RAFFLES BULLETIN OF ZOOLOGY **69**: 262–271
Date of publication: 9 July 2021
DOI: 10.26107/RBZ-2021-0021
http://zoobank.org/urn:lsid:zoobank.org:pub:2D3C5E29-66E6-4F1E-8FF4-9501658C5349

Phylogenetic relationships of *Eurema* butterflies from Peninsular Malaysia inferred from CO1 and 28S gene sequences with emphasis on *Eurema hecabe*

Noor Azrizal-Wahid[1,2*], Mohammed Rizman-Idid[3] & Mohd Sofian-Azirun[4]

Abstract. The phylogenetic relationships among species of the genus *Eurema* from Peninsular Malaysia were reconstructed using nucleotide sequences of mitochondrial CO1 (307 bp) and nuclear ribosomal 28S DNA (471 bp). A total of twenty-eight sequences generated through PCR amplification for each gene region were used to construct the Maximum Likelihood (ML) and Maximum Parsimony (MP) phylogenetic trees. The separate taxonomic grouping of the genus *Eurema* and the genus *Gandaca*, and their close association is tested here together with sequences of other pierid butterflies obtained from GenBank. All trees reveal a strongly supported monophyletic group of *Eurema* conspecifics and well-resolved interspecific genetic distances, indicating the usefulness of the genetic markers in local species identification. The combined phylogenetic analyses of CO1 and 28S genes strongly supports a close relationship of *E. hecabe* with *E. blanda*, while *E. andersonii* is recovered as a sister taxon to *E. ada*.

Key words. *Eurema*, CO1, 28S, relationships, phylogenetic, Peninsular Malaysia

INTRODUCTION

The butterflies of the genus *Eurema* are classified under family Pieridae and typically recognised by the bright to pale lemon yellow coloured ground wings, bordered with black margin on the apical side of both forewings (Corbet & Pendlebury, 1992). Since their discovery by Yata in 1989, nine species have been recorded in Peninsular Malaysia which can be identified using taxonomic keys developed by Corbet & Pendlebury (1992). However, despite the availability of well-developed taxonomic keys for this genus in Malaysia, members of *Eurema* butterfly are notoriously difficult to identify due to their close morphological resemblance (Mal et al., 2014). This factor has limited the use of morphological characteristics for accurate species identification.

There have been several conflicts about the morphological classification and taxonomic position of *Eurema* species as shown by studies done in Malaysia (Corbet & Pendlebury, 1992) and Thailand (Jeratthitikul et al., 2009). Both studies have created competing arguments on the number of species, taxonomic position, and nomination of numerous subspecies. The disparities between the studies are resulting from the different selection of morphological characters as their main species identification criteria. Although both studies were conducted in different countries, Thailand, Malaysia, and also Singapore, are all located within the same geographical region. Hence the use of different classification keys should be revised and a single most appropriate key established eventually.

Concerning the systematics of *Eurema*, the taxonomic position of *Eurema hecabe* Linnaeus, 1758 is of particular interest because it was reported to exhibit several morphological variations of the black apical border pattern, and wing marking pattern on forewing (underside). These patterns were reported to differ seasonally and geographically (Yata, 1989; Corbet & Pendlebury, 1992; Jeratthitikul et al., 2009), and also by elevation (Azrizal-Wahid et al., 2015). For these reasons, identification and relationship status of *E. hecabe* among its congeners are disputable. Moreover, *E. hecabe* is the most widely distributed species and has highly variable wing markings, resulting in frequent misidentifications (Ek-Amnuay et al., 2007).

Despite the morphological description of *E. hecabe* having been revised (Yata, 1994; Kato & Yata, 2005; Jeratthitikul et al., 2009), its status remains unconfirmed. Although most subspecies of *E. hecabe* basically have two cell spots on forewing underside, Khan & Sahito (2012) found that the number of cell spots could vary from two to one to none when reared under different environmental conditions. In fact,

Accepted by: Hwang Wei Song

[1]Institute of Biological Sciences, Faculty of Science, Universiti Malaya, 50603 Kuala Lumpur, Malaysia

[2]Department of Biology, Faculty of Science, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia; Email: azrizal_wahid@yahoo.com (*corresponding author)

[3]Institute of Ocean and Earth Sciences (IOES), C308 IAS Building, Universiti Malaya, 50603 Kuala Lumpur, Malaysia; Email: rizman@um.edu.my

[4]Institute of Biological Sciences, Faculty of Science, Universiti Malaya, 50603 Kuala Lumpur, Malaysia; Email: sofian@um.edu.my

ISSN 2345-7600 (electronic) | ISSN 0217-2445 (print)

