Exploration of SNP Markers in Shade Adaptation Genes and their Protein-Level Mutation Effects in Indonesian *Begonia*

Rafaelangelo Yudhistira Dharmawangsa
202108510014

Thesis

Department of Biotechnology
Faculty of Bioscience, Technology, and Innovation
Atma Jaya Catholic University of Indonesia
Jakarta
2025

Exploration of SNP Markers in Shade Adaptation Genes and their Protein-Level Mutation Effects in Indonesian *Begonia*

Rafaelangelo Yudhistira Dharmawangsa
202108510014

Thesis
As partial fulfillment for
The degree of Bachelor of Biotechnology
Faculty of Bioscience, Technology, and Innovation,
Atma Jaya Catholic University of Indonesia

Department of Biotechnology
Faculty of Bioscience, Technology, and Innovation
Atma Jaya Catholic University of Indonesia
Jakarta
2025

INTELLECTUAL PROPERTY STATEMENT FORM

I,

Name : Rafaelangelo Yudhistira Dharmawangsa
NIM : 202108510014

Certify that my thesis is my own original work and no portion of my thesis has been copyrighted previously unless properly referenced.

If there is a breach of items above, I will take full responsibility to Atma Jaya Catholic University of Indonesia for any legal action that might be caused

Jakarta, July 21st, 2025

Rafaelangelo Yudhistira Dharmawangsa

ABSTRACT

RAFAELANGELO YUDHISTIRA DHARMAWANGSA. Exploration of SNP Markers in Shade Adaptation Genes and their Protein-Level Mutation Effects in Indonesian *Begonia*. Under direction of LISTYA UTAMI KARMAWAN and SRI WAHYUNI. This study aims to identify the most informative SNPs to distinguish semi-shade and deep shade *Begonia*, using chloroplast genes that experienced positive selection in shade environment (*rpoC1, rpoB, psbE, psbK, petA, rpl2, rpl22, rps12*). These genes were parsed from 30 Asian *Begonia* chloroplast genome sequence and aligned using MUSCLE. The resulting sequences were analyzed to detect positive selection differentiating the semi-shade adapted from the deep shade adapted *Begonia.* Additionally, mutation impact analysis was performed on the obtained variants. Subsequently, leaf samples from seven Indonesian *Begonia* were collected for genomic DNA extraction. The aligned in silico samples were then used as a template for degenerate primer building. The target genes (*rpoC1*, *psbE*, and *petA*) were then amplified and sequenced using Sanger method. The sequences were then used for another positive selection analysis. Based on our findings, in *rpoC1,* SNP 671 A and 1679 A contributes to the semi-shade adaptation, while SNP 671 C and 1679 G contributes to deep shade adaptation *Begonia. In silico* analysis suggests that the H560R mutation stabilizes the local protein structure and may reduce solvent accessibility at site 560. Although direct evidence of improved transcriptional efficiency is lacking, this structural change may help maintain RNA polymerase function under energy-limited conditions common in deep-shade environments.

Keywords: *Begonia,* mutation, shade adaptation, single nucleotide polymorphism

ABSTRAK

RAFAELANGELO YUDHISTIRA DHARMAWANGSA. Eksplorasi Marka SNP pada Gen Tahan Naungan dan Efek Mutasi Protein di *Begonia* Indonesia. Pembimbing LISTYA UTAMI KARMAWAN dan SRI WAHYUNI. Penelitian ini ditujukan untuk mengidentifikasi marka SNPs yang dapat digunakan untuk membedakan *Begonia* naungan parsial dan naungan penuh menggunakan gen seleksi positif naungan (*rpoC1, rpoB, psbE, psbK, petA, rpl2, rpl22, rps12*). Gen seleksi positif diperoleh dari 30 sekuens genom kloroplas *Begonia* Asia dan dilakukan *alignment* dengan MUSCLE. Sekuens yang diperoleh digunakan untuk mendeteksi seleksi positif yang membedakan kedua sifat tanaman naungan parsial dan naungan penuh. Selain itu, sekuens tersebut juga digunakan untuk mengetahui dampak mutasi pada tingkat protein. Selanjutnya, sampel daun dari tujuh Begonia Indonesia dikumpulkan untuk isolasi DNA genom. Kemudian, sekuens yang sebelumnya telah dilakukan *alignment* digunakan sebagai *template* penyusunan primer dari gen *rpoC1*, *psbE*, and *petA*. Gen tersebut kemudian diamplifikasi dan dilakukan sekuensing menggunakan metode Sanger. Sekuens yang diperoleh akan digunakan kembali untuk analisis seleksi positif. Berdasarkan hasil dari analisis tersebut, pada gen *rpoC1*, marka SNP 671 A dan SNP 1679 A memiliki keterkaitan dengan sifat naungan parsial, sedangkan SNP 671 C dan SNP 1679 G memiliki keterkaitan dengan sifat naungan penuh. Hasil analisis *in silico* menunjukan bahwa mutasi H560R menyebabkan protein mengalami peningkatan stabilitas secara lokal, dan memiliki penurunan aksesibilitas terhadap pelarut pada titik 560. Meskipun bukti langsung mengenai peningkatan efisiensi transkripsi masih belum ada, perubahan struktur ini kemungkinan membantu mempertahankan fungsi RNA polimerase dalam kondisi terbatas energi yang umum terjadi di lingkungan dengan naungan yang sangat teduh.

Kata Kunci: adaptasi naungan, *Begonia,* mutasi, *single nucleotide polymorphism*

APPROVAL FORM

We hereby certify that:

Name : Rafaelangelo Yudhistira Dharmawangsa
NIM : 202108510014
Title : Exploration of SNP Markers in Shade Adaptation Genes and their Protein-Level Mutation Effects in Indonesian *Begonia*
Date of exam : July 21st, 2025

has passed the thesis exam and confirmed that this thesis had been thoroughly examined, improved, and approved by the advisor.

Approved by

Dr. Listya Utami Karmawan
Advisor

Sri Wahyuni, S.Si., M.Si.
Co-advisor

Acknowledge by

Dr. Listya Utami Karmawan
Head of Biotechnology Study Program

PREFACE

The author would like to express their heartfelt gratitude to the Almighty God as his divine blessings and guidance helped the completion of the thesis titled "Single Nucleotide Polymorphism (SNP) Exploration in Shade Adaptation Genes at Indonesian Begonia Hybrids." This research was performed to fulfill the requirements to acquire the bachelor's degree in biotechnology from the Department of Biotechnology, School of Bioscience, Technology, and Innovation and conducted at the Plant Cell and Tissue Culture Laboratory, spanning from February 2024 to June 2025.

In the preparation of this thesis, the author received valuable assistance from various individuals. Firstly, the author would like to extend their gratitude to their advisors, Dr. Listya Utami Karmawan and Sri Wahyuni, S.P., M.Si. for the opportunity, and all the support and guidance throughout this study. The author would like to express their gratitude to the laboratory technicians of Atma Jaya Catholic University of Indonesia, as well as the people in Bogor Botanical Gardens, National Research and Innovation Agency who aided assistance and support. The author also would like to express their gratitude to the lecturers at Atma Jaya that guided the author throughout this research, namely Dr. Adi Yulandi, and Jimmy Suryadi, Ph.D. for the support throughout the *in silico* process and the molecular techniques that aided this research. Special thanks are also extended to the parents and family for their unwavering support and prayers. Lastly, the author would also like to give thanks to friends and fellow colleagues researchers, both in the Plant Cell and Tissue Culture Laboratory (Nathanael Nanda Santosa, Steven, Hovan Gahara, Valdo Setiawan, Ariya Leesidro), and DNA technology Laboratory (Agnes Asyanty, Karina Christianti, Purwaningtyas, Govinda Adicitta), as well as friends from UKM FODIM (Christian Rainaldo Rahardja, Ivana Dharmaputri, Azzahra Amanda, Yosef Satrio Wicaksono, Felix Cahyo Nugroho, Nicolaus Gilang Yogabastono), MICROBES (Stella, Felicia Wihardjo, Caroline Vania Putrisantosa) and Dorm-mates (Chintya Titania, Gracia Amanda Sam Budisetyono, James Daniel, Steven Mavish).

Despite the author's utmost effort, this thesis acknowledges its imperfections. Nevertheless, the author hopes that this work can contribute to, and aid related to this topic. The author wishes best regards to those who will continue this research topic.

Sincerely,

Rafaelangelo Yudhistira Dharmawangsa

BIOGRAPHY

The author was born in Jakarta, 8th Agustus 2003, as a first-born child from dr. Regina Maria Irene Bhekti Putrantiwi and Antonius Herlambang Muliawan, S. E. The author graduated from St. Fransiscus Asisi Senior Highschool Jakarta in 2021 and continue his education at Atma Jaya Catholic University of Indonesia, Faculty of Bioscience, Technology, and Innovation until 2025. As a college student, the author served as teaching assistant for Biostatistics for Class of 2022 and 2023.

The author also took part in many organizations, namely MICROBES as a Staff of Graphic Design (2021-2023), and Journalist and Editor (2024). Besides MICROBES, he also joined Unit Kegiatan Mahasiswa Forum Diskusi Ilmiah Mahasiswa (UKM FODIM), as an active member of SIPOSTARU XXX (2021-2022), Coordinator of Content Creator (2022-2023), and Head of External Relations (2023-2024). Throughout his organization experienced, he contributed at many events as a committee, namely Chairman of Pelatihan Copywriting (PCW) UKM FODIM 2024, Committee Head for Layout Gaudeamus UKM FODIM 2024, Chaiman of Pelatihan Jurnalistik (Penjur) 2024, Supervisor and Reviewer of Pelatihan Karya Tulis 2024: ”Unleash Your Breakthrough with Program Kreativitas Mahasiswa (PKM)”, Liaison Officer for ”UAJ Goes To PKP2 and PIMNAS-37” with the Bureau of Career and Student Affairs AJCUI, and Supervisor of Atma Jaya Science Competition (AJSC).

He achieved the title of ONMIPA Finalist in 2023 at Biology and received a Gold Medalist at Kompetisi Sains Nasional (KSN) 2024 at Biology. He also has experience as a Product Specialist Intern at PT. Diastika Biotekindo and was assigned to quality control of lab instruments as well as PCR kit. He contributed to many voluntary projects, namely RAPD Fingerprinting of Rambutan Parakan Tangerang, and became the Head of Reviewer for Proposal PKM 2025 in affiliation with the Bureau of Career and Student Affairs AJCUI.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

INTRODUCTION

Begonia is one of the most diverse genera of Angiosperms plants, consisting of approximately 70 sections and around 2.000 species (Ningrum *et al.* 2020). Their variety of shapes and colors of leaves makes *Begonia* is considered as an ornamental plant with high economic value with their highly diverse leaf morphology. Indonesia is one of the countries with the largest number of endemic *Begonia* species, with 63 species recorded from Sumatra, 15 species from Java, 8 species from Kalimantan, 44 species from Sulawesi, and 77 species from Papua (Siregar 2017). Hence, hybridization efforts continue as plant breeders aim to develop cultivars with improved traits, one of which is light and shade tolerance.

Begonia typically thrives in fully shaded or semi-shaded environments but is highly susceptible to sunburn when exposed to full sun, although these traits vary among species. This is likely due to their habitats, which include dense-canopy rainforests, cave understories, and mountainous regions at elevations of up to 2,400 meters above sea level. Therefore, plant breeders consider the semi-shade cultivars to be of great value for landscaping, while the deep shade to be of great value for house plants with lower sunlight requirement. However, demands on *Begonia* continue to rise with emerging markets seeking versatile, shade-tolerant ornamentals adaptable to various microclimates. Developing cultivars with these traits can be challenging due to the limited information on selecting the suitable parental plant for hybridization. Therefore, molecular breeding markers should be utilized to identify the best candidate for a parental plant to further increase the probability of the desired trait cultivar.

Previous studies shown that *Begonia semperflorens* (which is a semi-shaded adapted *Begonia*) experienced stress with strong bright environment, with indication such as high level production of Reactive Oxygen Species (ROS), increasing number of anthocyanin, and reduction number of leaves with their chloroplast content (Zhang *et al.* 2018; Zhao *et al.* 2024). Others studies shown that *Begonia* have higher photosynthetic gene regulation under shaded environment (Emelianova & Kidner 2022). Li *et al.* (2022) study discovered that there are whole genome duplication events at the ancestral phylogeny of *Begonia*, that caused photosynthetic gene in chloroplast DNA. Furthermore, another study also found eight genes that are positively-selected under shade adapted Asian *Begonia*, such as RNA polymerase gene (*rpoC1*, *rpoB*), ribosome subunit genes (*rps12*, *rpl2*, and *rpl22*), and photosystem complex II (*psbE*, *psbK*, and *petA*) (Xiong *et al.* 2023). Those studies also showed multiple codon sites under positive selection, which had the potential for identifying shade-adaptative traits in *Begonia*.

Molecular breeding markers such as SNPs have the potential to be used to differentiate *Begonia* that are adapted in semi-open habitats to those which are adapted to deep full shade. In chloroplast DNA, these single base pair variation

could introduce non-synonymous mutation and be used as informative markers due to their mutations could cause alterations in protein functions (Kim *et al.* 2022; Kumar *et al.* 2023). Additionally, chloroplast DNA have higher conservative regions which allows mutations in coding regions to be more easily detected and interpreted their functional adaptation (Kim *et al.* 2022; Kumar *et al.* 2023). Furthermore, this research aims to identify the most informative SNPs to distinguish shade-adapted *Begonia* from those that can tolerate direct light, using chloroplast genes that experienced positive selection ($\omega > 1$), causing non-synonymous mutation in semi-shade and deep shade environment adapting plant according to Xiong *et al.* (2023). Therefore, these mutations could be used as a breeding marker to differentiate semi-shade and deep-shade plants. Meanwhile, chloroplast genes are inherited from the maternal plant; consequently, potential breeding markers can be used to select for desired traits from the maternal line (Park *et al.* 2021). This research aims to identify the most informative breeding SNPs to distinguish semi-shade adapted *Begonia*, using positive selected genes in shade environment using *in silico* methods, which were subsequently verified through *silico* experiments. This research also aims to study the impact of the mutation to the adaptation of the two different group of *Begonia*. Therefore, this analysis will aid in explaining the evolutionary mechanisms underlying their adaptive divergence.

METHODOLOGY

This research was conducted at Plant Cell and Tissue Culture Atma Jaya Catholic University of Indonesia from February 2024 to June 2025, which employed both *in silico* and *in silico* methods (Figure 1). Eight positive selected genes that are related to shade adaptation (*rpoC1, rpoB, psbE, psbK, petA, rps12, rpl2, rpl22*) were parsed and aligned using MUSCLE (Xiong *et al.* 2023). Aligned sequences were then used to build a phylogenetic tree, which were used for positive selection analysis. These markers were subsequently used in protein structure analyses to compare the impact of mutations between the two ecotypes Furthermore, degenerate primers for *rpoC1, psbE*, and *petA* designed from the alignment of 30 *Begonia* species, were tested via PCR on genomic DNA from seven Begonia species. The resulting PCR amplicons were sequenced and used to verify SNPs that differentiate between two ecotypes: semi-shade-adapted and deep-shade-adapted *Begonia*. In this study, the term semi-shade refers to *Begonia* species adapted to semi-open habitats or environments with bright and indirect sunlight. Conversely, deep shade describes *Begonia*, a species that thrives under low light growing conditions.

Parsing of shade adapted genes from *Begonia* Chloroplast DNA

Complete Chloroplast DNA sequences from 30 Begonia were collected from GenBank NCBI database (https://www.ncbi.nlm.nih.gov/genbank/) (Appendix 1, Appendix 6). Sequence of shade adapted genes alongside their corresponding intergenic spacer were parsed using BioPython ver. 1.8.4 in Google Collab (Appendix 2) (Cock *et al.* 2009). The intergenic regions that were parsed for this research were based on the adjacent genes located based on the chloroplast map (Figure 2). The sequences were then aligned with MUSCLE using MEGA11 (Tamura *et al.* 2021; Xiong *et al.* 2023).

Alignment and Phylogenetic Tree Building

The sequences were then aligned using MUSCLE, and maximum likelihood phylogenetic tree was constructed in MEGA11 (Tamura *et al.* 2021). The alignment files were then used to construct a Maximum Likelihood Phylogeny tree in MEGA11 using Tamura-Nei substitution model with no bootstrapping, as bootstrapping was not performed due to the low divergence among sequences. (Tamura *et al.* 2021).

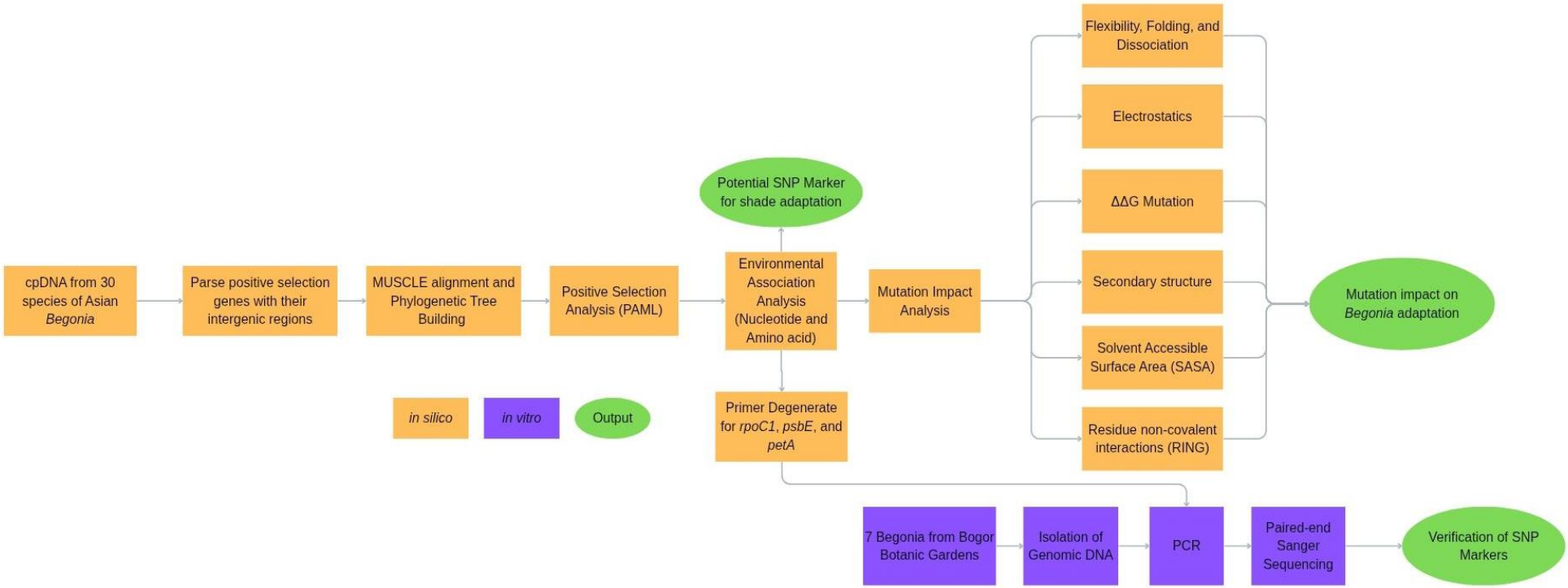

Figure 1 Diagram of Research Framework

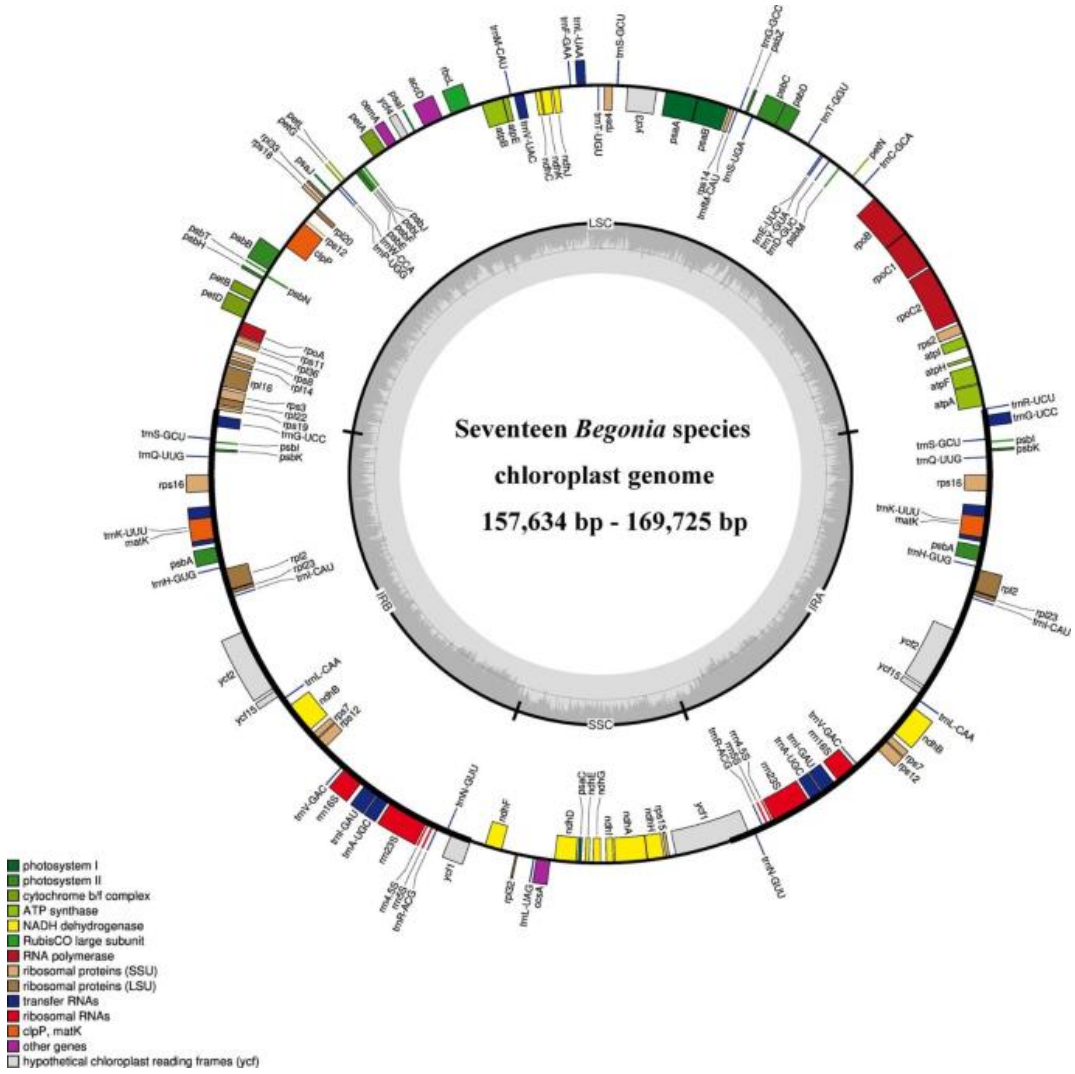

Figure 2 *Begonia* chloroplast genome map (Xiong *et al.* 2023).

