International Journal of Entomology Research
www.entomologyjournals.com
ISSN: 2455-4758
Received: 01-08-2022, Accepted: 18-08-2022, Published: 05-09-2022
Volume 7, Issue 9, 2022, Page No. 62-67

Molecular studies of eight species of butterflies (Lepidoptera: Insecta) using cytochrome c oxidase subunit i gene

Kanimozhi C, Ramesh V
Department of Zoology, Nehru Memorial College, Afflitated to Bharathidasan University, Puthanampatti, Tamilnadu, India

Abstract
Eight different Lepidopteran pests were selected for the molecular studies. The genomic DNA of these selected species was isolated by the Maggenome method, and their quantity and quality were tested through agarose gel electrophoresis. In the present study Cytochrome oxidase subunit I (COI) gene of eight species of butterflies was sequenced, aligned and used to construct phylogenetic trees. The molecular identification of these eight species was confirmed by comparing the related sequences in the National Centre for Biotechnology Information (NCBI) GenBank. A phylogenetic tree was constructed for the 8 COI gene sequences of different butterflies by the Neighbour-Joining method employing Bootstrap type. Thus the results besides showing the relationships and difference between the selected species also signify the merits of molecular phylogenetic studies based on evolutionarily conserved genes such as Cytochrome oxidase which predicts the phylogeny in contrast to the conventional methods which are based on visible traits like morphology, wing pattern, color, etc.

Keywords: lepidopteran pests, molecular identification, COI gene, phylogenetic trees

Introduction

DNA taxonomy is now used in harmony but in addition to other classical morphological data to delimit species (Tautz *et al*., 2002) [7]. Although it has been well accepted that DNA taxonomy can solve many taxonomic problems, it has still not got a central role in it. Presently, scientists are working on phylogeny and phylogeography of different species using DNA as the central theme of their analysis. Although morphological attributes are going to play a major role in the taxonomic description, DNA can be given a better position than what it has today. We believe the best way to give DNA its fair chance in taxonomy will be to implement DNA barcoding as an international unit for the identification of species.

Barcoding is international units for the identification of commercial products, having electronic barcodes, are used by the tradesman to get some information about the concerned product. This barcode acts as an identity for an article and it varies for each item that we purchase. This is the "universal product code" method (Savolainen *et al*., 2005) [8], which is known as a barcode in the retail business. Building upon this idea, Paul Hebert from the University of Guelph in Canada, developed the use of part of the mitochondrial gene as a "universal identification" marker for living organisms (Savolainen *et al*., 2005) [8]. The whole notion of product barcode is based on the arrangement of 10 alternate numbers in 11 positions to create 100 billion unique numbers, which can then be used as an individual product identity (Hebert *et al*., 2003) [9]. Genomic DNA can be used in the 17 same manners, but the problem faced is that only 4 bases are there to work with. This problem is solved when we look at the enormous size of the available DNA in the animal cell. It has been calculated by (Hebert *et al*., 2003) [9] just taking 15 sites of nucleotide positions can create a possibility of 15 codes, which is huge compared to the artificial barcode system. Some of the 4 sequences are very highly conserved, and other regions provide diversity to be checked for at least intraspecific levels. The present work was carried out to reveal the cytochrome oxidase subunit I (COI) gene sequences of eight species of butterflies for their species identification, genetic divergence and phylogenetic status.

Materials and Methods

DNA Isolation

Eight Lepidopteran pests were selected for molecular studies The thorax part of the butterflies should be removed in a sterile condition and washed thoroughly in sterile distilled water 3 times to remove any pollen or microbes attached. DNA was isolated from the butterfly thorax by the MagGenome method.

MagGenome provides an easy and flexible magnetic nanoparticles based method that allows extraction of genomic DNA from a variety of biological sources. It is designed to extract high molecular weight DNA that is free of protein, RNA, and PCR inhibitors. The sample once digested and lysed, is mixed with a solution containing magnetic nanoparticles, upon which the DNA is captured on its surface. The captured DNA is then washed with alcohol-based wash buffers to remove the contaminants before being dissolved in the elution buffer. It also requires the use of a magnetic separator (MagNa Stand) which allows the processing of several samples simultaneously during the washing and elution steps. MagNa Stand is made of rare earth magnets that provide effective magnetic strength to process.

