Molecular Identification of Hawkmoths (Lepidoptera: Sphingidae) in Selected Areas of Mt. Kitanglad Based on Cytochrome Oxidase Subunit I (COI) Gene Sequence

Michelle S. Suelo, Aprille Joy M. Luceño[2], Alma B. Mohagan[2] and [1], Reggie Y. Dela Cruz[2]

[1]Central Mindanao University (CMU)
[2]Faculty of the Biology Department, Central Mindanao University, 8714,
University Town Musuan, Maramag, Bukidnon, Philippines

DOI: https://doi.org/10.56293/IJASR.2022.5505

IJASR 2023
VOLUME 6
ISSUE 2 MARCH - APRIL **ISSN: 2581-7876**

Abstract: The study was conducted for the identification of selected species of family sphingidae through DNA Barcoding in Mt. Kitanglad Lirongan, Lantapan, Bukidnon, Philippines. Thirteen species were collected namely: *Acherontia lachesis, Agrius convolvuli, Ambulyx staudingeri, Amplypterus panopus mindanaoensis, Daphnis hypothous, Gnathothlibus erotus erotus, Hippotion brunneum, Hippotion echeclus, Psilogramma menephron, Theretra nessus, Theretra rhesus, Theretra manilae* and *Theretra sugii*. Isolation of the genomic DNA was carried out using the QIAGEN Blood & Tissue Kit. Mitochondrial cytochrome oxidase (COI) gene amplification was carried out using LepF1 (ATTCAACCAATCATAAAGATATTGG) and LepR1 (TAAACTTCTGGATGTCCAAAAAATCA) primers producing 656-666 base pairs were obtained from 30 samples of sphingid moth species. BLAST analyses were able to identify sphingid to the species level. BLAST hits of COI gene sequence of all 10 species ranged from 95%-99% similarity. Maximum likelihood (ML) and Bayesian inference (BI) was used to examine phylogenetic signals in COI with the highest bootstrap values. Sphingid moths formed a monophyletic group based on the clade.

Keywords: DNA Barcoding, Sphingidae, Mitochondrial cytochrome oxidase, Maximum likelihood, Bayesian inference

Introduction

Sphingid moth is also known as hummingbird moth or hawkmoth, belongs to order Lepidoptera, subfamily Bombycoidea. This species of moth has great impact on our biodiversity because they serve as pollinators and help on the reproduction of plant species. Identification system based on DNA has the potential to facilitate both the identification of known species and discovery of new ones. According to Singh [21], molecular tools have provided new opportunities to study questions in evolutionary biology and in phylogenetic systematics. DNA barcoding represents a straightforward way which is designed to provide automatable species identification by using a short-standardized gene regions as internal species tags [8]. It has become increasingly common since it was proposed in 2003 [10]. DNA barcoding is linked to taxonomy through integration of various types of data, such as morphological, ecological, physiological and molecular data which will improves species discovery and description processes. Integration of DNA barcoding with morphology has helped resolving taxonomically difficult groups of organisms [15, 18]. A robust phylogenetic framework for the family is currently lacking [13,3]. There have been no studies on the molecular phylogeny of moths in Mindanao, Philippines. Thus, this study focuses on sphingid moths which belong to the family Sphingidae in Mt. Kitanglad, Lirongan, Lantapan, Bukidnon. This study hypothesizes that the different species of the family Sphingidae in Mt. Kitanglad can be identified using their cytochrome oxidase subunit 1 (COI) gene barcode sequence.

Materials and Methods

Entry protocol and establishments of the sampling sites

A permit to conduct the study and collect samples was acquired from the authorities of the selected study sites. A Gratuitous permit (No. R10 2019-14) were also utilized prior to the conduct of study. The study was carried out across vegetation type of Mt. Kitanglad namely: agro ecosystem, dipterocarp forest, and montane forest.

Sampling procedures, collection and preservation of specimen samples

Different species of moths were collected across the vegetation types in Mt. Kitanglad. Light trapping technique was used for sampling night flying sphingid moths. The light traps were used for 10 hours from 6pm to 4am. Light traps used 500 watts 12 voltage tungsten bulbs powered from portable generator with source power of 220AC, where there was a white sheet where insects will be trapped. The collected moth species were paralyzed by pinching its thorax between the thumb and forefinger. They were transferred individually in glass jars containing absolute ethyl alcohol for preservation.

