DNA BARCODING COMPLEMENTING MORPHOLOGICAL TAXONOMIC IDENTIFICATION OF MOSQUITOES IN PENINSULAR MALAYSIA

A Noor Afizah[1], MM Torno[2], J Nur Jannah[1], AH Azahari[1], M Khairul Asuad[1], WA Nazni[1] and HL Lee[1]

[1]Medical Entomology Unit, WHO Collaborating Centre for Ecology, Taxonomy and Control of Vectors of Malaria, Filariasis and Dengue, Institute for Medical Research, Kuala Lumpur, Malaysia; [2]Department of Medical Entomology, Research Institute for Tropical Medicine, Muntinlupa City, Philippines

Abstract. Morphology-based mosquito identification is a gold standard, which is inexpensive and requires minimal tools and/or equipment. Nonetheless, it requires a high level of expertise and misdiagnosis is common. In addition, quality of samples is not always sufficient for identification at the species level. Hence, we explored the potential use of a gene-based DNA barcoding utilizing cytochrome c oxidase subunit I (COI) gene for mosquito identification. Ninety mosquito samples belonging to 13 species of 5 genera pre-identified using taxonomic keys were subjected to PCR and sequencing of COI gene. Specific anatomical parts, such as body (thorax and abdomen), leg and wing were also subjected to PCR to test for the utility of such parts for DNA barcoding, obtaining varying success with wings being the least successful. BLAST search showed sequences were 94-99% in agreement with morphological identification. Neighbor-joining tree analysis of COI gene fragment sequences allowed determination of phylogenetic relationships and reliable computation of divergence between taxa. Our findings suggested that COI-based DNA barcoding is a useful tool to complement taxonomy-based identification of mosquito species.

Keywords: cytochrome oxidase subunit I, DNA barcoding, mosquito identification, taxonomy, Malaysia

INTRODUCTION

Conventional method of mosquito identification is primarily based on morphology, a method both inexpensive and requiring minimal tools and/or equipment. However, accurate taxonomic identification based on morphology requires a high level of expertise, gained through years of experience, and misdiagnoses are common especially among less-experienced taxonomists. It is also time-consuming and the quality of samples is not always sufficient for identification at the species level (Besansky *et al,* 2003; Hebert *et al,* 2003; Cywinska *et al,* 2006). Samples of adult mosquito specimens can be damaged during collection, transportation from field to laboratory and storage, thereby losing important diagnostic

Correspondence: A Noor Afizah, Medical Entomology Unit, WHO Collaborating Centre for Ecology, Taxonomy and Control of Vectors of Malaria, Filariasis and Dengue, Institute for Medical Research, Jalan Pahang, 50588 Kuala Lumpur, Malaysia.
E-mail: afizah@imr.gov.my

morphological features. Furthermore, comparative analysis of morphological characteristics of certain species is sometimes insufficient to resolve issues of taxonomy and this approach is unable to establish phylogenetic relationships of the samples (Milankov *et al*, 2009).

In the context of epidemiology, establishing phylogenetic relationships among vector species can assist in evaluation of vector competency and in implementation of interventions in relation to disease prevention and control. The most currently used identification keys are mainly specific for a few and particular developmental stages, such as adult females and 4th instar larval stages. In addition, members of a species complex with similar morphological features make identification insufficient through taxonomic keys alone (Cagampang-Ramos and Darsie 1970; Hebert *et al*, 2003; Kumar *et al*, 2007; Jinbo *et al*, 2011). Therefore, there is a need for alternative methods to complement the conventional method for mosquito identification.

An increasingly popular method for identification of both vertebrate and invertebrate taxa is through DNA barcoding. The "DNA barcode" method utilizes partial mitochondrial DNA sequences that are unique for every species and therefore allows species differentiation (Besansky *et al*, 2003). This method can provide an objective means of species identification through the use of short, standardized gene regions as internal species tags. Unlike morphological identification, the DNA barcoding technique allows non-experts of taxonomy to objectively identify species even from small, damaged or previously unrecognizable life stage samples (Hebert and Gregory, 2005). Mitochondrial (mt) genes have the advantages of relative ease of isolation and amplification, even from marginally preserved specimens, due to their intracellular abundance. Mt genes uni-parental inheritance and haploid nature result in a conserved region exhibited by a general lack of recombination with minimal insertions and deletions (indels), which increases its utility in population genetic studies (Zardoya and Meyer, 1996; White *et al*, 2008).