Qualitative Analysis of DNA (Agarose gel Electrophoresis)

The quality of the isolated DNA was checked using agarose gel electrophoresis. 1µl of 6X gel-loading buffer (0.25 % bromophenol blue, 30 % sucrose in TE buffer pH-8.0) was added to 5 µl of DNA. The samples were loaded to 0.8 % agarose gel prepared in 0.5X TBE (Tris-Borate-EDTA)

buffer containing 0.5 µg/ ml ethidium bromide. Electrophoresis was performed with 0.5X TBE buffer at 75 V until the bromophenol dye front has migrated to the bottom of the gel.

PCR Amplification of Gene of Interest

The COI segment was amplified using the following universal primers,

Target Primer Direction Sequence (5' 3')

COI LCOI490 Forward 5'-GGTCA ACAAA TCATA AAGAT ATTGG-3'

HCO2198 Reverse 5'-TAAAC TTCAG GGTGA CCAAA AAATC A-3'

PCR amplification was performed (Applied Biosystems) and programmed to fulfill 40 cycles after an initial denaturation cycle for 10 min at 95°C. Each cycle consisted of a denaturation step at 95°C for 30 sec., an annealing step at 50°C for 20 sec. and an extension step at 72°C for 30 sec. The final extension was at 72°C for 15 mins. The resulting products were resolved by electrophoresis in 1.5% agarose gel and purified with a purification kit (Roche diagnostics). Sequencing was performed using the ABI BigDye terminator (Applied Biosystems).

Sequence Processing

The sequences obtained were assessed for errors, and the "N"s observed in the sequences were manually edited with bases corresponding to the peaks using the program Chromas Lite (http://technelysium.com.au/?page_id=13). Processed sequences were assembled using the Contig. Assembly Program version 3 (CAP3) (http://pbil.univ-lyon1.fr/cap3.php). The ends of the contigs were trimmed to get good quality sequences.

BLAST Analysis

The contigs obtained were compared to the sequences in the NCBI database (http://www/ncbi.nlm.nih.gov/) using the Basic Local Alignment Search Tool (BLAST) (Altschul *et al*., 1990) [1]. Mega blast option was chosen to find sequences with high similarity. All other parameters used were the default settings of the BLAST program. Sequences with more than 97% identity have been considered as the same species.

Gen Bank Submission

The processed sequences were then submitted to GenBank using the submission portal, BankIt for getting accession numbers for each sequence. All necessary information like author details, Genus and species names, Source modifiers, Title, coding sequences, etc. were keyed in during the submission process to complete the submission.

Multiple Sequence Alignment

All 8 sequences were aligned primarily with pair-wise alignment followed by multiple alignment. For the alignment of sequences, MEGA X software package was used which employs ClustalW algorithm (Thompson *et al*., 1994) [2].

Phylogenetic Analysis

Phylogenetic analysis was performed using the MEGAX package which uses its Phylogeny module for this purpose. The evolutionary history was inferred using the Neighbor-Joining method (Saitou N. and Nei M, 1987) [3]. The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of the taxa analyzed (Felsenstein J. 1985) [4]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura K., Nei M., and Kumar S. 2004) [5]. And are in the units of the number of base substitutions per site. This analysis involved 9 nucleotide sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 630 positions in the final dataset. Evolutionary analyses were conducted in MEGA X (Kumar S *et al*., 2018) [6].

Results

Molecular Studies

DNA isolation and sequencing

Eight different Lepidopteran pests were selected for the molecular studies. The selected species are given below (Table: 1)

Table 1: The butterfly species selected for molecular studies

Sr.No	Zoological Name
1	*Eurema hecabe* Linnaeus, 1758
2	*Catopsila pyranthe* Linnaeus, 1758
3	*Jamides celeno* Cramer, 1775
4	*Danaus chrysippus* Linnaeus, 1758
5	*Junonia lemonias* Linnaeus, 1758
6	*Graphium doson* C. & R. Felder, 1864
7	*Papilio demoleus* Linnaeus, 1758
8	*Chilades lajus* Stoll, 1780

The genomic DNA of these selected species was isolated by the Maggenome method, and their quantity and quality were tested through agarose gel electrophoresis. The gel image (Fig: 1) shows the DNA is of significant quality and could be used for any molecular techniques.

