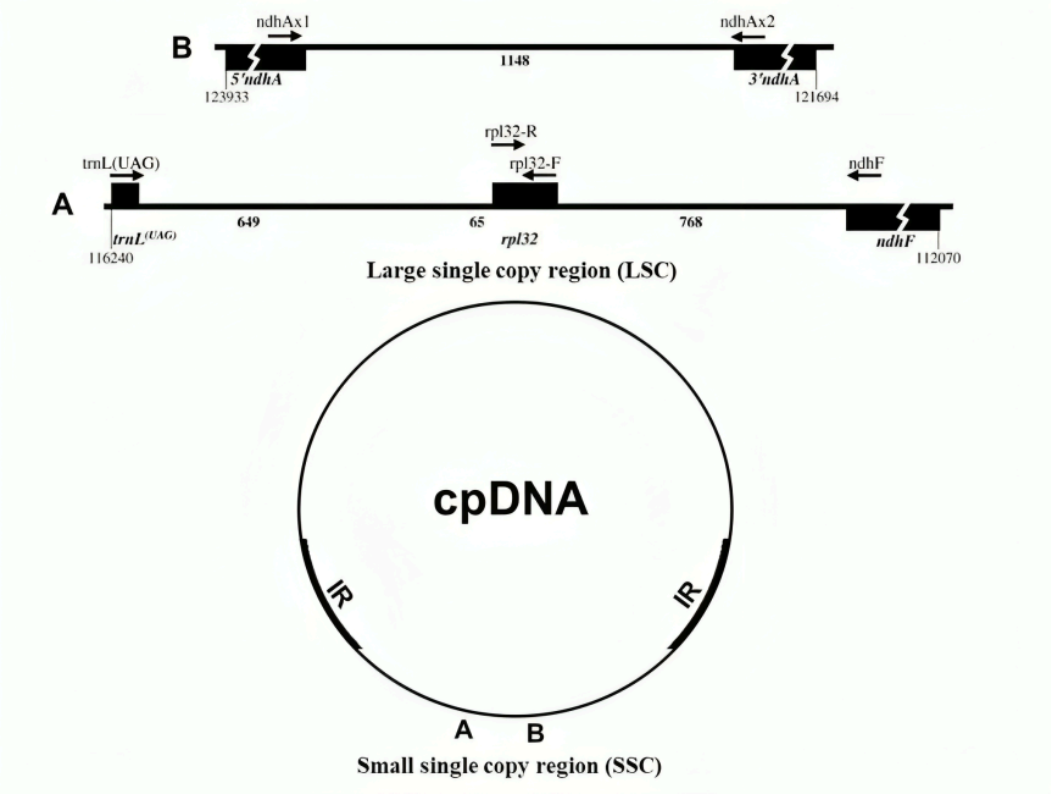


Figure 2.1. Location of *ndh*A intron, *ndh*F–*rpl*32 and *rpl*32–*trn*L region on the SSC region of chloroplast DNA (Shaw et al., 2007).

Moonlight et al. (2018) published the first classification of *Begonia* based on chloroplast markers (*ndh*A intron, *ndhF–rpl32*, and *rpl*32–*trn*L) where they sampled 574 species which spanned 57 of the 68 accepted sections at that time in hopes to produce a well-resolved phylogeny of *Begonia* as well as re-circumscribe the sectional classification of genus. These findings allowed for the erection of five new sections (*Begonia* sect. *Astrothrix* Moonlight, *Begonia* sect. *Ephemera* Moonlight, *Begonia* sect. *Jackia* M. Hughes, *Begonia* sect. *Kollmannia* Moonlight, and *Begonia* sect. *Stellandrae* Moonlight), bringing the number up to the 70 sections we have today. Since then, a more recent effort was made to reclassify the whole genus by Shui et al. (2019). Using the entire chloroplast genome as a base, the authors sorted the genus into 14 subgenera and 48 sections. However, this came with limitations such as the lumping of many species under a section, making it impractical, especially for regions with high *Begonia* diversity yet to be described, as well as the difficulty and higher cost of sequencing (Ardi et al., 2022; Hollingsworth, Graham and Little, 2011). To resolve these taxonomic conflicts, Ardi et al. (2022) analysed the phylogenetic incongruence based on nuclear, chloroplast and mitochondrial genomes to suggest a framework for future taxonomic development. They advised against the use of genetic data from only one genome and proposed that a combination of the ITS region, *ndh*A intron, *ndh*F–*rpl*32 spacer, and *rpl*32–*trn*L spacer is sufficient for species identification and sectional placement of *Begonia* due to their relative ease, convenience, and amount of already available data.

ITS is a popular choice for a non-coding nuclear DNA marker within plant molecular systematics due to its high amounts of variation and easy amplification. It mainly consists of two subregions, ITS1 and ITS2, but can also include the 5.8S loci located between the two spacers (Cheng et al., 2016). Despite its advantages, it has a few drawbacks keeping it from being a universal plant barcode. Namely, paralogues from the incomplete concerted evolution of the ITS region. Besides this, fungal contamination is also an issue, as ITS from fungi can also be amplified. Therefore, it is suggested that ITS be used in addition to other markers (Letsiou, 2024). In the context of *Begonia*, 26S and ITS were used in molecular systematics as early as 2003 (Forrest and Hollingsworth, 2003). A study by Cheng et al. (2016) designed new sets of primers to amplify the ITS region specifically in plants where the ITS-u1/ITS-u4 primers had the highest overall PCR success of 97.5% (Figure 2.2).

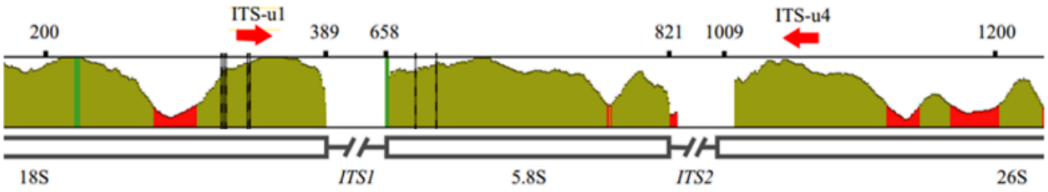


Figure 2.2. Location of the ITS-u1/ITS-u4 primers on the nuclear ribosomal DNA (Cheng et al., 2016).

2.2.3 *Begonia* Sect. *Jackia*

Begonia section *Jackia* was formed to house rhizomatous species within Malesia, which were previously classified under *Begonia* sect. *Reichenhemia*, to separate them from tuberous species (Moonlight et al., 2018). In addition to this morphological support, *Begonia* sect. *Jackia* is also molecularly supported as the section formed a strongly supported monophyletic clade distinct from other sections (Moonlight et al., 2018). Currently, the section comprises 76 accepted species (Hughes et al., 2015). *Begonia* sect. *Reichenhemia* and *Begonia* sect. *Platycentrum*, both of which can be found in Malaysia, have been seen to be closely related to sect. *Jackia* (Kono et al., 2023). The herbaceous characteristics belonging to species under this section include rhizomes which are glabrous or hairy and asymmetric leaves with palmate venation (Moonlight et al., 2018). Their inflorescence is axillary, where both male and female flowers have 2–4 free tepals. The anthers grow in an obovate structure, while their ovaries have three locules and an axile placenta.

Begonias under the section *Jackia* in Peninsular Malaysia are poorly documented and mostly only limited to morphological information. The phylogenetic tree constructed by Moonlight et al. (2018) was found to be incongruent with respect to *B. rajah* and *B. nurii*, two species that were inferred to be paraphyletic, likely a result of misidentification. Notably, numerous *Begonia* spp. extant in Peninsular Malaysia, i.e., *B. forbesii, B. yappii, B. yenyeniae*, etc., were also missing in that

study. Another study conducted by Tan, Tam and Kiew (2018) used only the *ndh*F–*rpl*32 spacer for phylogenetic inference of begonias in Peninsular Malaysia, which resulted in a poorly resolved and inconclusive tree. This corroborates Ardi et al. (2022) that more than one marker is needed for robust phylogenetic resolution of the begonias in this region. This highlights a significant gap in genetic information of Peninsular Malaysia *Begonia* species, as well as the effectiveness of the proposed genetic markers (ITS, n*dh*A intron, *ndh*F–*rpl*32 spacer, and *rpl*32–*trn*L spacer) in their phylogenetic resolution.

2.2.3.1 *Begonia rajah*

Begona rajah is a popular cultivated species belonging to *Begonia* sect. *Jackia* (Tan, Tam and Kiew, 2018). In 1892, the plant was first founded in the state of Terengganu before being brought to England, where it won a "First Class Certificate" from the Royal Horticultural Society and was featured in multiple magazines, gaining popularity in 1894 (Kiew, 2005). Ever since its discovery over 100 years ago, the species has never been rediscovered in the wild, causing it to earn its "extinct-in-the-wild" status. It was not until findings from Ho (2024) suggested that specimens resembling *B. rajah* from the wild were conspecific to cultured *B. rajah* based on foliar morphology and molecular data. However, the status of *B. rajah* could not able to be confirmed due to the lack of floral morphology from a fully grown sample.

The leaves of *B. rajah* are bullate with a bronzy-green colour between its green palmate veins. They grow in an asymmetric heart shape, where the margins are scalloped and ciliated. Its male flower has four white tepals, two of which are rotund-shaped while the other two are smaller and obovate. The female flower has three pale pink tepals and a whitish-green ovary with three locules and an axile placenta (Kiew, 2005).

2.3 Identifying and Understanding *Begonia* Phylogeny

2.3.1 Proper Documentation of *Begonia*

Improved identification and phylogeny of *Begonia* can help fill in the prevailing gap of proper documentation of the species within Malaysia. Species delimitation is required to avoid under- or overestimating the number of species found within Peninsular Malaysia through the lumping of cryptic species or the misinterpretation of environmentally affected variation as distinct (Bland et al., 2017). Genetic data generated through phylogenetic studies can also help supplement or act as the groundwork for future molecular studies within the genus (Tan, Tam and Kiew, 2018). Ultimately, the characterisation of *Begonia* biodiversity in Peninsular Malaysia will help in downstream applications such as future conservation and bioprospecting efforts (Crozier, 1997).

2.3.2 Conservation of *Begonia*

Conservation is crucial to *Begonia*, a genus that is highly endemic, with many species being limited to reside in single habitats prone to disturbances from anthropogenic activities (Tian et al., 2018). Therefore, accurate species identification along with well-resolved phylogenies can bring about urgent attention for the need of conservation efforts on the regional and species-specific level, according to the priority of extinction risk and evolutionary distinctiveness. *Ex situ* conservation can also be aided with the help of phylogenetic research, e.g., the addition of endangered *Begonia* into seed banks as well as the preservation of living collections in order to preserve their genetic diversity (Larkin et al., 2016).

2.3.3 Bioprospecting

Besides species identification and conservation of *Begonia*, phylogenetic studies can highlight evolutionary patterns of traits that are horticulturally and medicinally important (Thompson and Hawkins, 2024). Understanding the phylogenetic relationship between species can help identify the lineages with potential species that are rich in beneficial compounds such as terpenoids, flavonoids, anthocyanins, and many more (Borah et al., 2025). This is possible as closely related species often share secondary metabolites. This links bioprospecting with conservation to ensure that potentially useful resources are not lost to habitat destruction.

CHAPTER 3

METHODOLOGY

3.1 Sample Collection

The *Begonia* samples were collected from Peninsular Malaysia with proper permission, morphologically identified and supplied by *Begonia* taxonomists Ang Y. P. and Saw Y. Y. (Table 3.1). These samples consisted of taxonomically important species, namely *B. foxworthyi*, *B. rajah*, and *B. yappii*, or potentially new species. The samples were provided either in dried form (preserved in silica gel) or as fresh, viable leaf cuttings. The dried samples were kept in a dry and cool location in D203, Faculty of Science, UTAR, while the fresh samples (C9BMR, BGCK1, and BGCN1) were cultured in the Faculty of Science horticultural shed in plastic containers under artificial lighting.

Table 3.1: Samples of *Begonia* collected and used in this study.

No.	Voucher	Species	Location	Remark
1	25YC_BFOX6	*B. foxworthyi*	Sekayu, Terengganu	This study
2	25YC_BFOX9	*B. foxworthyi*	Chiku, Kelantan	
3	25YC_BFOX10	*B. foxworthyi*	Sungai Betis, Kelantan	
4	25YC_BYAP1	*B. yappii*	Kuala Airing, Kelantan	
5	25YC_BSP_KT1	*Begonia* sp.	Kelantan-Terengganu	
6	25YC_BSP_STNG1	*Begonia* sp.	Gunung Stong, Kelantan	
7	C9BMR	*B. rajah*	Kuala Terengganu, Terengganu	
8	BGCK1	*B*. sp. ‘Chiku 4/5’	Chiku, Kelantan	Ho, 2024
9	BGCN1	*B*. sp. ‘Creeping Nurii’	Galas, Kelantan	

3.2 Morphological Analysis

Plant habit was provided by Ang Y. P. Morphological identification, when applicable, was done based on the taxonomic keys by Doorenbos, Sosef, and Wilde (1998) and Tan, Tam and Kiew (2018). A live *ex situ Begonia rajah* (C9BMR) was characterised based on external morphology (i.e., stem and foliage) and microscopic examination of inflorescence using a stereomicroscope (Motic, China). This species was specifically investigated as a follow-up to Ho (2024), who identified sample C9BMR as *B. rajah* using only DNA data. All

photos were taken using a Samsung Galaxy S22 smartphone. Botanical illustrations were drawn on a Samsung Galaxy Tab S6 Lite.

3.3 DNA Extraction, Amplification, and Sequencing

3.3.1 DNA Extraction and Purification

DNA extraction was performed using Bio Basic Pte. Ltd. (Singapore) EZ-10 Column Plant Genomic DNA Purification Kit according to manufacturer's protocols. Approximately 20 mg of each sample was aseptically excised into smaller pieces using sterile forceps and placed into individual 1.5 mL microcentrifuge tubes. For each sample, 600 µL of PCB buffer was added before grinding the samples using a micropestle. After no large pieces remained, 12 µL of β-mercaptoethanol was added, and the samples were vortexed. Next, the samples are incubated at 65℃ for 30 minutes. Subsequently, 600 µL of chloroform was added prior to mixing via inversion. The samples were then centrifuged at 12,000 × g for 5 minutes. To a new 1.5 mL microcentrifuge tube, around 400 µL of the supernatant was extracted and transferred along with 200 µL of BD buffer and 200 µL of 100% ethanol. The solution was transferred into an EZ-10 column to be centrifuged at 9,000× g for 2 minutes. Next, the flow-through was disposed of, and 500 µL of PW solution was added. The samples were then centrifuged again using the same parameters, along with the discarding of the flow-through. After that, 500 µL of wash solution was added

before centrifugation. The spin column was moved to a new microcentrifuge tube to be eluted with 30 µL of TE buffer preheated to 60℃. After 30 minutes, the tubes were centrifuged again with the same parameters. Using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, USA), the DNA concentration and quality of the extracted samples were measured. The DNA of all six samples was then stored at -20℃ pending DNA amplification.

3.3.2 Polymerase Chain Reaction (PCR)

For each sample, four DNA markers were amplified. This consisted of one nuclear marker (ITS) and three chloroplast markers (*ndh*A intron, *ndh*F–*rpl*32 and *rpl*32–*trn*L) (Table 3.2). The *Taq* DNA polymerase was used to amplify the ITS and *rpl*32–*trn*L markers, while Phusion™ DNA polymerase was used for the markers *ndhA* intron and *ndh*F–*rpl*32. Phusion™ DNA polymerase has proofreading ability and can reduce errors such as slipped-strand mispairing (McInerney, Adams and Hadi, 2014). Its use for only two out of the four markers was based on preliminary amplification and sequencing showing that (i) it has no impact on the sequencing quality of ITS and *rpl*32–*trn*L, and (ii) it is more cost-efficient.

Table 3.2: The type of molecular markers used and their characteristics.

Origin	Region	Primer	Sequence	Length (bp)	Reference
Nuclear DNA	ITS	ITS-u1	5’ -GGAAGKARAAGTCGTAACAAGG- 3’	968-1160	Cheng et al., 2016
		ITS-u4	5’- RGTTTCTTTTCCTCCGCTTA- 3’		
Chloroplast DNA	*ndh*A intron	Beg_ndhAx1	5’ -GCYCAATCWATTAGTTATGAAATACC- 3’	729-1254	
		Beg_ndhAx2	5’ -GGTTGACGCCAMARATTCCA- 3’		
	*ndh*F–*rpl*32	Beg_ndhFBeg-F	5’ -TGGATGTGAAAGACATATTTTGCT- 3’	617-873	Shaw et al., 2007
		Beg_trnLBeg-R	5’- TTTGAAAAGGGTCAGTTAATAACAA- 3’		
	*rpl*32–*trn*L	Beg_rpL32-F	5’- CTGCTTCCTAAGAGCAGCGT- 3’	543-1417	
		Beg_trnL	5’- CAGTTCCAAAAAAACGTACTTC -3’		

Six Begonia samples extracted from this study and three samples (C9BMR, BGCK1, and BGCN1) from Ho (2024) were subjected to PCR amplification. The components required for the amplification of one PCR sample are listed in Table 3.3. PCR for all the samples was conducted using the FlexCycler2 Thermal Cycler (Analytik Jena, Germany).

Table 3.3: Reagents for one PCR using *Taq* DNA polymerase (ITS and *rpl*32–*trn*L) and Phusion™ DNA polymerase (*ndh*A intron and *ndh*F–*rpl*32).

Reagents	**Volume (µL)**	
	***Taq* DNA polymerase**	**Phusion™ DNA polymerase**
H_2O	15.75	11.8
10× *Taq* / HF buffer	2	4
dNTP	2	0.4
MgCl	1	0
Forward primer	1	1
Reverse primer	1	1
DMSO	0	0.6
Taq / Phusion™ DNA polymerase	0.25	0.2
DNA template	2	1
Total	25	20

The PCR temperature profile for ITS includes: template denaturation at 94℃ for 4 minutes, 34 cycles of denaturation at 94℃ for 30 seconds, primer annealing at 55℃ for 40 seconds, primer extension at 72℃ for 1 minute, and ending with 10 minutes at 72℃ for the final extension (Cheng et al., 2016). The PCR temperature profile for *ndh*A intron and *ndh*F–*rpl*32 includes: template denaturation at 95℃ for 30 seconds, 35 cycles of denaturation at 98℃ for 30 seconds, primer annealing at 50℃ for 30 seconds, and primer extension at 72℃ for 30 seconds, followed by 5 minutes at 72℃ for the final extension (Moonlight, 2015). Lastly, the PCR temperature profile for *rpl*32–*trn*L includes: template denaturation at 80℃ for 5 minutes, 30 cycles of denaturation at 95℃ for 1 minute, primer annealing at 50℃ for 1 minute, a ramp of 0.3 °C/s to 65 °C and primer extension at 65℃ for 4 minutes, and finally, final extension at 72℃ for 10 minutes (Thomas et al., 2011). The PCR temperature profiles for all the markers stated are summarised in Table 3.4. All amplicons were stored at -20℃ pending gel electrophoresis.

Table 3.4: Parameters for the PCR of each molecular marker.

Region	Template denaturation	Denaturation	Annealing	Ramp	Extension	Cycles	Final extension	Reference
ITS	94°C 4 minutes	94°C 30 seconds	55°C 1 minute	-	72°C 1 minute	30	72°C 10 minutes	(Cheng et al., 2016)
***ndh*A intron** ***ndh*F–*rpl*32**	95°C 30 seconds	98°C 30 seconds	50°C 30 seconds	-	72°C 30 seconds	35	72°C 10 minutes	(Moonlight et al., 2015)
***rpl*32–*trn*L**	80°C 5 minutes	95°C 1 minute	50°C 1 minute	0.3°C/s to 65°C	65°C 4 minutes	30	72°C 10 minutes	(Thomas et al., 2011)

3.3.3 Gel Electrophoresis

For gel electrophoresis, a 1% agarose gel was prepared by adding 0.6 g of agarose powder to 60 mL of 1× TAE buffer. The mixture was boiled using a microwave and was swirled until homogeneous. Into the molten agar, 4 µL of GelRed® nucleic acid gel stain (10,000× water) was added before the gel was cast. Following this, the gel was submerged in 1× TAE buffer within the gel box. For each amplicon, 2 µL of 6× Tri-Color loading dye (1st BASE, Singapore) was added to 3 µL of the sample. The mixture was then aliquoted into a well in the gel. Besides this, 1 µL of GeneRuler 1 kb DNA ladder (Thermo Fisher Scientific, USA) was mixed with 2 µL of 6× Tri-Color loading dye to be used as a scale for the size of the DNA fragments of the amplicons. Electrophoresis was conducted using the EnduroTM Gel XL Electrophoresis System (Labnet International, USA) at 350 mA and 110 volts for 11 minutes. Once completed, the gel was imaged using a Gel Doc XR+ Imaging System (Bio-Rad Laboratories Inc., USA) in order to visualise the DNA fragments under ultraviolet light.

3.3.4 DNA Sequencing

The amplified DNA and their respective primer pairs were sent to Bio Basic Asia Pacific Pte. Ltd. (Singapore) to be sequenced using Sanger sequencing.

3.4 Genetic Analysis

3.4.1 DNA Sequence Processing

ChromasPro V1.5 (Technelysium Pty. Ltd., Australia) was used to assemble and edit each DNA contig, during which disambiguites were corrected based on software computation and visual inspection. For each marker, all assembled DNA contigs were compiled in FASTA format prior to being subjected to Multiple Sequence Alignment (MSA) using Clustal X V2.0 (Larkin et al., 2007). Long DNA sequences were truncated to form an MSA block using BioEdit V7.0.9.0 (Hall, 1999).

3.4.2 BLASTN and Incorporation of GenBank Sequences

Each assembled contig was inputted into BLASTN in order to search for the best hits or closest species within GenBank. To be selected for use, the contigs were required to fulfil the criteria that it should be from a published journal article, have the highest percentage identity, have approximately 100% query coverage, and be verified against the latest taxonomic articles to identify changes/updates on nomenclature. For each DNA marker, these sequences were then downloaded and incorporated into their corresponding MSA block for subsequent analyses. Additionally, DNA sequences of *B. versicolor* and *B. grandis* were downloaded to serve as the outgroup, given their close phylogenetic relationship to the section

Jackia ingroup (Moonlight et al., 2018). Each MSA block was checked for indels or nucleotide discrepancies.

Finally, four individual MSA blocks were generated. Two additional combined (or concatenated) datasets were generated, i.e., using only three DNA markers from the chloroplast (i.e., *ndh*A intron + *ndh*F–*rpl*32 + *rpl*32–*trn*L; herein referred to as chloroplast markers), and using all four markers (i.e., ITS + *ndh*A intron + *ndh*F–*rpl*32 + *rpl*32–*trn*L; herein referred to as concatenated markers). Concatenated datasets are used as they generally produce phylogenetic trees with better resolution because of the larger amount of genetic information.

3.4.3 Phylogenetic Analysis and Genetic Distance

The finalised MSA blocks for each marker were input into ALTER (Glez-Peña, 2010) to convert them into PHYLIP and NEXUS formats to construct phylogenetic trees using maximum likelihood (ML) and Bayesian inference (BI) algorithms using IQTREE web server and MrBayes v3.2.6, respectively (Trifinopoulos et al., 2016; Ronquist et al., 2012). These trees were based on individual markers, chloroplast markers (*ndh*A intron + *ndh*F–*rpl*32 + *rpl*32–*trn*L) and concatenated markers (ITS + *ndh*A intron + *ndh*F–*rpl*32 + *rpl*32–*trn*L). The number of bootstraps for ML was set to 2,000 while the parameters for BI included 4 independent MCMC (Markov chain Monte Carlo) chains, 2 million generations, and sampling at each 500 generations, where the

first 20,000 samples were discarded as burn-in (Harrison, Harrison, and Kidner, 2016). The resulting trees were viewed and annotated using FigTree V1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/) and Microsoft PowerPoint.

