UTILIZING *COI* GENE FOR THE IDENTIFICATION OF THIRTEEN HESPERIIDAE BUTTERFLIES AND DETERMINING THEIR GENETIC RELATIONSHIP

Akter, K., S. M. Meem, S. Islam, M. A. A. Mamun, M. S. Abedin, F. S. Borsha, A. Ghosh[1] and M. M. Hossain[1]*

DNA Barcoding Laboratory, Department of Zoology, Jahangirnagar University, Savar, Dhaka-1342, Bangladesh; [1]Insect Biotechnology Division, IFRB, Atomic Energy Research Establishment, Dhaka, Bangladesh

*Corresponding author: monwar@juniv.edu

Abstract

In the present study, the *COI* genes of thirteen species were sequenced, and NCBI GenBank homology searches of these species revealed significant similarity across diverse countries of origin. The sequences of thirteen species of hesperiid butterflies were then submitted to NCBI's GenBank, and assigned accession number. The average nucleotide composition of 585 bp of these genes was analyzed, and the results showed the mean base compositions of the *COI* sequences were 40.01% T, 16.10% C, 30.21% A, and 13.66% G. The present investigation revealed a significant preference for Adenine+Thymine (A+T), with a prevalence of 70.22%. Afterward, the genetic divergence among thirteen species of hesperiid was analyzed using the Kimura 2-parameter of the MEGA10 program, and the results indicated that the interspecific nucleotide divergence among thirteen species of hesperiid ranged from 0.0532 to 0.1790%. Moreover, a phylogenetic tree was constructed using the Neighbor-Joining (NJ) method, and the result revealed three clusters within the Pyrginae, Hesperiinae, and Coeliadinae subfamilies under the family Hesperiidae. The present investigation revealed Pyrginae as paraphyletic, while Coeliadinae was shown to be the sister group to Hesperiinae. As a final point, the analyses of the data support the following relationships: [(*P. dan+T. japetus*) + (*P. guttata+P. bada+P. agna+B. cinnara+I. semamora+I. salsala+T. colon+U. folus* + (*O. angulatum*)) + (*B. oedipodea+S. gremius*)]. The findings may be applicable to future research in the disciplines of molecular taxonomy to identify hesperiid species and may reveal a higher level of phylogenetic relationship among the subfamilies of Hesperiidae.

Key words: *COI*; Butterfly; Hesperiidae; Phylogenetic; Bangladesh.

INTRODUCTION

The family Hesperiidae, commonly known as "skipper butterflies," has widespread distribution, with several species identified in every continent except Antarctica (Cock 2010). In Bangladesh, 56 species were enlisted as skippers for the threat assessment by IUCN, of which 23 were endangered, 16 vulnerable, 12 least concerned and 5 data deficient (IUCN Bangladesh 2015). These species are grouped into 3 subfamilies, *viz*. Pyrginae, Hesperiinae and Coeliadinae. Consequently, a review in 2023 on skipper butterflies revealed that Bangladesh is now home to the second-largest family after Lycaenidae (29.45%), and possesses remarkable abundance of butterflies (20.66%) across the country (Monwar, personal communication). However, the striking homogeneity of morphological structure among skipper species makes phenotype-based classification exceedingly difficult (Voss 1952, Warren *et al*. 2008). Many cryptic species, such as skippers, are difficult to identify precisely because of morphological diversity, evolution, sexual dimorphism, and other aspects (Xiao *et al*. 2010). While DNA barcodes can be used to identify cryptic species of skipper butterflies previously detected by classic taxonomic methods and to provide first clues to the existence of yet other cryptic species. In DNA barcoding, the mitochondrial cytochrome c oxidase I (*COI*) gene is utilized as a marker, which has about 648 base pairs (Hebert *et al*. 2003a). By using the DNA sequences of a

standardized region from an unknown species and comparing them to the sequence present in the database, species identification can be accomplished (Elías-Gutiérrez *et al.* 2018).

