AJAB

Asian J Agric & Biol. xxxx(x).
DOI: 10.35495/ajab.2024.015

Original Article

# Complete plastid genome of *Coelostegia griffithii* (Malvaceae): Structure, comparative and phylogenetic analysis

**Xue Jing Wong[1], Shiamala Devi Ramaiya[2], Wan Hee Cheng[1], Zheng-Feng Wang[3], Muhammad Syahmi Hishamuddin[4*], Shiou Yih Lee[1*]**

[1]Faculty of Health and Life Sciences, INTI International University, 71800 Nilai, Negeri Sembilan, Malaysia
[2]Department of Crop Science, Faculty of Agricultural and Forestry Science, Universiti Putra Malaysia Bintulu Sarawak Campus, 97008 Bintulu, Sarawak, Malaysia
[3]Guangdong Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, Guangdong, China
[4]Department of Forestry Science and Biodiversity, Faculty of Forestry and Environment, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

*Received:*
*February 8, 2024*
*Accepted:*
*September 17, 2024*
*Published Online:*
*October 13, 2024*

**Corresponding author email:*
syahmi@upm.edu.my
shiouyih.lee@newinti.edu.my

## Abstract

*Coelostegia griffithii* is a member of Malvaceae, which is native to the west Malaysia region. It has been exploited as a timber species to produce clogs, furniture, and construction materials. However, genetic studies of this species are limited. Thus, this study describes the first complete plastid genome (plastome) sequence of *C. griffithii*. The quadripartite-structured plastome was 163,159 bp long, consisting of large (95,536 bp) and small (20,435 bp) single-copy regions, which are separated by a pair of inverted repeats (IR) regions (each 23,594 bp). A total of 133 genes were annotated, including 88 protein-coding (CDS), 37 tRNA, and eight rRNA genes. Repeat analyses recorded 250 simple sequence repeats and 50 large repeats. The preferred amino acid often ended with codon A/T based on relative synonymous codon usage analysis. When compared to the plastid CDS of *Reevesia thyrsoidea*, a total of 12 genes displayed positive selection. There was no evidence of gene block rearrangement or inversion in comparison to four other closely related species of Helicteroideae. To ascertain its molecular placement, the phylogenetic analysis was carried out with the concatenated dataset of 79 shared unique CDS of 32 taxa of Malvaceae, using maximum likelihood (ML) and approximate Bayesian test (aBayes) methods. Both the ML and aBayes trees revealed a nearly resolved and well-resolved relationship within Malvaceae, respectively; *C. griffithii* is placed in the Helicteroideae clade and has a close relationship to the three *Durio* species.

**Keywords**: Chloroplast genome, Durian, Durioneae, Genetic resources, Phylogenomics

## How to cite this:

Wong XJ, Ramaiya SD, Cheng WH, Wang ZF, Hishamuddin MS and Lee SY. Complete plastid genome of *Coelostegia griffithii* (Malvaceae): Structure, comparative and phylogenetic analysis. Asian J. Agric. Biol. xxxx(x): 2024015. DOI: https://doi.org/10.35495/ajab.2024.015

## Introduction

The Malvaceae of Malvales consists of a wide range of flowering plants. It is believed to encompass around 244 genera and 4,225 documented species. Most of the species within Malvaceae are both commercially exploited and widely recognised as agricultural crops. According to the most recent classification proposed by the Angiosperm Phylogeny Group IV, the nine subfamilies of Malvaceae are Bombacoideae, Brownlowioideae, Byttnerioideae, Dombeyoideae, Grewioideae, Helicteroideae, Malvoideae, Sterculioideae, and Tilioideae (Stevens, 2019).

*Coelostegia*, a genus under Helicteroideae, comes with at least six recognised tropical species native to the west Malaysia region (POWO, 2024). Based on morphological and anatomical traits, *Coelostegia* exhibits a tight relationship with *Durio*, *Kostermansia*, and *Neesia* (Reksodihardjo, 2014). One of them, *C. griffithii*, has a broad distribution in Peninsular Malaysia, Sumatra, Java, Singapore, and Borneo (Nadiah and Soepadmo, 2011). *Coelostegia griffithii* was first described in 1862 and is the type species of the genus (Beccari, 1886). The tree of *C. griffithii* can grow up to 40 m tall and comes with a buttress and irregularly fissured dark brown or greyish brown bark. It comes with other names by the locals, such as punggai (Malay), durian hantu (Palembang), durian hutan (Bangka), etc. Similar to *Durio*, *Kostermansia*, and *Neesia*, the fruit of *C. griffithii* is a thorny capsule that consists of four lobes in general. In Malaysia, the species is considered one of the timber species (Thorogood et al., 2022), while like *Piper betle* and *Persicaria odorata*, the leaf is used by the locals as a traditional remedy for mastitis (Ong and Nordiana, 1999; Basit et al., 2023). *Coelostegia* is not considered a viable crop for agriculture because its aril content is significantly lower than that of *Durio*. Eventually, studies on *C. griffithii* are lacking; to date, the NCBI GenBank database recorded only two DNA sequences, namely the partial sequence of plastid *mat*K and *rbc*L genes (as of 15 January 2024). This raised the concern to expand its genomic information for better understanding of this species.

Plastids, i.e., chloroplasts, are organelles that semi-autonomously exist inside plant cells. They play a vital role in photosynthesis and the production of molecules necessary for the survival of plants. The size of the plastid genome (plastome) is relatively smaller compared to the mitochondrial and nuclear genomes, making it easier to assemble and have a lower recombination rate (Palmer, 1990). When compared to the nuclear genome, the plastome exhibits a slower rate of evolution and contains more conserved sequences. In general, most land plants have the plastome size typically ranged between 110 and 210 kilobase (kb), consisting of 110 to 130 genes and has a conserved circular structure in general. This structure is divided into two inverted repeats (IRs), which separate the plastome into a large single-copy (LSC) and a short single-copy (SSC) region. In general, and they are useful for phylogenetic studies because of their maternal inheritance and well conserved genomic structure (Sugiura, 1992). The analysis of the plastome sequence in various species of Malvaceae is a well-explored subject. Researchers have successfully decoded over 50 plastome sequences from different species in Malvaceae and have documented their application in creating the phylogenetic trees to study the classification and evolutionary pattern of the family. However, the limited sampling size could somewhat pose a challenge in revealing the phylogenetic relationship of Malvaceae at the genus and species level.

In this study, by using the next-generation sequencing technique and bioinformatics tools, the complete genome sequence of the plastome of *C. griffithii* was acquired. This was followed by the annotation and characterisation of the plastome sequence. Genome comparative analyses were carried out with other closely related taxa of Helicteroideae. Phylogenetic analysis was carried out to determine the molecular placement of *C. griffithii* in Malvaceae, as well as to reveal its phylogenetic relationship with other closely related species.

## Material and Methods

### Plant materials, DNA isolation, and next-generation sequencing

Fresh leaves of *C. griffithii* were collected from the Ayer Hitam Forest Reserve in Selangor of Malaysia (3° 0' 11.9" N, 101° 37' 26.0" E). The species was verified by both the forest ranger and the corresponding author, Prof. Lee Shiou Yih, based on the morphological description by Reksodihardjo (2014). A voucher specimen was prepared under the collection number LSY10014. The samples were stored in an aluminium ziplock bag containing silica gels and are transported to the laboratory for total genomic DNA isolation.

The total genomic DNA was isolated using the DNeasy Plant Mini Kit (Qiagen, Germany), following the protocol provided by the manufacturer. Gel electrophoresis using 1% agarose gel was conducted to verify the integrity of the genomic DNA visually. The DNA extract was quantified using a Qubit 4 Fluorometer (Thermo Fisher Scientific, USA). Genomic library preparation with an insert size of 300 bp was constructed using the TruSeq DNA Sample Prep Kit (Illumina, USA). Next-generation sequencing was conducted by Guangzhou Jierui Biotechnology (Guangdong, China) on the Illumina NovaSeq platform (Illumina, USA).

