Asian Australas. J. Biosci. Biotechnol. 2023, 8 (3), 49-55; https://doi.org/10.3329/aajbb.v8i3.68609

Asian-Australasian Journal of Bioscience and Biotechnology
ISSN 2414-1283 (Print) 2414-6293 (Online)
https://www.ebupress.com/journal/aajbb/

Article

DNA barcoding and phylogenetic relationships of ten butterfly caterpillars

Surma Mohiudden Meem, Kawsari Akter, Md Abdullah Al Mamun, Sijad Islam, Fahmina Sarkar Borsha, Muhammad Sohel Abedin and Md. Monwar Hossain*

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

*Corresponding author: Professor Dr. Md. Monwar Hossain, DNA Barcoding Laboratory, Department of Zoology, Jahangirnagar University, Savar, Dhaka-1342, Bangladesh. Phone: +8801926773511; E-mail: monwar@juniv.edu

Received: 02 September 2023/Accepted: 12 November 2023/Published: 16 November 2023

Abstract: Cytochrome C oxidase subunit I (COI), known as DNA barcodes, can be employed for the identification of an unknown specimen at the species level. The present study was aimed at performing COI gene-based identification of butterflies using butterfly caterpillars. Consecutively, ten caterpillars of ten different butterfly species from the families Papilionidae, Nymphalidae, Pieridae, Lycaenidae, and Hesperiidae were taken from respective host plants and used to generate COI gene sequences that were about 650 bp long. After BLAST analysis, the sequenced gene revealed 96–100% similarity to the same species from different regions. Then, sequences were submitted to NCBI's GenBank and obtained ten accession numbers. In order to elucidate genetic diversity and evolutionary relationships among ten species of butterfly caterpillars, pairwise distance analysis and the construction of a phylogenetic tree were performed using MEGA10 and BioEdit software. In the analysis, the interspecific genetic divergence among the caterpillars of butterflies was between 0.101-0.164%. A phylogenetic tree was constructed with the assistance of the Neighbor-Joining (NJ) algorithm, which identified two major clades, A and B, as well as indicated that butterflies caterpillars shared a common ancestor. All the species are included in these two clades, except *Catopsilia pomona*. This species is under the family Pieridae, and the phylogenetic position of Pieridae compared to other butterfly families is ambiguous, necessitating additional research to resolve this issue. The study demonstrated the general applicability of DNA barcoding for rapidly and accurately distinguishing butterfly species, even when using larval stages, as in the present study. Furthermore, it may disclose a higher taxonomic hierarchy of butterfly families.

Keywords: molecular identification; *COI* gene; butterfly; caterpillar; Bangladesh

1. Introduction

Butterflies are recognized as highly valuable pollinators and play an indispensable role in the food chain. The significance of their contribution to pollination and their status as indicators of ecosystem health render them valuable contributors to the operation and balance of diverse ecosystems. Butterflies substantial interactions with plants through herbivory and pollination account for their diversity and abundance (Hudewenz *et al*., 2012). In particular, butterflies are connected to certain plant species because they use those plants as host plants for their caterpillars (some larvae may only eat one species of plant) or as food sources for adults of different species. However, for conducting biodiversity censuses and understanding ecosystem health as well as the

overall conservation of butterflies, there is a need for precise species identification, which is accomplished by adult specimens. If the adult is not available, identification may proceed with immature stages like caterpillars or pupae. Conventionally, the majority of insect larvae are identified through the process of nurturing them to maturity and subsequently examining the morphology of the adult. This conventional technique is subject to numerous methodological constraints, which can at times be rather frustrating. In addition to the issue of identification, another obstacle known as the 'taxonomic impediment' exists, referring to unresolved taxonomy and cryptic diversity. This impediment often poses challenges in accurately analysing data (de Carvalho *et al*., 2007; Gossner and Hausmann, 2009). With these background, numerous countries have utilized DNA barcoding (COI gene) to identify lepidopteran larvae and other insects at the molecular level (Miller *et al*., 2007; Gossner *et al*., 2009; Hausmann *et al*., 2016), allowing for the simple, inexpensive, and speedy identification of larvae collected from their host plants. This form of work is limited in the Indian subcontinent, including Bangladesh (Daravath *et al*., 2015). Identification by DNA barcoding is feasible, even in the case of desiccated pupal exuviae subsequent to moulting and empty pupal skins following butterfly hatching (Lees *et al*., 2011; Hausmann *et al*., 2020). Recently, the molecular identification of butterfly species has been carried out with the eggs of butterflies in Bangladesh as well (Hossain *et al*., 2022a).