The family Hesperiidae has lots of conflicts in relationships at various taxonomic levels. This family includes around 4000 recognized species all around the world (Bridges 1994), which are currently distributed among 7 subfamilies and 567 genera (Warren *et al*. 2008, 2009). The seven subfamilies include Coeliadinae, Euschemoninae, Eudaminae, Pyrginae, Heteropterinae, Trapezitinae, and Hesperiinae. Nonetheless, the higher-level relationships among skipper butterflies have not yet been adequately resolved and there is still ambiguity regarding the relationships among the major clades within the subfamily Pyrginae (Sahoo *et al*. 2016). Furthermore, support for relationships among the monophyletic subfamilies Heteropterinae, Trapezitinae, and Hesperiinae was weak to moderate (Sahoo *et al*. 2016). In the last two decades, there has been an enormous increase in the use of molecular data (DNA sequences) to assess genetic variation among various taxa. DNA sequences have also been used to resolve the phylogenetic relationship between families and subfamilies (Wahlberg *et al*. 2005, Warren *et al*. 2008, 2009). In the present study, in addition to generating *COI* gene sequences for species identification, we attempted to reveal the subfamily-level phylogenetic relationships of the family Hesperiidae.

MATERIAL AND METHODS

Sample collection

Thirteen hesperiid butterfly specimens were collected from various locations in Bangladesh (Table 1 and Fig. 1). The samples were gathered in the wild using an insect net and preserved by dehydration in a little envelope. Using the keys outlined by Bingham (1905, 1907), Wynter-Blyth (1957) and Talbot (1939, 1947) the butterflies were identified morphologically. Voucher samples were created in accordance with Brower (1996).

DNA extraction, amplification and sequencing

The genomic DNA of thirteen species of hesperiid butterflies was isolated from their legs following the protocol outlined in the Wizard Genomic@ DNA Purification Kit. The amplification of the mitochondrial cytochrome c oxidase I (*COI*) gene area was performed by polymerase chain reaction (PCR) utilising the primers LCO 1490 (5′-GGTCAACAAATCATAAAGATATTGG-3′) and HCO 2198 (5′-TAAACTTCAGGGTGACCAAAAAATCA-3′). PCR was conducted using the Q2 Green PCR Master Mix on a Veriti thermal cycler from Applied Biosystems, USA. The PCR reaction mixture had a total volume of 20 µl, comprising of 10 µl of 2x master mix, 1 µl of forward primer (10 pM), 1 µl of reverse primer (10 pM), 1 µl of template DNA (quantity dependent on the amount of extracted DNA), and 7 µl of nuclease-free water. The parameters of the cycle were as follows: an initial denaturation step at a temperature of 95°C for duration of 5 minutes, followed by 35 cycles consisting of 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 cycle concluded with a final extension step at 72°C for 5 minutes. The effectiveness of the amplification was assessed through the utilisation of 1% agarose gel electrophoresis conducted under UV light conditions using a Bio Analyzer. The amplified product underwent sequencing analysis utilising an ABI 3500 sequencer.

Genetic distance and phylogenetic analysis

Chromas version 2.6.2 was used to edit the sequences of these thirteen butterfly species. Using the ClustalW multiple alignment program in BioEdit version 7.0, the assembled sequences were

aligned (Hall 1999). The Kimura 2-Parameter (K2P) model and the MEGA10 program were used to compute and summarize nucleotide compositions and estimate pairwise distances (Kimura 1980, Kumar *et al*. 2018). With 1000 bootstrap replications, a phylogenetic tree was constructed using Neighbor-Joining (NJ) method in MEGA10 (Saitou and Nei 1987, Kumar *et al*. 2018). In phylogenetic analyses, *Orthetrum sabina* (MF784360) served as an outgroup. In addition, the GenBank database provided additional genes that were incorporated into the analysis (Fig. 1).

Fig. 1. Voucher specimens of hesperiid butterflies collected from different locations in Bangladesh (Supplemental Figure).

RESULTS AND DISCUSSION

Thirteen species of hesperiid butterflies were collected from various regions of Bangladesh. Following the making of vouchers, genomic DNA was extracted from the voucher samples. Afterwards, an average of 659bp *COI* gene sequences was derived from this genomic DNA. Significant similarity confirmed the species identification after a BLAST homology search, and then submitted these genes to the NCBI's GenBank (Table 1).