**Plastid genome assembly and annotation**

The plastome sequence was assembled using NOVOWrap v1.20 (Wu et al., 2021) coupled with Geneious Prime v.2022.0.2 (Kearse et al., 2012). For NOVOWrap, the *rbc*L sequence of *C. griffithii* (GenBank accession no. LC736175) was assigned as the seed, while the complete plastome sequence of *Durio zibethinus* (GenBank accession no. MG138151) was included as the reference sequence in the "Map-to-reference" function in Geneious. Gene annotation and identification of inverted region junctions were carried out using GeSeq v2.03 (Tillich et al., 2017), and the Chloë program was activated as a third-party stand-alone annotator, while the remaining parameters were set as default. The assembled plastome sequence went through manual error checking, and the annotated plastome was visualised using OGDraw v1.3.1 (Greiner et al., 2019). The annotated plastome sequence of *C. griffithii* was deposited under the accession number PP082047 in the NCBI GenBank database.

**Repeat analysis**

The identification of the simple sequence repeats (SSRs) was conducted using MISA-web (Beier et al., 2017) for mono-, di-, tri-, tetra-, penta-, and hexanucleotides were based on the parameter of minimum number of repeats set at 10, 4, 4, 3, 3, and 3. REPuter programme (Kurtz et al., 2001) was employed to detect the presence of large repeats in the forms of forward, palindromic, reverse, and complement in the genome sequence. The Hamming distance was specified at 3, and the minimum repeat size was 30 bp.

**Relative synonymous codon usage and selective pressure analysis**

The protein-coding genes (CDSs) available in the plastome of *C. griffithii* was calculated for their relative synonymous codon usage (RSCU) using the RSCU analysis function embedded in PhyloSuite v1.2.3 (Zhang et al., 2020) based on the first codon of the gene sequence. Codons with an RSCU value more than 1 exhibit a positive bias, indicating that they are utilised more frequently compared to other synonymous codons. Conversely, codons with an RSCU value less than 1 demonstrate the opposite trend. An RSCU value of 1 indicates no bias. Selective pressures of the CDS in the plastome were estimated using the KaKs_Calculator 2.0 (Wang et al., 2010). The plastome of *R. thyrsoidea* (GenBank accession no. MH939148) was used as a reference. The sequences of the 79 unique shared CDSs between *R. thyrsoidea* and *C. griffithii* were extracted using PhyloSuite v1.2.3 (Zhang et al., 2020) and aligned using MACSE (Ranwez et al., 2018). The ratio of nonsynonymous to synonymous substitution (Ka/Ks) of the CDS between *R. thyrsoidea* and *C. griffithii* was determined using the Yang and Nielsen codon frequency model (YN). The settings for the initial ratio of transitions to transversion frequency (K) were adjusted to a range of $0.3 < K < 0.7$. All 79 CDSs were analysed separately.

**Genome synteny**

Genome synteny refers to the conserved arrangement of segments of DNA across different species. Based on the availability of published plastome data, four closely related taxa of Helicteroideae, including *Durio oxleyanus* (GenBank accession no. ON653424), *D. zibethinus* (GenBank accession no. MG138151), *Helicteres hirsuta* (GenBank accession no. OR806929), and *R. thyrsoidea* (GenBank accession no. MH939148), were selected and analysed together with *C. griffithii* on the presence of large-scale evolutionary events in their plastomes using MAUVE (Darling et al., 2004). The complete plastome sequences of all five taxa were aligned using progressiveMAUVE, with the plastome of *D. zibethinus* selected as the reference genome.

**Phylogenetic inference**

To identify the molecular placement of *C. griffithii*, phylogenetic analysis was conducted on the concatenated dataset of 79 shared unique CDS via two methods: maximum likelihood (ML) and approximate Bayesian test (aBayes). Identifying the best parameters for the Bayesian inference (BI) approach required an extensive amount of time and required determining generations needed to work optimally for the outcome's coverage. However, the

aBayes approach overcomes the BI method's bottleneck and generates results that were relatively reliable. A total of 32 complete plastome sequences representing 32 taxa of Malvaceae were obtained from the NCBI GenBank database, while four closely related taxa, including *Aquilaria malaccensis* (GenBank accession no. MH286934) and *Gonystylus affinis* (GenBank accession no. MN147872) of Thymelaeaceae, as well as *Neobalanocarpus heimii* (GenBank accession no. MH746730) and *Vatica odorata* (GenBank accession no. MZ160998) of Dipterocarpaceae, were included as outgroups. Prior to CDS extraction using PhyloSuite v1.2.3, the downloaded plastome sequences were reannotated using GeSeq to ensure the consistency of gene annotations. The extracted CDSs were MAFFT-aligned and concatenated using PhyloSuite v1.2.3 (Zhang et al., 2020). PartitionFinder v2.11 (Lanfear et al., 2017) that is available in PhyloSuite was used to determine the optimal data-partitioning scheme and nucleotide substitution models following the Bayesian information criterion (BIC). The unlinked branch lengths and the greedy search algorithm were also applied in the analysis. Using IQ-TREE v.1.6.8 (Nguyen et al., 2015), the ML branch supports were obtained using the Shimodaira-Hasegawa approximate likelihood-ratio test (SH-aLRT) and the ultrafast bootstrapping algorithm (UFboot), based on 1,000 bootstrap replicates, while the aBayes tree was calculated based on posterior probability (PP). The final trees were visualised using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

## Results

### Plastome organization and features

Approximately 6 Gb of raw data was generated and used for plastome assembly. With the minimum and average depth coverage of 56× and 335×, respectively, the size of complete plastome of *C. griffithii* was 163,159 bp. The plastome is quadripartite in structure, which consists of a 95,536-bp LSC region, a 20,435-bp SSC region, and a pair of 23,594-bp IRs (Figure 1). A total of 133 genes, including 88 CDS, 37 transfer RNA (tRNA), and eight ribosomal RNA (rRNA) genes, were annotated (Table 1). Among the unique CDSs, nine of them have one intron, including *atp*F, *ndh*A/B, *pet*B/D, *rpl*2/16, *rpo*C1, and *rps*16, while two of them contain two introns, namely *clp*P1 and *paf*1. The plastome has a base content of A (31.6%), T (32.6%), C (18.0%), and G (17.7%). The overall GC content was 35.7%.

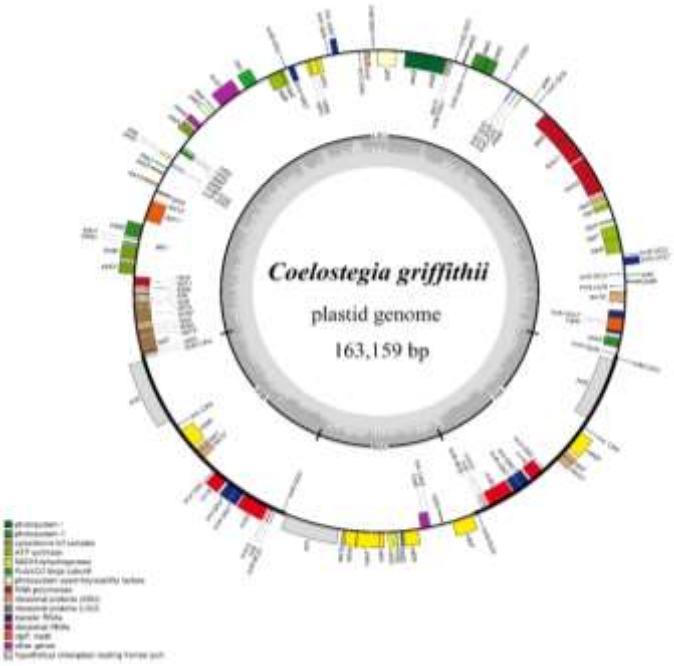


**Figure-1. Plastid genome of *Coelostegia griffithii*. The genes located in the inner circle are transcribed in a clockwise direction, whereas the genes in the outer circle are transcribed in a counter-clockwise direction. The colour codes correspond to distinct functional groupings of the genes. The boundaries of the inverted repeats are shown in thick black lines.**