IUCN-Bangladesh recorded and evaluated the threatened category of 305 species of butterflies in Bangladesh (IUCN Bangladesh, 2015). More than 150 species have been DNA barcoded from adult butterflies in Bangladesh, and the other species are being processed provided that adult samples are available (Ghosh *et al*., 2019; Hossain *et al*., 2022b; Akter *et al*., 2023). Researchers encounter caterpillars, pupae, and even eggs in the forest. Consequently, using the COI gene (molecular marker) of these immature stages, there are numerous opportunities for identifying the unknown adult butterflies. Therefore, the present study supports species-level molecular identification using butterfly caterpillars and also reveals evolutionary relationships among butterfly families.

2. Materials and Methods

2.1. Ethical statement

This study followed all the rules and instructions for ethics set by the Biosafety, Biosecurity, and Ethical Clearance Committee at Jahangirnagar University in Savar, Dhaka, Bangladesh, with approval number BBEC, JU/M 2023/11(70).

2.2. Sample collection

Ten caterpillars of ten different butterfly species were collected from different areas of the Jahangirnagar University campus in Bangladesh (Figure 1, 2). The specimens were manually collected by carefully selecting them from their respective host plants in the wild. Preliminary identification of butterfly caterpillars was carried out on the basis of available literature (Karmakar *et al*., 2018; Shahroni *et al*., 2022) and utilizing the online resource "Butterflies of India." (https://www.ifoundbutterflies.org/).

2.3. DNA extraction, amplification and sequencing

Following the instructions provided in the Wizard Genomic@ DNA Purification Kit by Promega (located in Madison, Wisconsin, USA), the genomic DNA of these butterfly caterpillars was extracted from body parts in line with the procedure. Primers LCO 1490 (5'-GGTCAACAAATCATAAAGATATTGG-3') and HCO 2198 (5' TAAACTTCAGGGTGACCAAAAAATCA-3') were used in PCR in order to amplify the mitochondrial COI gene region. 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. The first denaturation step was carried out at a temperature of 95°C for duration of 4 minutes. After this, 35 cycles were executed, with each cycle comprising 30 seconds of primer denaturation at 95°C, primer annealing at 49°C, and primer extension at 72°C. The process concluded with a 5 minutes extension at 72°C. 1% agarose gel electrophoresis was used to assess the PCR-amplified result under ultraviolet light (Bio Analyzer). The ABI 3500 Sequencer performed the sequence of the amplification products.