Table 1. List of the species from which *COI* genes were sequenced and their GenBank accession numbers.

Species name	Subfamily	GPS Coordinates	Voucher No.	GenBank Accession No.
Parnara guttata	Hesperiinae	23°87'68.3"N 90°26'83.9"E	BFBSV 032	OQ119722
Parnara bada	Hesperiinae	23°87'63.9"N 90°26'79.2"E	BFBSV 033	OQ727117
Pelopidas agna	Hesperiinae	23°87'68.3"N 90°26'82.5"E	BFBSV 042	OQ119721
Borbo cinnara	Hesperiinae	23°87'63.9"N 90°26'81.1"E	BFBSV 047	OQ727118
Telicota colon	Hesperiinae	23°87'59.4"N 90°26'69.7"E	BBV 0324	OR099719
Suastus gremius	Hesperiinae	23°87'68.6"N 90°26'85.0"E	BBV 0052	OR099711
Iambrix salsala	Hesperiinae	23°87'58"N 90°26'83"E	MBV 0051	MK014748
Udaspes folus	Hesperiinae	23°87'58"N 90°26'83"E	BBV 0049	MT602627
Iton semamora	Hesperiinae	24°32'89"N 91°78'53"E	BBV 0073	MT606168
Tagiades japetus	Pyrginae	23°87'58"N 90°26'83"E	BBV 0022	MN186875
Pseudocoladenia dan	Pyrginae	24°32'72"N 91°78'49"E	BBV 0026	MK757465
Odontoptilum angulatum	Pyrginae	24°32'64.7"N 91°78'35.6"E	BBV 0036	MK778436
Burara oedipodea	Coeliadinae	24°26'55.0"N 91°92'12.8"E	BFBSV315	OP703518

Nucleotide analysis of COI gene

The analysis focused on a 585 bp fragment of the *COI* gene. The average base compositions of the *COI* sequences were found to be 40.01% T, 16.10% C, 30.21% A, and 13.66% G. A significant bias towards the Adenine and Thymine (A+T) nucleotide base pair was observed, accounting for 70.22% of the total. The nucleotide composition of the first, second and third codon positions of the *COI* fragment consisted of 59.99%, 57.86%, and 92.91% A+T content, respectively, as shown in Table 2.

Table 2. Basic statistics of the *COI* gene sequences of the thirteen hesperiid butterfly species.

Position	No. of sites	Empirical base frequencies (%)				A+T%
		T	C	A	G	
All Positions	585	40.01	16.10	30.21	13.66	70.22
First Position	195	27.57	15.30	32.42	24.69	59.99
Second Position	195	42.60	27.10	15.26	15.02	57.86
Third Position	195	49.96	5.91	42.95	1.26	92.91

Insects typically exhibit a significant bias towards A+T in their mitochondrial genomes (Simon *et al*. 1994). Furthermore, a bias towards A+T content was observed in various Lepidopterans, from *Pieris* to *Catopsilia*, which is consistent with the findings presented in our study (Table 2) (DeSalle and Hunt 1987, Li *et al*. 2015, Nie *et al*. 2018).

Pairwise distance analysis

The pairwise distance among thirteen hesperiid butterflies was determined by the MEGA10 software. The analysis of interspecific genetic divergence among the thirteen species ranged from 0.0532 to 0.1790%. *Pseudocoladenia dan* and *Udaspes folus* were found to have the greatest genetic distance at 0.1790%. In Table 3, *Parnara guttata* and *Parnara bada* were found to have the smallest genetic distance of 0.0532%. Zakharov *et al*. (2004) concurred with our present findings. Despite the fact that interspecies hybridization is a common occurrence in many butterfly species, it is conceivable that this narrow range of interspecific divergence is the result of its frequency (Win *et al*. 2015). Numerous species of *Papilio* have sequence divergences ranging from 0% to 1.2%, whereas in lepidopterans, sequence divergences greater than 2% are used to differentiate species (Hebert *et al*. 2003b, Zakharov *et al*. 2004). In the present analysis, the narrow range of interspecific

divergence (0.0532%) may be a result of the prevalence of low interspecies hybridization, a common occurrence among numerous butterfly species (Win *et al*. 2015).