**Table-1. List of annotated genes in the plastid genome of *Coelostegia griffithii*.**

| Category | Group of Function | List of Genes |
|---|---|---|
| Photosynthesis related | ATP synthase | *atp*A, *atp*B, *atp*E, *atp*F, *atp*H, *atp*I |
| | Cytochrome b6/f complex | *pet*A, *pet*B, *pet*D, *pet*G, *pet*L, *pet*N |
| | Cytochrome c synthesis | *ccs*A |
| | NADH oxidoreductase | *ndh*A, *ndh*B(×2), *ndh*C, *ndh*D, *ndh*E, *ndh*F, *ndh*G, *ndh*H, *ndh*I, *ndh*J, *ndh*K |
| | Photosystem I | *psa*A, *psa*B, *psa*C, *psa*I, *psa*J |
| | Photosystem II | *psb*A, *psb*B, *psb*C, *psb*D, *psb*E, *psb*F, *psb*H, *psb*I, *psb*J, *psb*K, *psb*L, *psb*M, *psb*T, *psb*Z |
| | Photosystem assembly factors | *paf*I, *paf*II, *pbf*I |
| | Rubisco | *rbc*L |
| Self-replication related | DNA-dependent RNA polymerase | *rpo*A, *rpo*B, *rpo*C1, *rpo*C2 |
| | Large subunit of ribosome proteins | *rpl*2, *rpl*14, *rpl*16, *rpl*20, *rpl*22, *rpl*23, *rpl*32, *rpl*33, *rpl*36 |
| | Small subunit of ribosomal proteins | *rps*2, *rps*3, *rps*4, *rps*7(×2), *rps*8, *rps*11, *rps*12, *rps*14, *rps*15, *rps*16, *rps*18, *rps*19 |
| | transfer RNA | *trn*A-UGC(×2), *trn*C-GCA, *trn*D-GUC, *trn*E-UUC, *trn*F-GAA, *trn*G-GCC, *trn*G-UCC, *trn*H-GUG, *trn*I-GAU(×2), *trn*K-UUU, *trn*L-CAA(×2), *trn*L-UAA, *trn*L-UAG, *trn*M-CAU(×4), *trn*N-GUU(×2), *trn*P-UGG, *trn*Q-UUG, *trn*R-ACG(×2), *trn*R-UCU, *trn*S-GCU *trn*S-GGA, *trn*S-UGA, *trn*T-GGU, *trn*T-UGU, *trn*V-GAC(×2), *trn*V-UAC, *trn*W-CCA, *trn*Y-GUA |
| | ribosomal RNA | *rrn*4.5(×2), *rrn*5(×2), *rrn*16(×2), *rrn*23(×2) |
| Other | Envelope membrane protein | *cem*A |
| | Maturase | *mat*K |
| | Protease | *clp*P1 |
| | Subunit acetyl-CoA-carboxylase | *acc*D |
| | translational initiation factor 1 | *inf*A |
| Unknown function | Conserved hypothetical chloroplast reading frames | *ycf*1, *ycf*2(×2) |

Note: ×2 = comes in duplicates, ×4 = comes in four replicates

**Short and large sequence repeats**

A total of 250 SSRs were identified. For mononucleotides, the A/T repeats (n = 101) were more abundant than the C/G repeats (n = 1). For the dinucleotide repeats, two types of repeats were identified, namely AG/CT and AT/AT, which recorded only 14 and 85 units, respectively. The trinucleotide repeats present in the plastome sequence were AAG/CTT and AAT/ATT, which recorded two and eight units, respectively. Five types of tetranucleotide repeats were recorded, of which AAAT/ATTT was recorded with 11 units, while the other four types, AACT/AGTT, AATC/ATTG, ACTG/AGTC, AGAT/ATCT, had one unit each. For pentanucleotide repeats, AATAG/ATTCT had the most units (n = 13), followed by AAAAG/CTTTT with four units, AATAT/ATATT with two units, and AAAAG/CTTTT as well as AGGAT/ATCCT both had one unit each. There is one unit recorded for the hexanucleotide repeat AAAAAT/ATTTTT. For large sequence repeats, a total of 50 repeats were identified, of which 43 were forward repeats and the other seven were palindromic.

**Amino acids frequency and codon usage**

The overall CDS in the plastome of *C. griffithii* was 81,165 bp (27,055 codons). Among the amino acids, leucine (Leu) was the most frequent (n = 2,856),

followed by isoleucine (Ile), which was 2,322 units. Cysteine (Cys) was the least amino acid (n = 302), followed by tryptophan (Trp) (n = 468) (Table 2). RSCU revealed that the type of amino acids that had high preferences often ended with codon A or U, except for the amino acid combination UUG of Leu (RSCU = 1.24) (Figure 2). Two amino acids displayed no preferences, i.e. methionine (Met) and Trp. Between the start codons ATG and GTG, the start codon ATG was most preferred among the CDSs; only three CDSs start with GTG, i.e., *atp*I, *psb*C, and *ndh*D. For the stop codons, the amino acid combination UAA had a higher preference when compared to UAG and UGA.

**Table-2. The amino acids and relative synonymous codon usage for the plastid genome of *Coelostegia griffithii*.**

| Amino acid | Codon | Count | RSCU |
|---|---|---|---|
| Alanine (Ala) | GCA | 386 | 1.08 |
| | GCC | 240 | 0.67 |
| | GCG | 162 | 0.45 |
| | GCU | 639 | 1.79 |
| Arginine (Arg) | AGA | 493 | 1.83 |
| | AGG | 179 | 0.66 |
| | CGA | 381 | 1.41 |
| | CGC | 106 | 0.39 |
| | CGG | 121 | 0.45 |
| | CGU | 339 | 1.26 |
| Asparagine (Asn) | AAC | 302 | 0.45 |
| | AAU | 1036 | 1.55 |
| Aspartic acid (Asp) | GAC | 243 | 0.43 |
| | GAU | 892 | 1.57 |
| Cysteine (Cys) | UGC | 74 | 0.49 |
| | UGU | 228 | 1.51 |
| Glutamic acid (Glu) | GAA | 1077 | 1.49 |
| | GAG | 370 | 0.51 |
| Glutamine (Gln) | CAA | 735 | 1.53 |
| | CAG | 225 | 0.47 |
| Glycine (Gly) | GGA | 739 | 1.62 |
| | GGC | 179 | 0.39 |
| | GGG | 319 | 0.7 |
| | GGU | 591 | 1.29 |
| Histidine (His) | CAC | 160 | 0.51 |
| | CAU | 464 | 1.49 |
| Isoleucine (Ile) | AUA | 728 | 0.94 |
| | AUC | 440 | 0.57 |
| | AUU | 1154 | 1.49 |
| Leucine (Leu) | CUA | 387 | 0.81 |
| | CUC | 183 | 0.38 |
| | CUG | 184 | 0.39 |
| | CUU | 607 | 1.28 |
| | UUA | 906 | 1.9 |
| | UUG | 589 | 1.24 |
| Lysine (Lys) | AAA | 1104 | 1.5 |
| | AAG | 366 | 0.5 |
| Methionine (Met) | AUG | 614 | 1 |
| Phenylalanine (Phe) | UUC | 542 | 0.7 |
| | UUU | 1007 | 1.3 |
| Proline (Pro) | CCA | 337 | 1.2 |
| | CCC | 188 | 0.67 |

| | | | |
|---|---|---|---|
| | CCG | 163 | 0.58 |
| | CCU | 432 | 1.54 |
| Serine (Ser) | AGC | 125 | 0.37 |
| | AGU | 423 | 1.24 |
| | UCA | 419 | 1.23 |
| | UCC | 318 | 0.93 |
| | UCG | 174 | 0.51 |
| | UCU | 582 | 1.71 |
| stop | UAA | 48 | 1.66 |
| | UAG | 20 | 0.69 |
| | UGA | 19 | 0.66 |
| Threonine (Thr) | ACA | 410 | 1.21 |
| | ACC | 251 | 0.74 |
| | ACG | 155 | 0.46 |
| | ACU | 536 | 1.59 |
| Tryptophan (Trp) | UGG | 468 | 1 |
| Tyrosine (Tyr) | UAC | 213 | 0.41 |
| | UAU | 825 | 1.59 |
| Valine (Val) | GUA | 545 | 1.5 |
| | GUC | 195 | 0.53 |
| | GUG | 212 | 0.58 |
| | GUU | 506 | 1.39 |

