

***In silico* evolutionary analysis and phylogenetic relationships of *Gonepteryx* and related *Lepidoptera* based on the maximum likelihood approach**

Pallavi Vijendra Singh*

Department of Zoology, Bharatiya Mahavidyalaya, Amravati, Maharashtra, India

Corresponding Author: Pallavi Vijendra Singh

DOI: <https://doi.org/10.66856/ijzs.2026.11.3.11078>

Abstract

The morphology, geographical distribution, and evolutionary relationships of butterflies belonging to the *Pieridae* family exhibit significant variation. The current study analyzed the nucleotide substitution pattern and evolutionary relationships of specific *Gonepteryx* and related *Lepidoptera* sequences. The 16-nucleotide gene sequence of mitochondrial Cytochrome Oxidase subunit-1 (COI) was retrieved from the NCBI GenBank database. A significant bias toward transitions was revealed by the substitution analysis, with C→T/U exhibiting the highest relative substitution rate (34.88). A significant AT bias was also evident in the nucleotide composition, with A and T/U together comprising about 71.01% of the sequences. A number of strongly supported molecular affinities were identified using the Maximum Likelihood tree. While the *G. cleopatra*-associated group showed 99% support, the two *Gonepteryx farinosa* sequences formed a well-supported clade with 100% bootstrap support.

The *G. aspasia* and *G. mahaguru* sequences clustered with 65% bootstrap support, but their larger grouping with *G. rhamni* was 100% supported. One *Lepidoptera* sp. sequence demonstrated a tight relationship with the *G. cleopatra* group with an 88% bootstrap value, whereas two *Lepidoptera* sp. sequences constituted a well-supported group with 90% bootstrap support. On the other hand, certain *G. rhamni* sequences had bootstrap values of 23% and 29%, indicating weakly supported associations.

Conclusion: The work shows that while some deeper linkages are still unclear, the examined *Gonepteryx* sequences exhibit a number of firmly supported molecular relationships. The results offer valuable molecular insights into the evolutionary relationships between the selected taxa.

Keywords: Phylogenetic analysis, *Gonepteryx*, *Pieridae*, maximum likelihood

Introduction

Studies of biodiversity, ecology, and evolution frequently involve butterflies, a significant category of insects. Environmental factors, geographic distribution, and evolutionary history all have an impact on their diversity. Numerous frequently occurring butterflies belong to the *Pieridae* family, which exhibits significant variety in form, range, and ecological adaptation. The usefulness of molecular data for comprehending the evolutionary history of the family has been demonstrated by recent large-scale molecular studies that have increased knowledge of connections within *Pieridae* (Carvalho *et al.*, 2024) ^[1]. The genus *Gonepteryx* is a member of the Coliadinae subfamily of the *Pieridae* family. Among the species in this genus are *Gonepteryx rhamni*, *G. cleopatra*, *G. farinosa*, *G. mahaguru*, and *G. aspasia*. *Gonepteryx* is recognized as a genus within *Pieridae* by current taxonomic databases. A number of the genus's species have molecular data accessible, including full mitochondrial genome information for *G. mahaguru* and *G. rhamni*, which are helpful resources for evolutionary and comparative research. The morphological as well as molecular divergence of the four species of butterflies belonging to Nymphalidae and *Pieridae* from the Western part of the Himalayan ranges was studied by Singh V. *et al.* (2020) ^[4] and found to play an important role of morphological, morphometric, and molecular characters for correct identification of butterfly species. Maximum likelihood is one of the popular methods for reconstructing phylogenetic relationships. It assesses

several tree configurations and determines which topology, given a chosen nucleotide substitution model, best explains the observed sequence data. When nucleotide frequencies are uneven, and transition and transversion substitutions happen at variable rates, the Tamura-Nei model is especially helpful (Tamura and Nei, 1993) ^[5]. The Maximum Likelihood technique was used in the current work to examine molecular evolutionary links between certain *Gonepteryx* and related *Lepidoptera* sequences. The Tamura-Nei substitution model was used for the analysis, which comprised 16 nucleotide sequences with 677 locations. The study's objectives were to look at the nucleotide substitution pattern and use bootstrap analysis to ascertain the evolutionary relationships between the chosen sequences. It is anticipated that the findings will offer more molecular proof for comprehending the connections between the examined *Gonepteryx* taxa and related sequences.