Positive Selection Analysis

The Newick file and the alignment file were then used to count dN/dS ratio in Phylogenetic Analysis of Maximum Likelihood (PAML) ver. 4.9 (Yang *et al.* 2005; Yang 2007). The ω ratio were calculated using this formula:

$$\omega = \frac{dN}{dS}$$

Whereas: dN: Nonsynonymous substitution rate

dS: Synonymous substitution rate

$$dN = \frac{Number\ of\ non-synonymous\ mutation}{Number\ of\ non-synonymous\ site}$$

$$dS = \frac{Number\ of\ synonymous\ mutation}{Number\ of\ synonymous\ site}$$

The ω is interpreted as below:
ω > 1: Gene is on Positive/Rapid Evolution
ω = 1: Gene is on Neutral Evolution
ω < 1: Gene is on Negative/Constraint Evolution

To detect which site is positively selected, model = 0 (single dN/dS ratio (ω) is assumed for all branches), and Site models = M7 (beta) and M8 (beta & ω > 1) were used as a model in codeml (Appendix 5). M7 is used to constraint 0 < ω < 1, and M8 allowed M7 to detect ω > 1 (Yang 2007). The models were then compared using a likelihood ratio test (LRT) to detect evidence of diversifying selection (Yang 2007; Álvarez-Carretero *et al.* 2023). The Bayes Empirical Bayes (BEB) approach in model 8 was used to identify specific codon sites under positive selection. Posterior probabilities that are significant (p > 0.05) was selected as a candidate for detecting SNP marker (Xiong *et al.* 2023).

$$LR = 2 \times (\ln L_1 - \ln L_0)$$

Whereas: L_1 = Log-likelihood of the alternative (complex) model
L_0 = Log-likelihood of the null (simpler) model

Environmental Association Analysis for SNP Markers

Begonia samples growth habit, habitat, and shade adaptation data were collected mainly from Plants of The World Online (https://powo.science.kew.org/), and Flora of China (http://www.efloras.org/) database (Ding *et al.* 2014; Hoang & Lin 2023; Hughes & Miller 2002; Hughes & Girmansyah 2011; Indrakumar *et al.* 2013; Xiong *et al.* 2023). Any shade adaptation that is not recorded by both databases and other journal publications were inferenced ecologically based on their habitat (Appendix 6). Analysis of potential SNP marker was performed using SPSS ver. 25. To determine which site fits the appropriate model, the nucleotide data was used in Fisher-Exact test, and Adjusted Residual test (Huang *et al.* 2023; Purcell *et al.* 2007).

Protein Homologous Modelling and Mutation Impact Analysis

Begonia arachnoidea was chosen as a control for mutation impact analysis. *Begonia arachnoidea* is native to semi-shaded environments and represents the typical structural and functional profile under moderate light conditions (Peng *et al.* 2008). Therefore, in this Mutation Impact Analysis, semi-shade plant group were chosen as a Wild-Type, while deep shade plant group were chosen as a mutant.

Homologous Protein Modelling

The protein model was built using SWISS-MODEL (https://swissmodel.expasy.org/), and the positive sites were visualized as licorice sticks in PyMoL ver. 2.5.7 (Waterhouse *et al.* 2018). Models were chosen based on GMQE value (≥ 0.8), QMEANDisCo (≥ 0.7), sequence identity (≥ 90%) and the highest coverage. This parameter was chosen to select the best homologous protein model. GMQE and QMEANDisCo were used to evaluate the quality of the protein models, while sequence identity was used to show the similarity between the template and the protein sequence. Meanwhile, coverage showed the matching portion of the protein sequence in the template.

Secondary Structural, Network Protein Changes and Stability Changes

Secondary structures were verified under Ramachandran Plot using DynoPlot. The two different group of shades secondary structures were then tested statistically using Fisher Exact Test to determine whether the mutation significantly altered the protein's secondary structure.

Residue Interaction Network Generator (RING) (https://ring.biocomputingup.it/) was also performed in all the protein models to find any changes in amino acid network. Then, the connected amino acid residues were compared between the semi-shaded *Begonia* and the Deep shade *Begonia*. Meanwhile, stability calculation was performed under ΔΔG (delta-delta Gibbs) calculation with DUET (https://biosig.lab.uq.edu.au/duet/) (Pires *et al.* 2014). These tests will be used for detecting changes in stability of the related amino acids mutations. $\Delta\Delta G < 1$ will be considered as destabilizing mutations, while $\Delta\Delta G > 1$ will be considered as stabilizing mutations, and $\Delta\Delta G = 0$ will be considered as no change in stabilization. The values of ΔΔG will be tested in one sample T-test to check whether the mutation changes are significant enough to change the protein stability.

Solvent Accessible Surface Area (SASA) and Electrostatics Changes

SASA of protein models were calculated with GETAREA (https://curie.utmb.edu/area_man.html) (Fraczkiewicz & Braun 1998). SASA values of apolar value, backbone and sidechain were recorded. Electrostatics Changes were carried out by Adaptive Poisson-Boltzmann Solver (APBS)(https://server.poissonboltzmann.org/) (Jurrus *et al.* 2018). PDB2PQR were used to record pKa, desolvation regular, local reaction electrostatic field energy (Local RE), Columbic charge interactions and global net energy were recorded. These tests values were then statistically tested using two independent sample T-tests to determine whether the mutation caused a significant change to the mutant *Begonia* protein significantly.

Flexibility, Disorder, and Early Protein Folding Changes Analysis

Flexibility, disorder, and early protein folding regions was tested in Dynamine, Disomine, and EFoldMine from Bio2Bytes Tools (https://bio2byte.be/tool/) (Cilia *et al.* 2014; Orlando *et al.* 2022; Raimondi *et al.* 2017). Protein flexibility (backbone, sidechain, sheet, helix, and coils), protein folding, and disorder value were recorded between the two plant traits. Interpretation of Dynamine, EFoldMine, and Disomine were shown at Appendix 3. Afterwards, these values were statistically tested using Independent two sample T-test to determine whether the value of the mutant *Begonia* significantly changed.

Primer Design of *rpoC1*, *petA*, and *psbE*

Degenerate primers were then design according to the conserved region within the gene. Afterwards, the primer characteristics, namely melting temperature (Tm) and GC% were then analysed using NetPrimer (https://www.premierbiosoft.com/netprimer/) (Table 1; Figure 3). NetPrimer reaction conditions were optimized following these settings: Oligo concentration 10.000.000 pM, Monovalent Ion Concentration 50 mM, Free Mg2+ Ion Concentration 1.5 mM, Total Na+ Equivalent 204.92 mM, Temperature for Free Energy Calculation 25 °C. Consecutively, degeneracy (D) was calculated using the following equation:

$$D = \prod_{i=1}^{n} d_i$$

whereas d_i : number of nucleotides present at site i in the alignment (Appendix 4).

Table 1 Primers designed for amplification of *rpoC1, petA* and *psbE*

Amplicon	Primers	Length (bp)	Tm (°C)	GC %	D	Amplicon size (bp)
*rpoC1*_R1	F: GGAACTAMGGTCCTCGTTCAACSGG	25	68	52	4	927
	R: GGTCCCTACTTCACCTCGACGGTTATC	27	68	55	1	
*rpoC1*_R2	F: GCAGGAAGGAGTAACATGAAGCTCAG	26	63	50	1	877
	R: CGATCTTCCCATTCATTACCGGG	23	63	52	1	
*rpoC1*_R3	F: CGACAGGAGCGGGTGCTATCC	21	68	66	1	758
	R: CCCCATAACTTCCTGAAGTATTTCCSATAC	30	63	40	2	
*rpoC1*_R4	F: GGACGTTCTGTTATTGTCGTAGGTCC	26	65	50	1	960
	R: CGTTTCGACCCCCAACAAGCAT	22	65	54	1	
petA	F: CAGGGATAGGGAACTATACTAGCAACCTAC	30	63	46	1	1274
	R: GCGATACAAGTAATTCCCGAAATCC	25	60	44	1	
psbE	F: CAATACAGACGTGGTGACCAGTTGG	29	65	52	1	535
	R: GGATGAACTGCATTGCTGATATTGATC	27	60	41	1	

Abbreviation: R1, 2, 3, 4 = Read 1, 2, 3, 4

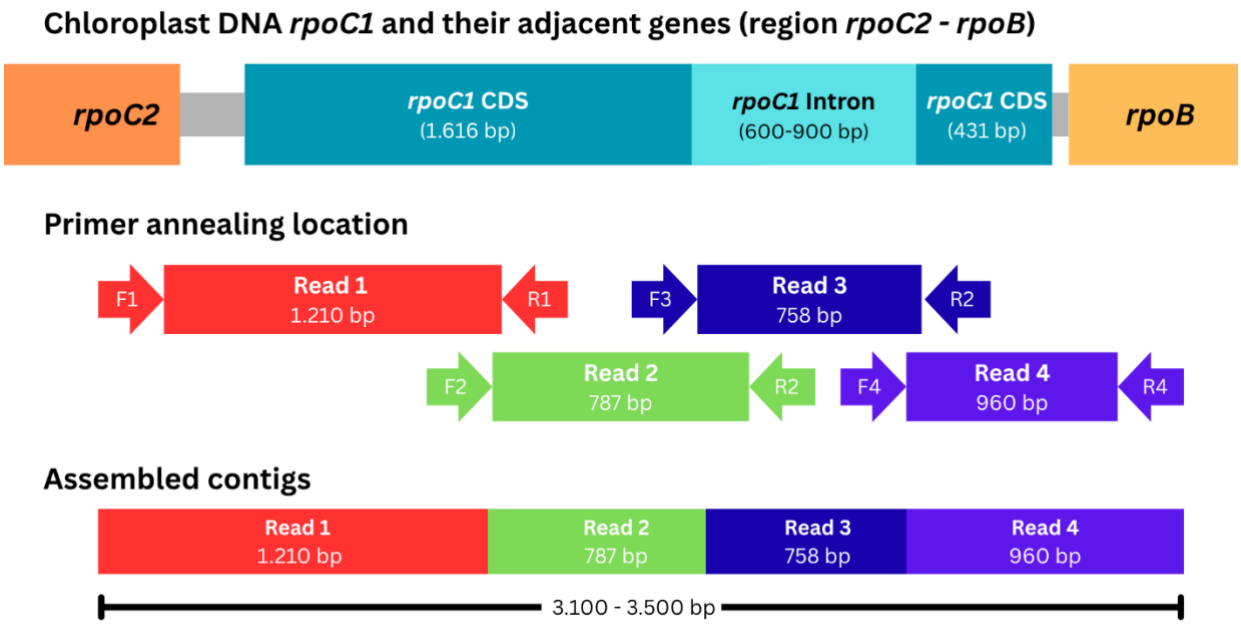

Figure 3 *rpoC1* gene reads amplification map

Sample Collection

Samples of *Begonia* leaves were collected from Kebun Raya Bogor (Appendix 8). *B. cucullata*, and *B. mexicana* are selected as the semi-shade Begonias, while *B. hijauvenia*, *B.* 'Ivory White' (*B. hijauvenia* ♀ × *B. kudoensis* ♂), *B.* 'Eternal Flames' (*B. hijauvenia* ♀ × *Begonia sp.* ♂), *B. isoptera*, *B.* 'Silver Pixie' (Gamma-irradiated *B. isoptera*) are selected as shade tolerant *Begonia* (Table 2).

Table 2 *Begonia* samples used and their respective natural habitat, growth habit, and shade adaptation

Begonia Samples	Growth Habit	Habitat	Shade adaptation
B. mexicana H.Karst. ex C.Chev.	Cane	NA	Semi-shade
B. cucullata	Cane	Grows in wet tropical forest borders or open areas (Handoyo Ardi & Thomas 2022; Moonlight & Fuentes 2022).	Semi-shade
B. 'Eternal Flame'	Rhizomatous	*B. hijauvenia* ♀ × *B. kudoensis* ♂	Deep shade
B. 'Ivory White'	Rhizomatous	*B. hijauvenia* ♀ × *Begonia sp.* ♂ (Davin 2023).	Deep shade
Begonia hijauvenia	Rhizomatous	Lowland forest, North Sumatera (Girmansyah *et al.* 2022).	Deep shade
Begonia isoptera Dryand. ex Sm.	Cane	Lowland forest, East and West Java (Ayu *et al.* 2019)	Deeep shade
Begonia Silver Pixie	Cane	Gamma-irradiated *B. isoptera* (Wahyuni *et al.* 2023)	Deep shade

Genomic DNA Extraction, Quantification and Quality Check

Genomic isolations were carried using Tiangen Plant Genomic DNA Kit protocol (Tiangen Beijing, Cat. no. 4992201) and Geneaid Plant Genomic DNA Kit protocol (Geneaid Taiwan, Cat. no. GP100). Quantification and Purity of the extracted DNA were checked under NanoDrop™ 2000, then verified using 1% agarose gel electrophoresis under 80V in 60 minutes in a 1× Tris-acetate-EDTA (TAE) buffer solution. To determine the correct size of amplicon, we used 1 kb DNA Ladder ExcelBand™ Smobio (Cat no. DM3100) for comparison. The gel was then stained using Ethidium Bromide (EtBr) and viewed under UV with Gel Documentation Model Infinity CX5WL (Vilber, France), and BioSoft application.

PCR Amplification and Visualization

PCR was performed using GoTaq® Green Mastermix (Promega), with a 40 µL volume reaction with components listed on Table 3. Five pairs of primers were used to amplify *rpoC1, psbE,* and *petA* chloroplast genes (Table 3). The PCR protocol was acquired with these steps accordingly (Table 4, Table 5). The annealing temperature (Ta) were calculated with NEB calculator (with primer concentration set to 5000 nM) and optimized to be 3–5 °C lower than the lowest. Meanwhile, the extension time were adjusted with the extension speed of Taq Polymerase used (~1 kb/minute) to amplify the intended size of the PCR product.

Table 3 PCR Component Reaction (40 µL) for Sanger Sequencing

Component	Concentration	Volume (µL)
2× GoTaq® Green Mastermix	1×	20
Forward Primer	0.5 ng/µL	1
Reverse Primer	0.5 ng/µL	1
Genomic DNA	(adjusted to 200 ng)	1-5
Nuclease Free Water (NFW)	-	(adjusted to 40 µL)

Table 4 PCR Protoccol of each pair of primer used

Reads	Primers	Tm (°C)	Product size (bp)	Ta (°C)	Extension time (s)
rpoC1 R1	F: GGAACTAMGGTCCTCGTTCAACSGG R: GGTCCCTACTTCACCTCGACGGTTATC	68 68	927	63	57
rpoC1 R2	F: GCAGGAAGGAGTAACATGAAGCTCAG R: CGATCTTCCCATTCATTACCGGG	63 63	877	61	48
rpoC1 R3	F: CGACAGGAGCGGGTGCTATCC R: CCCCATAACTTCCTGAAGTATTTCCSATAC	68 63	758	58	48
rpoC1 R4	F: GGACGTTCTGTTATTGTCGTAGGTCC R: CGTTTCGACCCCCAACAAGCAT	65 65	960	61	59
petA	F: CAGGGATAGGGAACTATACTAGCAACCTAC R: GCGATACAAGTAATTCCCGAAATCC	63 60	1274	58	78
psbE	F: CAATACAGACGTGGTGACCAGTTGG R: GGATGAACTGCATTGCTGATATTGATC	65 60	535	57	36

Abbreviation: R1, 2, 3, 4 = Read 1, 2, 3, 4

Table 5 PCR Protocol for Sanger Sequencing

Steps	Temperature (°C)	Time	Cycle
Pre-Denaturation	94	5 min	1
Denaturation	95	30 s	
Annealing	Adjusted	30 s	40
Extension	72	Adjusted	
Post-Extension	72	10 min	1
Hold	4	∞	1

Positive control primers, which are the universal primer ITS-Y4 (5'-CCC GCCTGACCTGGGGTCGC-3') and ITS-Y5 (5'-CCCGCCTGACCTGGGGTCGC-3'), were used to verify the PCR setup, and to ensure the DNA is amplified correctly with the PCR components remain functional. These controls are also essential for testing the annealing capability of the primers, as they were designed based on conserved regions across multiple species. Moreover, the 20 µL volume reaction followed these components (Table 6) and PCR protocol (Table 7Table 7) (Kita & Ito 2000).

Table 6 PCR Component Reaction (20 µL) for ITSY4-Y5 Amplification

Component	Concentration	Volume (µL)
2× GoTaq® Green Mastermix	1×	10
ITSY4 Primer	1 ng/µL	2
ITSY5 Primer	1 ng/µL	2
Genomic DNA	(adjusted to 100 ng)	1-5
Nuclease Free Water (NFW)	-	(adjusted to 20 µL)

Table 7 PCR Protocol for ITSY4-Y5 Amplification (Kita & Ito 2000).

Steps	Temperature (°C)	Time	Cycle
Pre-Denaturation	94	5 min	1
Denaturation	94	30 s	
Annealing	55	30 s	35
Extension	72	30 s	
Post-Extension	72	10 min	1
Hold	4	∞	1

Visualization of all the PCR products were done under 2% agarose gel electrophoresis under 80V in 90 minutes in a 1× Tris-acetate-EDTA (TAE) buffer solution. To determine the correct size of amplicon, we used 1 kb DNA Ladder ExcelBand™ Smobio (Cat no. DM3100) for comparison. the gel was then stained using Ethidium Bromide (EtBr) and viewed under UV with Gel Documentation Model Infinity CX5WL (Vilber, France), and BioSoft application.

Sanger Sequencing, Contigs Assembly and SNP Verification

PCR amplicons were then sent to be purified and paired-end Sanger sequenced at Macrogen, Inc., Singapore. All reads were assembled using SeqTrace ver. 0.9.0, and contigs were built based on a Q30 threshold, which put a limit for 1 in 1000 chance of a false base call (Stucky 2012). To verify the amplicon, the consensus sequences were checked in BLASTx. Scaffolds of *rpoC1* from the four contigs which were previously assembled were built using usegalaxy (https://usegalaxy.org.au/) merger tool.

RESULTS

Positive Selection Analysis

Based on LRT and BEB result, *rpoC1* (not included with the six *Begonia* samples isolated before), *rpoB, rpl22* and *rps12* showed significant evidence of positive selection. Based on the *in silico* BEB test result (Table 8), under the branch-site model, there are 3 significant positive sites for *rpoC1* (224, 560, and 566), 1 site for *rpoB* (594), 3 sites *rpl22* (37, 105, 129), and 3 sites for *rps12* (116, 117, 118). These sites are inferred to be evolving under adaptive pressure from the shade environment difference. Based on Appendix 6, most *Begonia* tend to have mutation E224T and H560R on *rpoC1*, and L37F on *rpl22*, while other mutations considered to be rare.

Table 8 BEB test results

Gene	Site (*)	p-value for $\omega > 1$	Post Mean $\omega \pm$ SE
rpoC1	224 E	0.968*	5.204 ± 1.702
	560 H	0.983*	5.267 ± 1.622
	566 K	0.994**	5.308 ± 1.558
rpoB	594 Y	0.951*	2.611 ± 0.705
	37 L	1.000**	7.618 ± 1.825
rpl22	105 E	0.954*	7.319 ± 2.284
	129 T	0.967*	7.384 ± 2.170
	116 Q	0.999**	9.165 ± 1.475
rps12	117 Y	0.985*	9.047 ± 1.784
	118 G	0.976*	8.976 ± 1.942

(*) Note: All amino acids position refers to *B. arachnoidea*

Environmental Association Analysis

The SNP 671 A/C and SNP 1679 A/G in *rpoC1* is the most significant associated with the shade adaptation on the Adjusted Residuals result (Figure 4). Allele A contributes to semi-shade adaptation, while allele C at site 671 or G at site 1679 contributes to deep shade adaptation ($p < 0.05$). Therefore, these two SNP markers in *rpoC1* may be a potential marker to differentiate between the shade adaptability of *Begonia*. Meanwhile, Glu 224 and His 560 are significantly linked to Semi-shade, while Arg 560 is significantly linked to deep shade (Figure 4). These differences might be biologically important as these mutations impact the proteins stability and functions.