Genetic p-distance matrices were calculated for individual markers, chloroplast, and concatenated markers using PAUP* V.4.0b10 (Swofford, 2001).

3.4.4 Species Delimitation

Species delimitation is a process of determining whether specimens are conspecific (Wiens, 2007). This is generally performed as a more objective and reliable approach than species delineation based solely on morphology or phylogenetic trees, especially for species that do not conform to the biological species concept. Species delimitation was performed using the MSA blocks generated from Section 3.4.2, but without the GenBank and outgroups. To identify the number of "species" present for all individual and concatenated datasets, four common species delimitation algorithms were used, i.e., assemble species by automatic partitioning (ASAP), automatic barcode gap discovery (ABGD), Bayesian Poisson tree processes (bPTP), and generalised mixed Yule Coalescent (GMYC). The identification of "species" by ASAP and ABGD is based on genetic distance matrices and the presence of a "barcoding gap". On the other hand, species delimitation by bPTP and GMYC is based on consensus non-ultrametric and ultrametric phylogenetic trees, respectively. Using default

parameters for all algorithms, ASAP was performed using ASAPy 1.0 (Puillandre, Brouillet and Achaz, 2021) while ABGD was done with ABGDpy 0.1 (Puillandre et. al., 2012). BEAST v.2.7.6 (Bouckaert et al., 2019) was used for GMYC, where the inference of the ultrametric gene trees was done with a relax molecular clock model and Yule prior. A total of 50 million generations were produced, with trees being saved every 1,000 generations and the burn-ins discarded using tracers. The webserver at https://species.h-its.org/ptp/ was then used to conduct PTP and GMYC analyses (Pons et. al., 2006; Zhang et. al., 2013). The results were then annotated using Microsoft PowerPoint, and the species delimitation ability of every marker was assessed against that of the concatenated dataset (i.e., ITS + *ndh*A intron + *ndh*F–*rpl*32 + *rpl*32–*trn*L), which is considered to be the most accurate.

3.5 Marker Assessment

All four markers (ITS, *ndh*A intron, *ndh*F–*rpl*32, and *rpl*32–*trn*L) were assessed based on their PCR performance and genetic attributes to determine their suitability as a DNA barcode for *Begonia* species within section *Jackia*. PCR performance includes the ease of amplification, success rate during the first run, and the requirement of Phusion™ DNA polymerase. Marker performance was based on the number of phylogenetically-informative characters, the barcoding gap, i.e., the range between intra- and interspecific genetic variance, length of MSA block and other genetic parameters.

CHAPTER 4

RESULTS

4.1 Morphological Observations

4.1.1 Gross Morphology

The morphology of the samples used in this study was briefly described based on *in situ* images provided by the collaborators.

4.1.1.1 Morphology of 25YC_BFOX6

The leaves of 25YC_BFOX6 are asymmetrical with a slightly oblique form as their diameter spans around 12.4 cm. The margins of the leaves are slightly undulate where their tip is acute, and the base is cordate and minimally overlapped. While its adaxial surface is consistently shiny and pale green throughout, the abaxial surface is lighter with a yellow hue. The venation is palmate, where slight bullation can be seen between the veins. The petiole of which is dark red while being covered by fine hair.

4.1.1.2 Morphology of 25YC_BFOX9

The asymmetric leaves of 25YC_BFOX9 span approximately 10 cm in diameter. They have an orbicular to reniform shape where their basal lobes were cordate and overlapped, while they had an eccentric acute leaf tip (Figure 4.1). Their veins were palmate with bullation between them. The adaxial surface takes on a dark bronze red colour when young, which later takes on a dark green colour. However, they have a pale pink underside. The leaves are ciliated throughout as well as along their slightly crenate margins. Besides this, their petioles are red and hairy. Their inflorescences have a dichasial cyme arrangement where red stalks with sparse fine hairs carry monoecious flowers. The male flowers have two white orbicular tepals and an elliptical cluster of stamens. The female flowers have the same tepals but with 3 styles and stigmas, which spiral. Their ovaries are yellowy green where they have three locules with a placenta, each of that are axile.

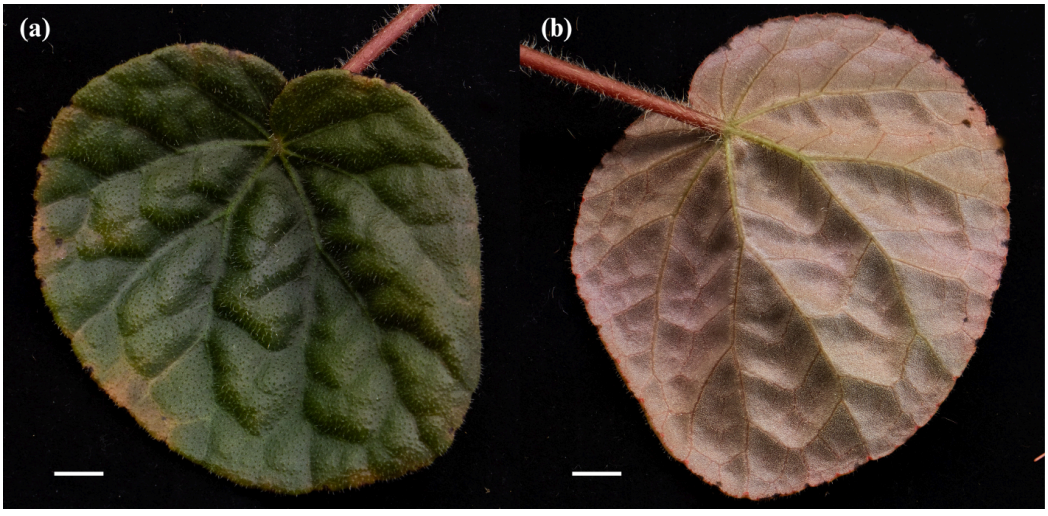


Figure 4.1. Foliar morphology of 25YC_BFOX9: **(a)** top view of mature leaf, **(b)** bottom view of mature leaf. (photo by Ang, Y.P.) Scale bar: 1 cm.

4.1.1.3 Morphology of 25YC_BFOX10

25YC_BFOX10 has leaves that are around 11.5 cm in diameter. They are asymmetrical with a slightly oblique form (Figure 4.2). They have an acute leaf tip and an overlapped cordate base, while their leaf margins are slightly undulate. The adaxial surface is light green and is ciliated all over and at the edge of the leaf, while the abaxial surface is paler. Bullation can be seen between its palmate veins. The petiole of the leaf is light brown and hairy.

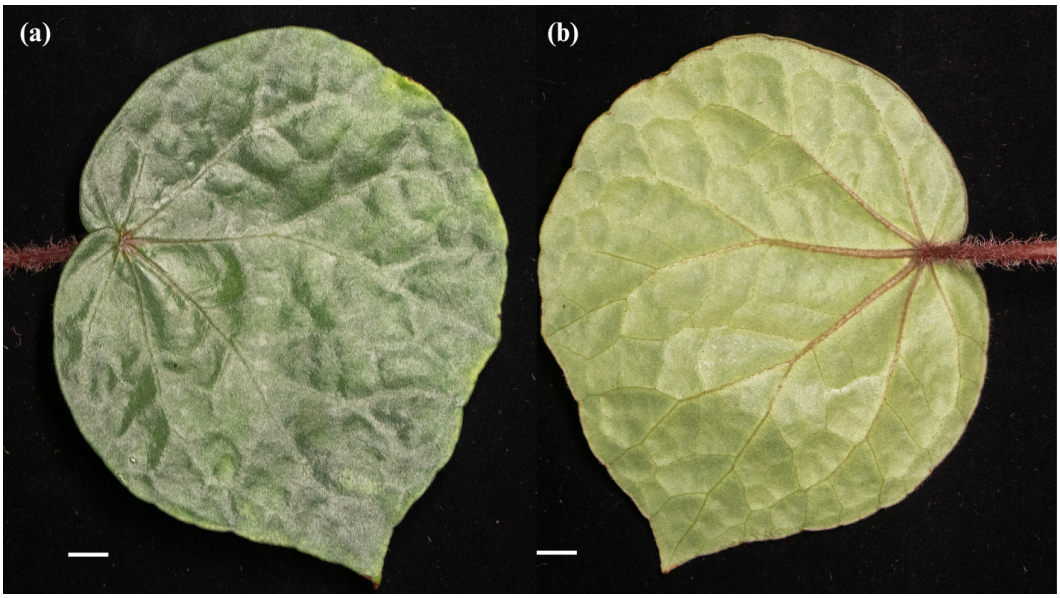


Figure 4.2. Foliar morphology of 25YC_BFOX10: **(a)** top view of mature leaf, **(b)** bottom view of mature leaf. (photo by Ang, Y.P.) Scale bar: 1 cm.

4.1.1.4 Morphology of 25YC_BSP_KT1

The leaves of 25YC_BSP_KT1 are orbicular to reniform in shape with a diameter of around 14 cm (Figure 4.3). They have a cordate base with a slightly crenated edge, which is ciliated. The adaxial surface of the leaves is green and bronzy green between its palmate veins. The leaves are smooth with no bullation and have a hairy dark red petiole.

Figure 4.3. 25YC_BSP_KT1 in its natural habitat. (photo by Ang, Y.P.)

4.1.1.5 Morphology of BGCK1

The leaves of BGCK1 are smaller than the rest with an approximate diameter of 6 cm. The slightly oblique leaves have a cordate and overlapped base, where its leaf tip is eccentric and acute. The margins of the leaves are crenated and ciliated. They have palmate venation with slight interveinal bullation. The adaxial surface of their leaves is a shiny bronzy green with green veins. They also have a petiole, which is dark red in colour and hairy. The floral characteristics of the sample are monoecious with a dichasial cyme arrangement. The red peduncles with sparse fine hairs bear inflorescence with two to four white tepals which were ovate orbicular or elliptic. The stamens were in globous clusters in the male flower. The female flowers have yellowy green ovaries with three locules and placentas that are axile.

4.1.2 Morphology of C9BMR

The gross morphology of C9BMR is shown in Figure 4.4, followed by a botanical illustration of it (Figure 4.5). The stem of C9BMR is rhizomatous and around 6 mm thick. It is yellowy green in colour, which grows pink stipules that have a lanceolate shape. Its 9×7.5 cm succulent leaves take on a slightly oblique form with an overlapped cordate base and a slight acuminate leaf tip. The lead edges are ciliated and scalloped, especially towards its apex. Bullation can be seen between their palmate venations. The adaxial surface is shiny and a dark bronze colour with green veins. On the contrary, the abaxial side is paler. Younger shoots

are smooth with no bullation and have a brownish pink colour with green veins and a redder underside. The petiole is yellowy green but pink when young. The inflorescence is axillary with a cyme arrangement. Its pink peduncles carry monoecious flowers. The male flower had four white tepals, two ovate orbicular outer tepals, 8×8 mm, and two inner obovate tepals, 7×3 mm. It has many stamens which take shape as a globous cluster with a 2 mm diameter. The female flower is white to pale pink with three tepals, two larger ovate tepals, 8×5 mm, and one smaller linear-shaped tepal, 6×1 mm, in between. Its ovary is pale pink with three wings spanning approximately 8 mm. It has three locules with an axile placenta each. It also has three styles and stigmas, 2 mm, which spiral.

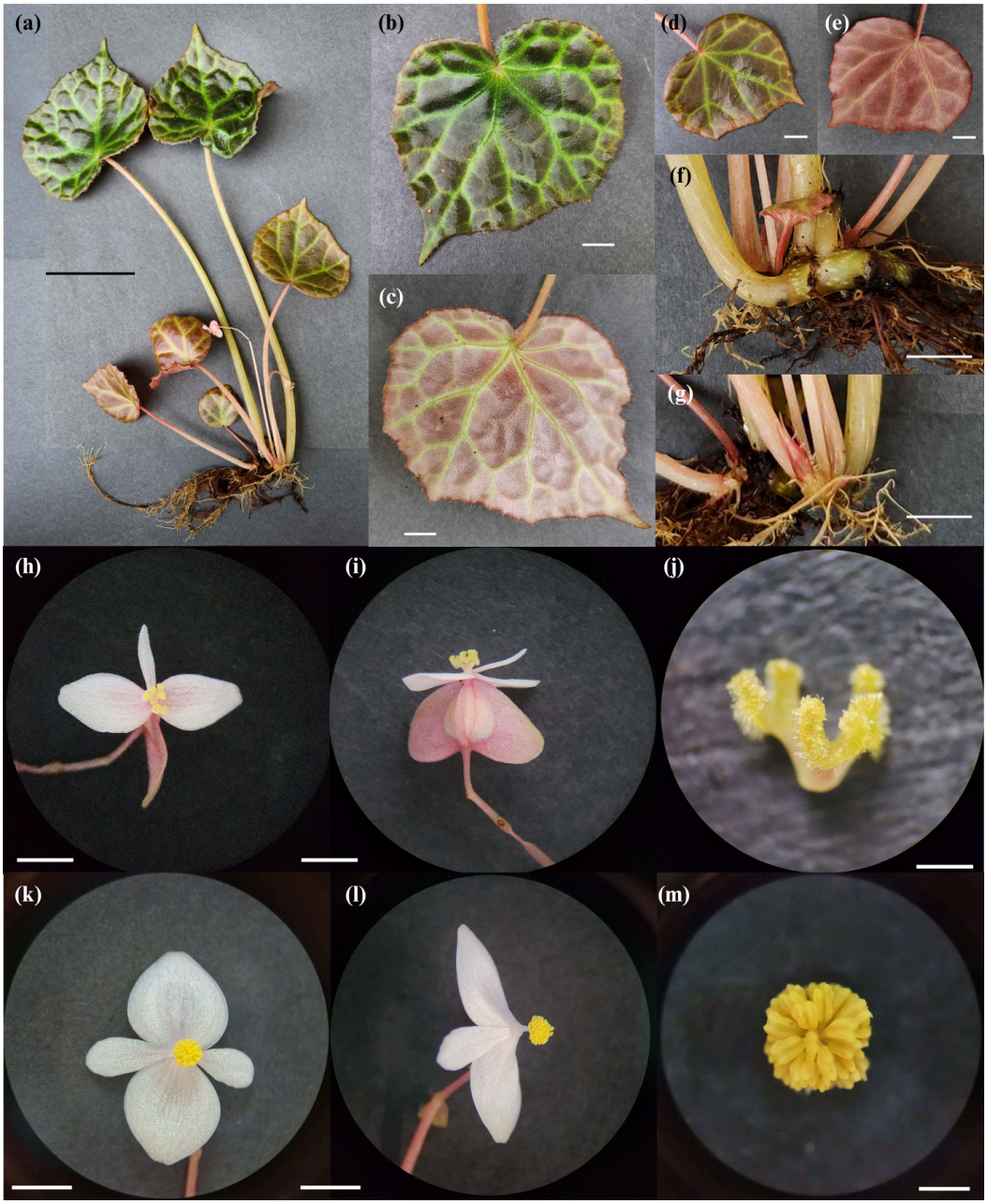


Figure 4.4. Gross morphology of C9BMR: **(a)** front view of the entire plant, **(b)** top view of mature leaf, **(c)** bottom view of mature leaf, **(d)** top view of young leaf, **(e)** bottom view of young leaf, **(f)** rhizomatous stem, **(g)** stipule, inflorescence; male: **(h)** front view, **(i)** side view, and **(j)** stamen, female: **(k)** front view, **(l)** side view, and **(m)** stigma. Scale bars: **(a)** 5cm, **(b–g)** 1cm, **(h, i, k** and **l)** 5 mm, and **(j** and **m)** 1 mm.

Figure 4.5. Botanical illustration of C9BMR. Scale bar: 1 cm.

4.2 Molecular Analysis

4.2.1 DNA Extraction and Amplification

The DNA concentration and purity of the six extracted *Begonia* samples are shown in Table 4.1.

Table 4.1: The DNA concentrations and purity of all the extracted samples.

No.	Sample	DNA concentration (ng/µL)	DNA purity (260/280)
1	25YC_BFOX6	3.4	-3.14
2	25YC_BFOX9	3.8	-6.31
3	25YC_BFOX10	3.1	-2.97
4	25YC_BYAP1	2.9	-2.40
5	25YC_BSP_KT1	3.2	-3.65
6	25YC_BSP_STNG1	2.8	-2.15

Gel electrophoretograms showing the resulting DNA bands for ITS, *ndh*A intron, *ndh*F–*rpl*32, and *rpl*32–*trn*L amplicons are shown in Figure 4.6. ITS amplicon had bright bands with a size between 750–1,000 bp for all samples, except that of 25YC_BSP_STNG, which had a weaker band at 1,500 bp. Amplicons for *ndh*A intron had bright bands at 1,500 bp, except for 25YC_BFOX10 and

25YC_BSP_STNG had weak bands. The bands for *ndh*F–*rpl*32 amplicons had a size between 750–1,000 bp, which were not as bright as the rest, especially for 25YC_BFOX10, 25YC_BSP_STNG, and BGCN1. The amplicons for *rpl*32–*trn*L spacer had bright bands at 1,500 bp, except for 25YC_BFOX9.

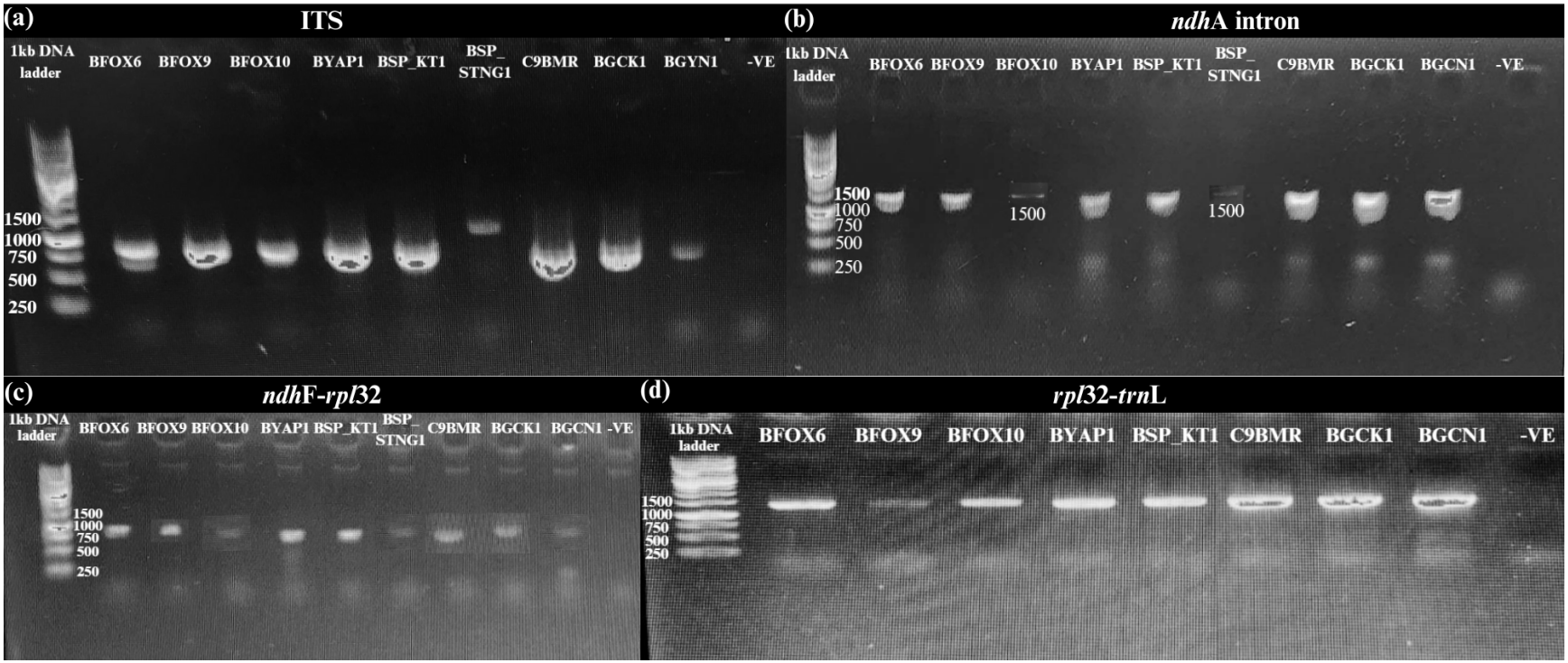


Figure 4.6. Gel electrophoretogram of amplified regions from the *Begonia* species studied: **(a)** ITS, **(b)** *ndh*A intron, **(c)** *ndh*F–*rpl*32, and **(d)** *rpl*32–*trn*L.

Of the 30 of PCR amplicons sent for sequencing (six for ITS and *rpl*32–*trn*L and nine for *ndh*A intron and *ndh*F–*rpl*32), *rpl*32–*trn*L had the most successful sequenced samples (100%) (Table 4.2). This is followed by ITS, *ndh*F–*rpl*32, and *ndh*A intron with a sequencing success rate of 83.3%, 77.8%, and 66.7%, respectively.

Table 4.2: The sequencing success rate of all the DNA markers.

Marker	**Sequencing success rate (%)**	**Species failed to be sequenced**
ITS	83.3	25YC_BSP_STNG1
ndhA intron	66.7	25YC_BFOX10, 25YC_BSP_STNG1, and BGCN1
*ndh*F–*rpl*32	77.8	25YC_BSP_STNG1 and BGCN1
*rpl*32–*trn*L	100	

4.2.2 Marker Performance

Details of the ease of amplification, multiple sequence alignment block length, DNA sequencing quality and genetic attributes of each genetic marker, the chloroplast markers and concatenated markers are shown in Table 4.3. The barcoding gaps found in the ASAP histogram of each marker is shown in Appendix A.

Table 4.3: Characteristics and performance of all four individual markers, chloroplast markers, and concatenated markers.

Marker(s)	Amplification success (first run)	Phusion™ polymerase requirement	MSA block length (nt)	No. of phylogenetically -informative characters	Pairwise identity (%)	Identical sites (%)	GC content (%)
Single							
ITS	100% amplified and sequenced	No	788	87	93.3	73.5	56.1
ndhA intron	78% amplified, difficult to amplify certain samples	Yes	1208	25	96.2	86.1	26.6
*ndh*F–*rpl*32	67% amplified, difficult to amplify certain samples	Yes	836	17	92.8	79.7	21.5
*rpl*32–*trn*L	89% amplified, can't amplify with Phusion™	No	1210	22	94.5	64.8	20.4
Concatenated							
*ndh*A intron + *ndh*F–*rpl*32+ *rpl*32–*trn*L	-	-	3254	65	94.4	76.2	23
ITS+ *ndh*A intron + *ndh*F–*rpl*32+ *rpl*32–*trn*L	-	-	4042	152	94.2	75.7	29.6

4.2.3 BLASTN Identification

The BLASTN results of the assembled DNA contigs for all four markers are listed in Tables 4.4 to 4.7. For ITS region, the BLASTN results identified all samples as *B. muricata* with query percentages ranging from 91% to 92% and percentage identity values of 91.48% to 92.73%. As for the *ndh*A intron, 25YC_BFOX6 was identified as *B. foxworthyi*, while the rest of the samples were either identified as *B. geogeonsis*, *B. forbesii*, or *B. muricata*. The BLASTN results for *ndh*F–*rpl*32 spacer saw 25YC_BFOX6 being identified as *B. foxworthyi* and C9BMR identified as *B. rajah*. The rest of the samples for the marker were identified as either *B. sublobata, B. ignorata*, or *B. rajah*. The *rpl*32–*trn*L spacer identified 25YC_BFOX10 as *B. foxworthyi*, while the others were identified as either *B. ignorata* or *B. nurii*. The identity of 25YC_BSP_KT1 is inconclusive, with a percentage identity of only 92.73–99.18% over a query cover of 92–100%. However, the sample was more closely related to *B. muricata, B. goegoensis, B. rajah,* and *B. nurii.* Of all the samples, only four identified genetically 'tallied' with their morphological identification, 25YC_ BFOX6 for the *ndh*A intron, 25YC_ BFOX6 and C9BMR for the *ndh*F–*rpl*32 spacer, and 25YC_ BFOX10 for the *rpl*32–*trn*L spacer.

