*Bangladesh J. Zool.* ***53***(2): 149-157, 2025 ***Article***
ISSN: 0304-9027
eISSN: 2408-8455

# COI GENE–BASED MOLECULAR IDENTIFICATION OF NINETEEN BUTTERFLY CATERPILLARS (ORDER: LEPIDOPTERA) OF BANGLADESH

Shraboni Das, Nur A Afsari, Surma Mohiudden Meem, Sijad Islam, Kawsari Akter, Md. Abdullah Al Mamun and Md. Monwar Hossain*

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

**ABSTRACT:** The application of mitochondrial Cytochrome c oxidase subunit I (COI) gene-based molecular techniques for species identification and phylogenetic analysis offers distinct advantages when applied to butterfly caterpillars (larvae), compared to traditional methods reliant on adult morphology. In this study, we successfully amplified and sequenced an average of 626 base pairs of the COI gene from caterpillars representing nineteen different butterfly species. Sequence similarity searches using BLAST confirmed high matches with corresponding species entries in the GenBank database, validating the accuracy of our molecular identifications. All sequences were subsequently submitted to GenBank and assigned accession numbers. Phylogenetic analysis revealed two well-supported clades (Clade A and Clade B), with high bootstrap values (81–99), indicating distinct evolutionary lineages and a shared common ancestor among the species. Sequences clustered with conspecific data from other regions, reinforcing the reliability of COI-based phylogenetics. This study underscores the effectiveness of DNA barcoding as a non-invasive and accurate method for species identification across life stages, supporting advancements in taxonomy, biodiversity monitoring, and conservation biology, particularly when adult specimens are unavailable or lack distinguishing morphological features.

***Key words****:* DNA barcoding, Butterfly caterpillars, Phylogenetic analysis, Species identification

## INTRODUCTION

Butterfly caterpillars play a vital role in the ecology of forests, serving as an essential link in the food web and contributing significantly to the health and biodiversity of forest ecosystems. As herbivores, caterpillars are primary consumers that feed on a wide variety of plant species, including many trees and shrubs that are crucial to forest structure and regeneration (Wagner and Hoyt 2022). Through their feeding, caterpillars help control plant populations and

*Author for correspondence: <monwar@juniv.edu>
 DOI: https://doi.org/10.3329/bjz.v53i2.85041

influence plant community dynamics. Additionally, caterpillars are a key food source for a variety of predators, such as birds, amphibians, and other insects, thus supporting higher trophic levels in the ecosystem (Sinha and Dutta 2024). Furthermore, butterfly caterpillars are integral to pollination networks, as they eventually transform into adult butterflies, which are important pollinators of numerous flowering plants. The presence of healthy butterfly populations in a forest is often indicative of a thriving ecosystem, making the study and conservation of these species vital for forest sustainability (Aluri 2022).

However, identifying butterfly species based on their caterpillars presents significant challenges. Unlike adult butterflies, which can be easier to identify due to their distinct wing patterns and colors, caterpillars often exhibit subtle morphological variations that make visual identification difficult (Meem *et al.* 2023). In many cases, species of caterpillars that look similar may belong to entirely different butterfly species. This makes proper identification essential, not only for understanding the ecology of butterfly populations but also for effective conservation efforts, as different species may have specific habitat, food plant, and environmental requirements (Joan *et al.* 2022). Without accurate identification, it becomes difficult to assess species diversity, population status or implement appropriate conservation strategies.

In recent years, COI (Cytochrome Oxidase I) gene-based molecular identification has emerged as a powerful tool for overcoming these challenges (Hebert *et al.* 2003, Hossain 2022, Abedin *et al.* 2023, Meem *et al.* 2023, Islam *et al.* 2024). The COI gene, commonly used in DNA barcoding, provides a unique genetic fingerprint for species identification that is reliable, even for immature life stages like caterpillars (Hausmann *et al.* 2020, Meem *et al.* 2023). Unlike traditional morphological methods, which rely on visible traits that may be hard to discern in caterpillars, molecular techniques offer a non-invasive and precise means of identifying species from small or fragmented body parts (Islam *et al.* 2023). Moreover, COI gene based identification can provide clear results even when the caterpillar is in an early or damaged stage of development, ensuring accurate identification that might otherwise be overlooked (Meem *et al.* 2023). Therefore, by analyzing the mitochondrial COI gene of nineteen different butterfly caterpillars, we aimed to assess the efficiency of molecular identification for this often-overlooked life stage and to establish a reliable genetic database for Lepidoptera insects. The findings of this study will contribute to improving our understanding of butterfly biodiversity, particularly within forest ecosystems.