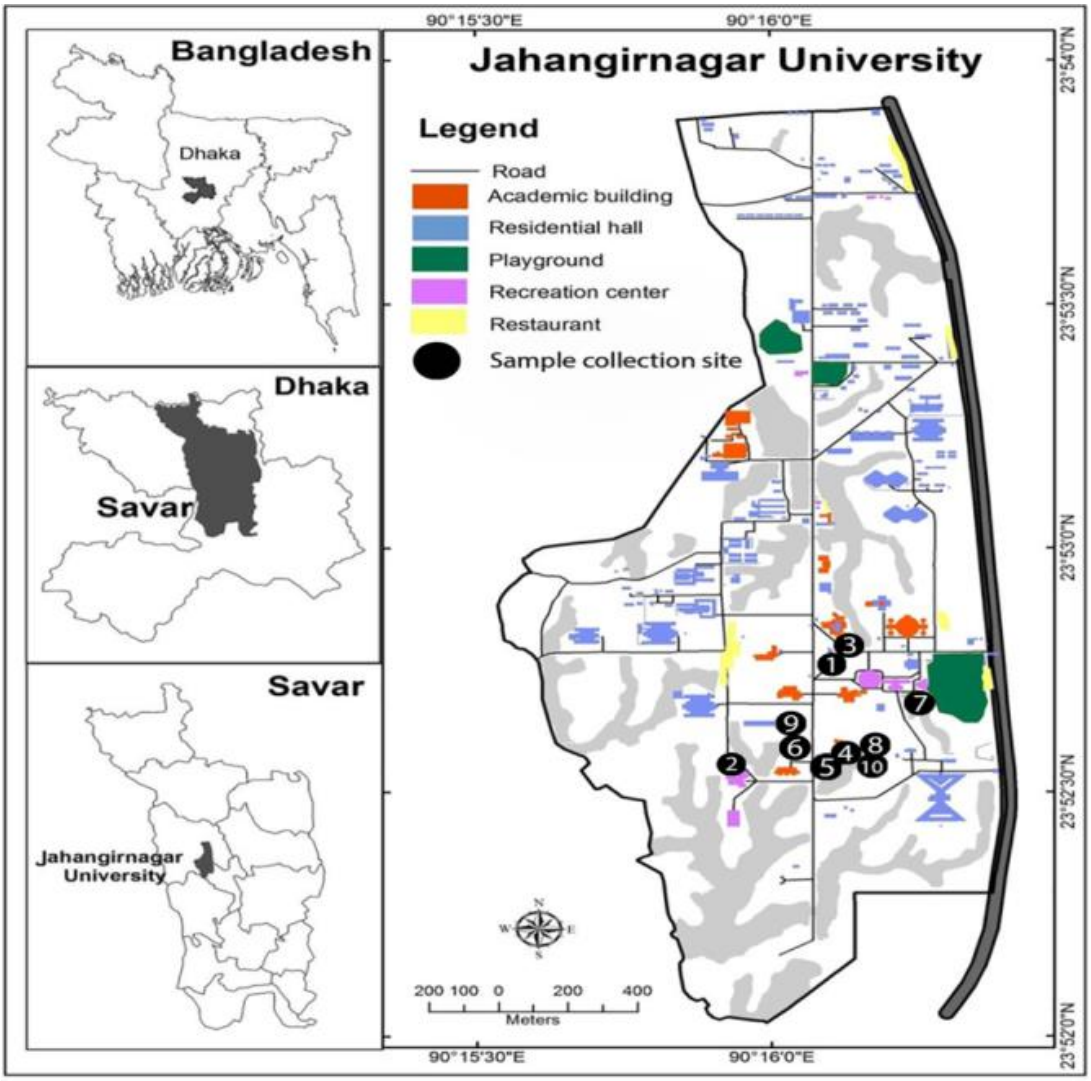

Figure 1. The sample collection site at Jahangirnagar University, Bangladesh (locations 1-10 denote compliance with the species serials specified in Table 1).