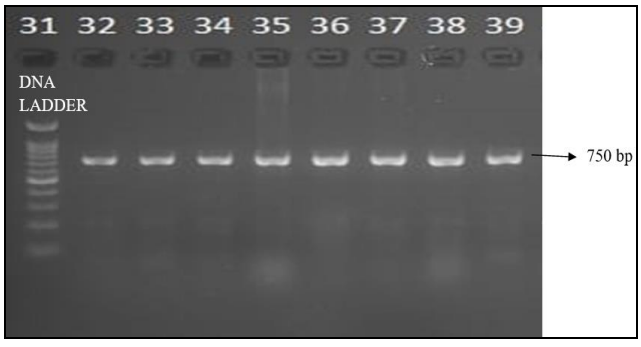


Fig 1: Agarose (1.5%) with genomic DNAs isolated from butterflies

32. Common Grass Yellow 33. Mottled Emigrant 34. Common Cerulean

35. Plain Tiger 36. Lemon Pansy 37. Common Jay 38. Lime Butterfly

39. Lime Blue

The PCR amplified products of the COI gene using forward and reverse primers. The amplifications revealed that there are 750 bp products. The PCR amplicons were purified and

then sequenced by the Sanger sequencing method. The PCR products were checked in 1.5% agarose gels and the molecular standard was a 2-log DNA ladder. The gels are visualized in a UV illuminator and the image was captured under UV light using a Gel documentation system.

DNA Sequence Analysis

The sequences obtained out of sequencing runs were aligned using the CAP3 program. CAP3 program aligns 2 different reads into a single contig based on the overlapping sequences present in both the reads. The obtained contigs were then searched against the non-redundant public database, Nucleotide database of NCBI using the BLAST program. Based on the search results, the end sequences were trimmed to get high-quality contigs, which were analyzed further.

BLAST Analysis

BLAST (Basic local alignment search tool) is a Bioinformatics tool used to align rapidly and compute a DNA sequence with a database of sequences. The sequence given is called query sequence and the sequences matched from the databases are called the subject sequences. The genus and species names are closet match from the database were identified based on BLAST parameters like identity, query coverage, and E-value. The identity of all the sequences was close to 100% (98-99%) and its database base hits along their similarity match are given below (Table: 2).

Table 2: Results of BLAST analysis

S.No	Common Name	Database Hit	Identity
1	Common Grass Yellow	*Eurema hecabe*	99%
2	Mottled Emigrant	*Catopsilia pyranthe*	100%
3	Common Cerulean	*Jamides celeno*	99%
4	Plain Tiger	*Danaus chrysippus*	99.67%
5	Lemon Pansy	*Junonia lemonias*	99.3%
6	Common Jay	*Graphium doson*	99.2%
7	Lime Butterfly	*Papilio demoleus*	99%
8	Lime Blue	*Chilades lajus*	100%

Submission of sequences in GenBank

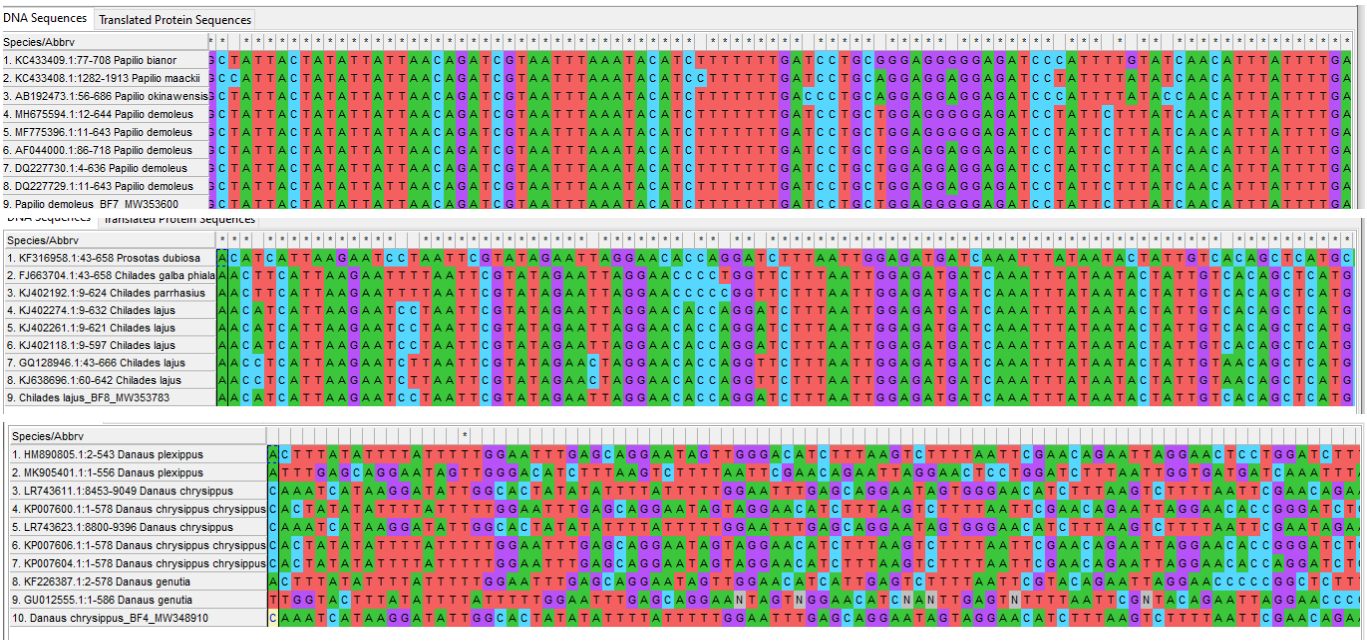
The aligned sequences were submitted to GenBank (http: //www.ncbi. nlm.nih.gov), BankIt-NCBI. The corresponding BankIt accession numbers were obtained. The provided accession numbers by GenBank are tabulated below (Table: 3).