Materials and Methods

Sequence Data Collection

For the current phylogenetic study, 16 nucleotide gene sequences of mitochondrial Cytochrome Oxidase subunit-1 (COI) were obtained from the NCBI GenBank database. Selected taxa of the genus *Gonepteryx* and related *Lepidoptera* were represented by the sequences. In addition to sequences classified as *Lepidoptera* sp. and *Hesperiidae* sp., the dataset contained sequences identified to *Gonepteryx mahaguru*, *G. aspasia*, *G. rhamni*, *G. aspasia*

alvinda, *G. farinosa*, and *G. cleopatra*. Additionally, the current NCBI taxonomy database was used to verify the taxonomic placement of representative *Gonepteryx* species.

Sequence Alignment and Dataset Preparation

Phylogenetic analysis began by aligning the selected nucleotide sequences. There were 677 nucleotide positions in the final dataset. The first, second, and third codon positions were taken into account, along with noncoding and coding regions. To make sure the locations utilized for phylogenetic reconstruction were similar among the chosen taxa, the aligned sequences were analyzed.

Estimation of Nucleotide Substitution

The Tamura-Nei model was used to predict the nucleotide substitution pattern. This model allows varying rates of transition and transversion substitutions and takes into account variations in nucleotide frequency. As a result, it works well with sequence datasets where substitution rates and nucleotide composition differ (Tamura and Nei, 1993) [5].

Phylogenetic Analysis

The Maximum Likelihood (ML) approach with the Tamura-Nei substitution model was used to determine the evolutionary relationships between the 16 nucleotide sequences. MEGA11 was used for the analysis (Tamura *et al.*, 2021) [6]. MEGA11 offers techniques for Maximum Likelihood phylogenetic reconstruction, substitution-model analysis, evolutionary distance estimation, and sequence analysis. The pairwise distance matrix calculated under the Tamura-Nei model was utilized to create initial trees for the tree search using the Neighbor-Joining (NJ) and BioNJ techniques. The topology with the highest log-likelihood was selected as the final Maximum Likelihood tree after multiple possible tree topologies were examined.

Bootstrap Analysis

Using 1000 bootstrap iterations, the estimated phylogenetic relationships' reliability was evaluated. By repeatedly resampling nucleotide characters from the original dataset, the bootstrap technique estimates the consistency of particular branches (Felsenstein, 1985) [2]. Each branch in the phylogenetic tree was accompanied by the percentage of duplicate trees in which a certain group was recovered. According to MEGA11's analysis settings, branches with less than 50% bootstrap support were collapsed (Tamura *et al.*, 2021) [6]. Stronger support for the respective grouping was indicated by higher bootstrap values.

Observations and Results

Substitution Matrix

The maximum-likelihood substitution analysis demonstrated a strong bias in favor of transition substitutions, especially C→T/U, which had the greatest relative substitution rate (34.88). While A→G transitions occurred at relatively lower rates, the reverse T/U→C substitution was likewise rather high (12.99). In general, the transversion rates were lower, while the figures for A→T/U replacements were rather higher. T/U (40.21%) and A (30.80%) jointly accounted for 71.01% of the nucleotides, compared to 28.99% for G+C, indicating a substantial AT bias in the nucleotide composition. The use of the Tamura-Nei model to estimate

evolutionary relationships in the dataset is supported by these results, which show significant variability in nucleotide composition and substitution rates.

Table 1: Maximum Likelihood Estimate of Substitution Matrix

	A	T/U	C	G
A	-	8.98	3.35	2.33
T/U	6.88	-	12.99	3.13
C	6.88	34.88	-	3.13
G	5.12	8.98	3.35	-

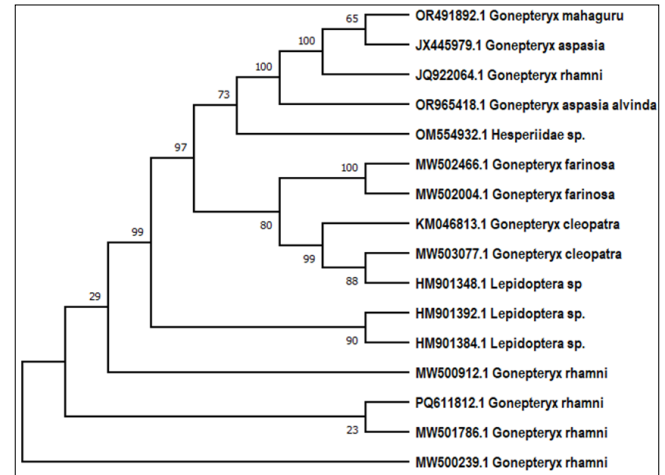


Fig 1: Phylogenetic tree inferred using the maximum likelihood method with the Tamura-Nei model