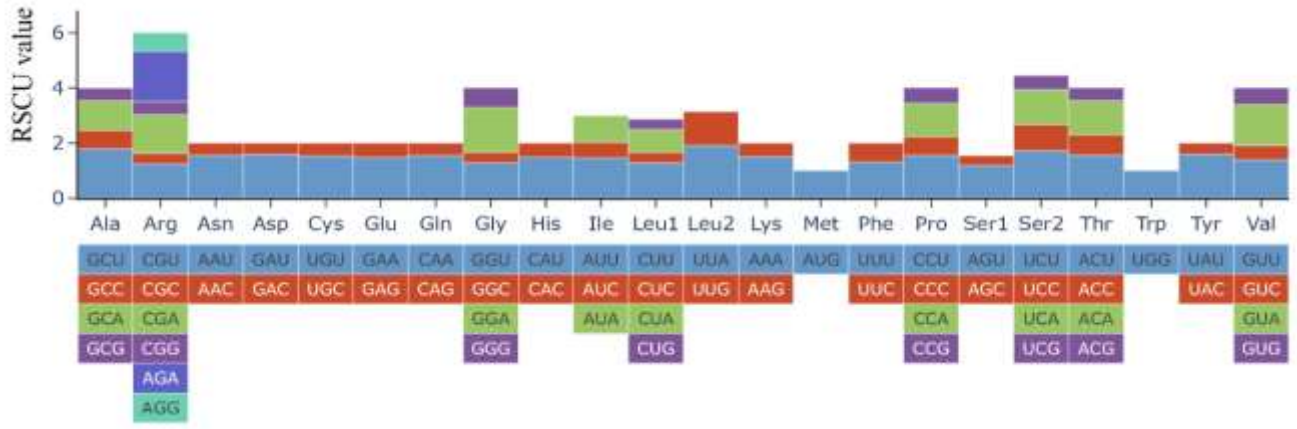


**Figure-2. Relative synonymous codon usage of protein-coding genes in the plastid genome of *Coelostegia griffithii* and the amino acids encoded by these codons.**

By using the 79 CDSs of *R. thyrsoidea* as the reference, due to the constraints of the used model, 20 of them had their Ka and/or Ks = 0 (Table 3). Among all the CDSs, 47 showed signals of purifying selection, which these genes have prevented an excess of synonymous substitutions over nonsynonymous substitutions by indicating alterations in amino acid residues. On the other hand, 12 CDSs, specifically *atp*E, *cem*A, *clp*P1, *inf*A, *paf*II, *psb*I, *rpl*14, *rpl*22, *rpl*32, *rps*11, *ycf*1, and *ycf*2, exhibited positive selection, with a Ka/Ks ratio of at least and more than 1.00.

**Structure comparison**

Genome structure comparison using *D. zibethinus* as the reference genome showed that there were only one gene block present in all five selected taxa included in the analysis, illustrating high degree of similarity of the gene content and gene order (Figure 3). There was no gene block inversion was detected among these four taxa when compared to the plastome sequence of *D. zibethinus*.

**Table-3. Synonymous (Ks), nonsynonymous (Ka) nucleotide substitution rate and Ka/Ks ratios of 79 protein-coding genes from the plastid genome of *Coelostegia griffithii*.**

| CDS | Length | Ka | Ks | Ka/Ks | p-value |
|---|---|---|---|---|---|
| *acc*D | 1428 | 0.10 | 0.16 | 0.59 | 0.02 |
| *atp*A | 1521 | 0.01 | 0.04 | 0.26 | 0.00 |
| *atp*B | 1494 | 0.01 | 0.01 | 0.81 | 0.67 |
| *atp*E | 399 | 0.04 | 0.02 | 2.09 | 0.52 |
| *atp*F | 552 | 0.01 | 0.02 | 0.45 | 0.33 |
| *atp*H | 243 | 0.01 | 0.02 | 0.73 | 0.59 |
| *atp*I | 684 | 0.00 | 0.01 | 0.30 | 0.23 |
| *ccs*A | 963 | 0.02 | 0.03 | 0.51 | 0.16 |
| *cem*A | 687 | 0.03 | 0.01 | 4.82 | 0.13 |
| *clp*P1 | 591 | 0.29 | 0.13 | 2.28 | 0.00 |
| *inf*A | 207 | 0.03 | 0.02 | 1.55 | 0.80 |
| *mat*K | 1506 | 0.02 | 0.03 | 0.75 | 0.43 |
| *ndh*A | 1089 | 0.01 | 0.05 | 0.18 | 0.00 |
| *ndh*B | 1530 | 0.01 | 0.01 | 0.88 | 0.96 |
| *ndh*C | 360 | 0.00 | 0.01 | 0.00 | 0.00 |
| *ndh*D | 1500 | 0.01 | 0.04 | 0.15 | 2.87E-05 |
| *ndh*E | 303 | 0.00 | 0.02 | 0.00 | 0.00 |
| *ndh*F | 2202 | 0.01 | 0.08 | 0.17 | 7.54E-10 |
| *ndh*G | 528 | 0.01 | 0.03 | 0.49 | 0.27 |
| *ndh*H | 1179 | 0.01 | 0.06 | 0.08 | 1.25E-06 |
| *ndh*I | 501 | 0.00 | 0.02 | 0.24 | 0.17 |
| *ndh*J | 474 | 0.01 | 0.03 | 0.27 | 0.12 |
| *ndh*K | 675 | 0.01 | 0.03 | 0.30 | 0.06 |
| *paf*I | 504 | 0.00 | 0.02 | 0.00 | 0.00 |
| *paf*II | 552 | 0.01 | 0.01 | 1.10 | 0.70 |
| *pbf*1 | 129 | 0.00 | 0.03 | 0.00 | 0.00 |
| *pet*A | 960 | 0.01 | 0.04 | 0.18 | 0.00 |
| *pet*B | 645 | 0.00 | 0.03 | 0.13 | 0.01 |
| *pet*D | 480 | 0.00 | 0.01 | 0.27 | 0.33 |
| *pet*G | 111 | 0.00 | 0.00 | n/a | 0.00 |
| *pet*L | 93 | 0.01 | 0.05 | 0.28 | 0.35 |
| *pet*N | 87 | 0.00 | 0.00 | n/a | 0.00 |
| *psa*A | 2250 | 0.00 | 0.03 | 0.02 | 3.67E-10 |
| *psa*B | 2202 | 0.00 | 0.01 | 0.14 | 0.00 |
| *psa*C | 243 | 0.00 | 0.04 | 0.00 | 0.00 |
| *psa*I | 99 | 0.01 | 0.00 | n/a | 0.00 |
| *psa*J | 126 | 0.00 | 0.03 | 0.00 | 0.00 |
| *psb*A | 1059 | 0.00 | 0.01 | 0.10 | 0.03 |
| *psb*B | 1524 | 0.00 | 0.03 | 0.06 | 3.23E-05 |
| *psb*C | 1383 | 0.00 | 0.04 | 0.00 | 0.00 |
| *psb*D | 1059 | 0.00 | 0.02 | 0.00 | 0.00 |
| *psb*E | 249 | 0.00 | 0.04 | 0.00 | 0.00 |
| *psb*F | 117 | 0.00 | 0.00 | n/a | 0.00 |
| *psb*H | 219 | 0.01 | 0.06 | 0.19 | 0.08 |
| *psb*I | 108 | 0.12 | 0.10 | 1.17 | 0.71 |
| *psb*J | 120 | 0.00 | 0.04 | 0.00 | 0.00 |
| *psb*K | 189 | 0.01 | 0.03 | 0.54 | 0.51 |
| *psb*L | 114 | 0.00 | 0.00 | n/a | 0.00 |
| *psb*M | 102 | 0.00 | 0.04 | 0.00 | 0.00 |
| *psb*T | 99 | 0.00 | 0.04 | 0.00 | 0.00 |