Table 3. Percentage of pairwise distances among thirteen hesperiid species.

Species Name	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Parnara guttata												
Parnara bada	0.0532											
Pelopidas agna	0.1045	0.1084										
Borbo cinnara	0.1083	0.1163	0.0949									
Telicota colon	0.1184	0.1264	0.0966	0.1163								
Suastus gremius	0.1385	0.1324	0.1304	0.1283	0.1183							
Iambrix salsala	0.1203	0.1123	0.1065	0.1245	0.0947	0.1163						
Udaspes folus	0.1468	0.1573	0.1410	0.1471	0.1263	0.1468	0.1533					
Iton semamora	0.1123	0.1123	0.0988	0.1271	0.1244	0.1469	0.1124	0.1535				
Tagiades japetus	0.1447	0.1468	0.1531	0.1450	0.1406	0.1283	0.1347	0.1679	0.1325			
Pseudocoladenia dan	0.1552	0.1594	0.1386	0.1659	0.1468	0.1658	0.1470	0.1790	0.1641	0.1428		
Odontoptilum angulatum	0.1304	0.1344	0.1264	0.1304	0.1063	0.1449	0.1086	0.1266	0.1308	0.1124	0.1473	
Burara oedipodea	0.1406	0.1345	0.1489	0.1386	0.1203	0.1263	0.1263	0.1447	0.1345	0.1284	0.1574	0.1163

Phylogenetic analysis

A phylogenetic tree was constructed using the Neighbor-Joining (NJ) method by MEGA10. In the analysis, three clusters were obtained with the first cluster formed by *P. dan* and *T. japetus* of Pyrginae subfamily. The second cluster was formed by *P. guttata, P. bada P. agna, B. cinnara, I. semamora, I. salsala, T. colon* and *U. folus* of Hesperiinae subfamily (Fig. 2). While in the same cluster, *O. angulatum* of Pyrginae with *U. folus* of Hesperiinae is resided with weak support (Fig. 2).

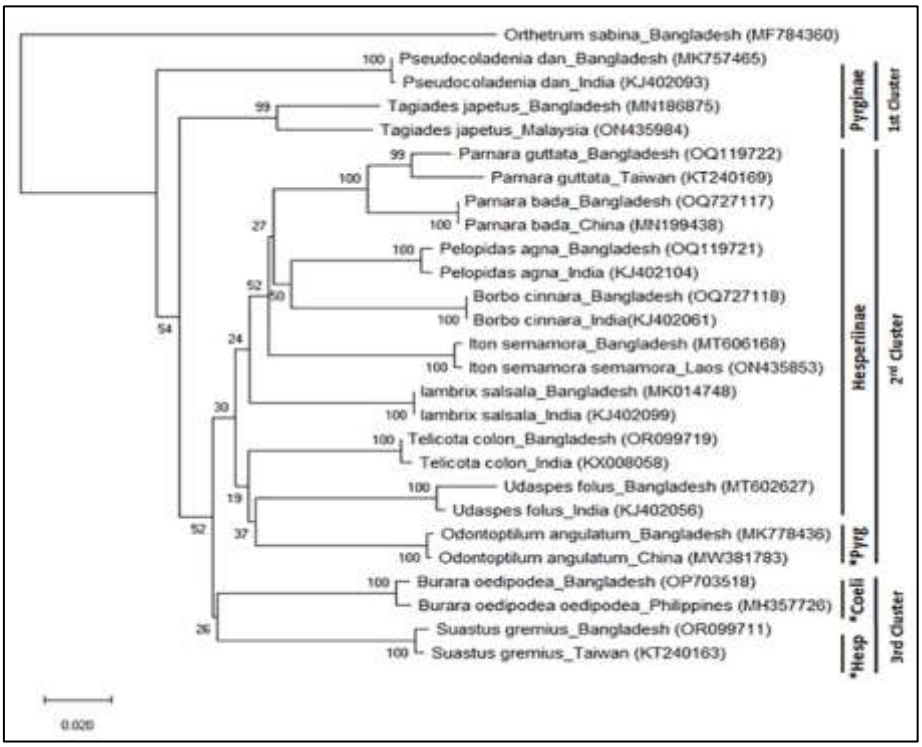

Fig. 2. Neighbor-Joining (NJ) tree of the hesperiid butterflies based on *COI* gene sequences. Bootstrap values are shown at the branching points. (*Pyrg= Pyrginae, *Coeli= Coeliadinae, *Hesp= Hesperiinae).

