PENSOFT
SENCKENBERG world of biodiversity
Leibniz Association

DNA barcodes and species boundaries of black flies (Diptera: Simuliidae) in Malaysia

Noor Izwan-Anas[1], Van Lun Low[1], Zubaidah Ya'cob[1], Emmanuel Y. Lourdes[1], Mohamad Rasul Abdullah Halim[2], Mohd Sofian-Azirun[2], Hiroyuki Takaoka[1], Peter H. Adler[3]

1 *Higher Institution Centre of Excellence, Tropical Infectious Diseases Research and Education Centre (TIDREC), Universiti Malaya, 50603, Kuala Lumpur, Malaysia*

2 *Institute of Biological Sciences, Faculty of Science, Universiti Malaya, 50603, Kuala Lumpur, Malaysia*

3 *Department of Plant and Environmental Sciences, Clemson University, Clemson, SC, 29634, USA*

https://zoobank.org/28C799F0-9AB1-4F42-AAAE-6CFB9FE2D8EB

Corresponding authors: Van Lun Low (vanlun_low@um.edu.my), Zubaidah Ya'cob (zyacob@um.edu.my)

Received 03 April 2023
Accepted 21 September 2023
Published 08 December 2023

Academic Editors Brian Wiegmann, Martin Wiemers

Citation: Izwan-Anas N, Low VL, Ya'cob Z, Lourdes EY, Halim MRA, Sofian-Azirun M, Takaoka H, Adler PH (2023) DNA barcodes and species boundaries of black flies (Diptera: Simuliidae) in Malaysia. Arthropod Systematics & Phylogeny 81: 931–943. https://doi.org/10.3897/asp.81.e104426

Abstract

Black flies play a prominent role in public health and the epidemiology of parasitic diseases of humans, domesticated and wild animals. Correct identification and a comprehensive survey are required to identify vector and pest species and thus understand their biological attributes which play a vital role in the monitoring program. DNA barcoding is an established molecular tool that provides rapid and accurate species identification. Our study strengthens the molecular database for black flies in Malaysia by adding 59 cytochrome *c* oxidase I sequences for 22 species, of which 14 are included for the first time. These sequences, combined with those in public databases, represent a total of 338 sequences for 52 Malaysian species, nearly 50% of which were collected from type localities. At the subgeneric level, barcode gap analysis most accurately identified species in the subgenus *Nevermannia* (92%), followed by *Simulium* s. l. (91%), and *Gomphostilbia* (81%). The remaining sequences were ambiguous and could not be distinguished from those of nearest neighbour species due to an overlap in genetic divergence and low genetic diversity, especially between insular species. Tree analyses indicate that certain species had incomplete lineage sorting and low mitochondrial signals. Possible cryptic species were indicated in the *Simulium* (*Gomphostilbia*) *batoense* and *S.* (*G.*) *epistum* species groups. Species delimitations were consistent with morphological identifications except in large species groups such as the *S.* (*G.*) *asakoae*, *S.* (*G.*) *batoense*, *S.* (*G.*) *epistum*, and *S.* (*Simulium*) *melanopus* groups. The use of type specimens or specimens collected from type localities (topotypes) in barcoding is strongly recommended for reference sequences to increase the reliability of the molecular database.

Keywords

cryptic species, cytochrome c oxidase I, *Gomphostilbia*, *Simulium*, vector

1. Introduction

Black flies are important in public health and play a significant role in the epidemiology of parasitic diseases of humans, domesticated animals, and wildlife. Certain species of the genus *Simulium* are vectors of *Onchocerca volvulus*, the sole causative agent of human onchocerciasis. This disease infects more than 15 million people in

Africa, Yemen, and Latin America, with approximately a million having lost sight (WHO 2022). In Southeast Asia, several pest species or species complexes have been reported, particularly in Thailand: *Simulium asakoae*, *S. chamlongi*, *S. doipuiense* complex, *S. tenebrosum* complex, *S. umphangense*, *S. nigrogilvum*, *S. nodosum*, *S. khelangense*, and *S. chumpornense* (Saeung et al 2020). Some of these black flies transmit filarial nematodes, such as *Onchocerca* spp., and blood protozoa, such as *Leucocytozoon* spp., among domestic animals and wildlife (Fukuda et al. 2003; Ishii et al. 2008; Jumpato et al. 2019; Pramual et al. 2020; Saeung et al. 2020; Takaoka et al. 2003; Thaijarern et al. 2019).

Baseline taxonomic information is available for many black flies in Southeast Asian countries (Takaoka et al. 2017a, 2017b, 2017c, 2018a, 2019), and their biodiversity has been documented in Indonesia, Malaysia, Thailand, and Vietnam (Adler et al. 2016; Takaoka and Davies 1995; Takaoka et al. 2015, 2018b; Hadi and Takaoka 2018). Currently, approximately 96 species of black flies have been recorded in Malaysia (Adler, 2022). However, the specific biting and vector species remain unknown. Additionally, there is insufficient molecular information available for many species, despite the availability of robust DNA barcode databases in neighboring countries such as Thailand, Vietnam, and Indonesia (database (Pramual et al. 2021; Hew et al. 2023; Putt et al. 2023). Ensuring accurate identification of Simuliidae is crucial for future public health research in this region, including Malaysia. By eliminating doubts of misidentification, researchers can confidently conduct studies and make informed decisions to address potential health concerns related to these insects. A reliable reference library for species ensures that each barcode can be used globally for identifications and comparisons (Ratnasingham and Hebert 2007; Sayers et al. 2020). In this case, type specimens or specimens collected from type localities (topotypes) are the most reliable for establishing the reference database (Kvist et al. 2010).

In the workflow application of DNA barcoding, organisms need to be identified and verified by taxonomists before DNA sequences can be deposited in the reference library (Hajibabaei et al. 2007; Janzen et al. 2005). This procedure ensures the reliability of the sequences for other researchers, especially non-taxonomists. With the combination of high-quality taxonomic descriptions and a strong reference library of specific taxa, DNA barcode applications can be swiftly performed with ease (Hajibabaei et al. 2006; Hebert et al. 2004; Janzen et al. 2005). Black flies are ideal candidates for DNA barcoding because they have been taxonomically studied throughout the world (Adler et al. 2004; Takaoka 2012, 2017; Hernández-Triana et al. 2015, 2017) and the taxonomy and distribution per country of each species have been summarized (Adler 2022). A molecular database for black flies has been developed in selected areas of the Nearctic Region (Rivera and Currie 2009), Mesoamerican countries (Hernández-Triana et al. 2015), Europe (Hernández-Triana et al. 2017; Ruiz-Arrondo et al. 2018), Thailand (Pramual et al. 2016, 2021; Pramual and Adler 2014), Indonesia (Hew et al. 2023) and Vietnam (Putt et al. 2023). To fill the knowledge gaps of species boundaries, the present study aims to establish a complete DNA barcode reference for Malaysian black flies, using topotypes and verified voucher specimens.