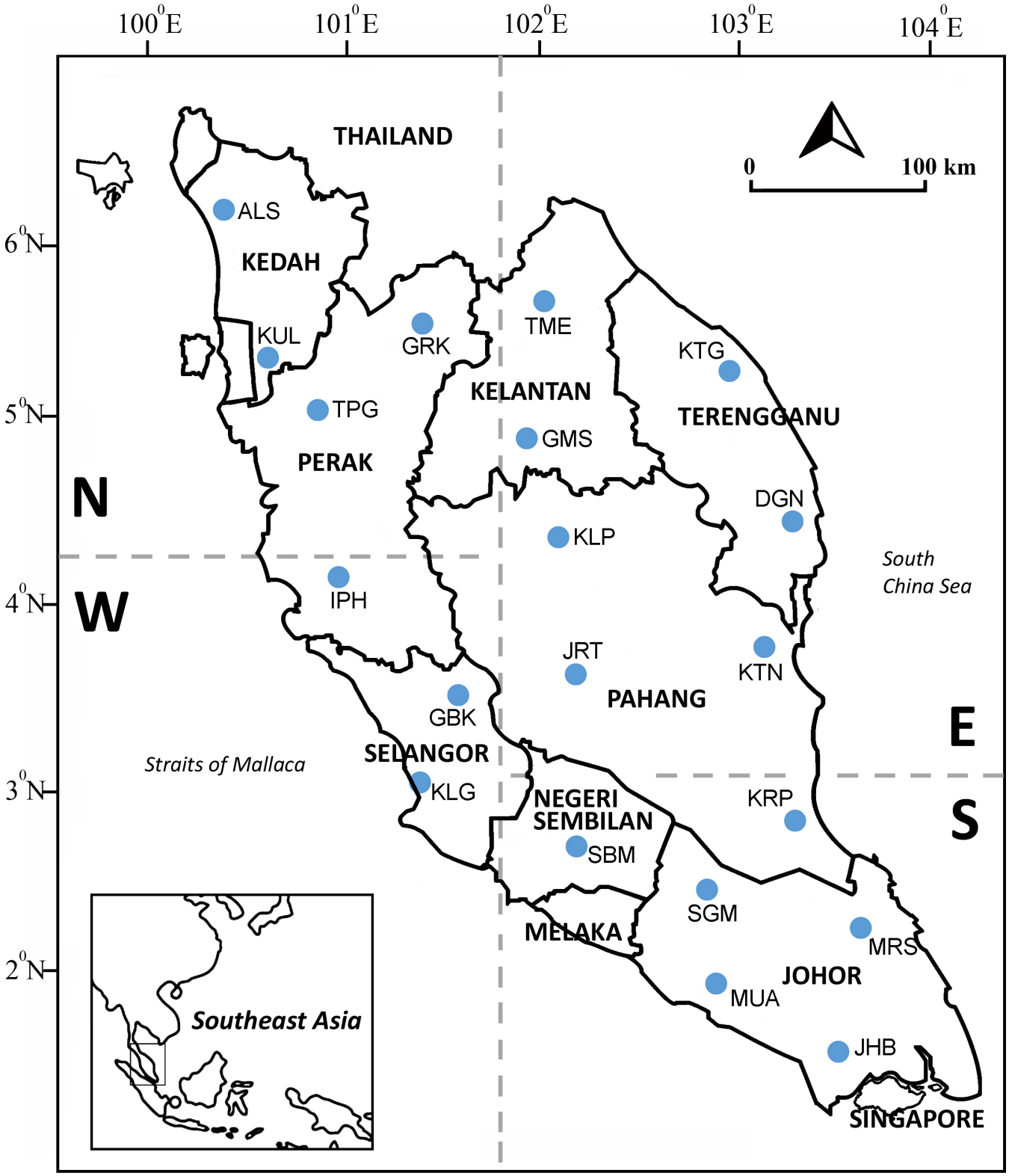

Fig. 1. The geographical sites where samplings have been conducted in Peninsular Malaysia. **N**, northern area; **E**, eastern area; **W**, western area; **S**, southern area. The dots indicate the distribution of various sampling sites in this study. Triplet letter represents the site code.

such morphological variations have resulted in the description of numerous subspecies by previous researchers, such as *Eurema hecabe mandarina* from Japanese mainland (Numata et al., 1972), *Eurema hecabe hobsoni* from Taiwan (Yata, 1989), and *Eurema hecabe contubernalis* from Malaysia and Singapore (Corbet & Pendlebury, 1992).

Nevertheless, Yata (1994) has proposed an integration of the *E. hecabe* subspecies into a single subspecies, namely *Eurema hecabe hecabe*. However, in more recent studies, *E. hecabe* was divided into two good species: the Y-type, *Eurema mandarina*, which has yellow fringe on forewing and is distributed around the mainland Japan (also in the mountains of Taiwan), and the B-type, *Eurema hecabe*, that has black fringe on the forewing and is widely distributed in tropical Asia including the Ryukyu Archipelago (Kato & Yata, 2005; Narita et al., 2007).

In Malaysia, much of the current understanding of the higher classification and interrelationships of *Eurema* butterflies was based on detailed morphological works conducted over twenty years ago (Corbet & Pendlebury, 1992). Since then, there has been a lack of studies to update the current status of the butterflies, particularly the validation of previously developed taxonomic keys for species identification.

In this study, therefore, we employed the analyses of two genes derived from mitochondrial cytochrome oxidase I (mtDNA CO1) and nuclear ribosomal 28S DNA (28S rDNA) in order to reveal the genetic relationships and taxonomic groupings of *Eurema* butterflies of Peninsular Malaysia, with particular emphasis on the taxonomic position of *E. hecabe*. This study is the first attempt to provide phylogenetic analysis on most *Eurema* species from a wide sampling range within Peninsular Malaysia, which also facilitates the DNA barcoding effort for species identification.

MATERIAL AND METHODS

Sample collection and preservation. Samples were collected from twenty sampling sites located within four different regions: northern, eastern, western, and southern Peninsular Malaysia (Fig. 1). The division of four distinct regions was designed in an attempt to test the effect of existing geographical barriers on genetic variation between sampled populations. The regions were divided according to the presence of mountain ranges and isolated patches separated by human developments that probably served as geographical barriers. The sampling sites were randomly distributed to maximise the coverage of the region.

Butterflies were caught by using insect sweep nets following the methods as described by Orr (2003) from the period of February 2015–March 2016. Caught butterflies were immediately kept in insect envelopes, and the front legs were removed and immediately preserved in absolute ethanol for molecular work. All the collected samples were preserved following standard pinning procedures to get the best display for species identification. Specimens were identified morphologically by using the classification keys derived by Corbet & Pendlebury (1992). The specimens are deposited as voucher specimens in Museum of Zoology, University of Malaya. Specimens of *Gandaca harina* Horsfield, 1829 were also included in analyses for group comparison due to their close resemblance and possibly sharing a common origin (Yamauchi & Yata, 2000).

DNA extraction, PCR amplification and sequencing. DNA was extracted from ethanol-preserved legs by using DNeasy® Blood and Tissue Extraction Kit (Qiagen, German), following the manufacturer's instructions. The PCR amplification of the CO1 gene was done using primers MLepF1 (5′-GCT TTC CCA CGA ATA AAT AAT A-3′) (Hajibabaei et al., 2006) and LepR1 (5′-TAA ACT TCT GGA TGT CCA AAA AAT CA-3′) (Hebert et al., 2004), whereas the 28S gene was amplified using primers None(D2)_F (5′-AGA GAG AGT TCA AGA GTA CGT G-3′) and None(D2)_R (5′-TTG GTC CGT GTT TCA AGA CGGG-3′) (Brower & DeSalle, 1998).