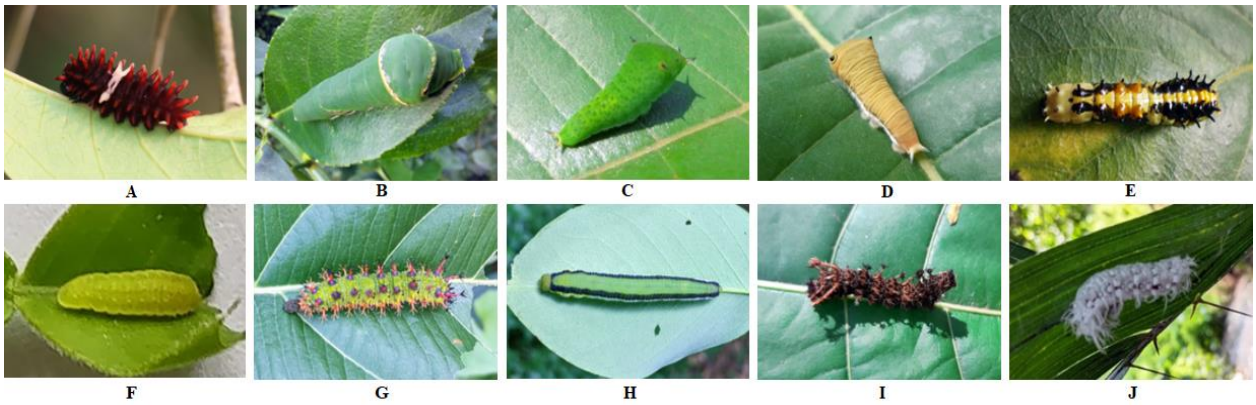

Figure 2. Caterpillars of butterflies used for genomic DNA extraction in the investigation. A. *Pachliopta aristolochiae*, B. *Papilio polytes*, C. *Graphium agamemnon*, D. *Graphium doson*, E. *Papilio clytia*, F. *Chilades lajus*, G. *Athyma perius*, H. *Catopsilia pomona*, I. *Moduza procris*, J. *Gangara thyrsis*.

2.4. Phylogenetic analysis

All of the sequences of these ten species of butterfly caterpillars were edited using Chromas 2.6.2. ClustalW, a multiple alignment tool included in BioEdit version 7.0 (Hall, 1999), was used to align the assembled

sequences. The Kimura 2-Parameter (K2P) model implemented in the software package MEGA10 was used to compute and summarize nucleotide compositions and to estimate pairwise distances (Kimura, 1980; Kumar *et al*., 2018). MEGA10 was used to generate a phylogenetic tree following the K2P model and 1000 NJ bootstrap replicates (Kumar et al., 2018). As an outgroup, *Orthetrum sabina* (MF784360) was used in phylogenetic analysis. Additional genes utilised in the analysis were obtained from the GenBank database (Figure 3).

