

Mitochondrial Barcodes of Three Malaysian Butterflies Originating from Taman Negara Endau Rompin Johor, Malaysia

J Y Oon¹, K R Kamarudin², N Ismail², M S S Omar², L Tokiman³, P N S Jahari^{1*}, N Ihsan¹, F Mohd Salleh¹

¹ Department of Biosciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia

² Faculty of Applied Sciences and Technology, Universiti Tun Huessein Onn Malaysia, 84600 Muar, Johor, Malaysia

³ Johor National Parks Corporation, Kota Iskandar, 79575 Iskandar Puteri, Johor, Malaysia

*Corresponding author: puterijahari22@gmail.com

Abstract. Butterflies are projected as reliable and economical biodiversity indicator. Traditionally, taxonomists identified and classified butterfly species based on highly similar and ambiguous morphological appearances which can result in problematic species identification process leading to misidentification of species. DNA barcoding has been developed for taxonomic identification of butterflies to species level. Nonetheless, this approach is hampered by the paucity of reference barcodes encompassing butterflies of all families and species in Malaysia. This study reported four novel DNA barcodes (two cytochrome oxidase subunit I (COI) and two cytochrome b, *cytb*) of Malaysian butterflies generated using Sanger sequencing. One barcode (DIB032 COI) supported the species identified by taxonomists whereas the other three barcodes (DIB034 COI, DIB034 *cytb* and DIB046/049 *cytb*) deduced that the butterfly species were either misidentified or unidentified up to species level due to the lack of reference barcodes in GenBank. The four novel DNA barcodes were deposited in GenBank under the accession number MT210226.1 (DIB032 COI), MT210227.1 (DIB034 COI), MT210228.1 (DIB034 *cytb*) and MT210229.1 (DIB046/049 *cytb*).

1. Data Description

To date, a comprehensive DNA barcode library was established by Wilson, Sing & Sofian-Azirun (2013) employing voucher specimens acquired from Museum of Zoology, Universiti of Malaya [4]. This library stored DNA barcodes for over 200 species of butterfly. COI barcodes from this database were deposited in GenBank. However, for the same butterfly species, it was found out only 18 *cytb* DNA barcodes are available in GenBank database including partial and complete mitogenomes. Hence, the issue of DNA reference barcode paucity in GenBank database was addressed to urge for continual development and expanding of reference database to cater the need for species identification [1]. This study provided four novel DNA barcodes for Malaysian butterflies with additional information presented in Table 1.

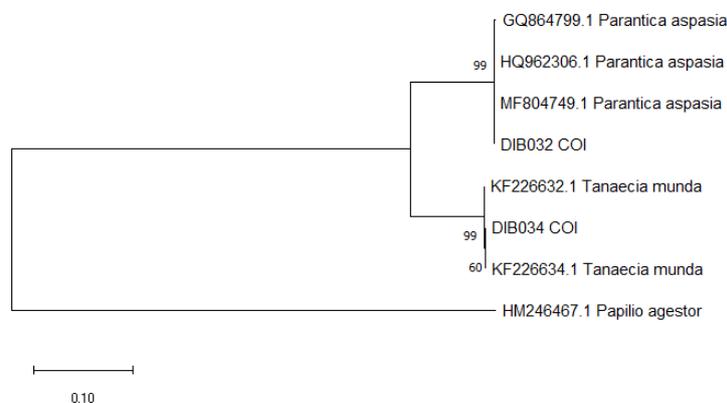


Table 1. Information of 4 novel DNA barcodes with reference species retrieved from BLASTn.