## MATERIAL AND METHODS

*Sample collection*: A total of nineteen butterfly caterpillars were collected from different locations of Bangladesh, using sweep nets and handpicking from their habitat (Fig. 1, Table 1). Only one species, molting cage of caterpillar was used (Fig. 1, Q). The preliminary identification of the butterfly caterpillars was conducted using the existing literatures and reference articles (Bingham and Blanford 1905, Karmakar *et al.* 2018, Shahroni *et al.* 2022) and Foundation of Indian Butterflies (https://www.ifoundbutterflies.org/).

*DNA extraction, amplification and sequencing:* Genomic DNA was extracted from the body segments of nineteen butterfly caterpillars following the Wizard Genomic DNA Purification Kit methodology (Promega, Madison, WI, USA). PCR was utilised to amplify the mitochondrial cytochrome c oxidase I (COI) gene region with the designated primers LCO 1490 (5′-GGTCAACAAATCATAAAGATATTGG-3′) and HCO 2198 (5′ TAAACTTCAGGGTGACCAAAAAATCA-3′). In a thermal cycler (Veriti, USA), PCR was carried out in 20 µL of Q2 Green PCR Master Mix. The cycling conditions included an initial denaturation at 95°C for 4 minutes, succeeded by 35 cycles comprising denaturation at 95°C for 30 seconds, primer annealing at 49°C for 30 seconds, and primer extension at 72°C for 45 seconds. The concluding extension phase was conducted at 72°C for a duration of 5 minutes. The PCR-amplified product was assessed via 1% agarose gel electrophoresis under UV illumination (Bio Analyser). The amplification product was sequenced utilising the ABI 3500 Sequencer.

*Phylogenetic analysis*: The COI gene sequences of these nineteen butterfly species underwent modifications utilising Chromas version 2.6.2. The ClustalW multiple alignment algorithm was employed to align the assembled sequences, utilising BioEdit version 7.0 (Hall 1999). The estimation and assessment of nucleotide compositions were conducted utilising the Kimura 2 Parameter (K2P) model through the MEGA10 program (Kimura 1980, Kumar *et al.* 2018). The phylogenetic tree was constructed utilising MEGA10 through Maximum Likelihood (ML) methods, incorporating 1000 bootstrap replications (Saitou and Nei 1987, Kumar *et al.* 2018).

## RESULTS AND DISCUSSION

In the present study, nineteen butterfly caterpillars from nine distinct families were used to generate mitochondrial COI (mtCOI) gene sequences, namely Acraeidae, Amathusiidae, Danaidae, Lycaenidae, Nymphalidae, Papilionidae, Pieridae, Riodinidae, and Satyridae (Fig. 1; Table 1). The gene we sequenced was compared to other sequences in NCBI's GenBank, and the BLAST analysis showed that there were strong similarities (97.34-100%) among the same species found in different places. Subsequently, all sequences were submitted to NCBI's GenBank, resulting in the acquisition of nineteen accession numbers (Table 1).