Table 3: GenBank accession numbers for submitted sequences

Sr.No	Common Name	Zoological Name	Accession Number
1	Common Grass Yellow	*Eurema hecabe*	MW342150
2	Mottled Emigrant	*Catopsilia pyranthe*	MW342707
3	Common Cerulean	*Jamides celeno*	MW346679
4	Plain Tiger	*Danaus chrysippus*	MW348910
5	Lemon Pansy	*Junonia lemonias*	MW349495
6	Common Jay	*Graphium doson*	MW353288
7	Lime Butterfly	*Papilio demoleus*	MW353800
8	Lime Blue	*Chilades lajus*	MW353783

Multiple Sequence Alignment

Multiple sequence alignment (MSA) is a global alignment program where each one of the given sequences is aligned from start to end with all the other sequences. ClustalW algorithm was used for this multiple sequence alignment, which comes as a part of the MEGAX software package. MSA helps to identify the sequences that are closely related to each other and it is a prerequisite for any phylogenetic analysis for tree construction. Also, it helps to locate the conserved and variable regions in the sequences. If a sequence doesn't have a matching base at a given position, a gap will be introduced.



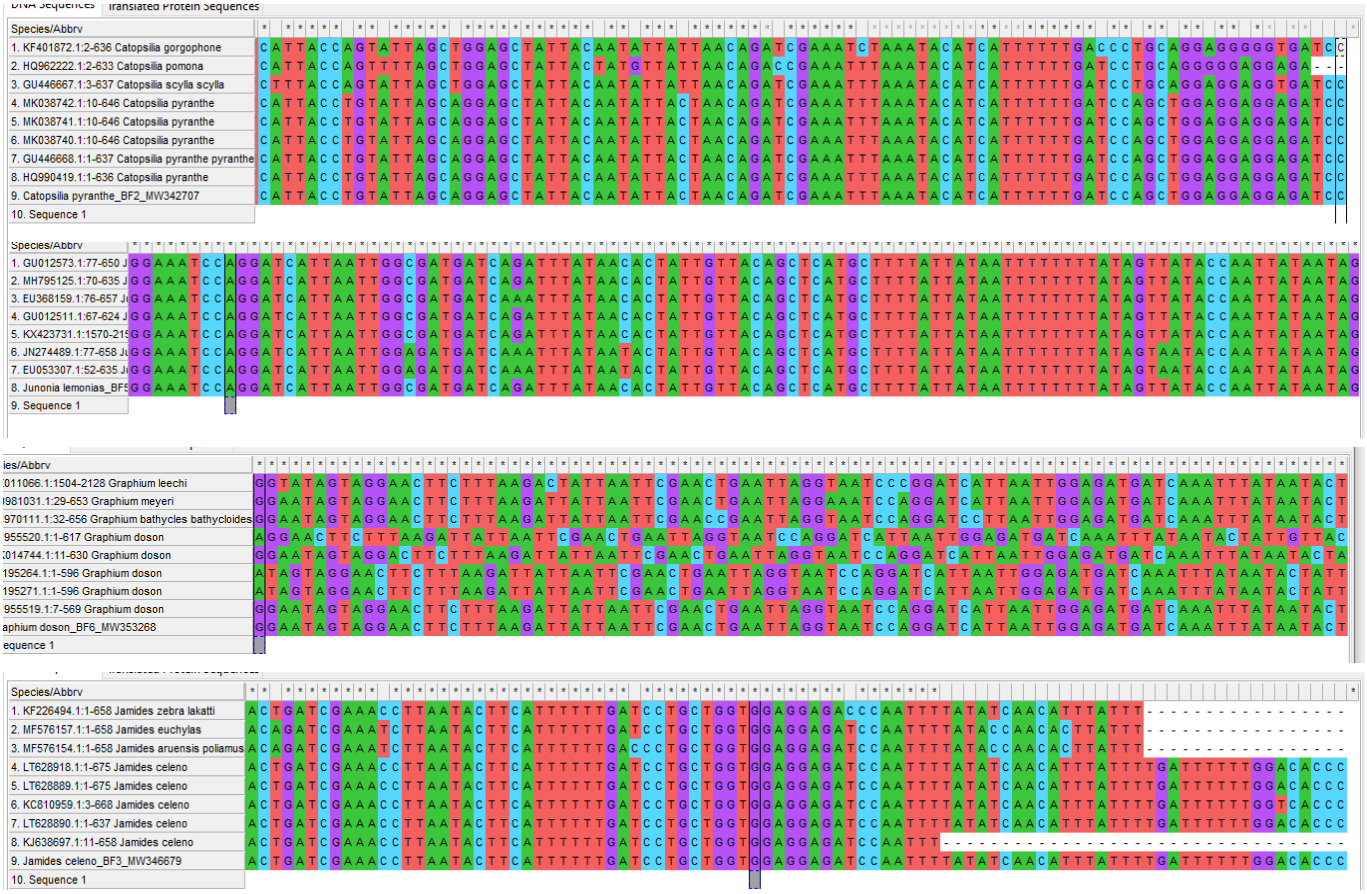


Fig 2: Multiple sequence alignment of COI gene sequences from 8 different butterflies

Phylogenetic Analysis

Phylogenetics is a study that deals with the evolutionary relationships or differences between organisms based on their genetic makeup are called phylogenetics. The difference or relationships are usually represented by a tree-like structure called a phylogenetic tree. The branches of the phylogenetic tree represent the point of divergence and the organisms coming under the same branch are said to be evolutionarily related to each other than those that are in different branches. In this study, a phylogenetic tree was constructed for the 8 COI gene sequences of different butterflies by the Neighbor-Joining method employing Bootstrap type.