The PCR mixture was prepared for 25 µl containing 2.0 µl of DNA template, 1.0 µl of each forward and reverse primers (10 µM), 12.5 µl of PCR Premix (My Taq Red Mix), and 8.5 µl of double distilled water (ddH_2O). PCR amplifications for both CO1 and 28S were performed in Veriti® Thermal Cycler (Applied Biosystems) using the thermal cycling adapted from CO1_fast method (Wilson et al., 2011): initial heating at 98°C for 30 seconds, denaturing of DNA at 95°C for 2 min in 5 cycles, annealing of primers at 94°C for 30 sec, 45°C for 40 sec, 72°C for 1 min (performed for 35 cycles), DNA extension at 94°C for 30 sec, 51°C for 40 sec, 72°C for 1 min (35 cycles), final elongation at 72°C for 10 minutes and held at 4°C. DNA sequencing of PCR products was outsourced to MyTACG Bioscience Enterprise (Malaysia) and were sequenced in both directions (forward and reverse) using the same respective PCR primers.

Sequence variation and phylogenetic analyses. DNA sequence chromatograms were checked and edited using ChromasPro V7 (Technelysium Pty Ltd). Contiguous sequences were assembled from forward and reverse sequence reads using ClustalW software programme (Thompson et al., 1994) in MEGA7 (Tamura et al., 2007) with default parameters. There are a total of 40 sequences of *Eurema* butterflies including the isolates from outside the Malaysia region (Table 1), four sequences of the genus *Gandaca*, eleven sequences representing other pierid butterflies, and four sequences of outgroup species used for phylogenetic analyses of both genes. The sequences for other *Eurema* isolates, other pierid butterflies, and outgroup species of both genes were obtained from GenBank (Table 2). The outgroup species used for rooting both CO1 and 28S phylogenetic trees were from the genus *Graphium* (Wilson et al., 2014). Multiple sequences alignment and analysis for both genes were performed using MEGA7.

The aligned sequences were analysed for their properties to determine the nucleotide composition, variation, and genetic divergence by using MEGA7. The aligned DNA sequences

Table 1. List of *Eurema* species and corresponding GenBank accession numbers for the CO1 and 28S sequences used in phylogenetic analyses. Region codes represented as [NPM]: North of Peninsular Malaysia, [WPM]: West of Peninsular Malaysia, [EPM]: East of Peninsular Malaysia, [SPM]: South of Peninsular Malaysia, [KLM]: Kuala Lumpur, Malaysia, [CHN]: China, [IND]: India, [MYA]: Myanmar, [THA]: Thailand, [AUS]: Australia, [JPN]: Japan.

Genus	Species	Region code	Locality	Locality code	GenBank accession number	
					CO1	28S
Eurema	*blanda(1)*	NPM	Taiping, Perak	TPG	KT222700	KT222744
	blanda(2)	EPM	Jerantut, Pahang	JRT	KT222701	KT222749
	blanda(3)	WPM	Klang, Selangor	KLG	KT222702	KT222743
	blanda(4)	SPM	Mersing, Johor	MRS	KT222706	KT222746
	*blanda(5)**	JPN	Okinawa, Japan	–	AB969804	–
	*blanda(6)**	IND	Maharashtra, India	–	KJ423049	–
	*blanda(7)**	CHN	Hainan, China	–	HM175719	–
	*blanda(8)**	CHN	Shanxi, China	–	–	KM669510
	hecabe(1)	EPM	Dungun, Terengganu	DGN	KT222716	KT222753
	hecabe(2)	WPM	Gombak, Selangor	GBK	KT222710	KT222752
	hecabe(3)	NPM	Taiping, Perak	TPG	KT222715	KT222758
	hecabe(4)	SPM	Seremban, N. Sembilan	SBM	KT222714	KT222756
	*hecabe(5)**	IND	Tamil Nadu, India	–	HM386417	–
	*hecabe(6)**	CHN	Anhui, China	–	EF068257	–
	ada(1)	EPM	Kuala Rompin, Pahang	KRP	KT222727	KT222778
	ada(2)	SPM	Segamat, Johor	SGM	KT222718	KT222730
	ada(3)	NPM	Alor Setar, Kedah	ALS	KT222724	KT222732
	ada(4)	WPM	Gombak, Selangor	GBK	KT222721	KT222728
	sari(1)	SPM	Muar, Johor	MUA	KT222720	KT222766
	sari(2)	WPM	Gombak, Selangor	GBK	KT222680	KT222764
	sari(3)	NPM	Gerik, Perak	GRK	KT222681	KT222761
	sari(4)	EPM	Tanah Merah, Kelantan	TME	KT222677	KT222762
	*sari(5)**	MYA	Tanintharyi, Myanmar	–	MF804648	–
	simulatrix(1)	NPM	Kulim, Kedah	KUL	KT222697	KT222773
	simulatrix(2)	EPM	Kuala Lipis, Pahang	KLP	KT222694	KT222770
	simulatrix(3)	SPM	Mersing, Johor	MRS	KT222698	KT222772
	simulatrix(4)	WPM	Klang, Selangor	KLG	KT222695	KT222771
	andersonii(1)	NPM	Gerik, Perak	GRK	KT222685	KT222737
	andersonii(2)	WPM	Gombak, Selangor	GBK	KT222691	KT222739
	andersonii(3)	SPM	Johor Bahru, Johor	JHB	KT222689	KT222741
	andersonii(4)	EPM	Kuantan, Pahang	KTN	KT222686	KT222736
	andersonii(5)	THA	Trang, Thailand	–	HM395582	–
	*andersonii(6)**	CHN	Shanxi, China	–	–	KM669509
	tilaha(1)	NPM	Kulim, Kedah	KUL	KT222708	KT222776
	tilaha(2)	SPM	Seremban, N. Sembilan	SBM	KT222709	KT222775
	*brigitta(1)**	CHN	Hainan, China	–	HM175720	–
	*brigitta(2)**	CHN	Shanxi, China	–	–	KM669512
	*brigitta(3)**	IND	Tamil Nadu, India	–	KP119870	–
	*brigitta(4)**	AUS	Queensland, Australia	–	KF400836	–

*sequence obtained from GenBank

Table 2. List of other pierid species and outgroup sequences used in phylogenetic analyses and their corresponding GenBank accession numbers for the CO1 and 28S.