Phylogenetic Relationship

Using the Maximum Likelihood approach based on the Tamura-Nei model in MEGA11, the phylogenetic relationships between the 16 nucleotide sequences were deduced. First, second, and third codon positions as well as noncoding areas comprised the 677 nucleotide sites used in the investigation. Several strongly supported molecular clades within the examined taxa were identified in the resultant tree. While the *G. cleopatra*-associated group has 99% bootstrap support, the two *Gonepteryx farinosa* sequences constituted a robustly supported clade with 100% bootstrap support. With intermediate bootstrap support (65%), the sequences of *G. mahaguru* and *G. aspasia* clustered together. This group was then linked to *G. rhamni* with high support (100%).

Additionally, this highly supported upper *Gonepteryx* assemblage included the *G. aspasia alvinda* lineage. While one *Lepidoptera* sp. sequence had a strong link with the *G. cleopatra* group (88%), two *Lepidoptera* sp. sequences constituted a well-supported group with 90% bootstrap support. Low bootstrap values of 23 and 29% show that although several *G. rhamni* sequences were found in distinct locations within the tree, their deeper relationships were not well supported. Overall, the phylogeny shows a number of links between the examined *Gonepteryx* sequences that are strongly supported, while some deeper and interspecific relationships are still unclear.

Discussion

The selected *Gonepteryx* and related *Lepidoptera* sequences revealed distinct variations in nucleotide substitution patterns and evolutionary links, according to the current study. The Tamura-Nei model (Tamura & Nei, 1993) [5] is supported by the high C→T/U substitution rate (34.88) and

the AT-rich nucleotide composition (71.01%), which show a non-uniform pattern of nucleotide evolution. While *G. mahaguru* and *G. aspasia* showed intermediate support (65%), the Maximum Likelihood tree showed strong support for several groups, most notably *G. farinosa* (100%) and the *G. cleopatra*-associated group (99%). Similar research has shown how molecular data can resolve relationships among closely related butterfly taxa (Singh *et al.*, 2020; Carvalho *et al.*, 2024) ^[1, 4]. While the low support values of 23–29% for some *G. rhamni* relationships imply that these branches are still unresolved and should be interpreted carefully, the 88–90% bootstrap support seen for several Lepidoptera sequences further suggests robust molecular links (Felsenstein, 1985) ^[2]. Overall, the results show that Maximum Likelihood analysis is a valuable tool for evaluating molecular interactions; however, the poorly supported relationships may be clarified with further sampling and molecular markers (Tamura *et al.*, 2021; Schoch *et al.*, 20) ^[3, 6].

Conclusion

The study demonstrates the maximum likelihood method identified the strongly supported clades, as reflected in bootstrap values and substitution patterns; for elucidation of evolutionary relationships between members of the families selected for this work, a larger dataset is needed.

Conflict of Interest

The author declared that there is no conflict of interest

References

1. Carvalho APS, Owens HL, St Laurent RA, Earl C, Dexter KM, Messcher RL, *et al.* A comprehensive phylogeny of *Pieridae* butterflies reveals strong correlation between diversification and temperature. *iScience*, 2024;27(4):109336.
2. Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, 1985;39(4):783–791.
3. Schoch CL, Ciufo S, Domrachev M, Chetvernin V, Federhen S, Geer LY, *et al.* NCBI Taxonomy: A comprehensive update on curation, resources and tools. *Database*, 2020, baaa062.
4. Singh VK, Joshi PC, Gupta SK. Molecular and morphometric divergence of four species of butterflies (Nymphalidae and *Pieridae*) from the Western Himalaya, India. *Molecular Biology Reports*, 2020;47(11):8687–8699.
5. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, 1993;10(3):512–526.
6. Tamura K, Stecher G, Kumar S. MEGA11: Molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution*, 2021;38(7):3022–3027.