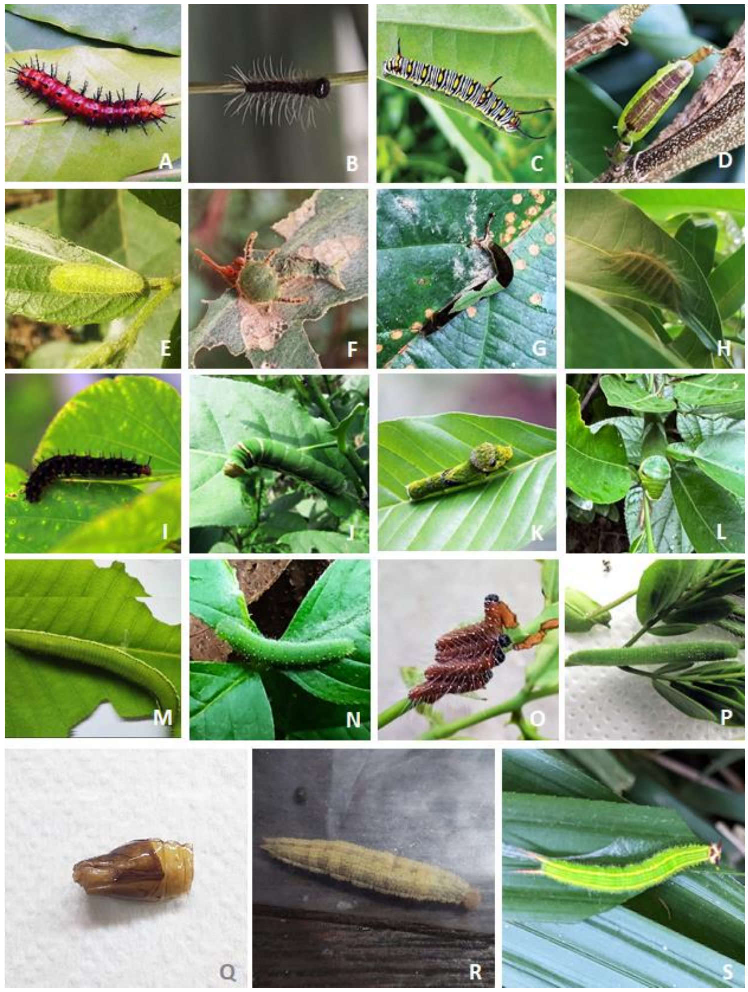


Fig.1. Butterfly caterpillars were used to get genomic DNA for the study. (A) *Acraea violae*, (B) *Discophora sondaica*, (C*) Danaus chrysippus*, (D) *Anthene emolus*, (E) *Castalius rosimon*, (F) *Charaxes bernardus*, (G) *Stibochiona nicea*, (H) *Euthalia aconthea*, (I) *Hypolimnas bolina*, (J) *Papilio demoleus*, (K) *Papilio nephelus*, (L) *Papilio memnon*, (M) *Catopsilia pyranthe*, (N) *Leptosia nina*, (O) *Delias eucharis*, (P) *Eurema hecabe*, (Q) *Zemeros flegya*, (R) *Ypthima baldus*, and (S) *Elymnias hypermnestra.*

*Nucleotide analysis*: In the present study, the average nucleotide composition of the sequenced COI gene of nineteen caterpillars was as follows: 30.6% A, 39.2% T, 14.1% G, 16.1% C, and the average total length was 626 bp. All nineteen sequences analyzed exhibited a pronounced A+T bias, with adenine and thymine comprising 69.8% of the total nucleotide composition. Such A+T richness is a well-documented feature of insect mitochondrial genomes, reflecting the mutational and structural constraints typical of mitochondrial DNA (Cameron 2014). Consistently, high A+T content has also been reported in the mitochondrial genomes of various Lepidopteran species, supporting the patterns observed in our study (Foster *et al.* 2010, Nie *et al.* 2018).