Extraction and Isolation of Genomic DNA

The wings of 10 species with three replicates each were used for the extraction of genomic DNA. First, the wings were placed in individual sterile two (2)-ml microcentrifuge tube with label. DNA extraction followed the manufacturer's protocol of the QIAGEN DNeasy Blood & Tissue Kit.

Determination of the Quality and Quantity of genomic DNA

The quantity and quality of undiluted gDNA samples were determined using microdrop plate in the microplate reader (MultiSKAN Go, Thermo Scientific). UV absorbances at 230 nm, 260 nm and 280 nm were determined.

PCR Amplification and Gel Electrophoresis

A set of primer, LepF1 (ATTCAACCAATCATAAAGATATTGG) and Lep R1 (TAAACTTCTGGATGTCCAAAAAATCA) [23, 9, 6, 15, 25] was used in the study for amplification of cytochrome oxidase 1 gene.

COI was amplified with the gDNA as template using Applied Biosystem, Inc Veriti Thermocycler. The polymerase chain reaction (PCR) components of 15 µl reaction containing 1x PCR buffer-$MgCl_2$, 25 mM $MgCl_2$, 0.3 mM dNTP mix, 0.3 mM forward primer, 0.3 mM reverse primer, 0.5 U Taq polymerase and 4.0 µl DNA sample. PCR profile consist of PCR conditions: 94 °C (1 min); 5 cycles of 94 °C (30 s), 45 °C (40 s), 72 °C (1 min); 35 cycles of 94 °C (30 s), 51 °C (40 s), 72 °C (1 min); and final extension of 72 °C(10 min) [2].

The PCR products were visualized using gel electrophoresis. One (1) % agarose gel was prepared. 0.35g of agarose was dissolved into 35 ml of 0.25X Tris-Boric EDTA (TBE) and was heated in microwave oven for 1 minute. It was cooled and added with 3.5 µl Biotium Gel Red DNA Stain. The mixture was poured into the molder with aluminium foil cover and allowed to harden. Once hardened, the gel was set into the electrophoresis machine. Five (5) µl of PCR product of gDNA was added with 2 µl of 2x loading dye and was loaded into the wells of the agarose gel. The gel was subjected to run in 100V for 30 minutes using Mupid-One Electrophoresis system. The gel was then viewed using Bio-Rad Doc EZ imager.

Sequence Analysis and Construction of Phylogenetic Tree

An online account was made at the Macrogen website (http://dna.macrogen. com/eng/). The PCR products of moth species were submitted to Macrogen, Inc. in Korea. The sequences were downloaded at Macrogen website. In aligning the sequences given, Bioedit in NCBI was used. After the final sequence for the selected moths was made, the sequences were aligned and analyzed using MEGA7 to obtain its phylogenetic tree. BLAST (BLASTN) analysis [1], were carried out to determine their homology to available sequences in the National Center of Biotechnology Information (NCBI) database. Phylogenetic tree was constructed using Molecular Evolutionary Genetics Analysis (MEGA). The MEGA aligns the sequences, estimates the evolutionary distances and builds trees by using the maximum likelihood method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test with 1000 replicates which will be computed.

Results and Discussions

DNA Isolation

Genomic DNA was successfully isolated from the species of sphingid moths using a QIAGEN DNeasy Blood and Tissue kit. Figure 1 shows the genomic DNA bands of thirty (30) samples (10 species with three (3) replicates each). The quality of the DNA was determined by the ratio of absorbance 260 nm divided by absorbance 280 nm. The absorbance ratio reading ranged from 1.03-1.432. Good quality DNA will range to 1.7-2.0. Based on the result, DNA purity is lower in ratio thus indicating that more contaminants are present. Absorbance reading lower than 1.7 implies the presence of proteins and aromatic amino acids [14].

All 30 samples extracted DNA has low purity, but all were successfully amplified during PCR by increasing the DNA concentration and with the addition of MgCl2 (Figure 1).