| | | | | | |
|---|---|---|---|---|---|
| *psb*Z | 186 | 0.00 | 0.03 | 0.00 | 0.00 |
| *rbc*L | 1431 | 0.01 | 0.03 | 0.46 | 0.05 |
| *rpl*2 | 822 | 0.00 | 0.02 | 0.08 | 0.01 |
| *rpl*14 | 366 | 0.01 | 0.01 | 1.17 | 0.72 |
| *rpl*16 | 405 | 0.01 | 0.01 | 0.95 | 0.66 |
| *rpl*20 | 276 | 0.07 | 0.08 | 0.83 | 0.74 |
| *rpl*22 | 501 | 0.05 | 0.05 | 1.00 | 0.99 |
| *rpl*23 | 279 | 0.01 | 0.04 | 0.26 | 0.19 |
| *rpl*32 | 162 | 0.08 | 0.05 | 1.58 | 0.84 |
| *rpl*33 | 198 | 0.01 | 0.02 | 0.29 | 0.35 |
| *rpl*36 | 111 | 0.01 | 0.04 | 0.32 | 0.38 |
| *rpo*A | 981 | 0.01 | 0.05 | 0.16 | 0.00 |
| *rpo*B | 3210 | 0.01 | 0.03 | 0.33 | 0.00 |
| *rpo*C1 | 2043 | 0.01 | 0.03 | 0.40 | 0.02 |
| *rpo*C2 | 4107 | 0.01 | 0.03 | 0.39 | 0.00 |
| *rps*2 | 708 | 0.02 | 0.02 | 0.94 | 0.96 |
| *rps*3 | 654 | 0.01 | 0.05 | 0.28 | 0.03 |
| *rps*4 | 603 | 0.01 | 0.05 | 0.17 | 0.01 |
| *rps*7 | 465 | 0.04 | 0.00 | n/a | 0.01 |
| *rps*8 | 402 | 0.02 | 0.03 | 0.51 | 0.29 |
| *rps*11 | 414 | 0.04 | 0.04 | 1.08 | 0.98 |
| *rps*12 | 369 | 0.01 | 0.00 | n/a | 0.13 |
| *rps*14 | 300 | 0.02 | 0.03 | 0.60 | 0.42 |
| *rps*15 | 270 | 0.02 | 0.04 | 0.66 | 0.45 |
| *rps*16 | 264 | 0.00 | 0.09 | 0.05 | 0.00 |
| *rps*18 | 282 | 0.03 | 0.04 | 0.72 | 0.48 |
| *rps*19 | 276 | 0.04 | 0.09 | 0.51 | 0.12 |
| *ycf*1 | 5217 | 0.13 | 0.11 | 1.10 | 0.34 |
| *ycf*2 | 6687 | 0.04 | 0.03 | 1.75 | 0.00 |

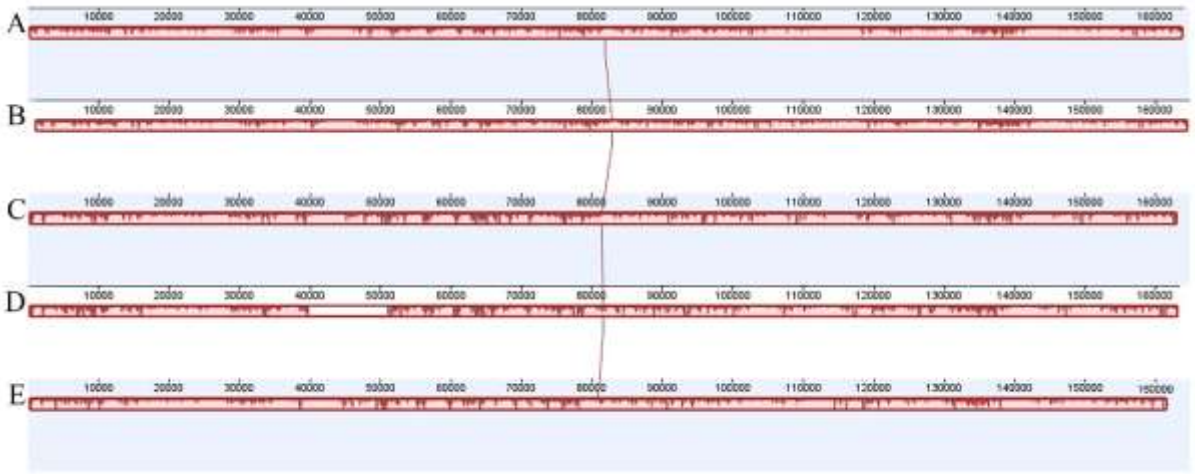


**Figure-3. Structures of locally collinear blocks (LCB) using MAUVE with *Durio zibethinus* shown on top as references genome. The coloured LCB boxes represent the highly conserved collinear sections, while the block histograms display the level of sequence similarity. (A: *D. zibethinus*; B: *Durio oxleyanus*; C: *Coelostegia griffithii*; D: *Helicteres hirsuta*; E: *Reevesia thyrsoidea*)**

**Phylogenetic analysis**

Phylogenetic analysis based on the 79 shared unique CDS of 32 taxa of Malvaceae revealed a nearly resolved relationship in Malvaceae (Figure 4). The branch nodes are considered well supported when the bootstrap value for SH-aLRT (left) is the

same as or more than 80%; PP (centre) is the same as or more than 0.95; and UFboot (right) is the same as or more than 95%; and posterior probability (PP) is greater than 0.95. Based on ML, branch nodes that are not reliable include the split for the clade *Firmiana danxiaensis+Heritiera parvifolia+Sterculia nobilis* (SH-aLRT = 66%, UFboot = 73%), the divergence of *Theobroma cacao* (SH-aLRT = 55%, UFboot = 72%), the divergence of *Gossypium hirsutum* (79%, UFboot = 82%), and the divergence of *Ochroma pyramidale* (SH-aLRT = 79%, UFboot = 78%). Based on the current sampling size, *C. griffithii* was clustered with *Durio dulcis*, *D. oxleyanus*, *D. zibethinus*, *Helicteres hirsuta*, and *R. thyrsoidea*; *C. griffithii* is sister to the *Durio* clade.