Genus	Species	Region code	Locality	Locality code	GenBank accession number	
					CO1	28S
Gandaca	*harina(1)*	EPM	K. Terengganu, Terengganu	KTG	KT222726	KT222777
	harina(2)	NPM	Taiping, Perak	TPG	KT222727	–
	*harina(3)**	THA	Trang, Thailand	–	HQ962120	–
	*harina(4)**	CHN	Yunnan, China	–	HM175728	–
Prioneris	*thestylis**	KLM	Kuala Lumpur, Malaysia	–	KF226607	–
	*thestylis**	CHN	Shanxi, China	–	–	KM669535
	*philonome**	KLM	Kuala Lumpur, Malaysia	–	KF226606	–
Cepora	*nadina**	KLM	Kuala Lumpur, Malaysia	–	KF226337	–
	*nadina**	CHN	Shanxi, China	–	–	KM669483
	*iudith**	KLM	Kuala Lumpur, Malaysia	–	KF226335	–
	*iudith**	CHN	Shanxi, China	–	–	KM669484
Delias	*agostina**	KLM	Kuala Lumpur, Malaysia	–	KF226391	–
	*agostina**	CHN	Shanxi, China	–	–	KM669463
	*hyparete**	KLM	Kuala Lumpur, Malaysia	–	KF226392	–
	*hyparete**	CHN	Shanxi, China	–	–	KM669462
Graphium	*sarpedon**	KLM	Kuala Lumpur, Malaysia	–	KC970131	KC970148
	*doson**	KLM	Kuala Lumpur, Malaysia	–	KC970122	KC970144
	*eurypylus**	KLM	Kuala Lumpur, Malaysia	–	KC970124	KC970145
	*agamemnon**	KLM	Kuala Lumpur, Malaysia	–	KC970100	KC970137

*sequence obtained from GenBank

were then subjected to evolutionary model selection by the jModeltest programme version 3.7 (Posada & Buckley, 2004). jModeltest result suggested the General Time Reversible model with gamma distribution (GTR+G; G=0.23) as the most suitable evolutionary model for the CO1 alignment data. For 28S alignment data, the programme suggested the Tamura-3-parameter model with gamma distribution (T92+G; G=2.58). The parameters of the best evolutionary model selected were incorporated into the phylogenetic tree analysis. In order to determine whether the two gene regions had different phylogenetic signals, we also analysed combined CO1-28S and portioned data separately. For the combined tree, to avoid any potential bias stemming from chimeric taxa, *E. brigitta* was excluded from the analyses due to unavailability of the CO1 and 28S sequences generated from the same individual in the GenBank database.

The phylogenetic tree analyses were performed using Maximum-Likelihood (ML) and Maximum Parsimony (MP) analysis methods by using MEGA7 with 1,000 bootstrapping replications. For ML, the tree was generated using the best model selected by jModeltest with default number of substitution type and transversion/transition ratio. MP tree was obtained using the Close-Neighbour-Interchange algorithm (CNI) (Nei & Kumar, 2000) with search level 3, in which the initial trees were obtained with the random addition of sequences (1,000 replicates). All positions containing gaps and missing data were eliminated. Only branches with over 70% bootstrap values are considered for analysis in all trees.

RESULTS

Collection of samples. Although sampling efforts were relatively extensive, this study only managed to obtain seven out of nine recorded *Eurema* species in Malaysia: *Eurema sari* Horsfield, 1829, *Eurema andersonii* Moore, 1886, *Eurema simulatrix* Semper, 1891, *Eurema blanda* Boisduval, 1836, *Eurema ada* Distant & Pryer, 1887, *Eurema tilaha* Horsfield, 1829, and *Eurema hecabe*. The two species not collected were *Eurema lacteola* Distant, 1886 and *Eurema brigitta* Cramer, 1780. For this reason, these two species were considered as rare or least commonly encountered species in this study because of the difficulty to find or capture during field sampling.

DNA sequence variation. A total of 26 sequences of *Eurema* for each gene region were generated from this study. The final sequence alignment length was 306 and 471 nucleotide

Table 3. Average value of genetic distance pairwise among sequences of seven *Eurema* species generated from this study, showing interspecies value obtained from CO1 sequences (values below diagonal) and 28S sequences (values above diagonal), and intraspecies value as indicated in bold for CO1 (values on left) and 28S (values on right). **[1]**: *E. sari*, **[2]**: *E. andersonii* **[3]**: *E. simulatrix*, **[4]**: *E. blanda*, **[5]**: *E. tilaha*, **[6]**: *E. hecabe*, **[7]**: *E. ada*.

Species	[1]	[2]	[3]	[4]	[5]	[6]	[7]
[1]	**0.002 / 0.000**	0.046	0.091	0.068	0.037	0.049	0.049
[2]	0.087	**0.002 / 0.001**	0.073	0.043	0.049	0.038	0.003
[3]	0.109	0.159	**0.003 / 0.001**	0.072	0.095	0.078	0.069
[4]	0.062	0.112	0.113	**0.001 / 0.000**	0.046	0.023	0.040
[5]	0.073	0.114	0.126	0.075	**0.000 / 0.000**	0.034	0.052
[6]	0.048	0.131	0.120	0.036	0.069	**0.003 / 0.001**	0.040
[7]	0.065	0.101	0.122	0.069	0.070	0.080	**0.003 / 0.000**

bases for partial CO1 and 28S sequences, respectively. There were no indels in the CO1 sequences with average base composition of A=29.9%, T=38.7%, C=17.2%, G=14.2%. Seventy-nine (25.6%) sites were variable, of which 64 sites were parsimony-informative. CO1 sequences were also observed to have AT bias (68.6%). Aligned 28S sequences showed the observation of indels at several nucleotide sites with most sizes ranging between 3 to 8 nucleotide bases. The average nucleotide composition was A=13.7%, T=20.6%, C=33.0% and G=32.8%. A total of 128 (29.6%) sites were variable and 65 sites were parsimony-informative.