This result indicated that Pyrginae is paraphyletic, which is consistent with the findings obtained by Sahoo *et al*. (2016). They worked with 270 hesperiid genera, and their studies showed that the Pyrginae subfamily displayed inconsistent topologies because it was paraphyletic. Besides, in the

cladogram some exceptions for Pyrginae were described also by Warren *et al*. (2009). On the other hand, the third cluster was formed by *B. oedipodea* and *S. gremius* of Coelidae and Hesperiinae subfamilies, respectively with weak support (Fig. 2). Sahoo *et al*. (2016) showed in their analysis that Coeliadinae was sister to the rest of the subfamilies of Hesperridae. The same pattern of results depicted from our analysis, as Coeliadinae was sister to Hesperiinae, a major family in Hesperiidae, where *B. oedipodea* and *S. gremius* also remain together. In the present study, the analyses of the data support the following relationships: [(*P. dan*+*T. japetus*) + (*P. guttata*+*P. bada*+*P. agna*+*B. cinnara*+*I. semamora*+*I. salsala*+*T. colon*+*U. folus*+(*O. angulatum*) + (*B. oedipodea*+*S. gremius*)] (Fig. 2).

Nonetheless, the higher-level relationships among hesperiid butterflies, which comprise more than 4,000 species in approximately 567 genera, remain unresolved (Sahoo *et al*. 2016). According to Yuan *et al*. (2015), there is still considerable debate on the position of taxa belonging to the Eudaminae and Pyrginae subfamilies. Additionally, the question of whether Pyrginae is monophyletic or paraphyletic remains unresolved (Warren *et al*. 2009, Sahoo *et al*. 2016). Nevertheless, only 35% of the hesperiid genera in the world were employed by Warren *et al*. (2009) to examine molecular systematics studies because molecular data (mitochondrial DNA) is still not available. Finally, more molecular data are required to get a comprehensive scenario, and to reveal higher phylogenetic relationship among these subfamilies for Hesperiidae.

ACKNOWLEDGEMENTS

The authors express their gratitude for the financial assistance provided by the University Grants Commission of Bangladesh (CP No. 3424) for the project titled "Enhancement of Entomological Research Capability Using DNA Barcoding". Thanks to Sabbir Hussain Khan for photographing specimens. Sincere appreciation is also extended to the Wazed Miah from Science Research Center at Jahangirnagar University for providing the laboratory space.

REFERENCES

Bingham, C. T. 1905. *Fauna of British India. Butterflies*. Vol. I. Taylor and Francis, London, UK. 511 pp.

Bingham, C. T. 1907. *Fauna of British India. Butterflies*. Vol. II. Taylor and Francis, London, UK. 480 pp.

Bridges, C. A. 1994. *Catalogue of the Family-group, Genus-group and Species-group Names of the Hesperioidea (Lepidoptera) of the World.* Urbana, Ill, C. A. Bridges. 7 pp.

Brower, A. V. Z. 1996. A new mimetic species of Heliconius (Lepidoptera: Nymphalidae), from southeastern Colombia, as revealed by cladistic analysis of mitochondrial DNA sequences. *Zool. J. Linn. Soc*. **116**: 317-332.

Cock, M. J. W. 2010. The skipper butterflies (Hesperiidae) of Trinidad Part 17, Hesperiinae, Anthoptini and the remainder of Evans' Genera Group I. *J. Trinidad and Tobago Field Naturalists*. **2010**: 11-33.

DeSalle, R. and J. A. Hunt. 1987. Molecular evolution in Hawaiian drosophilids. *Trends Ecol. Evol*. **2**: 212-216.