Table 4.4: BLASTN results from sequenced ITS region of the *Begonia* samples in this study.

No.	Sample	Base number (bp)	Query coverage (%)	Percent identity (%)	Closest identification entry from GenBank	Reference
1	25YC_ BFOX6 *B. "foxworthyi"*	781	92	92.32		
2	25YC_ BFOX9 *B. foxworthyi*	783	92	91.66		
3	25YC_ BFOX10 *B. foxworthyi*	783	92	91.66	*B. muricata* (AY753725)	(Tebbitt et al., 2006)
4	25YC_ BYAP1 *B. yappii*	783	91	91.48		
5	25YC_ BSP_KT1 *B*. sp.	780	92	92.73		

Table 4.5: BLASTN results from sequenced *ndh*A intron of the *Begonia* samples in this study.

No.	Sample	Base number (bp)	Query coverage (%)	Percent identity (%)	Closest identification entry from GenBank	Reference
1	25YC_ BFOX6 *B. "foxworthyi"*	1118	100	99.11	*B. foxworthyi* (KR186468)	(Hughes et al., 2015)
2	25YC_ BFOX9 *B. foxworthyi*	1128	100	98.85		
3	25YC_ BFOX10 *B. foxworthyi*	1135	100	98.85		
4	25YC_ BYAP1 *B. yappii*	1138	100	99.03	*B. goegoensis* (JF756376)	(Thomas et al., 2011)
5	25YC_ BSP_KT1 *B*. sp.	1080	100	99.18		
6	C9BMR *B. rajah*	1204	100	98.51	*B. forbesii* (KR186467)	(Hughes et al., 2015)
7	BGCK1 *B*. sp.	940	100	99.26	*B. muricata* (JF756378)	(Thomas et al., 2011)

Table 4.6: BLASTN results from sequenced *ndh*F–*rpl*32 spacer of the *Begonia* samples in this study.

No.	Sample	Base number (bp)	Query coverage (%)	Percent identity (%)	Closest identification entry from GenBank	Reference
1	25YC_ BFOX6 *B. "foxworthyi"*	818	99	98.89	*B. foxworthyi* (KR186555)	(Hughes et al., 2015)
2	25YC_ BFOX9 *B. foxworthyi*	788	100	98.73	*B. sublobata* (KR186602)	
3	25YC_ BFOX10 *B. foxworthyi*	808	100	98.76	*B. ignorata* (KR186565)	
4	25YC_ BYAP1 *B. yappii*	738	100	98.24	*B. rajah* (MH454103)	(Tan, Tam and Kiew, 2018)
5	25YC_ BSP_KT1 *B*. sp.	737	100	98.64		
6	C9BMR *B. rajah*	608	100	99.18	*B. rajah* (MH207741)	(Moonlight et al., 2018)
7	BGCK1 *B*. sp.	821	100	97.93	*B. ignorata* (KR186565)	(Hughes et al., 2015)

Table 4.7: BLASTN results from sequenced *rpl*32–*trn*L spacer of the *Begonia* samples in this study.

No.	Sample	Base number (bp)	Query coverage (%)	Percent identity (%)	Closest identification entry from GenBank	Reference
1	25YC_ BFOX6 *B. "foxworthyi"*	932	100	99.04	*B. ignorata* (KR186738)	(Hughes et al., 2015)
2	25YC_ BFOX9 *B. foxworthyi*	955	100	97.12		
3	25YC_ BFOX10 *B. foxworthyi*	1191	100	98.74	*B. foxworthyi* (KR186728)	
4	25YC_ BYAP1 *B. yappii*	1028	95	99.08	*B. nurii* (MH208082)	(Moonlight et al., 2018)
5	25YC_ BSP_KT1 *B*. sp.	1151	95	98.35		
6	25YC_ BSP_STNG3 *B*. sp.	1102	100	97.58	B. ignorata (KR186738)	(Hughes et al., 2015)

4.2.4 Specimens Used for Phylogenetic Inference

Following BLASTN analysis, specimens from Ho (2024), along with relevant GenBank sequences, were used for the phylogenetic reconstruction of Begonia section Jackia. These are summarised in Table 4.8.

Table 4.8: Sequences of related *Begonia* species retrieved from GenBank and Ho (2024) which were used in the current study.

Species	Voucher	Markers			
		ITS	*ndh*A intron	*ndh*F–*rpl*32	*rpl*32–*trn*L
GenBank Samples					
B. fluvialis	M. Hughes 1489	-	-	-	MH207975
B. forbesii	Peng P22685 HAST	-	KR186467	-	KR186727
B. foxworthyi	Peng P22721	JX656702	KR186468	-	KR186728
B. goegoensis	L.L. Forrest 132	AF485138	-	-	-
	E DCT 08-107	-	JF756376	-	JF756544
B. ignorata	Peng P22725	-	KR186478	-	KR186738
B. muricata	W.S. Hoover 901	AY753725	-	-	-
	E WHA 27	-	JF756378	-	JF756546
B. nurii	FRI 824	-	-	MH207673	-
	C. Salamone 50627	-	-	-	MH208081

Table 4.8: Sequences of related *Begonia* species retrieved from GenBank and Ho (2024) which were used in the current study. (continued).

Species	Voucher	Markers			
		ITS	***ndh*A intron**	***ndh*F–*rpl*32**	***rpl*32–*trn*L**
B. rajah	L.L. Forrest 130	AF485136	-	-	-
	FRIM47082	-	-	MH207742	-
	FRI 88682	-	-	MH454103	-
	LBG 880168	-	-	MH207741	-
B. reginula	FRIM82471	-	-	MH207744	-
	FRI 82471	-	-	MH454104	-
B. stictopoda	C. Puglesi & al. 239	-	-	-	MH208198
B. sublobata	DEDEN1486 E	-	KR186515	-	KR186775
B. sudjanae	E DCT 08-109	-	JF756377	-	JF756545
B. tigrina	Peng P22720	JX656703	KR186526	-	KR186786
B. trichopoda	SUBOE 97	-	-	MH207827	MH208218
B. yappii	Peng 20243	KF636491	-	-	-

Table 4.8: Sequences of related *Begonia* species retrieved from GenBank and Ho (2024) which were used in the current study. (continued).

Species	Voucher	Markers			
		ITS	***ndh*A intron**	***ndh*F–*rpl*32**	***rpl*32–*trn*L**
B. yenyeniae	FRI470822	-	-	MH454102	-
B. versicolor	-	AF485090	NC047450	NC047450	NC047450
B. grandis	-	MK541080	OP618125	OP618125	OP618125
Samples from Ho (2024)		**Location**			
B. rajah	C1BMR	Kuala Terengganu, Terengganu			
B. rajah	C7BMR				
B. rajah	C8BMR				
B. rajah	C11BMR				
B. rajah	C12BMR				
B. nurii	BGNR1	Gua Ikan, Kelantan			
B. nurii	BGNG1	Gua Musang, Kelantan			
B. nurii	BGNM1				

Table 4.8: Sequences of related *Begonia* species retrieved from GenBank and Ho (2024) which were used in the current study. (continued).

Species	Voucher	Location
B. nurii	BGFK1	Kelantan
B. foxworthyi	BGFR1	Chiku, Kelantan
B. foxworthyi	BGFC1	
B. reginula	BGRE1	UM Rimba Ilmu
B. reginula	BGRE2	
B. reginula	BGRE3	
B. yenyeniae	BGYN1	

4.2.5 Phylogenetic Trees

After merging both the Maximum Likelihood (ML) and Bayesian Inference (BI) phylograms, a total of six phylogenetic trees were generated—four based on individual markers, one on chloroplast markers (*ndh*A intron + *ndh*F–*rpl*32 + *rpl*32–*trn*L) and one based on the concatenated markers (ITS + *ndh*A intron + *ndh*F–*rpl*32 + *rpl*32–*trn*L). Comparisons in topology and resolution amongst all analysed trees revealed the concatenated tree to have the highest phylogenetic resolution, while the ITS tree has a topology most similar to that of the concatenated tree. As such, these two trees are elaborated in detail here. The other trees, i.e., *ndh*A intron, *ndh*F–*rpl*32, *rpl*32–*trn*L, and chloroplast trees are provided as Appendix B–E. The trees generated from individual chloroplast markers were highly incongruent and demonstrated more polytomous characteristics. The chloroplast tree was more phylogenetically resolved in comparison to the individual marker trees; however, there were more clades that had poor nodal support compared to the concatenated tree.

4.2.5.1 Concatenated Tree

The concatenated (ITS + *ndh*A intron + *ndh*F–*rpl*32 + *rpl*32–*trn*L) phylogenetic tree showed the best resolution fo*r Begonia* species within section *Jackia* in comparison to the rest of the trees (Figure 4.7). When rooted with *B. grandis* and *B. versicolor*, Section *Jackia* forms a strongly supported monophyletic ingroup (ML=100 | PP=1). Four main clades can be inferred from the tree. Clade A

(ML=96 | PP=1) is monophyletic with strong nodal support, consisting of *B. reginula, B. yenyeniae, B. yappii, B. rajah*, as well as sample 25YC_BSP_KT1. Clades B and C were each poorly supported, with Clade B (PP=0.94) consisting of samples BGCN1 and 25YC_BFOX6, which formed a sister taxon. Clade C (ML=55 | PP=0.99) consisted of a fully supported *Begonia nurii* subclade (BGNM1 and BGNG1), as well as specimens BGFK1 and BGNR1 that are closely related to, but not conspecific with *B. nurii*. *Begonia foxworthyi* was inferred to be monophyletic (ML=100 | PP=1), although two different genotypes were identified even within Kelantan.

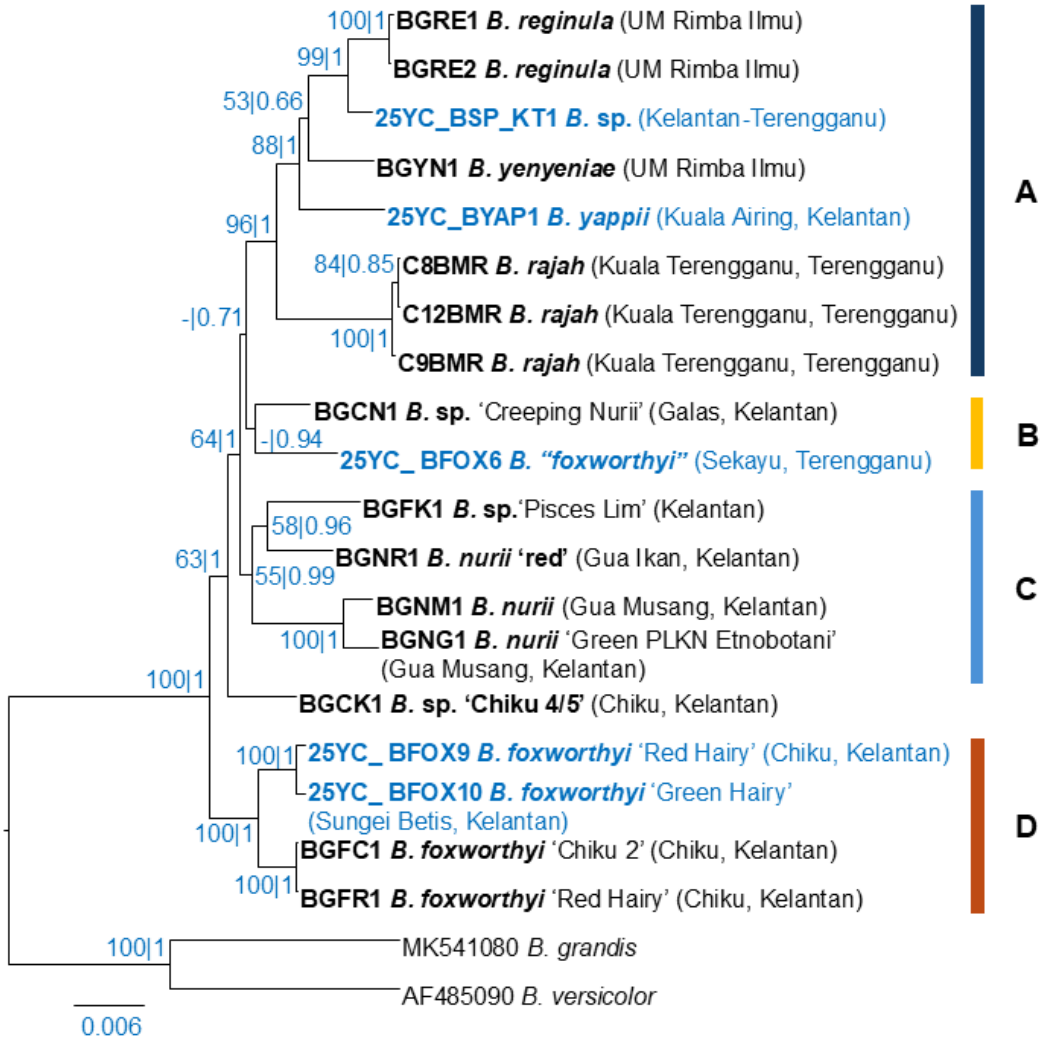


Figure 4.7. Bayesian tree showing phylogenetic relationship between *Begonia* species based on the fully concatenated dataset. Nodal values are arranged as: Ultrafast ML bootstrap support | Bayesian posterior probability. Scale bar corresponds to 0.006 nucleotide substitutions per site. Specimens in blue indicate specimens from this study. Hyphens indicate Ultrafast ML bootstrap support <50.

4.2.5.2 ITS Tree

The ITS dataset generated the most resolved phylogram amongst the four individual markers (Figure 4.8). Clade A consists of *B. reginula*, *B. rajah*, *B. foxworthyi*, as well as 25YC_BSP_KT1 and 25YC_BFOX6. Both *B. reginula* (ML=99 | PP=1) and *B. rajah* (ML=100 | PP=1) samples individually form strongly supported clades. Clade B exhibits polytomous character as many taxa can be seen to stem from the same ancestral line. Therefore, the relationship between them remains unresolved. The presumed-to-be *B. foxworthyi* samples (25YC_BFOX9, 25YC_BFOX10, BGFC1, and BGFR1) are paraphilic to the *B. foxworthyi* samples from GenBank since they can be found in different clades across the phylogram. Besides this, 25YC_BYAP1 was seen to group up with KF636491.

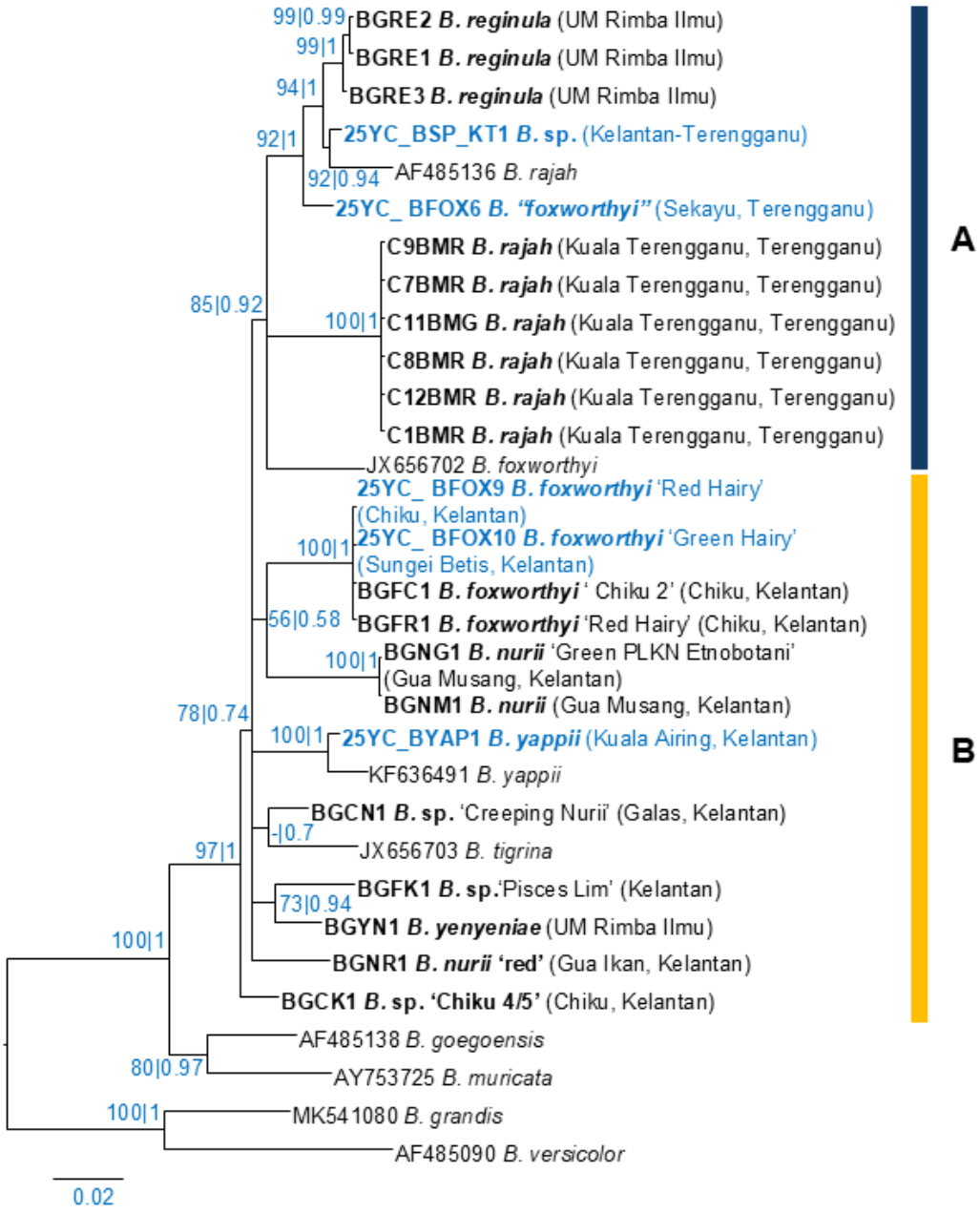


Figure 4.8. Bayesian tree showing phylogenetic relationship between *Begonia* species based on the ITS dataset. Nodal values are arranged as: Ultrafast ML bootstrap support | Bayesian posterior probability. Scale bar corresponds to 0.02 nucleotide substitutions per site. Specimens in blue indicate specimens from this study. Hyphens indicate Ultrafast ML bootstrap support <50.

4.2.5.3 Species Delimitation

The species delimitation of the concatenated ITS + *ndh*A intron + *ndh*F–*rpl*32 + *rpl*32–*trn*L dataset is shown here as it is the most phylogenetically informative (Figure 4.9). The remaining phylograms are shown in Appendix F–J. The species delimitation analyses of the concatenated dataset reveal that there are at least 14 different "species" identified within the dataset based on ASAP, ABGD and GMYC. BGRE1 and BGRE2 are of the same species, while C9BMR, C8BMR, and C12BMR are the same. All four algorithms agree that 25YC_BFOX9 and 25YC_BFOX10 are a separate "species" from BGFC1 and BGFR1. The results from bPTP differ from the rest of the algorithms, where it suggests that there are 15 "species" where BGFC1 and BGFR1 are not conspecific. Comparison of species delimitations results against the concatenated one revealed that the *rpl*32–*trn*L dataset showed the closest results to the concatenated dataset, where all four algorithms revealed at least 14 "species" which were the same as that the concatenated dataset results.

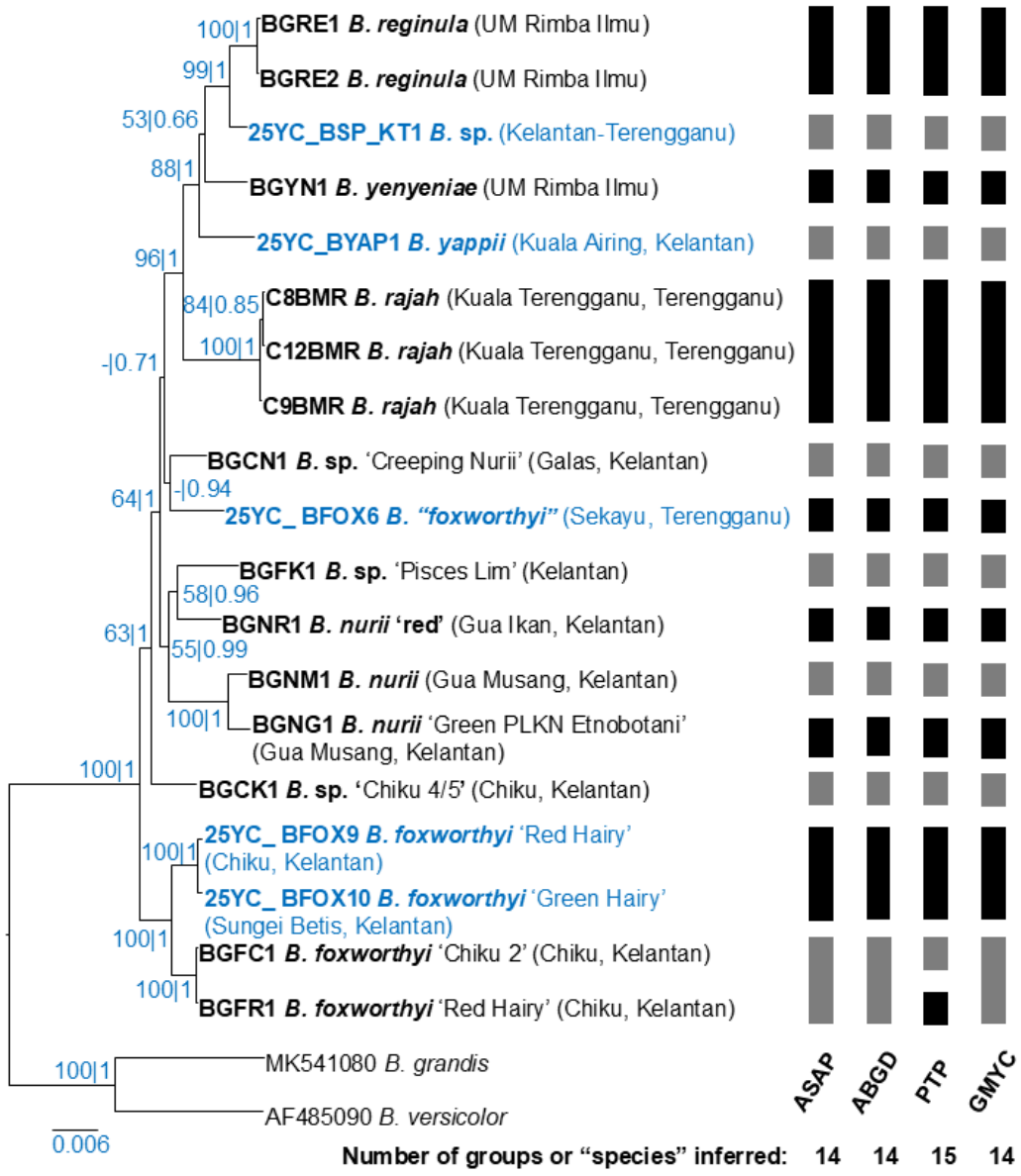


Figure 4.9. Bayesian tree showing phylogenetic relationship between *Begonia* species based on the fully concatenated dataset and its species delimitation analysis using ASAP, ABGD, PTP, and MNYC. Nodal values are arranged as: Ultrafast ML bootstrap support | Bayesian posterior probability. Scale bar corresponds to 0.006 nucleotide substitutions per site. Specimens in blue indicate specimens from this study. Hyphens indicate Ultrafast ML bootstrap support <50.

