Bangladesh J. Bot. **55**(1): 37-43, 2026 (March) DOI: https://doi.org/10.3329/bjb.v55i1.88596

MOLECULAR IDENTIFICATION OF BUTTERFLY LARVAL HOST PLANTS USING CHLOROPLAST *MATK* AND *RBCL* GENES

MD MONWAR HOSSAIN*, SHRABONI DAS, MUNTAHENA RUHI AND MUHAMMAD SOHEL ABEDIN

DNA Barcoding Laboratory, Department of Zoology, Jahangirnagar University, Savar, Dhaka-1342, Bangladesh

Keywords: Butterfly, DNA barcoding, Identification, Evolution, Genbank, Conservation

Abstract

This study used DNA barcoding with *matK* and *rbcL* genes to identify thirteen butterfly larval host (BLH) plants. BLAST analysis of *matK* and *rbcL* gene sequences confirmed 99.04-100% similarity with global records. The *matK* gene showed significantly more variable (624) and parsimony-informative sites (521), making it suitable for resolving relationships at lower taxonomic levels. In contrast, *rbcL* had more conserved sites (351), highlighting its effectiveness at higher taxonomic levels. These differences underscore the complementary roles of *matK* and *rbcL* in phylogenetic studies. Phylogenetic trees were constructed with *matK* and *rbcL* gene sequences which showed consistent topologies and evolutionary relationships. The variable *matK* gene proved effective for species-level resolution, while the conserved *rbcL* gene offered stronger support at broader taxonomic classification. Robust bootstrap values (98-100) confirmed the reliability of both markers. Their complementary features make them valuable tools for DNA barcoding and understanding the evolutionary history of BLH plant species.

Introduction

Butterflies are essential components of terrestrial ecosystems, playing critical roles as pollinators, indicators of environmental health, and contributors to biodiversity. Their survival and distribution are intricately linked to the availability of specific larval host plants, which provide food and habitat for the caterpillar stage of their life cycle (Tiple *et al*. 2011). Documentation and identification butterfly larval host (BLH) plants are thus crucial for conservation, ecological research, and understanding plant-insect interactions (Dennis *et al*. 2008, Tiple *et al*. 2011). However, accurate identification of these host plants is often challenging due to the reliance on morphological characteristics, which may vary with environmental conditions or developmental stages (Jurado-Rivera *et al*. 2009).

In recent years, molecular tools have revolutionized in plant identification and phylogenetic studies, providing reliable alternatives to traditional morphology-based methods (Hollingsworth *et al*. 2011). DNA barcoding, using specific loci such as the *Maturase K* (*matK*) and *ribulose-1,5-bisphosphate carboxylase large subunit* (*rbcL*), has emerged as a robust approach for plant identification (Burgess *et al*. 2011, Hollingsworth *et al*. 2011). These chloroplast genes are widely used for their complementary properties: *matK* offers high variability, making it suitable barcode for species-level discrimination, while *rbcL* is more conserved and effective for resolving higher taxonomic relationships (CBOL 2009). Studies across the globe, including in India, Sri-Lanka, China, and tropical regions, have demonstrated the effectiveness of these loci in identifying and classifying diverse plant species (Huang *et al*. 2015, Kumar *et al*. 2018a, Rajphriyadharshini and Weerasena 2020).

In Bangladesh, butterflies are an integral part of the country's rich biodiversity, with approximately 421 species recorded to date (Larsen 2004, IUCN Bangladesh 2015, Hossain 2023). However, the larval host plants for many of these species remain undocumented, and very recently

*Author for correspondence: <monwar@juniv.edu>.