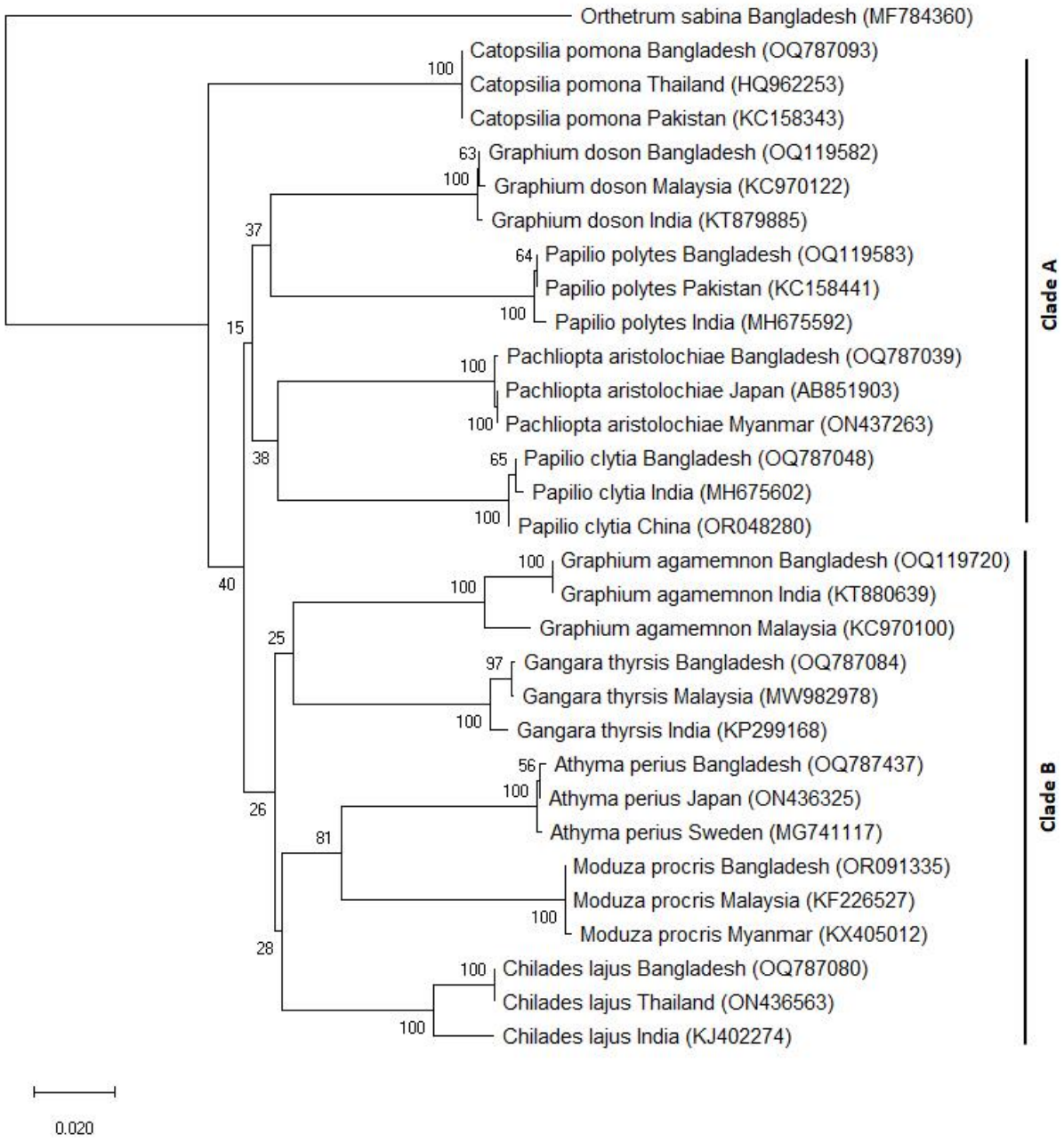

Figure 3. **The Neighbor-Joining tree of the ten butterfly caterpillars is based on partial sequences of the mitochondrial COI gene. *O. sabina* was used as the out group.**

3. Results and Discussion

The current investigation involved the generation of mitochondrial COI gene sequences (mtCOI) for ten butterfly caterpillars belonging to five different families, namely Papilionidae, Nymphalidae, Pieridae, Lycaenidae, and Hesperiidae. The sequenced gene was compared to the available sequences in the NCBI's GenBank, and the BLAST analysis revealed remarkable similarities (96-100%) between the same species from different geographic locations. Then, sequences were submitted to NCBI's GenBank and obtained ten accession numbers (Table 1).

Table 1. Butterfly caterpillar species and GenBank accession numbers from which COI genes were sequenced.

Species name	Family	GPS coordinates	Date of collection	Voucher no.	GenBank accession no.
1. *Graphium doson*	Papilionidae	23.87933 N 90.26814 E	22.08.22	BFBSV320	OQ119582
2. *Graphium agamemnon*	Papilionidae	23.87547 N 90.26606 E	24.08.22	BFBSV322	OQ119720
3. *Papilio polytes*	Papilionidae	23.87942 N 90.26825 E	22.08.22	BFBSV321	OQ119583
4. *Papilio clytia*	Papilionidae	23.87586 N 90.26844 E	04.09.22	BBV-CP225	OQ787048
5. *Pachliopta aristolochiae*	Papilionidae	23.87578 N 90.26831 E	04.09.22	BBV-CP235	OQ787039
6. *Athyma perius*	Nymphalidae	23.87644 N 90.26767 E	04.09.22	BBV-CP182	OQ787437
7. *Moduza procris*	Nymphalidae	23.87847 N 90.27078 E	08.09.22	BBV-CP169	OR091335
8. *Catopsilia pomona*	Pieridae	23.87622 N 90.26850E	24.08.22	BBV-CP258	OQ787093
9. *Chilades lajus*	Lycaenidae	23.87656 N 90.26758 E	04.09.22	BBV-CP108	OQ787080
10. *Gangara thyrsis*	Hesperiidae	23.87617 N 90.26875 E	30.08.22	BBV-CP058	OQ787084

3.1. Genetic distance analysis

Based on the K2P model, the nucleotide divergence values of ten species ranged from 0.101% to 0.164%. A distance of 0.101 was found between the caterpillars of *Athyma perius* and *Moduza procris*, indicating that the genetic composition of these two species is the most similar. The highest distance, 0.164, was found between *Papilio polytes* and *Moduza procris* caterpillars (Table 2). In the case of *Papilio*, it has already been reported to have sequence divergences ranging from 0% to 1.2% (Zakharov *et al*., 2004). The restricted range of interspecific divergence could be the result of rare interbreeding between species (Win *et al*., 2015). Consequently, their genetic uniqueness is maintained, and they are not exchanging genetic material through hybridization. In addition, genetic distance analysis provides researchers with a note that enables them to determine how the genes of organisms are related to one another and to draw conclusions regarding their evolution and genetic diversity (Mackintosh *et al*., 2019).