In this study, the evolutionary relationship of constructed phylogenetic trees for each genus individually (highlighted as *) (Fig. 3- 10).

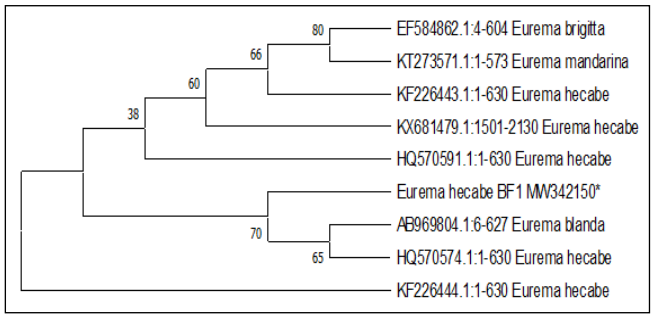


Fig 3: Construction of phylogenetic tree based COI genes of *Eurema hecabe* and it's closely related species

To understand the evolutionary divergence of the interrelated species, a phylogenetic tree was constructed using Clustal W for multiple sequence alignment and the Neighbor-Joining method with maximum composite likelihood method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates). From (Fig: 3) the results showed two monophyletic clades were in clade 1 *Eurema brigitta* and *Eurema mandarina* (sister groups) showed support values of 0.8, while the support values for *Eurema hecabe* ranged from 0.38 to 0.66. This indicated *Eurema brigitta* and *Eurema mandarina* were closely linked while (*E.brigitta* and *E.mandarina*) were relatively distantly linked to *Eurema hecabe.* In clade II, *Eurema blanda* and *Eurema hecabe* (HQ57054.1) were closely linked with a support value of 0.65 while (*Eurema blanda* and *Eurema hecabe*) sister group were relatively distantly linked with *Eurema hecabe* (MW342150) with a support value of 0.70. *Eurema hecabe* (KF226444.1) is distantly related to others, so it is the outgroup in the given phylogenetic tree on the overall, the tree constructed showed interspecific evolutionary similarity as well as divergence among *Eurema.*