2. Materials and methods

2.1. Sample collection and identification

All specimens in this study were collected from streams across Malaysia (Table 1). Larvae and pupae were removed with fine forceps from their substrates (aquatic plants, rocks, twigs, fallen leaves, and plant roots). Live pupae were individually maintained in damp screw-capped tubes until adult emergence. All specimens were stored in 1.5-ml vials with 80% ethanol and were kept in the -20°c freezer at the Black Fly Gallery, Tropical Infectious Diseases Research & Education Centre, Universiti Malaya. Field sampling procedures and species identifications were carried out according to Takaoka (2003) and Adler et al. (2004).

2.2. DNA extraction and PCR amplification

A total of 85 black fly specimens were subjected to DNA extraction using the G-spin Total DNA Extraction Mini Kit (iNtRON™ Biotechnology, Inc., Seongnam, South Korea), according to the animal tissues protocol provided by the manufacturer. Polymerase chain reaction (PCR) was performed as described by Low et al. (2015). Briefly, COI sequences were amplified using primers LCO1490 (5'-GGT CAA CAA ATC ATA AAG ATA TTG G3') and HCO2198 (5'-TAA ACT TCA GGG TGA CCA AAA AAT CA-3') (Folmer et al. 1994). The 25-µL reaction volume contained 9.5 µL of sterile distilled water, 12.5 µL of 5x MyTaq™ Red Reaction Buffer (Bioline GmbH, Germany), 1 µL of each primer, and 1 µL of template DNA.

PCR amplification was performed with Applied Biosystems Veriti 96-Wll Thermal Cycler (Applied Biosystems, Inc., Foster City, CA, USA). All amplifications were confirmed using 1.5% agarose gel pre-stained with SYBR Safe (Invitron Corp., Carlsbad, CA, USA) run using a 100 bp DNA ladder (GeneDireX, Inc., Taiwan). Successful PCR products (59 specimens) were confirmed approximately at 700 bp and sent to Apical Scientific Sdn. Bhd., Selangor, Malaysia, for sequencing.

2.3. Tree analyses

A total of 361 sequences were included in data analyses. Of these, 338 were Malaysian sequences (Table 2), with 59 newly generated. These newly generated sequences were deposited in the NCBI GenBank under accession

Table 1. Species and collection details for 22 species of black flies from Malaysia used for barcoding.

Subgenus/species	n	Location	Coordinates	Date
***Nevermania* Enderlein**				
Simulium aureohirtum Brunetti,1991	3	Cameron Highland, Pahang	4°31.258′N, 101°24.247′E	23 Feb 2013
†*Simulium ledangense* Ya'cob, Takaoka & Sofian-Azirun, 2014	3	Mount Ledang, Johor	2°22.76′N, 102°36.615′E	11 Apr 2013
***Gomphostilbia* Enderlein**				
†*Simulium auratum* Takaoka, 2009	2	Murud, Sarawak	3°57.097′N, 115°33.075′E	11 Jun 2013
Simulium aziruni Takaoka, Hashim & Chen, 2012	1	Tasik Kenyir, Terengganu	5°0.5518′N, 102°42.1472′E	23 Sep 2017
Simulium barioense Takaoka, 2008	3	Mesilau, Sabah	6°2.133′N, 116°35.835′E	19 Jun 2014
†*Simulium hiroyukii* Ya'cob & Sofian-Azirun, 2015	2	Murud, Sarawak	3°55.365′N, 115°30.5066′E	13 Jun 2013
Simulium kelabitense Takaoka, 2008	3	Bakalalan, Sarawak	3°57.419′N, 115°37.057′E	16 Jun 2013
Simulium pegalanense Smart & Clifford, 1969	3	National Park, Pahang	4°33.282′N, 102°18.995′E	13 Sep 2013
†*Simulium sarawakense* Takaoka, 2001	3	Pueh, Sarawak	NA	28 Aug 2008
†*Simulium terengganuense* Takaoka, Sofian-Azirun & Ya'cob, 2012	3	Pasir Raja, Terengganu	4°33.985′N, 102°57.429′E	7 Jun 2013
Simulium varicorne Edwards, 1929	1	Negeri Sembilan	NA	29 Dec 2010
***Simulium* Latreille**				
Simulium alberti Takaoka, 2008	3	Murud, Sarawak	3°57.277′N, 115°33.308′E	10 Jun 2013
Simulium beludense Takaoka, 1996	1	Mesilau, Sabah	3°50.104′N, 115°36.521′E	9 Apr 2014
Simulium bishopi Takaoka & Davies, 1995	1	Cameron Highland, Pahang	4°18.420′N, 101°19.658′E	24 Dec 2012
	2		4°26.723′N, 101°22.979′E	25 Dec 2012
Simulium brevipar Takaoka & Davies, 1995	2	Raub, Pahang	4°23.715′N, 101°36.443′E	27 Dec 2012
	1		4°26.723′N, 101°22.979′E	22 Feb 2013
Simulium grossifilum Takaoka & Davies, 1995	1	Cameron Highland, Pahang	4°23.165′N, 101°22.334′E	25 Dec 2012
†*Simulium hackeri* Edwards, 1928	1	Cameron Highland, Pahang	4°34.956′N, 101°20.717′E	27 May 2012
	1			31 Mac 2012
	1			27 Jun 2012
	1			26 Dec 2012
†*Simulium hirtinervis* Edwards, 1928	1	Cameron Highland, Pahang	4°22.220′N, 101°21.512′E	25 Dec 2012
	2			27 May 2013
Simulium jeffreyi Takaoka & Davies, 1995	3	Tapah, Perak	4°16.316′N, 101°19.022′E	24 Dec 2012
Simulium malayense Takaoka & Davies, 1995	3	Cameron Highland, Pahang	4°22.220′N, 101°21.512′E	25 Dec 2012
†*Simulium murudense* Takaoka, Ya'cob & Sofian-Azirun, 2015	3	Murud, Sarawak	3°55.6084′N, 115°30.8434′E	13 Jun 2013
Simulium perakense Takaoka, Ya'cob & Sofian-Azirun, 2018	3	Batu Gajah, Kelantan	5°45.075′N, 101°58.816′E	2 Feb 2015
	2	Janda, Baik, Pahang	3°18.2167′N, 101°52.5′E	25 Jul 2011
n, total number of sequences. †, samples collected from type localities. NA, not available				

numbers OQ152321–OQ152378 and OQ540484. The deposited sequences and associated information are also accessible and available at the Global Biodiversity Information Facility (GBIF) database (https://www.gbif.org/dataset/c80987f7-f87a-4ae3-a2cc-ccd59bc951e8). Reference sequences of Thailand and Vietnam specimens

Table 2. Malaysian black flies (n = 52) used for DNA barcoding, according to subgenus, with intraspecific genetic distances. Species with newly generated sequences are in bold.