4.2.5.4 Genetic Distance

A summary of intra- and interspecific genetic variance of the six datasets (i.e., four individual markers, chloroplast markers and concatenated markers) is shown in Table 4.9.

Table 4.9: Summary of intra-and interspecific genetic variation between Begonia species for all four individual markers, chloroplast markers, and concatenated markers.

Markers	**Intraspecific GD (%)**	**Interspecific GD (%)**
ITS	0 – 1.33	1.03 – 9.86
*ndh*A intron	0 – 1.02	0.18 – 1.73
*ndh*F–*rpl*32	0 – 0.97	0.27 – 2.95
*rpl*32–*trn*L	0 – 0.90	0.28 – 2.04
Chloroplast markers	0 – 0.13	0.31 – 1.30
Concatenated markers	0 – 0.10	0.51 – 2.16

For simplicity, only the genetic distance matrix of the concatenated dataset is shown here as Table 4.10. The remaining genetic distance data are provided as Appendix K to O. Notably, the genetic distance between the B. rajah samples (C8BMR, C9BMR, and C12BMR) is low, ranging from 0–0.03%. 25YC_BSP_KT1 can be seen to be closely related to B. reginula as the genetic distance ranges from 0.53–0.56%.

Table 4.10: Genetic distance (%) between *Begonia* species based on fully concatenated dataset.

No.	Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	**BGFC1 *B. foxworthyi* 'Chiku 2'**																		
2	**BGFR1 *B. foxworthyi* 'Red Hairy'**	0.00																	
3	**BGRE1 *B. reginula***	1.84	1.84																
4	**BGRE2 *B. reginula***	1.82	1.82	0.03															
5	**BGYN1 *B. yenyeniae***	1.86	1.86	1.12	1.09														
6	**C8BMR *B. rajah***	1.96	1.96	1.53	1.51	1.67													
7	**C9BMR *B. rajah***	1.99	1.99	1.56	1.54	1.69	0.03												
8	**C12BMR *B. rajah***	1.96	1.96	1.53	1.51	1.67	0.00	0.03											
9	**BGFK1 *B.* sp. 'Pisces Lim'**	1.60	1.60	1.69	1.65	1.58	1.85	1.87	1.85										
10	**BGNR1 *B. nurii* 'red'**	1.49	1.49	1.65	1.63	1.44	1.70	1.73	1.70	1.23									
11	**BGNM1 *B. nurii***	1.64	1.64	1.94	1.92	1.94	1.96	1.98	1.96	1.75	1.54								

Table 4.10: Genetic distance (%) between *Begonia* species based on fully concatenated dataset (continued).

No.	Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
12	**BGNG1 *B. nurii*'Green PLKN Etnobotani'**	1.64	1.64	1.84	1.82	1.99	2.01	2.04	2.01	1.67	1.54	0.51							
13	**BGCN1 *B.* sp. 'Creeping Nurii'**	1.29	1.29	1.47	1.45	1.52	1.64	1.62	1.65	1.29	1.24	1.41	1.46						
14	**25YC_ BFOX9 *B. foxworthyi* 'Red Hairy'**	0.67	0.67	2.01	1.98	1.95	2.08	2.06	2.08	1.74	1.57	1.81	1.86	1.39					
15	**25YC_ BFOX10 *B. foxworthyi* 'Green Hairy'**	0.67	0.67	2.00	1.98	1.89	2.02	1.99	2.02	1.70	1.57	1.74	1.85	1.33	0.10				
16	**25YC_ BFOX6 *B. "foxworthyi"***	1.53	1.53	1.30	1.27	1.61	1.68	1.65	1.69	1.46	1.47	1.67	1.72	1.12	1.62	1.58			
17	**25YC_BYAP1 *B. yappii***	1.97	1.97	1.43	1.41	1.22	1.80	1.77	1.80	2.02	1.58	2.08	2.16	1.62	1.96	1.90	1.83		
18	**25YC_BSP_KT1 *B.* sp. 'Kelantan-Terengganu'**	1.86	1.86	0.56	0.53	1.04	1.36	1.33	1.36	1.74	1.48	1.86	1.94	1.38	1.90	1.83	1.07	1.12	
19	**BGCK1 *B.* sp. 'Chiku 4/5'**	1.34	1.34	1.63	1.61	1.69	1.76	1.73	1.76	1.53	1.24	1.65	1.63	1.11	1.40	1.39	1.37	1.74	1.45

CHAPTER 5

DISCUSSION

5.1 BLASTN Inconsistent in Species Identification

BLASTN identification results were not very accurate in the present study for all four markers (Tables 4.4–4.7), mainly due to the lack of reference sequence in GenBank. An example of which can be seen through the lack of B. yappii sequences for the *ndh*A intron, *ndh*F–*rpl*32 spacer, and *rpl*32–*trn*L spacer, where it was first sequenced in this study. The ITS region resulted in the lowest percentage identities (91.48–92.73%) as well as query coverage, leading all of the samples to be identified as *B. muricata* (Table 4.4). This was due to how hits that were better matches were unpublished and therefore, not considered to be reliable enough (Tan, Tam and Kiew, 2018). The lack of hits could be reflected from how many taxonomical *Begonia* studies prefer to use other molecular markers due to the high variability of the ITS region and its difficulty in sequence alignment (Moonlight et al., 2018; Thomas, 2010). This is not unexpected as many Malaysian *Begonia* species are yet to be genetically characterised, and one of the driving reasons for this project. The DNA data generated from this study (journal article in preparation) will be added to the GenBank database for more accurate BLASTN identification in the future. Notably, potential misidentification in

GenBank entries was also observed, i.e., *B. nurii* accession MH207673, highlighting the importance of being cautious in data retrieval, as well as the importance of a polyphasic approach in identification by adding in other information such as morphology, physiology, biochemistry, etc.

For samples that could be identified using BLASTN, an identification threshold of >99% is recommended for all markers. This threshold also takes into consideration the genetic distance between samples that have been confirmed to be the same species (Ross, Murugan and Li, 2008). However, the intra- and interspecific genetic variation overlaps across all of the individual markers, making it harder to distinguish and, in turn, reducing the confidence of species delimitation based on BLASTN. Even with a strict threshold (percentage identity >99%), a polyphasic approach to identification is still strongly recommended. The *ndh*A intron has the most hits over the 99% threshold; however, most of the identities assigned were still not aligned with the original identities identified using morphology (Table 4.5). Despite this, these GenBank results are unreliable due to the lack of metadata on the samples, especially locality information, making it difficult to further support the results. Given that begonias are highly endemic, it is unlikely that samples from different localities are conspecific, as local environmental pressures and geographic isolation reduce the genetic variance found between the species. This highlights the importance of location data in supplementing *Begonia* species identification.

As such, a more conservative approach is adopted in this study as species being left unassigned is preferred to their misidentification. As aforementioned, misidentified sequences in the GenBank can lead to confusion and erroneous conclusions in future taxonomic studies on Southeast Asian *Begonia* (Tan, Tam and Kiew, 2018)

5.2 Phylogenetic Relationship of *Begonia* Section *Jackia*

The concatenated phylogenetic tree (ITS + *ndh*A intron + *ndh*F–*rpl*32 + *rpl*32–*trn*L; Figure 4.7) resulted in the best resolution compared to individual chloroplast markers. This is within expectations as combining all DNA markers not only increases the number of phylogenetically informative sites (Table 4.3) but also offsets the different evolutionary rates of each DNA marker. The phylogeny also supported the monophyly of *Begonia* sect. *Jackia* as erected by Moonlight et al. (2018). Nevertheless, poor resolution was still observed, particularly in clades B and C, which mainly contained samples that are assumed to be *B. nurii*. The clades show weak to no nodal support, which is not uncommon for phylogenetic studies in *Begonia* studies (Moonlight et al., 2018; Tan, Tam and Kiew, 2018). This is due to how plants growing in tropical regions tend to evolve at a faster rate in comparison to their temperate counterparts (Wang et al., 2020). The faster evolution rate leads to an increase in variance throughout the sequences. Greater amounts of variance, although desirable for phylogenetic

reconstruction, may make it more challenging to resolve phylogenetic relationships, especially when it occurs in a shorter timeframe (Duchêne et al., 2021). A contributing factor to this would be substitution saturation, where the same nucleotide site evolves and alters multiple times, as it masks over phylogenetic information (Xia et al., 2003).

Despite this, the concatenated tree also resulted in a well-supported clade that included *B. rajah*, *B. reginula*, and *B. yenyeniae*. This is congruent with that suggested by Tan, Tam and Kiew (2018), further supporting that these species are closely related and share a recent common ancestor. The phylogenetic tree also shows several samples (25YC_BFOX6 and 25YC_BSP_KT1) as potentially distinct species, which are worth investigating further. Notably, there are still a number of species from Section *Jackia* unaccounted for (i.e., *B. ignorata*, *B. forbesii*, *B. tigrina*, etc.), highlighting the need for intensified sampling and genetic characterisation efforts.

5.3 Species Delimitation

Species delimitation based on ASAP, ABGD and GMYC identified 14 different "species" within the concatenated dataset, while bPTP suggested 15 "species" (Figure 4.9). This indicated that there are potentially more novel species than originally anticipated. Discrepancies in species delimitation are common between

the species delimitation models used, as they each emphasise different data. According to this dataset, In addition to BGCK1 and BGCN1, there are likely three potentially new species of *Begonia* under the section *Jackia,* namely, 25YC_BSP_KT1, 25YC_BFOX6, 25YC_BFOX9, and 25YC_BFOX10. Despite this, it is important to consider multiple sources of information, including morphology, genetic data, and phylogenetic analyses, to avoid misidentification or the risk of redescribing a species.

25YC_BSP_KT1 can be seen to be closely related to *B. reginula* and *B. yenyeniae* (Figure 4.7). This is further supported by the low genetic distance ranging 0.53–0.56% for *B. reginula* and 1.04% for *B. yenyeniae*, both of which fall below the interspecific genetic variance threshold of 0.51–2.16% (Table 4.10). Besides this, common foliar morphology can be observed between them, such as bronzy interveinal bullation as well as ciliated leaf edges (Tan, Tam and Kiew, 2018; Figure 4.3). 25YC_BFOX6 is the most closely related to BGCN1, as they have a genetic distance of 1.12% and also form sister taxa based on the concatenated tree.

Furthermore, 25YC_BFOX9 and 25YC_BFOX10 (GD= 0.10%) were consistently identified as the same species by all four species delimitation algorithms (Figure 4.9). Although both samples are slightly different in terms of morphology, such as leaf colour and shape, where 25YC_BFOX9 is a dark red to

green, while 25YC_BFOX10 has leaves in a light green colour, they still share important morphological characteristics, such as leaves that have bullation and are completely ciliated (Figures 4.1–4.2). The differences in morphology are likely a result of phenotypic plasticity as they have different localities where they are subjected to different environmental conditions (Schlichting, 1986). Additionally, color is often considered an unreliable trait in distinguishing species (Kendal et al., 2013). This comparison is limited as there is a lack of inflorescence morphology for 25YC_BFOX10. Returning to its locality to collect this data would be useful in solidifying that they are conspecific since floral morphology is a critical characteristic in *Begonia* taxonomy (Doorenbos, Sosef, and Wilde, 1998). In future instances, fresh samples should be collected and cultured to avoid additional time and resources spent on returning to the original site. When compared to the morphology of *B. foxworthyi* specimens (BGFR1 and BGFC1), it is suggested that they are conspecific despite the results from species delimitation showing otherwise. This brings about an issue amongst taxonomists, where 'lumpers' prefer to generally classify individuals with a broad range of characteristics, leading to the description of fewer species; whereas, 'splitters' prefer to classify species based on distinct characteristics (Thaxton et al., 2022). It is challenging to identify what should be considered as a new species since the definition of species, the ability to produce fertile offspring, is not accurately reflected in plants as they are capable of hybridisation (Stace, 2005). There is potential in the overestimation of species by species delimitation algorithms, especially as the four markers used in this study are non-coding regions, which

tend to evolve faster due to their lack of expression (Kostka, Hahn and Pollard, 2010). This shows that more data, particularly the reproductive morphology, is required to make an informed decision on whether to split or lump 25YC_BFOX9 and 25YC_BFOX10 with BGFR1 and BGFC1.

Notably, despite these findings, the risk of misidentification still exists due to the lack of genetic data for several *Begonia* species within section *Jackia* from Peninsular Malaysia. This again underscores the need to increase sampling and genetic characterisation efforts, particularly by expanding the number of individuals studied for BGCK1, BGCN1, 25YC_BSP_KT1, 25YC_BFOX6, 25YC_BFOX9, and 25YC_BFOX10, to achieve more robust morphological and genetic insights.

5.4 Marker Assessment

5.4.1 Ease of Extraction, Amplification, and Sequencing

Comparison of the sequences C9BMR, BGCK1, and BGCN1 amplified using Phusion™ DNA polymerase, for *ndh*A intron, *ndh*F–*rpl*32, against those from Ho (2024) found no difference in sequencing quality. The amplification of the chloroplast markers (*ndh*A intron, *ndh*F–*rpl*32 and *rpl*32–*trn*L) was more

challenging compared to the ITS marker as they have a high AT content (Table 4.3). AT-rich sequences are harder to amplify since they have a lower melting stability as AT base pairs are only connected by two hydrogen bonds compared to the three hydrogen bonds in GC base pairs (Rajewska, Wegrzyn and Konieczny, 2012). This increases the chances for errors such as DNA slippage or the formation of secondary structures, which causes the sequence to stall and dissociate (New England Biolabs, 2017). In this study, the amplification of the *rpl*32–*trn*L marker, the sequence with the highest AT content, was unsuccessful when using Phusion™ DNA polymerase, requiring it to be amplified using *Taq* DNA polymerase instead. However, it is not necessary that *Taq* DNA polymerase is better at amplifying sequences with high AT content (Oyola et al., 2012). Rather, it is because the PCR protocol for Phusion™ DNA polymerase requires more optimisation to amplify AT-rich sequences, whereas the *Taq* DNA polymerase had already been optimised from previous studies (Dhatterwal, Mehrotra and Mehrotra, 2017). Out of the four markers, *ndh*A intron and *rpl*32–*trn*L had the worst sequencing quality (i.e., higher amounts of noise) and required more manual corrections.

5.4.2 Recommendation of ITS for DNA Barcoding of *Begonia* Section *Jackia*

This project represents the first effort to evaluate the performance of four DNA markers recommended by Ardi et al. (2022) in inferring the phylogeny and delimiting species of *Begonia* in section *Jackia* from Peninsular Malaysia.

Amongst all four DNA markers tested, the ITS region proves to be the best individual marker within this study. The amplification of the region was relatively easy (i.e., all amplified in the first attempt), and it had a sequencing success rate of 100%. It also did not require Phusion™ DNA polymerase to produce good-quality nucleotide sequences, highlighting its ease of use and practicality. Besides this, it resulted in the most resolved phylogenetic tree and the most distinct barcoding gap in comparison to the other individual markers (Figure 4.8; Appendix A). However, the gap is still very small in contrast to conventional DNA barcodes for plants (Besse, Da Silva and Grisoni, 2020). Similar results were seen by Jiao and Shui (2013), where the ITS1/ITS2 region demonstrated significant intra- and interspecific genetic divergence, which allows for more accurate species delimitation for species of *Begonia* in China. Additionally, ITS is the universal DNA barcode for fungi, demonstrating its capability in DNA barcoding (Schoch et al., 2012). However, results from its species delimitation analyses did not complement those of the concatenated markers, as only revealed at most 12 different "species", where it considered 25YC_BFOX9, 25YC_BFOX10, and the *B. foxworthyi* samples as conspecific (Appendix F). This demonstrates how the use of a single marker may lack the resolution to discriminate against species that are very closely related - an uncommon scenario. As such, the use of the ITS region is recommended for the DNA barcoding of *Begonia* species in Peninsular Malaysia.

Additionally, the DNA sequences generated in this study will also serve as valuable GenBank references (manuscript in preparation), in which *B. yappii* was DNA sequenced for the first time for the *ndh*A intron, *ndh*F–*rpl*32 spacer, and *rpl*32–*trn*L spacer. This will aid in the proper documentation and species identification for *Begonia* under sect. *Jackia*, allowing for improvements in conservation and bioprospecting. Categorising the biodiversity and their phylogenetic relationship will help bring attention to those that require priority of conservation and identify *Begonia* species that have potential horticultural and medicinal value. These results can be extended to *Begonias* throughout Southeast Asia, serving as a groundwork for future *Begonia* taxonomic studies.

5.5 Conformation on the Identity of *B. rajah*

Due to the lack of inflorescence morphology by Ho (2024), the wild specimen of *B. rajah* (C9BMR) was cultured and described (Figure 4.4). The inflorescence characteristics displayed by C9BMR are almost identical to the descriptions of *B. rajah.* The only disparity is in the colour of the ovaries, where C9BMR had pink ovaries, while Kiew (2005) described *B. rajah* with whitish green ovaries. However, Tan, Tam and Kiew (2018) have more recently described *B. rajah* with the same pink ovaries as C9BMR. However, colour is not a critical character in the identification and delineation of species (Kendal et al., 2013). Besides this, the foliar differences from the previous study by Ho (2024) between the suspected and cultivated *B. rajah*, such as the lack of surface bullation, were only observed

in younger leaves on C9BMR, whereas mature leaves showed all the typical morphology of *B. rajah*. With the addition of the low genetic distance from the previous study, it is safe to conclude that the suspected wild samples of *B. rajah*, are conspecific to cultured *B. rajah* making it no longer extinct in the wild, where it is found in Kuala Terengganu, Terengganu. Additionally, genetically different individuals of *B. rajah* C9BMR, C8BMR and C11BMR (GD= 0.00–0.03%) collected from the same location are also being cultivated in UTAR's Horticultural Shed as part of germplasm conservation efforts.

5.6 Limitations and Recommendations

The main limitation of this study stemmed from unreliable references from GenBank. Many of the sequences within GenBank are unpublished (i.e., lack credibility), and also lack useful metadata, such as the specimen location, while some were misidentified (i.e., *B. nurii* accession MH207673). This resulted in issues in species identification and phylogenetic inferences. A key recommendation to aid this would be to use sequences from the type species as references. Some of which can be derived from reputable sources like museum collections, herbaria, and botanical gardens, although notably aged specimens often do not harbor sufficient intact DNA for PCR and sequencing, in addition to policies of collections which discouraged destructive sampling (Drábková, 2020). For species without type specimens, an option would be to track down the type locality, the location from which the species was originally collected (Caldwell,

Lima and Keller, 2002). The collection and sequencing of samples from the type locality would provide specimens that were highly similar to the type specimens.

The limited availability of samples and morphological data has led to the uncertainty of indemnification and phylogenetic placement of certain samples (e.g., 25YC_BFOX9 and 25YC_BFOX10). This is due to the exhaustive effort required to obtain the samples. However, to address this, additional sampling, amplification, and sequencing of the samples is required. Sequencing more samples from the same species would result in a more reliable intraspecific genetic variation range. With the incorporation of more comprehensive morphological data, these efforts would result in a more robust phylogeny for *Begonia* under the section *Jackia* in Peninsular Malaysia.

CHAPTER 6

CONCLUSION

As a conclusion to this study, the phylogenetic relationships between species of *Begonia* sect. *Jackia* within Peninsular Malaysia were inferred based on the concatenated phylogenetic tree (ITS + *ndh*A intron + *ndhF-rpl32* + *rpl32-trnL*) which showed the best topology and species delimitation despite having inconsistent resolution for a few *Begonia* species, specifically BGFK1, BGCK1, and *B*. *nurii* samples. Whereas well supported clades formed where clade A grouped up *B. reginula*, *B. yenyeniae*, *B. yappii*, *B. rajah*, and 25YC_BSP_KT1 while Clade D consisted of *B. foxworthyi* samples. The phylogeny also supported the monophyly of *Begonia* sect. *Jackia*.

Besides this, at least two species (25YC_BFOX6 and 25YC_BSP_KT1) are potentially novel, while the taxonomic status of others remains uncertain as species delimitation algorithms based on the concatenated dataset (ITS + *ndh*A intron + *ndh*F–*rpl*32 + *rpl*32–*trn*L) reveal at least 14 different "species", with the inconsistency of bPTP suggesting 15 "species". On the other hand, the separation of 25YC_BFOX9 and 25YC_BFOX10 from BGFR1 and BGFC1 is inconclusive,

as morphological data suggest they are conspecific despite molecular data revealing otherwise. Therefore, further investigation is required, specifically, additional sampling, or retrieval of their type specimen and its information.

The present study represents the first effort to assess potential DNA markers to be used for DNA barcoding of *Begonia* in Peninsular Malaysia, ITS marker was identified as the best individual marker as it best reflects the fully concatenated dataset and is able to reliably identify *Begonia* species. It should be used in future Begonia projects where these properties would likely allow it to be further used throughout Southeast Asia. The chloroplast markers (*ndh*A intron, *ndh*F–*rpl*32 and *rpl*32–*trn*L spacers) were more challenging to amplify due to their high AT content, even with the use of Phusion™ DNA polymerase. The DNA sequences generated in this study will serve as valuable GenBank references, where *B. yappii* was DNA sequenced for the first time for the *ndh*A intron, *ndh*F–*rpl*32 spacer, and *rpl*32–*trn*L spacer. Collectively, these findings can contribute as a groundwork for *Begonia* studies in other parts of Southeast Asia, where many *Begonai* species are also threatened and understudied.

Lastly, the morphology of wild samples of *B. rajah* collected from the state of Terengganu, as a follow-up to Ho (2024), was observed. The inflorescence of which matched that of cultivated *B. rajah*, when accompanied with molecular data (i.e., phylogenetic relationship and genetic distance), signifies that they are

conspecific. This signifies that *B. rajah* remains extant in its natural environment in Peninsular Malaysia, overturning its long-held extinct-in-the-wild status of over a century.

REFERENCES

Alexander, D.C., Levett, P.N. and Turenne, C.Y., 2015. Molecular taxonomy. *Molecular Medical Microbiology*, pp.369–379. https://doi.org/10.1016/b978-0-12-397169-2.00021-4.

Al-Samarai, F.R. and Al-Kazaz, A.A., 2015. Molecular markers: an introduction and applications. *European Journal of Molecular Biotechnology*, 9(3), pp.118–130. https://doi.org/10.13187/ejmb.2015.9.118.

Ardi, W.H., Chikmawati, T., Witono, J.R. and Thomas, D., 2018. A synopsis of *Begonia* (Begoniaceae) of Southeastern Sulawesi including four new species. 381(1), pp.27–27. https://doi.org/10.11646/phytotaxa.381.1.7.

Ardi, W., Campos, L., Chung, K.F., Dong, W.-K., Drinkwater, E., Fuller, D., Gagul, J., Garnett, G., Girmansyah, D., Goodall-Copestake, W., Hughes, M., Jacques, E., Jara Muaoz, O.A., Sang, J., Kidner, C., Kiew, R., Krishna, N., Li, R. and Kumarage, L., 2022. Resolving phylogenetic and taxonomic conflict in *Begonia*. *Edinburgh Journal of Botany*, 79, pp.1–28. https://doi.org/10.24823/ejb.2022.1928

Atkins, H.J., Bramley, G.L.C., Nishii, K., Möller, M., Jay, Kartonegoro, A. and Hughes, M., 2021. Sectional polyphyly and morphological homoplasy in Southeast Asian *Cyrtandra* (Gesneriaceae): consequences for the taxonomy of a mega-diverse genus. *Plant Systematics and Evolution*, 307(6). https://doi.org/10.1007/s00606-021-01784-x.