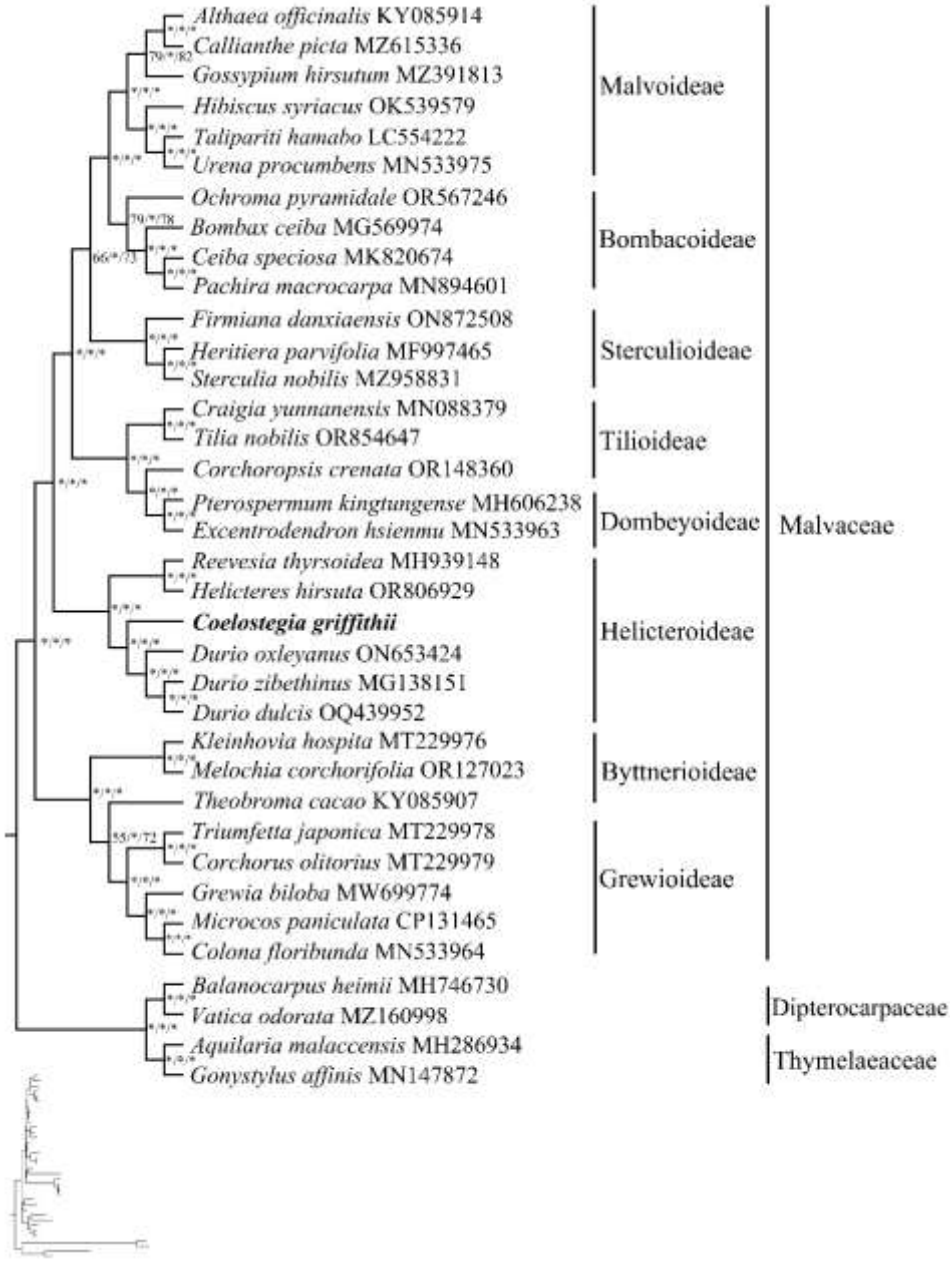


**Figure-4. Phylogenetic tree reconstructed using maximum likelihood (ML) and approximate Bayesian test (aBayes). Branch node that comes with a reliable support value for a non-parametric variant and approximate likelihood-ratio test (SH-aLRT ≥ 80%, left); posterior probability (PP ≥ 0.95, centre); and ultrafast bootstrap (UFBoot ≥ 95%, right) is indicated with an asterisk (*).**

## Discussion

We assembled and characterised the first complete plastome sequence of *C. griffithii*. The plastome of *C. griffthii* displayed similarity to those of other angiosperms by having a quadripartite structure. When compared to other published plastome sequences of Durioneae, *C. griffithii* also has a smaller genome size than *D. oxleyanus* and *D. zibethinus* (Cheon et al., 2017; Wong et al., 2022), but is greater than those of *Reevesia* taxa (Quan et al., 2019; Zhang et al., 2022). Eventually, the overall GC content for the plastome sequence of *C. griffithii* was similar to that of *D. oxleyanus* and *D. zibethinus* (i.e., 35.8%), yet it was comparably lower when compared to *Reevesia* (i.e., 36.8%). There was a difference in total gene number between *C. griffithii* and the other four published plastomes of *D. oxleyanus*, *D. zibethinus*, *R. thyrsoidea*, and *R. pycnantha*. Eventually, all five species had the same amount of tRNA and rRNA genes, except for *D. zibethinus* (i.e., 30 and four, respectively), while *D. oxleyanus* has the most CDS (n = 90), followed by *C. griffithii*. The variation in the total gene count in the plastome of these closely related species can be explained by the use of different annotation tools (Tonti et al., 2017). It is worth noting that a closely related species, *D. zibethinus* also exhibited a stereotypical tripartite plastome structure when assembled using third-generation sequencing data (Shearman et al., 2020). In most cases, we would assume that the assembly process would have been misconducted; however, assembly results from long read sequencing techniques are usually considered highly reliable (Ly et al., 2020). Plastomes are functionally divided into three parts because the IR evolves as a cohesive unit, with equal amounts of molecules in the single-copy regions becoming inverted in relation to each other (Knox, 2014). As tripartite plastome structures were also reported in other plant species from Rubiaceae, in which only one IR copy was present (Ly et al., 2020), as well as in Aristolochiaceae, in which the SSC region was incorporated into the IR regions (Lim et al., 2018), we do not exclude the possibility that *Coelostegia* species would have undergone a similar evolution pattern like *Durio*.

SSRs and large repeats are significant sites of genetic mutations, as they can cause changes in the DNA sequence and rearrangements in the genome. These mutations occur because of errors in DNA replication and recombination (Borsch and Quandt, 2009). In some cases, these repeated loci could be developed into specific markers that contribute to the identification of species and the study of population (Li et al., 2020). In *C. griffithii*, the mononucleotide repeats had the highest count among the other repeats, which have a greater number of motifs. Among the mono-, di-, tri-, and tetranucleotide repeats, the dominant repeat combinations for each type of repeat exhibit strong A/T preferences, i.e., A/T for mononucleotide repeats, AT/AT for dinucleotide repeats, AAT/AAT for trinucleotide repeats, and AAAT/ATTT for tetranucleotide repeats. Based on previous works, the A and T combination is usually the richest among all SSR types identified (Vieira et al., 2016). The combination of A/T in most SSRs would be the reason for the high AT content in the complete plastome. However, for pentanucleotide repeats, the dominant motif combination was not an A/T combination, i.e., AATAG/ATTCT, while the only hexanucleotide repeat still belonged to an A/T combination.

Eukaryotes commonly exhibit a preference for certain synonymous codons that encode an amino acid, which is caused by differences in their usage frequency (Lyu and Liu, 2020). Other research suggested that the codon use frequency remains consistent in numerous plant plastomes, and the codon use in the plastome is significantly influenced by two primary factors: selection and mutation, which are demonstrated in members of Asteraceae (Nie et al., 2014) and Euphorbiaceae (Wang et al., 2020). However, mutation exhibits varying effects on certain genes within the genome, while at the same time applying selective restrictions to codon usage (Kalkus et al., 2021). The codon usage patterns in *C. griffithii* could be influenced by the composition bias resulting from the significant fraction of A/T, which is comparable to previous reported plastomes. Nonetheless, Leu is the most abundant amino acid in *C. griffithii*, whereas Cys is the least abundant. This finding is consistent with prior studies, i.e., He et al. (2021) and Xiao et al. (2022).

Most protein-coding areas are more prone to experiencing synonymous substitutions as opposed to non-synonymous alterations. (Rono et al., 2020). Under a purifying selection, the synonymous substitution results in no change to the amino acid, while the change in the sequence of the amino acid is a result of a positive selection. By assuming that *C. griffthii* has a wide natural habitat distribution, it is expected that the CDSs were under purifying

selection. However, 12 CDSs showed signals of positive selection when compared to those of *R. thyrsoidea*, a subtropical tree species. Among the 12 genes, three are photosynthesis-related (i.e., *atp*E, *paf*II, and *psb*I), four are involved in self-replication function (i.e., *rpl*14/22/32, and *rps*11), while the rest are either genes with individual "housekeeping" functions (i.e., *cem*A, *clp*P1, and *inf*A) or come with an unknown function (i.e., *ycf*1/2). To our knowledge, there is no analysis carried out on the pressure selection between tropical and subtropical species; we deduce that the genes that are subjected to positive selection could be related to the environmental adaptation to the tropical region. On a second note, MAUVE analysis showed that the plastome of *C. griffithii* exhibits a high level of conservation in terms of gene block arrangement when compared to *D. zibethinus*. Such events align with the gradual rates of sequence and structural evolution observed in plant plastomes, especially between closely related species (Gruenstaeudl et al., 2017).