Subgenus	Species	Intraspecific distance (average)	n
Nevermannia	***S. aureohirtum***†	0.25–1.02 (0.68)	3
	S. borneoense‡	4.21	2
	S. ledangense†‡	0.00–0.25 (0.10)	5
	S. pairoti‡	0.00–1.54 (0.71)	16
Gomphostilbia	*S. angulistylum*‡	0.00–0.25 (0.17)	3
	S. asakoae‡	0.00	3
	S. auratum‡	8.38	2
	S. aziruni†	—	1
	S. barioense	2.60–3.94 (3.31)	3
	S. brinchangense‡	0.25–0.77 (0.51)	3
	S. cheongi	0.00–2.59 (0.90)	45
	S. decuplum	0.00–1.28 (0.85)	3
	S. duolongum	0.00	4
	S. gombakense	0.51–2.33 (1.72)	3
	S. hiroyukii‡	1.02	2
	S. izuae‡	0.25–1.02 (0.68)	3
	S. johorense‡	—	1
	S. kelabitense	0.00	3
	S. leparense‡	0.00	3
	S. lurauense‡	0.77–1.28 (1.02)	3
	S. parahiyangum	0.00	4
	S. pegalanense	0.51–7.24 (4.99)	3
	S. roslihashimi	0.00–0.51 (0.34)	3
	S. sarawakense‡	0.77–3.67 (2.43)	3
	S. sazalyi‡	0.00–2.33 (1.27)	8
	S. sheilae‡	0.00	3
	S. sofiani‡	0.00	3
	S. tanahrataense‡	0.00	3
	S. terengganuense‡	0.00–3.11 (2.08)	3
	S. trangense	0.00–0.77 (0.51)	3
	S. varicorne	—	1
	S. whartoni	0.51–1.28 (0.84)	3
Simulium	***S. alberti***	0.25–2.33 (1.55)	3
	S. argentipes	0.51	2
	S. beludense	0.00–0.25 (0.17)	3
	S. bishopi	1.02–1.54 (1.28)	3
	S. brevipar	0.51–0.77 (0.68)	3
	S. crassimanum	—	1
	S. grossifilum	—	1
	S. hackeri†‡	0.00–0.51 (0.20)	8
	S. hirtinervis‡	0.25–0.77 (0.51)	3
	S. jeffreyi†	0.00–2.33 (0.99)	36
	S. kiuliense	0.00	6
	S. laterale	—	1
	S. maklarini	—	1
	S. malayense†	0.00–3.38 (1.02)	8
	S. mirum‡	0.00–2.33 (0.93)	11
	S. murudense‡	0.00	3
	S. nigripilosum	0.00–1.80 (1.20)	4
	S. perakense	0.25–2.60 (1.29)	5
	S. tani	—	1
	S. vanluni	0.00–2.59 (0.73)	84

n, total number of sequences. †, species with newly generated sequences and sequences retrieved from GenBank. ‡, samples from type localities.

retrieved from GenBank were included for comparison. All sequences were assembled, aligned, and trimmed to 395bp in BioEdit 2.7.5 (Hall et al. 2011). *Simulium* (*Asiosimulium*) *furvum* (MZ508532) was used as an outgroup. Neighbor Joining (NJ) analysis was performed using Kimura's two-parameter substitution model in MEGA 11 version 11.0.11 (Tamura et al. 2021) with 1000 bootstrap replicates. Maximum likelihood (ML) analysis was performed using the webserver RAxML (Kozlov et al. 2019), with a default setting of the GTR substitution model with proportion invariant sites and 100 bootstrap replicates. Bayesian analysis was run using MrBayes version 3.2.7 (Ronquist et al. 2012). The GTR + G + I substitution model was suggested as the best model by jModeltest2 (Darriba et al. 2012). Bayesian analysis was performed on 30 million generations of Markov Chain Monte Carlo (MCMC) and the tree was sampled every 100th generation, with the first 25% of trees discarded as burn-in. Both jModeltest2 and MrBayes were performed on the online server Cyber Infrastructure for Phylogenetic Research (CIPRES) (Miller et al. 2011). All trees were visualized in FigTree v1.4.4 and edited in Interactive Tree of Life (iTOL) (Letunic and Bork 2021) and Adobe Illustrator 2020.

2.4. Genetic distance and species discrimination analyses

Pairwise genetic distance was calculated using the Kimura 2-parameter model in MEGA 11 version 11.0.11 (Tamura et al. 2021). The distribution of genetic distances was calculated by using the pairwise distance function in TaxonDNA (Meier et al. 2006). The presence of a barcode gap for each species was determined by plotting the farthest conspecific distance versus nearest neighbor (NN) and the mean conspecific distance versus NN (Robinson et al. 2009). The accuracy of each species barcode was tested using Best-Matching (BM), Best-Close Matching (BCM), and All Species Barcode (ASB) functions in TaxonDNA, based on the Kimura 2-parameter model.

2.5. Species delimitation analyses

Two different species delimitation methods were used: Assemble Species by Automatic Partitioning (ASAP) (Puillandre et al. 2021) and Generalized Mixed Yule-coalescent (GMYC) (Fujisawa and Barraclough 2013), which are distance- and tree-based methods (Guo and Kong 2022), respectively. ASAP was run on a web-based server using a Kimura 2-parameter (K80) model with default settings. GMYC analysis was performed using the ultrametric tree generated from BEAST2 version 2.6.6 (Bouckaert et al. 2014, 2019) with the GTR + G + I substitution model selected using jModeltest2. The analysis was confirmed using Tracer v1.7.1 (Rambaut et al. 2018) with an Effective Sampling Size (ESS) of more than 200. The output tree was analyzed in TreeAnnotator 2.6.6, with 10% burn-in. All analyses were carried out on the CIPRES online server. GMYC analysis was performed with the software package "splits" in RStudio 2022.02.1 Build 461.

3. Results

3.1. Pairwise distances and DNA barcode gaps

Pairwise intraspecific divergences for all Malaysian sequences ranged from 0.00% to 8.38% (Fig. 1). The lowest divergence (0%) was found in nine species: *Simulium kelabitense*, *S. murudense*, *S. tanahrataense*, *S. asakoae*, *S. sofiani*, *S. leparense*, *S. kiuliense*, *S. parahiyangum*, and *S. duolongum*. The highest intraspecific divergence was recorded in *S. auratum* at 8.38%, followed by *S. pegalanense* with a mean of 4.99%. Interspecific genetic distances ranged from 0.00% to 22.73%, with the lowest interspecific distances (0%) between *S. mirum* and *S. kiuliense*, *S. cheongi* and *S. whartoni*, and *S. sazalyi* and *S. parahiyangum*. The highest divergence (22.73%) was between *S. borneoense* and *S. pegalanense*. Overlapping of intra- and interspecific divergence was 0.00–8.38%, covering 3.34% of all intra- and interspecific regions.

A scatter plot of the data demonstrates that almost 80% of all species had a DNA barcode gap, whereas the rest did not due to overlap of the farthest conspecific with the nearest neighbour species (Fig. 2). Overall, nine species did not express a DNA barcode gap (seven species in the subgenus *Gomphostilbia* and two in the subgenus *Simulium*), whereas all species in the subgenus *Nevermannia* had a gap. The seven species of *Gomphostilbia* included one from the *S. asakoae* species group (*S. lurauense*), three from the *S. batoense* species group (*S. parahiyangum*, *S. pegalanense*, and *S. sazalyi*), and three from the *S. epistum* species group (*S. auratum*, *S. cheongi*, and *S. whartoni*). The two species in the subgenus *Simulium* were *S. kiuliense* (*S. nobile* species group) and *S. mirum* (*S. melanopus* species group).