Badcock, Z., 1998. *A phylogenetic investigation of Begonia L, section Knesebeckia (Klotzsch)* A.DC. PhD. University of Glasgow.

Bawono, P., Dijkstra, M., Pirovano, W., Feenstra, A., Abeln, S. and Heringa, J., 2016. Multiple sequence alignment. *Methods in Molecular Biology*, pp.167–189. https://doi.org/10.1007/978-1-4939-6622-6_8.

Bazinet, A.L., Ondov, B.D., Sommer, D.D. and Ratnayake, S., 2018. BLAST-based validation of metagenomic sequence assignments. *PeerJ*, 6, p.e4892. https://doi.org/10.7717/peerj.4892.

Besse, P., Da Silva, D. and Grisoni, M., 2020. *Plant DNA barcoding principles and limits: A case study in the genus Vanilla.* In Molecular plant taxonomy: methods and protocols (pp. 131-148). New York, NY: Springer US.

Bidyananda, N., Jamir, I., Nowakowska, K., Varte, V., Vendrame, W.A., Devi, R.S. and Nongdam, P., 2024. Plant genetic diversity studies: insights from DNA marker analyses. *International Journal of Plant Biology*, 15(3), pp.607–640. https://doi.org/10.3390/ijpb15030046.

Bland, L.M., Bielby, J., Kearney, S., Orme, C.D.L., Watson, J.E.M. and Collen, B., 2017. Toward reassessing data-deficient species. *Conservation Biology*, 31(3), pp.531–539. https://doi.org/10.1111/cobi.12850.

Borah, D., Dutta, R., Majumdar, S. and Mili, C., 2025. *Begonia* species: a review on its ethnobotany, phytochemicals, and biological activities. *Discover Plants*, [online] 1(182). https://doi.org/10.1007/s44372-025-00272-7.

Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F.K., Müller, N.F., Ogilvie, H.A., du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I. and Suchard, M.A., 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology*, 15(4), p.e1006650. https://doi.org/10.1371/journal.pcbi.1006650.

Briggs, B.G., 1991. One hundred years of plant taxonomy, 1889-1989. *Annals of the Missouri Botanical Garden*, 78(1), p.19. https://doi.org/10.2307/2399588.

Bromham, L. and Penny, D., 2003. The modern molecular clock. *Nature Reviews Genetics*, 4(3), pp.216–224. https://doi.org/10.1038/nrg1020

Buyun, L., Tkachenko, H., Kurhaluk, N., Góralczyk, A., Tomin, V., Osadowski, Z., 2019. Screening for antimicrobial activity of nine ethanolic extracts obtained from leaves of *Begonia* plant: a possible alternative in the treatment of infections caused by *Citrobacter freundii*. *Agrobiodiversity for Improving Nutrition Health and Life Quality*, pp.312–322. https://doi.org/10.15414/agrobiodiversity.2019.2585-8246.312-322.

Caldwell, J.P., Lima, A.P. and Keller, C., 2002. Redescription of *Colostethus marchesianus* from its type locality. *Copeia*, 2002(1), pp.157-165.

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

Choi, B., Crisp, M.D., Cook, L.G., Meusemann, K., Edwards, R.D., Toon, A. and Külheim, C., 2019. Identifying genetic markers for a range of phylogenetic utility–From species to family level. *PloS one*, 14(8), pp.e0218995–e0218995. https://doi.org/10.1371/journal.pone.0218995.

Chua, L.S.L., Kiew, R. and Chan, Y.M., 2009. Assessing conservation status of Peninsular Malaysian *Begonias* (*Begoniaceae*). *Blumea - Biodiversity, Evolution and Biogeography of Plants*, 54(1), pp.94–98. https://doi.org/10.3767/000651909x474131.

Chung, K.-F., Leong, W.-C., Rubite, R.R., Repin, R., Kiew, R., Liu, Y. and Peng, C.-I., 2014. Phylogenetic analyses of *Begonia* sect. *Coelocentrum* and allied limestone species of China shed light on the evolution of Sino-Vietnamese karst flora. *Botanical Studies*, 55(1). https://doi.org/10.1186/1999-3110-55-1.

Crozier, R.H., 1997. Preserving the information content of species: genetic diversity, phylogeny, and conservation worth. *Annual Review of Ecology and Systematics*, 28(1), pp.243–268. https://doi.org/10.1146/annurev.ecolsys.28.1.243.

de Vere, N., Rich, T.C.G., Trinder, S.A. and Long, C., 2014. DNA barcoding for plants. *Methods in Molecular Biology*, [online] 1245, pp.101–118. https://doi.org/10.1007/978-1-4939-1966-6_8.

Dewitte, A., Twyford, A.D., Thomas, D.C., Kidner, C.A. and Van Huylenbroeck, J., 2011. The origin of diversity in *Begonia*: genome dynamism, population processes and phylogenetic patterns. *The dynamical processes of biodiversity-case studies of evolution and spatial distribution*, pp.27-52.

Drábková, L.Z., 2020. Herbarium specimens: a treasure for DNA extraction, an update. *Methods in Molecular Biology*, pp.69–88. https://doi.org/10.1007/978-1-0716-0997-2_4.

Dhatterwal, P., Mehrotra, S. and Mehrotra, R., 2017. Optimization of PCR conditions for amplifying an AT-rich amino acid transporter promoter sequence with high number of tandem repeats from *Arabidopsis thaliana*. *BMC Research Notes*, [online] 10(1). https://doi.org/10.1186/s13104-017-2982-1.

Dias, A.A., 2019. Mapping *Begonia* hotspots on Borneo. PhD. University of Edinburgh.

Dolgova, A.S. and Stukolova, O.A., 2017. High-fidelity PCR enzyme with DNA-binding domain facilitates de novo gene synthesis. *3 Biotech*, 7(2). https://doi.org/10.1007/s13205-017-0745-2.

Doorenbos, J., Sosef, M.S.M. and Wilde, J.J.F.E. de, 1998. The sections of *Begonia*: including descriptions, keys and species lists. *Studies in Begoniaceae*. Wageningen: Agricultural University.

Duchêne, D.A., Mather, N., Van Der Wal, C. and Ho, S.Y.W., 2021. Excluding loci with substitution saturation improves inferences from phylogenomic data. *Systematic Biology*, 71(3), pp.676–689. https://doi.org/10.1093/sysbio/syab075.

Edgar, G.J., 2025. IUCN Red List criteria fail to recognise most threatened and extinct species. *Biological Conservation*, 301, p.110880. https://doi.org/10.1016/j.biocon.2024.110880.

Emelianova, K., Martínez Martínez, A., Campos-Dominguez, L. and Kidner, C., 2021. Multi-tissue transcriptome analysis of two *Begonia* species reveals dynamic patterns of evolution in the chalcone synthase gene family. *Scientific Reports*, 11(1). https://doi.org/10.1038/s41598-021-96854-y.

Garcia-Diaz, M. and Bebenek, K., 2007. Multiple functions of DNA polymerases. *Critical Reviews in Plant Sciences*, [online] 26(2), pp.105–122. https://doi.org/10.1080/07352680701252817.

Glez-Peña, D., Gómez-Blanco, D., Reboiro-Jato, M., Fdez-Riverola, F. and Posada, D., 2010. ALTER: program-oriented conversion of DNA and protein alignments. *Nucleic Acids Research*, 38(suppl_2), pp.W14–W18. https://doi.org/10.1093/nar/gkq321.

Gohl, D.M., Auch, B., Certano, A., LeFrançois, B., Bouevitch, A., Doukhanine, E., Fragel, C., Macklaim, J., Hollister, E., Garbe, J. and Beckman, K., 2021. Dissecting and tuning primer editing by proofreading polymerases. *Nucleic Acids Research*, 49(15), pp.e87–e87. https://doi.org/10.1093/nar/gkab471.

Goodall-Copestake, W.P., Pérez-Espona, S., Harris, D.J. and Hollingsworth, P.M., 2010. The early evolution of the mega-diverse genus *Begonia* (Begoniaceae) inferred from organelle DNA phylogenies. *Biological journal of the Linnean society*, 101(2), pp.243-250. https://doi.org/10.1111/j.1095-8312.2010.01489.x

Guindon, S., Delsuc, F., Dufayard, J.-F. and Gascuel, O., 2009. Estimating maximum likelihood phylogenies with PhyML. *Methods in Molecular Biology*, pp.113–137. https://doi.org/10.1007/978-1-59745-251-9_6.

Haider, N., 2018. A brief review on plant taxonomy and its components. *The Journal of Plant Science Research*, 34(2), pp.277-292.

Hall, B.G., 2013. Building phylogenetic trees from molecular data with MEGA. *Molecular Biology and Evolution*, 30(5), pp.1229–1235. https://doi.org/10.1093/molbev/mst012.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows. 41, pp.95–98.

Harrison, N., Harrison, R.J. and Kidner, C.A., 2016. Comparative analysis of *Begonia* plastid genomes and their utility for species-level phylogenetics. *PLOS ONE*, 11(4), p.e0153248. https://doi.org/10.1371/journal.pone.0153248.

Hebert, P.D.N., Cywinska, A., Ball, S.L. and deWaard, J.R., 2003. Biological identifications through DNA barcodes. *Proceedings. Biological Sciences*, 270(1512), pp.313–21. https://doi.org/10.1098/rspb.2002.2218.

Hickman, E.J., Yates, C.J. and Hopper, S.D., 2017)\. Botanical illustration and photography: a southern hemisphere perspective. Australian Systematic Botany, 30(4), p.291. https://doi.org/10.1071/sb16059.

Ho, A.E.Z., 2024. *Morphological and molecular characterization of 'extinct in the wild' Begonia rajah (Ridley, 1894) and potential new Begonia spp. from Peninsular Malaysia*. Undergraduate. Universiti Tunku Abdul Rahman. Unpublished.

Hollingsworth, P.M., Graham, S.W. and Little, D.P., 2011. Choosing and using a plant DNA barcode. *PLoS ONE*, 6(5), p.e19254. https://doi.org/10.1371/journal.pone.0019254.

Huang, Q., Baum, L. and Fu, W.-L., 2010. Simple and practical staining of DNA with GelRed in agarose gel electrophoresis. *Clinical Laboratory*, [online] 56(3-4), pp.149–152. Available at: <https://pubmed.ncbi.nlm.nih.gov/20476647/> [Accessed 6 September 2025]

Huelsenbeck, J.P. and Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17(8), pp.754-755.

Hughes, M., Moonlight, P.W., Jara-Muñoz, A., Tebbitt, M.C., Wilson, H.P., Zhong, K., Mills, T. and Pullan, M., 2025. *Begonia*. [online] *Begonia* Resource Centre. Available at: <https://padme.rbge.org.uk/Begonia/> [Accessed 6 September 2025]

Hughes, M., Rubite, R.R., Blanc, P., Chung, K.-F. and Peng, C.-I., 2015. The Miocene to Pleistocene colonization of the Philippine archipelago by *Begonia* sect. *Baryandra* (Begoniaceae). *American Journal of Botany*, 102(5), pp.695–706. https://doi.org/10.3732/ajb.1400428.

Ishino, S. and Ishino, Y., 2014. DNA polymerases as useful reagents for biotechnology - the history of developmental research in the field. *Frontiers in Microbiology*, [online] 5(465). https://doi.org/10.3389/fmicb.2014.00465.

IUCN, 2025. *The IUCN Red List of Threatened Species*. [online] IUCN Red List. Available at: <https://www.iucnredlist.org/> [Accessed 3 September 2025]

Jiao, L.J. and Shui, Y.M., 2013. Evaluating candidate DNA barcodes among Chinese *Begonia* (Begoniaceae) species. *Plant Diversity*, 35(6), pp.715-724. https://doi.org/10.7677/ynzwyj201313064

Jinbo, U., Kato, T. and Ito, M., 2011. Current progress in DNA barcoding and future implications for entomology. *Entomological Science*, 14(2), pp.107–124. https://doi.org/10.1111/j.1479-8298.2011.00449.x.

Jörger, K.M. and Schrödl, M., 2013. How to describe a cryptic species? Practical challenges of molecular taxonomy. *Frontiers in Zoology*, 10(1), p.59. https://doi.org/10.1186/1742-9994-10-59.

Kendal, D., Hauser, C.E., Garrard, G.E., Jellinek, S., Giljohann, K.M. and Moore, J.L., 2013. Quantifying plant colour and colour difference as perceived by humans using digital images. *PLoS ONE*, 8(8), p.e72296. https://doi.org/10.1371/journal.pone.0072296.

Khehra, N., Padda, I.S. and Swift, C.J., 2023. Polymerase chain reaction (PCR). [online] *PubMed*. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK589663/> [Accessed 3 September 2025]

Kiew, R., 2005. *Begonias of Peninsular Malaysia*. Sabah, Malaysia: Natural History Publications Sdn. Bhd.

Kono, Y., Peng, C.-I., Kazuo Oginuma, Yang, H.-A. and Chung, K.-F., 2023. Cytological study of *Begonia* sections *Jackia*, *Platycentrum* and *Reichenheimia* (Begoniaceae) with chromosome numbers 2*n*=34 and 38. *CYTOLOGIA*, 88(2), pp.161–166. https://doi.org/10.1508/cytologia.88.161.

Kostka, D., Hahn, M.W. and Pollard, K.S., 2010. Noncoding sequences near duplicated genes evolve rapidly. *Genome Biology and Evolution*, 2, pp.518–533. https://doi.org/10.1093/gbe/evq037.

Kress, W.J. and Erickson, D.L., 2008. DNA barcodes: Genes, genomics, and bioinformatics. *Proceedings of the National Academy of Sciences of the United States of America*, [online] 105(8), pp.2761–2762. https://doi.org/10.1073/pnas.0800476105.

Laferrière, J.E., 1992. Begonias as food and medicine. *Economic Botany*, 46(1), pp.114-116.

Lalawmpuii, L. and Tlau, L., 2021. *Begonia roxburghii*: A potentially important medicinal plant. *Science Vision*, 21(1), pp.22–25. https://doi.org/10.33493/scivis.21.01.05.

Larkin, D.J., Jacobi, S.K., Hipp, A.L. and Kramer, A.T., 2016. Keeping all the pieces: phylogenetically informed *ex situ* conservation of endangered species. *PLOS ONE*, 11(6), p.e0156973. https://doi.org/10.1371/journal.pone.0156973.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. and Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)*, 23(21), pp.2947–8. https://doi.org/10.1093/bioinformatics/btm404.

Lauri, A. and Mariani, P.O., 2009. Potentials and limitations of molecular diagnostic methods in food safety. *Genes & Nutrition*, [online] 4(1), pp.1–12. https://doi.org/10.1007/s12263-008-0106-1.

Lee, P.Y., Costumbrado, J., Hsu, C.-Y. and Kim, Y.H., 2012. Agarose gel electrophoresis for the separation of DNA fragments. *Journal of Visualized Experiments*, [online] 3923(62). https://doi.org/10.3791/3923.

Luscombe, N.M., Greenbaum, D. and Gerstein, M., 2001. What is bioinformatics? An introduction and overview. *Yearbook of medical informatics*, 10(01), pp.83-100.

Latiff, A., 2018. Biodiversity in Malaysia. *Apple Academic Press eBooks*, pp.307–349. https://doi.org/10.1201/9780429487743-10.

Mahardika, I.W.A., 2016. *The origin of Sumatran Begonia flora*. PhD. The University of Edinburgh and Royal Botanic Garden Edinburgh.

Marchioni, I., Taglieri, I., Dimita, R., Ruffoni, B., Zinnai, A., Venturi, F., Sanmartin, C. and Pistelli, L., 2022. Postharvest treatments on sensorial and biochemical characteristics of *Begonia cucullata* willd edible flowers. *Foods*, 11(10), p.1481. https://doi.org/10.3390/foods11101481.

McInerney, P., Adams, P. and Hadi, M.Z., 2014. Error rate comparison during polymerase chain reaction by DNA polymerase. *Molecular Biology International*, 2014, pp.1–8. https://doi.org/10.1155/2014/287430.

Mishra, P., Kumar, A., Nagireddy, A., Mani, D.N., Shukla, A.K., Tiwari, R. and Sundaresan, V., 2015. DNA barcoding: an efficient tool to overcome authentication challenges in the herbal market. *Plant Biotechnology Journal*, [online] 14(1), pp.8–21. https://doi.org/10.1111/pbi.12419.

Moonlight, P.W., Ardi, W.H., Padilla, L.A., Chung, K.F., Fuller, D., Girmansyah, D., Hollands, R., Jara-Muñoz, A., Kiew, R., Leong, W.C., Liu, Y., Mahardika, A., Marasinghe, L.D.K., O'Connor, M., Peng, C.I., Pérez, Á.J., Phutthai, T., Pullan, M., Rajbhandary, S. and Reynel, C., 2018. Dividing and conquering the fastest-growing genus: Towards a natural sectional classification of the mega-diverse genus *Begonia* (Begoniaceae). *Taxon*, 67(2), pp.267–323. https://doi.org/10.12705/672.3.

Moonlight, P.W., Richardson, J.A., Tebbitt, M.C., Thomas, D., Hollands, R., Peng, C.-I. and Hughes, M., 2015. Continental-scale diversification patterns in a megadiverse genus: the biogeography of Neotropical *Begonia*. *Journal of Biogeography*, 42(6), pp.1137–1149. https://doi.org/10.1111/jbi.12496.

Neale, S., Goodall-Copestake, W. and Kidner, C.A., 2006. The evolution of diversity in *Begonia*. *Floriculture, ornamental and plant biotechnology*, 4, pp.606-611.

Naive, M.A.K., Calimbo, L.G.L., Cudal, M.G., Alejandro, G.J.D. and Yu, W.B., 2022. Taxonomy of the genus *Begonia* (Begoniaceae) in Mindanao, Philippines IV: *Begonia bangsamoro* subsp. *bagasa* (*Begonia* section *Petermannia*), a new subspecies from Zamboanga del Sur. *Phytotaxa*, 559(1), pp.88–94. https://doi.org/10.11646/phytotaxa.559.1.10.

New England Biolabs, 2017. [online] *Neb.com*. Available at: <https://www.neb.com/en/tools-and-resources/feature-articles/anatomy-of-a-polymerase-how-structure-effects-function?srsltid=AfmBOop6uqPbt2zQz0_pkdoE33hUeSCdRugc3HPWGPXTU8S4IfoWxlGL> [Accessed 5 September 2025].

Oyola, S.O., Otto, T.D., Gu, Y., Maslen, G., Manske, M., Campino, S., Turner, D.J., MacInnis, B., Kwiatkowski, D.P., Swerdlow, H.P. and Quail, M.A., 2012. Optimizing illumina next-generation sequencing library preparation for extremely at-biased genomes. *BMC Genomics*, 13(1), p.1. https://doi.org/10.1186/1471-2164-13-1.

Padial, J.M., Miralles, A., De la Riva, I. and Vences, M., 2010. The integrative future of taxonomy. *Frontiers in Zoology*, 7(1), p.16. https://doi.org/10.1186/1742-9994-7-16.

Pappalardo, P., Hemmi, J.M., Machida, R.J., Leray, M., Collins, A.G. and Osborn, K.J., 2025. Taxon‐specific BLAST percent identity thresholds for identification of unknown sequences using metabarcoding. *Methods in Ecology and Evolution*. https://doi.org/10.1111/2041-210x.70147.

Permata, D.A. and Susandarini, R., 2022. Morphological diversity and phenetic relationship of wild and cultivated *Begonia* based on morphology and leaf venation. *Biodiversitas Journal of Biological Diversity*, 23(2). https://doi.org/10.13057/biodiv/d230235.

Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S., Kamoun, S., Sumlin, W.D. & Vogler, A.P., 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, 55, 595–609.

Puillandre, N., Lambert, A., Brouillet, S. & Achaz, G., 2012. ABGD, automatic barcode gap discovery for primary species delimitation. *Molecular Ecology*, 21, 1864–1877.

Puillandre, N., Brouillet, S. & Achaz, G., 2021. ASAP: assemble species by automatic partitioning. *Molecular Ecology Resources*, 21: 609-62

Radbouchoom, S., Marjorie, Phutthai, T. and Schneider, H., 2024. Towards zero extinction—A case study focusing on the plant genus *Begonia* in Thailand. *Integrative Conservation*. https://doi.org/10.1002/inc3.67.

Radbouchoom, S., delos Angeles, M.D., Ngarega, B.K., Phutthai, T. and Schneider, H., 2025. Forecasting habitat suitability of tropical karst plants in a warmer world — Thailand's *Begonia* diversity as a key example. *Frontiers in Plant Science*, 16. https://doi.org/10.3389/fpls.2025.1496040.

Rajbhandary, S., 2015. Traditional uses of *Begonia* species (Begoniaceae) in Nepal. *Journal of Natural History Museum*, 27, pp.25–34. https://doi.org/10.3126/jnhm.v27i0.14150.

Rajewska, M., Wegrzyn, K. and Konieczny, I., 2012. AT-rich region and repeated sequences – the essential elements of replication origins of bacterial replicons. *FEMS Microbiology Reviews*, 36(2), pp.408–434. https://doi.org/10.1111/j.1574-6976.2011.00300.x.

Rjosk, A., Neinhuis, C. and Lautenschläger, T., 2022. Anatomy and biomechanics of peltate *Begonia* leaves—comparative case studies. *Plants*, 11(23), p.3297. https://doi.org/10.3390/plants11233297.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A. and Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61(3), pp.539–542. https://doi.org/10.1093/sysbio/sys029.

Ross, H.A., Murugan, S. and Li, S.W.L., 2008. Testing the reliability of genetic methods of species identification via simulation. *Systematic Biology*, 57(2), pp.216–230. https://doi.org/10.1080/10635150802032990.

Sang, J., and Kiew, R., 2014. Diversity of *Begonia* (Begoniaceae) in Borneo – how many species are there? *Reinwardtia*, 14(1), pp.233−236. Available at: <https://www.mybis.gov.my/pd/281> [Accessed 5 September 2025].

Satyanti, A., 2012. The natural habitat of *Begonia* on the island of Bali: microclimate preference in two forest sites in Bedugul. *Biotropia*, 19(2).

Schleifer, K.-H. and Ludwig, W., 1994. Molecular taxonomy: classification and identification. *Bacterial Diversity and Systematics*, pp.1–15. https://doi.org/10.1007/978-1-4615-1869-3_1.

Schlichting, C.D., 1986. The evolution of phenotypic plasticity in plants. *Annual review of ecology and systematics*, pp.667-693.

Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., Chen, W., Bolchacova, E., Voigt, K., Crous, P.W., Miller, A.N., Wingfield, M.J., Aime, M.C., An, K.-D. ., Bai, F.-Y. ., Barreto, R.W., Begerow, D., Bergeron, M.-J. ., Blackwell, M. and Boekhout, T., 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences*, 109(16), pp.6241–6246. https://doi.org/10.1073/pnas.1117018109.

Shaw, J., Lickey, E.B., Schilling, E.E. and Small, R.L., 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany*, 94(3), pp.275–288. https://doi.org/10.3732/ajb.94.3.275.

Simpson, M.G., 2019. *Plant systematics*. Academic press.

Stayton, C.T., 2015. What does convergent evolution mean? The interpretation of convergence and its implications in the search for limits to evolution. *Interface focus*, [online] 5(6), p.20150039. https://doi.org/10.1098/rsfs.2015.0039.