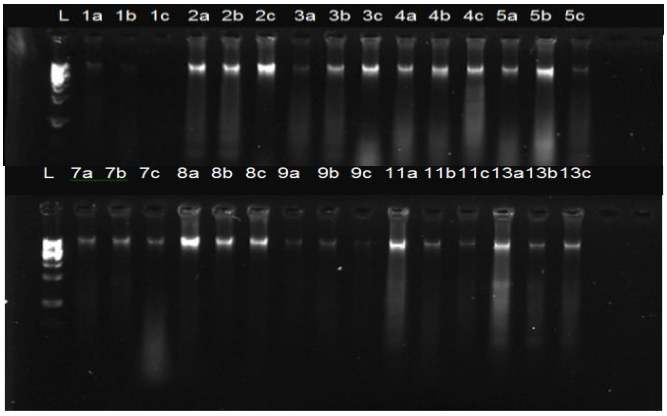

Figure 1. Agarose gel electrophoresis showing the undiluted genomic DNA of 30 samples. L-ladder

PCR Amplification

Polymerase chain reaction (PCR) is used to amplify a single copy or a few copies of DNA across several orders of magnitude, generating thousands to millions of copies of a particular DNA sequence [19, 12]. All 30 samples of sphingid moths were successfully amplified. Smears were also observed during the process. Smears may indicate too much DNA template, MgCl2 not optimal or DNA degradation or contamination of reagents [1, 7]. The approximate size of the amplicons is ~650 base pair. Figure 2 revealed the PCR products of 30 samples.

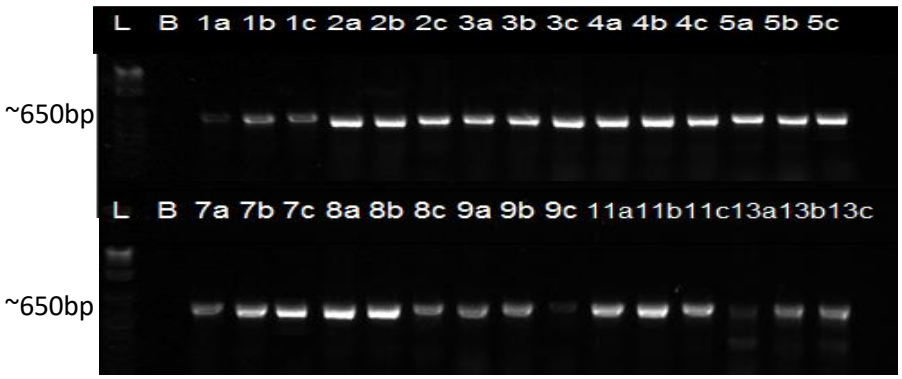

Figure 2. Agarose gel electrophoresis showing the PCR products of 30 samples. L-ladder, B- blank

The amount of gDNA that gave product ranged from 1 µl to 4 µl. Optimization was done trying to generate a positive result of PCR product such as addition of MgCl2, amount of DNA used, dilution of DNA and adjusting the temperature of the PCR protocol. The interaction between the primers and DNA template were influenced by the concentration of MgCl2. Magnesium chloride is essential for the Taq polymerase to function. If the concentration is high, it can stabilize the interaction of DNA template and primer; however, it can result to

erroneous PCR product formation and nonspecific binding. If the MgCl2 is too low, it can help eliminate nonspecific priming, however, it can result in low yield of PCR product [19]. No band or faint bands were also observed during the amplification, it may result if annealing time was too short or annealing temperature was too high. It may also indicate that the DNA was already degraded or contained inhibitor. Primer concentration if it's too high will also result to nonspecific binding. Too much DNA template can inhibit PCR by binding all the primers, but amplification may not be detectable if DNA template is too little template. If there are complementary sequences in two primers used, the primers will hybridize with each other thus forming primer dimers and will not be available for binding with template [4]. Basic Local Alignment Search Tool (BLASTN) is a program that tries to find the identical region of the sequence in the database to the "words" in the words list. The identical region is called a "hit" [10]. The sequence was highly similar to the following COI sequences in GenBank. Based on the analysis of BLASTN, percentage identity ranged 95% - 99% (Table 1). This may indicate that there are already sequences in the public database for the 10 species in this study. A lower E-value (near to 0) indicates higher significance of the similarity [10]. Barcode of Life Data System (BOLD) is an informatics workbench which aids in the storage, analysis and publication of DNA barcode records [20]. Based on the analysis of Barcode of Life Database (BOLD System), percentage similarity ranged 99.22%-100%. According to BOLD database, *Psilogramma medicieloi* was previously identified as *Psilogramma menephron.*