3.2. Efficacy of species discrimination

The overall percentage of correct identification was 88.46% for Best Match (BM) and 87.27% for Best-Closed Match (BCM) but only 75% for All Species Barcode (ASB) (Table 3). Twenty-four sequences (7.1%) were ambiguous for BM and BCM, whereas 67 sequences (19.82%) were ambiguous for ASB. The subgenus *Nevermannia* had the highest correct percentage in all functions, with more than 90% for BCM and ASB and 100% for BM. Only *S. borneoense* sequences were below the threshold value at 3.95%.

The subgenus *Gomphostilbia* had the lowest percentage of correct species identifications for all three functions, particularly ASB with less than 50%. All 45 sequences of *S. cheongi* and *S. whartoni* were ambigu-

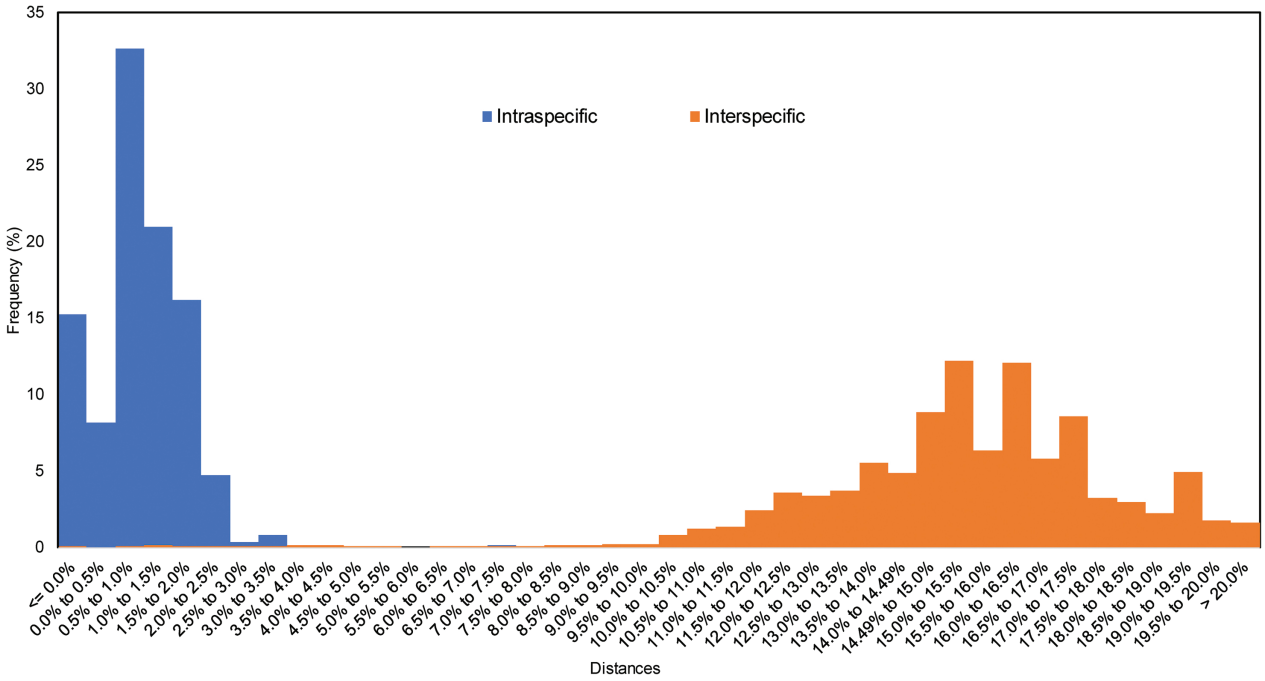

Figure 1. Distribution of intra- and interspecific genetic distances of 52 species of Malaysian black flies, based on 338 COI sequences.

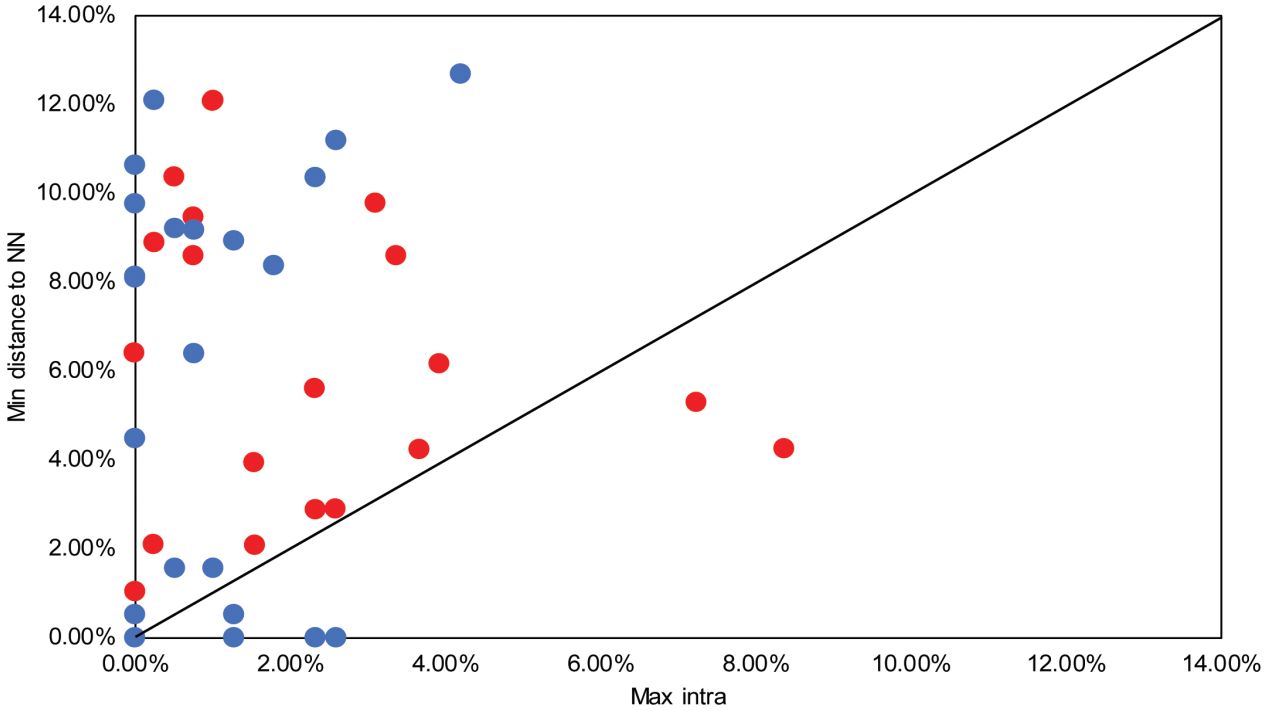

Figure 2. DNA barcoding gap represented by a scatter plot of all 338 Malaysian black fly sequences: Maximum intra-distance versus minimum distance to nearest neighbour (NN). The gap exists for species above the 1:1 line. The DNA barcode gap was present in 79.54% of all species. Single sequence species were excluded. Red dots include newly generated sequences, whereas blue dots represent sequences from GenBank.

ous. *Simulium lurauense* was the only incorrect sequence for ASB. For BM, 10 sequences (8.0%) were incorrectly identified, whereas only four sequences were incorrectly identified for BCM (3.2%). Eight sequences were below the threshold value of 3%.