Siregar, H.M., Sudarmono, Witono, J.R. and Wawangningrum, H., 2021. Pollen morphology of six Indonesian *Begonia* (Begoniaceae) Species. *Annual Research & Review in Biology*, pp.7–13. https://doi.org/10.9734/arrb/2021/v36i530370.

Stace, C.A., 2005. Plant taxonomy and biosystematics — does DNA provide all the answers? *TAXON*, 54(4), pp.999–1007. https://doi.org/10.2307/25065484.

Stern, D.L., 2013. The genetic causes of convergent evolution. *Nature reviews. Genetics*, [online] 14(11), pp.751–64. https://doi.org/10.1038/nrg3483.

Swofford, D.L., 2001. *PAUP*. Phylogenetic analysis using parsimony (*and other methods)*. Sinauer associates, Sunderland, Massachusetts.

Tan, J.P.C., Tam, S.M. and Kiew, R., 2018. *Begonia yenyeniae* (Begoniaceae), a new species from Endau Rompin National Park, Johor, Malaysia. *PhytoKeys*, 110, pp.23–37. https://doi.org/10.3897/phytokeys.110.25846.

Tebbitt, M.C., Lowe-Forrest, L., Santoriello, A., Clement, W.L. and Swensen, S.M., 2006. Phylogenetic relationships of Asian *Begonia*, with an emphasis on the evolution of rain-ballist and animal dispersal mechanisms in sections *Platycentrum, Sphenanthera* and *Leprosae*. *Systematic Botany*, 31(2), pp.327–336. https://doi.org/10.1600/036364406777585784.

Thaxton, C., Goldstein, J., DiStefano, M., Wallace, K., Witmer, P.D., Haendel, M.A., Hamosh, A., Rehm, H.L. and Berg, J.S., 2022. Lumping versus splitting: how to approach defining a disease to enable accurate genomic curation. *Cell Genomics*, 2(5), pp.100131–100131. https://doi.org/10.1016/j.xgen.2022.100131.

Thomas, D.C., 2010. Phylogenetics and historical biogeography of southeast Asian *Begonia* L.(Begoniaceae). PhD. University of Glasgow.

Thomas, D.C., Hughes, M., Phutthai, T., Rajbhandary, S., Rubite, R., Ardi, W.H. and Richardson, J.E., 2011. A non-coding plastid DNA phylogeny of Asian *Begonia* (Begoniaceae):
Evidence for morphological homoplasy and sectional polyphyly. *Molecular Phylogenetics and Evolution*, 60(3), pp.428–444.
https://doi.org/10.1016/j.ympev.2011.05.006.

Thompson, J.B. and Hawkins, J.A., 2024. Phylogeny and bioprospecting: The diversity of medicinal plants used in cancer management. *Plants, People, Planets*. https://doi.org/10.1002/ppp3.10566.

Tian, D., Xiao, Y., Tong, Y., Fu, N., Liu, Q. and Li, C., 2018. Diversity and conservation of Chinese wild *Begonias*. *Plant Diversity*, 40(3), pp.75–90. https://doi.org/10.1016/j.pld.2018.06.002.

Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A. and Minh, B.Q., 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research*, 44(W1), pp.W232–W235. https://doi.org/10.1093/nar/gkw256.

Tseng, Y.-H., Hsieh, C.-L., Campos-Domínguez, L., Hu, A., Chang, C., Hsu, Y.-T., Kidner, C., Hughes, M., Moonlight, P.W., Hung, C., Wang, Y.-C., Wang, Y.-T., Liu, S., Deden Girmansyah and Chung, K.-F., 2022. Insights into the evolution of the chloroplast genome and the phylogeny of *Begonia*. *Edinburgh Journal of Botany*, 79, pp.1–32. https://doi.org/10.24823/ejb.2022.408.

Wiens, J.J., 2007. Species delimitation: new approaches for discovering diversity. *Systematic Biology*, 56(6), pp.875–878. https://doi.org/10.1080/10635150701748506.

Xia, X., Xie, Z., Salemi, M., Chen, L. and Wang, Y., 2003. An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution*, 26(1), pp.1–7. https://doi.org/10.1016/s1055-7903(02)00326-3.

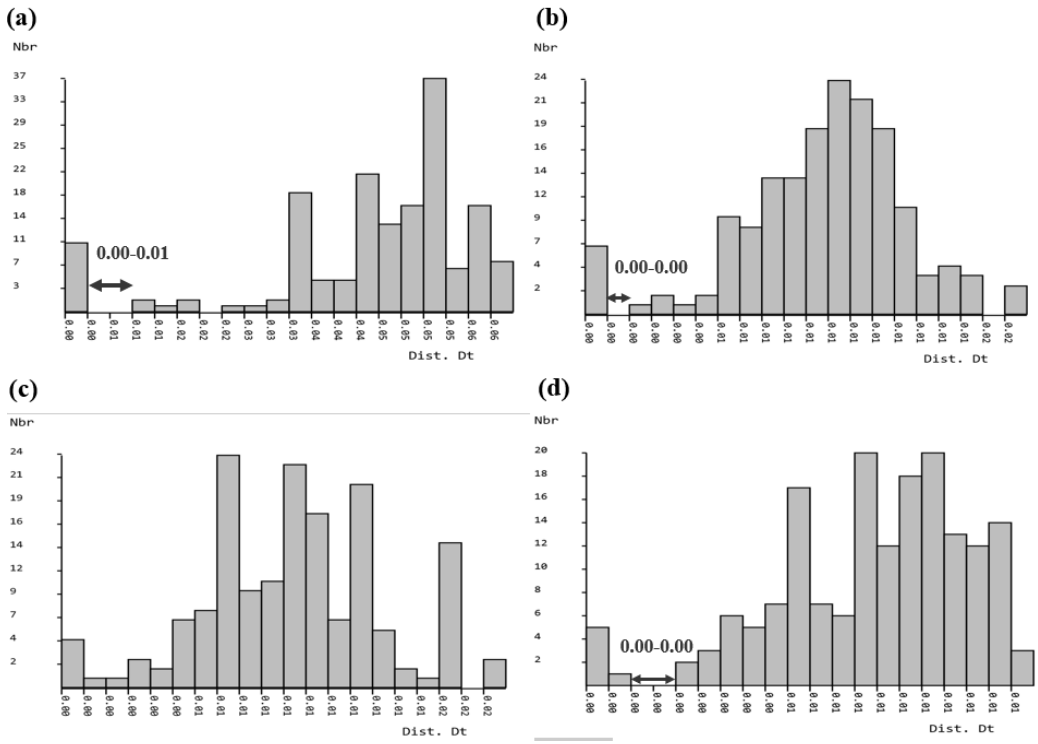
Xiong, C., Huang, Y., Li, Z., Wu, L., Liu, Z., Zhu, W., Li, J., Xu, R. and Hong, X., 2023. Comparative chloroplast genomics reveals the phylogeny and the adaptive evolution of *Begonia* in China. *BMC Genomics*, 24(1). https://doi.org/10.1186/s12864-023-09563-3.

Zhang, J., Kapli, P., Pavlidis, P. & Stamatakis, A., 2013. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, 29, 2869–2876.

Zou, Y., Zhang, Z., Zeng, Y., Hu, H., Hao, Y., Huang, S. and Li, B., 2024. Common methods for phylogenetic tree construction and their implementation in R. *Bioengineering*, 11(5), p.480. https://doi.org/10.3390/bioengineering11050480.

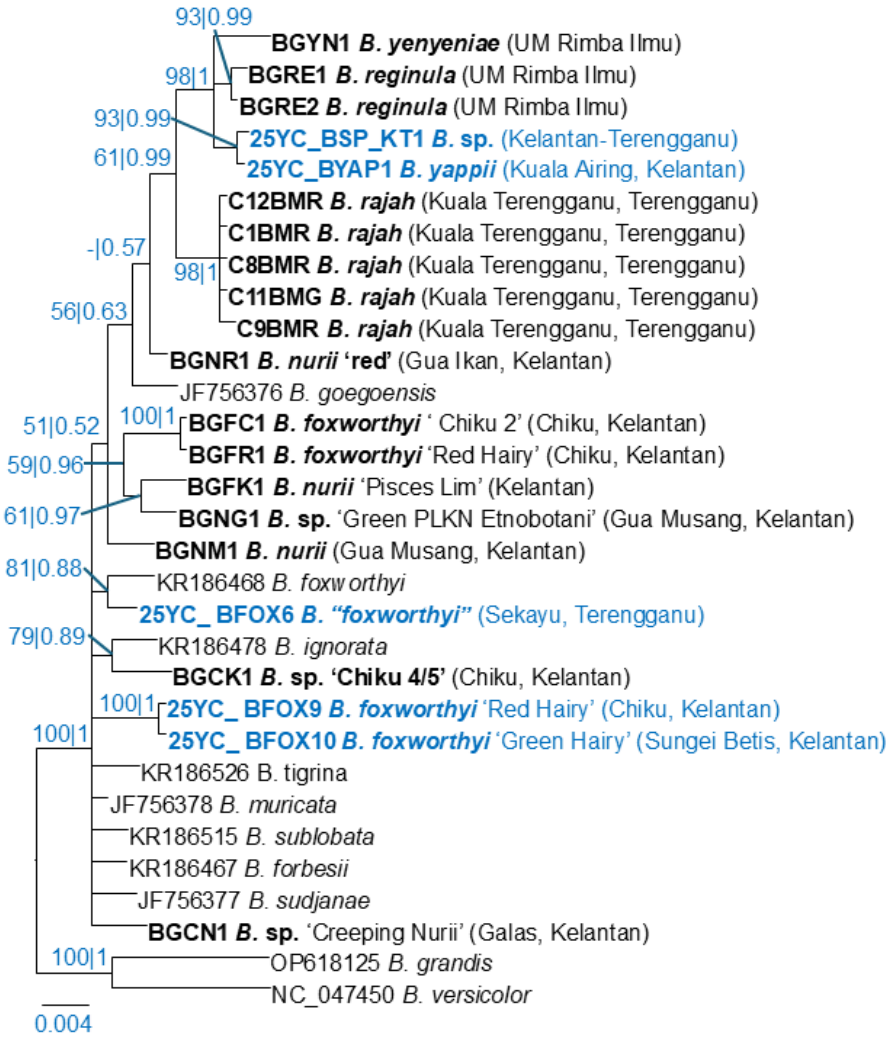
APPENDICES

APPENDIX A



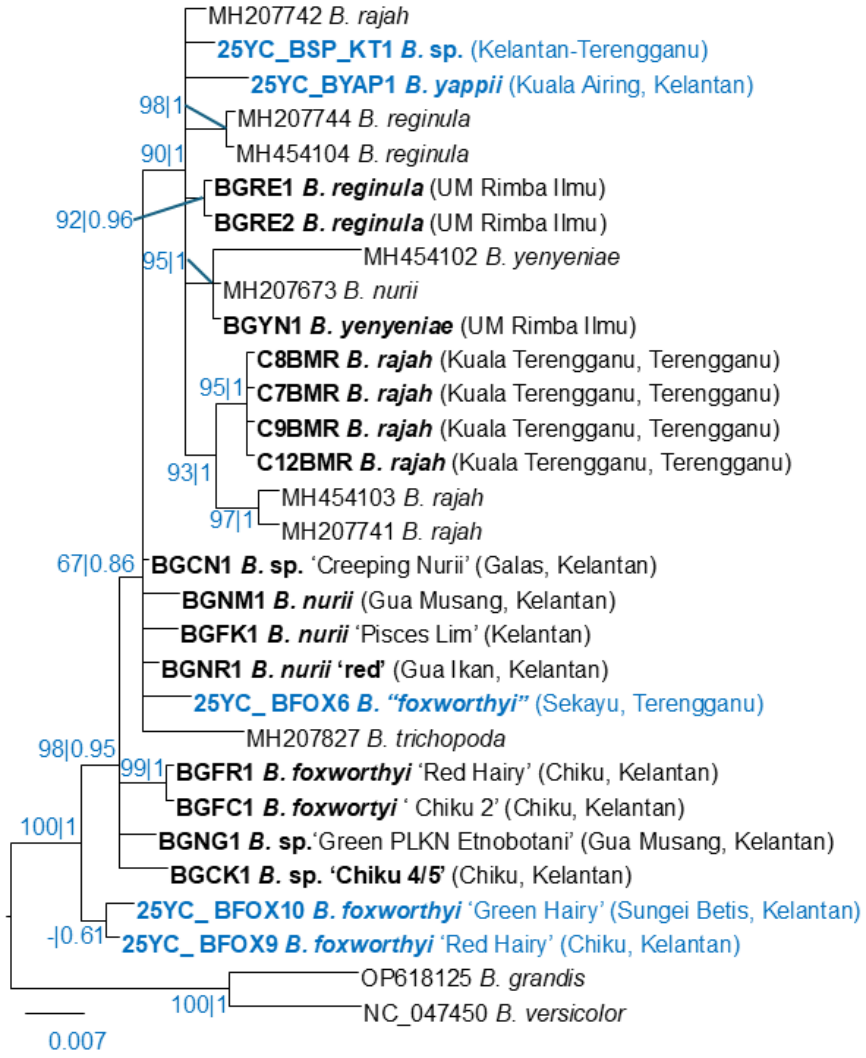
Appendix A: ASAP histogram of the species delimitation of **(a)** ITS, **(b)** *ndhA* intron, **(c)** *ndhF-rpl32*, and **(d)** *rpl32-trnL* dataset.

APPENDIX B



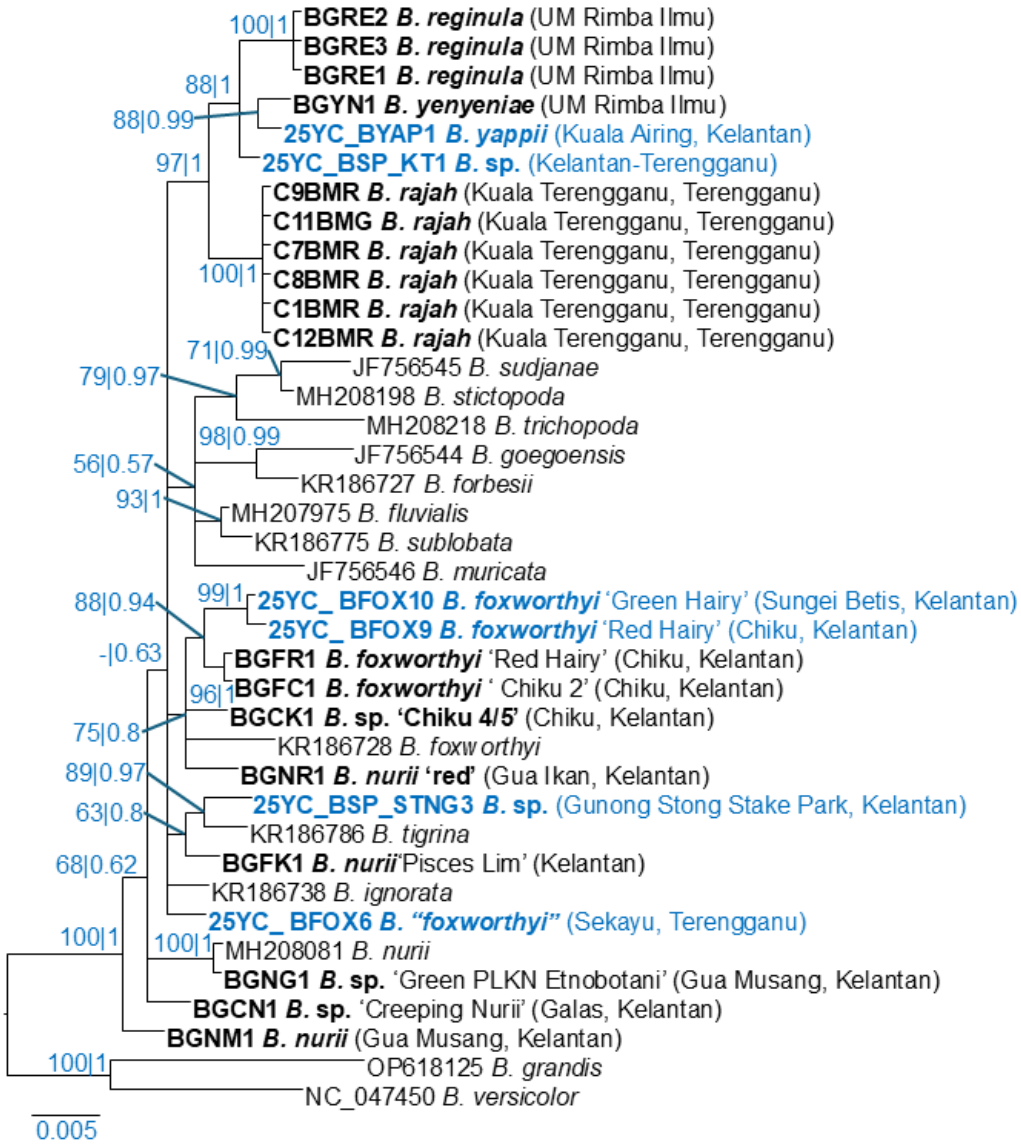
Appendix B: Bayesian tree showing phylogenetic relationship between *Begonia* species based on the *ndh*A intron dataset. Nodal values are arranged as: Ultrafast ML bootstrap support | Bayesian posterior probability. Scale bar corresponds to 0.004 nucleotide substitutions per site. Specimens in blue indicate specimens from this study. Hyphens indicate Ultrafast ML bootstrap support <50.

APPENDIX C



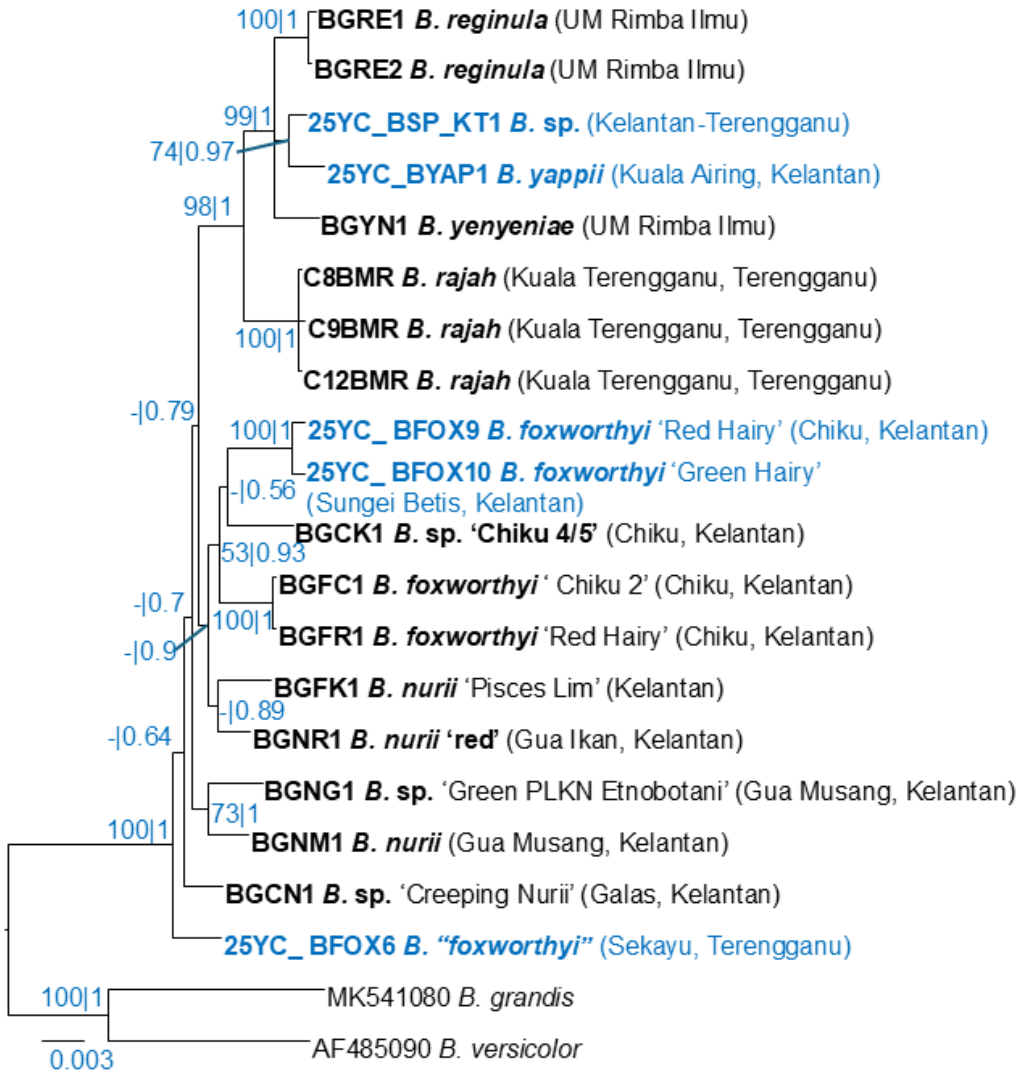
Appendix C: Bayesian tree showing phylogenetic relationship between *Begonia* species based on the *ndh*F–*rpl*32 dataset. Nodal values are arranged as: Ultrafast ML bootstrap support | Bayesian posterior probability. Scale bar corresponds to 0.007 nucleotide substitutions per site. Specimens in blue indicate specimens from this study. Hyphens indicate Ultrafast ML bootstrap support <50.

APPENDIX D



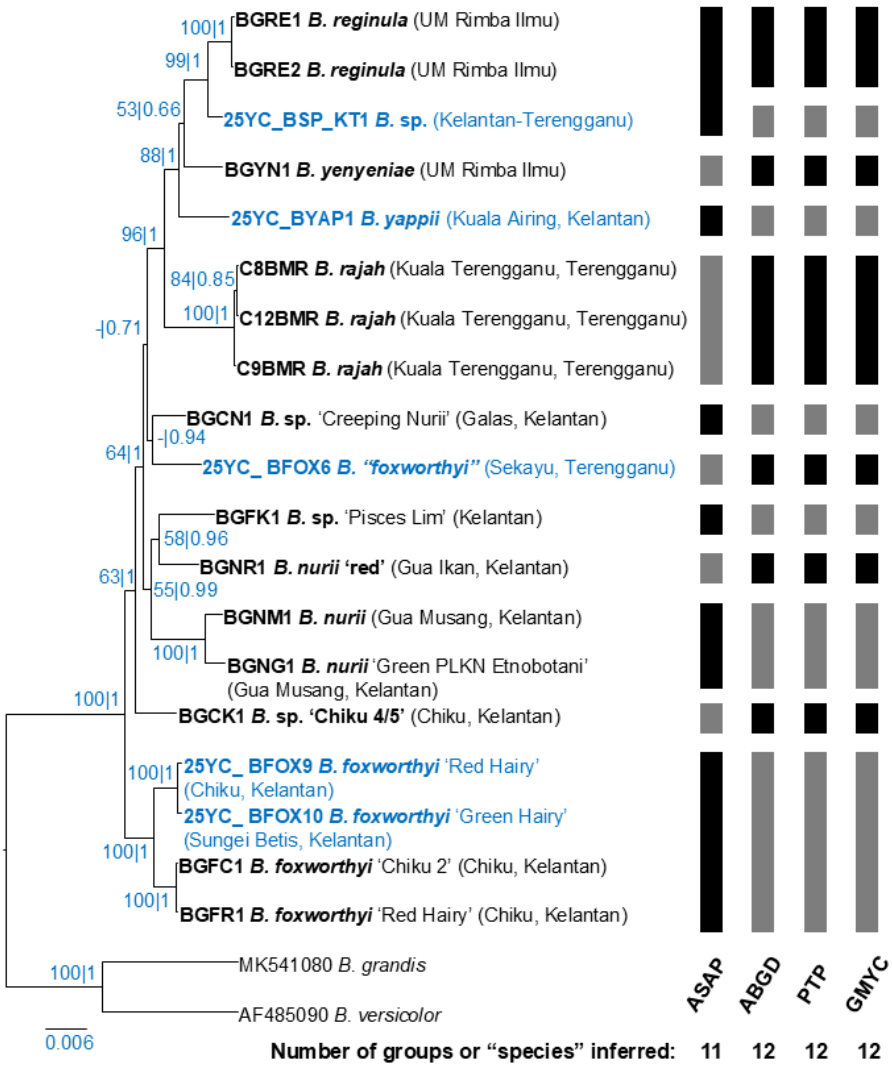
Appendix D: Bayesian tree showing phylogenetic relationship between *Begonia* species based on the *rpl*32–*trn*L dataset. Nodal values are arranged as: Ultrafast ML bootstrap support | Bayesian posterior probability. Scale bar corresponds to 0.005 nucleotide substitutions per site. Specimens in blue indicate specimens from this study. Hyphens indicate Ultrafast ML bootstrap support <50.