For CO1 sequences, the intraspecies genetic distance ranged from 0.000 to 0.003 (Table 3). The highest interspecies genetic distance of 0.159 (15.9%) was observed between *E. andersonii* and *E. simulatrix* while the lowest interspecies pairwise value was observed between *E. blanda* and *E. hecabe* (3.6%). In 28S sequences, the genetic distance for intraspecies pairwise comparisons ranged from 0.000–0.001. The mean genetic distance value for interspecies pairwise was 0.049, with the highest value shown between pairwise of *E. simulatrix* and *E. tilaha* (0.095) and the lowest value between pairwise of *E. andersonii* and *E. ada* (0.003).

Phylogenetic tree analysis. The phylogenetic trees of both ML and MP analysis methods of CO1 sequences revealed eight distinct clades representing the eight *Eurema* species that are monophyletic with strong bootstrap scores (>95%) (Fig. 2). The only strongly supported sister taxa relationship was observed between *E. blanda* and *E. hecabe* with a bootstrap score of 87%. Although the relationships of other *Eurema* species were not well supported, the formation of sister taxon groups serves as useful preliminary hypotheses that deserve further testing. *Eurema ada* was found to be a sister taxon to *E. tilaha*, while *E. andersonii* was a sister taxon to *E. sari*, and *E. simulatrix* was a sister taxon to *E. brigitta*.

Alternatively, the analysis of 28S phylogeny also showed the formation of eight clades representing the eight *Eurema* species as observed in CO1 analyses, with the exception of a polyphyletic *E. andersonii* (Fig. 2). Herein, the tree reveals a close relationship supported with strong bootstrap score between *E. andersonii* and *E. ada*, with *E. ada* nested within an unresolved polytomy of *E. andersonii*. The genus *Gandaca* appears to be the next closest lineage to a monophyletic *Eurema* in both CO1 and 28S analyses, however the sister relationship to *Eurema* was not well supported as bootstrap scores were low (<50%).

Combined analysis of concatenated CO1-28S sequences showed similar grouping pattern as partitioned analysis of CO1 and 28S sequences. The tree strongly supported the monophyletic groups of all seven *Eurema* species from sequences generated in this study. The tree also reveals the sister taxa relationships between *E. hecabe* and *E. blanda*, and between *E. andersonii* and *E. ada*, which were supported with strong bootstrap scores (>80%) (Fig. 3). The sister-group association of the genus *Gandaca* to the genus *Eurema*, however, was no longer recovered, but a monophyletic *Eurema* remains strongly supported.

DISCUSSION

Sequence analysis and phylogenetic inferences. The nucleotide compositions for *Eurema* CO1 sequences have strong AT bias similarly as observed in partial CO1 sequences of the genus *Delias* (Müller et al., 2013) which is also found in most insect mtDNA CO1 sequences (Kandul et al., 2004; Tan et al., 2010; Jiang et al., 2015). There were sufficient magnitudes of barcoding gap as observed in high values of interspecies genetic distances and low values of intraspecies genetic distances of both CO1 and 28S sequences. Moreover, the monophyly of all *Eurema* species was well-established in most analyses, indicating the usefulness of the utilised molecular markers in delineation of *Eurema* species in this study.

The CO1 and 28S trees also revealed that conspecifics were recovered in their respective clades regardless of their sampling origin, except for one *E. andersonii* 28S sequence

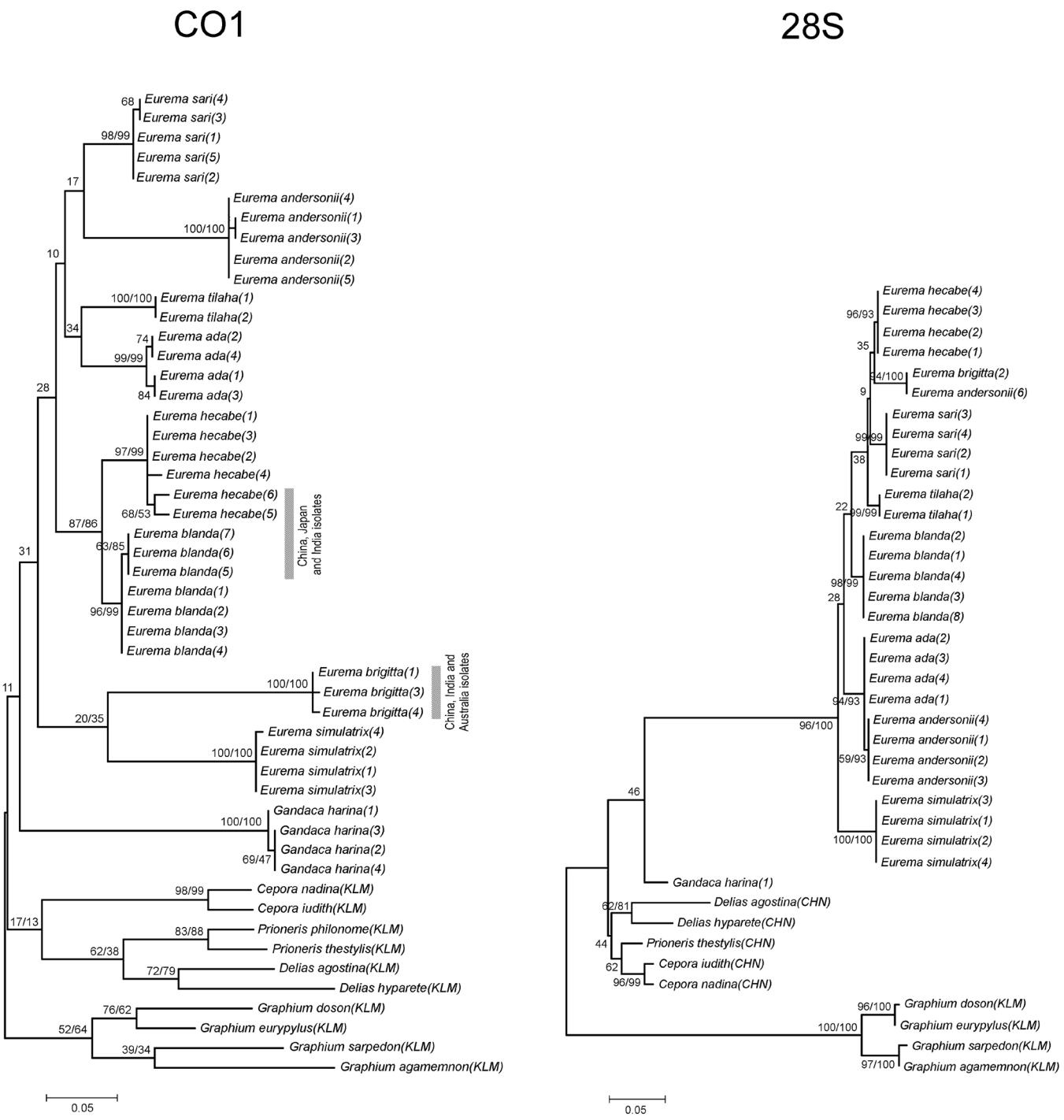

Fig. 2. Phylogenetic tree of Maximum-Likelihood method showing the comparison of phylogram as inferred from partial sequences of mtDNA CO1 and 28S rDNA genes. The bootstrap scores obtained from 1,000 replicates for ML/MP analyses are shown at the branching point. The trees were rooted with the genus *Graphium*.