Table 2. Interspecific genetic distance among caterpillars of ten butterfly species.

Species name	1	2	3	4	5	6	7	8	9	10
1. *Graphium doson*										
2. *Graphium agamemnon*	0.113									
3. *Papilio polytes*	0.122	0.133								
4. *Pachliopta aristolochiae*	0.113	0.115	0.127							
5. *Papilio clytia*	0.138	0.138	0.115	0.122						
6. *Athyma perius*	0.125	0.140	0.140	0.136	0.144					
7. *Catopsilia pomona*	0.129	0.134	0.138	0.125	0.134	0.125				
8. *Chilades lajus*	0.123	0.113	0.132	0.132	0.138	0.125	0.151			
9. *Gangara thyrsis*	0.120	0.111	0.145	0.134	0.122	0.132	0.138	0.122		
10. *Moduza procris*	0.138	0.129	0.164	0.149	0.162	0.101	0.143	0.140	0.132	-

3.2. Phylogenetic analysis

The Neighbor-Joining phylogenetic tree using the *COI* sequences of the ten butterfly caterpillars showed monophyletic entities that shared a common ancestor (Figure 3). The phylogenetic tree revealed two main clades, A and B. Clade A consisted of four species that included *Graphium doson, Papilio polytes, Pachliopta aristolochiae* and *Papilio clytia.* While Clade B consisted of *Graphium agamemnon, Gangara thyrsis, Athyma perius, Moduza procris* and *Chilades lajus.* On the other hand, *Catopsilia pomona is* situated distantly compared to Clades A and B. *C. pomona* is a member of the Pieridae that has worldwide distribution and contains

approximately 1100 species (Ackery *et al*., 1999; Vane-Wright, 2003). The precise phylogenetic relationship between Pieridae and the remaining butterfly families remains ambiguous (Robbins, 1988; Vane-Wright, 2003; Wahlberg *et al*., 2005). This hypothesis is consistent with our results, as *C. pomona* is outside of the two clades, where these two clades consist of species of Papilionidae, Nymphalidae, Lycaenidae, and Hesperiidae (Figure 3). It is believed that the Pieridae is the sister family of the Papilionidae (Scott, 1985) or, most probably, sister to Nymphalidae + (Riodinidae + Lycaenidae) (de Jong *et al*., 1996; Weller *et al*., 1996; Ackery *et al*., 1999; Wahlberg *et al*., 2005). These results are consistent with our findings, despite the fact that we were unable to include Riodinidae and added Hesperiidae species in the current investigation. Nevertheless, further investigation is required, utilizing extensive data, in order to examine the unequivocal evolutionary connections among the various families.

4. Conclusions

The COI gene was sequenced from larvae of ten butterfly species in this work. After BLAST analysis, gene sequences submitted to NCBI's GenBank received ten unique accession numbers. Pairwise distance analysis and phylogenetic trees were used to evaluate ten butterfly caterpillar species' evolutionary relationships and genetic diversity. The results showed that butterfly caterpillars have little interspecific genetic diversity. NJ-assisted phylogenetic analysis revealed two major clades, A and B, and showed that all butterfly caterpillars have a common ancestor. The research highlighted the broad utility of DNA barcoding in efficiently and precisely differentiating butterfly species, even when employing larval stages, as observed in the current investigation. Moreover, it has the potential to reveal a more comprehensive taxonomic classification of butterfly families.