Among the mt protein-coding genes, cytochrome oxidase subunit I (COI) gene is the most commonly used barcode region in the identification process of many arthropods due to the higher rate of molecular evolution allowing differentiation between closely allied species (Young and Hebert, 2015). This gene is considered easier to align because it is a protein coding sequence that has no gaps within the alignment sequence (Zhang *et al*, 2012). Although COI gene is not applicable to identify all animal species, previous studies showed its applicability in identifying insects, including mosquitoes and ticks (Ruiz-Lopez *et al*, 2012; Taira *et al*, 2012; Chan *et al*, 2014; Che Lah *et al*, 2016).

In the present study, we explored the applicability of COI-based DNA barcoding in the identification of mosquito species in Malaysia by comparing the results obtained from 13 mosquito species of 5 genera to the morphology-based identification. The study also examined different mosquito parts *eg*, body (abdomen and thorax), leg and wing, for their utility for DNA barcoding purposes as a means of establishing a more reliable identification system for mosquito species found in peninsular Malaysia.

MATERIALS AND METHODS

Mosquito collection

A variety of mosquito species (adult stage) were collected from a number of

localities in peninsular Malaysia (Fig 1) using human landing catch and CDC light trap baited with CO_2 (Li *et al*, 2015). All samples collected were stored in 1.5 ml tubes and transported to the lab. All mosquito samples were identified morphologically by an experienced taxonomist using specific illustrated morphological taxonomic keys (Jeffery *et al*, 2012). Identified samples were stored in -20˚C freezer prior DNA extraction. Laboratory-reared mosquitoes *(Aedes aegypti*, *Anopheles maculatus* and *Culex quinquefasciatus)* maintained in the insectarium of the Institute for Medical Research (IMR), Kuala Lumpur, were also used as references specimens.

PCR amplification and DNA sequencing of mosquito COI gene

Individual mosquito or the respective mosquito parts [body (abdomen and thorax), leg and wing] were removed using clean sterile forceps and placed in individual tubes. Samples were homogenized using a sterile polypropylene pestle attached to hand-operated tissue grinder, then were incubated in 20 µl of (600 mAU/ml) proteinase K at 56°C with shaking for 3 hours. DNA was extracted using QIAamp® DNA Mini Kit (Qiagen™, Hilden, Germany) and stored at -20°C until used.

Primers used were C1-J-1718 (5′-GGAGGATTTGGAAATTGATTAGT-TCC-3′)