**Table 1. List of nineteen butterflies' caterpillars, their geo-location, voucher and Genbank accession numbers**

| Species Name | Family | Geo Location | Voucher No. | Accession No. |
|---|---|---|---|---|
| *Acraea violae* | Acraeidae | 23°52'39.5"N 90°16'07.1"E | BBV-CP001 | PP732660 |
| *Discophora sondaica* | Amathusiidae | 23°52'21.4"N 90°16'01.3"E | BBV-CP002 | PV019352 |
| *Danaus chrysippus* | Danaidae | 23°59'15.8"N 90°39'13.7"E | BBV-CP014 | PV019042 |
| *Anthene emolus* | Lycaenidae | 22°30'12.8"N 92°09'10.1"E | BBV-CP102 | PP733317 |
| *Castalius rosimon* | Lycaenidae | 23°52'29.5"N 90°16'11.1"E | BBV-CP089 | PP732674 |
| *Charaxes bernardus* | Nymphalidae | 23°52'34.0"N 90°16'06.4"E | BBV-CP306 | PP733555 |
| *Stibochiona nicea* | Nymphalidae | 22°29'22.5"N 92°10'42.7"E | BBV-CP218 | PP733554 |
| *Euthalia aconthea* | Nymphalidae | 23°52'47.8"N 90°15'40.4"E | BBV-CP161 | PP980689 |
| *Hypolimnas bolina* | Nymphalidae | 23°52'33.9"N 90°14'27.4"E | BBV-CP162 | PV449465 |
| *Papilio demoleus* | Papilionidae | 23°52'34.8"N 90°16'05.5"E | BBV-CP238 | PP732658 |
| *Papilio nephelus* | Papilionidae | 22°29'48.4"N 92°11'14.5"E | BBV-CP239 | PP732676 |
| *Papilio memnon* | Papilionidae | 24°15'21.4"N 91°54'19.3"E | BBV-CP236 | PV019097 |
| *Catopsilia pyranthe* | Pieridae | 24°15'32.7"N 91°54'11.6"E | BBV-CP260 | PV019084 |
| *Leptosia nina* | Pieridae | 23°53'35.4"N 90°16'02.3"E | BBV-CP268 | PV020668 |
| *Delias eucharis* | Pieridae | 24°23'49.9"N 88°33'02.6"E | BBV-CP269 | PV449460 |
| *Eurema hecabe* | Pieridae | 23°53'34.7"N 90°16'02.9"E | BBV-CP264 | PV449463 |
| *Zemeros flegyas* | Riodinidae | 22°30'26.8"N 92°08'28.2"E | BBV-CP283 | PP733319 |
| *Ypthima baldus* | Satyridae | 23°53'15.0"N 90°16'04.4"E | BBV-CP286 | PP732659 |
| *Elymnias hypermnestra* | Satyridae | 23°52'33.9"N 90°16'06.0"E | BBV-CP287 | PV019348 |

*Phylogenetic analysis*: The construction of phylogenetic tree was carried out using the Maximum Likelihood (ML) method that included two clades A and B respectively (Fig. 2). Clade A consisted of *Acraea violae* (Bangladesh and India), *Discophora sondaica* (Bangladesh and China), *Ypthima baldus* (Bangladesh and India), *Hypolimnas bolina* (Bangladesh and China), *Danaus chrysippus* (Bangladesh and India), *Stibochiona nicea* (Bangladesh and China), *Leptosia nina* (Bangladesh and Thailand), *Eurema hecabe* (Bangladesh and Korea), *Delias eucharis* (Bangladesh and India) and *Elymnias hypermnestra* (Bangladesh and India). This larger clade encompasses a diverse group of butterfly species, primarily belonging to the families Nymphalidae, Amathusiidae, Pieridae, Acraeidae, Satyridae and Danaidae. Within Clade A, individuals belonging to the same species from different countries consistently form strongly supported monophyletic entity (high bootstrap values, 81-99).

While Clade B consisted of *Anthene emolus* (Bangaladesh and China) *Castalius rosimon* (Bangladesh and India), *Zemeros flegyas* (Bangladesh and