The topology of the ML tree generated from our study is congruent with the recent findings on the phylogenetic analysis of Malvaceae using the CDS dataset by Abdullah et al. (2021). A monophyly of Malvaceae was revealed; our ML tree did not display strong backbone support when based on the SH-aLRT and UFboot methods but was well-supported when using the aBayes method, however our ML tree is in accordance with the classification proposed by APGIV at the subfamily level, in which the 32 taxa used in this study were clustered into eight different clades that corresponded to the eight different subfamilies and were distinctly separated. Based on the CDS dataset, *C. griffithii* is well resolved as a member of Helicteroideae. Despite the limited sampling size used to represent Helicteroideae, the sistership between the Helictereae tribe, which was represented by two species, *Helicteres hirsuta* and *R. thyrsoidea*, and the Durioneae tribe, which includes *C. griffithii* and the three *Durio* species, is convincing based on the strong branch supports indicated in the ML tree. Based on the phylogenetic tree reconstructed using a single plastid gene, *ndh*F, *Coelostegia* is sister to *Kostermansia*; *Neesia* was the first to diverge in Durioneae, followed by the *Coelostegia+Kostermansia* clade (Nyffeler and Baum, 2001). Eventually, the divergence of the *Coelostegia+Kostermansia* clade was not well-supported at that time. The relationship between *Coelostegia* and *Cullenia*, *Kostermansia*, as well as *Durio*, remained uncertain while analysing the nuclear ribosomal DNA internal transcribed spacer (ITS) region using the gap as an informative site. Although we were only able to include *Coelostegia* and *Durio* samples in our study, the well supported relationship between *C. griffithii* and the three *Durio* species indicated that the use of the plastid CDS dataset in reconstructing a well-resolved phylogenetic tree of Durioneae could be promising. This was also being demonstrated by other published works on other plant families at the tribe level using the plastid CDS dataset, i.e., Aquilarieae of Thymelaeaceae (Lee et al., 2022). Nevertheless, the addition of samples from *Cullenia*, *Kostermansia*, and *Neesia* would better reveal the phylogenetic relationship in Durioneae.

## Conclusion

This study offers a novel insight into the complete plastome of *C. griffithii* and thus, expanded the genomic information of this understudied species. The comparative studies revealed that the plastome was highly conserved, showing similar genome structure, gene content, and gene order when compared to other closely related species. Moreover, classifications and genetic diversity analysis of Helicteroideae were made possible by SSRs, long sequence repeats, amino acid frequency, and relative synonymous codon usage analysis. Phylogenetic reconstruction that was based on CDS confirms the molecular placement of *C. griffithii*, which was grouped under a monophyly relationship with *Durio* species. Collectively, these findings laid the foundation for further genomic studies in Malvaceae, as well as providing substantial implications for taxonomic identification, phylogenetic studies, and evolutionary research in *Coelostegia*.

## Acknowledgment

The authors thank Sultan Idris Shah Forestry Education Centre (SISFEC) of Universiti Putra Malaysia (UPM) for the approval and assistance in sample collection from the Ayer Hitam Forest Reserve, Selangor, Malaysia. This work is supported by the Ministry of Education, Malaysia, through the Fundamental Research Grant Scheme (FRGS) Fund, under Grants FRGS/1/2023/ST01/INTI/01/1 and FRGS/1/2020/STG03/UPM/02/10.

**Disclaimer:** None.
**Conflict of Interest:** None.
**Source of Funding:** The Ministry of Education, Malaysia.

## Contribution of Authors

Wong XJ: Formal analysis, software and original draft preparation
Ramaiya SD: Resources, supervision, validation, manuscript review and editing
Cheng WH: Project administration, supervision, validation, manuscript review and editing
Wang ZF: Formal analysis, software and validation
Hishamuddin MS: Conceptualization, resources, manuscript review and editing
Lee SY: Conceptualization, resources, supervision, manuscript review and editing

## References

Abdullah, Mehmood F, Shahzadi I, Ali Z, Islam M, Naeem M, Mirza B, Lockhart PJ, Ahmed I and Waheed MT, 2021. Correlations among oligonucleotide repeats, nucleotide substitutions, and insertion–deletion mutations in chloroplast genomes of plant family Malvaceae. *Journal of Systematics and Evolution 59(2): 388-402. DOI: https://doi.org/10.1111/jse.12585.*

Basit MA, Kadir AA, Chwen LT, Salleh A, Kaka U, Idris SB, Farooq AA, Javid MA and Murtaza S, 2023. Qualitative and quantitative phytochemical analysis, antioxidant activity and antimicrobial potential of selected herbs *Piper betle* and *Persicaria odorata* leaf extracts. Asian Journal of Agriculture and Biology 2023(3). DOI: 10.35495/ajab.2023.038.

Beccari O, 1886. Le Bombacaceae Malesi descritte ed illustrate. Malesia 3(4): 202-280.

Beier S, Thiel T, Münch T, Scholz U and Mascher M, 2017. MISA-web: a web server for microsatellite prediction. Bioinformatics 33: 2583-2585. DOI: https://doi.org/10.1093/bioinformatics/btx198.

Borsch T and Quandt D, 2009. Mutational dynamics and phylogenetic utility of noncoding chloroplast DNA. Plant Systematics and Evolution 282: 169-199. DOI: 10.1007/s00606-009-0210-8.

Cheon SH, Jo S, Kim HW, Kim YK, Sohn JY and Kim KJ, 2017. The complete plastome sequence of Durian, *Durio zibethinus* L. *(*Malvaceae). *Mitochondrial DNA Part B 2(2): 763-764. DOI:* https://doi.org/10.1080/23802359.2017.1398615.

Darling ACE, Mau B, Blattner FR and Perna NT, 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Research 14: 1394-1403. DOI: https://doi.org/10.1101/gr.2289704.

Greiner S, Lehwark P and Bock R, 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Research 47(1): 59-64. DOI: https://doi.org/10.1093/nar/gkz238.

Gruenstaeudl M, Nauheimer L and Borsch T, 2017. Plastid genome structure and phylogenomics of Nymphaeales: conserved gene order and new insights into relationships. Plant Systematics and Evolution 303: 1251-1270. DOI: https://doi.org/10.1007/s00606-017-1436-5.

He L, Zhang Y and Lee SY, 2021. Complete plastomes of six species of *Wikstroemia* (Thymelaeaceae) reveal paraphyly with the monotypic genus *Stellera*. *Scientific Reports 11(1): 13608.* DOI: https://doi.org/10.1038/s41598-021-93057-3.

Kalkus A, Barrett J, Ashok T and Morton BR, 2021. Evidence from simulation studies for selective constraints on the codon usage of the angiosperm *psb*A gene. *PLoS Computational Biology 17(10): e1009535.* DOI: https://doi.org/10.1371/journal.pcbi.1009535.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowits S, Duran C, Thierer T, Ashton B, Meintjes P and Drummond A, 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12): 1647-1649. DOI: https://doi.org/10.1093/bioinformatics/bts199.

Knox EB, 2014. The dynamic history of plastid genomes in the Campanulaceae sensu lato is unique among angiosperms. *Proceedings of the National Academy of Sciences 111(30): 11097-11102.* DOI: https://doi.org/10.1073/pnas.1403363111.

Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J and Giegerich R, 2001. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Research 29:

4633–4642. DOI: https://doi.org/10.1093/nar/29.22.4633.

Lanfear R, Frandsen PB, Wright AM, Senfeld T and Calcott B, 2017. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Molecular Biology and Evolution 34: 772-773. DOI: https://doi.org/10.1093/molbev/msw260.