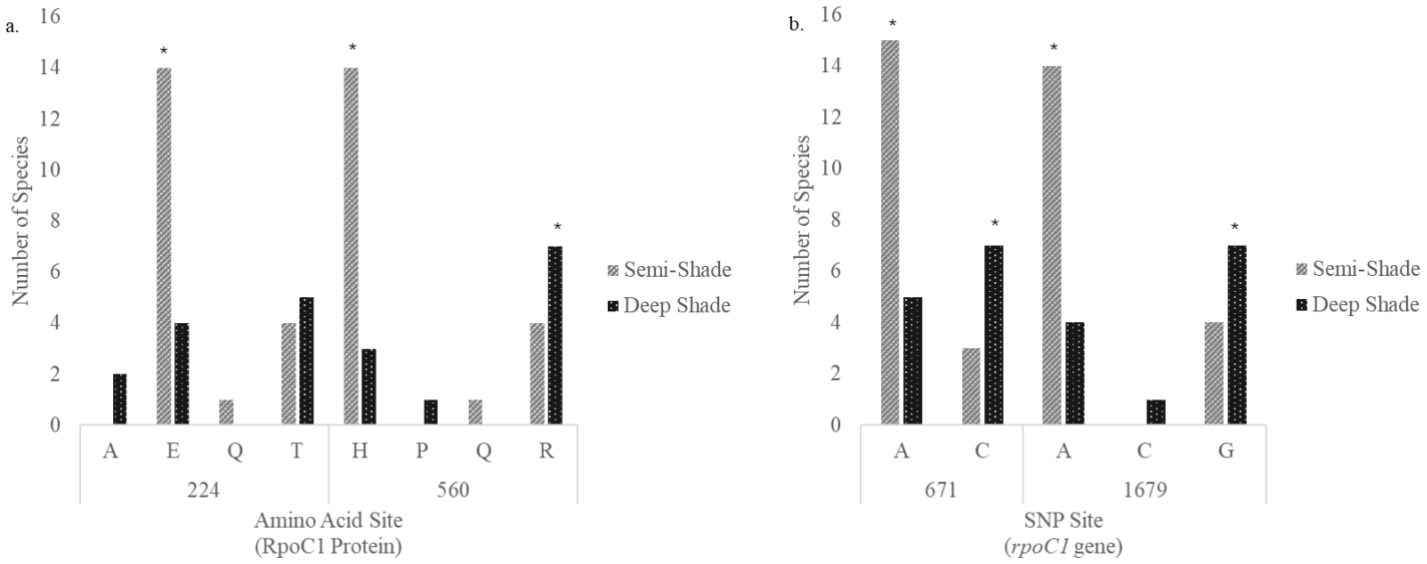

Figure 4 Adjusted residual test results of SNP sites in (a) RpoC1 protein and (b) *rpoC1* gene; Note: * significant contribution

Protein Modelling and Mutation Impact Analysis

Homologous Protein Modelling

RpoC1 proteins were modelled based on DNA-directed RNA polymerase subunit gamma from cryo-EM structure of the *Nicotiana tabacum* PEP-PAP-TEC2, with 100% sequence coverage while other the other parameter scores were shown on Appendix 7. This means that the models used have highly similar template with strong structural reliability and confidence in the predicted conformation.

Secondary Structural, Network Protein Changes and Stability Changes

The mutation H560R in *rpoC1* neither did change secondary structure nor the stability of the RpoC1 (RNA polymerase β' subunit) protein significantly (Table 9, Figure 5). The mutations also changed residue network interactions between the two different amino acids, notably with Pro 557, shifting from Van Der Waals interaction to Hydrogen Bond interaction.

Solvent Accessible Surface Area (SASA) and Electrostatics Changes

However, based on the SASA Result (Table 9), there was a significant difference between the two traits, specifically on the solvent accessible area in sidechain and polarity of the residue ($p < 0.01$). This was caused by the two different properties of amino acid, where Histidine is polar, partially charged residue, while Arginine is a polar, positively charged residue. Consequently, the electrostatics were also impacted with there being a significant change in the pKa, global net energy, and effects RE ($p < 0.01$). These alterations indicate a shift in the local electrostatic environment. Therefore, the substitution may influence not only structural stability but also the protein's local functional site.

Table 9 RpoC1 mutation H560R SASA, RING, and Electrostatics Result

Amino acid	ΔΔG (H560R)	Electrostatics	SASA	Interactions
H	-0.0024 ± 0.0486 kcal/mol ^{ns} (Slightly Destabilizing)	pKa: 6.1929 ± 0.0381** Global net energy: 6.12×10^5 kJ/mol** Local RE: 231.2353 ± 3.1384** Columbic Interaction: Mostly with Lys 577	Total: 121.7931 ± 8.4166 Å$^{2\ ns}$ Apolar: 91.6281 ± 6.6838 Å^{2} ** Backbone: 24.7394 ± 2.0513 Å^{2} ** Sidechain: 97.0231± 6.9020 Å$^{2\ ns}$	Mostly Van Der Waals Interaction with Pro 557**
R		pKa: 12.3336 ± 0.024** Global net energy: 6.148×10^5 kJ/mol** Local RE: 244.9091 ± 4.3452** Columbic Interaction: Mostly with Glu 580	Total: 124.0391 ± 11.3766 Å$^{2\ ns}$ Apolar: 51.6355 ± 6.6838 Å^{2} ** Backbone: 10.3355 ± 0.6667 Å^{2} ** Sidechain: 113.7036 ± 10.9453 Å$^{2\ ns}$	Mostly Hydrogen Bond Interaction with Pro 557**

Note: ^{ns} not significant; * significant differences ($p < 0.05$); ** very significant differences ($p < 0.01$)

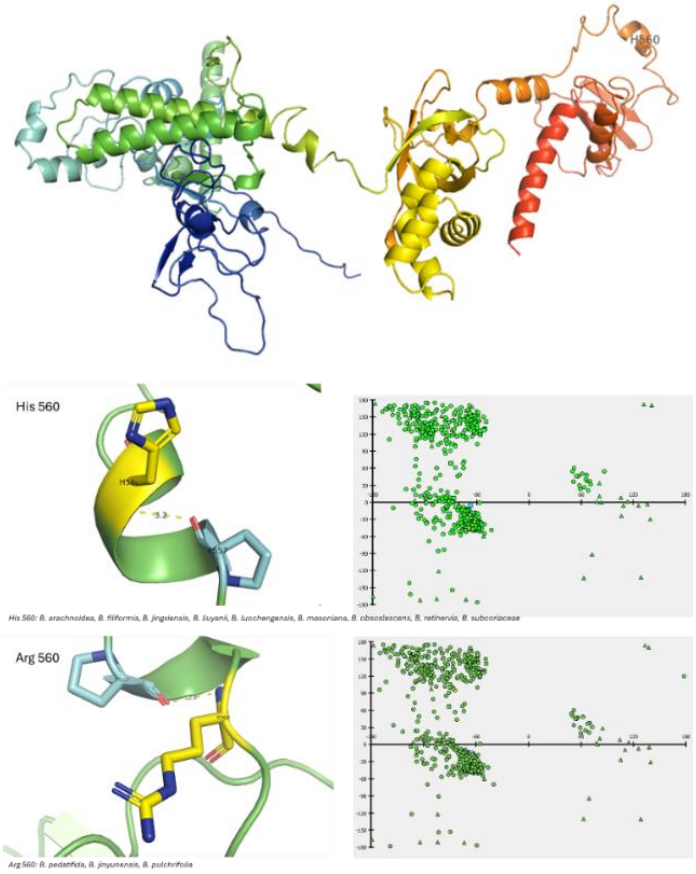

Figure 5 RpoC1 protein, His 560, Arg 560, and their respective Ramachandran Plot

Flexibility, Disorder, and Early Protein Folding Changes Analysis

The flexibility and folding of the protein were also impacted significantly, including the protein's backbone, sidechain, helix, coils flexibility (Table 10). These changes suggest that amino acid substitutions influence the dynamic behaviour of the protein, potentially altering its structural stability and functional performance. Furthermore, the H560R mutation might play a significant role in altering the transcription efficiency of the RNA polymerase.

Table 10 RpoC1 mutation flexibility, folding, and disorder changes

Amino acid	Flexibility	Propensities	Folding	Disorder
H	Backbone: 0.7284 ± 0.002** (Context Dependent) Sidechain: 0.5807 ± 0.001**	Sheet: 0.0845 ± 0.002^{ns} Helix: 0.2416 ± 0.005** Coil: 0.5771 ± 0.002**	0.0062 ± 0.000** (Low Folding)	0.1911 ± 0.001^{ns} (Low Disorder)
R	Backbone: 0.7357 ± 0.000** (Context Dependent) Sidechain: 0.3767 ± 0.000**	Sheet: 0.0869 ± 0.000^{ns} Helix: 0.2815 ± 0.001** Coil: 0.5535 ± 0.001**	0.0051 ± 0.000** (Low Folding)	0.1906 ± 0.003^{ns} (Low Disorder)

Note: ^{ns} not significant; * significant differences ($p < 0.05$); ** very significant differences ($p < 0.01$)

From the *in silico* test results, we chose *rpoC1* to verify further with *petA* and *psbE*. Both *petA* and *psbE* encodes protein related to light harvesting and photosystem electron transfer, which *psbE* encodes part of a core subunit protein to Photosystem II, and *petA* encodes cytochrome f in cytochrome b6f complex that play a role in membrane electron transport (Johnson & Pakrasi 2022; Lan *et al.* 2021).

Sanger Sequencing, Contigs Assembly, and Phylogenetic Tree Building

Based on the result of Sanger Sequencing of *petA* (Table 11, Figure 6, Appendix 14), the *petA* amplification, the result shown that the amplicon is Cytochrome f (PetA, Photosynthetic Electron Transport A), meaning that we've successfully amplified the *petA* gene. Meanwhile, The same clustering pattern result was also observed at *rpoC1* Read 2 and Read 3 (Table 12, Figure 7, Appendix 15). However, *B. mexicana* sequence is mostly like marking nut tree's (*Semecarpus reticulatus*) RNA polymerase β' subunit. This was because of Read 2 *rpoC1* is also sequenced with their intron, which the region may varied across species. Hence, any ambiguous base used in consensus sequences building caused the different result with the other species.

Table 11 BLASTx sequencing result of *petA* sequence

Species	Accession No.	Protein	Organism	Query Coverage (%)	E-value	Identity (%)
B. cucullata	YP_010629210.1	Cytochrome f	*Begonia heydei*	85	0.0	96.88
B. mexicana	YP_010629210.1		*Begonia heydei*	88	0.0	97.81
B. hijauvenia	XMH34472.1		*Begonia luochengensis*	91	0.0	97.81
B. 'Ivory White'	YP_009733650.1		*Begonia guangxiensis*	82	0.0	100.00
B. isoptera	YP_009733650.1		*Begonia guangxiensis*	85	1×10^{-154}	95.44
B. 'Silver Pixie'	YP_009775141.1		*Begonia versicolor*	89	0.0	99.69

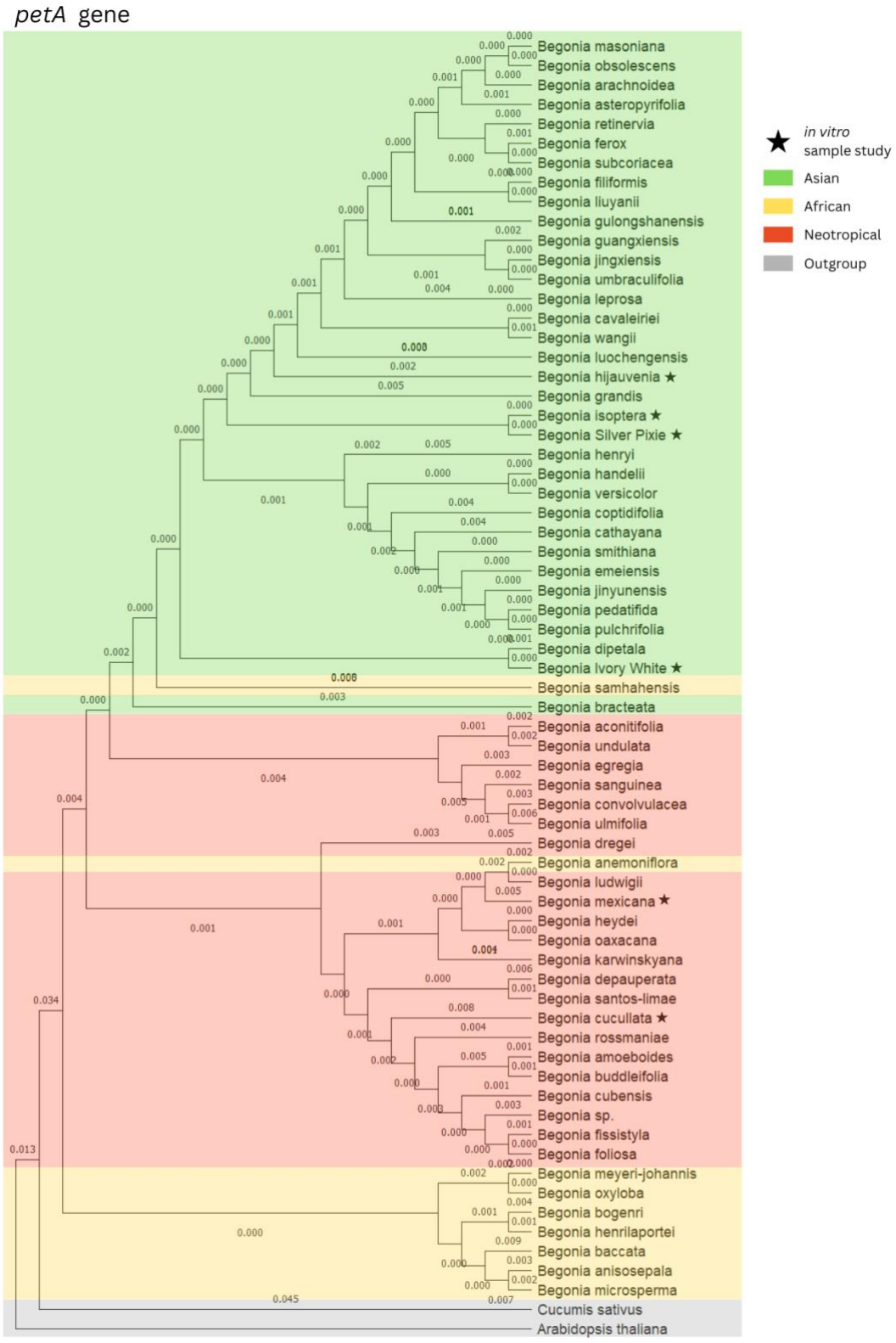

Figure 6 Phylogenetic tree alignment of *petA* with 59 sequences of *Begonia* Outgroup: *Cucumis sativa* (Order Cucurbitales) and *Arabidopsis thaliana*. The 59 sequence *Begonia* consists of 30 Asian *Begonia*, 20 Neotropical *Begonia*, and 9 African *Begonia*.

Table 12 BLASTx sequencing result of partial *rpoC1* (Merged R2 and R3) sequence

Species	Accession No.	Protein	Organism	Query Coverage (%)	E-value	Identity (%)
B. cucullata	YP_010744610.1	RNA polymerase β' subunit	*Begonia grandis*	46	1×10^{-150}	97.99
B. mexicana	YP_010595673.1		*Semecarpus reticulatus*	58	1×10^{-112}	75.90
B. hijauvenia	YP_009775163.1		*Begonia versicolor*	50	9×10^{-67}	97.12
B. 'Ivory White'	YP_010744520.1		*Begonia cathayana*	31	9×10^{-67}	97.12
B. isoptera	YP_010744520.1		*Begonia cathayana*	61	0.0	95.22
B. 'Silver Pixie'	YP_010744520.1		*Begonia cathayana*	61	0.0	96.44

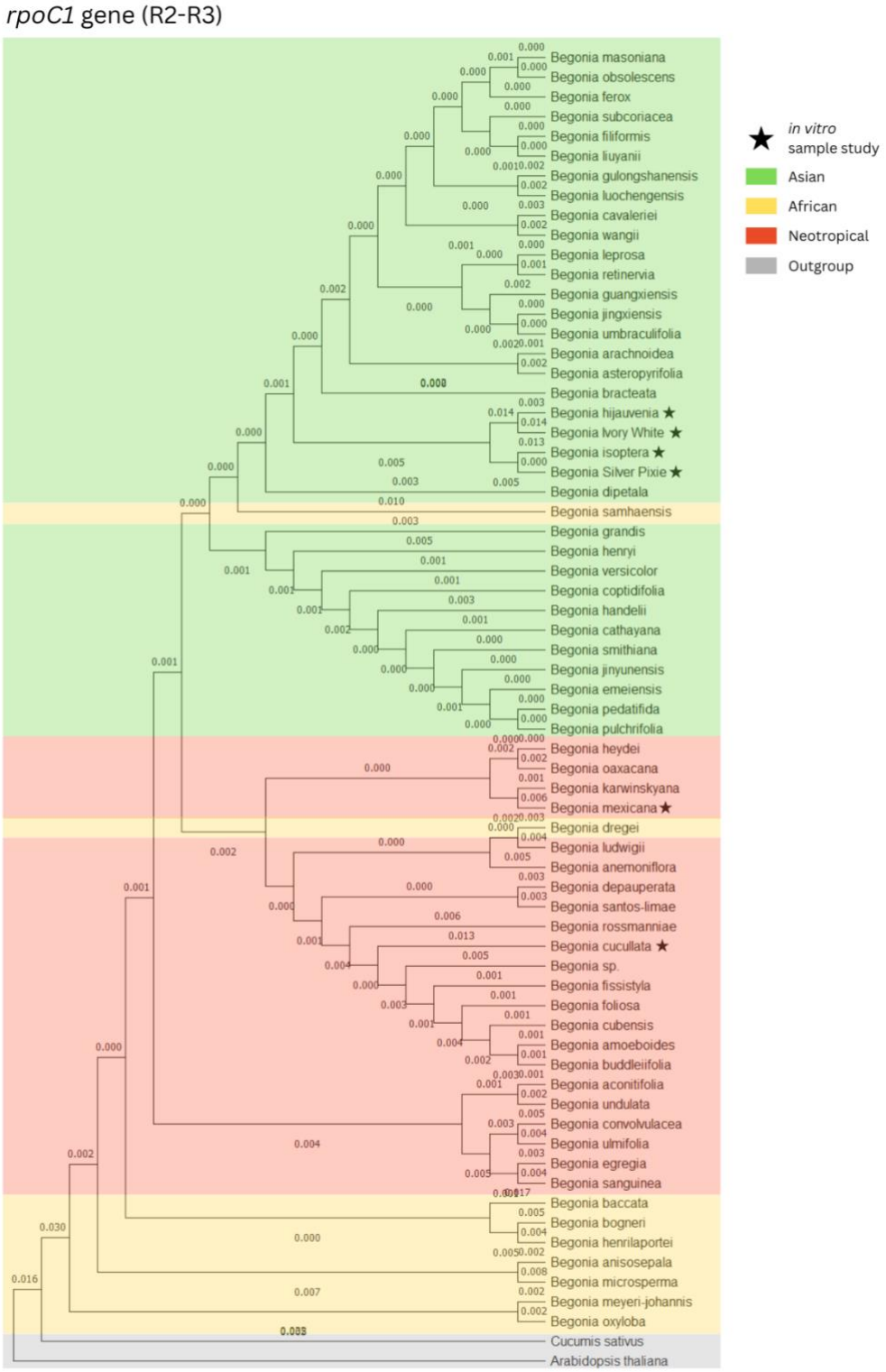

Figure 7 Phylogenetic tree alignment of partial *rpoC1* (R2-R3) with 59 sequences of *Begonia.* Outgroup: *Cucumis sativa* (Order Cucurbitales) and *Arabidopsis thaliana*. The 59 sequence Begonia consists of 30 Asian *Begonia*, 20 Neotropical *Begonia*, and 9 African *Begonia.*

DISCUSSION

Based on our findings, we can conclude that the *rpoC1* gene SNP, might be useful to differentiate between the semi-shade and deep shade *Begonia*. SNP 671 A/C and SNP 1679 A/G were associated with the two different traits, with SNP 671 A and SNP 1679 A related to semi-shade adaptations, while SNP 671 C and SNP 1679 G is related to deep shade adaptations. These two different sites were used as a template for Glu 224 and His 560 which played a significant role in semi-shade adaptation, while Arg 560 contributes to deep shade adaptability.

In *Panax ginseng*, wild-type variants showed an increase in *rpoC1* regulation, which caused the plant to be much more sensitive in light changes compared to cultivated Ginseng (Lee *et al.* 2012). Meanwhile, in Amaryllidaceae, Cheng *et al.* 2022 found that *rpoC1* experienced positive selection alongside other photosynthetic genes, including *petA* and *psbE*. Moraceae also experienced positive selection in *rpoC1* due to adaptation to low light environment (Zhang *et al.* 2022). This selection did not only happen in plants, but also in algae. *Ostreobium quekettii* which can be found in the depths below 200 m, also have *rpoC1* in positive selection due to the low light presence habitat (Marcelino *et al.* 2016). Other studies also found that *rpoC1* can be used as a marker for SNP. Tiwari *et al.* 2024 found out that *rpoC1* mutation (N571K) in wild barley and their SNP 1713 G/T is related to plasticity of photosynthetic rhythmicity.