grouped together with a *E. brigitta* sequence in the 28S phylogeny. These results confirm that most of the species defined based on morphological characters here are also monophyletic, thereby further supporting the validity of their species status and also the accuracy of the morphological diagnoses. The polyphyly of *E. andersonii* 28S sequences, which is not present in the CO1 phylogeny, is likely to be reflecting discordance between gene trees and species trees. This discordance can further be explained by incomplete lineage sorting due to past rapid speciation events within closely related *Eurema* species, or alternatively, a more recent genetic exchange due to a hybridisation event also cannot be ruled out here.

The lack of further resolution or phylogenetic structure within each species clade can be due to multiple reasons: a) the species studied here all had a relatively recent spread across Peninsular Malaysia, b) that genetic exchange is occurring on a regular level, suggesting that geographic distance and proposed physical barriers do not pose a problem, c) the speciation event between all the species studied here took place relatively recently, or d) the genetic markers used here

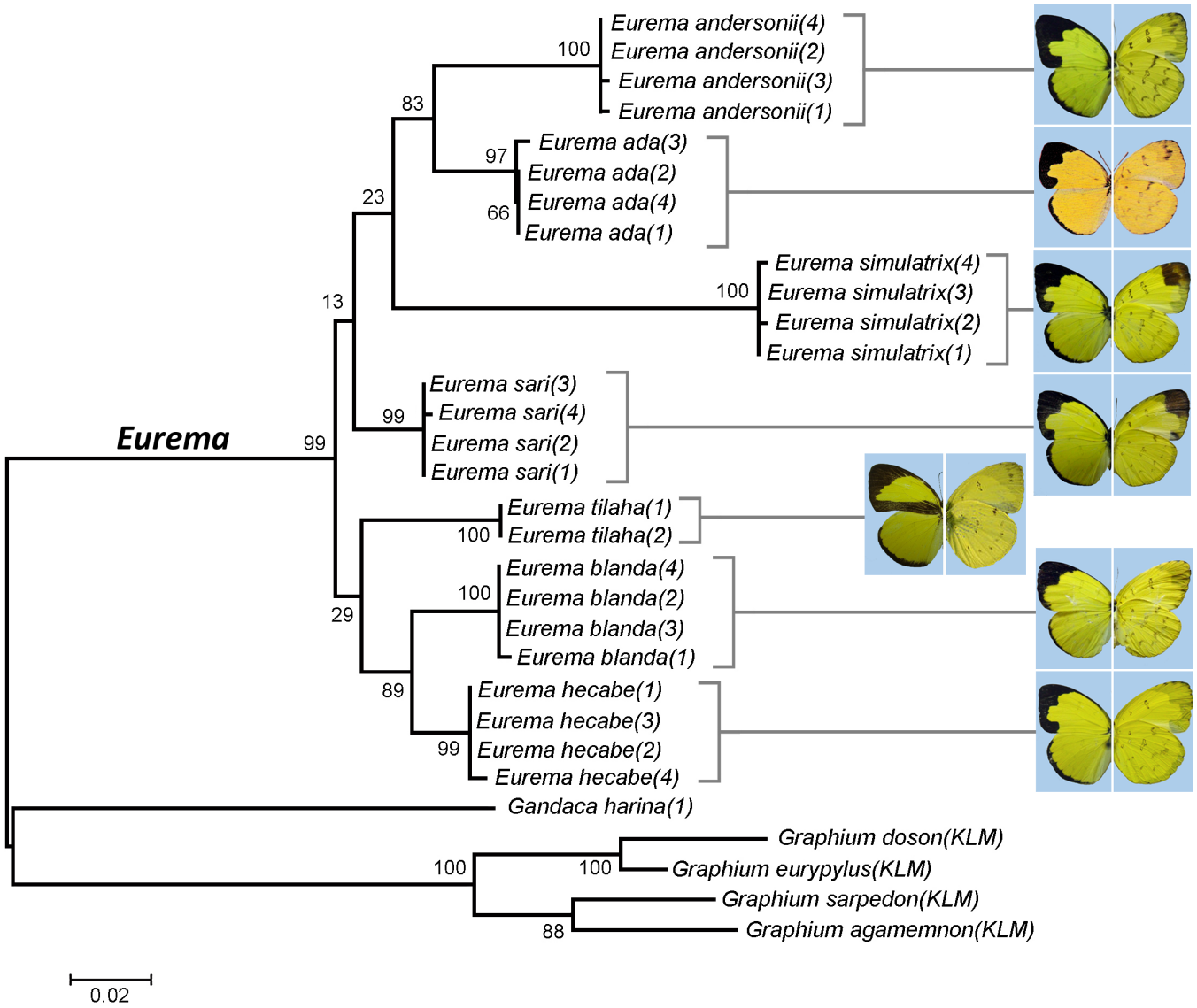

Fig. 3. Maximum Likelihood output phylogram for CO1-28S concatenated analysis showing seven major clades representing the seven *Eurema* species obtained from this study. Bootstrap scores are shown at the branching points. The tree was rooted with the genus *Graphium*. The butterfly figures show the comparison of morphology among the species corresponding to their respective clades. Figures of butterflies provided as upperside of the wings (left) and downside of wings (right).

are not fast evolving enough to capture population-level differences within this group of *Eurema* species. High gene flow among the *Eurema* populations in Peninsular Malaysia could be related to their migratory behaviour. Migratory butterflies are known to exploit host plants outside their normal distribution (Yata, 1989; Braby, 2000), and have been known to overcome geographical barriers, increasing the chance of gene flow between populations. Examples of reported migratory behaviour in *Eurema* include migration of *Eurema alitha* over the mountain ranges in Queensland, Australia (Dunn, 2007) and migration of approximately 135 kilometres into southern Australia (Nielsen, 2015).