a list of 107 BLH plants were recorded from Jahangirnagar University campus, Bangladesh (Chowdhury and Hossain 2013, Shihan 2018, Das *et al*. 2025). These plants are facing a challenge for conservation planning and need habitat restoration efforts. Identification of these plants is practically critical in the context of rapid deforestation, agricultural expansion, development of roads and highways and urbanization, which threaten both butterfly species and their habitats (Tiple *et al*. 2011, Munoz-Galicia *et al*. 2023, Hossain 2023). Despite the growing recognition of their ecological importance, molecular studies focusing on butterfly larval host plants in Bangladesh are limited, creating a significant knowledge gap in this field. Therefore, this study aims to address this gap by using DNA barcoding and phylogenetic analysis to identify and classify thirteen larval host plants of butterflies from Bangladesh using *matK* and *rbcL* loci. The findings also to be compared with similar studies from other countries to provide a global perspective on the phylogenetics of larval host plants. Therefore, this research may contribute to understand plant-insect interactions in Bangladesh and also provide a foundation for conservation strategies.

Materials and Methods

A total of thirteen plant species were collected from the Jahangirnagar University (JU) campus (23°52'47.6"N 90°16'04.5"E) and labeled (Table 1). The leaves were cleaned, stored in sealable plastic packs and kept at room temperature (24°C) in the DNA Barcoding Laboratory.

For isolation of genomic DNA the leaf samples were crushed in liquid nitrogen using a mortar and pestle. A Wizard Genomic DNA Purification Kit (Promega, Madison, WI, USA) was used. The concentration and quality of the extracted DNA were determined using a NanoDrop 1000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA). The isolated genomic DNA was stored at -20°C. The target DNA regions, namely *matK* and *rbcL*, were amplified with respective universal DNA barcoding primers (CBOL 2009). For the *matK* gene, matK-Forward: 5'-CGTACAGTACTTTTGTGTTTACGAG-3' and matK-Reverse: 5'-ACCCAGTCCATCTGGAA ATCTTGGTTC-3'; for the *rbcL* gene, rbcL-Forward: 5'-ATGTCACCACAAACAGAGA CTAAAGC-3' and rbcL-Reverse: 5'-GTAAAATCAAGTCCACCRCG-3' were used. PCR was conducted within a Veriti, USA-manufactured thermal cycler, utilising 20 µL of Q2 Green PCR Master Mix. The experimental protocol involved a series of cycle conditions. Denaturation step was carried out at 95°C for 4 min. After this, 35 cycles were executed, with each cycle comprising 30 sec of primer denaturation at 95°C, primer annealing at 49-53°C, and primer extension at 72°C. The process concluded with a 5 min extension at 72°C. The amplified PCR product was run through 1% agarose gel and visualised under ultraviolet light (Bio Analyzer). The DNA sequences of the amplification PCR products were generated by a ABI 3500 sequencer.

All sequences from the BLH plant species were initially edited using Chromas version 2.6.2. Multiple sequence alignment was performed using ClustalW BioEdit version 7.0 (Hall 1999). Nucleotide compositions were calculated and summarized using MEGA version 10 (Kimura 1980, Kumar *et al*. 2018b). Phylogenetic analysis was also conducted in MEGA10, employing the Kimura 2-Parameter (K2P) model and 1000 maximum likelihood (ML) bootstrap replicates to assess branch support (Kumar *et al*. 2018b). Additional gene sequences used in the analysis were retrieved from the GenBank database.

Results and Discussion

A total of twenty six chloroplast gene sequences, representing *matK* and *rbcL*, were generated from 13 butterfly larval host plants belonging to ten families (Table 1). The gene sequences were subsequently submitted to NCBI GenBank for Accession number. The BLAST analysis revealed 99.04 to 100% similarity with corresponding species worldwide (Table 1). These results

underscore the reliability of DNA barcoding for accurate plant identification, particularly when morphological identification is difficult (Jurado-Rivera *et al*. 2009, Hollingsworth *et al*. 2011).

Table 1. Host plant species with voucher numbers, geo-locations, identities and GenBank accession numbers for the *matK* and *rbcL* genes.