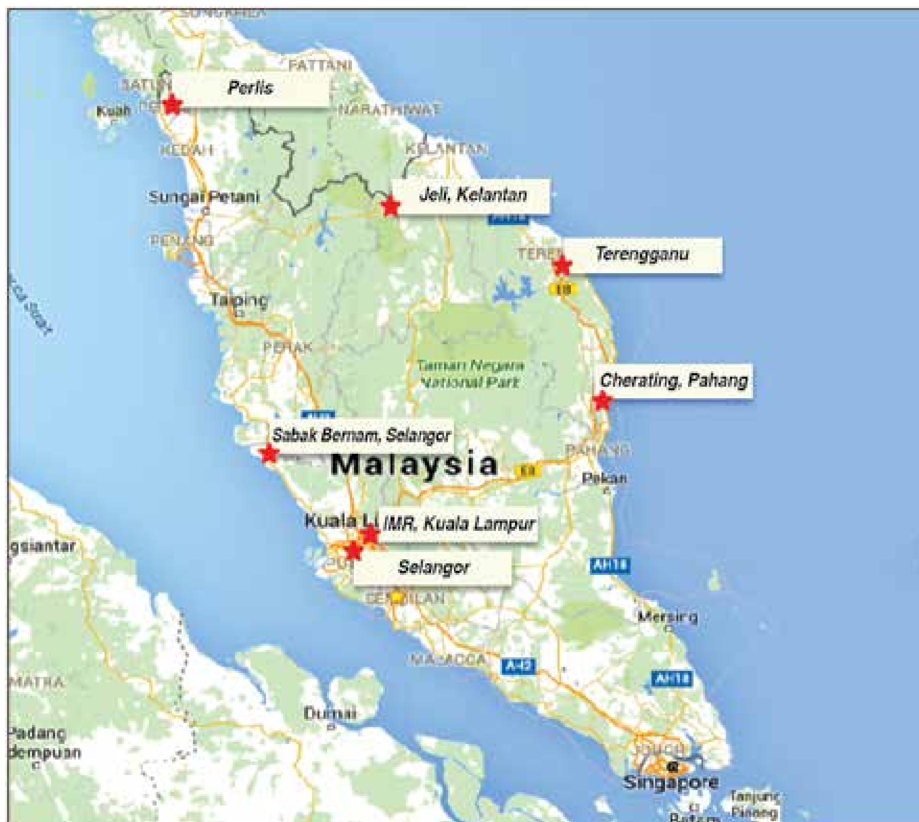

Fig 1-Map of peninsular Malaysia showing mosquito collection sites (indicated by stars).

and C1-N-2329 (5′-ACTGTAAATATATGATGTGCTCA-3′ (Simon *et al*, 1994). PCR mixture (25 µl) contained 5 µl of extracted DNA, 12.5 µl of MyTaq™ Mix (Bioline, Taunton, MA), 1 µl of each primer (10 µM) and 5.5 µl of deionized distilled (dd) water. For each PCR reaction, a negative control containing ddwater was included. Thermocycling was performed in Eppendorf EP S thermal cycler (Eppendorf, Hamburg, Germany) as follows: 95°C for 1 minute, followed by 30 cycles of 95°C for 15 seconds, 45°C for 15 seconds and 72°C for 15 seconds. Amplicons were separated by 1.5% agarose gel-electrophoresis and stained with GelStar™ Nucleic Acid Gel Stain (Lonza, Salisbury, MD). Amplicon was purified using QIAquick® Gel Extraction Kit (Qiagen™, Hilden, Germany) and sequenced by Medigene Sdn Bhd, Petaling Jaya, Selangor, Malaysia.

Phylogenetic analysis

All sequences were compared against those deposited at GenBank database using BLAST program (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Mosquito COI gene fragment sequences (610-bp) were aligned using Clustal-W algorithm (Thomson *et al*, 1994). Evolutionary distances of each species isolates based on COI gene fragment sequences were constructed using neighbor-joining tree based on Kimura-2P analysis with 1,000 bootstrap replicates performed in a MEGA 6.0 software program (Tamura *et al*, 2013). Sequences of previously published data from GenBank were aligned simultaneously. Sequences obtained from the study were deposited at GenBank with accession numbers shown in Fig 2.

RESULTS

Mosquito samples

Of 90 mosquito samples belonging to 13 species of 5 genera (3 *Aedes* spp, 2 *Anopheles* spp, 1 *Armigeres* sp, 4 *Culex* spp, and 3 *Mansonia* spp) analyzed, 75 samples were used as whole samples and 15 as specific anatomical parts [body (abdomen and thorax), leg and wing] to evaluate the utility of such parts for DNA barcoding (Table 1).

Molecular identification and phylogenetic analysis

After alignment and trimming, the 610-bp mosquito COI sequences showed high percent (94-100) homologies with previously published data from the GenBank. Our analysis indicted the collection of 3 *Aedes* spp (n = 18), 2 *Anopheles* spp (n = 26), 1 *Armigeres* sp (n = 6), 4 *Culex* spp (n = 34), and 3 *Mansonia* spp (n = 6) (Table 2). Phylogenetic analysis revealed individual mosquito species were clustered as distinct groups in accordance to their genus with strong bootstrap support (Fig 2). The phylogenetic tree showed clear separation of each mosquito genus present in the study samples, namely, *Aedes*, *Anopheles*, *Armigeres*, *Culex*, and *Mansonia*. The results are in agreement with morphology-based identification, the gold standard for taxonomy (Fig 2).

Comparison of mosquito parts for COI DNA barcoding

The utility of different mosquito parts [body (abdomen and thorax), leg and wing] in detection of COI gene across different mosquito species was evaluated. Amplifications of the 610-bp COI gene fragment from mosquito bodies, legs and wings were mostly successful (see Fig 3 for representative results) with body, leg and wing generating bright, faint and no bands in 1, 4 and 7; 8, 6 and 11; and 9, 3 and 12 specimens, respectively. This can be explained by different amounts of DNA and/or PCR inhibitor(s) in each preparation.