Table 1. Shows the BLAST and BOLD Database Identification

Sample Code	BLAST Search Match	E-value	BLAST Identification %	Accession Number	BOLD System Identification	BOLD System Similarity %	BIN (Cluster ID)
MSS1	*Acherontia lachesis*	0.0	98.77%	MG20017	*Acherontia lachesis*	99.69%	BOLD:AAb1746
MSS2	*Agrius convolvuli*	0.0	99.54%	MG783983	*Agrius convolvuli*	100%	BOLD:AAA2393
MSS3	*Ambulyx staudingeri*	0.0	99.69%	JN677726	*Ambulyx staudingeri*	99.69%	BOLD:ABY9088
MSS4	*Amplyterus panopus*	0.0	95.84%	KP720027	*Amplypterus mindanaoensis*	99.22%	BOLD:AAD2277
MSS5	*Daphnis hypothous*	0.0	99.70%	JN677869	*Daphnis hypothous*	100%	BOLD:AAE7433
MSS7	*Hippotion brunnea*	0.0	99.39%	JN678016	*Hippotion brunnea*	99.85%	BOLD:AAE3980
MSS8	*Hippotion echeclus*	0.0	99.55%	AJ749425	*Hippotion echeclus*	99.85%	BOLD:ABZ6316

MSS9	*Psilogramma menephron*	0.0	98.01%	KJ168433	*Psilogramma medicieloi*	100%	BOLD:ACE3526
MSS11	*Theretra nessus*	0.0	99.84%	KJ168076	*Theretra nessus*	99.82%	BOLD:AAB3024
MSS13	*Theretra sugii*	0.0	99.69%	JN678632	*Theretra sugii*	99.84%	BOLD:AAW6734

Construction of Phylogenetic Tree

A total of 37 DNA sequences were considered in the construction of phylogenetic tree. This is sorted from the 10 samples, 27 samples from NCBI website including 3 outgroups. The 3 outgroups *are Pachyrrhynchus forsteni, Adalia bipunctata* (Two-spotted Lady Beetle) and *Acanthepeira stellata* (Starbelliedorbweaver). The criteria for choosing the DNA sequence from the NCBI website were, 1) must be a sphingid moth, 2) must have a voucher specimen, 3) must not be less than 650bp and 4) must be identified to the species level.