The phylogenetic association of *Gandaca harina* with the genus *Eurema* was also noteworthy as observed in both single-gene tree analyses, as their close genetic relatedness was consistent with their morphological resemblance. These morphological similarities have led to the classification of members of the genus *Gandaca* (Horsfield) in the past under *Terias* Swainson, 1821, a subgenus of the genus *Eurema* Hubner, 1819. However, based on relatively slight differences in wing-shape and venation, Moore (1906) removed *G. harina* from *Terias* and established the genus *Gandaca* (Yamauchi & Yata, 2000). The present study shows possible genetic evidence for the separation of *Gandaca* from *Eurema*, but more complete taxon sampling and additional genes are needed for further confirmation.

Relationships amongst *Eurema* species. Although relationships were poorly supported between the *Eurema* species in CO1 analysis, the strongly supported monophyletic grouping of each *Eurema* species indicated the usefulness of the gene as an ideal barcoding marker in molecular identification of the species. Close relationships of these congeners, however, were contradicting with their appearance in which several morphological characters of their wings are distinct enough for species delimitation (Corbet & Pendlebury, 1992; Azrizal-Wahid et al., 2015). Further

study on integration between morphological character and genetic signal of these *Eurema* species is needed for clearer clarification on their relationships.

In the 28S phylogenetic analysis, the data revealed a strong sister taxa relationship between *E. andersonii* and *E. ada* which is also corroborated by the combined data analysis. The sister relationship between these two species may not be surprising as they are sometimes misidentified as the same species. Morphologically, both species share similar black distal margin pattern on the hindwing upperside, and pattern of apical border and apical patch on the upperside and underside of forewings (Corbet & Pendlebury, 1992). The close relationship between these species was also supported by the morphometric study of *Eurema* by using characters from body and wing regions (Azrizal-Wahid et al., 2016).

Relationship status of *Eurema hecabe*. The present study provides strong support for the sister relationship of *E. hecabe* and *E. blanda*. Morphologically, *E. hecabe* is different from *E. blanda* based on a reduced pattern of black apical border towards basal part (Corbet & Pendlebury, 1992). The black apical border towards basal part is a character unique to and possessed only by *E. blanda* among its congeners. However, the variation in the extent of the black apical border pattern found in *E. hecabe* can sometimes lead to its misidentification as *E. blanda* (Azrizal-Wahid et al., 2015).

Although morphological comparisons suggest *E. hecabe* to be closely related to *E. ada* due to the sharing of many aspects in wing elements including pattern of black apical border, number of cell spots, pattern of brown apical patch, and pattern of black distal margin (Corbet & Pendlebury, 1992), the combined CO1-28S phylogenetic analysis from our present study indicates a closer relationship between *E. hecabe* and *E. blanda*, strongly supported by the high bootstrap score. The close relationship between these two species was also reported in several previous studies using a similar fragment of the CO1 gene (Odagiri & Yata, 2005; Rajpoot et al., 2018). The studies yielded phylogenetic trees that grouped *E. hecabe* and *E. blanda* into the same clade consistently, strongly supporting their close relationship, in line with the findings of this study. Moreover, in another study, analyses using RAPD-PCR technique also reveal close affinity between *E. blanda* and *E. hecabe* (Tiple et al., 2010).

Although the present work involves the data from partial mtDNA CO1 and 28S rRNA, the information generated is invaluable in studies directed at the molecular identification of *Eurema* species. Phylogenetic analyses of both CO1 and 28S genes showed reciprocal monophyly of all *Eurema* species when combined. The combined analysis of CO1 and 28S data was able to clarify the relationship status of the most variant and controversial *E. hecabe* to be close to species *E. blanda* and confirmed the close position of highly remarkable *Eurema* species, *E. ada* and *E. andersonii*. Both genes appear to be useful markers for reconstructing the phylogeny of *Eurema* butterflies and are also appropriate DNA barcoding genes to identify species.

ACKNOWLEDGEMENTS

We sincerely thank the staff of Institute of Biological Sciences (ISB), University of Malaya, for the help towards completing this research. This research was fully funded by the University of Malaya, Kuala Lumpur, Malaysia through IPPP Grant (PV085-2011A). We also would like to thank the anonymous reviewers for the great comments in helping to improve the manuscript quality.

LITERATURE CITED

Azrizal-Wahid N, Sofian-Azirun M & Rizman-Idid M (2015) A review on morphological characterization, variation and distribution pattern of *Eurema* butterflies of Peninsular Malaysia. Journal of Biodiversity and Environmental Sciences, 6(3): 359–372.

Azrizal-Wahid N, Sofian-Azirun M & Rizman-Idid M (2016) The significance of wing and body morphometry in discriminating of six *Eurema* butterflies (Lepidoptera: Pieridae) of Peninsular Malaysia. Sains Malaysiana, 45(10): 1413–1422.

Braby MF (2000) The butterflies of Australia: their identification, biology and distribution. CSIRO Publishing, Collingwood, 976 pp.

Brower AVZ & DeSalle R (1998) Patterns of mitochondrial versus nuclear DNA sequence divergence among Nymphalid butterflies: The utility of wingless as a source of characters for phylogenetic inference. Insect Molecular Biology, 7: 73–82.

Corbet AS & Pendlebury HM (1992) The butterflies of the Malay Peninsula. Fourth Edition. Malayan Nature Society, Kuala Lumpur, 595 pp.

Dunn KL (2007) An extension to the known distribution of *Eurema alitha* (C. Felder & R. Felder) into south-eastern Queensland (Lepidoptera: Pieridae). Victorian Entomologist, 37(4): 58–61.

Ek-Amnuay P, Chiba H, Kimura Y, Inayoshi Y, Saito K, Seki Y & Uémura Y (2007) Corrigenda to "Butterflies of Thailand" (Ek-Amnuay, 2006). Yodoriga, 213: 2–20.