Sample ID	Type of barcode	Species identified by taxonomists	Size (bp)	Reference species retrieved from BLAST				
				Species name	Identity	Query cover	Size (bp)	GenBank accession no.
DIB032	COI	<i>Parantica aspasia</i>	649	<i>Parantica aspasia</i>	99.84%	99%	1450	GQ864799.1
				<i>Parantica aspasia</i>	100%	98%	658	MF804749.1
				<i>Parantica aspasia</i>	100%	98%	658	HQ962306.1
DIB034	COI	<i>Tanaecia palguna</i>	646	<i>Tanaecia munda</i>	100%	98%	658	KF226634.1
				<i>Tanaecia munda</i>	99.84%	98%	658	KF226632.1
DIB034	COI	<i>Tanaecia palguna</i>	423	<i>Euthalia monina</i>	99.05%	95%	420	EF095686.1
				<i>Euthalia duda</i>	94.29%	99%	420	EF095683.1
DIB046 /049	<i>cytb</i>	<i>Allotinus leogoron</i>	426	<i>Ariadne ariadne</i>	98.57%	98%	420	EF095678.1
				<i>Ariadne ariadne</i>	91/55%	100%	15179	KF990123.1
				<i>Ariadne ariadne</i>				

A phylogenetic tree was constructed for COI barcodes (Figure 1) and *cytb* barcodes (Figure 2) each with the novel barcodes generated from this study and other reference barcodes available in GenBank. The barcode sequences were aligned with Clustal W software followed by construction of Neighbour-Joining (NJ) phylogenetic tree using MEGA X program [2].

In the COI NJ phylogenetic tree (Figure 1.1), DIB032 COI barcode (accession no.: MT210226.1) formed a discrete clade supported 99% bootstrap value with 3 COI reference barcodes for *Parantica aspasia* (accession no.: GQ864799.1, HQ962306.1, MF804749.1) whereas DIB034 COI barcode (accession no.: MT210227.1) barcodes formed a clade of 99% clade node value with 2 COI reference barcodes for *Tanaecia munda* (accession no.: KF226634.1, KF226632.1). *Papilio agestor* was selected as an outgroup (accession no.: HM246467.1).

**Figure 1.** COI Neighbour-Joining phylogenetic tree.

In the *cytb* NJ phylogenetic tree (Figure 2), DIB034 *cytb* barcode (accession no.: MT210228.1) formed a 91% bootstrap value supported clade with *Euthalia monina* (accession no.: EF095686.1); consequently, forming a bigger clade of 68% with *Euthalia duda* (accession no.: EF095683.1). DIB046/049 *cytb* barcode (accession no.: MT210229.1) formed a discrete clade of 99% node value with 2 *cytb* reference barcodes for *Ariadne ariadne* (accession no.: KF095678.1, KF990123.1). *Papilio agestor* was selected as an outgroup (accession no.: HM246445.1).

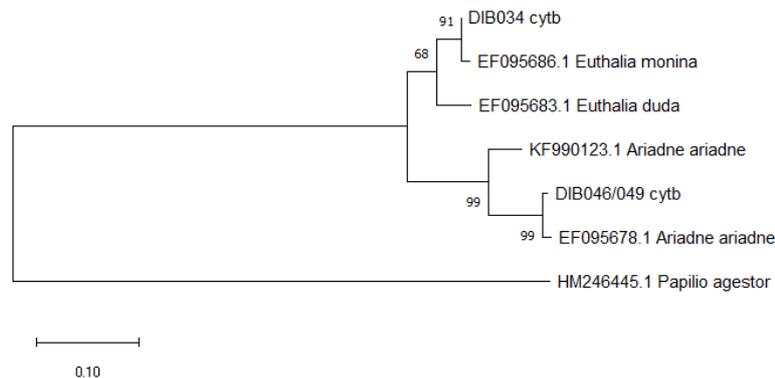


Figure 2. *cytb* Neighbour-Joining phylogenetic tree.

From the BLASTn results in Table 1, DIB032 COI barcode (MT210226.1) achieved a match with 99.84% identity similarity and 99% query coverage with *Parantica aspasia* (GQ864799.1). Hence, the BLASTn result supported the species of DIB032 identified by taxonomists from Universiti Tun Hussein Onn Malaysia (UTHM). The origin of the reference barcode was deduced as Southeast Asia as the next nearest match belong to the same species, MF804749.1 and HQ962306.1 were sourced from Thailand and Myanmar, respectively. Besides, this finding also supported that genetic diversity is present in the species complex due to geographical distribution [3].