Five singleton sequences (2.06%) for species in the subgenus *Simulium* were incorrect for BM but not for BCM: *Simulium crassimanum*, *S. laterale*, *S. maklarini*, *S. grossifilum*, and *S. tani*. All these sequences were below the threshold value, including a sequence from *S. malayense.* Twelve sequences (6.41%), comprised of *S. mirum* (six) and *S. kiuliense* (six), were ambiguous for both BM and BCM.

Table 3. DNA barcode identifications of all 338 Malaysian black fly sequences, according to subgenus, using TaxonDNA functions: Best Match (BM), Best Close Match (BCM), and All Species Barcode (ASB).

	Best Match (%)			Best Close Match (%)				All Species Barcode (%)			
	Correct	Ambiguous	Incorrect	Correct	Ambiguous	Incorrect	No match	Correct	Ambiguous	Incorrect	No match
Gomphostilbia	82.40	9.60	8.00	80.80	9.60	3.20	6.40	49.60	43.20	0.80	6.40
Nevermannia	100	0.00	0.00	92.31	0.00	0.00	7.69	92.31	0.00	0.00	7.69
Simulium	90.91	6.42	2.67	90.91	6.42	0.00	2.67	90.37	6.95	0.00	2.67
All	88.46	7.10	4.43	87.27	7.10	1.18	4.43	75.44	19.82	0.29	4.43

3.3 Tree analyses and species delimitation

Overall, 59 species of black flies were used in trees inferred based on maximum likelihood (ML): 52 species from Malaysia and eight reference species from Thailand and two from Vietnam. All species were members of one of three subgenera: *Gomphostilbia* (seven species groups), *Nevermannia* (two species groups), and *Simulium* (nine species group).

For species delimitation, 57 and 58 operational taxonomic units (OTUs) were recognised in ASAP and GMYC, respectively. The results were slightly different when compared against species groups (Table 4).

For the subgenus *Nevermannia*, *S. ledangense* and *S. pairoti* were recognised as the same entity, whereas *S. borneoense* consisted of two entities (Fig. 3). The result was consistent with the tree in which *S. borneoense* formed a strong monophyletic clade but with slightly high intraspecific distance (Table 2), whereas *S. pairoti* and *S. ledangense* shared the same clade. All three species were in the *S. feuerborni* species group. On the other hand, *S. aureohirtum* from Malaysia and Thailand were considered different entities by both species delimitation methods, although they were in the same monophyletic clade with strong support.

There were 24 nominal species in the subgenus *Simulium* (Fig. 4), including four reference species from Thailand (*S. bullatum*, *S. chamlongi*, *S. sansahoense*, and *S. tani*) and one from Vietnam (*S. rosliramlii*), but ASAP and GMYC only recognised 20 and 22 entities, respectively. *Simulium bullatum* was treated as the same entity as Malaysian *S. malayense* by ASAP but not by GMYC. This outcome was the same for *S. jeffreyi* and *S. perakense*. On the other hand, the *S. melanopus* species group was divided into two clades, the first clade consisting of *S. nigripilosum*, *S. crassimanum*, *S. bishopi*, *S. laterale*, and *S. maklarini*, which were successfully distinguished by the two species delimitation approaches. The second clade consisted of three species in two species groups of East Malaysia, namely *S. murudense* and *S. mirum*, both in the *S. melanopus* species group, and *S. kiuliense* in the *S. nobile* species group; all three species were recognised as a single entity by both ASAP and GMYC, with strong support. *Simulium vanluni*, a species from Peninsular Malaysia formerly known as *S. nobile*, was in a different clade. All species in the *S. grossifilum*, *S. nitidithrox*, *S. tuberosum*, and *S. variegatum* species groups were successfully distinguished by both ASAP and GMYC.

Table 4. Number of species and operational taxonomic units (OTUs), using two different species delimitation methods, ASAP & GMYC, according to species group in the genus *Simulium*

Subgenus	Species group	Morphology	ASAP	GMYC
Nevermannia	*Aureohirtum*	1	2	2
	Feuerborni	3	3	3
	Total	4	5	5
Gomphostilbia	*Asakoae*	7	5	5
	Batoense	8	11	10
	Ceylonicum	3	3	3
	Darjeelingense	1	1	1
	Epistum	7	8	8
	Gombakense	3	3	3
	Varicorne	1	1	1
	Total	30	32	31
Simulium	*Argentipes*	3	3	3
	Grossifilum	1	1	1
	Melanopus	7	6	6
	Multistriatum	3	2	3
	Nitidithorax	1	1	1
	Nobile	2	1	1
	Striatum	2	1	2
	Tuberosum	3	3	3
	Variegatum	2	2	2
	Total	24	20	22
Overall		58	57	58

Thirty morphologically identified nominal species of the subgenus *Gomphostilbia*, including two species from Thailand (*S. parahiyangum* and *S. decuplum*) and one species from Vietnam (*S. yvonneae*), represented the *S. batoense* species group (Fig. 5). Molecular identification recognised 32 and 31 entities under ASAP and GMYC, respectively. The *S. batoense* species group was divided into three distinct clades, with *S. sazalyi* sharing the same clade with *S. parahiyangum* from Malaysia and recognised as a single entity based on both species delimitation methods.

In the second clade, *S. decuplum* from Malaysia and Thailand was distinguished as two distinct entities. DNA

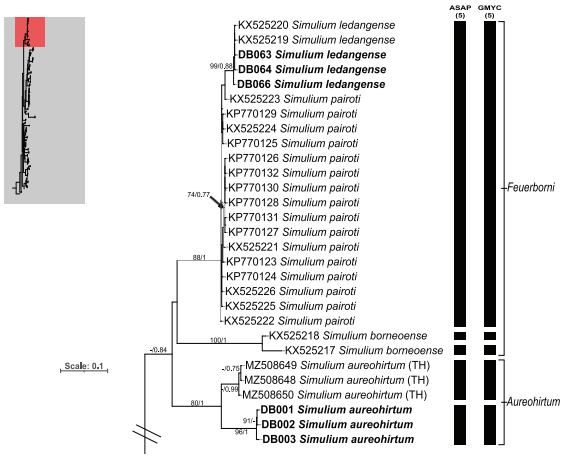

Figure 3. Maximum likelihood tree for the subgenus *Nevermannia*. Full view of the tree is in the top left corner. The bootstrap value and Bayesian Inference (BI) are shown on the branches. Sequences generated from this study are in bold. Vertical bars on the right are the result of species delimitation, with the species groups indicated to the right.