Scientific name	Family	Voucher No.	GPS Coordinates	Genbank accession (*matK*)	% Identity (*matK*)	Genbank accession (*rbcL*)	% Identity (*rbcL*)
Aristolochia indica	Aristolochiaceae	PBV0005	23°52'33.2"N 90°16'05.6"E	PQ111806	99.64	PP342054	99.81
Calotropis gigantea	Apocynaceae	PBV0011	23°52'32.5"N 90°16'05.7"E	PQ111809	99.89	PP352606	99.83
Citrus aurantiifolia	Rutaceae	PBV0014	23°52'36.8"N 90°16'06.1"E	PQ111812	99.77	PP352609	100
Flacourtia indica	Salicaceae	PBV0009	23°52'35.2"N 90°16'02.1"E	PQ111807	99.36	PP352604	100
Glycosmis pentaphylla	Rutaceae	PBV0006	23°52'30.4"N 90°15'52.4"E	PQ111805	99.04	PP342052	100
Holarrhena pubescens	Apocynaceae	PBV0004	23°52'34.7"N 90°16'02.3"E	PP555181	100	PP342053	100
Litsea glutinosa	Lauraceae	PBV0003	23°52'32.9"N 90°16'06.5"E	PQ111804	100	PP342051	100
Magnolia champaca	Magnoliaceae	PBV0013	23°52'32.9"N 90°16'05.3"E	PQ111811	99.89	PP352608	99.82
Mimosa pudica	Fabaceae	PBV0017	23°52'41.1"N 90°16'06.2"E	PQ530511	100	PQ530512	100
Monoon longifolium	Annonaceae	PBV0012	23°52'31.1"N 90°16'04.5"E	PQ111810	100	PP352607	99.83
Passiflora foetida	Passifloraceae	PBV0001	23°52'34.0"N 90°16'06.0"E	PQ111803	99.62	PP342049	99.13
Senna occidentalis	Fabaceae	PBV0002	23°53'34.5"N 90°16'02.7"E	PP555180	100	PP342050	99.83
Ziziphus oenopolia	Rhamnaceae	PBV0010	23°52'35.3"N 90°16'02.8"E	PQ111808	99.36	PP352605	100

Table 2 describes nucleotide composition of *matK* and *rbcL* genes, along with it is compared the nucleotide composition of these two genes in thirteen BLH plants. The *matK* gene has more nucleotide sites (761) than *rbcL* (455). The *matK* also has significantly more variable sites (624 vs. 104) and parsimony-informative sites (521 vs. 81), indicating it is more diverse and suitable for phylogenetic studies. In contrast, *rbcL* has more conserved sites (351 vs. 178), reflecting its stability and functional conservation (Table 2). The average AT content is higher in *matK* (65.41%) than in *rbcL* (53.57%), which might affect its structural properties. *matK*'s high variability makes it ideal barcode for studying species-level evolutionary relationships. On the other hand, the conserved nature of *rbcL* makes it better suited for broader taxonomic studies. These differences highlight the complementary roles of these genes in molecular research (Burgess *et al*. 2011).

Table 2. Nucleotides composition of *matK* and *rbcL* genes sequences across BLH plants.

Gene	No. of sites	No. of variable	Parsimony informative sites (Pi)	No. of conserved sites	Average base frequencies (%) AT
matK	761	624	521	178	65.41
rbcL	455	104	81	351	53.57

These findings suggest that *matK* generally exhibits higher sequence variability, making it more suitable for species-level identification and phylogenetic studies. In contrast, *rbcL* tends to be more conserved, which can be advantageous for broader taxonomic identifications. However, the effectiveness of these genes as DNA barcodes can vary depending on the plant group under study, and in some cases, *rbcL* may provide better discrimination (Hollingsworth *et al*. 2011).