For DIB034, the COI barcode (MT210227.1) had a closest match with *Tanaecia munda* (KF226634.1) sourced from Negeri Sembilan [4] with 100% identity similarity. However, the species was previously identified as *Tanaecia palguna*. This discrepancy may result from misidentification caused by the inconsistency of morphological features subjected to climate and temperature differences [5]. Thus, this finding indicate that DNA barcoding is reliable in butterfly species identification to overcome morphological-based approach limitations [6]. The morphological comparison can be seen in Figure 3. On the other hand, the *cytb* barcode (MT210228.1) of DIB034 reached the highest identity similarity of 99.05% and 95 query coverage with *Euthalia monina* (EF095686.1) collected from China. It should be noted that there is no COI and *cytb* reference barcode for *Tanaecia palguna* available in GenBank.

Lastly, DIB046/049 *cytb* barcode had the highest DIB identity similarity of 98.57% and 98% query coverage with *Ariadne Ariadne* (EF095678.1) sourced from China. All 4 novel barcodes were deposited in GenBank to augment the available reference database [1].



Figure 3. *Tanaecia palguna* (left) and *Tanaecia munda* (right).

2. Experimental Design

The three butterfly samples were collected from Taman Negara in March 2019 identified by UTHM taxonomists based on their morphological features. A pair of forelegs were used as the tissue sample for genomic DNA extraction using QIAGEN DNeasy® Blood & Tissue Kit. The gDNA were subjected to amplification via polymerase chain reaction (PCR) using COI [6] and *cytb* [7] primers. The gDNA and PCR products were analysed by agarose gel electrophoresis. The successful PCR amplicons were subjected to Sanger sequencing on forward sequences. The DNA barcodes were first revised using BioEdit and subjected to BLASTn in GenBank. Next, the DNA sequences were aligned using Clustal W software and viewed on Jalview program. Lastly, phylogenetic analysis was conducted with the construction of phylogenetic tree using Neighbour-Joining approach in MEGA X program. The inferred phylogenetic trees were tested for robustness by 1000 bootstrap re-samplings and presented as the node values.

3. Conclusion

In short, DNA barcoding plays an important role as an auxiliary approach to provide molecular-based evidence to overcome limitations of morphological-based evaluations in developing a reliable and accurate species identification methodology [6].

Acknowledgment

This study was supported by Ministry of Education, Malaysia under the Fundamental Research Grant Scheme (R.J13000.7845.4F963), Universiti Teknologi Malaysia (UTM) Research University Grant Tier 1 (Q.J13000.2545.18H49) and UTM Transdisciplinary Research Grant (Q.J13000.3554.05G69).

References

- [1] Madden M J, Young R G, Brown J W, Miller S E, Frewin A J and Hanner R H 2019 Using DNA barcoding to improve invasive pest identification at US ports-of-entry *PLoS One* **14**
- [2] Kumar S, Stecher G, Li M, Knyaz C and Tamura K 2018 MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution* **35** 1547-49
- [3] Gompert Z, Lucas L K, Buerkle C A, Forister M L, Fordyce J A and Nice C C 2014 Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants *Molecular ecology* **23** 4555-73
- [4] Wilson J-J, Sing K-W and Sofian-Azirun M 2013 Building a DNA barcode reference library for the true butterflies (Lepidoptera) of Peninsula Malaysia: What about the subspecies? *PLoS One* **8** e79969
- [5] MacLean H 2015 Adaptation to Climate and Climate Change in Rocky Mountain Butterflies: Morphology, Physiology, and Behavior
- [6] Čandek K and Kuntner M 2015 DNA barcoding gap: reliable species identification over morphological and geographical scales *Molecular ecology resources* **15** 268-77
- [7] Hebert P D, Penton E H, Burns J M, Janzen D H and Hallwachs W 2004 Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator* *Proceedings of the National Academy of Sciences* **101** 14812-17
- [8] Menezes R S T, Brady S G, Carvalho A F, Del Lama M A and Costa M A 2015 Molecular phylogeny and historical biogeography of the neotropical swarm-founding social wasp genus *Synocca* (Hymenoptera: Vespidae) *PLoS One* **10** e0119151