Elías-Gutiérrez, M., M. V. Moreno, J. Topan and M. R. Young. 2018. Improved protocols to accelerate the assembly of DNA barcode reference libraries for freshwater zooplankton. *Ecol. Evol*. **8**(5): 3002-3018.

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

Hebert, P. D. N., A. Cywinska, S. L. Ball and J. R. deWaard. 2003a. Biological identifications through DNA barcodes. *Proc. R. Soc. Lond*. B. **270**: 313-321.

Hebert, P. D. N., S. Ratnasingham and J. R. Dewaard. 2003b. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. B*. **270**: 96-99.

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

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.

Li, N., H. Jia, Z. Liu, J. Tao, S. Chen, X. Li, Y. Deng, X. Jin, J. Song, L. Zhang, Y. Liang, W. Wang and J. Zhu. 2015. Molecular characterisation of phenylketonuria in a Chinese mainland population using next-generation sequencing. *Sci. Rep*. **5**: 157-169.

Nie, L., Y. Wang, D. Huang, R. Tao, C. Su, J. Hao and C. Zhu. 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.

Sahoo, R. K., A. D. Warren, N. Wahlberg, A. V. Brower, V. A. Lukhtanov and U. Kodandaramaiah. 2016. Ten genes and two topologies: An exploration of higher relationships in skipper butterflies (Hesperiidae). *Peer J*. **4**: e2653.

Saitou, N. and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mole. Biol. Evol*. **4**: 406-425.

Simon, C., F. Frati, A. Beckenbach, B. Crespi, H. Liu and P. Flook. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am*. **87**: 651-701.

Talbot, G. 1939. *The Fauna of British India Including Ceylon and Burma. Butterflies*. Vol. I. Taylor and Francis, London, USA. 600 pp.

Talbot, G. 1947. *The Fauna of British India Including Ceylon and Burma. Butterflies*. Vol. II. Taylor and Francis, London, USA. 506 pp.

Voss, E. G. 1952. On the classification of the Hesperiidae. *Ann. Entomol. Soc. America*. **45**: 246-258.

Warren, A. D., J. R. Ogawa and A. V. Z. Brower. 2008. Phylogeneticre lationships of subfamilies and circumscription of tribes in the family Hesperiidae (Lepidoptera: Hesperioidea). *Cladistics*. **24**: 642-676.

Warren, A. D., J. R. Ogawa and A. V. Z. Brower. 2009. Revised classification of the family Hesperiidae (Lepidoptera: Hesperioidea) based on combined molecular and morphological data. *Syst. Entomol.* **34**: 467-523.

Wahlberg, N., M. F. Braby, A. V. Z. Brower, R. D. Jong, M. M. Lee, S. Nylin, N. E. Pierce, F. A. H. Sperling, R. Vila, A. D. 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*. **272**: 1577-1586.

Wynter-Blyth, M. A. 1957. *Butterflies of the Indian Region*. Bombay Natural History Society, Bombay, India. 523 pp.

Win, N. Z., E. Y. Choi, D. J. Jang, J. Park and J. K. Park. 2015. Molecular comparison of the genus *Junonia* (Lepidoptera: Nymphalidae) in Myanmar. *J. Asia-Pacific Biod*. **8**(4): 287-294.

Xiao, J. H., N. X. Wang, Y. W. Li, R. W. Murphy, D. G. Wan, L. M. Niu, H. Y. Hu, Y. G. Fu and D. W. Huang. 2010. Molecular approaches to identify cryptic species and polymorphic species within a complex community of Fig Wasps. *PLoS ONE*. **5**(11): e15067.

Yuan, X., K. Gao, F. Yuan, P. Wang and Y. Zhang. 2015. Phylogenetic relationships of subfamilies in the family Hesperiidae (Lepidoptera: Hesperioidea) from China. *Sci. Rep*. **5**(1): 1-7.

Zakharov, E. V., M. S. Caterino and F. A. H. Sperling. 2004. Molecular phylogeny, historical biogeography, and divergence time estimates for swallowtail butterflies of the genus *Papilio* (Lepidoptera: Papilionidae). *Sys. Biol*. **53**: 193-215.

(Manuscript received on 29 August, 2023)