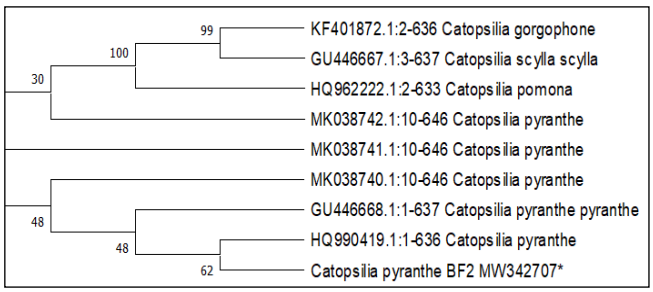


Fig 4: Construction of phylogenetic tree based COI genes of *Catopsilia pyranthe* and its closely related species

The phylogenetic analysis shows two monophyletic clades in clade I (KF401872.1 to MK038742.1) are closely related together and in clade II (MK038740.1 to BF2 MW34270) are closely linked together. In the given phylogenetic tree on the overall, the tree constructed showed interspecific evolutionary similarity as well as divergence among *Catopsilia.* (Fig: 4).

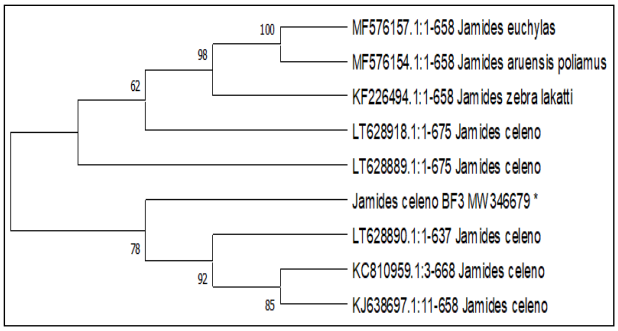


Fig 5: Construction of phylogenetic tree based COI genes of *Jamides celeno* and it's closely related species

From (Fig: 5) the phylogenetic analysis shows two monophyletic clades in clade I (MF576157.1 to LT628889.1) are closely related together and in clade II (BF3 MW346679.1 to KJ638697.1) are closely linked together. In the given phylogenetic tree on the overall, the tree constructed showed interspecific evolutionary similarity as well as divergence among *Jamides.*

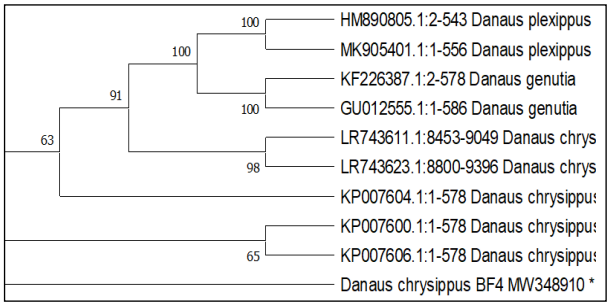


Fig 6: Construction of phylogenetic tree based COI genes of *Danaus chrysippus* and its closely related species

The phylogenetic analysis showed two monophyletic clade in clade I consists of (HMB90805.1 to KP007604.1) are closely linked together whereas clade II consists of (KP007600.1 to KP007606.1) are sister groups. *Danaus chrysippus* (BF 4 MW348910) is distantly related to others, so it is the outgroup in the given phylogenetic tree on the overall, the tree constructed showed interspecific evolutionary similarity as well as divergence among *Danaus* (Fig: 6).