Maximum likelihood (ML) phylogenetic analyses based on *matK* and *rbcL* gene sequences revealed congruent topologies and consistent evolutionary relationships among the thirteen BLH plant species (Figs 1 and 2). In the *matK*-based ML phylogeny, the thirteen species were resolved into two major clades (A and B), reflecting their evolutionary divergence. Clade A comprises *Passiflora foetida*, *Citrus aurantifolia*, *Calotropis gigantea*, *Litsea glutinosa*, *Magnolia champaca* and *Flacourtia indica*, all of which exhibit strong bootstrap support (99-100) for conspecific individuals sampled from different geographic regions (Fig. 1). Clade B includes *Aristolochia indica*, *Monoon longifolium*, *Holarrhena pubescens*, *Senna occidentalis*, *Mimosa pudica*, *Ziziphus oenopolia*, and *Glycosmis pentaphylla*, also showing high bootstrap support (98-100) at the species level (Fig. 1). These results indicate a high degree of sequence similarity in *matK* across geographically disparate specimens, thereby supporting its utility in accurate species identification. The high nodal support (bootstrap values typically 99-100) for species-level clades underscores the effectiveness of the *matK* marker in resolving both intra- and interspecific relationships, with implications for understanding ecological specialization among larval host plants **(**Teklemariam *et al*. 2023).

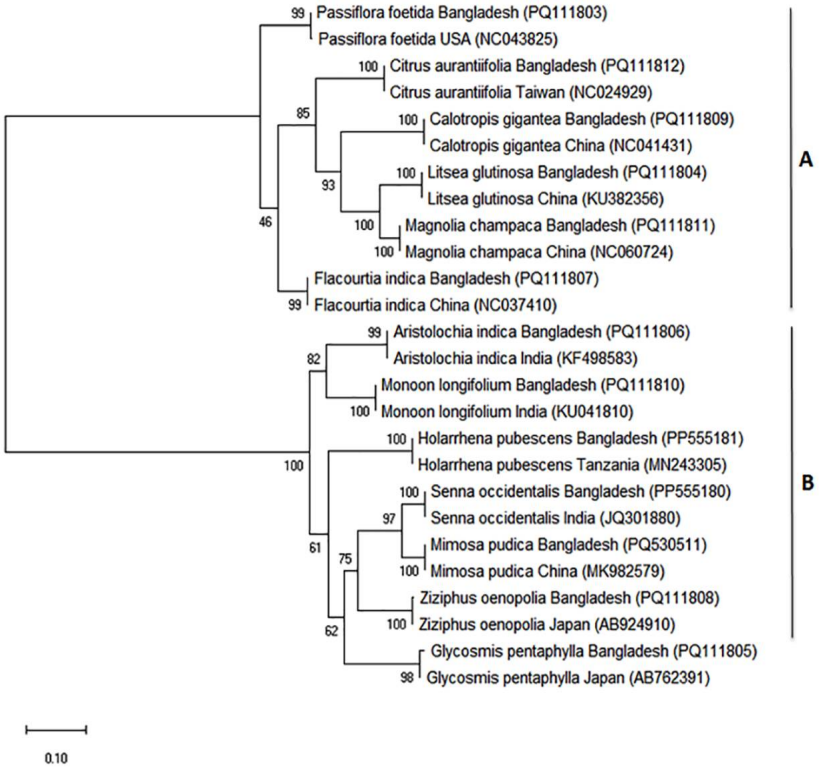

Fig. 1. Maximum likelihood (ML) tree showing inter-relationships among thirteen BLH plants using 1000 bootstrap replicates based on *matK* gene.

Similarly, the *rbcL*-based ML phylogeny provides a robust framework for resolving higher-level taxonomic relationships among the same set of species (Fig. 2). As with the *matK* tree, the *rbcL* phylogeny resolved two major clades (A and B), likely representing monophyletic lineages derived from distinct common ancestors. High bootstrap support (98-100) across clades confirms species identities and reveals strong genetic affinity between Bangladeshi specimens and reference sequences from Brazil, China, India, Indonesia and Taiwan (Fig. 2). While the *rbcL* gene shows slightly lower resolution at the species level compared to *matK*, it is informative for broader evolutionary relationships. Notably, Clade A encompasses members of the families Passifloraceae (*P. foetida*), Salicaceae (*F. indica*), Fabaceae (*S. occidentalis* and *M. pudica*), Rhamnaceae (*Z. oenopolia*), and Rutaceae (*G. pentaphylla* and *C. aurantiifolia*), whereas Clade B includes Apocynaceae (*H. pubescens* and *C. gigantea*), Lauraceae (*L. glutinosa*), Aristolochiaceae (*A. indica*), Magnoliaceae (*M. champaca*), and Annonaceae (*M. longifolium*). Although based on a single locus, this phylogenetic framework contributes valuable insights into the evolutionary history of BLH plant species and highlights the utility of molecular data in both species identification and the inference of macroevolutionary patterns.