Photosynthesis consisted of light and dark reactions. In light reaction, water molecules will be split by photolysis in Photosystem II (PSII) (Li & Kim 2022; Sun & Guo 2016). As this process continue, any excess photons will be considered as Reactive Oxidative Species (ROS), which are going to be used for gene expression, and chloroplast modification (Li & Kim 2022; Sun & Guo 2016). However, ROS accumulation from overactivation of PSII may cause plants to get oxidative stress which might be destructive to cell components (Khalid *et al.* 2023). Plant produces transcriptome of anthocyanin or Superoxide Dismutase (SOD) to decrease the overaccumulation of ROS (Cerqueira *et al.* 2023; Khalid *et al.* 2023). Therefore, any mutation in RNA polymerase would cause decreased transcription efficiency.

Like cyanobacteria or other prokaryotic photosynthetic organism's transcription mechanism, the Plastid Encoded Polymerase (PEP) is encoded by several genes including *rpoA*, *rpoB*, *rpoC1* and *rpoC2* (which are mostly similar to *rpoC* gene in bacteria) (Zhang *et al.* 2023). However, they also used nuclease that were made in the cell's nucleus to splice introns inside the chloroplast (Zhang *et al.* 2023). Although RpoC1 does not directly bind DNA, it contains conserved catalytic motifs essential for RNA synthesis and contributes the structural and functional stability of the chloroplast PEP core (do Prado *et al.* 2024). Furthermore, any mutation in this gene is mostly part of fine-tuning for the organism adaptation to

environmental pressure, including survivability in low light habitat (Börner *et al.* 2015; Du *et al.* 2021).

Most *Begonia* tend to adapt at shady environment. This was due to their habitat being in understory of rainforest canopy, limestone caves or karst, which mostly tend to have minimum light exposure (Li *et al.* 2022). This caused *Begonia* to have higher efficiency transcription under low light, which makes them have more survivability in deep shaded areas (Zhang *et al.* 2018). In other words, *Begonia* that are placed under brighter sunlight are more prone to oxidative stress, as mentioned before with signs such as a decrease in number of leaves, and overproduction of anthocyanin (Zhang *et al.* 2018).

While this research may have found the potential SNP sites to be used in determining the different traits of *Begonia* and able to predict the impact of the H560R mutation, we are not able to verify our *in silico* SNP findings due to the lack of full *rpoC1* CDS as mentioned above. The *rpoC1* gene contains an intron that is relatively more variable than the CDS, which caused primers such as our Read 1 reverse to not properly intact to the DNA template and Read 2 having more than one band (Figure 17). Meanwhile, we also failed to amplify the Read 4 of *rpoC1*, which should have contained the SNP 1679 site due to the false annealing position since the *rpoB* template (which located adjacent to *rpoC1* mentioned before) have a similar sequence to those from within the *rpoC1* gene, causing the unexpected amplicon size.

rpoC1 contained 2 exons, which size is estimated to be 431 bp, and 1.616 bp (Downie *et al.* 1996; Xiong *et al.* 2023). Partial exons sequenced might not be sufficient to predict positive selection analysis since not all the functional sites will be detected. Therefore, we need the full CDS sequence to predict the codon sequence and the translated protein sequence (Yang 2007; Álvarez-Carretero *et al.* 2023). Furthermore, whole genome sequencing DNA chloroplast, and transcriptome of the 6 *in silico* samples might be necessary to obtain the full CDS of *rpoC1*.

We were also not succeeding to amplify *psbE* from the six *in silico* samples. While the gene is not significant in the positive selection test, the differing climates of China, India, and the Indochinese region may still influence gene conservation. These regions tend to have seasonal climates with periods of drought or temperature fluctuations while, Indonesia has a consistently humid, tropical climate with relatively stable temperatures year-round. During the Oligocene to Miocene, regions like India and Indochina experienced seasonal climates and monsoon-driven dry periods due to the uplift of the Himalayas and increased aridity (Morley 2018). In contrast, areas such as Sunda land (including present-day Indonesia) came under perhumid tropical conditions, remaining consistently humid year-round, especially after Australian–Sunda landmass collisions (Morley 2018). Therefore, *psbE* may have different sequence in Indonesian species, which affected PCR amplification. This reason may also apply to the rest of rpoC1 that we failed to

amplify (e.g. *Begonia* 'Eternal Flame'). Therefore, further primer optimization or Whole Genome Sequencing for chloroplast is essential to successful PCR amplification.

This research also only provides SNP sites that are still considered as a potential site to differentiate the two shades adaptability. A mutation can be considered as an SNP if the number of mutations occurrence surpass >1% of the population number (Gambino *et al.* 2017; Fedorova *et al.* 2022). Otherwise, the variation is considered only as a Single Nucleotide Variation (SNV), which may represent a rare or individual-specific mutation without broader evolutionary or adaptive significance (Gambino *et al.* 2017; Fedorova *et al.* 2022). Meanwhile, we don't have the population data of Asian *Begonia*, since the *in silico* plant *Begonia* template chosen were randomly distribute mostly in East Asian regions and India (Appendix 6). Therefore, population study is also necessary to confirm whether this mutation can be considered as a predictive SNP.

In our study, we showed that SNP might be the right choice for plant cultivars to detect the difference of traits between semi and deep shaded adapted *Begonia* and how the two *Begonia* have different adaptation to light regulation. Based on our research result and mutation impact analysis, we know that the mutation H560R caused a significant change in protein SASA, flexibility, and electrostatics. This caused the entrance to PEP core to be more less accessible and RNA product exits the core less easily as shown in the SASA result. Electrostatic analysis further revealed that the mutation significantly increased the local domain energy, making the pore more positively charged. This change likely changes the efficiency of transcription. Interestingly, flexibility analysis showed that the local domain became more rigid due to the H560R mutation. This finding while not directly aligned with Xiong *et al.* 2023, who suggested that the mutation may affect transcriptional accuracy, it provides new insight into how positive selection might contribute to the adaptation of deep-dwelling plants. Furthermore, it raises the possibility that the mutation change the enzyme functionality as a response to adaptive change under specific light conditions.

Based on our analysis above, we might conclude that semi-shade *Begonia* tends to have more flexible RNA polymerase, allowing the plant to better adapt more in sunlight or semi-open habitat. This caused the plant produce higher number of gene regulation, including the SOD enzyme, anthocyanin or other genes that can reduced the number of excessive photons. In contrast, the deep shade *Begonia* have much more rigid and accessible enzymes for much more energy conserved transcription. Thus, the smaller number of transcripts is sufficient to reduce lower number of ROS accumulation. However, when exposed to brighter light, deep-shade-adapted plants experience excess excitation energy, leading to overproduction of reactive oxygen species (ROS), causing the plants to be more vulnerable to oxidative stress and their tendency to become more prone to sunburn.

This contributes to their lower survival rates in semi-open or high-light environments.

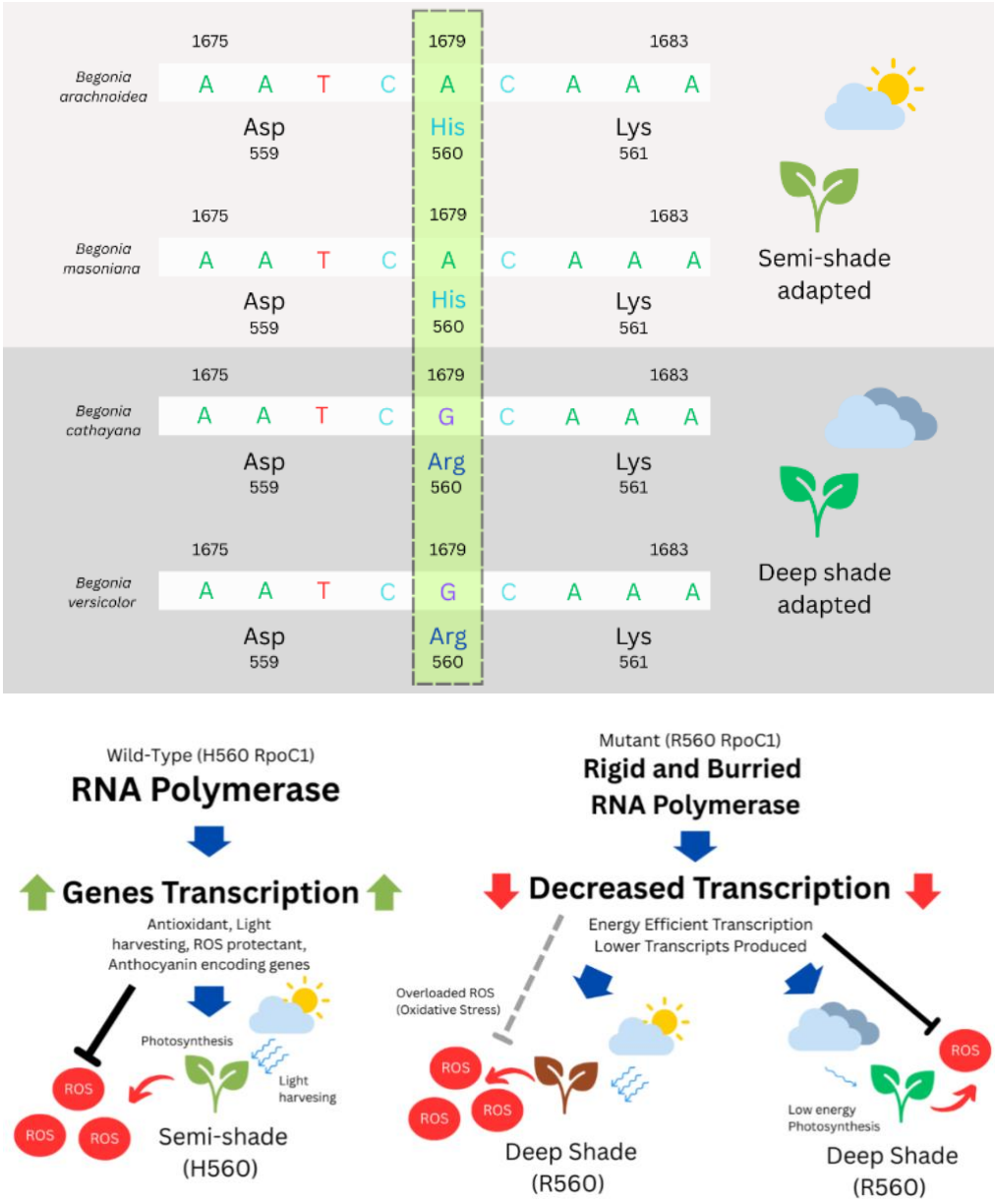

Figure 8 Different mechanisms light perception on semi-shade and deep shade *Begonia*

This difference in adaptation mechanisms (Figure 8) provides insight into how plants modulate their plastid transcription machinery to cope with environmental stress, particularly low-light conditions. It highlights the role of chloroplast-encoded genes like *rpoC1* in enabling fine-scale regulatory changes that contribute to survival and ecological success. Furthermore, it suggests that molecular adaptation in core transcriptional components. However, these findings still require further verification from *in vitro* studies.

While this research may have found the potential SNP sites to be used in determining the different traits of *Begonia* and able to predict the impact of the H560R mutation, we are not able to verify our *in silico* SNP findings due to the lack of full *rpoC1* CDS as mentioned above. The *rpoC1* gene contains an intron that is relatively more variable than the CDS, which caused primers such as our Read 1 reverse to not properly intact to the DNA template and Read 2 having more than one band (Figure 17). *rpoC1* gene contained 2 exons, which size is estimated to be 431 bp, and 1.616 bp (Downie *et al.* 1996; Xiong *et al.* 2023). Partial exons sequenced might not be sufficient to predict positive selection analysis since not all the functional sites will be detected. Therefore, we need the full CDS sequence to predict the codon sequence and the translated protein sequence (Álvarez-Carretero *et al.* 2023; Yang 2007). We were also failed to amplify *psbE* from the six *in silico* samples. While the gene is not significant in the positive selection test, the differing climates of China, India, and the Indochinese region may still influence gene conservation. These regions tend to have seasonal climates with periods of drought or temperature fluctuations while, Indonesia has a consistently humid, tropical climate with relatively stable temperatures year-round (Morley 2018). Furthermore, whole genome sequencing DNA chloroplast, and transcriptome of the 6 *in silico* samples might be necessary to obtain the full CDS of *rpoC1* and *psbE*.

Based on our findings, we conclude that SNPs in the *rpoC1* gene, particularly SNP 671 A/C and SNP 1679 A/G can potentially differentiate semi-shade and deep-shade *Begonia*, with semi-shade plants carrying the A alleles and deep-shade plants carrying the C and G alleles, respectively. These mutations, especially H560R, are predicted to influence the plastid RNA polymerase's surface accessibility, electrostatics, and rigidity, which may fine-tune transcription efficiency under varying light conditions. This study gives us a novel understanding on how plastid-encoded genes like *rpoC1* contribute to light adaptation, offering the first evidence in *Begonia* that structural SNPs in chloroplast transcription machinery. However, the lack of full coding sequences (CDS), experimental validation, and population-level data limits the ability to confirm whether these SNPs are fixed adaptive markers or their precise functional impact. Future work should include complete sequencing of *rpoC1*, transcriptomic validation of the mutation's regulatory effects, and broader population studies to confirm allele frequency and ecological relevance. This will help determine whether these mutations are conserved adaptations or individual variations and enable cultivars to detect light-resilient *Begonia* parental varieties, especially the female parental since chloroplast is maternally inherited. Hence, by choosing the right parental, plant breeders can find the suitable plant to breed plants with favoured traits.

CONCLUSION

Our research objective was to identify potential SNP markers that can distinguish semi-shade adapted *Begonia* from the deep shade adapted *Begonia,* and to predict how the two traits differ in their adaptation to varying sunlight. Based on our findings, we discovered that SNP 671 A/C and SNP 1679 A/G in *rpoC1* might have the potential to differentiate between the two plant groups. Among the two discovered sites, A is significantly associated with semi-shade plants, while C at site 671 or G at site 1679 is significantly associated with deep shade plant. Additionally, we identified that the H560R mutation plays a significant role in determining shade adaptation, with histidine (H) contributing to the semi-shade trait, while arginine (R) contributing to deep shade adaptation. The mutation has a notable impact on the flexibility and surface accessibility of the RNA polymerase β' subunit (RpoC1) resulting in more energy-efficient RNA polymerization under deep shade conditions.

These findings can assist plant breeders in selecting suitable parental lines to develop *Begonia* cultivars with favorable shade adaptability, while also providing insights into adaptation mechanisms in other shade-adapted plants species. However, the lack of full CDS sequences and population-level data limits the ability to verify these SNPs as fixed adaptive markers. Future work should include complete sequencing of *rpoC1*, transcriptomic validation, experimental mutagenesis verification, and broader population studies to confirm allele frequency and ecological relevance. This research provides candidate SNP markers for shade adaptation in *Begonia*, offering practical tools for breeding and knowledge into light-related plant evolutionary mechanisms.

REFERENCES

Álvarez-Carretero S, Kapli P, Yang Z. 2023. Beginner's guide on the use of PAML to detect positive selection. Molecular Biology and Evolution. 40(4):1–18. DOI:10.1093/molbev/msad041.

Ayu SAD, Putrika A, Girmansyah D. 2019. Analisis morfologi *Begonia isoptera* Dryand. ex Sm (Begoniaceae) Kompleks di Jawa. Floribunda. 6(2). DOI:10.32556/floribunda.v6i2.2019.263.

Blum M, Andreeva A, Cavalcanti Florentino L, Rocio Chuguransky S, Grego T, Hobbs E, Lazaro Pinto B, Orr A, Pa ysan-Laf osse T, Ponamareva I, *et al*. 2025. InterPro: the protein sequence classification resource in 2025. Nucleic Acids Research. 53(D1):D444–D456. DOI:10.1093/NAR/GKAE1082.

Börner T, Aleynikova AY, Zubo YO, Kusnetsov V V. 2015. Chloroplast RNA polymerases: role in chloroplast biogenesis. Biochimica et Biophysica Acta - Bioenergetics. 1847(9):761–769. DOI:10.1016/j.bbabio.2015.02.004.

Cerqueira JVA, de Andrade MT, Rafael DD, Zhu F, Martins SVC, Nunes-Nesi A, Benedito V, Fernie AR, Zsögön A. 2023. Anthocyanins and reactive oxygen species: a team of rivals regulating plant development? Plant Molecular Biology. 112(4–5):213–223. DOI:10.1007/s11103-023-01362-4.

Cheng RY, Xie DF, Zhang XY, Fu X, He XJ, Zhou SD. 2022. Comparative plastome analysis of three Amaryllidaceae subfamilies: Insights into variation of genome characteristics, phylogeny, and adaptive evolution. BioMed Research International. 2022. DOI:10.1155/2022/3909596.

Cilia E, Pancsa R, Tompa P, Lenaerts T, Vranken WF. 2014. The DynaMine webserver: predicting protein dynamics from sequence. Nucleic Acids Research. 42(W1):W264–W270. DOI:10.1093/NAR/GKU270.

Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, et al. 2009. Biopython: freely available python tools for computational molecular biology and bioinformatics. Bioinformatics. 25(11):1422–1423. DOI:10.1093/bioinformatics/btp163.

Cramer P, Bushnell DA, Kornberg RD. 2001. Structural basis of transcription: RNA polymerase II at 2.8 ångstrom resolution. Science. 292(5523):1863–1876. DOI:10.1126/SCIENCE.1059493.

Ding B, Nakamura K, Kono Y, Ho MJ, Peng CI. 2014. *Begonia jinyunensis* (Begoniaceae, section Platycentrum), a new palmately compound leaved species from Chongqing, China. Botanical Studies. 55(1):1–8. DOI:10.1186/s40529-014-0062-6.

Downie SR, Katz-Downie DS, Cho KJ. 1996. Phylogenetic analysis of Apiaceae subfamily Apioideae using nucleotide sequences from the chloroplast *rpoC1* intron. Molecular Phylogenetics and Evolution. 6(1):1–18. DOI:10.1006/mpev.1996.0053.

Du Y, Mo W, Ma T, Tang W, Tian L, Lin R. 2021. A pentatricopeptide repeat protein DUA1 interacts with sigma factor 1 to regulate chloroplast gene expression in Rice. Photosynthesis Research. 147(2):131–143. DOI:10.1007/s11120-020-00793-0.

Emelianova K, Kidner C. 2022. Comparative transcriptome analysis of two closely

related begonia species reveals divergent patterns in key light-regulated pathways. Edinburgh Journal of Botany. 79:1–18. DOI:10.24823/ejb.2022.398.

Engel C, Sainsbury S, Cheung AC, Kostrewa D, Cramer P. 2013. RNA polymerase i structure and transcription regulation. Nature. 502(7473):650–655. DOI:10.1038/NATURE12712.

Fedorova L, Khrunin A, Khvorykh G, Lim J, Thornton N, Mulyar OA, Limborska S, Fedorov A. 2022. Analysis of common SNPs across continents reveals major genomic differences between human populations. Genes. 13(8). DOI:10.3390/genes13081472.

Fraczkiewicz R, Braun W. 1998. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. Journal of Computational Chemistry. 19(3):319–333. DOI:10.1002/(SICI)1096-987X(199802)19:3<319::AID-JCC6>3.0.CO;2-W.

Gambino G, Dal Molin A, Boccacci P, Minio A, Chitarra W, Avanzato CG, Tononi P, Perrone I, Raimondi S, Schneider A, *et al*. 2017. Whole-genome sequencing and SNV genotyping of "Nebbiolo" (*Vitis vinifera* L.) clones. Scientific Reports. 7(1):1–15. DOI:10.1038/s41598-017-17405-y.

Girmansyah D, Hughes M, Sulistijorini, Rugayah, Ardi WH, Chikmawati T. 2022. Six new species of *Begonia* (Sect. Jackia, Begoniaceae) from Sumatra, Indonesia. Taiwania. 67(1):97–109. DOI:10.6165/tai.2022.67.97.

Handoyo Ardi W, Thomas D. 2022. Synopsis of *Begonia* (Begoniaceae) from the northern arm of Sulawesi and Sangihe Island, Indonesia, including three new species. Edinburgh Journal of Botany. 79:1–50. DOI:10.24823/ejb.2022.405.

Hoang TS, Lin C-W. 2023. "*Begonia filiformis* Irmscher 1939", a new species *Begonia wiformis*, and a new record of *B. filiformis* (B. sect. Coelocentrum, Begoniaceae) from nothern Vietnam. DOI:10.5281/zenodo.7835561.

Huang X, Li J, Pang X, Zhu J, Pan J, Li Y, Tang J. 2023. Gene polymorphism and prediction of toxicity to platinum-based chemotherapy in patients with gynecologic cancer. Clinical and Translational Science. 16(12):2519–2529. DOI:10.1111/CTS.13642,.

Hughes M, Girmansyah D. 2011. Searching for sumatran *Begonia* described by William Jack: following in the footsteps of a 19th century scottish botanist. Gardens' Bulletin Singapore. 63:83–96.

Hughes M, Miller AG. 2002. A new endemic species of *Begonia* (Begoniaceae) from the Socotra archipelago. Edinburgh Journal of Botany. 59(2):273–281. DOI:10.1017/s0960428602000082.

Indrakumar I, Karpagam S, Jayaraman P. 2013. Anatomical protocol of *Begonia dipetala* Graham for the specific identity of the plant. International Journal of Plant Research. 2013(3):27–38. DOI:10.5923/j.plant.20130303.03.