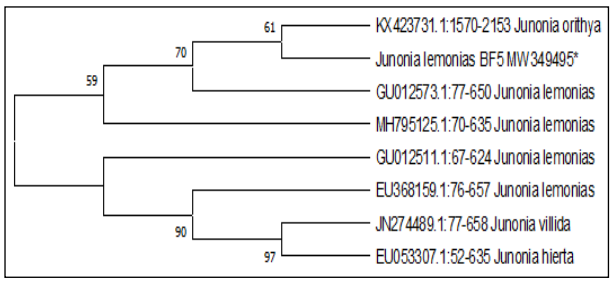


Fig 7: Construction of phylogenetic tree based COI genes of *Junonia lemonias* and its closely related species

From (Fig: 7) the phylogenetic analysis shows two monophyletic clades in clade I (KX423731.1 to MH795125.1) are closely related together and in clade II (GU012511.1 to EU053307.1) are closely linked together. *Junonia lemonias* (BF5 MW349495) were closely linked with *Junonia orithya* with a support value 0.61. In the given phylogenetic tree on the overall, the tree constructed showed interspecific evolutionary similarity as well as divergence among *Junonia.*

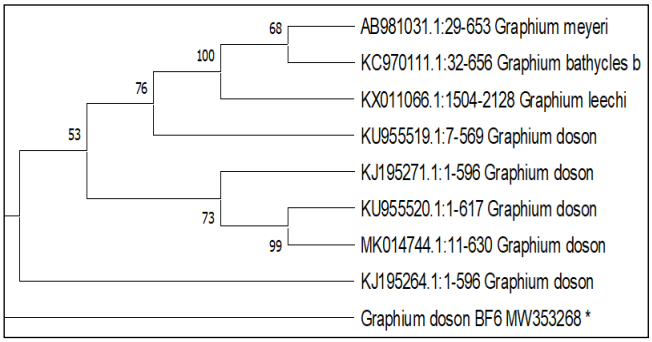


Fig 8: Construction of phylogenetic tree based COI genes of *Graphium doson* and its closely related species

The phylogenetic analysis revealed three monophyletic clades. Clade I consists of (*Graphium meyeri* AB981031.1 to KU955519.1) are sister group to each other. Clade II consists of (KJ195271.1 to MK014744.1) are closely linked together. Clade III contains only one species (*Graphium doson* KJ195284.1). *Graphium doson* (BF 6 MW363268) is distantly related to others, so it is the outgroup in the given phylogenetic tree on the overall, the tree constructed showed interspecific evolutionary similarity as well as divergence among *Graphium* (Fig: 8).

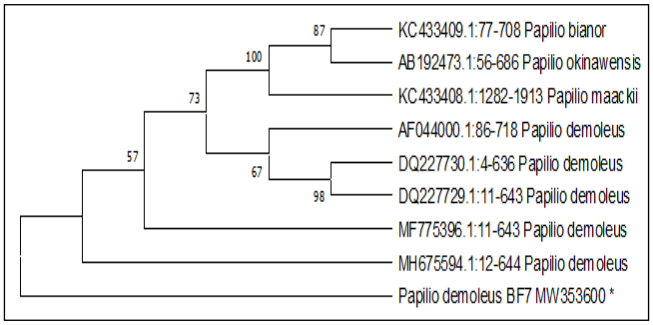


Fig 9: Construction of phylogenetic tree based COI genes of *Papilio demolus* and its closely related species

From (Fig: 9) the phylogenetic analysis reveals four monophyletic clades. Clade I consist of (KC433409.1 to KC433408.1) are closely related to each other. In clade II *Papilio demolus* (DQ227730.1) and *Papilio demoleus* (DQ227729.1) are sister groups while *Papilio demoleus* (AF044000.1) are distantly related group. Clade III contain *Papilio demoleus* (MF775396) and Clade IV contain *Papilio demoleus* (MH675594). *Papilio demoleus* (BF 7 MW353600) is distantly related to others, so it is the outgroup in the given phylogenetic tree on the overall, the tree constructed showed interspecific evolutionary similarity as well as divergence among *Papilio.*

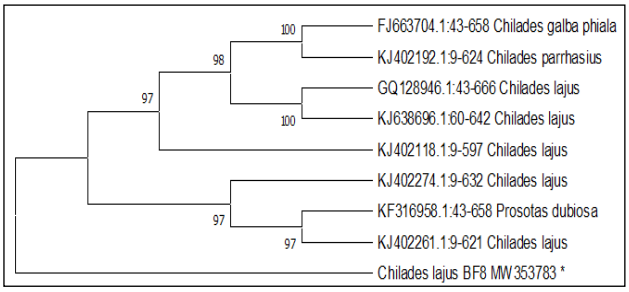


Fig 10: Construction of phylogenetic tree based COI genes of *Chilades lajus* and its closely related species