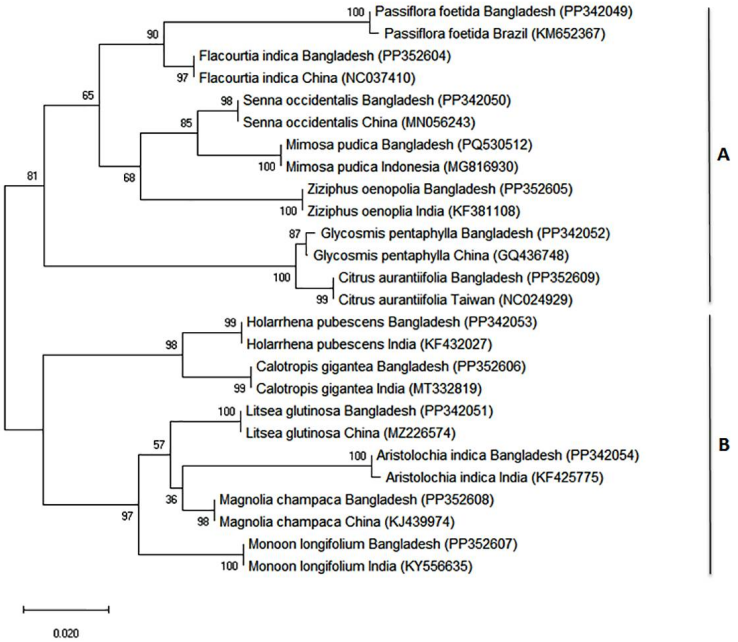

Fig. 2. Maximum likelihood (ML) tree showing inter-relationships among thirteen BLH plants using 1000 bootstrap replicates based on *rbcL* gene.

Analyzing both *matK* and *rbcL* phylogenies provides a robust framework for identifying the thirteen BLH plant species, as shown by the consistent, high-support clustering of conspecifics across global samples (Figs 1 and 2) (Teklemariam *et al.* 2023). This dual-marker approach combines *matK*'s finer resolution at lower taxonomic levels with *rbcL*'s strength in resolving deeper divergences (Li *et al*. 2011). Concordant tree topologies increase confidence in the evolutionary relationships, while minor discrepancies highlight areas for further sampling or additional markers. Overall, these results offer a strong molecular basis for understanding the identity and evolutionary history of BLH plant species.

This study highlights the importance of accurately identifying butterfly larval host plants to support conservation and ecological research. The use of DNA barcoding with *matK* and *rbcL* genes proved effective for BLH plant identification, with high sequence similarity (99.04-100%) and successful GenBank accession. Comparative analyses demonstrated the higher variability and phylogenetic resolution of *matK* at lower taxonomic levels, while *rbcL* provided conserved sequences valuable for higher-level classification. The complementary nature of these markers reinforces their utility in plant DNA barcoding and contributes to a better understanding of the evolutionary relationships and ecological significance of BLH plant species.

Acknowledgements

The authors are very grateful to the Wazed Miah Science Research Centre at Jahangirnagar University for providing the laboratory space. Additionally, they express their gratitude to Mr. Abdur Rahim of the Botany Department at JU for his assistance in the identification of plant species.