Johnson VM, Pakrasi HB. 2022. Advances in the understanding of the lifecycle of photosystem II. Microorganisms. 10(5). DOI:10.3390/microorganisms10050836.

Jurrus E, Engel D, Star K, Monson K, Brandi J, Felberg LE, Brookes DH, Wilson L, Chen J, Liles K, *et al*. 2018. Improvements to the APBS biomolecular solvation software suite. Protein Science. 27(1):112–128. DOI:10.1002/PRO.3280,.

Khalid MF, Huda S, Yong M, Li Lihua, Li Li, Chen ZH, Ahmed T. 2023. Alleviation of drought and salt stress in vegetables: crop responses and mitigation strategies. Plant Growth Regulation. 99(2):177–194. DOI:10.1007/s10725-022-00905-x.

Kim KW, Nawade B, Nam J, Chu SH, Ha J, Park YJ. 2022. Development of an inclusive 580K SNP array and its application for genomic selection and genome-wide association studies in rice. Frontiers in Plant Science. 13:1–15. DOI:10.3389/fpls.2022.1036177.

Kita Y, Ito M. 2000. Nuclear ribosomal ITS sequences and phylogeny in east asian aconitum subgenus *Aconitum* (Ranunculaceae), with special reference to extensive polymorphism in individual plants. Plant Systematics and Evolution. 225(1):1–13. DOI:10.1007/BF00985455.

Kumar K, Gupta P, Singh KN, Nirgude MS, Srivastava H, Sharma S, Sevanthi AM, Durgesh K, Jain PK, Gaikwad K. 2023. Whole chloroplast genome-specific non-synonymous SNPs reveal the presence of substantial diversity in the pigeonpea mini-core collection. 3 Biotech. 13(11):1–15. DOI:10.1007/s13205-023-03785-8.

Lee K-H, Kwon K-R, Kang W-M, Jeon E-M, Jang J-H. 2012. Identification and analysis of the chloroplast *rpoC1* gene differentially expressed in wild ginseng. Journal of Pharmacopuncture. 15(2):20–23. DOI:10.3831/kpi.2012.15.2.020.

Li L, Chen X, Fang D, Dong S, Guo X, Li N, Campos-Dominguez L, Wang Wenguang, Liu Y, Lang X, et al. 2022. Genomes shed light on the evolution of *Begonia*, a mega-diverse genus. New Phytologist. 234(1):295–310. DOI:10.1111/nph.17949.

Li M, Kim C. 2022. Chloroplast ROS and stress signaling. Plant Communications. 3(1):100264. DOI:10.1016/j.xplc.2021.100264.

Marcelino VR, Cremen MCM, Jackson CJ, Larkum AAW, Verbruggen H. 2016. Evolutionary dynamics of chloroplast genomes in low light: A case study of the endolithic green alga *Ostreobium quekettii*. Genome Biology and Evolution. 8(9):2939–2951. DOI:10.1093/gbe/evw206.

Moonlight P, Fuentes AF. 2022. an Updated Checklist and key to the species of bolivian *Begonia*, including one new species. Edinburgh Journal of Botany. 79:1–66. DOI:10.24823/ejb.2022.407.

Morley RJ. 2018. Assembly and division of the South and South-East Asian flora in relation to tectonics and climate change. Journal of Tropical Ecology. 34(4):209–234. DOI:10.1017/S0266467418000202.

Orlando G, Raimondi D, Codicè F, Tabaro F, Vranken W. 2022. Prediction of disordered regions in proteins with recurrent neural networks and protein dynamics. Journal of Molecular Biology. 434(12):167579. DOI:10.1016/j.jmb.2022.167579.

Park HS, Lee WK, Lee SC, Lee HO, Joh HJ, Park JY, Kim S, Song K, Yang TJ. 2021. Inheritance of chloroplast and mitochondrial genomes in cucumber revealed by four reciprocal F1 hybrid combinations. Scientific Reports. 11(1):1–11. DOI:10.1038/s41598-021-81988-w.

Peng CI, Ku SM, Kono Y, Chung KF, Liu Y. 2008. Two new species of *Begonia* (sect. Coelocentrum, Begoniaceae) from limestone areas in Guangxi, China: *B. arachnoidea* and *B. subcoriacea*. Botanical Studies. 49(4):405–418.

Peters JM, Vangeloff AD, Landick R. 2011. Bacterial transcription terminators: The RNA 3′-end chronicles. Journal of Molecular Biology. 412(5):793–813. DOI:10.1016/j.jmb.2011.03.036.

Pires DEV, Ascher DB, Blundell TL. 2014. DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. Nucleic Acids Research. 42(W1):W314–W319. DOI:10.1093/NAR/GKU411.

do Prado PFV, Ahrens FM, Liebers M, Ditz N, Braun HP, Pfannschmidt T, Hillen HS. 2024. Structure of the multi-subunit chloroplast RNA polymerase. Molecular Cell. 84(5):910-925.e5. DOI:10.1016/j.molcel.2024.02.003.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, De Bakker PIW, Daly MJ, *et al*. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. american journal of human genetics. 81(3):559. DOI:10.1086/519795.

Raghothama S. 2010. REVIEWS NMR of peptides. Journal of the Indian Institute of Science. 90(1):90.

Raimondi D, Orlando G, Pancsa R, Khan T, Vranken WF. 2017. Exploring the sequence-based prediction of folding initiation sites in proteins. Scientific Reports. 7(1). DOI:10.1038/S41598-017-08366-3,.

Severinov K, Mustaev A, Kukarin A, Muzzin O, Bass I, Darst SA, Goldfarb A. 1996. Structural modules of the large subunits of RNA Polymerase: introducing archaebacterial and chloroplast split sites in the β and β′ subunits of *Escherichia coli* RNA polymerase. Journal of Biological Chemistry. 271(44):27969–27974. DOI:10.1074/JBC.271.44.27969.

Stucky BJ. 2012. Seqtrace: a graphical tool for rapidly processing DNA sequencing chromatograms. Journal of Biomolecular Techniques. 23(3):90–93. DOI:10.7171/jbt.12-2303-004.

Sun AZ, Guo FQ. 2016. Chloroplast retrograde regulation of heat stress responses in plants. Frontiers in Plant Science. 7:1–16. DOI:10.3389/fpls.2016.00398.

Tamura K, Stecher G, Kumar S. 2021. MEGA11: molecular evolutionary genetics analysis version 11. Molecular Biology and Evolution. 38(7):3022–3027. DOI:10.1093/molbev/msab120.

Tiwari LD, Bdolach E, Prusty MR, Bodenheimer S, Be'ery A, Faigenboim-Doron A, Yamamoto E, Panzarová K, Kashkush K, Shental N, *et al*. 2024. Cytonuclear interactions modulate the plasticity of photosynthetic rhythmicity and growth in wild barley. Physiologia Plantarum. 176(1):1–16. DOI:10.1111/ppl.14192.

Wahyuni S, Siregar H-M, Isnaini Y, Widiarsih S, Dwimahyani I, Witono JR, Ritonga AW. 2023. Increasing variability of *Begonia isoptera* dryand. ex Sm. by gamma-ray irradiation for developing new varieties. AIP Conference Proceedings. 2972(1):60008. DOI:10.1063/5.0182836.

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, De Beer TAP, Rempfer C, Bordoli L, et al. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Research. 46(W1):W296–W303. DOI:10.1093/nar/gky427.

Xiong C, Huang Y, Li Z, Wu L, Liu Z, Zhu W, Li J, Xu R, Hong X. 2023. Comparative chloroplast genomics reveals the phylogeny and the adaptive evolution of *Begonia* in China. BMC Genomics. 24(1).

DOI:10.1186/s12864-023-09563-3.
Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution. 24(8):1586–1591. DOI:10.1093/molbev/msm088.
Yang Z, Wong WSW, Nielsen R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. Molecular Biology and Evolution. 22(4):1107–1118. DOI:10.1093/molbev/msi097.
Zdobnov EM, Apweiler R. 2001. InterProScan – an integration platform for the signature-recognition methods in InterPro. Bioinformatics. 17(9):847–848. DOI:10.1093/BIOINFORMATICS/17.9.847.
Zhang Y, Liu A, Zhang X, Huang S. 2018. Effects of shading on some morphological and physiological characteristics of *Begonia semperflorens*. Pak J Bot. 50(6):2173–2179.
Zhang Y, Tian L, Lu C. 2023. Chloroplast gene expression: Recent advances and perspectives. Plant Communications. 4(5). DOI:10.1016/j.xplc.2023.100611.
Zhang ZR, Yang X, Li WY, Peng YQ, Gao J. 2022. Comparative chloroplast genome analysis of ficus (Moraceae): Insight into adaptive evolution and mutational hotspot regions. Frontiers in Plant Science. 13(September):1–17. DOI:10.3389/fpls.2022.965335.
Zhao K, Liu A, Zhang Y, Liu W, Zhao Z, Yang S. 2024. Comparative transcriptome analysis reveals changes in gene expression associated with anthocyanin metabolism in *Begonia semperflorens* under light conditions. Horticulturae. 10(1):96. DOI:10.3390/HORTICULTURAE10010096/S1.

APPENDICES

Appendix 1 Biopython code for parsing gene from chloroplast (example shown for *rpoC1*)

```
#Install Biopythom
!pip install biopython

#Check the installed Biopython version
import Bio
print(Bio.__version__)

import os
from Bio import Entrez, SeqIO

# Fetch a GenBank record
def fetch_genbank_record(accession):
    Entrez.email = "rafaelangeloyudhistira@gmail.com"   # Always
provide your email
    handle    =    Entrez.efetch(db="nucleotide",    id=accession,
rettype="gb", retmode="text")
    record = SeqIO.read(handle, "genbank")
    handle.close()
    return record

# Function to extract rpoC1 gene sequence
def extract_rpoC1_sequence(record):
    for feature in record.features:
        if feature.type == "gene" and "gene" in feature.qualifiers:
            if feature.qualifiers["gene"][0].lower() == "rpoc1":
                return feature.location.extract(record).seq
    return None

# List of accession numbers for different species
accessions = ["NC_063512.1","NC_065014.1","NC_070310.1",
"NC_073119.1","NC_073117.1","NC_056110.1",
"NC_070315.1","NC_061410.1","NC_067030.1",
"PQ675783.1","NC_073120.1","NC_046385.1",
"NC_063513.1","NC_065245.1","NC_070321.1",
"PQ572754.1","NC_068748.1","NC_073118.1",
"PQ619426.1","PQ675781.1","OP618127.1",
"NC_073122.1","OR288087.1","NC_045096.1",
"NC_088496.1","NC_070330.1","NC_073123.1",
"PQ619425.1","NC_073121.1","NC_047450.1","PQ675782.1"]  # Add your
accession numbers here

# Corresponding species names for the accession numbers
species_names = {
    "NC_063512.1" : "Begonia arachnoidea",
```

```
    "NC_065014.1" : "Begonia asteropyrifolia",
    "NC_070310.1" : "Begonia bracteata voucher Peng23521",
    "NC_073119.1" : "Begonia cathayana",
    "NC_073117.1" : "Begonia cavaleriei",
    "NC_056110.1" : "Begonia coptidifolia",
    "NC_070315.1" : "Begonia dipetala voucher Peng22521",
    "NC_061410.1" : "Begonia emeiensis",
    "NC_067030.1" : "Begonia ferox",
    "PQ675783.1" : "Begonia filiformis",
    "NC_073120.1" : "Begonia grandis",
    "NC_046385.1" : "Begonia guangxiensis plastid",
    "NC_063513.1" : "Begonia gulongshanensis",
    "NC_065245.1" : "Begonia handelii",
    "NC_070321.1" : "Begonia henryi voucher RBGE20141517",
    "PQ572754.1" : "Begonia jingxiensis",
    "NC_068748.1" : "Begonia jinyunensis",
    "NC_073118.1" : "Begonia leprosa",
    "PQ619426.1" : "Begonia liuyanii",
    "PQ675781.1" : "Begonia luochengensis",
    "OP618127.1" : "Begonia masoniana",
    "NC_073122.1" : "Begonia obsolescens",
    "OR288087.1" : "Begonia pedatifida",
    "NC_045096.1" : "Begonia pulchrifolia",
    "NC_088496.1" : "Begonia retinervia",
    "NC_070330.1" : "Begonia samhaensis voucher RBGE19990398",
    "NC_073123.1" : "Begonia smithiana",
    "PQ619425.1" : "Begonia subcoriacea",
    "NC_073121.1" : "Begonia umbraculifolia",
    "NC_047450.1" : "Begonia versicolor",
    "PQ675782.1" : "Begonia wangii",
}

# Folder to store the output files
output_folder = "rpoC1_output"
os.makedirs(output_folder, exist_ok=True)

# Fetch records and extract rpoc1 sequences, writing each to a
separate file
for accession in accessions:
    print(f"Processing accession: {accession}")
    record = fetch_genbank_record(accession)
    rpoC1_sequence = extract_rpoC1_sequence(record)
    species_name = species_names.get(accession, "Unknown species")

    output_file_path         =         os.path.join(output_folder,
f"{species_name}_{accession}_rpoC1.txt")
    with open(output_file_path, "w") as output_file:
        if rpoC1_sequence:
            output_file.write(f">{species_name}|{accession}|rpoC1
gene\n{rpoC1_sequence}\n")
        else:
```

```
            output_file.write("No rpoC1 gene found.\n")

    print(f"Output written to {output_file_path}")

#Zip the output file for easy download
!zip -r /content/rpoC1_output.zip /content/rpoC1_output/
```

Appendix 2 Biopython code for parsing intergenic regions from chloroplast (example shown for *rpoC1-rpoB*)

```
from Bio import Entrez, SeqIO

# Function to fetch a GenBank record
def fetch_genbank_record(accession):
    Entrez.email = "rafaelangeloyudhistira@gmail.com"   # Always
provide your email
    handle    =    Entrez.efetch(db="nucleotide",    id=accession,
rettype="gb", retmode="text")
    record = SeqIO.read(handle, "genbank")
    handle.close()
    return record

# Function to extract intergenic regions between specified genes
def extract_intergenic_regions(record, gene_list):
    intergenic_regions = []
    gene_positions = []

    # Collect the positions of the specified genes
    for feature in record.features:
        if feature.type == "gene":
            gene_name = None
            for qualifier in ["gene", "locus_tag"]:
                if qualifier in feature.qualifiers:
                    gene_name = feature.qualifiers[qualifier][0]
                    break
            if gene_name in gene_list:
                gene_positions.append((feature.location.start,
feature.location.end, gene_name))

    # Sort the gene positions by their start positions
    gene_positions.sort()

    # Extract intergenic regions between the specified genes
    for i in range(len(gene_positions) - 1):
        start = gene_positions[i][1]
        end = gene_positions[i + 1][0]
        if start < end:
            intergenic_region = record.seq[start:end]
            intergenic_regions.append((gene_positions[i][2],
gene_positions[i + 1][2], intergenic_region))

    return intergenic_regions

# List of accession numbers for different species
accessions = ["NC_063512.1","NC_065014.1","NC_070310.1",
"NC_073119.1","NC_073117.1","NC_056110.1",
"NC_070315.1","NC_061410.1","NC_067030.1",
"PQ675783.1","NC_073120.1","NC_046385.1",
```

```
"NC_063513.1","NC_065245.1","NC_070321.1",
"PQ572754.1","NC_068748.1","NC_073118.1",
"PQ619426.1","PQ675781.1","OP618127.1",
"NC_073122.1","OR288087.1","NC_045096.1",
"NC_088496.1","NC_070330.1","NC_073123.1",
"PQ619425.1","NC_073121.1","NC_047450.1","PQ675782.1"]  # Add your
accession numbers here

# Corresponding species names for the accession numbers
species_names = {
    "NC_063512.1" : "Begonia arachnoidea",
    "NC_065014.1" : "Begonia asteropyrifolia",
    "NC_070310.1" : "Begonia bracteata voucher Peng23521",
    "NC_073119.1" : "Begonia cathayana",
    "NC_073117.1" : "Begonia cavaleriei",
    "NC_056110.1" : "Begonia coptidifolia",
    "NC_070315.1" : "Begonia dipetala voucher Peng22521",
    "NC_061410.1" : "Begonia emeiensis",
    "NC_067030.1" : "Begonia ferox",
    "PQ675783.1" : "Begonia filiformis",
    "NC_073120.1" : "Begonia grandis",
    "NC_046385.1" : "Begonia guangxiensis plastid",
    "NC_063513.1" : "Begonia gulongshanensis",
    "NC_065245.1" : "Begonia handelii",
    "NC_070321.1" : "Begonia henryi voucher RBGE20141517",
    "PQ572754.1" : "Begonia jingxiensis",
    "NC_068748.1" : "Begonia jinyunensis",
    "NC_073118.1" : "Begonia leprosa",
    "PQ619426.1" : "Begonia liuyanii",
    "PQ675781.1" : "Begonia luochengensis",
    "OP618127.1" : "Begonia masoniana",
    "NC_073122.1" : "Begonia obsolescens",
    "OR288087.1" : "Begonia pedatifida",
    "NC_045096.1" : "Begonia pulchrifolia",
    "NC_088496.1" : "Begonia retinervia",
    "NC_070330.1" : "Begonia samhaensis voucher RBGE19990398",
    "NC_073123.1" : "Begonia smithiana",
    "PQ619425.1" : "Begonia subcoriacea",
    "NC_073121.1" : "Begonia umbraculifolia",
    "NC_047450.1" : "Begonia versicolor",
    "PQ675782.1" : "Begonia wangii",
}

# List of genes of interest
genes_of_interest = ["rpoC1", "rpoB"]  # Replace with actual gene
names

# Fetch records and extract intergenic regions
for accession in accessions:
    print(f"Processing accession: {accession}")
    record = fetch_genbank_record(accession)
```

```
    intergenic_regions = extract_intergenic_regions(record,
genes_of_interest)
    species_name = species_names.get(accession, "Unknown species")
    print(f"Intergenic regions for {species_name} ({accession}):")
    if not intergenic_regions:
        print("No intergenic regions found between the specified
genes.")
    for idx, (gene1, gene2, region) in
enumerate(intergenic_regions):
        print(f">{species_name}|{accession}|region between {gene1}
- {gene2} \n{region}\n")

import os

# Folder to store the output files
output_folder = "rpoC1-rpoB_output"
os.makedirs(output_folder, exist_ok=True)

# Fetch records and extract intergenic regions, writing each to a
separate file
for accession in accessions:
    print(f"Processing accession: {accession}")
    record = fetch_genbank_record(accession)
    intergenic_regions = extract_intergenic_regions(record,
genes_of_interest)

    species_name = species_names.get(accession, "Unknown species")
# Move this line inside the loop

    output_file_path = os.path.join(output_folder,
f"{species_name} {accession}_rpoC1-rpoB.txt")
    with open(output_file_path, "w") as output_file:
        if not intergenic_regions:
            output_file.write("No intergenic regions found between
the specified genes.\n")
        else:
            for idx, (gene1, gene2, region) in
enumerate(intergenic_regions):

output_file.write(f">{species_name}|{accession}|region between
{gene1} - {gene2} \n{region}\n")

    print(f"Output written to {output_file_path}")

#Create a zip file
!zip -r /content/rpoC1-rpoB_output.zip /content/rpoC1-rpoB_output/
```

Appendix 3 Bio2Byte tools values interpretation

Bio2Byte Tools	Values Interpretation
Dynamine (Backbone) (Cilia *et al.* 2014).	Above 1.0: Membrane spanning regions Above 0.8: Rigid 0.8 – 0.69: 'Context' dependent Below: 0.69: Flexible
Dynamine (Sidechain) (Cilia *et al.* 2014).	Higher values mean more likely rigid (Depends on the amino acid type)
Dynamine (Conformational propensities sheet, helix, coil) (Cilia *et al.* 2014).	Higher values mean higher propensities
EFoldMine (Raimondi *et al.* 2017).	Above 0.169: likely to start the protein folding process based on only local interactions with other amino acid
Disomine (Orlando *et al.* 2022).	Above 0.5: Disordered residue

Appendix 4 d_i score for every possible degenerate base

Base	Nucleotide	d_i
A	A	1
T	T	1
C	C	1
G	G	1
R	A atau G	2
Y	C atau T	2
S	G atau C	2
W	A atau T	2
K	G atau T	2
M	A atau C	2
B	C, G, atau T	3
D	A, G, atau T	3
H	A, C, atau T	3
V	A, C, atau G	3
N	A, T, C, atau T	4