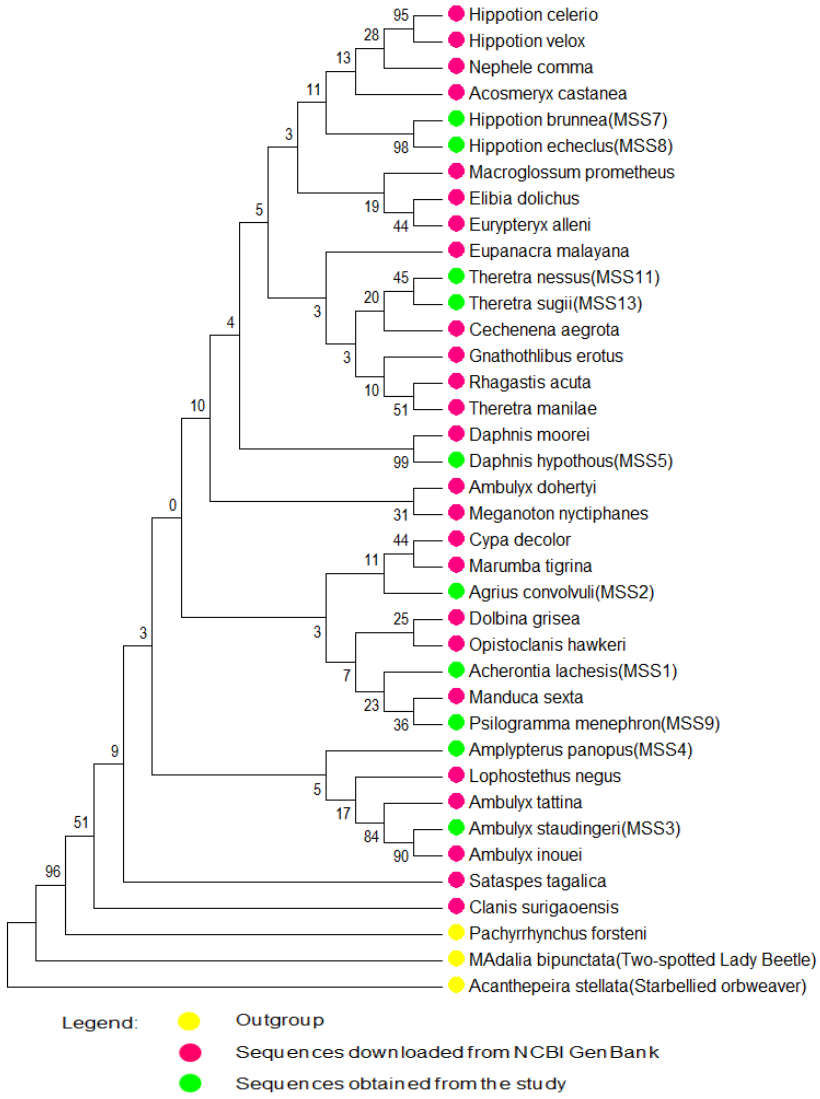

Figure 3. Molecular Phylogenetic Analysis by Maximum Likelihood Method

MEGA 7 which includes MUSCLE is popular software because of its user-friendly interphase. It is widely used since it can find the best substitution model and use the model to infer phylogenetic tree by maximum likelihood method [10, 24,16]. Bayesian method is a character state method for inferring phylogenetic tree. Models with the lowest BIC scores (Bayesian Information Criterion) are considered to describe the best substitution pattern [10].

Maximum likelihood is a statistical methodology for estimating unknown parameters in a model. For each model, AICc value (Akaike Information Criterion, corrected), Maximum Likelihood value (lnL), and the number of parameters (including branch lengths) are also presented (Nei& Kumar, 2000). Non-uniformity of evolutionary rates among sites may be modeled by using a discrete Gamma distribution (+G) with 5 rate categories and by assuming that a certain fraction of sites are evolutionarily invariable (+I). Whenever applicable, estimates of gamma shape parameter and/or the estimated fraction of invariant sites are shown. Assumed or estimated values of transition/transversion bias (R) are shown for each model, as well. They are followed by nucleotide frequencies (f) and rates of base substitutions (r) for each nucleotide pair. Relative values of instantaneous r should be considered when evaluating them. For simplicity, sum of r values is made equal to 1 for each model. For estimating ML values, a tree topology was automatically computed. The analysis involved 254 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There was a total of 32 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 Figure 3 [17].

The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura 3-parameter model [22]. The bootstrap consensus tree inferred from 1000 replicates [5] is taken to represent the evolutionary history of the taxa analyzed [5]. 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 (1000 replicates) are shown next to the branches [5]. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.3766)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 32.1444% sites). The analysis involved 38 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 254 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 [17].

To determine the position of a root, out group should be added to the analysis data. The chosen out group are species which are closely related to group under study but less closely related than the relationship among the species understudy [10]. The bootstrap test is commonly used method for evaluation the reliability of specific clades in the tree. If an internal branch has high proportion (90% or more) (bootstrap value), the internal branch is thought to be reliable [24]. This result confirms that the study of molecular phylogeny of the family sphingidae form a monophyletic family because they form from one ancestor which means that it consists of an ancestral species and all its descendants.

Conclusion

Mitochondrial cytochrome oxidase 1 (COI) sequence obtained from the 10 species (each with 3 replicates) amplified using the primers LepF1 and LepR1. COI gene is useful in identifying the sphingid moths species up to species level successfully. The phylogenetic tree revealed that the ten species form a monophyletic group which means that all organisms collected share common ancestry.

Acknowledgement

The author would like to thank Department of Science and Technology – Science Education Institute (DOST-SEI) for the funding. To Sir Pasu Daniel F. Somera for facilitating our gratuitous permit. Thank you very much.

References

1. Altshuler, M. L. (2006). PCR Troubleshooting: The essential Guide. *Caiser Academic Press, 2*, 1-80. Retrieved from https://www.caister.com/pcr2