Hajibabaei M, Janzen DH, Burns JM, Hallwachs W & Hebert PDN (2006) DNA barcodes distinguish species of tropical Lepidoptera. Proceedings of the National Academy of Sciences of the United States of America, 103: 968–971.

Hebert PDN, Penton EH, Burns JM, Janzen DH & 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 of the United States of America, 101: 14812–14817.

Jeratthitikul E, Lewvanich A, Butcher BA & Lekprayoon C (2009) A taxonomic study of the genus *Eurema* Hübner [1819] (Lepidoptera: Pieridae) in Thailand. Natural History Journal of Chulalongkorn University, 9(1): 1–20.

Jiang Y, Yang Z, Wang X & Hou Y (2015) Molecular identification of sibling species of *Sclerodermus* (Hymenoptera: Bethylidae) that parasitize Buprestid and Cerambycid beetles by using partial sequences of mitochondrial DNA cytochrome oxidase subunit 1 and 28S ribosomal RNA gene. Plos One, 10(3): e0119573.

Kandul NP, Lukhtanov VA, Dantchenko AV, Coleman JWS, Sekercioglu CH, Haig D & Pierce NE (2004) Phylogeny of *Agrodiaetus* Hübner 1822 (Lepidoptera: Lycaenidae) inferred from mtDNA sequences of COI and COII and nuclear sequences of EF1-α: karyotype diversification and species radiation. Systematic Biology, 53(2): 278–298.

Kato Y & Yata O (2005) Geographic distribution and taxonomical status of two types of *Eurema hecabe* (L.) (Lepidoptera, Pieridae) in south-western Japan and Taiwan. Transactions of the Lepidopterological Society of Japan, 56: 171–183.

Khan D & Sahito ZA (2012) *Eurema hecabe* (Linnaeus, 1758) [Lepidoptera: Pieridae], a serious seedling pest of *Acacia stenophylla* A. Cunn. Ex. Benth., in Karachi. International Journal of Biology and Biotechnology, 9(3): 307–312.

Mal B, Memon N, Memon SA, Ali shah M & Turk JK 2014 First record and redescription of *Eurema hecabe simulata* M. (Lepidoptera: Pieridae) from Sanghar, Sindh, Pakistan. Journal of Entomology and Zoology Studies, 2(4): 93–98.

Moore F (1906) *Gandaca*. Lepidoptera Indica, 7(75): 33–36, pl. 536.

Müller CJ, Matos Maraví PF & Beheregaray LB (2013) Delving into *Delias* Hübner (Lepidoptera: Pieridae): Fine scale biogeography, phylogenetics and systematics of the world's largest butterfly genus. Journal of Biogeography, 40(5): 881–893.

Narita S, Nomura M, Kato Y, Yata O & Kageyama D (2007) Molecular phylogeography of two sibling species of *Eurema* butterflies. Genetica, 131: 241–253.

Nei M & Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, New York, 348 pp.

Nielsen JE (2015) A range extension for *Eurema alitha* (C. & R. Felder) (Lepidoptera: Pieridae) in Australia, with notes on the migratory behaviour of *Eurema* species in South-east Queensland. Australian Entomologist, 42(1): 15–18.

Numata A, Hokimoto K, Shimada A, Yamaguchi H & Takaishi K (1979) Plant constituents biologically active to insects. I. Feeding stimulants for the larvae of the yellow butterfly, *Eurema hecabe mandarina*. Chemical Pharmacy Bulletin, 27: 602–608.

Odagiri K & Yata O (2005) Phylogenetic relationships and biogeography of the *sari* group of Old World *Eurema* (Lepidoptera: Pieridae) based on the mitochondrial CO1 and ND5 sequences. Report on Insect Inventory Project in Tropic Asia (TAIIV): 451–460.

Orr D (2003) The constitution of nature. Conservation Biology, 17: 1478–1484.

Posada D & Buckley TR (2004) Model selection and model averaging in phylogenetics: Advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Systematic Biology, 53: 793–808.

Rajpoot A, Kumar VP & Bahuguna A (2018) DNA barcodes and insights into the phylogenetic relationships of butterflies of the genus *Eurema* (Pieridae) from Uttarakhand, India. BioRxiv, PPR20165. doi: 10.1101/242263

Tamura K, Dudley J, Nei M & Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution, 24: 1596–1599.

Tan SH, Rizman-Idid M, Mohd-Aris E, Kurahashi H & Mohamed Z (2010) DNA-based characterisation and classification of forensically important flesh flies (Diptera: Sarcophagidae) in Malaysia. Forensic Science International, 199(1): 43–49.

Thompson JD, Higgins DG & Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research, 22(22): 4673–4680.

Tiple AD, Padwad SV & Deshmukh VP (2010) Molecular characterization of morphologically similar four Pieridae butterflies (Lepidoptera: Insecta) by RAPD-PCR technique. International Journal of Pharma and Bio Sciences, 2: 1–10.

Wilson J, Karen-Chia H, Sing K & Sofian-Azirun M (2014) Towards resolving the identities of the *Graphium* butterflies (Lepidoptera: Papilionidae) of Peninsular Malaysia. Journal of Asia-Pacific Entomology, 17(3): 333–338.

Wilson JJ, Rougerie R, Schonfeld J, Janzen DH, Hallwachs W, Kitching I, Haxaire J, Hajibabaei M & Hebert PDN (2011) When species matches are unavailable are DNA barcodes correctly assigned to higher taxa? An assessment using sphingid moths. BMC Ecology, 11: 1–14.

Yamauchi T & Yata O (2000) Systematics and biogeography of the genus *Gandaca* Moore (Lepidoptera: Pieridae). Entomological Science, 3(2): 331–343.

Yata O (1989) A revision of the Old World species of the genus *Eurema* Hübner (Lepidoptera, Pieridae). Part I. Phylogeny and zoogeography of the subgenus *Terias* Swainson and description of the subgenus *Eurema* Hübner. Bulletin Kitakyushu Museum of Natural History, 9: 1–103.

Yata O (1994) A revision of the Old World species of the genus *Eurema* Hübner (Lepidoptera, Pieridae). Part IV. Description of the *hecabe* group (part). Bulletin of the Kitakyushu Museum of Natural History, 13: 59–105.