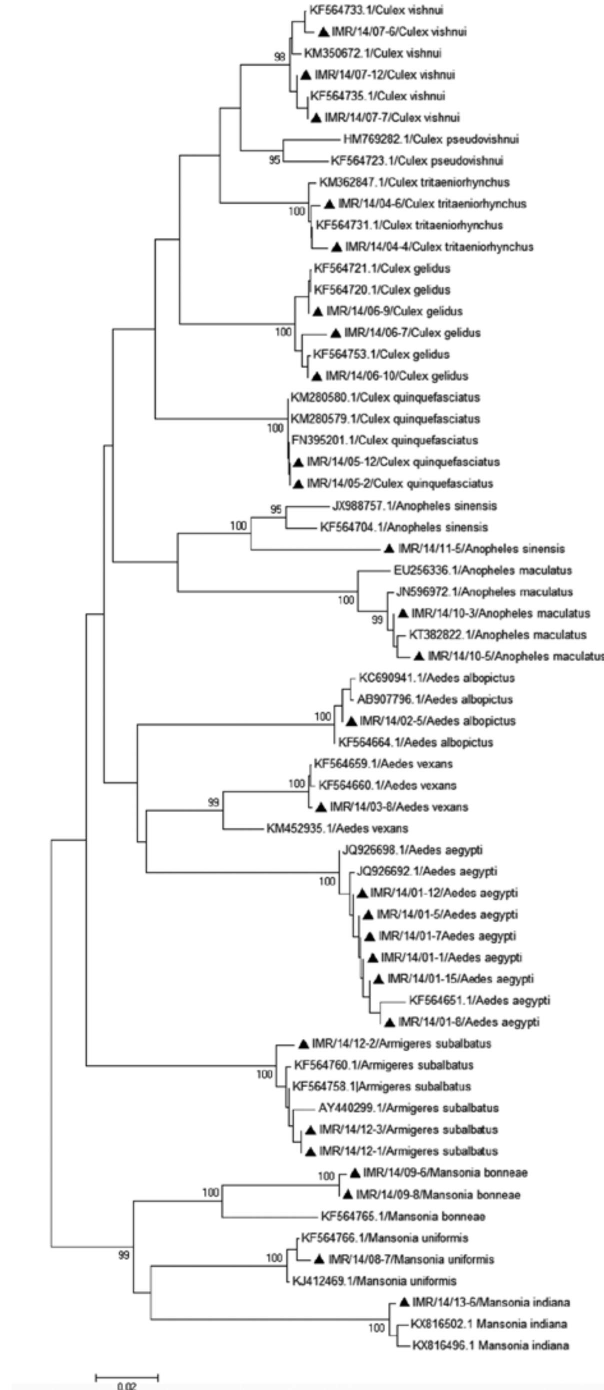

Fig 2-Phylogenetic tree of mosquito species based on COI gene 610-bp region. The tree was constructed using neighbor-joining method computed using Kimura 2-parameter method in MEGA 6.0 software (Tamura *et al,* 2013). Percent similarity is indicated by number at branch node (1,000 bootstraps). Scale bar denotes nucleotide substitution per site. GenBank accession number is shown in front of species name.

Table 1
Mosquito species, collection sites in peninsular Malaysia and number of specimens per species in the study.

Mosquito species	Collection site	Number of specimens
Aedes aegypti (lab strain)	IMR, Kuala Lumpur	7
Ae. albopictus	Selangor	5
Ae. vexans	Jeli, Kelantan	6
Anopheles sinensis	Sabak Bernam, Selangor	17
An. maculatus (lab strain)	IMR, Kuala Lumpur	9
Armigeres subalbatus	Cherating, Pahang	6
Culex tritaeniorhynchus	Selangor	10
Cx. quinquefasciatus (lab strain)	IMR, Kuala Lumpur	5
Cx. gelidus	Perlis	7
Cx. vishnui	Sabak Bernam, Selangor	12
Mansonia bonneae	Jeli, Kelantan	2
Ma. uniformis	Jeli, Kelantan	1
Ma. indiana	Terengganu	3
Total		90

DISCUSSION

Morphological characters approach utilizes direct observation of external phenotype differences and key morphological features between individual specimens. In this study, in order to provide a powerful and precise comparison between the conventional morphological-based identification technique and COI-based DNA barcoding approach, the species of collected mosquito were pre-identified prior to subjecting the samples for PCR amplification and sequencing of a 610-bp COI gene fragment and subsequent species identification by BLAST and phylogenetic analysis in comparison with known sequences deposited in GenBank. This method allowed excellent (94-100%) species identification based on GenBank database, comparable with results from morphology identification procedure.