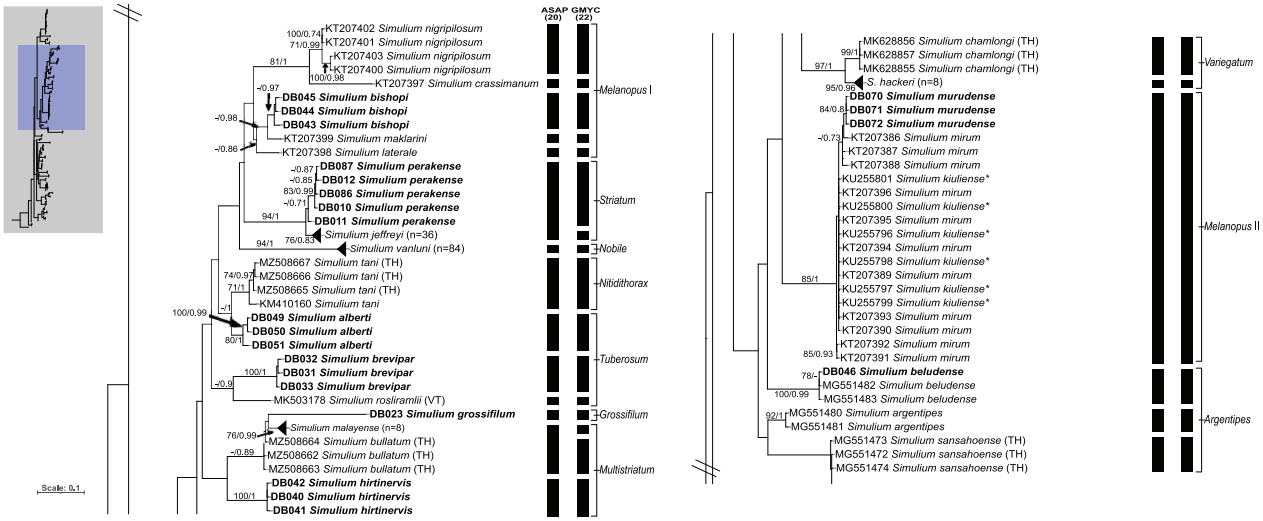

Figure 4. Maximum likelihood tree of subgenus *Simulium*. Full view of the tree is in the top left corner. The bootstrap value and Bayesian Inference (BI) are shown on the branches. Sequences generated from this study are in bold. Vertical bars on the right are the result of species delimitation, with the species group indicated to the right. Species with an asterisk (*) came from a different species group: *Simulium kiuliense* belongs to the *S. nobile* species group but was assigned to the *S. melanopus* species group by species delimitation.

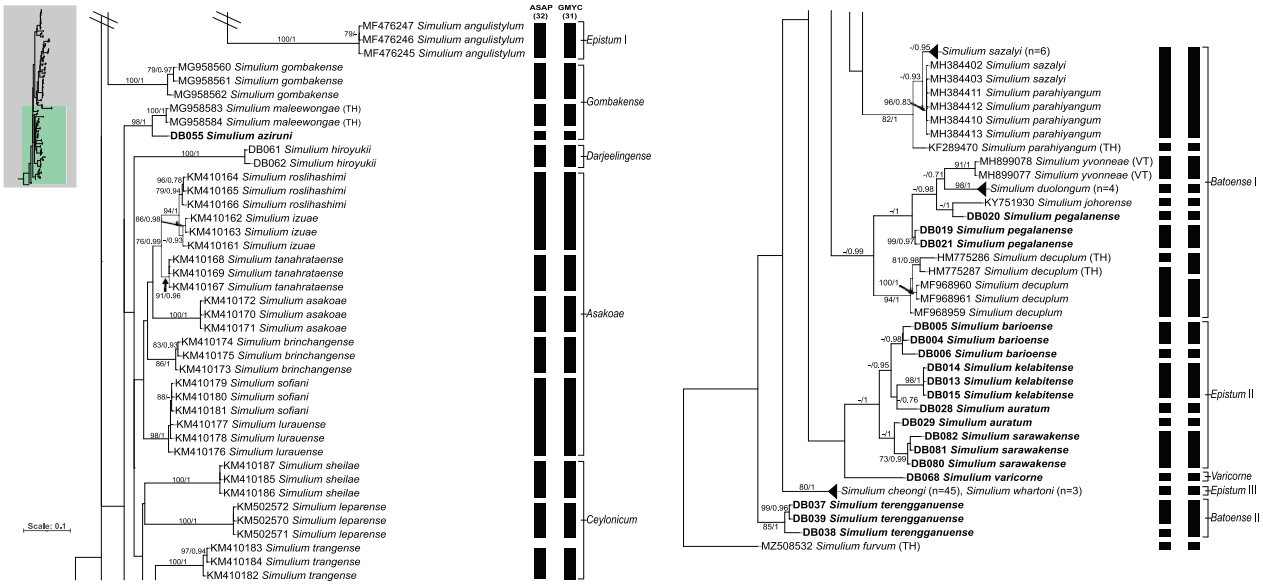

Figure 5. Maximum likelihood tree of subgenus *Gomphostilbia*. Full view of the tree is in the top left corner. The bootstrap value and Bayesian Inference (BI) are shown on the branches. Sequences generated from this study are in bold. Vertical bars on the right are the result of species delimitation, with the species group indicated to the right.

barcode sequences of *S. pegalanense* are reported for the first time and the species was placed in two different lineages and considered two distinct entities. The third clade of the *S. batoense* species group consisted only of *S. terengganuense*, recognized by ASAP and GMYC as two and one entities, respectively. Seven species in the *S. asakaoe* species group were clustered together, with *S. asakoae*, *S. brinchangense*, and *S. tanahrataense* forming monophyletic clades with strong support. *Simulium izuae* and *S. roslihashimi* formed a paraphyletic group and were recognised as a single entity by species delimitation; the same occurred with *S. sofiani* and *S. lurauense*. ASAP and GMYC were both consistent with morphological identification, with three species each in the *S. ceylonicum* and *S. gombakense* species groups. Seven nominal species comprised the *S. epistum* species group, which was split into three clades. The first clade consisted only of *S. angulistylum*, which was positioned near the subgenus *Simulium*. The second clade consisted of four nominal species: *Simulium barioense* (two OTUs), *S. kelabitense* (one OTU), *S. auratum* (two OTUs), and *S. sarawakense* (one OTU). *Simulium whartoni* and *S. cheongi* were a single entity sharing the same strong bootstrap and Bayesian support in the third clade.

4. Discussion

An increasing demand for accurate and timely species identification, and rapid advances in genetic methodology, have spurred progress in establishing molecular identification tools for black flies. By using short standardized DNA markers from mitochondrial COI, molecular species identification is becoming standard practice in many regions, world wide. The increasing importance of black flies in Malaysia makes the development of comprehensive and reliable species databases essential for accurate identification and comparison of regional faunas. A total of 52 species of black flies from Malaysia are represented in our study, accounting for more than 50% of the total species recorded in the country (Adler 2022) and including newly generated sequences for 14 species. The remaining species are unavailable, mostly owing to their rarity. Yet, their barcodes could be available in the future when more specimens are found or through non-destructive DNA extraction from museum specimens.

The distance-based approach successfully identified almost 90% of the species, in parallel with morphological identification, except for several species considered ambiguous or misidentified due to overlapping genetic distance with the nearest-neighbour species and high intraspecific distance. Despite the small number of samples per species, the genetic divergence overlap in this study (8.38%) is less than in the comprehensive study of black flies in Thailand (12.6%) (Pramual et al. 2021).