References

Burgess KS, Fazekas AJ, Kesanakurti PR, Graham SW, Husband BC and Newmaster SG 2011. Discriminating plant species in a local temperate flora using the *rbcL+matK* DNA barcode. Meth. Ecol. Evol. **2**: 333-340.

CBOL Plant Working Group 2009. A DNA barcode for land plants. Proc. Natl. Acad. Sci. USA. **4**: 106(31): 12794-7.

Chowdhury SH and Hossain M 2013. Butterflies of Bangladesh-A Pictorial Handbook. Skylark Printers, Dhaka, Bangladesh. pp. 210.

Das S, Ruhi M, Abedin MS and Hossain MM 2025. Checklist of butterfly larval host plants and their spatial distribution on the campus of Jahangirnagar University, Bangladesh. Asian-Aust. J. Bio. Biotech. **10**(1): 22-33.

Dennis RLH, Hardy PB and Shreeve TG 2008. The importance of resource databanks for conserving insects: a butterfly biology perspective. J. Insect Conserv. **12**: 711-719.

Hall TA 1999. Bioedit, a user-friendly biological sequences alignment editor and analysis program for windows 95/98/NT. Nucleic Acids Sym. Ser. **41**: 95-98.

Hollingsworth PM, Graham SW and Little DP 2011. Choosing and using a plant DNA barcode. PLoS ONE **6**(5): e19254.

Hossain MM 2023. A review on the diversity of butterfly (Insecta: Lepidoptera) fauna from Bangladesh. Bangladesh J. Zool. **51**(1): 03-34.

Huang X-C, Ci X-Q, Conran JG and Li J 2015. Application of DNA barcodes in asian tropical trees – A case study from Xishuangbanna nature reserve, southwest China. PLoS ONE **10**(6): e0129295.

IUCN Bangladesh 2015. Red List of Bangladesh. Butterflies. IUCN, international union for conservation of nature, Bangladesh country office, Dhaka, Bangladesh. **7**: 1-400.

Jurado-Rivera JA, Vogler AP, Reid CAM, Petitpierre E and Gomez-Zurita J 2009. DNA barcoding insect-host plant associations. Proc. Biol. Sci. **276**: 639-648.

Kimura M 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**: 111-120.

Kumar JUS, Krishna V, Seethapathy GS, Ganesan R, Ravikanth G and Shaanker RU 2018a. Assessment of adulteration in raw herbal trade of important medicinal plants of India using DNA barcoding. 3 Biotech. **8**:135.

Kumar S, Stecher G, Li M, Knyaz C and Tamura K 2018b. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. **35**: 1547-1549.

Larsen TB 2004. Butterflies of Bangladesh: an annotated checklist: IUCN, the world conservation union, Bangladesh country office. pp. 158.

Li DZ, Gao LM, Li HT, Wang H, Ge XJ and Liu JQ 2011. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. Proc. Natl. Acad. Sci. USA **108**(49): 19641-6.

Munoz-Galicia D, Lara C, Castillo-Guevara C, Cuautle M and Rodríguez-Flores C 2023. Impacts of land use change on native plant-butterfly interaction networks from central Mexico. Peer J. **11**: e16205.

Rajphriyadharshini R and Weerasena OVDSJ 2020. DNA barcoding of medicinal plant: a systemic review. Intl. J. Pharma. Sci. Inven. **9**(6): 6-16.

Shihan TR 2018. Records of new larval host plants of some common butterflies of Bangladesh. J. Threat. Taxa. **10**(2): 11303-11311.

Teklemariam DM, Gailing O, Siregar IZ, Amandita FY and Moura CCM 2023. Evaluating the effectiveness of rbcL and matK markers for species identification in Burseraceae. Ecol. Evol. **13**(3): e9935.

Tiple AD, Khurad AM and Dennis RLH 2011. Butterfly larval host plant use in a tropical urban context: Life history associations, herbivory, and landscape factors. J. Insect Sci. **11**(65): 1-13.

(Manuscript received on 20 August, 2025; revised on 29 November, 2025)