Acknowledgements

The authors express their gratitude for the financial assistance provided by Jahangirnagar University Research Grants (Fiscal Year 2022-23). Sincere appreciation is also extended to the Wazed Miah Science Research Center at Jahangirnagar University, Bangladesh for providing the laboratory space.

Data availability

All relevant information is included in the manuscript.

Conflict of interest

None to declare.

Authors' contribution

Kawsari Akter, Sijad Islam, and Md. Abdullah Al Mamun were responsible for collecting the sample and preparing voucher specimens. The experiments were conducted by Surma Mohiudden Meem. The experiment was developed by Fahmina Sarkar Borsha, while the data analysis and initial text drafting were performed by Muhammad Sohel Abedin. Md. Monwar Hossain provided supervision and made some edits to the final manuscript. The final manuscript has reviewed and endorsed by all authors.

References

Ackery PR, R de Jong and RI Vane-Wright, 1999. The Butterflies: Hedyloidea, Hesperioidea and Papilionoidea, In: Kristensen (ed.), Lepidoptera, Moths and Butterflies, Volume 1: Evolution, Systematics and Biogeography, Walter de Gruyter, Berlin and New York, 263-300.

Akter K, SM Meem, S Islam, MAA Mamun, MS Abedin, FS Borsha, A Ghosh and MM Hossain, 2023. Utilizing COI gene for the Identification of thirteen Hesperiidae butterflies and determining their genetic relationship. J. Biodivers. Conserv. Bioresour. Manag., 9: 21-28 (In press).

Daravath SS, M Siddaiah and BR Naik, 2015. Molecular characterization and phylogenetic analysis of *Culex quinquefasciatus* by DNA barcoding. Adv. Entomol., 3: 118-124.

de Carvalho MR, FA Bockmann, DS Amorim, CRF Brandão, M de Vivo, JL de Figueiredo, HA Britski, MCC de Pinna, NA Menezes, FPL Marques, N Papavero, EM Cancello, JV Crisci, JD McEachran, RC Schelly, JG Lundberg, AC Gill, R Britz, QD Wheeler, MLJ Stiassny, LR Parenti, LM Page, WC Wheeler, J Faivovich, RP Vari, L Grande, CJ Humphries, R DeSalle, MC Ebach and GJ Nelson, 2007. Taxonomic impediment or impediment to taxonomy? A commentary on systematics and the cybertaxonomic-automation paradigm. Evol. Biol., 34: 140-143.

de Jong R, RI Vane-Wright and PR Ackery, 1996. The higher classification of butterflies (Lepidoptera): problems and prospects. Entomol. Scand., 27: 65-101.

Ghosh A, MS Abedin, AJ Howlader and MM Hossain, 2019. Molecular identification and phylogenetic relationships of seven Satyrinae butterflies in Bangladesh using Cytochrome c oxidase subunit I gene. Jahangirnagar Univ. J. Biol. Sci., 8: 67-74.

Gossner MM and A Hausmann, 2009. DNA barcoding enables the identification of caterpillars feeding on native and alien oak (Lepidoptera:Geometridae). Mitt. Munch Ent. Ges., 99: 135-140.

Hall TA, 1999. Bioedit, a user-friendly biological sequences alignment editor and analysis program for windows95/98/NT. Nucleic Acids Sym. Ser., 41: 95-98.

Hausmann A and S Scalercio, 2016. Host-plant relationships of 29 Mediterranean Lepidoptera species in forested ecosystems unveiled by DNA Barcoding. SHILAP Soc. Hispano Luso. Am. Lepid., 44: 463-471.

Hausmann A, J Diller, J Moriniere, A Hocherl, A Floren and G Haszprunar, 2020. DNA barcoding of fogged caterpillars in Peru: A novel approach for unveiling host-plant relationships of tropical moths (Insecta, Lepidoptera). PLoS ONE., 15: e0224188

Hossain MM, K Akter, SM Meem, S Islam, MAA Mamun, MS Abedin and FS Borsha, 2022a. Papilio polytes voucher BFBSV237 (egg sample) cytochrome c oxidase subunit I (COX1) gene submission at NCBI Genbank (https://www.ncbi.nlm.nih.gov/nuccore/2413839588).