The relatively low (94%) similarity of BLAST results for *Ma. bonneae* compared to other species was probably caused by intraspecific variation and in some cases, could be due to poor sequence quality. The latter seems the most likely reason in our study as the aligned sequence consists of several unassigned nucleotides, which may affect BLAST algorithm and thereby yielding a poorer percent identification. The limited number of samples especially for this species was also a limitation. Hence, as DNA sequence analysis relies heavily on the robustness of the available information for the target species in a gene library or BLAST database more samples need to be collected, sequenced and aligned for species confirmation, especially for *Mansonia* spp. Nonetheless, the sequences obtained in the present study

Table 2
Morphological identification and BLAST results of COI gene 610-bp fragment sequences against available sequences in GenBank database of mosquitoes collected in peninsular Malaysia.

ID code	Mosquito species (morphological identification)	Stage	Sex	Percent similarity with sequences from GenBank database
***Aedes* sp**				
IMR/14/01-12	*Ae. aegypti*	Adult	Female	100
IMR/14/01-5	*Ae. aegypti*	Adult	Male	100
IMR/14/01-7	*Ae. aegypti* (red eye)	Adult	Female	100
IMR/14/01-1	*Ae. aegypti* (red eye)	Adult	Male	100
IMR/14/02-5	*Ae. albopictus*	Adult	Female	100
IMR/14/03-8	*Ae. vexans*	Adult	Female	99
***Anopheles* sp**				
IMR/14/10-3	*An. maculatus*	Adult	Female	100
IMR/14/11-5	*An. sinensis*	Adult	Female	100
***Armigeres* sp**				
IMR/14/12-1	*Ar. subalbatus*	Adult	Female	100
***Culex* sp**				
IMR/14/04-6	*Cx. tritaeniorhynchus*	Adult	Female	99
IMR/14/05-12	*Cx. quinquefasciatus*	Adult	Female	100
IMR/14/06-7	*Cx. gelidus*	Adult	Female	100
IMR/14/07-6	*Cx. vishnui*	Adult	Female	100
***Mansonia* sp**				
IMR/14/08-7	*Ma. uniformis*	Adult	Female	99
IMR/14/09-6	*Ma. bonneae*	Adult	Female	94
IMR/14/13-6	*Ma. indiana*	Adult	Female	99

for *Ma. bonneae* which revealed only 94% sequence homology maybe compensated for the lack of information of genetic data to the BLAST library for this species. The accuracy and specificity of COI gene barcoding could be increased by combining with other DNA markers, *viz* 12S rDNA, 16S rDNA and internal transcribed spacer subunit 2 (Puslednik *et al*, 2012; Lv *et al*, 2014). In addition, a robust analysis of phylogeny of mosquito species with regards to other samples from different