Lee SY, Turjaman M, Chaveerach A, Subasinghe S, Fan Q and Liao W, 2022. Phylogenetic relationships of *Aquilaria* and *Gyrinops* (Thymelaeaceae) revisited: evidence from complete plastid genomes. *Botanical Journal of the Linnean Society 200(3): 344-359.* DOI: https://doi.org/10.1093/botlinnean/boac014.

Li C, Zheng Y and Huang P, 2020. Molecular markers from the chloroplast genome of rose provide a complementary tool for variety discrimination and profiling. Scientific Reports 10: 12188. DOI: 10.1038/s41598-020-68092-1.

Lim CE, Lee SC, So S, Han SM, Choi JE and Lee BY, 2018. The complete chloroplast genome sequence of *Asarum sieboldii* Miq. (Aristolochiaceae), a medicinal plant in Korea. Mitochondrial DNA Part B 3(1): 118-119. DOI: https://doi.org/10.1080/23802359.2018.1424577.

Ly SN, Garavito A, Block PD, Asselman P, Guyeux C, Charr JC, Janssens S, Mouly A, Hamon P and Guyot R, 2020. Chloroplast genomes of Rubiaceae: comparative genomics and molecular phylogeny in subfamily Ixoroideae. *PLoS One 15(4): e0232295.* DOI: https://doi.org/10.1371/journal.pone.0232295.

Lyu X and Liu Y, 2020. Nonoptimal codon usage is critical for protein structure and function of the master general amino acid control regulator CPC-1. *Mbio 11(5): 10-1128.* DOI: https://doi.org/10.1128/mbio.02605-20.

Nadiah I and Soepadmo E, 2011. A synopsis of *Coelostegia* (Bombacaceae/Malvaceae: Helicteroideae: Durioneae) and new records from Borneo. The Gardens' Bulletin Singapore 63 (1 and 2): 125-135.

Nie X, Deng P, Feng K, Liu P, Du X, You FM and Weining S, 2014. Comparative analysis of codon usage patterns in chloroplast genomes of the Asteraceae family. *Plant Molecular Biology Reporter 32: 828-840.* DOI: https://doi.org/10.1007/s11105-013-0691-z.

Nyffeler R and Baum DA, 2001. Systematics and character evolution in *Durio* s.lat. (Malvaceae/Helicteroideae/Durioneae or Bombacaceae-Durioneae). Organisms Diversity and Evolution 1(3): 165-178. DOI: https://doi.org/10.1078/1439-6092-00015.

Nguyen LT, Schmidt HA, Haeseler A and Minh BQ, 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution 32: 268-274. DOI: https://doi.org/10.1093/molbev/msu300.

Ong HC and Nordiana M, 1999. Malay ethno-medico botany in Machang, Kelantan, Malaysia. *Fitoterapia 70(5): 502-513.* DOI: https://doi.org/10.1016/s0367-326x(99)00077-5.

Palmer JD, 1990. Contrasting modes and tempos of genome evolution in land plant organelles. Trends in Genetics 6: 115-120. DOI: https://doi.org/10.1016/0168-9525(90)90125-p.

POWO, 2024. Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. Available online: http://www.plantsoftheworldonline.org/.

Quan G, Zou P, Liu G, Sun M, Wang W and Dai S, 2019. The complete chloroplast genome of *Reevesia thyrsoidea* (Malvaceae). *Mitochondrial DNA Part B 5(1): 292-293.* DOI: https://doi.org/10.1080/23802359.2019.1674206.

Ranwez V, Douzery EJP, Cambon C, Chantret N and Delsuc F, 2018. MACSE v2: toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. Molecular Biology and Evolution 35(10): 2582-2584. DOI: https://doi.org/10.1093/molbev/msy159.

Reksodihardjo WS, 2014. The genus *Coelostegia* Benth. (Bombac.). *Reinwardtia 5(3): 269-291.*

Rono PC, Dong X, Yang JX, Mutie FM, Oulo MA, Malombe I, Kirika PM, Hu GW and Wang QF, 2020. Initial complete chloroplast genomes of *Alchemilla* (Rosaceae): comparative analysis and phylogenetic relationships. *Frontiers in Genetics 11: 560368.* DOI: https://doi.org/10.3389/fgene.2020.560368.

Shearman JR, Sonthirod C, Naktang C, Sangsrakru D, Yoocha T, Chatbanyong R, Vorakuldumrongchai S, Chusri O, Tangphatsornruang S and Pootakham W, 2020. Assembly of the durian chloroplast genome using

long PacBio reads. *Scientific Reports 10(1): 15980*. DOI: https://doi.org/10.1038/s41598-020-73549-4.

Stevens PF, 2019. Angiosperm Phylogeny Website. Available online: http://www.mobot.org/MOBOT/research/APweb/.

Sugiura M, 1992. The chloroplast genome, pp. 149-168. In R.A. Schilperoort and L. Dure (eds), 10 Years Plant Molecular Biology, Springer Netherlands 19. DOI: https://doi.org/10.1007/978-94-011-2656-4_10.

Thorogood CJ, Ghazalli MN, Siti-Munirah MY, Nikong D, Kusuma YWC, Sudarmono S, and Witono JR, 2022. The king of fruits. Plants, People, Planet, Early view, 4(6), 538-547. DOI: https://doi.org/10.1002/ppp3.10288.

Tonti FJ, Nevill PG, Dixon K and Small I, 2017. What can we do with 1000 plastid genomes? The Plant Journal 90(4): 808-818. DOI: https://doi.org/10.1111/tpj.13491.

Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R and Greiner S, 2017. GeSeq-versatile and accurate annotation of organelle genomes. Nucleic Acids Research 45(1): 6-11. DOI: https://doi.org/10.1093/nar/gkx391.

Vieira MLC, Santini L, Diniz AL and Munhoz CF, 2016. Microsatellite markers: what they mean and why they are so useful. Genetics and Molecular Biology 39(3): 312-328. DOI: 10.1590/1678-4685-GMB-2016-0027.

Wang D, Zhang Y, Zhang Z, Zhu J and Yu J, 2010. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics, Proteomics and Bioinformatics 8(1): 77-80*. DOI: https://doi.org/10.1016/s1672-0229(10)60008-3.

Wang Z, Xu B, Li B, Zhou Q, Wang G, Jiang X, Wang C and Xu Z, 2020. Comparative analysis of codon usage patterns in chloroplast genomes of six Euphorbiaceae species. *PeerJ 8: e8251*. DOI: https://doi.org/10.7717/peerj.8251.

Wong XJ, Law D, Wang ZF, Ramaiya SD and Lee SY, 2022. The complete chloroplast genome sequence of *Durio oxleyanus* (Malvaceae) and its phylogenetic position. *Mitochondrial DNA Part B 7(9): 1709-1712*. DOI: https://doi.org/10.1080/23802359.2022.2123256.

Wu P, Xu C, Chen H, Yang J, Zhang X and Zhou S, 2021. NOVOWrap: an automated solution for plastid genome assembly and structure standardization. *Molecular Ecology Resources 21(6): 2177-2186*. DOI: https://doi.org/10.1111/1755-0998.13410.

Xiao T, He L, Yue L, Zhang Y and Lee SY, 2022. Comparative phylogenetic analysis of complete plastid genomes of *Renanthera* (Orchidaceae). *Frontiers in Genetics 13: 998575*. DOI: https://doi.org/10.3389/fgene.2022.998575.

Zhang YH, Chen PD, Chen DJ, Ke XR and Wang HF, 2022. Complete plastome sequence of *Reevesia pycnantha* Y. Ling (Malvaceae): a rare medicinal tree species in South Asia. *Mitochondrial DNA Part B 7(2): 343-344*. DOI: https://doi.org/10.1080/23802359.2022.2032439.

Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX and Wang GT, 2020. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Molecular Ecology Resource 20: 348-355. DOI: https://doi.org/10.1111/1755-0998.13096.