From (Fig: 10) the results showed two monophyletic clades were in clade 1 is divided into two subclades were *Chilades galba phiale* and C*hilades parrhasius* (sister groups) showed support value of 0.10. In clade 2 *Prosotus dubiosa* and *Chilades lajus* were closely linked with a support value of 0.9. *Chilades lajus* (BF8 MW 353783) is distantly related to others, so it is the outgroup in the given phylogenetic tree on the overall, the tree constructed showed interspecific evolutionary similarity as well as divergence among *Chilades.*

Discussion

Eight different Lepidopteran pests were selected for the molecular studies. The genomic DNA of these selected species was isolated by the Maggenome method, and their quantity and quality were tested through agarose gel electrophoresis. The PCR amplified products of the COI gene using forward and reverse primers. The amplifications revealed that there are 750 bp products. The PCR amplicons were purified and then sequenced by the Sanger sequencing method. The PCR products were checked in 1.5% agarose gels and the molecular standard was a 2-log DNA ladder. The gels are visualized in a UV illuminator and the image was captured under UV light using a Gel documentation system. The sequences obtained out of sequencing runs were aligned using the CAP3 program. CAP3 program aligns 2 different reads into a single contig based on the overlapping sequences present in both the reads. The obtained contigs were then searched against the non-redundant public database, Nucleotide database of NCBI using the BLAST program. Based on the search results, the end sequences were trimmed to get high-quality contigs, which were analyzed further.

BLAST (Basic local alignment search tool) is a Bioinformatics tool used to align rapidly and compute a DNA sequence with a database of sequences. The sequence given is called query sequence and the sequences matched from the databases are called the subject sequences. The genus and species names are closet match from the database were identified based on BLAST parameters like identity, query coverage, and E-value. The identity of all the sequences was close to 100% (98-99%). Multiple sequence alignment (MSA) is a global alignment program where each one of the given sequences is aligned from start to end with all the other sequences. ClustalW algorithm was used for this multiple sequence alignment, which comes as a part of the MEGAX software package. MSA helps to identify the sequences that are closely related to each other and it is a prerequisite for any phylogenetic analysis for tree construction. Also, it helps to locate the conserved and variable regions in the sequences. If a sequence doesn't have a matching base at a given position, a gap will be introduced.

Phylogenetic is a study that deals with the evolutionary relationships or differences between organisms based on their genetic makeup are called phylogenetics. The difference or relationships are usually represented by a tree-like structure called a phylogenetic tree. The branches of the phylogenetic tree represent the point of divergence and the organisms coming under the same branch are said to be evolutionarily related to each other than those that are in different branches. In this study, a phylogenetic tree was constructed for the 8 COI gene sequences of different butterflies by the Neighbor-Joining method employing Bootstrap type.

Similar findings were carried out by Hajibabaei *et al*., 2006 [10] for DNA barcoding for tropical Lepidopteran species. This result found that 97.9% of the 521 species recognized by prior taxonomic work possess distinctive cytochrome C oxidase I barcodes and that the few instances of interspecific sequence overlap involve very similar species. Thus the results besides showing the relationships and difference between the selected species also signify the merits of molecular phylogenetic studies based on evolutionarily conserved genes such as Cytochrome oxidase which predicts the phylogeny in contrast to the conventional methods which are based on visible traits like morphology, wing pattern, colour, etc.

References

1. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of Molecular Biology,1990:215(3):403-410.
2. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic acids research,1994:22(22):4673-4680.
3. Saitou N, Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution,1987:4:406-425.
4. Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. Evolution,1985:39:783-791.
5. Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proceedings of the National Academy of Sciences (USA),2004:101:11030-11035.
6. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution,2018:35:1547-1549.
7. Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP. DNA points the way ahead in taxonomy. Nature,2002:418(6897):479.
8. Savolainen V, Cowan R, Vogler A, Roderick G, Lane R. Towards writing the encyclopaedia of life: an introduction to DNA barcoding. Philosophical Transactions of the Royal Society,2005:B360:1805-1811.
9. Hebert PD, Cywinska A, Ball SL, Dewaard JR. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B: Biological Sciences,2003:270(1512):313-321.
10. Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PD. DNA barcodes distinguish species of tropical Lepidoptera. Proceedings of the National Academy of Sciences,2006:103(4):968-971.