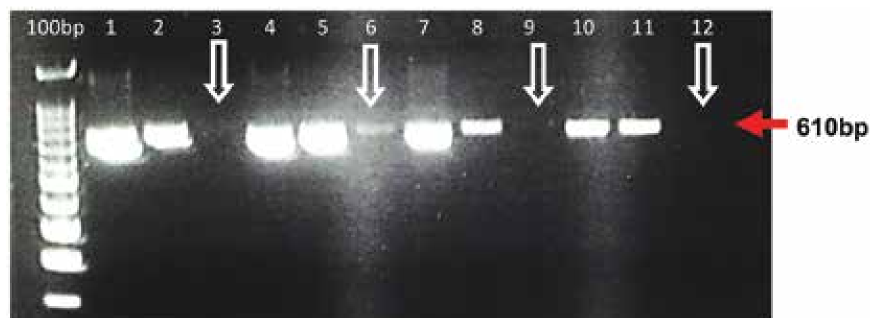

Fig 3-Gel-electrophoresis of cytochrome oxidase subunit I (COI) 610-bp amplicons from different parts of a mosquito. DNA was extracted using a commercial kit, amplified employing universal primers for mosquito COI gene target region, separated by 1.5% agarose gel-electrophoresis and visualized with GelStar™ Nucleic Acid Gel Stain. Left lane, DNA size markers; lane 1, *Aedes aegypti* body; lane 2, *Ae. aegypti leg*; lane 3, *Ae. aegypti* wing; lane 4, *Ae. albopictus* body; lane 5, *Ae. albopictus* leg; lane 6, *Ae. albopictus* wing; lane 7, *An. maculatus* body; lane 8, *An. maculatus* leg; lane 9, *An. maculatus* wings; lane 10, *Culex vishnui* body; lane 11, *Cx. vishnui* leg; lane 12, *Cx. vishnui* wing. Vertical arrow indicate faint or no band.

geographical regions will be useful in elucidating inheritance relationships of genes for refractoriness, insecticide resistance and genetically-determined ecological and behavioral traits important to disease transmission (Harbach, 2013).

Different parts of the mosquito were examined for their ability to generate the desired COI gene fragment. It is suggested that legs yielded sufficient amount of DNA that can be used for molecular studies. Wings were the least reliable mosquito part probably due to its chitinous nature with tissues limited only to the larger veins such as the costal vein. On the other hand, despite the smaller mass of the legs in relation to the body, both generated similar results with regards to production of distinct amplicon bands on gel-electrophoresis. This finding suggests DNA barcoding of rare mosquito specimens where samples need to be preserved, the legs can be used for molecular identification purposes as this anatomical part is almost equally successful in detection of DNA sequences as the body.

In conclusion, our studies on the use of COI gene barcoding for identification of mosquitoes from peninsular Malaysia highlights the advantages of DNA-based identification system over that of morphology-based method: (i) utilizes samples from all stages of the life cycle, (ii) recovers sufficient DNA from minimal parts of the specimen, (iii) identifies species from relatively short DNA sequences, and (iv) provides data for objective taxonomic decision.

ACKNOWLEDGEMENTS

The authors thank the Director-General of Health, Malaysia and the Director, Institute for Medical Research (IMR) for permission to publish this study, which was supported by a grant from SEAMEO-TROPMED and the Ministry of Health, Malaysia.

Table 3
DNA sequences obtained from GenBank for phylogenetic analysis.

Sample	Country	GenBank acession no.
Aedes aegypti	Mexico	JQ926698.1
Ae. aegypti	Singapore	KF564651.1
Ae. aegypti	Thailand	JQ926692.1
Ae. albopictus	USA	KC690941.1
Ae. albopictus	Singapore	KF564664.1
Ae. albopictus	Costa Rica	AB907796.1
Ae. vexans	Singapore	KF564659.1
Ae. vexans	Singapore	KF564660.1
Ae. vexans	Hungary	KM452935.1
Anopheles maculatus	China	KT382822.1
An. maculatus	India	JN596972.1
An. maculatus	India	EU256336.1
An. sinensis	India	JX988757.1
An. sinensis	Singapore	KF564704.1
Armigeres subalbatus	Singapore	KF564758.1
Ar. subalbatus	Singapore	KF564760.1
Ar. subalbatus	USA	AY440299.1
Culex gelidus	Singapore	KF564721.1
Cx. gelidus	Singapore	KF564720.1
Cx. gelidus	Singapore	KF564753.1
Cx. pseudovishnui	India	HM769282.1
Cx. pseudovishnui	Singapore	KF564723.1
Cx. quinquefasciatus	India	FN395201.1
Cx. quinquefasciatus	Austria	KM280580.1
Cx. quinquefasciatus	Austria	KM280579.1
Cx. tritaeniorhynchus	Singapore	KF564731.1
Cx. tritaeniorhynchus	India	KM362847.1
Cx. vishnui	Singapore	KF564733.1
Cx. vishnui	India	KM350672.1
Cx. vishnui	Singapore	KF564735.1
Mansonia bonneae	Singapore	KF564765.1
Ma. uniformis	India	KJ412469.1
Ma. uniformis	Singapore	KF564766.1
Ma. indiana	Thailand	KX816502.1
Ma. indiana	Thailand	KX816496.1

REFERENCES

Besansky NJ, Severson DW, Ferdig MT. DNA barcoding of parasites and invertebrate disease vectors: what you don't know can hurt you. *Trends Parasitol* 2003;1:545-6.