APPENDIX E



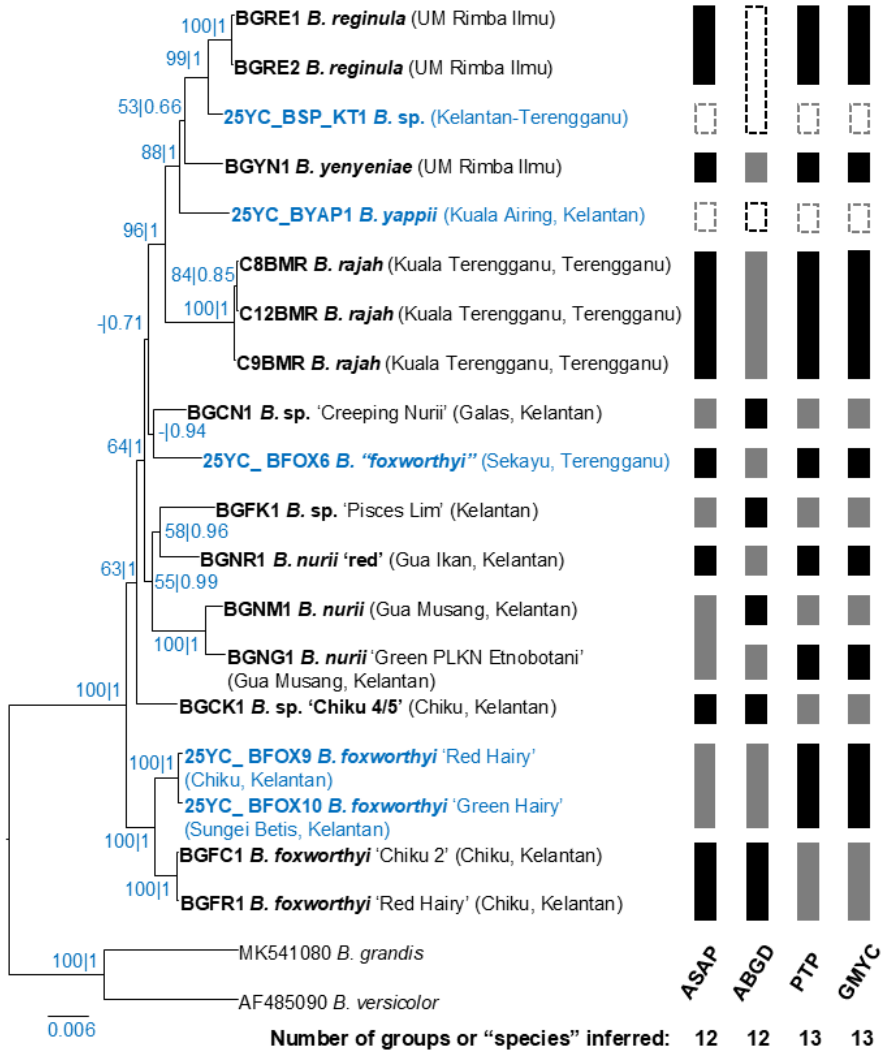
Appendix E: Bayesian tree showing phylogenetic relationship between *Begonia* species based on the chloroplast dataset. Nodal values are arranged as: Ultrafast ML bootstrap support | Bayesian posterior probability. Scale bar corresponds to 0.003 nucleotide substitutions per site. Specimens in blue indicate specimens from this study. Hyphens indicate Ultrafast ML bootstrap support <50.

APPENDIX F



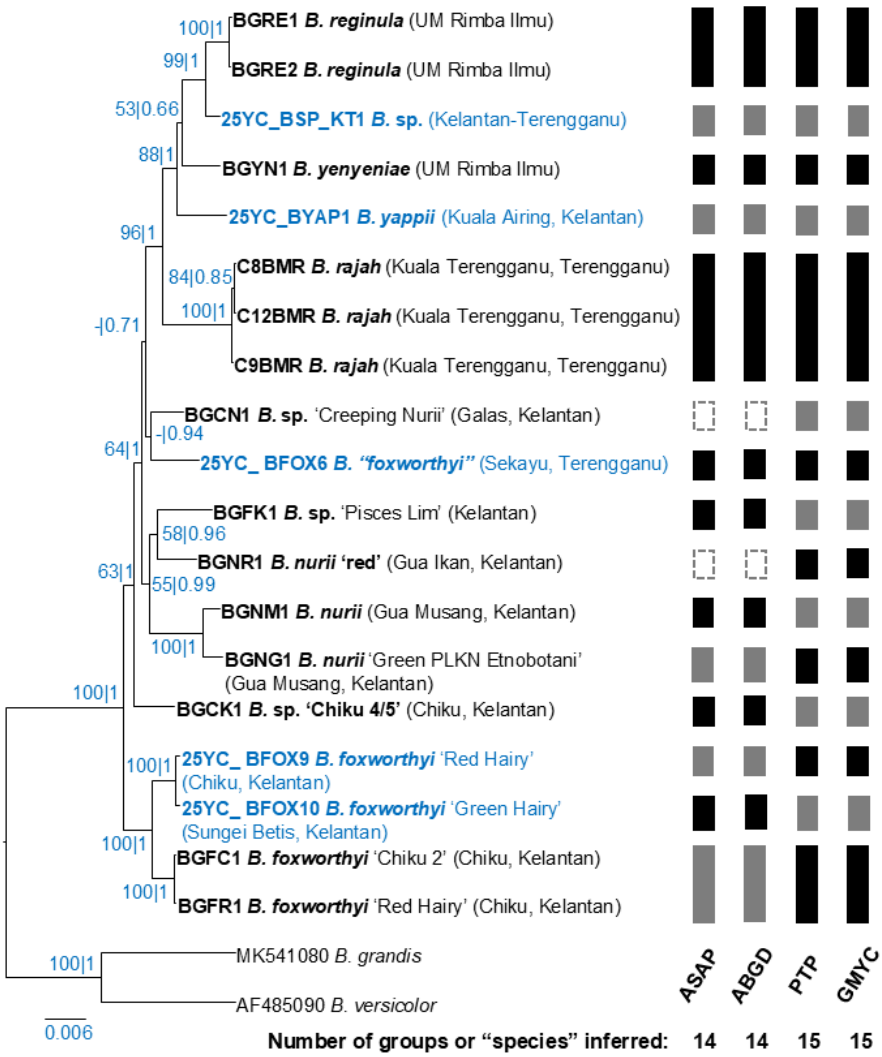
Appendix F: Bayesian tree showing phylogenetic relationship between *Begonia* species based on the ITS dataset and its species delimitation analysis using ASAP, ABGD, PTP, and MNYC. Nodal values are arranged as: Ultrafast ML bootstrap support | Bayesian posterior probability. Scale bar corresponds to 0.006 nucleotide substitutions per site. Specimens in blue indicate specimens from this study. Hyphens indicate Ultrafast ML bootstrap support <50.

APPENDIX G



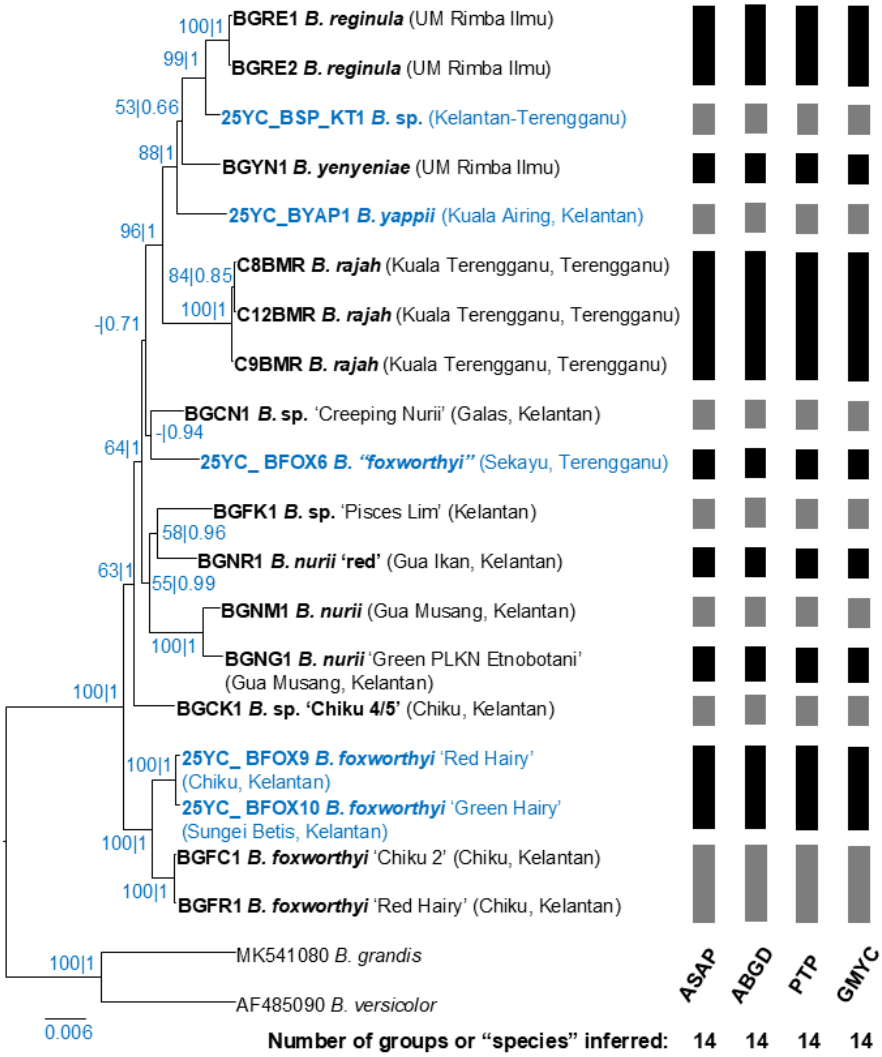
Appendix G: Bayesian tree showing phylogenetic relationship between *Begonia* species based on the *ndh*A intron dataset and its species delimitation analysis using ASAP, ABGD, PTP, and MNYC. Nodal values are arranged as: Ultrafast ML bootstrap support | Bayesian posterior probability. Scale bar corresponds to 0.006 nucleotide substitutions per site. Specimens in blue indicate specimens from this study. Hyphens indicate Ultrafast ML bootstrap support <50.

APPENDIX H



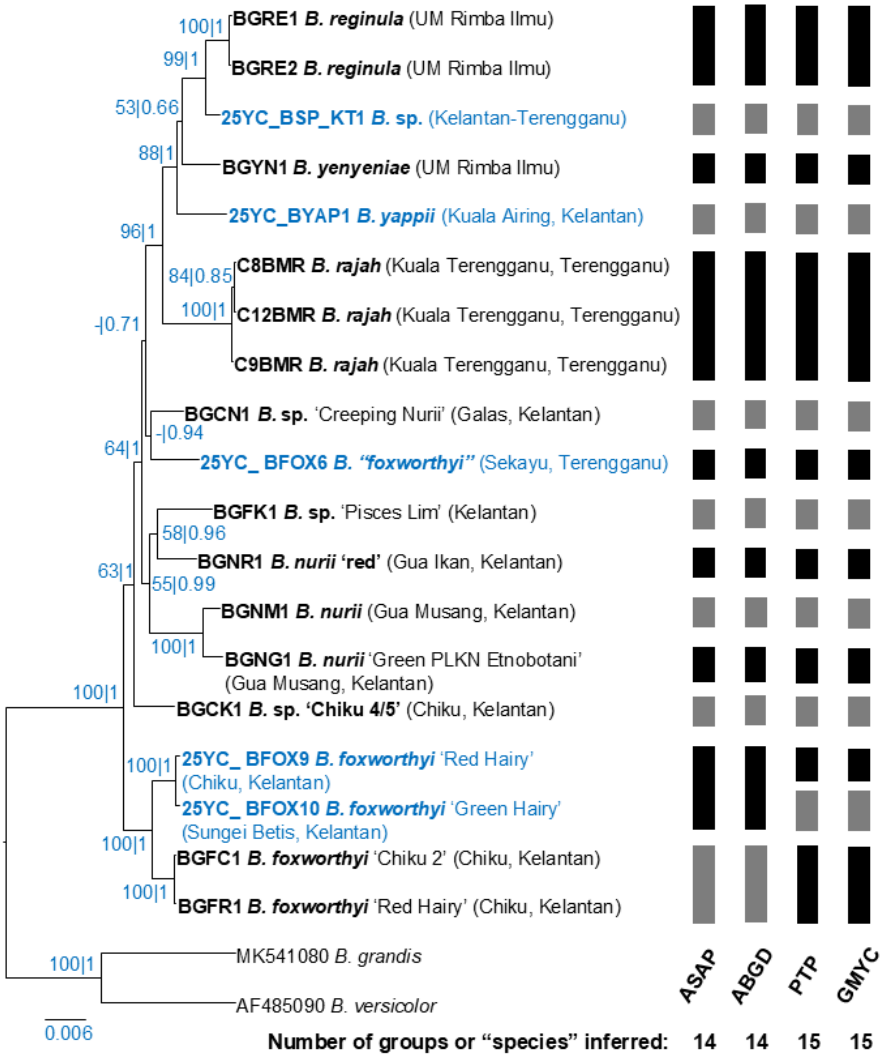
Appendix H: Bayesian tree showing phylogenetic relationship between *Begonia* species based on the *ndh*F–*rpl*32 dataset and its species delimitation analysis using ASAP, ABGD, PTP, and MNYC. Nodal values are arranged as: Ultrafast ML bootstrap support | Bayesian posterior probability. Scale bar corresponds to 0.006 nucleotide substitutions per site. Specimens in blue indicate specimens from this study. Hyphens indicate Ultrafast ML bootstrap support <50.

APPENDIX I



Appendix I: Bayesian tree showing phylogenetic relationship between *Begonia* species based on the *rpl*32–*trn*L dataset and its species delimitation analysis using ASAP, ABGD, PTP, and MNYC. Nodal values are arranged as: Ultrafast ML bootstrap support | Bayesian posterior probability. Scale bar corresponds to 0.006 nucleotide substitutions per site. Specimens in blue indicate specimens from this study. Hyphens indicate Ultrafast ML bootstrap support <50

APPENDIX J



Appendix J: Bayesian tree showing phylogenetic relationship between *Begonia* species based on the chloroplast dataset and its species delimitation analysis using ASAP, ABGD, PTP, and MNYC. Nodal values are arranged as: Ultrafast ML bootstrap support | Bayesian posterior probability. Scale bar corresponds to 0.006 nucleotide substitutions per site. Specimens in blue indicate specimens from this study. Hyphens indicate Ultrafast ML bootstrap support <50

APPENDIX K

Appendix K: Genetic distance (%) between *Begonia* species based on ITS dataset.

No	Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	**BGFK1 *B.* sp. 'Pisces Lim**																												
2	**BGYN1 *B. yenyeniae***	3.34																											
3	**BGNR1 *B. nurii* 'red'**	3.97	3.73																										
4	**25YC_BYAP1 *B. yappii***	5.24	3.85	4.35																									
5	**BGRE2 *B. reginula***	4.74	3.47	5.00	4.87																								
6	**BGRE1 *B. reginula***	4.74	3.47	5.00	4.87	0.00																							
7	**BGRE3 *B. reginula***	4.61	3.35	4.87	4.74	0.13	0.13																						
8	**25YC_BSP_KT1 *B.* sp. 'Kelantan-Terengganu'**	4.48	3.35	4.23	4.10	1.03	1.03	0.90																					

Appendix K: Genetic distance (%) between *Begonia* species based on ITS dataset (continued).

No	Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
9	**25YC_ BFOX6 *B. "foxworthyi"***	3.84	3.09	4.48	4.49	1.79	1.79	1.67	1.54																				
10	**C9BMR *B. rajah***	5.19	5.11	5.20	5.91	4.66	4.66	4.52	4.12	4.39																			
11	**C7BMR *B. rajah***	5.19	5.11	5.20	5.91	4.66	4.66	4.52	4.12	4.39	0.00																		
12	**C11BMR *B. rajah***	5.19	5.11	5.20	5.91	4.66	4.66	4.52	4.12	4.39	0.00	0.00																	
13	**C8BMR *B. rajah***	5.19	5.11	5.20	5.91	4.66	4.66	4.52	4.12	4.39	0.00	0.00	0.00																
14	**C12BMR *B. rajah***	5.19	5.11	5.20	5.91	4.66	4.66	4.52	4.12	4.39	0.00	0.00	0.00	0.00															
15	**C1BMR *B. rajah***	5.19	5.11	5.20	5.91	4.66	4.66	4.52	4.12	4.39	0.00	0.00	0.00	0.00	0.00														
16	**25YC_ BFOX9 *B. foxworthyi* 'Red Hairy'**	5.12	4.37	4.87	4.74	5.00	5.00	4.87	4.74	4.10	5.62	5.62	5.62	5.62	5.62	5.62													
17	**25YC_ BFOX10 *B. foxworthyi* 'Green Hairy'**	5.12	4.37	4.87	4.74	5.00	5.00	4.87	4.74	4.10	5.62	5.62	5.62	5.62	5.62	5.62	0.00												
18	**BGFC1 *B. foxworthyi* 'Chiku 2'**	5.12	4.37	4.87	4.74	5.00	5.00	4.87	4.74	4.10	5.62	5.62	5.62	5.62	5.62	5.62	0.00	0.00											

Appendix K: Genetic distance (%) between *Begonia* species based on ITS dataset (continued).

No	Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
19	BGFR1 *B. foxworthyi* 'Red Hairy'	5.12	4.37	4.87	4.74	5.00	5.00	4.87	4.74	4.10	5.62	5.62	5.62	5.62	5.62	5.62	0.00	0.00	0.00										
20	BGCN1 *B.* sp. 'Creeping Nurii'	3.71	3.21	3.84	3.97	3.21	3.21	3.08	3.21	3.07	4.54	4.54	4.54	4.54	4.54	4.54	3.46	3.46	3.46	3.46									
21	BGCK1 *B.* sp. 'Chiku 4/5'	4.10	3.21	3.20	3.72	3.34	3.34	3.21	2.82	3.46	4.13	4.13	4.13	4.13	4.13	4.13	3.34	3.34	3.34	3.34	2.44								
22	BGNG1 *B. nurii* 'Green PLKN Etnobotani'	5.66	5.04	5.02	5.79	5.29	5.29	5.41	5.03	5.28	5.92	5.92	5.92	5.92	5.92	5.92	5.15	5.15	5.15	5.15	4.51	4.38							
23	AF485138 *B. goegoensis*	8.12	6.42	7.58	7.33	7.06	7.06	6.93	6.79	6.93	7.70	7.70	7.70	7.70	7.70	7.70	7.74	7.74	7.74	7.74	6.13	5.88	8.70						
24	BGNM1 *B. nurii*	5.90	5.26	5.23	6.04	5.52	5.52	5.65	5.25	5.52	6.20	6.20	6.20	6.20	6.20	6.20	5.37	5.37	5.37	5.37	4.70	4.56	0.00	8.78					
25	AF485136 *B. rajah*	5.63	4.98	6.16	6.02	2.28	2.28	2.41	2.00	3.07	6.05	6.05	6.05	6.05	6.05	6.05	6.43	6.43	6.43	6.43	4.55	4.81	6.47	8.37	6.38				
26	AY753725 *B. muricata*	8.58	7.92	7.87	8.31	7.05	7.05	6.91	7.19	7.33	8.26	8.26	8.26	8.26	8.26	8.26	8.42	8.42	8.42	8.42	7.05	6.64	9.33	5.97	9.34	8.30			
27	KF636491 *B. yappii*	6.64	5.47	5.63	1.33	6.62	6.62	6.48	5.73	6.03	7.03	7.03	7.03	7.03	7.03	7.03	6.35	6.35	6.35	6.35	5.31	4.91	7.00	9.05	7.01	6.75	9.86		
28	JX656702 *B. foxworthyi*	5.67	5.14	5.41	5.69	4.86	4.86	4.72	4.30	4.03	5.85	5.85	5.85	5.85	5.85	5.85	5.67	5.67	5.67	5.67	4.44	4.45	6.27	8.38	6.29	5.73	8.60	6.46	
29	JX656703 *B. tigrina*	5.27	4.59	5.14	5.27	4.31	4.31	4.45	4.30	4.59	5.68	5.68	5.68	5.68	5.68	5.68	5.39	5.39	5.39	5.39	3.47	3.76	5.44	7.67	5.45	4.89	8.47	5.76	5.84

APPENDIX L

Appendix L: Genetic distance (%) between *Begonia* species based on *ndh*A intron dataset.

No.	Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	**BGRE1 *B. reginula***																												
2	**BGRE2 *B. reginula***	0.08																											
3	**BGYN1 *B. yenyeniae***	0.59	0.51																										
4	**C12BMR *B. rajah***	0.76	0.68	1.02																									
5	**C1BMR *B. rajah***	0.76	0.68	1.02	0.00																								
6	**C8BMR *B. rajah***	0.76	0.68	1.02	0.00	0.00																							
7	**C11BMR *B. rajah***	0.76	0.68	1.02	0.00	0.00	0.00																						
8	**C9BMR *B. rajah***	0.84	0.76	1.10	0.08	0.08	0.08	0.08																					

Appendix L: Genetic distance (%) between *Begonia* species based on *ndh*A intron dataset (continued).

No.	Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
9	**KR186526 *B. tigrina***	1.18	1.10	1.44	0.93	0.93	0.93	0.93	0.85																				
10	**JF756378 *B. muricata***	0.93	0.85	1.19	0.68	0.68	0.68	0.68	0.59	0.42																			
11	**KR186515 *B. sublobata***	1.10	1.02	1.36	0.85	0.85	0.85	0.85	0.76	0.59	0.34																		
12	**KR186467 *B. forbesii***	1.10	1.02	1.28	0.77	0.77	0.77	0.85	0.68	0.60	0.34	0.51																	
13	**KR186468 *B. foxworthyi***	1.27	1.19	1.53	1.02	1.02	1.02	1.02	0.93	0.76	0.51	0.68	0.68																
14	**JF756377 *B. sudjanae***	1.19	1.10	1.44	0.93	0.93	0.93	0.93	0.85	0.68	0.43	0.60	0.51	0.77															
15	**BGCN1 *B.* sp. 'Creeping Nurii'**	1.27	1.18	1.52	1.02	1.02	1.02	1.02	0.93	0.76	0.51	0.68	0.68	0.85	0.77														
16	**BGFC1 *B. foxworthyi* 'Chiku 2'**	1.10	1.02	1.53	1.02	1.02	1.02	1.02	1.10	0.93	0.68	0.85	0.77	1.02	0.94	1.02													
17	**BGFR1 *B. foxworthyi* 'Red Hairy'**	1.10	1.02	1.53	1.02	1.02	1.02	1.02	1.10	0.93	0.68	0.85	0.77	1.02	0.94	1.02	0.00												
18	**BGFK1 *B.* sp. 'Pisces Lim'**	1.10	1.01	1.52	1.02	1.02	1.02	1.02	1.10	0.93	0.59	0.85	0.77	1.02	0.93	1.01	0.85	0.85											

Appendix L: Genetic distance (%) between *Begonia* species based on *ndh*A intron dataset (continued).