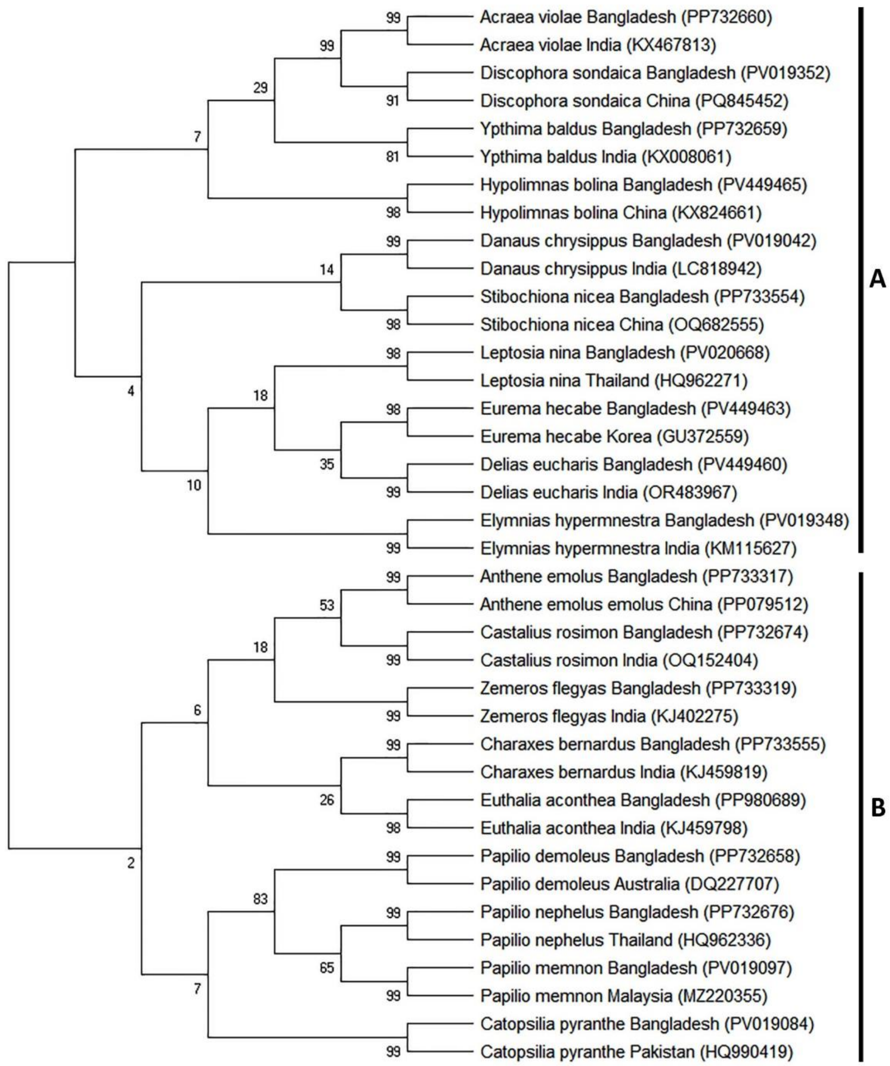


Fig. 2. Phylogenetic tree created using the ML technique with nineteen COI gene sequences. Bootstrap values are displayed at the nodes of the tree.

India), *Charaxes bernardus* (Bangladesh and India), *Euthalia aconthea* (Bangladesh and India), *Papilio demoleus* (Bangladesh and Australia), *Papilio nephelus* (Bangladesh and Thailand), *Papilio memnon* (Bangladesh and Malaysia) and *Catopsilia pyranthe* (Bangladesh and Pakistan). This clade primarily consists of butterfly species from the families Papilionidae, Riodinidae, Pieridae, Nymphalidae and Lycaenidae. Similar to Clade A, the member of Clade B conspecific individuals from different countries within Clade B form highly supported monophyletic groups (98-99 bootstrap value).

The consistent clustering of individuals belonging to the same species from different geographical locations with high bootstrap support strongly validates the use of the COI gene as an effective marker for species identification (DNA barcoding) for these butterflies' taxa (Hebert *et al.* 2003, Hajibabaei *et al.* 2006). It confirms that the COI gene can reliably distinguish these species across their distribution ranges. The ML tree supports the current taxonomic classification of these butterflies at the species level. The monophyletic grouping of conspecifics indicates that the COI gene data aligns well with established species boundaries (Dinca *et al.* 2015). Accurate species identification is essential for effective biodiversity conservation and additional genetic research (Janzen *et al.* 2005). These findings contribute to a better understanding of butterfly biodiversity and can aid conservation efforts by improving species documentation and monitoring. Further studies with larger datasets and additional genetic markers could enhance the resolution of phylogenetic relationships and provide deeper insights into the evolutionary history of these Lepidopteran insects.