Cagampang-Ramos A, Darsie RF. Illustrated keys to the Anopheles mosquitoes of the Philippine islands. Vol 2. (Flight Used 5th ed). Manila: Malaria Eradication Training Center Manila, 1970.

Chan A, Chiang L-P, Hapuarachchi HC, *et al.* DNA barcoding: complementing morphological identification of mosquito species in Singapore. *Parasit Vectors* 2014;7:569.

Che Lah EF, Yaakop S, Ahamad M, George E, Md Nor S. Precise identification of different stages of a tick, *Ixodes granulatus* Supino, 1897 (Acari: Ixodidae). *Asian Pac J Trop Biomed* 2016;6:597-604.

Cywinska A, Hunter FF, Hebert PD. Identifying Canadian mosquito species through DNA barcodes. *Med Vet Entomol* 2006;20:413-24.

Harbach RE. The phylogeny and classification of *Anopheles*. In: *Anopheles* mosquitoes - New insights into malaria vectors. Rijeka: InTech, 2013:3-55.

Hebert PD, Cywinska A, Ball SL, deWaard JR. Biological identification through DNA barcodes. *Proc R Soc Lond Ser B Biol Sci* 2003;270:313-21.

Hebert PD, Gregory TR. The promise of DNA barcoding for taxonomy. *Syst Biol* 2005;5:852-9.

Jeffery J, Rohela M, Muslimin M, *et al.* Illustrated keys: some mosquitoes of peninsular Malaysia. Kuala Lumpur: University of Malaya Press, Malaysia, 2012.

Jinbo U, Kato T, Ito M. Current progress in DNA barcoding and future implications for entomology. *Entomol Sci* 2011;1:107-24.

Kumar NP, Rajavel AR, Natarajan R, Jambulingam P. DNA barcodes can distinguish species of Indian mosquitoes (Diptera: Culicidae). *J Med Entomol* 2007; 4:1-7.

Li CX, Smith ML, Fulcher A, Kaufman PE, Zhao TY, Xue RD. Field evaluation of three new mosquito light traps against two standard light traps to collect mosquitoes (Diptera: Culicidae) and non-target insects in northeast Florida. *Fla Entomol* 2015;98:114-7.

Lv J, Wu S, Zhang Y, *et al.* Development of a DNA barcoding system for the Ixodida (Acari: Ixodida). *Mitochondr DNA* 2014;25:142-9.

Milankov V, Petric D, Vujic A, Vapa L. Taxonomy, biology, genetic vulnerability and medical importance of *Ochlerotatus caspius* (Pallas, 1771) and *O. dorsalis* (Meigen, 1830) (Diptera: Culicidae). *Acta Entomol Serbica* 2009;1:195-207.

Puslednik L, Russell RC, Ballard JWO. Phylogeography of the medically important mosquito *Aedes (Ochlerotatus) vigilax* (Diptera: Culicidae) in Australasia. *J Biogeogr* 2012;3:1333-46.

Ruiz-Lopez F, Wilkerson R, Conn J, *et al.* DNA barcoding reveals both known and novel taxa in the Albitarsis group (*Anopheles*: *Nyssorhynchus*) of neotropical malaria vectors. *Parasit Vectors* 2012;5:44.

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

Taira K, Toma T, Tamashiro M, Miyagi I. DNA barcoding for identification of mosquitoes (Diptera: Culicidae) from the Ryukyu Archipelago, Japan. *Med Entomol Zool* 2012;63:289-306.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular biology and evolution. *Mol Biol Evol* 2013;30:2725-9.

Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673-80.

White DJ, Wolff JN, Pierson M, Gemmell NJ.

Revealing the hidden complexities of mtDNA inheritance. *Mol Ecol* 2008;17:4925-42.

Young MR, Hebert PDN. Patterns of protein evolution in cytochrome c oxidase 1 (COI) from the class Arachnida. *PLOS One* 2015;10:1-15.

Zardoya R, Meyer A. Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. *Mol Biol Evol* 1996;13:933-42.

Zhang A, Feng J, Ward RD, *et al.* A new method for species identification via protein-coding and non-coding DNA barcodes by combining machine learning with bioinformatic methods. *PLOS One* 2012;7:1-13.