No.	Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
19	**BGNG1 *B. nurii* 'Green PLKN Etnobotani'**	1.02	0.93	1.44	0.93	0.93	0.93	0.93	1.02	0.85	0.59	0.77	0.77	0.93	0.85	0.93	0.76	0.76	0.59										
20	**BGNM1 *B. nurii***	1.10	1.01	1.36	0.85	0.85	0.85	0.85	0.93	0.76	0.51	0.68	0.60	0.85	0.77	0.85	0.85	0.85	0.85	0.59									
21	**KR186478 *B. ignorata***	1.18	1.10	1.44	0.93	0.93	0.93	0.93	0.85	0.85	0.51	0.68	0.68	0.85	0.77	0.85	1.02	1.02	1.02	0.93	0.85								
22	**BGNR1 *B. nurii* 'red'**	0.68	0.59	0.93	0.59	0.59	0.59	0.59	0.68	0.68	0.42	0.59	0.60	0.76	0.68	0.76	0.76	0.76	0.59	0.51	0.59	0.68							
23	**JF756376 *B. goegoensis***	0.93	0.85	1.19	0.85	0.85	0.85	0.85	0.77	0.77	0.51	0.68	0.51	0.85	0.77	0.85	1.02	1.02	1.02	0.94	0.85	0.76	0.43						
24	**25YC_ BFOX9 *B. foxworthyi* 'Red Hairy'**	1.42	1.33	1.69	1.16	1.16	1.16	1.16	1.06	0.89	0.62	0.81	0.80	0.98	0.89	0.89	1.16	1.16	1.16	1.08	0.99	0.98	0.90	0.80					
25	**25YC_ BFOX10 *B. foxworthyi* 'Green Hairy'**	1.42	1.33	1.69	1.16	1.16	1.16	1.16	1.06	0.89	0.62	0.81	0.80	0.98	0.89	0.89	1.16	1.16	1.16	1.08	0.99	0.98	0.90	0.80	0.00				
26	**25YC_BSP_KT1 *B.* sp. 'Kelantan -Terengganu'**	0.37	0.28	0.55	0.74	0.74	0.74	0.83	0.64	1.11	0.84	0.93	0.84	1.21	1.03	1.11	1.31	1.31	1.30	1.31	1.13	1.11	0.76	0.84	1.28	1.28			
27	**BGCK1 *B.* sp. 'Chiku 4/5'**	1.46	1.36	1.73	1.19	1.19	1.19	1.19	1.09	0.91	0.64	0.73	0.83	1.01	0.92	0.91	1.20	1.20	1.19	1.11	1.02	0.74	0.92	1.01	1.09	1.09	1.28		

Appendix L: Genetic distance (%) between *Begonia* species based on *ndh*A intron dataset (continued).

No.	Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
28	**25YC_ BFOX6 *B. "foxworthyi"***	1.17	1.08	1.44	0.90	0.90	0.90	0.90	0.81	0.63	0.36	0.45	0.55	0.54	0.64	0.63	0.91	0.91	0.91	0.82	0.73	0.72	0.64	0.73	0.81	0.81	1.01	0.82	
29	**25YC_BYAP1 *B. yappii***	0.26	0.18	0.53	0.71	0.71	0.71	0.71	0.62	0.97	0.71	0.89	0.80	1.06	0.97	0.97	1.24	1.24	1.23	1.15	1.06	0.97	0.62	0.62	1.07	1.07	0.00	1.18	0.90

APPENDIX M

Appendix M: Genetic distance (%) between *Begonia* species based on *ndh*F–*rpl*32 dataset.

No.	Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	MH454102 B. yenyeniae																											
2	MH207673 B. nurii	1.67																										
3	BGFR1 *B. foxworthyi* 'Red Hairy'	2.95	1.24																									
4	BGFC1 *B. foxworthyi* 'Chiku 2'	2.95	1.24	0.00																								
5	BGNG1 *B. nurii* 'Green PLKN Etnobotani'	2.68	0.97	0.86	0.86																							
6	BGCN1 *B.* sp. 'Creeping Nurii'	2.40	0.69	0.61	0.61	0.49																						
7	BGNM1 *B. nurii*	2.68	0.97	0.99	0.99	0.62	0.37																					

Appendix M: Genetic distance (%) between *Begonia* species based on *ndh*F–*rpl*32 dataset (continued).

No.	Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
8	BGFK1 *B.* sp. 'Pisces Lim'	2.40	0.69	0.98	0.98	0.86	0.37	0.74																				
9	BGNR1 *B. nurii* 'red'	2.40	0.69	0.74	0.74	0.61	0.12	0.50	0.49																			
10	25YC_ BFOX6 *B. "foxworthyi"*	2.96	1.25	1.10	1.10	0.98	0.49	0.86	0.86	0.62																		
11	25YC_ BFOX10 *B. foxworthyi* 'Green Hairy'	2.81	1.11	1.12	1.12	1.12	0.74	1.00	0.87	0.63	1.24																	
12	BGCK1 *B.* sp. 'Chiku 4/5'	2.81	1.10	0.98	0.98	0.86	0.73	1.11	0.98	0.74	0.98	1.11																
13	25YC_BYAP1 *B. yappii*	2.52	0.84	1.74	1.74	1.47	1.08	1.21	1.34	1.21	1.62	1.62	1.61															
14	25YC_BSP_KT1 *B.* sp. 'Kelantan-Terengganu'	1.96	0.42	1.33	1.33	1.06	0.66	0.93	0.93	0.79	1.07	1.21	1.20	0.81														
15	C8BMR *B. rajah*	2.52	0.84	1.60	1.60	1.32	0.93	1.20	1.19	1.06	1.21	1.60	1.59	1.08	0.68													
16	C7BMR *B. rajah*	2.52	0.84	1.60	1.60	1.32	0.93	1.20	1.19	1.06	1.21	1.60	1.59	1.08	0.68	0.00												
17	C9BMR *B. rajah*	2.52	0.84	1.60	1.60	1.32	0.93	1.20	1.19	1.06	1.21	1.60	1.59	1.08	0.68	0.00	0.00											

Appendix M: Genetic distance (%) between *Begonia* species based on *ndh*F–*rpl*32 dataset (continued).

No.	Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
18	C12BMR *B. rajah*	2.52	0.84	1.61	1.61	1.33	0.93	1.20	1.20	1.07	1.21	1.61	1.59	1.09	0.69	0.00	0.00	0.00										
19	BGYN1 *B. yenyeniae*	1.53	0.00	1.21	1.21	0.94	0.54	0.81	0.80	0.67	1.09	1.09	1.08	0.82	0.41	0.69	0.69	0.69	0.69									
20	BGRE1 *B. reginula*	1.96	0.42	0.95	0.95	0.67	0.55	0.82	0.81	0.68	1.23	1.22	0.94	0.82	0.55	0.69	0.69	0.69	0.69	0.27								
21	BGRE2 *B. reginula*	1.97	0.42	0.97	0.97	0.69	0.56	0.84	0.69	0.70	1.26	1.26	0.97	0.84	0.56	0.71	0.71	0.71	0.71	0.28	0.00							
22	25YC_ BFOX9 *B. foxworthyi* 'Red Hairy'	2.97	1.25	1.02	1.02	1.02	0.89	1.15	0.89	0.77	1.39	0.38	1.02	1.81	1.38	1.79	1.79	1.79	1.79	1.25	1.11	1.11						
23	MH207827 *B. trichopoda*	3.25	1.70	1.66	1.66	1.38	1.11	1.39	1.11	1.11	1.67	1.66	1.65	2.11	1.70	2.12	2.12	2.12	2.12	1.58	1.71	1.70	1.80					
24	MH454103 *B. rajah*	2.81	1.12	1.93	1.93	1.65	1.24	1.52	1.38	1.38	1.25	1.94	1.92	1.54	1.13	0.84	0.84	0.84	0.84	0.99	0.99	0.99	2.07	2.40				
25	MH207741 *B. rajah*	2.61	0.87	1.84	1.84	1.70	1.28	1.57	1.42	1.42	1.15	1.73	1.71	1.30	0.88	0.57	0.57	0.57	0.57	1.02	1.03	1.02	1.86	2.20	0.29			
26	MH207742 *B. rajah*	1.86	0.31	1.20	1.20	1.08	0.62	0.92	0.76	0.78	1.24	1.25	1.21	0.89	0.47	0.62	0.62	0.62	0.62	0.31	0.30	0.30	1.37	1.71	1.06	1.07		
27	MH207744 *B. reginula*	2.33	0.73	1.43	1.43	1.29	0.86	1.16	1.00	1.00	1.59	1.45	1.58	1.17	0.59	0.89	0.89	0.89	0.89	0.58	0.58	0.59	1.73	2.09	1.33	1.32	0.48	
28	MH454104 *B. reginula*	2.33	0.73	1.60	1.60	1.29	0.86	1.17	1.01	1.01	1.44	1.45	1.45	1.19	0.59	0.89	0.89	0.89	0.89	0.59	0.59	0.59	1.74	2.09	1.04	1.05	0.49	0.00

APPENDIX N

Appendix N: Genetic distance (%) between *Begonia* species based on *rpl*32–*trn*L dataset.

No.	Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
1	**JF756545 *B. sudjanae***																																			
2	**JF756544 *B. goegoensis***	1.74																																		
3	**JF756546 *B. muricata***	1.63	1.83																																	
4	**YCL25_STNG3_RT *B.* sp. 'Gunong Stong State Park'**	1.20	1.51	1.40																																
5	**MH208198 *B. stictopoda***	0.47	1.51	1.31	0.76																															
6	**MH208218 *B. trichopoda***	1.41	1.99	1.88	1.42	1.08																														

Appendix N: Genetic distance (%) between *Begonia* species based on *rpl*32–*trn*L dataset (continued).

No.	Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
7	**MH207975** *B. fluvialis*	0.84	1.22	0.94	0.57	0.54	1.27																														
8	**MH208081** *B. nurii*	1.31	1.69	1.40	0.84	0.99	1.53	0.81																													
9	**C9BMR** *B. rajah*	1.54	1.56	1.47	1.01	1.17	1.73	0.81	1.07																												
10	**C11BMR** *B. rajah*	1.54	1.56	1.47	1.01	1.17	1.73	0.81	1.07	0.00																											
11	**C7BMR** *B. rajah*	1.54	1.56	1.47	1.01	1.17	1.73	0.81	1.07	0.00	0.00																										
12	**C8BMR** *B. rajah*	1.54	1.56	1.47	1.01	1.17	1.73	0.81	1.07	0.00	0.00	0.00																									
13	**C1BMR** *B. rajah*	1.54	1.56	1.47	1.01	1.17	1.73	0.81	1.07	0.00	0.00	0.00	0.00																								
14	**C12BMR** *B. rajah*	1.54	1.56	1.47	1.01	1.17	1.73	0.81	1.07	0.00	0.00	0.00	0.00	0.00																							

Appendix N: Genetic distance (%) between *Begonia* species based on *rpl*32–*trn*L dataset (continued).

No.	Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
15	BGRE2 *B. reginula*	1.59	1.51	1.60	0.83	1.20	1.58	1.02	1.10	0.86	0.86	0.86	0.86	0.86	0.86																					
16	BGRE3 *B. reginula*	1.59	1.51	1.60	0.83	1.20	1.58	1.02	1.10	0.86	0.86	0.86	0.86	0.86	0.86	0.00																				
17	BGRE1 *B. reginula*	1.59	1.51	1.60	0.83	1.20	1.58	1.02	1.10	0.86	0.86	0.86	0.86	0.86	0.86	0.00	0.00																			
18	BGYN1 *B. yenyeniae*	1.68	1.51	1.52	1.21	1.39	1.76	1.02	1.29	0.77	0.77	0.77	0.77	0.77	0.77	0.60	0.60	0.60																		
19	BGCK1 *B.* sp. 'Chiku 4/5'	1.48	1.67	1.73	0.84	1.40	1.66	0.64	0.91	0.92	0.92	0.92	0.92	0.92	0.92	1.11	1.11	1.11	1.03	0.00																
20	BGFK1 *B.* sp. 'Pisces Lim'	1.21	1.41	0.98	0.51	0.70	1.29	0.50	0.73	0.92	0.92	0.92	0.92	0.92	0.92	0.75	0.75	0.75	0.92	0.45																
21	25YC_BFOX10 *B. foxworthyi* 'Green Hairy'	1.38	1.57	1.12	0.74	0.91	1.55	0.55	1.00	0.84	0.84	0.84	0.84	0.84	0.84	1.03	1.03	1.03	0.94	0.59	0.47															

Appendix N: Genetic distance (%) between *Begonia* species based on *rpl*32–*trn*L dataset (continued).

No.	Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
22	25YC_BFOX9 *B. foxworthyi* 'Green Hairy'	1.29	1.67	1.22	0.84	0.82	1.65	0.65	1.09	0.93	0.93	0.93	0.93	0.93	0.93	1.12	1.12	1.12	1.03	0.67	0.56	0.08														
23	BGFR1 *B. foxworthyi* 'Red Hairy'	1.10	1.48	1.32	0.56	0.73	1.27	0.37	0.81	0.84	0.84	0.84	0.84	0.84	0.84	1.03	1.03	1.03	0.95	0.42	0.37	0.34	0.42													
24	BGFC1 *B. foxworthyi* 'Chiku 2'	1.10	1.48	1.32	0.56	0.73	1.27	0.37	0.81	0.84	0.84	0.84	0.84	0.84	0.84	1.03	1.03	1.03	0.95	0.42	0.37	0.34	0.42													
25	KR186728 *B. foxworthyi*	1.84	2.04	1.59	1.21	1.36	1.83	1.07	1.62	1.26	1.26	1.26	1.26	1.26	1.26	1.46	1.46	1.46	1.37	0.84	0.90	0.92	1.01	0.76	0.76											
26	BGNR1 *B. nurii* 'red'	1.39	1.58	1.32	0.84	0.92	1.38	0.74	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.03	1.03	1.03	0.94	0.59	0.55	0.67	0.59	0.50	0.50	0.93										
27	KR186727 *B. forbesii*	1.50	0.93	1.51	1.23	1.10	1.76	0.74	1.29	1.45	1.45	1.45	1.45	1.45	1.45	1.57	1.57	1.57	1.48	1.19	1.12	1.19	1.28	1.02	1.02	1.54	1.28									
28	KR186775 *B. sublobata*	1.11	1.20	1.02	0.75	0.72	1.28	0.88	1.00	1.01	1.01	1.01	1.01	1.01	1.01	1.02	1.02	1.02	1.03	0.84	0.63	0.76	0.84	0.59	0.59	1.18	0.76	0.94								
29	BGNM1 *B. nurii*	1.38	1.48	1.40	0.75	0.91	1.65	0.55	0.72	1.02	1.02	1.02	1.02	1.02	1.02	1.21	1.21	1.21	1.12	0.76	0.65	0.68	0.76	0.51	0.51	1.10	0.85	1.02	0.76							

Appendix N: Genetic distance (%) between *Begonia* species based on *rpl*32–*trn*L dataset (continued).

No.	Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
30	BGNG1 *B. nurii* 'Green PLKN Etnobotani'	1.47	1.75	1.48	0.83	1.00	1.56	0.82	0.00	1.09	1.09	1.09	1.09	1.09	1.09	1.11	1.11	1.11	1.20	0.84	0.67	0.93	1.01	0.76	0.76	1.18	0.93	1.36	1.01	0.67						
31	BGCN1 *B.* sp. 'Creeping Nurii'	1.46	1.56	1.30	0.83	0.99	1.55	0.64	0.72	0.92	0.92	0.92	0.92	0.92	0.92	1.12	1.12	1.12	1.23	0.67	0.56	0.77	0.84	0.59	0.59	1.01	0.76	1.19	0.85	0.59	0.67					
32	KR186738 *B. ignorata*	1.20	1.58	1.12	0.74	0.72	1.47	0.55	0.82	0.84	0.84	0.84	0.84	0.84	0.84	0.03	0.03	0.03	0.94	0.59	0.45	0.68	0.59	0.51	0.51	0.93	0.51	1.11	0.76	0.51	0.76	0.59				
33	KR186786 *B. tigrina*	1.37	1.65	1.20	0.55	0.90	1.46	0.72	0.81	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	0.75	0.55	0.67	0.76	0.67	0.67	0.99	0.84	1.28	0.92	0.84	0.75	0.75	0.67			
34	25YC_BSP_KT1 *B.* sp. 'Kelantan-Terengganu'	1.43	1.36	1.36	0.95	1.20	1.58	0.84	1.01	0.61	0.61	0.61	0.61	0.61	0.61	0.43	0.43	0.43	0.35	0.87	0.75	0.78	0.87	0.79	0.79	1.02	0.79	1.15	0.79	0.96	1.04	0.87	0.78	0.95		
35	25YC_BYAP1 *B. yappii*	1.29	1.19	1.09	0.83	1.13	1.44	0.72	1.02	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.29	0.96	0.85	0.87	0.99	0.87	0.87	1.36	0.87	1.08	0.78	0.78	0.97	1.07	0.97	0.77	0.96	0.28
36	25YC_BFOX6 *B. "foxworthyi"*	1.33	1.33	1.10	0.72	0.92	1.05	0.47	0.58	0.87	0.87	0.87	0.87	0.87	0.87	1.11	1.11	1.11	0.99	0.64	0.64	0.77	0.76	0.55	0.55	0.58	0.59	0.76	0.66	0.54	0.43	0.55	0.65	0.43	0.64	0.78	0.89

APPENDIX O

Appendix O: Genetic distance (%) between *Begonia* species based on chloroplast dataset.

No.	Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	**25YC_ BFOX9 *B. foxworthyi* 'Red Hairy'**																		
2	**25YC_ BFOX10 *B. foxworthyi* 'Green Hairy'**	0.13																	
3	**BGRE1 *B. reginula***	1.23	1.23																
4	**BGRE2 *B. reginula***	1.20	1.20	0.03															
5	**BGYN1 *B. yenyeniae***	1.33	1.26	0.52	0.49														
6	**C8BMR *B. rajah***	1.22	1.14	0.78	0.75	0.85													
7	**C9BMR *B. rajah***	1.18	1.11	0.81	0.78	0.88	0.03												
8	**C12BMR *B. rajah***	1.22	1.15	0.78	0.75	0.85	0.00	0.03											
9	**BGFC1 *B. foxworthyi* 'Chiku 2'**	0.84	0.84	1.04	1.01	1.23	1.09	1.12	1.09										

Appendix O: Genetic distance (%) between *Begonia* species based on chloroplast dataset (continued).

No.	Sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
10	**BGFR1 B. foxworthyi 'Red Hairy'**	0.84	0.84	1.04	1.01	1.23	1.09	1.12	1.09	0.00									
11	**BGFK1 B. sp. 'Pisces Lim'**	0.87	0.83	0.90	0.84	1.13	1.02	1.06	1.03	0.72	0.72								
12	**BGNG1 *B. nurii* 'Green PLKN Etnobotani'**	1.04	1.03	0.97	0.94	1.22	1.09	1.12	1.09	0.79	0.79	0.68							
13	**BGNM1 *B. nurii***	0.94	0.87	1.07	1.04	1.14	1.00	1.03	1.00	0.76	0.76	0.75	0.63						
14	**BGCN1 *B.* sp. 'Creeping Nurii'**	0.87	0.80	1.04	1.01	1.10	0.96	0.93	0.96	0.75	0.75	0.68	0.72	0.63					
15	**BGNR1 *B. nurii* 'red'**	0.75	0.74	0.81	0.78	0.87	0.86	0.89	0.86	0.66	0.66	0.55	0.69	0.66	0.60				
16	**BGCK1 *B.* sp. 'Chiku 4/5'**	0.91	0.90	1.20	1.17	1.30	1.18	1.15	1.19	0.84	0.84	0.87	0.94	0.94	0.77	0.75			
17	**25YC_BSP_KT1 *B.* sp. 'Kelantan-Terengganu'**	1.15	1.08	0.44	0.41	0.44	0.67	0.64	0.67	1.11	1.11	1.01	1.14	1.01	0.91	0.78	1.10		
18	**25YC_ BFOX6 *B. "foxworthyi"***	0.95	0.91	1.16	1.13	1.20	0.97	0.93	0.97	0.85	0.85	0.81	0.77	0.67	0.60	0.67	0.81	0.95	
19	**25YC_BYAP1 *B. yappii***	1.22	1.14	0.52	0.48	0.52	0.76	0.72	0.76	1.24	1.24	1.13	1.20	1.07	0.99	0.85	1.22	0.31	1.09

Universiti Tunku Abdul Rahman			
Form Title : Supervisor's Comments on Originality Report Generated by Turnitin for Submission of Final Year Project Report (for Undergraduate Programmes)			
Form Number: FM-IAD-005	Rev No.: 0	Effective Date: 01/10/2013	Page No.: 1of 1

FACULTY OF SCIENCE

Full Name(s) of Candidate(s)	YONG CAYLEE
ID Number(s)	23ADB00540
Programme / Course	BACHELOR OF SCIENCE (HONS) AGRICULTURAL SCIENCE (AG)
Title of Final Year Project	MARKER ASSESSMENT, MOLECULAR CHARACTERISATION AND PHYLOGENY OF PENINSULAR MALAYSIAN *Begonia* spp. UNDER THE SECTION *Jackia*

Similarity	**Supervisor's Comments (Compulsory if parameters of originality exceeds the limits approved by UTAR)**
Overall similarity index: 10 % **Similarity by source** Internet Sources: 5 % Publications: 4 % Student Papers: 7 %	Values within accepted range
Number of individual sources listed of more than 3% similarity: 0	N/A
Parameters of originality required and limits approved by UTAR are as follows: **(i) Overall similarity index is 20% and below, and** **(ii) Matching of individual sources listed must be less than 3% each, and** **(iii) Matching texts in continuous block must not exceed 8 words** ***Note: Parameters (i) – (ii) shall exclude quotes, bibliography and text matches which are less than 8 words.***	

Note Supervisor/Candidate(s) is/are required to provide softcopy of full set of the originality report to Faculty/Institute

Based on the above results, I hereby declare that I am satisfied with the originality of the Final Year Project Report submitted by my student(s) as named above.

Signature of Supervisor
Name: Dr. Tan Ji
Date: 8/10/2025

Signature of Co-Supervisor
Name: _____
Date: _____

10% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Match Groups

85 Not Cited or Quoted 5%
Matches with neither in-text citation nor quotation marks

68 Missing Quotations 4%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

5% Internet sources
4% Publications
7% Submitted works (Student Papers)

Match Groups

- **85** Not Cited or Quoted 5%
  Matches with neither in-text citation nor quotation marks
- **68** Missing Quotations 4%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

5% Internet sources
4% Publications
7% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Student papers	
Taylor's Education Group on 2020-10-14		<1%
2	Internet	
writtle.repository.guildhe.ac.uk		<1%
3	Internet	
theses.gla.ac.uk		<1%
4	Student papers	
Taylor's Education Group on 2021-06-24		<1%
5	Student papers	
Taylor's Education Group on 2017-07-25		<1%
6	Student papers	
Taylor's Education Group on 2019-06-17		<1%
7	Internet	
academic.oup.com		<1%
8	Internet	
eprints.utar.edu.my		<1%
9	Internet	
www.researchgate.net		<1%
10	Student papers	
Royal Botanic Garden Edinburgh on 2023-08-18		<1%