## CONCLUSION

This study shows the efficacy of mitochondrial COI gene-based molecular methods as a reliable substitute for conventional morphological techniques in the identification of butterfly caterpillars. The effective amplification and sequencing of COI gene regions from nineteen butterfly caterpillars, along with significant sequence similarity to GenBank entries and the establishment of robust phylogenetic clades, validates the precision and dependability of DNA barcoding in elucidating species-level taxonomy and evolutionary relationships. These findings highlight the effectiveness of COI gene-based methods in biodiversity research, particularly when developmental stage or phenotypic plasticity impedes morphological identification.

*Acknowledgments*: The University Grants Commission of Bangladesh is acknowledged by the authors for providing financial support for this research. The authors are also grateful to Jahangirnagar University's Wazed Miah Science Research Centre for allowing them their lab space. Moreover, authors also extended their appreciations to Md. Towhidur Rahman for his support in collecting the sample.

## LITERATURE CITED

ABEDIN, M.S., GHOSH, A., HOWLADER, A.J. and HOSSAIN, M.M. 2023. Molecular identification and genetic diversity of ten Pierid butterflies based on mitochondrial COI gene. *Bangladesh Journal of Zoology*, **51**(3): 289–299. https://doi.org/10.3329/bjz.v51i3.72027

ALURI, J.S.R. 2022. Butterfly-flower interactions and biodiversity in the changing environments. *Journal of Biodiversity & Endangered Species*, **8**(2): 1–2.

BINGHAM, C.T. and BLANFORD, W.T. 1905. *The fauna of British India, including ceylon and burma: Butterflies*. Taylor and Francis.

CAMERON, S.L. 2014. Insect mitochondrial genomics: Implications for evolution and phylogeny. *Annual Review of Entomology,* **59**: 95–117. https://doi.org/10.1146/annurev-ento-011613-162007

DINCA, V., ZAKHAROV, E.V., HEBERT, P.D.N. and VILA, R. 2015. Complete DNA barcode reference library for a country's butterfly fauna reveals high performance for temperate Europe. *Proceedings of the Royal Society B: Biological Sciences,* **282**(1805): 20142335.

FOSTER, P.G., JERMIIN, L.S. and HICKEY, D.A. 2010. Nucleotide composition bias affects amino acid content in proteins coded by animal mitochondria. *Journal of Molecular Evolution,* **50**(3): 284–290. https://doi.org/10.1007/s00239-004-0178-2

FOUNDATION OF INDIAN BUTTERFLIES (https://www.ifoundbutterflies.org/) (Accessed on 09.05.25)

HAJIBABAEI, M., JANZEN, D.H., BURNS, J.M., HALLWACHS, W. and HEBERT, P.D.N. 2006. DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences,* **103**(4): 968–971. https://doi.org/10.1073/pnas.0510466103

HEBERT, P.D.N., CYWINSKA, A., BALL, S.L. and DEWAARD, J.R. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, ***270***(1512): 313-321.

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

HOSSAIN, M.M., GHOSH, A., BORSHA, F.S., HASAN, M.K., ABEDIN, M.S. and HOWLADER, A.J. 2022. Molecular phylogeny of swallowtail butterflies (Lepidoptera: Papilionidae) based on mitochondrial cytochrome c oxidase I (COI) gene from Bangladesh. *Nepalese Journal of Zoology*, **6**(S1): 1–6. https://doi.org/10.3126/njz.v6iS1.50523

HAUSMANN, A., DILLER, J., MORINIERE, J., HÖCHERL, A., FLOREN, A. and HASZPRUNAR, G. 2020. DNA barcoding of fogged caterpillars in Peru: A novel approach for unveiling host-plant relationships of tropical moths (Insecta, Lepidoptera). *PLoS ONE*, **15**(1): e0224188.