Hossain MM, A Ghosh, FS Borsha, MK Hasan, MS Abedin and AJ Howlader, 2022b. Molecular phylogeny of swallowtail butterflies (Lepidoptera: Papilionidae) based on mitochondrial cytochrome c oxidase I (COI) gene from Bangladesh. Nepalese J. Zool., 6: 1-6.

Hudewenz A, A Klein, C Scherber, L Stanke, T Tscharntke, A Vogel, A Weigelt, W W Weisser and A Ebeling, 2012. Herbivore and pollinator responses to grassland management intensity along experimental changes in plant species richness. Biol. Conserv., 150: 42-52.

IUCN Bangladesh, 2015. Red List of Bangladesh. Butterflies. IUCN, International Union for Conservation of Nature, Bangladesh Country Office, Dhaka, Bangladesh. 7: 1-400.

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 S, G Stecher, M Li, C Knyaz and K Tamura, 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol. Biol. Evol., 35: 1547-1549.

Karmakar T, R Nitn, V Sarkar, S Baidya, S Mazumder, VK Chandrasekharan, R Das, GSG Kumar, S Lokhande, J Veino, L Veino, R Veino, Z Mirza, RV Sanap, B Sarkar and K Kunte, 2018. Early stages and larval host plants of some northeastern Indian butterflies. J. Threat. Taxa., 10: 11780-11799.

Lees DC, HW Lack, R Rougerie, A Hernandez-Lopez, T Raus, ND Avtzis, S Augustin and C Lopez-Vaamonde, 2011. Tracking origins of invasive herbivores through herbaria and archival DNA: the case of the horse-chestnut leaf miner. Front. Ecol. Environ., 9: 322–328.

Mackintosh A, DR Laetsch, A Hayward, B Charlesworth, M Waterfall, R Vila and K Lohse, 2019. The determinants of genetic diversity in butterflies. Nat. Commun., 10: 3466.

Miller MA, GC Muller, VD Kravchenko, A Junnila, KK Vernon, CD Matheson and A Hausmann, 2007. DNA-based identification of Lepidoptera larvae and plant meals from their gut content. Russ. Entomol. J., 15: 427-432.

Robbins RK, 1988. Comparative morphology of the butterfly foreleg coxa and trocanter (Lepidoptera) and its systematic implications. Proc. Entomol Soc. Wash., 90: 133-154.

Scott JA, 1985. The phylogeny of butterflies (Papilionoidea and Hesperioidea). J. Res. Lepid., 23: 241-281.

Shahroni AM, FIA Shidiq, Soenarko, L Irwansyah and D Peggie, 2022. Record on some Caterpillars of Butterflies (Lepidoptera: Papilionoidea) in East Java, Indonesia. Treubia, 49: 25-40.

Vane-Wright RI, 2003. Evidence and identity in butterfly systematics.In: Boggs CL, Watt WB, Ehrlich PR, eds. Butterflies: ecology and evolution taking flight. Chicago: University of Chicago Press, 477-513.

Wahlberg N, MF Braby, AVZ Brower, R de Jong, M Lee, S Nylin, NE Pierce, FA Sperling, R Vila, AD Warren and E Zakharov, 2005. Synergistic effects of combining morphological and molecular data in resolving the phylogeny of butterflies and skippers. Proc. R. Soc. B. Biol. Sci., 272: 1577-1586.

Weller SJ, DP Pashley and JA Martin, 1996. Reassessment of butterfly family relationships using independent genes and morphology. Ann. Entomol. Soc. Am., 89: 184-192.

Win NZ, EY Choi, D-JJang, J Park and JK Park, 2015. Molecular comparison of the genus Junonia (Lepidoptera: Nymphalidae) in Myanmar. J. Asia Pac. Biodivers., 8: 287-294.

Zakharov EV, MS Caterino and FAH Sperling, 2004. Molecular phylogeny, historical biogeography, and divergence time estimates for swallowtail butterflies of the genus Papilio (Lepidoptera: Papilionidae). Syst. Biol., 53: 193-215.