Appendix 5 PAML installation and codeml command used to run in terminal (Github Bash)

```
# Make a directory for software
mkdir -p /workspaces/path/to/PAML && cd /workspaces/path/to/PAML

# Clone the GitHub mirror of PAML
git clone https://github.com/abacus-gene/paml.git

# Enter the source directory
cd paml/src

# Compile the source code
make

# Optionally, move executables to your ~/bin directory
mkdir -p ~/bin
cp codeml baseml pamp evolver yn00 chi2 ~/bin/

# Add ~/bin to your PATH if it's not already
echo 'export PATH="$HOME/bin:$PATH"' >> ~/.bashrc
source ~/.bashrc

#Verify the installation
codeml

#Create codeml.ctl file in terminal
nano codeml.ctl

#Insert this command to codeml.ctl
```

```
seqfile = #Path to phylip file
treefile = #Path to tree file
outfile = results_(gene).txt

* Model specifications:
seqtype = 1      # Codon sequences
runmode = 0                              * 0 = pairwise, 1 = small
dataset, 2 = large dataset
model = 0                              * 0 = one ratio, 1 = nearly
neutral, 2 = positive selection, etc.
NSsites = 7 8                       * Site model: 0 = one-ratio, 1
= neutral, 2 = selection, 8 = beta&ω # 0: No site models (for
branch models)

* Omega settings (dN/dS ratio):
fix_omega = 0                          * 0: Estimate omega, 1: Fix
omega # 0: Allow ω to vary
omega = 1.5                          # Initial guess for ω
```

```
* Codon frequency:
CodonFreq = 2                       * 0: Equal, 1: F1X4, 2: F3X4,
3: Codon table

* Other settings:
fix_kappa  =  0                                    *  0:  Estimate
transition/transversion ratio
kappa = 2.0                         * Initial kappa value

* Output control:
cleandata = 1                       * 1 = Remove sites with gaps,
0 = Keep all sites
```

```
#Run codeml in terminal
path/to/codeml codeml.ctl
```

Appendix 6 List of Asian *Begonia* with their chloroplast genome accession number, habitat, and shade adaptation

No.	Name	Accesion Code	Country	Growth Type	Location	Habitat	Shade Adaptation
1	*B. arachnoidea*	NC_063512.1	China	Rhizomatous	Guangxi Zhuangzu, Daxin Xian, Encheng	Encheng Nature Reserve, elev. ca. 200 m, on semishaded rocky slope.	Semi-shade
2	*B. asteropyrifolia*	NC_065014.1	China	Rhizomatous	Guangxi, Donglan	Foot of limestone hill, alt. 340 m (Xiong *et al.* 2023).	Semi-shade
3	*B. bracteata* voucher Peng23521	NC_070310.1	Indonesia	Rhizomatous	Bengkulu, Mount Bungkuk	Near the foot of Gunong Bunko (Mount Bungkuk) in Bencoolen (Bengkulu) (Hughes & Girmansyah 2011).	Semi-shade
4	*B. cathayana* voucher GX202201	NC_073119.1	China	Cane	Fangchenggang Guangxi, China	In forest, slopes and valley, gully bottoms in deep shady with alt: 1200-1500 m	Deep shade
5	*B. cavaleriei* voucher GX202202	NC_073117.1	China	Rhizomatous	Baise Guangxi, China	Ravines, foot of the mountains, valleys in deep shady; alt. 700-1,000 m (Xiong *et al.* 2023).	Semi-shade
6	B. coptidifolia	NC_056110.1	China	Rhizomatous	Guangdong, Yangchun city	Ehuangzhang Natural Reserve, Honghuatan, alt. 600 m, in a ravine, on rocks at streamside	Semi-shade
7	*B. dipetala* voucher Peng22521	NC_070315.1	India	Rhizomatous	Bombay	Western Ghats mountains: moist crevices rock (Indrakumar *et al.* 2013).	Semi-shade

No.	Name	Accesion Code	Country	Growth Type	Location	Habitat	Shade Adaptation
8	*B. emeiensis*	NC_061410.1	China	Rhizomatous	Sichuan, Mount Emei	Steam banks alt. 900-1000 m; prefers indirect sunlight	Semi-shade
9	*B. ferox*	NC_067030.1	China	Rhizomatous	Guangxi Zhuangzu	Chunxiu Headwater Forest Nature Reserve, elev. 130 m, on forest floor, limestone rock surface	Deep shade
10	*B. filiformis*	PQ675783.1	China & Vietnam	Rhizomatous	Guangxi, Longzhou (China), Minh Tan, Ha Giang (Vietnam)	On limestone rocks or in moist rocky caves of forest understories; Growing in cracks of mossy rocks on semi-shaded limestone cliffs in evergreen broadleaved forest	Semi-shade
11	*B. grandis voucher GX202203*	NC_073120.1	China	Rhizomatous	Nanning Guangxi	Wet stone, densely forested rocks in valleys, alt. 100-1,100 m (Xiong *et al.* 2023).	Deep shade
12	*B. guangxiensis plastid*	NC_046385.1	China	Rhizomatous	Guangxi, Du'an, Yishan	Limestone hills, on moist rocky slopes, forest understories, in caverns, alt. 200-300 m	Deep shade
13	*B. gulongshanensis*	NC_063513.1	China	Rhizomatous	Guangxi, Zhuang, Jingxi, Gulongshan	In the deep valley, growing on the moist, surface of steep cliffs, alt. 286 m	Semi-shade

No.	Name	Accesion Code	Country	Growth Type	Location	Habitat	Shade Adaptation
14	*B. handelii*	NC_065245.1	China	Rhizomatous	Guangdong, Guangxi, Hainan, Yunnan	Forests, variable vegetation, shaded moist environments alt. 100-1500 m.	Deep shade
15	*B. henryi voucher RBGE20141517*	NC_070321.1	China	Rhizomatous	Guangxi, Guizhou, Yichang, Yunnan	On rocks or in fissures, shaded moist environments; 800-2600 m	Deep shade
16	*B. jingxiensis*	PQ572754.1	China	Rhizomatous	Hurun, Jingxi	In forest limestone hill, alt. 450 m	Semi-shade
17	*B. jinyunensis*	NC_068748.1	China	Rhizomatous	Chongqing: Mount Jinyun	Jinyun Mountain National Nature Reserve, 29° 49' N, 106° 20' E, alt. 789 m. The species live on limestone faces of shaded moist environments such as valleys and broadleaved forests (Ding *et al.* 2014).	Deep shade
18	*B. leprosa voucher GX202204*	NC_073118.1	China	Rhizomatous	Guangxi, Guangdong	Semi-open forests or scrubby vegetation, on limestone rocks; 100-800 m (Xiong *et al.* 2023).	Semi-shade

No.	Name	Accesion Code	Country	Growth Type	Location	Habitat	Shade Adaptation
20	*B. luochengensis*	PQ675781.1	China	Rhizomatous	Guangxi Zhuangzu, Luocheng	Huaiqun Zhen, 180°34’ 12’’ E, 24°50’ 35’’N; elev. 250 m; semi shaded, dry or slightly moist limestone hills.	Semi-shade
21	*B. masoniana*	OP618127.1	China & Vietnam	Rhizomatous	Guangxi, Daxin, Pingxiang	Rocky limestone slopes, in caves under dense forests or shrubberies; 100-300 m.	Semi-shade
22	*B. obsolescens* voucher GX202205	NC_073122.1	China	Rhizomatous	Liuzhou, Guangxi	Rocky crevices; ca. 1,200 m	Semi-shade
23	*B. pedatifida*	OR288087.1	China & Vietnam	Rhizomatous	Guizhou, Hubei, Hunan, Sichuan	Broad-leaved forests, shaded moist environments; 300-1700 m.	Deep shade
24	*B. pulchrifolia*	NC_045096.1	China	Rhizomatous	Sichuan, Meifeng	Leshan, Sichuan, 29°20′57″N, 103°31′47″E, alt. 1160 m	Semi-shade
25	*B. retinervia*	NC_088496.1	China	Rhizomatous	Guangxi, Du’an	Limestone slopes, moist caves; alt. 200-600 m.	Semi-shade
26	*B. smithiana* GX202206	NC_073123.1	China	Rhizomatous	Hechi Guangxi	Gullies, valleys, foot of the mountains; ca. 700-1,320 m (Xiong *et al.* 2023).	Semi-shade

No.	Name	Accesion Code	Country	Growth Type	Location	Habitat	Shade Adaptation
27	*B. subcoriacea*	PQ619425.1	China	Rhizomatous	Guangxi, Zhuangzu	Daxin Xian, between Longhua and Niandi, elev. ca. 250 m, N-facing, broadleaf forest with scattered bamboo groves, on semishady limestone slope	Semi-shade
28	*B. umbraculifolia* voucher GX202207	NC_073121.1	China	Rhizomatous	Nannin g Guangxi	Limestone rocks, understory or valley; ca. 170–500 m (Xiong *et al.* 2023).	Deep shade
29	*B. versicolor*	NC_047450.1	China	Rhizomatous	Yunnan	Forests, shaded moist environments on slopes or along stream banks; alt. 1800-2100 m.	Deep shade
30	*B. wangii*	PQ675782.1	China	Rhizomatous	Yunnan, Malipo	Scrubby vegetation, on limestone rocks; alt. 800-1000 m.	Semi-shade

Appendix 7 RpoC1 homologous protein 3D models parameters score

Accession	GMQE	QMEANDisCo Global	Sequence Identity	Amino acid	Adaptation
Begonia arachnoidea	0.81	0.75	92.21%	560H	Semi-shaded
Begonia bracteata	0.81	0.75	92.06%	560R	Semi-shaded
Begonia cathayana	0.8	0.74	90.15%	564R	Deep Shade
Begonia cavaleriei	0.82	0.76	91.76%	560H	Semi-shaded
Begonia coptidifolia	0.82	0.76	91.47%	560R	Semi-shaded
Begonia dipetala	0.82	0.76	92.06%	560H	Semi-shaded
Begonia emeiensis	0.82	0.76	91.34%	560R	Semi-shaded
Begonia ferox	0.8	0.74	92.35%	560H	Deep Shade
Begonia filiformis	0.8	0.74	92.35%	560H	Semi-shaded
Begonia grandis	0.82	0.76	91.91%	560R	Deep Shade
Begonia guangxiensis	0.81	0.75	92.06%	560H	Deep Shade
Begonia gulongshanensis	0.82	0.76	92.06%	560H	Semi-shaded
Begonia handelii	0.82	0.76	91.47%	560R	Deep Shade
Begonia jingxiensis	0.8	0.74	92.35%	560H	Semi-shaded
Begonia jinyunensis	0.82	0.76	91.34%	560R	Deep Shade
Begonia leprosa	0.81	0.75	92.21%	560H	Semi-shaded
Begonia liuyanii	0.8	0.74	92.35%	560H	Semi-shaded
Begonia luochengensis	0.81	0.75	92.06%	560H	Semi-shaded
Begonia masoniana	0.81	0.75	92.21%	560H	Semi-shaded
Begonia obsolescens	0.81	0.75	92.21%	560H	Semi-shaded
Begonia pedatifida	0.82	0.76	91.34%	560R	Deep Shade
Begonia pulchrifolia	0.82	0.76	91.34%	560R	Deep Shade
Begonia retinervia	0.81	0.76	92.21%	560H	Semi-shaded
Begonia smithiana	0.82	0.77	91.48%	560R	Semi-shaded
Begonia subcoriaceae	0.8	0.74	92.35%	560H	Semi-shaded
Begonia umbraculifolia	0.81	0.75	92.50%	560H	Deep Shade
Begonia versicolor	0.81	0.75	90.59%	564R	Deep Shade
Begonia wangii	0.81	0.76	92.06%	560H	Semi-shaded

Appendix 8 *Begonia* samples taken from Bogor Botanical Garden samples

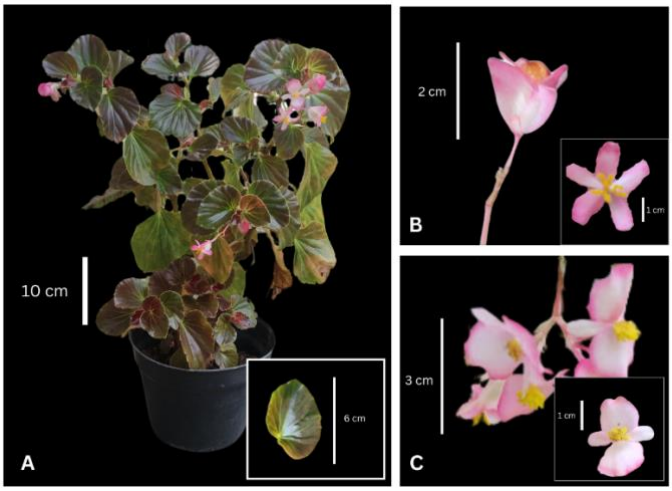

Figure 9 *Begonia cucullata*. a) plant, b) female flower, c) male flower

Figure 10 *Begonia mexicana*

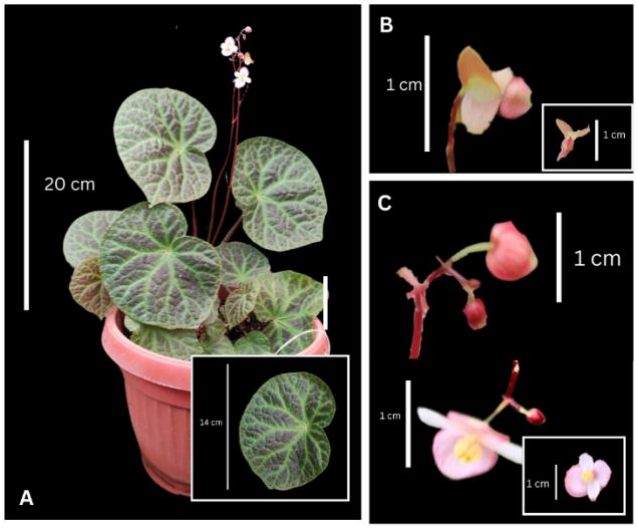

Figure 11 *Begonia hijauvenia*. a) plant, b) female flower, c) male flower

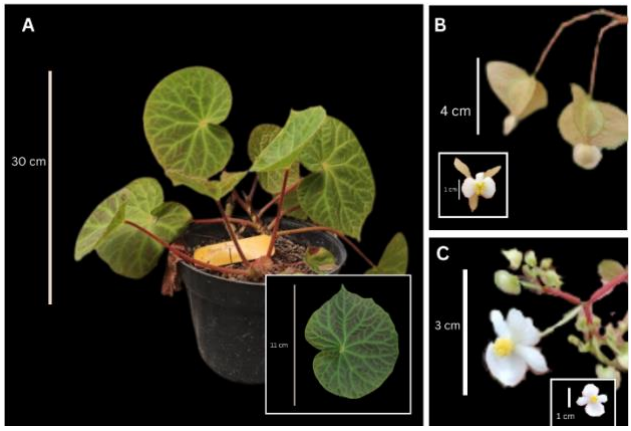

Figure 12 *Begonia* 'Ivory White' (*B. hijauvenia* ♀ × *B. kudoensis* ♂). a) plant, b) female flower, c) male flower

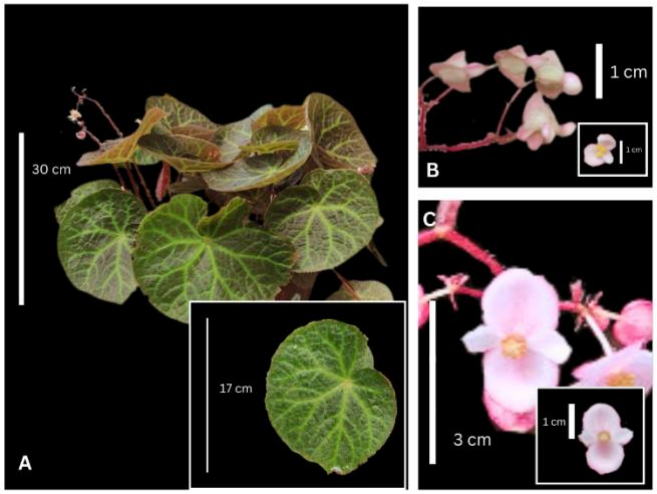

Figure 13 *Begonia* 'Eternal Flame' (*B. hijauvenia* ♀ × *Begonia sp*. ♂). a) plant, b) female flower, c) male flower

Figure 14 *Begonia isoptera*

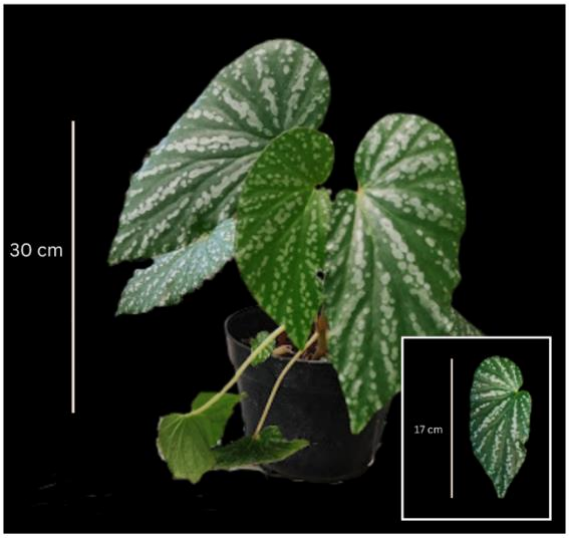

Figure 15 *Begonia* 'Silver Pixie' (Gamma-irradiated *B. isoptera*)

Appendix 9 Plant Variety Certification *B*. 'Ivory White'

REPUBLIK INDONESIA
KEMENTERIAN PERTANIAN
PUSAT PERLINDUNGAN VARIETAS TANAMAN DAN PERIZINAN PERTANIAN

Tanda Daftar Varietas Tanaman

VARIETAS HASIL PEMULIAAN
NOMOR : : 982/PVHP/2022

Berdasarkan Undang-undang Nomor 29 Tahun 2000 tentang Perlindungan Varietas Tanaman; Peraturan Pemerintah Nomor 13 Tahun 2004 tentang Penanaman, Pendaftaran dan Penggunaan Varietas Asal untuk Pembuatan Varietas Turunan Esensial; Peraturan Menteri Pertanian Nomor 29 Tahun 2021 tentang Penamaan dan Pendaftaran Varietas Tanaman, bersama ini kami memberikan Tanda Daftar Varietas Hasil Pemuliaan Kepada :

Nama Pemohon	:	**ORGANISASI RISET HAYATI DAN LINGKUNGAN BRIN**
Alamat	:	Jalan Raya Jakarta-Bogor Km 46, Cibinong Bogor-Jawa Barat
Tanggal Penerimaan Pendaftaran	:	13 September 2022
Nama Pemulia	:	Dra. Hartutiningsih-M.Siregar , Sri Wahyuni, MSi., Yupi Isnaini, M.Si. (Pusat Riset Konservasi Tumbuhan, Kebun Raya dan Kehutanan, OR Hayati dan Lingkungan-BRIN), Dr. Joko Ridho Witono, M.Si., Wisnu Handoyo Ardi, MSi. (Pusat Riset Biositematika dan Evolusi, OR Hayati dan Lingkungan-BRIN), Dr. Sasanti Widiarsih, MS. (Pusat Riset Teknologi Aplikasi Isotop dan Radiasi, OR Tenaga Nuklir-BRIN), Dr. Arya Widura Ritonga, M.Si (Fakultas Pertanian, Institut Pertanian Bogor), Sopian, AMd., Faturahman (Direktorat Pengelolaan Koleksi Ilmiah, Kedeputian Bidang Infrastruktur Riset dan Inovasi BRIN)
Kewarganegaraan Pemulia	:	**Indonesia**
Jenis Tanaman	:	Begonia (Begonia species Linnaeus)
Nama Varietas	:	Begonia Ivory White

Dengan Demikian varietas tersebut telah terdaftar di Pusat Perlindungan Varietas Tanaman dan Perizinan Pertanian, dan menjadi milik masyarakat di wilayah yang bersangkutan sesuai dengan Peraturan Perundang-undangan yang berlaku dan dicatat dalam Daftar Umum PVT, serta diumumkan dalam Berita Resmi PVT.

Jakarta, 26 September 2022
Kepala Pusat

Gunawan, S.P. M.Si
NIP 197407081998031002

Appendix 10 Plant Variety Certification *B*. ‘Eternal Flame’

REPUBLIK INDONESIA
KEMENTERIAN PERTANIAN
PUSAT PERLINDUNGAN VARIETAS TANAMAN DAN PERIZINAN PERTANIAN

Tanda Daftar Varietas Tanaman

VARIETAS HASIL PEMULIAAN
NOMOR : 922/PVHP/2021

Berdasarkan Undang-undang Nomor 29 Tahun 2000 tentang Perlindungan Varietas Tanaman; Peraturan Pemerintah Nomor 13 Tahun 2004 tentang Penanaman, Pendaftaran dan Penggunaan Varietas Asal untuk Pembuatan Varietas Turunan Esensial; Peraturan Menteri Pertanian Nomor 01 Tahun 2006 tentang Syarat Penanaman dan Tata Cara Pendaftaran Varietas Tanaman, bersama ini kami memberikan Tanda Daftar Varietas Hasil Pemuliaan Kepada :

Nama Pemohon	:	**Pusat Riset Konservasi Tumbuhan dan Kebun Raya - BRIN**
Alamat	:	Jl. Ir. H. Juanda No. 13 Bogor, Bogor - Jawa Barat
Tanggal Penerimaan Pendaftaran	:	22 Oktober 2021
Nama Pemulia	:	**Wisnu Handoyo Ardi, MSi; Sri Wahyuni, MSi.; Dra. Hartutiningsih-M.Siregar (Pemulia Pusat Riset Konservasi Tumbuhan, Organisasi Riset Ilmu Pengetahuan Hayati-Badan Riset dan Inovasi Nasional)**
Kewarganegaraan Pemulia	:	**Indonesia**
Jenis Tanaman	:	**Begonia**
Nama Varietas	:	**Begonia Ethernal Flame**

Dengan Demikian varietas tersebut telah terdaftar di Pusat Perlindungan Varietas Tanaman dan Perizinan Pertanian, dan menjadi milik masyarakat di wilayah yang bersangkutan sesuai dengan Peraturan Perundang-undangan yang berlaku dan dicatat dalam Daftar Umum PVT, serta diumumkan dalam Berita Resmi PVT.