ISLAM, S., MAMUN, M.A.A., AKTER, K., MEEM, S.M., ABEDIN, M.S., BORSHA F.S. and HOSSAIN, M.M. 2023. DNA barcoding of eggs from *Papilio polytes*: A molecular approach for species identification using immature stage. 23rd National Conference and AGM 2022 of the Zoological Society of Bangladesh, pp 105.

ISLAM, S., SAKAL, I.S., ABEDIN, M.S., GHOSH, A., BORSHA, F.S., HASAN, M.K., MAMUN M.A.A., AKTER, K., MEEM, S.M. and HOSSAIN, M.M. 2024. Molecular identification of twenty-six moth species based on mitichondrial COI gene sequences. *Bangladesh Journal of Zoology*, **52**(1): 67–79. https://doi.org/10.3329/bjz.v52i1.74728

JANZEN, D.H., HAJIBABAEI, M., BURNS, J.M., HALLWACHS, W., REMIGIO, E. and HEBERT, P.D.N. 2005. Wedding biodiversity inventory of a large and complex Lepidoptera fauna with

DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences,* ***360***(1462): 1835–1845.

JOAN, C. H., TOTH, J.P. and MONASTERIO, Y. 2022. Luis Sánchez Mesa, Miguel G. Muñoz Sariot, Ruth Escobés, Roger Vila. "Integrative Taxonomy Reveals a New *Melitaea* (Lepidoptera: Nymphalidae) Species Widely Distributed in the Iberian Peninsula," *Insect Systematics and Diversity*, **6**(2): 1-9.

KARMAKAR, T., NITN, R., SARKAR, V., BAIDYA, S., MAZUMDER, S., CHANDRASEKHARAN, V.K., DAS, R., KUMAR, G.S.G., LOKHANDE, S., VEINO, J., VEINO, L., VEINO, R., MIRZA, Z., SANAP, R.V., SARKAR, B. and KUNTE, K. 2018. Early stages and larval host plants of some northeastern Indian buterfies. Journal of Threatened Taxa, **10**(6): 11780–11799; htp://doi.org/10.11609/jot.3169.10.6.11780-11799

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., STECHER, G., LI, M., KNYAZ, C. and TAMURA, K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol. Biol. Evol. **35**(6): 1547-1549.

MEEM, S.M., AKTER, K., MAMUN, M.A.A., ISLAM, S., BORSHA, F.S., ABEDIN M.S. and HOSSAIN, M.M. 2023. DNA barcoding and phylogenetic relationships of ten butterfly caterpillars. *Asian-Australasian J. Biosci. Biotechnol.* **8**(3): 49-55.

NIE, L., WANG, Y., HUANG, D., TAO, R., HAO, C.S.J. and ZHU, C. 2018. Mitochondrial genomes of four pierid butterfly species (Lepidoptera: Pieridae) with assessments about Pieridae phylogeny upon multiple mitogenomic datasets. *Zool. Sys.* **43**(4): 387-409. https://doi.org/10.11865/zs.201834

SAITOU, N. and NEI, M. 1987. The neighbor-joining method:a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406-425.

SHAHRONI, A.M., SHIDIQ, F., IRWANSYAH, S.L. and PEGGIE, D. 2022. Record on some Caterpillars of Butterflies (Lepidoptera: Papilionoidea) in East Java, Indonesia. *Treubia*, **49**: 25-40.

SINHA, D. and DUTTA, A.J. 2024. Butterfly-Host plant relationships ecological significance and conservation needs- A review. *International Journal of Entomology Research*, **9**(8): 230-235.

WAGNER, D.L. and HOYT, A.C. 2022. On Being a Caterpillar: Structure, Function, Ecology, and Behavior. In: Marquis, R.J., Koptur, S. (eds) Caterpillars in the Middle. Fascinating Life Sciences. Springer, *Cham.* https://doi.org/10.1007/978-3-030-86688-4_2

*(Manuscript received on 1 June 2025 revised on 12 August 2025)*