The barcode gap is the difference between intra- and interspecific genetic distances for each species (Robinson et al. 2009). A larger gap reflects greater accuracy of the DNA barcoding identification (Meyer and Paulay 2005). Absence of this gap leads to failed species separation due to overlap with the nearest-neighbour species. In certain cases, genetic overlap is due to the species having large genetic diversity (DeSalle et al. 2005) or closely related species with incomplete lineage sorting that may lead to non-monophyly (Funk and Omland 2003). In our study, almost 17% of the species lacked barcode gaps, and the sequences were identified as ambiguous after a failure to differentiate species. This result might be due to inadequate mitochondrial signals (Low et al. 2015) and incomplete lineage sorting (Low et al. 2016a). The nuclear 28S gene, however, can resolve some relationships, for example, between *S. lurauense* and *S. sofiani* but not of other members of the *S. asakoae* species group (Low et al. 2015), supporting the use of multi-locus genes as an alternative for distinguishing species rather than using a COI gene alone. Yet, the COI gene is still the gold standard for molecular identification of black flies. It has been used extensively in public databases for comparison.

A genetic distance for black fly species of more than 3% suggests the presence of cryptic species (Pramual and Adler 2014; Rivera and Currie 2009). By this metric, possible cryptic species are present in *S. auratum* and *S. pegalanense*, where the mean intraspecific differences are high (> 4%) and the species are placed in different lineages of the DNA barcode tree. Morphological re-examination is needed after a population study is done with more samples from various geographic distributions. Cryptic and pseudocryptic species are common in black flies, the former referring to two or more morphologically indistinguishable taxa within a nominal species (Bickford et al. 2006) and the latter referring to taxa that are distinguishable after thorough morphological examination coupled with other diagnostic techniques such as population genetics and cytogenetics (Low et al. 2014, 2016a, 2016b; Pramual and Kuvangkadilok 2012).

In the subgenus *Nevermannia*, *S. aureohirtum* in Malaysia shows high divergence (> 7%) from populations in Thailand. Delimitation of both species shows them as different entities. This generalist species has a wide geographic distribution, which could affect the species genetically. Chromosomal and molecular studies show differences among populations (Pramual et al. 2008; Thaijarern et al. 2014). Hence, morphological re-examination is suggested to evaluate the possibility of two separate entities. For the *S. feuerborni* species group, the low genetic difference between *S. ledangense* and *S. pairoti* is similar to the results of the study by Ya'cob et al. (2017a). However, distinct morphological characters are present between these two species and *S. pairoti* and *S. ledangense* are considered pseudocryptic species (Ya'cob et al. 2017a).

In the subgenus *Simulium*, *S. kiuliense* of the *S. nobile* species group is in the same lineage with species of the *S. melanopus* species group (*S. mirum* and *S. murudense*). Even species delineation indicates they are a single entity. *Simulium kiuliense* was revalidated after re-examination of *S. nobile* s. l., which revealed morphological differences and a large molecular distinction between populations in Java and mainland Asia along with a third species, *S.*

vanluni (Low et al. 2016a; Ya'cob et al. 2017b). *Simulium kiuliense* is known only from the island of Borneo. The low interspecific divergence between these species might be related to an island effect, that is, low genetic diversity due to genetic drift and a possible bottleneck (Boessenkool et al. 2007; Eldridge et al. 2004). This possibility is in line with the genetic distances found in our study, whereby a large genetic distance (> 11%) exists within the *S. nobile* species group between *S. kiuliense* and *S. vanluni* but not (< 2%) between *S. kiuliense* and two species of the *S. melanopus* species group (*S. mirum* and *S. murudense*). Despite the high genetic similarity, morphological differences are present between the *S. nobile* and *S. melanopus* groups (Takaoka 2017).

Subgenus *Gomphostilbia* is the largest Simuliidae taxon in Malaysia, consisting of 50 species. In total, 58% of the *Gomphostilbia* species known from Malaysia were available for our study. In the *S. batoense* species group, *S. parahiyangum* and *S. sazalyi* are in the same lineage and are considered one entity due to incomplete lineage sorting (Sriphirom et al. 2014). This similarity is a consequence of one of four populations of *S. sazalyi* analysed from Peninsular Malaysia, which shares a COI sequence with *S. parahiyangum* from Sarawak (Takaoka et al. 2018c). Yet, these two nominal species can be separated morphologically in each life stage above the egg. On the other hand, sequences from seven of eight species in the *S. asakoae* species group are correctly identified in our study, with the exception of *S. lurauense* and *S. sofiani*, which are ambiguous and molecularly misidentified. All species in the *S. epistum* species group show a DNA barcode gap, except *S. cheongi* and *S. whartoni*, which are considered ambiguous due to high sequence similarity. Even the morphological characteristics of these species are similar (Takaoka and Davies 1995). *Simulium whartoni* is widely distributed from low to high elevations, but *S. cheongi* has been found only at low elevations (Takaoka and Davies 1995; Ya'cob et al. 2016). All insular species known only from East Malaysia in the *S. epistum* species group (*S. auratum*, *S. barioense*, *S. kelabitense*, and *S. sarawakense*) are correctly identified in our study and distinguishable from mainland species in the DNA barcode tree. This distinction is supported by the large genetic distance (> 13%) between mainland and insular species, suggesting that no genetic migration had occurred.

Overall, 22% of the species in our study are non-monophyletic and almost half of their sequences are ambiguous at the species level. This finding is consistent with that of Funk and Omland (2003), who reported nearly 26% of arthropods as non-monophyletic due to incomplete lineage sorting, lack of phylogenetic information, and imperfect taxonomy. These molecular issues are common in black fly studies (Low et al. 2016b; Sriphirom et al. 2014), although the species can often be differentiated morphologically. The remaining species in our study are correctly identified molecularly, suggesting that DNA barcoding of Malaysian black flies is acceptable with the support of species delimitation approaches. A good DNA barcode database can be established using specimens that have been identified morphologically before undergoing molecular analysis to increase accuracy for future referencing. A large sample size from various localities also will help increase the reliability of the species sequences (Meyer and Paulay 2005). The use of variable genetic markers, such as the fast-evolving nuclear big zinc finger (BZF) and elongation complex protein 1 (ECP1) genes, is also encouraged to improve identification efficiency, especially for problematic taxa (Aupalee et al. 2022; Low et al. 2016b; Pangjanda and Pramual 2017; Srisuka et al. 2022).

In summary, we report the DNA barcode for black flies in Malaysia, with high (90%) accuracy for species identification. An increase in the number of sequences per species deposited in DNA barcode databases, particularly when based on correct species identifications, will enhance the possible applications, such as monitoring vectors and other species of public health importance.

5. Author Contributions

Noor Izwan-Anas, Van Lun Low and Zubaidah Ya'cob designed research and analyzed the data.

Zubaidah Ya'cob, Mohamad Rasul Abdullah Halim, Van Lun Low, Mohd Sofian-Azirun and Hiroyuki Takaoka collected the samples in the field.

Noor Izwan-Anas, Emmanuel Yogan Lourdes, Van Lun Low and Zubaidah Ya'cob performed the research.

Noor Izwan-Anas, Van Lun Low, Zubaidah Ya'cob, Hiroyuki Takaoka and Peter H. Adler wrote the paper.