Jakarta, 03 November 2021
Kepala Pusat,

Prof(Riset). Dr. Ir. Erizal Jamal, M.Si.
NIP 196303011989031002

Appendix 11 Plant Variety Certification *B*. 'Silver Pixie'

REPUBLIK INDONESIA
KEMENTERIAN PERTANIAN
PUSAT PERLINDUNGAN VARIETAS TANAMAN DAN PERIZINAN PERTANIAN

Tanda Daftar Varietas Tanaman

VARIETAS HASIL PEMULIAAN
NOMOR : 114/A.9/08/2024

Berdasarkan Undang-undang Nomor 29 Tahun 2000 tentang Perlindungan Varietas Tanaman; Peraturan Pemerintah Nomor 13 Tahun 2004 tentang Penamaan, Pendaftaran dan Penggunaan Varietas Asal untuk Pembuatan Varietas Turunan Esensial; Peraturan Menteri Pertanian Nomor 29 Tahun 2021 tentang Penamaan dan Pendaftaran Varietas Tanaman, bersama ini kami memberikan Tanda Daftar Varietas Hasil Pemuliaan kepada :

Nama Pemohon	: **BADAN RISET DAN INOVASI NASIONAL**
Alamat	: Gedung B.J. Habibie, Jl. M.H. Thamrin No. 8, Jakarta Pusat
Tanggal Penerimaan Pendaftaran	: 26 Agustus 2024
Nama Pemulia	: **Dra. Hartutiningsih-M. Siregar, Sri Wahyuni, SP., M.Si., Dra. Inggit Puji Astuti, M.Si., R. Vitri Garvita, M.Si., Mutiara Kusumaningtyas Pitaloka, S.P., M.Sc., Ph.D. (PR Botani Terapan, OR Hayati dan Lingkungan-BRIN), Ir. Mustaid Siregar, M.Si., Angga Yudaputra, S.Si., M.Sc. (PR Ekologi dan Etnobiologi, OR Hayati dan Lingkungan-BRIN), Sasanti Widiarsih, S.P., M.Sc., Ph.D., Ir. Ita Dwimahyani (PR Tanaman Pangan, OR Pertanian dan Pangan-BRIN), Dr. Ridho Kurniati, SP., M.Si. (PR Hortikultura, OR Pertanian dan Pangan-BRIN)**
Kewarganegaraan Pemulia	: **Indonesia**
Jenis Tanaman	: **Begonia (*Begonia isoptera Dryand. ex Sm*)**
Nama Varietas	: **Silver Pixie**

Dengan demikian varietas tersebut telah terdaftar di Pusat Perlindungan Varietas Tanaman dan Perizinan Pertanian sesuai dengan Peraturan Perundang-undangan yang berlaku dan dicatat dalam Daftar Umum PVT, serta diumumkan dalam Berita Resmi PVT.

Jakarta, 29 Agustus 2024
Kepala Pusat,

Dr. Ir. Leli Nuryati, M.Sc
NIP 196802251993032002

Appendix 12 Concentration, purity and electrophoresis result of genomic DNA isolates

Table 13 Concentration and purity result of genomic DNA isolates

Begonia samples	Genomic Isolation kit	Nucleic Acid Concentration (ng/µL)	$A_{260/280}$	$A_{260/230}$
B. cucullata	Geneaid	4.5	2.93	0.29
	Tiangen	125.2	2.83	2.19
B. mexicana	Geneaid	7.5	2.52	0.73
	Tiangen	176.3	2.16	2.60
B. hijauvenia	Geneaid	7.3	3.23	0.26
	Tiangen	67.7	2.01	1.50
B. 'Ivory White'	Geneaid	3.8	2.90	0.54
	Tiangen	56.9	2.20	3.13
B. 'Eternal Flame'	Geneaid	5.5	1.77	0.59
	Tiangen	59.4	2.19	3.39
B. isoptera	Geneaid	4.6	2.54	0.69
	Tiangen	70.4	2.16	2.53
B. 'Silver Pixie'	Geneaid	6.3	1.02	0.16
	Tiangen	99.6	2.54	10.07

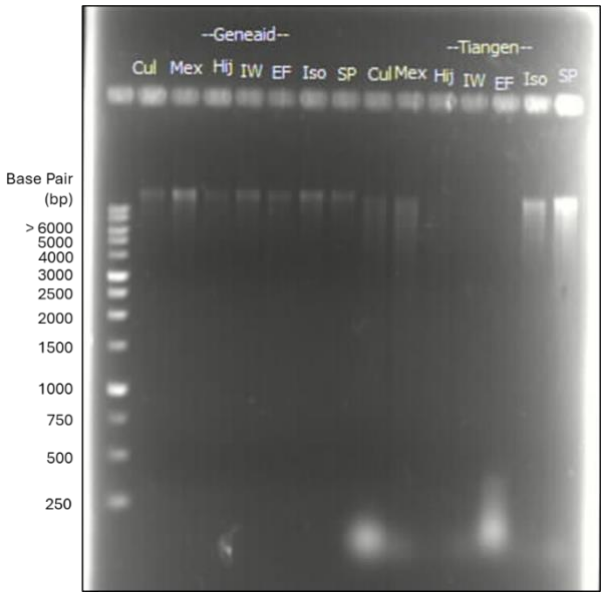

Figure 16 Gel electrophoresis result of Genomic DNA isolation Abbreviations: *B. cucullata* (Cul), *B. mexicana* (Mex), B. *hijauvenia* (Hij), *B.* 'Ivory White' (IW), *B.* 'Eternal Flame' (EF), *B. isoptera* (Iso), *B.* 'Silver Pixie' (SP)

Appendix 13 Gel electrophoresis result amplification of ITSY4-Y5, *rpoC1* Read 1 – 4, *psbE* and *pet*

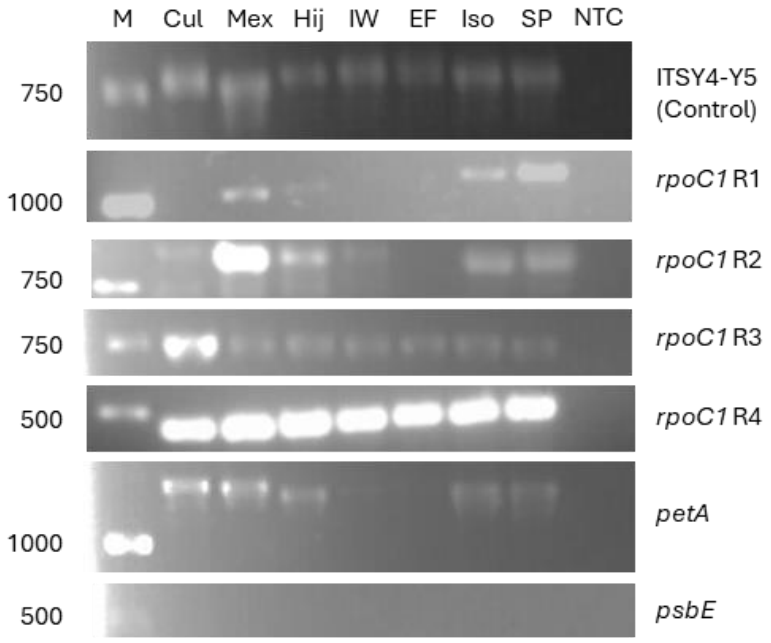

Figure 17 Gel electrophoresis result of ITSY4-Y5, *rpoC1* Read 1 – 4, *psbE* and *petA* amplification. Abbreviations: *B. cucullata* (Cul), *B. mexicana* (Mex), B. *hijauvenia* (Hij), *B.* 'Ivory White' (IW), *B.* 'Eternal Flame' (EF), *B. isoptera* (Iso), *B.* 'Silver Pixie' (SP)

Appendix 14 Sequence of *petA* gene of Indonesian *Begonia* (FASTA)

>petA|Begonia cucullata
GNNTGGANCATGCNAACTAGAAATACCTTTTCTTGGNNAANNNNAGAGANTACTCAATCCATTTTC
CTATCGCTCATGATATATATAATAACTGGGGCACCTGTTTCAAATGCATATCCCATTTTTGCACAG
CAGGGTTATGAAAATCCACGAGAAGCGACCGGTCGTATTGTCTGTGCCAACTGCCATTTAGCTAAT
AAGCCCGTGGATATTGAGGTTCCACAAGCTGTACTTCCGGATAGTGTATTTGAGGCAGTTGTTCGA
ATTCCTTATGATAAGCAATTGAAACAAGTTCTTGGGAATGGTAAAAAGGGAGCTTTGAATGTGGGA
GCTGTTCTTATTTTACCTGAGGGGTTTGAATTAGCCCCCCCCGATCGTATTTCGCCCGAGATTAAA
GAAAAGATAGGTAATCTGTCTTTTCAGAACTATCGTCCCACTAAAAAAAATATTCTTGTGATAGGT
CCTGTTCCTGGTCAGAAATATAGTGAAATCACCTTTCCTATTCTTTCTCCGGACCCTGCTACTAAG
AAAGATGTTTACTTCTTAAAATATCCCATATATGTAGGCGGGAACAGAGGAAGGGGCCAGATTTAT
CCCGACGGGAGTAAGAGTAATAATAATGTTTATAATGCTACAGTAGCGGGCATAGTAAGCAAAATA
ATACGAAAAGAAAAGGGAGGATATGAAATAACCCTAGTGGATACATTGGNTGGGCGTCAAGTGGTG
GAAATTATCCCTCCAGGACCAGAACTTCTTATTTCAGAGGGCGAATCTATCAAACTTGATCAACCA
TTAACGAGCAATCCTAATGTGGGTGGATTTGGTCAGAGGGAGGCAGAAATAGTACTTCAAGATCCA
TTACGTGTCCAAGGCCTTTTGTTCTTTTTTGCATCTATNNTTTTAGCACAAATCTTTTTGGTTCTT
AAAAAGAAACAATTTGAAAAGGTTCAATTGTCCGAAATGAATTTCTAGATTGGTGGATTTCTCAAC
ATCAAACTCGTAAAAAGAACCAAATTCTTGTTGGAAATGCATTCTATAATTATGTATGACCCATAA
TTNTGTNTGCNCCAAAAAAAGGNNTNGAAAAGCTTTTTACTGGTTTTNNNTTTTTTNNNNNNNNTT
T

>petA|Begonia mexicana
NNTGGACCATGCNANCTAGAAATATCTTTTCTTGGATAAAGGCAGAGATTACTCGATCCATTTTCC
TATCGCTCATGATATATATAATAACTGGGGCACCTGTTTCAAATGCATATCCCATTTTTGCACAGC
AGGGTTATGAAAATCCACGAGAAGCGACCGGCCGTATTGTCTGTGCCAACTGCCATTTAGCTAATA
AGTCCGTGGATATTGAGGTTCCACAAGCTGTACTTCCGGATAGTGTATTTGAAGCAGTTGTTCGAA
TTCCTTATGATAAGCAATTGAAACAAGTTCTTGGGAATGGTAAAAAAGGAGCTTTGAATGTGGGGG
CTGTTCTTATTTTACCTGAGGGGTTTGAATTAGCCCCCCCCGATCGTATTTCGCCCGAGATTAAAG
AAAAGATAGGCAATTTGTCTTTTCAGAACTATCGTCCCACTAAAAAAAATATTCTTGTGATAGGTC
CTGTTCCTGGTCAGAAATATAGTGAAATCACCTTTCCTATTCTTTCTCCGGACCCTGCTACTAAGA
AAGATGTTTATTTCTTAAAATATCCCATATATGTAGGCGGGAACAGAGGAAGGGGCCAGATTTATC
CCGACGGTAGTAAGAGTAATAATAATGTTTATAATGCTACAGTAGCGGGTATAGTAAGCAAAATAA
TACGAAAAGAAAAGGGAGGATATGAAATAACCCTAGTGGATACATTGGATGGGCGTCAAGTGGTGG
ATATTATCCCTCCAGGACCAGAACTTCTTATTTCAGAGGGCGAATCTATCAAACTTGATCAACCAT
TAACGAGCAATCCTAATGTGGGTGGATTTGGTCAGAGGGAGGCGGAAATAGTACTTCAAGATCCAT
TACGTGTCCAAGGCCTTTTGTTNTTTTTTGCATCTATTATTTTAGCACAAATNTTTTTGGTTNTTA
AAAAGAAACAATTTGAAAAGGTTCAATTGTCCGAAATGAATTTCTAGATCGGTGGATTTCTCAACA
TCAAGTTCGTAAAAAGAACCAAATTNTTGTTGGAAATTAATTNNATAATTANGTATGACCCATAAT
TAGTATGACNAAAAAAAGTANGAAAAGNCTNT

>petA|Begonia hijauvenia
NTGGNCCATGCAAACTAGAAATACCTTTTCTTGGATAAAGGCAGAGATTACTNGATCCATTTTCCT
ATCGCTCATGATATATATAATAACCGGGGCACCTGTTTCAAATGCATATCCCATTTTTGCACNGCA
GGGTTATGAAAATCCACGAGAAGCGACCGGCCGTATTGTCTGTGCCAACTGCCATTTAGCTAATAA
GCCCGTGGATATTGAGGTTCCACAAGCTGTACTTCCGGATAGTGTATTTGAAGCAGTTGTTCGAAT
TCCTTATGATAAGCAATTGAAACAAGTTCTTGGGAATGGTAAAAAAGGAGCTTTGAATGTGGGGGC
TGTTCTTATTTTACCTGAGGGGTTTGAATTAGCCCCCCCAGATCGTATTTCGCCCGAGATTAAAGA
AAAGATAGGCAATCTGTCTTTTCAGAACTATCGTCCCACTAAAAAAAATATTCTTGTAATAGGTCC
TGTTCCTGGTCAGAAATATAGTGAAATCACCTTTCCTATTCTTTCTCCGGACCCTGCTACTAAGAA
AGATGTTTACTTCTTAAAATATCCCATATATGTAGGTGGGAACAGAGGAAGGGGCCAGATTTATCC

CGACGGGAGTAAGAGTAATAATAATGTTTATAATGCTACNGTAGCGGGTATAGTAAGCAAAATAAT
ACGAAAAGAAAAGGGAGGATATGAAATAACCCTAGTGGATGCATTGGATGGGCGTCAAGTGGTGGA
TATTATCCCTCCAGGACCAGAACTTCTTATTTCAGAGGGCGAATCTATCAAACTTGATCNACCATT
AACGANCAATCNTAATGTGGGTGGATTTGGTCAGAGGGAGGCGGAAATAGTACTTCAAGATCCATT
ACGTGTCCAAGGCCTTTTGTTCTTTTTTGCATCTATTATTTTAGCACAAATNTTTTTGGTTCTTAA
AAAGNAACAATTTGAAAAGGTTCAATTGTCCAAAATGAATTTCTAGATCGGTGGATTTATCAACAT
CAAGTTNGTAAAAAGAACCAAATTCTTGTTGGAAATTTTATGTATGACCAAAAAAAAGTANGAN

>petA|Begonia Ivory White
NNTAGCAACCTANCTNANTTTNTTGTNNNAANTTTCAGGATCAATGATTGGACCATGCAAACTAGA
AATACCTTTTCTTGGATAAAGGCAGAGATTACTCGATCCATTTTCCTATCGCTCATGATATATATA
ATAACTGGGGCACCTGTTTCAAATGCATATCCCATTTTTGCACAGCAGGGTTATGAAAATCCACGA
GAAGCGACCGGCCGTATTGTCTGTGCCAACTGCCATTTAGCTAATAAGCCCGTGGATATTGAGGTT
CCACAAGCTGTACTTCCGGATAGTGTATTTGAAGCAGTTGTTCGAATTCCTTATGATAAGCAATTG
AAACAAGTTCTTGGGAATGGTAAAAAAGGAGCTTTGAATGTGGGGGCTGTTCTTATTTTACCTGAG
GGGTTTGAATTAGCCCCCCCAGATCGTATTTCGCCCGAGATTAAAGAAAAGATAGGCAATCTGTCT
TTTCAGAACTATCGTCCCACTAAAAAAAATATTCTTGTAATAGGTCCTGTTCCTGGTCAGAAATAT
AGTGAAATCACCTTTCCTATTCTTTCTCCGGACCCTGCTACTAAGAAAGATGTTTACTTCTTAAAA
TATCCCATATATGTAGGTGGGAACAGAGGAAGGGGCCAGATTTATCCCGACGGGAGTAAGAGTAAT
AATAATGTTTATAATGCTACAGTAGCGGGTATAGTAAGCAAAATAATACGAAAAGAAAAGGGAGGA
TATGAAATAACCCTAGTGGATGCATTGGATGGGCGTCAAGTGGTGGATATTATCCCTCCAGGACCA
GAACTTCTTATTTCAGAGGGCGAATCTATCAAACTTGATCAACCATTAACGAGCAATCCTAATGTG
GGTGGATTTGGTCAGAGGGAGGCGGAAATAGTACTTCAAGATCCATTACGTGTCCAAGGCCTTTTG
TTCTTTTTTGCATCTATTATTTTAGCACAAATCTTTTTGGTTCTTAAAAAGAAACAATTTGAAAAG
GTTCAATTGTCCGAAATGAATTTCTAGATCGGTGGATTTATCAACATCAAGTTCGTAAAAAGAACC
AAATTCTTGTTGGAAATTCATTCTATAATTATGTATGACCAAAAAAAAGTATGAAAAGCTTTTTAC
TTGTTTTTTTNNNNANNCTTTTTANNNNCNNGNNNNNTNCGG

>petA|Begonia isoptera
CAACCNNNNNAANNTNNTTNTNTGTAGNAATTTTCGGGATCAATGATTGGACCATGCAAACTAGAA
ATACTTTTTCTTGGATAAAGGCAGAGATTACTCGATCCATTTTCCTATCGCTCATGATATATATAA
TAACTGGGGCACCTGTTTCAAATGCATATCCCATTTTTGCACAGCAGGGTTATGAAAATCCACGAG
AAGCGACCGGCCGTATTGTCTGTGCCAACTGCCATTTAGCTAATAAGCCCGTGGATATTGAGGTTC
CACAAGCTGTACTTCCGGATAGTGTATATTGAAGCAGTTGTTCGAATTCCTTATGATANGCAATTG
AAACAAGTTCTTGGGAATGGTAAAAAAGGAGCTTTGAATGTGGGGGCTGTTCTTATTTTACCTGNG
GGGTTTGAATTAGNCCCCCCAGATNGTATTTCNNNGCCCGAGATTAAAGAAAAGATAGGCAATCTG
TCTTTTCAGAACTATNGTCCCACTAAAAAAANTATTCTTGTAATAGGTCCTGTTCCTGGTCAGAAA
TATAGTGAAATCACCTTTCCTATTCTTTCTCCGGACCCTGCTACTAAGAAAGATGTTTACTTCTTA
AAATATCCCATATATGTAGGTGGGAACAGAGGAAGGGGCCAGATTTATCCCGACGGGAGTAAGAGT
AATAATAATGTTTATAATGCTACAGTAGCGGGTATAGTAAGCAAAATAATACGAAAAGAAAAGGGA
GGATATGAAATAACCCTAGTGGATGCATTGGATGGGCGTCAAGTGGTGGATATTATCCCTCCAGGA
CCAGAACTTNTTATTTCAGAGGGCGAATNTATCAAACTNGATCAACCNTTAACGAGCAATCCTAAT
GTGGGTGGATTTGGTCAGAGGGAGGCGGAAATAGTACTTCAAGATCCATTACGTGTCCAAGGCCTT
TTGTTCTTTTTTGCATNTATTATTTTAGCACAAATCTTTTTGGTTCTTAAAAAGAAACAATTTGAA
AAGGTTCAATTGTCCGAAATGAATTTNTAGATCGGTGGATTTATCAACNTCAAGTTCGTAAAAAGA
ACCAAATTCTTGTTGGAAATTCATTCTATAATTATGTATGACCAAAAAAAAGTANNNNAAANNNNN
NNNA

>petA|Begonia Silver Pixie
NNATTGGACCATGCAAACTAGAAATACTTTTTCTTGGATAAAGGCAGAGATTACTCGATCCATTTT
CCTATCGCTCATGATATATATAATAACTGGGGCACCTGTTTCAAATGCATATCCCATTTTTGCACA

GCAGGGTTATGAAAATCCACGAGAAGCGACCGGCCGTATTGTCTGTGCCAACTGCCATTTAGCTAA
TAAGCCCGTGGATATTGAGGTTCCACAAGCTGTACTTCCGGATAGTGTATTTGAAGCAGTTGTTCG
AATTCCTTATGATAAGCAATTGAAACAAGTTCTTGGGAATGGTAAAAAAGGAGCTTTGAATGTGGG
GGCTGTTCTTATTTTACCTGAGGGGTTTGAATTAGCCCCCCCAGATCGTATTTCGCCCGAGATTAA
AGAAAAGATAGGCAATCTGTCTTTTCAGAACTATCGTCCCACTAAAAAAAATATTCTTGTAATAGG
TCCTGTTCCTGGTCAGAAATATAGTGAAATCACCTTTCCTATTCTTTCTCCGGACCCTGCTACTAA
GAAAGATGTTTACTTCTTAAAATATCCCATATATGTAGGTGGGAACAGAGGAAGGGGCCAGATTTA
TCCCGACGGGAGTAAGAGTAATAATAATGTTTATAATGCTACAGTAGCGGGTATAGTAAGCAAAAT
AATACGAAAAGAAAAGGGAGGATATGAAATAACCCTAGTGGATGCATTGGATGGGCGTCAAGTGGT
GGATATTATCCCTCCAGGACCAGAACTTCTTATTTCAGAGGGCGAATCTATCAAACTCGATCAACC
ATTAACGAGCAATCCTAATGTGGGTGGATTTGGTCAGAGGGAGGCGGAAATAGTACTTCAAGATCC
ATTACGTGTCCAAGGCCTTTTGTTCTTTTTTGCATCTATTATTTTAGCACAAATCTTTTTGGTTCT
TAAAAAGAAACAATTTGAAAAGGTTCAATTGTCCGAAATGAATTTCTAGATCGGTGGATTTATCAA
CATCAAGTTCGTAAAAAGAACCAAATTCTTGTTGGAAATTCATTCTATAATTATGTATGACCAAAA
AAAAGTATGAAAAGCTTTTTACTTNN