2. Ashfaq, M., Akhtar, S., Khan, A. M., Adamowicz, S. J., & Hebert, P. (2013).DNA barcode analysis of butterfly species from Pakistan points towards regional endemism. *Molecular Ecology Resources*, *13,* 832-843. doi:10.1111/1755-0998.12131
3. Cho, S., Epstein, S. W., Mitter, K., Hamilton, C. A., Plotkin, D., Mitter, C., & Kawahara, A. Y. (2016). Preserving and vouchering butterflies and moths for large-scale museum-based molecular research. *PeerJ*. Doi 10.7717/peerj.2160
4. Dubey, V. K. (n. d.). Lecture 37: Polymerase Chain Reaction. *Proteonomics and Genomics,* 1-9. Retrieved from https://nptel.ac.in>courses/102103017/pdf/lecture
5. Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution 39*:783-791
6. Hajibabaei, M., Singer, G., Herbert, P., & Hickey, D. (2007). DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Science Direct*, *23*(4), 167-172. doi:10.1016/j.tig.2007.02.001
7. Hashim, H. O. (2006). PCR Troubleshooting. *Research Gate*. Doi:10.13140/RG.2.2.21694.48967
8. Hebert, P. D. N., & Gregory, T. R. (2005). The Promise of DNA Barcoding for Taxonomy. *Sytematic Biology*, *54*(5), 852-859.doi:10.1080/10635150500354886
9. Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H., &Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptesfulgerator*. *PNAS,* 101(41), 14817. Doi:10.1073pnas.0406166101
10. Horiike, T. (2016). An Introduction to Molecular Phylogenetic Analysis. *Reviews in Agricultural Science, 4,* 36-45.doi:10.7831/ras.4.36
11. Jinbo, U., Kato, T., & Ito, M. (2011). Current progress in DNA Barcoding and future implications for entomology. *Entomological Science, 14*, 107-124. doi:10.1111/j.1479-8298.2011.00449.x
12. Joshi, M., &Dreshpande, J. D. (2010). Polymerase Chain Reaction: Methods, Principles and Application. *International Journal of Biomedical Research,1*(5), 81-97. Retrieved from www.ssjournal.com
13. Kawahara, A. Y., Mignault, A. A., Regier, J. C., Kitching, I. J., &Mitter, C. (2009). Phylogeny and Biogeography of Hawkmoths (Lepidoptera: Sphingidae): Evidence from Five Nuclear Genes. *PLoS ONE, 4*(5). doi:10.1371/journal.pone.0005716
14. Kloos, R. (2016). DR6000 for Quantitative and Qualitative Determination of DNA and RNA. Retrieved from at:hach.com.
15. Koch, H. (2010). Combining Morphology and DNA Barcoding Resolves theTaxonomy of Western Malagasy Liotrigona Moure, 1961 (Hymenoptera: Apidae: Meliponini). *African Invertebrates, 51*(2): 413-421 Retrieved from https://doi.org/10.5733/afin.051.0210
16. Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology Evolution*, *33*, 1870-1874.
17. Nei, M. & Kumar, S. (2000). Molecular Evolution and Phylogenetics. Oxford University Press, New York.
18. Packer, L., Gibbs, J., Sheffield, C., & Hanner, R. (2009). DNA barcoding and the mediocrity of morphology. *Molecular Ecology Resources* 9,42–50
19. Patel, S. V., Bosamia, T. C., Bhalani, H. N., Singh, P., & Kumar, A. (2015). Polymerase Chain Reaction (PCR). *A monthly Magazine of Agricultural and Biological Sciences,13*(9), 10-12. Retrieved from https://www.researchgate.net/publication/272295921
20. Ratnasingham, S.&Hebert, P. D. (2007). BOLD: The Barcode of Life Data System. *Molecular Ecology Notes, 7,*355-364. Doi: 10.1111/j.1471-8286.2006.01678.x
21. Singh, D. (2017). DNA barcoding of Some Indian species of hawkmoths based on COI gene (Lepidoptera: Sphingidae*). Journal of Entomology and Zoology Studies*, *5*(4), 35-40. Retrieved from www.entomoljournal.com
22. Tamura, K. (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C-content biases. *Molecular Biology and Evolution* 9:678-687.
23. Wilson, J. (2012). DNA Barcodes for Insects. *Methods in Molecular Biology*, 858. Doi:10.1007/978-1-61779-591-6_3
24. Yang, Z., &Rannala, B. (2012). Molecular Phylogenetics: principle and practice. *Nature reviews|genetics, 13,303*-314. Retrieved from www.nature.com/reviews/genetics
25. Zhang, R., Huang, X., Jiang, L., &Qiao, G. (2011). Phylogeny and apecies differentiation of *Mollitrichosiphum* spp. (Aphididae, Greenideinae) based on mitochondrial COI and Cyt b genes. *Current Zoology*, 57(6), 806-815