Appendix 15 Sequence of partial *rpoC1* gene of Indonesian *Begonia* (FASTA)

>rpoC1_R2_R3|B.cucullata
TTTTTNNNNNNNNNNNNNNNNANNNNNANNNNNANNANNACNNNNGGGGGNTTTNNTA
ATAAAAAAAAAGAATTTTTAGTTTCTACCTCGNGAAANAAAAANAGAATTNTNTTTAAAG
AAGCAAATATGTCATGGTTGNNGGNGTCTATCCATCGCATATAGGCTTTAAGTACNNCAN
GGGATAACCGTCGNGGTGAAGTAGGGNCCTAAAAGATCGAATGGAACAATACAAAGACAA
GTCAATCCCTTATGAATTTCAAGGTATTCCTTTTAATTAACTAATTAAAGGTATTTTTCC
TTCGAAGGGAAGTAGACTACTCAAAAATTGAACATTTTTTTTTTNTTTTTTTTTTTTTTT
TTNNNNNNNNNAAAAAANNCNAAAAANNNNAAAAAAAAAAAAAANAATCAATNAAATGTTG
TTTTGCTTGGTCTTAATTNAAAANTTNGAGTAAAGANTANTTAGAGGTTTTANANAATTT
GAAAGCGAGAACTCTTTTNTTTNATTTNANANAACNNCTTGNGCCGGATNANAANAAATT
TTCATGTCCGATTTTNAAAGGGGGANANCCNANANTANCCNNTCCAAATTTTTNNTTTGN
NNGGCCCCNAGTNAAAAAACCTANTTTTTTTNNNANNNNNGGGGCNCNTTTNAANANNAA
ANANAANCTNGGAAANANAGCNNCCCNCNTTTTTTTTTACTACCCAAGGNTTCNANACANT
TTNNAAATCGANAAATTTCNACAGGNGNGGGNGNTATCCNANAANANTTNTNNCNNNNNG
GNNNTNNNCNNNNNNNNNNNANGAAGGANTNACNNNNGAAGNNNAGNAAAANNNNNGGGG
AAAAANNNNCCCCNNNNGANNNNNNNNCAGGNNNNNNNTGCTATCCGAGAACAATTAGCC
GATCTGGATTTGCGACTTATTATAGATTATTCATTGGTAGAATGGAAAGAATTAGGGGAA
GAAGTTCCCCCCGGTAATGAATGGGAAGATCGAAAAGTTGGAAGAAGAAAAGATTTTTTG
GTTAGACGTATGGAATTAGCTAAGTATTTTATTCGAACAAATATAGAACCTGAATGGATG
GTTTTATGTCTATTACCAGTTCTTCCCCCCGAGTTGAGACCAATCATTCAGATAGATGGG
GGTAAATTAATGAGCTCGGATATTAATGAACTCTATAGAAGAGTTATCTATCGGAACAAT
ACTCTTATTGATCTATTAACAACAAGTAGATCTACGCCAGGAGAATTAGTAATGTGTCAG
GAGAAATTGGTACAAGAAGCCGTGGATACACTTCTTGATAATGGAATCCGCGGACAACCA
ATGAGGGATGGTCATAATAAGGTTTACAAGTCATTTTCAGATGTAATTGAAGGCAAAGAG
GGAAGATTTCGTGAGACTCTTCTTGGCAAACGGGTCGATTATTCAGGACGTTCTGTTATT
GTTGTAGGTCCTTCACTTTCATTACATCGCTGTGGATTGCCTCGCGAAATCGCAATAGAG
CTTTTCCAGACATTTGTAATTCGTGGTCTAATTAGACAACATTTTGCTTCGAACATAGGA
GTTGCTAAGAGTAAAATTCGGGAAAAAGAACCGATTGTATCGGAAATACTTCNNNNNNNN
NNNGGGGG

>rpoC1_R2_R3|B. mexicana
TTTNNNNNNNNNNNGGNNNNNNNNNNNGAATNNNNANNNNNNTNATNCTNAAAAATANA
AAGGGAANNGNGTNNTATGGTCGATTTAGTAATAAAAAAAAGGAATTTTTAGTTTGTACC
TCGTGAAAAAAAAAAAATTGAATTATGTTCAAATAAGCAAATATGTCANGGTTGCAGGAG
TCTATCCATCGCATATAGGCTTTAAGTACATAANGGGATAACCGTCGAGGTGAAGTAGGG
ACCTAAAAGATCGAANGGAACAATACAAAGACAAGTCAATCCCTTATGAATTCCAAGGTA
TTCCTTTTAATTAACTAATTAAAGGTATTTTTCCTTCGAAGGGAAGTANACTACTCAAAA
ATTTAACNTTTTTTTTTTCTTTTATGTCGTATAAANAACTCAGAAAATNTAAAAAAAAAA
GAAAAAGAATCAATGAAATGTTGTTTTGCTTGGTCTTAATTGAAAACTTGAGTAAAGAGT
AGTTAGGGGTTTTATAGAATTTGAAAGTGGGAACTCTTTTATTTGATTTGATATAACTAC
TTGAGCCGGATGAGAAGAAACTTTCATGTCCGATTTTGAAAGGGGGAGATCCTATAGTAT
CCTATCCAAATTTTTCGTTTGCTAGGCCCCTAGCTAAAAAACCTACTTTCTTACGATTAC
GGGGCTCATTTGAATATGAAATACAATCTTGGAAATATAGCATCCCACTTTTTTTTACTA
CTCAAGGCTTCGATACATTTCGAAATCGAGAAATTTCGACAGGCGCGGGTGCTANCCGAG
AACAATTAGCCGATCTGGATTTGCGACTTATTATNNNNNNNNNNNNNNNNANNNNNNNNN
NGGNNNNNNCCCNNNNNNNGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNTATANGANTANTCATTGGTAGAATGGAAAGAATTAGGGGAAGAAGGTCCCCCCG
GTAATGAATGGGAAGATNGAAAAGTTGGAAGAAGAAAAGATTTTTTGGTTAGACGTATGG
AATTAGCTAAGTATTTTATTCGAACAAATATAGAACCTGAATGGATGGTTTTATGTCTAT

TACCAGTTCTTCCCCCCGAGTTGAGACCAATCATTCAGATAGATGGGGGTAAACTAATGA
GCTCGGATATTAATNNACTCTATAGAAGAGTTATCTATCGGAACANANACTCTTATTGAT
CTATTAACAACANGTAGATCTACGCCAGGAGAATTAGTAATGTGTCAGGAGAAATTGGTA
CAAGAAGCCGTGGATACACTTCTTGATAATGGAATCCGCGGACAACCAATGAGGGATGGT
CATAATAAGGTTTACAAGTCATTTTCAGATGTAATTGAAGGCAAAGAGGGAAGATTTCGT
GAGACTCTTCTTGGCAAACGGGTCNNNNTANTCANGACGTTCTGTTATTGTCGTAGGTCC
TTCACTTTCATTACATCGATGTGGATTGCCTCGCGAAATCGCAATAGAGCTTTTCCAGAC
ATTTGTAATTCGTGGTCTAATTAGACAACNTTTTGCTTCGAACATANGAGNTG

>rpoC1_R2_R3|B. hijauvenia
TTNNNNNNNNNNNNNNNNNNNNNNNNNNNCNNAATTTNNGNNNTTTNCNNNNCNNAAAA
ANNGGGGGGAATTGNTNTNTGGNCGNTTTNNTAATNAAAAAAAAAGGAATTTTTAGTTTN
TNCCTCGNGAAAAAAAAAAANGGAATTNNGTTNAAATAANNCNAATNNGTCNNGGTTGCN
GGNNNTNTNNCCTNNNNNNATNGGNNTTAANNNCNTNCNGGGNNNACCNNNNNNGGNGAA
NTNGGGNCCNNAAANATNGNAAGGGANCNANACAAANNNCAANTNCANNCCCTTNTGAAN
TNCNNNGNNNTTNTTTTNANTTAANTAATTNAANGNNNTTTTTCCTTNNAAGGGNAANNN
NANNNCNNCAAAAATTTNACNTTTTTTTTTTTTTTCTTTTTTTTATGTCGTATAAAGAAC
TCAGAAAATCTAAATAAAAAAGAAAAAGAATCAATGAAATGTTGTTTTGCTTGGTCTTAA
TTGAAAACTTGAGTAAAGAGTAGTTAGGGGTTTTATAGAATTTGAAAGCGGGAACTCTTT
TATTTGATTTGATGTAACTACTTGAGCCGGATGAGAAGAAACTTTCATGTCCGATTTTGA
AAGGGGGAGATCCTATAGTATCCTATCCAAATTTTTCGTTTGCTAGGCCCCTAGCTAAAA
AACCTACTTTCTTACGATTACGGGGCTCATTTGAATATGAAATACAATCTTGGAAATACA
GCATCCCACNTTTTTTTACTACCCAAGGCTTCGATACATTTCGAAATCGAGAAATTTCGA
CAGGAGCGGGTGCTATCCGAACAATNAGCNCCGATCTGGATTTGCGACTTATTATAGATT
ATTCATTGGTAGAATGGAAANAATTAGGNNNGGAAGNAGNNCCCCCCGGNTAANNNNTNN
AANTNNNAAGNTCNAAANNNTGGNAGANGAAAAGANNNNTNGGNTAGNCNTATGNNATNN
NGANTAAGTATTTNNTATTCNGANANNCNNAATNATNNAGAACNNCNNGAANTGGATGGT
TTTATGTCTATTACCAGTTCTTCCCCCCGAGTTGAGACCAATCATTCAGATAGATGGGGG
TAAACTAATGAGCTCGGATATTAATNNACTCTATAGAAGAGTTATCTATCGGAACAATAC
TCTTATTGATCTATTAACAACAAGTAGATCTACGCCGNNGAGAATTAGNAATGTGTCAGG
AGAAATTGGTACAAGAAGCCGTGGATACACTTCTTGATAATGGAATCCGCGGACAACCAA
TGAGGGATGGTCATAATAAGGTTTACAAGTCATTTTCAGATGTAATTGAAGGCAAAGAGG
GAAGATTTCGTGAGACTCTTCTTGGCAAACGGGTCGATTATTCAGGACGTTCTGTTATTG
TCGTAGGTCCTTCACTTTCATTACATCGATGTGGATTGCCTCGCGAAATCGCAATNNNGC
TTTTCCAGACATTTGTAATTCGTGGTCTAATTAGACAACATTTTGCTTCGAACATNNGGA
NTGNTNNGA

>rpoC1_R2_R3|B. Ivory White
NNNNNNNNNNNNNNNNNNNNAAANAANGGGNNNNNNGNNNNNNNNGNNNTTTNNAAAAAA
AAAAAGGAATTTTTNNTTNNNCCNNNNNAAAAAAAAAAAAANGGNANTNNNNTNNAANNAN
NNNANANNNNNCCNGGGNTNNNGGNNNNNTTNCCNNNCCANNNNGGNNTTNANNNNNNNN
NNGGNNNNACCNNNNCCCNNNNGGGNNNNNNGGGGNCNNAAAANNNNNANGGGNNNANNN
NNNNNNNNNNCNNNNCCCTTNNNNNTNCCNNGNNTTNTTTTTTNNNTNNNNNANNNAANG
GNTTTTTNCCNNNNNNGGGAANNNNNNNNNNNNNNAAANTTNNNNNTTTTTTTTTTTTTN
TTTTTTTTANGTNGNANAAANAANCNCNNAAAATNNAAANAAAAAANAAAAANAANCNNN
GAAATGTTNTTTTNNTTGGTNTTAANNNAAAANTTGAGNAAANANNANTTNGGGGTTTTA
NANAANTTNAAANGCGGGANNNNTTTTTTNTTTNNTTTGATGTNNCNNCNNGNGCCGGNN
NAGNAANAAANTTTNNNNTNCGNTTNNNNTNAAAGNGGGAGATNCNNTANTNTCCNATCC
NAANTTTTCGTTTGNTNGNCCCCTNANCNTAAAAAACCCNNCNNTNNNTNNNGANTNCGG
GGNNNANNTGNANNNNAAANNCNNTNNTNNGAANNANANCNNCNNCNNTNTNTTNTNNTA
CCCNAANGCNTNGANNNCNNNTNGAAATCNNGNNANTTCGNNNCNGGNGCGGGNGCTNTC

CNNGNNCNNNTNNCCNNNNGNNNNNNNNNNNNNNNNNNNNNNNNNCNNNNGNNTNCGAG
AACAATNAGCNNCGATCTNGATNTGNGACTTATTATAGATTATTCATTGGTAGAATGGAA
ANAATTAGGNNNGGAAGNAGNNCCCCCCGGNTAANNNNTNNAANTNNNAAGNTCNAAANN
NTGGNAGANGAAAAGANNNNTNGGNTAGNCNTATGNNATNNNGANTAAGTATTTNNTATT
CNGANANNCNNAATNATNNAGAACNNCNNGAANTGGATGGTTTTATGTCTATTACCAGTT
CTTCCCCCCGAGTTGAGACCAATCATTCAGATAGATGGGGGTAAACTAATGAGCTCGGAT
ATTAATNNACTCTATAGAAGAGTTATCTATCGGAACAATACTCTTATTGATCTATTAACA
ACAAGTAGATCTACGCCGNNGAGAATTAGNAATGTGTCAGGAGAAATTGGTACAAGAAGC
CGTGGATACACTTCTTGATAATGGAATCCGCGGACAACCAATGAGGGATGGTCATAATAA
GGTTTACAAGTCATTTTCAGATGTAATTGAAGGCAAAGAGGGAAGATTTCGTGAGACTCT
TCTTGGCAAACGGGTCGATTATTCAGGACGTTCTGTTATTGTCGTAGGTCCTTCACTTTC
ATTACATCGATGTGGATTGCCTCGCGAAATCGCAATNNNGCTTTTCCAGACATTTGTAAT
TCGTGGTCTAATTAGACAACATTTTGCTTCGAACATNNGGANTGNTNNGA

>rpoC1_R2_R3|B. Isoptera
NNNNNTTTNNNNNNNGGNNGGGGGNNNNNNNNNNNNNTTNNNNNNNTNNNNNNNNAAAAA
AAANNGGNNNNNNNNNTNTGGNCGATTTNNTNNNNNAAAANAAAAAAGGAANTTTTNNNT
TTNNCCNNNNGNAAAAAAAAAAANGAATTTNNNTNNAANNANCNAATANNTNATNGTNGC
NGGNGNNGNNTNNNCCCNNNNCCNNNNNGGNTTTNANTTNANCNTNNNCNNNNNNNNNCG
TCGNNGNNNAGNNNGGGNCCNNNAANNTNNNNANANNGGACNNAANACNAAGANAANNCC
AGNCCCNNNNCAATNTNANNNGGNCNNNTTTTTNNTTNCCTNANTAAAANTANNTTTTCC
NTNNNNANNGNANNGGNNNNNCNCANNAANNAAANNTTTTNNTTTTTTTTTTTTTTTTAT
GTNGTATAAAGAACTCAGAAAATCTAAATAAAAAANAAAAAGAATCAATGAAATGTTGTT
TTGCTTGGTCTTAATTGAAAACTTGAGTAAAGAGTAGTTAGGGGTTTTATAGAATTTGAA
AGCGGGAACNCTTTTATTTGATTTGATGTAACTACTTGAGCCGGATNAGAAGAAACTTTC
ATGTCCGATTTTGAAAGGGGGAGATCCTATAGTATCCTATCCAAATTTTTCGTTTGCTAG
GCCCCTAGCTAAAAAACCTACTTTCTTACGATTACGGGGCTCNTTTGAATATGAAATACA
ATCTTGGNAATACAGCATCCCACTTTTTTTTACTACCCAAGGCTTCGATACATTTCGAAA
TCGAGAAATTTCGACAGGAGCGGGTGCTATCCGAGAACAATTAGCCGATCTGGATTTGCG
ACTTATTATAGATTATTCATTGGTAGAATGGAAAGAATTAGGGGAAGAAGGTCCCCCCGG
TAATGAATGGGAAGATCGAAAAGTTGGAAGAAGAAAAGATTTTTTGGTTAGACGTATGGA
ATTAGCTAAGTATTTTATTCGAACAAATATAGAACCTGAATGGATGGTTTTATGTCTATT
ACCAGTTCTTCCCCCCGAGTTGAGACCAATCATTCAGATAGATGGGGGTAAACTAATGAG
CTCGGATATTAATGAACTCTATAGAAGAGTTATCTATCGGAACAATACTCTTATTGATCT
ATTAACAACAAGTAGATCTACGCCAGGAGAATTAGTAATGTGTCAGGAGAAATTGGTACA
AGAAGCCGTGGATACACTTCTTGATAATGGAATCCGCGGACAACCAATGAGGGATGGTCA
TAATAAGGTTTACAAGTCATTTTCAGATGTAATTGAAGGCAAAGAGGGAAGATTTCGTGA
GACTCTTCTTGGCAAACGGGTCGATTATTCAGGACGTTCTGTTATTGTCGTAGGTCCTTC
ACTTTCATTACATCGATGTGGATTGCCTCGCGAAATCGCAATAGAGCTTTTCCAGACATT
TGTAATTCGTGGTCTAATTAGACAACATTTTGCTTCGAACATAGGAGTTGCTAAGAGTAA
AATTCGGGAAAAAGAACCGATTGTANNNGAAATACTTCANNNNNNNTATGGGG

>rpoC1_R2_R3|B. Silver Pixie
NNNNTTTTNNNNNNNNNNNGNNNNNNGNNNNNNNAATTTTNNNTTTTTTNNANNNNNAAA
NNNNANGGGGAANNNNNTNTGGTCGATTTAGTAATAAAAAAAAAAAGGAATTTTTAGTTT
GTACCTNNTGAAAAAAAAAANGGAATTATGTTCAAATAANCAAAGCAAATATGTCANGGT
TGCNGGAGTNTNTCCCTCGCANATAGGCTTTAAGTACATCNNGGGATAACCGTCGNGGGG
AAGTAGGGACCTAAAANATNGAANGGAACAATACAAAGACAAGTCAATCCCTTNNGAATT
CCNAGGNNTTNTTTTTAATTACCTAATTAAAGGTATTTTTCCTTNGAAGGGAAGNANNCT
NCTCNAAAATTTAACNTTTTTTTTTTTTTTTTATGTCGTATAAAGAACTCAGAAAATCTA
AATAAAAAAGAAAAAGAATCAATGAAATGTTGTTTTGCTTGGTCTTAATTGAAAACTTGA

GTAAAGAGTAGTTAGGGGTTTTATAGAATTTGAAAGCGGGAACTCTTTTATTTGATTTGA
TGTAACTACTTGAGCCGGATGAGAAGAAACTTTCATGTCCGATTTTGAAAGGGGGAGATC
CTATAGTATCCTATCCAAATTTTTCGTTTGCTAGGCCCCTAGCTAAAAAACCTACTTTCT
TACGATTACGGGGCTCGTTTGAATATGAAATACAATCTTGGAAATACAGCATCCCACNTT
TTTTTACTACCCAAGGCTTCGATACATTTCGAAATCGAGAAATTTCGACAGGAGCGGGTG
CTATCCGAGAACAATTAGCCGATCTGGATTTGCGACTTATTATAGATTATTCATTGGTAG
AATGGAAAGAATTAGGGGAAGAAGGTCCCCCCGGTAATGAATGGGAAGATCGAAAAGTTG
GAAGAAGAAAAGATTTTTTGGTTAGACGTATGGAATTAGCTAAGTATTTTATTCGAACAA
ATATAGAACCTGAATGGATGGTTTTATGTCTATTACCAGTTCTTCCCCCCGAGTTGAGAC
CAATCATTCAGATAGATGGGGGTAAACTAATGAGCTCGGATATTAATGAACTCTATAGAA
GAGTTATCTATCGGAACAATACTCTTATTGATCTATTAACAACAAGTAGATCTACGCCAG
GAGAATTAGTAATGTGTCAGGAGAAATTGGTACAAGAAGCCGTGGATACACTTCTTGATA
ATGGAATCCGCGGACAACCAATGAGGGATGGTCATAATAAGGTTTACAAGTCATTTTCAG
ATGTAATTGAAGGCAAAGAGGGAAGATTTCGTGAGACTCTTCTTGGCAAACGGGTCGATT
ATTCAGGACGTTCTGTTATTGTCGTAGGTCCTTCACTTTCATTACATCGATGTGGATTGC
CTCGCGAAATCGCAATAGAGCTTTTCCAGACATTTGTAATTCGTGGTCTAATTAGACAAC
ATTTTGCTTCGAACATAGGAGTTGCTAAGAGTAAAATTCGGGAAAAAGAACCGATTGTAN
NNGAAATACTTCNN