6. Acknowledgement

We extend our gratitude to Prof. Dr. Sazaly Abu Bakar, the Director of the Higher Institution Centre of Excellence, Tropical Infectious Diseases Research & Education Centre (TIDREC), University of Malaya, for his dedicated support. This study was financially supported by the Fundamental Research Grant Scheme (Ref code: FRGS/1/2019/STG03/UM/02/15) (UM Ref. code: FP024-2019A), Malaysia Ministry of Higher Education for Higher Institution Centre of Excellence (Ref. code: MO002-2019) and Ministry of Environment, Government of Japan under GBIF Biodiversity Information Fund for Asia (BIFA) program (BIFA6_017). The work by P.H.A. was supported by NIFA/USDA under project number SC-1700596 and is Technical Contribution No. 7151 of the Clemson University Experiment Station. The authors declare that no competing interests exist.

7. References

Adler PH (2022) World Blackflies (Diptera: Simuliidae): A comprehensive revision of the taxonomic and geographical inventory [2022]. Available from: https://biomia.sites.clemson.edu/pdfs/blackflyinventory.pdf

Adler PH, Currie DC, Wood DM (2004) The black flies (Simuliidae) of North America. USA. Cornell University Press, 941 pp.

Adler PH, Takaoka H, Sofian-Azirun M, Low VL, Ya'cob Z, Chen CD, Lau KW, Pham XD (2016). Vietnam, a hotspot for chromosomal diversity and cryptic species in black flies (Diptera: Simuliidae).

PLoS ONE 11(10), e0163881. https://doi.org/10.1371/journal.pone.0163881

Aupalee K, Srisuka W, Saeung A, Takaoka H (2022) A new species of the *Simulium* (*Simulium*) *striatum* species-group (Diptera: Simuliidae) from Thailand: Description and phylogenetic relationship with its Thai members. Acta Tropica 230: 106387. https://doi.org/10.1016/j.actatropica.2022.106387

Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I (2007) Cryptic species as a window on diversity and conservation. Trends in Ecology and Evolution 22: 148–155. https://doi.org/10.1016/j.tree.2006.11.004

Boessenkool S, Taylor SS, Tepolt CK, Komdeur J, Jamieson IG (2007) Large mainland populations of South Island robins retain greater genetic diversity than offshore island refuges. Conservation Genetics 8: 705–714. https://doi.org/10.1007/s10592-006-9219-5

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ (2014) BEAST 2: A software platform for Bayesian evolutionary analysis. PLoS Computational Biology 10: e1003537. https://doi.org/10.1371/journal.pcbi.1003537

Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D, de Maio N, Matschiner M, Mendes FK, Müller NF, Ogilvie HA, du Plessis L, Popinga A, Rambaut A, Rasmussen D, Siveroni I, Suchard MA, Wu CH, Xie D, Zhang C, Stadler T, Drummond AJ (2019) BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Computational Biology 15: e1006650. https://doi.org/10.1371/journal.pcbi.1006650

Darriba D, Taboada GL, Doallo R, Posada D (2012) JModelTest 2: More models, new heuristics and parallel computing. Nature Methods 9: 772. https://doi.org/10.1038/nmeth.2109

DeSalle R, Egan MG, Siddall M (2005) The unholy trinity: Taxonomy, species delimitation and DNA barcoding. Philosophical Transactions of the Royal Society B: Biological Sciences 360: 1905–1916. https://doi.org/10.1098/rstb.2005.1722

Eldridge MDB, Kinnear JE, Zenger KR, McKenzie LM, Spencer PBS (2004) Genetic diversity in remnant mainland and "pristine" island populations of three endemic Australian macropodids (Marsupialia): *Macropus eugenii*, *Lagorchestes hirsutus* and *Petrogale lateralis*. Conservation Genetics 5: 325–338. https://doi.org/10.1023/B:COGE.0000031148.59923.aa

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular marine biology and biotechnology 3: 294–299. https://doi.org/10.1071/ZO9660275

Fujisawa T, Barraclough TG (2013) Delimiting species using single-locus data and the generalized mixed yule coalescent approach: A revised method and evaluation on simulated data sets. Systematic Biology 62: 707–724. https://doi.org/10.1093/sysbio/syt033

Fukuda M, Choochote W, Bain O, Aoki C, Takaoka H (2003) Natural infections with filarial larvae in two species of black flies (Diptera: Simuliidae) in northern Thailand. Japanese Journal of Tropical Medicine and Hygiene 31: 99–102. https://doi.org/10.2149/tmh1973.31.99

Funk DJ, Omland KE (2003) Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annual Review of Ecology, Evolution, and Systematics 34: 397–423. https://doi.org/10.1146/annurev.ecolsys.34.011802.132421

Guo B, Kong L (2022) Comparing the efficiency of single-locus species delimitation methods within Trochoidea (Gastropoda: Vetigastropoda). Genes 13: 2273. https://doi.org/10.3390/genes13122273

Hadi UK, Takaoka H (2018) The biodiversity of black flies (Diptera: Simuliidae) in Indonesia. Acta Tropica 185: 133–137. https://doi.org/10.1016/j.actatropica.2018.02.013

Hajibabaei M, Singer GAC, Hebert PDN, Hickey DA (2007) DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends in Genetics 23: 167–172. https://doi.org/10.1016/j.tig.2007.02.001

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 103: 968–971. https://doi.org/10.1073/pnas.0510466103

Hall T, Biosciences I, C. Carlsbad (2011) BioEdit: An important software for molecular biology. GERF Bulletin of Biosciences 2: 60–61.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004) Identification of birds through DNA barcodes. PLoS Biology 2: e312. https://doi.org/10.1371/journal.pbio.0020312

Hernández-Triana LM, Chaverri LG, Rodríguez-PÉREZ MA, Prosser SWJ, Hebert PDN, Gregory TR, Johnson N (2015) DNA barcoding of Neotropical black flies (Diptera: Simuliidae): Species identification and discovery of cryptic diversity in Mesoamerica. Zootaxa 3936: 93–114. https://doi.org/10.11646/zootaxa.3936.1.5

Hernández-Triana LM, Montes Oca F de, J Prosser SW, N Hebert PD, Ryan Gregory T, Mcmurtrie S, Hernández-Triana LM (2017) DNA barcoding as an aid for species identification in Austral black flies (Insecta: Diptera: Simuliidae). Genome 60: 348–357. https://doi.org/10.1139/gen-2015-0168

Hew YX, Ya'cob Z, Adler PH, Chen CD, Lau KW, Sofian-Azirun M, Muhammad-Rasul AH, Putt QY, Izwan-Anas N, Hadi UK, Suana IW, Takaoka H, Low VL (2023) DNA barcoding of black flies (Diptera: Simuliidae) in Indonesia. Parasites Vectors 16: 248. https://doi.org/10.1186/s13071-023-05875-1

Ishii Y, Choochote W, Bain O, Fukuda M, Otsuka Y, Takaoka H (2008) Seasonal and diurnal biting activities and zoonotic filarial infections of two *Simulium* species (Diptera: Simuliidae) in northern Thailand. Parasite 15: 121–129. https://doi.org/10.1051/parasite/2008152121

Janzen DH, Hajibabaei M, Burns JM, Hallwachs W, Remigio E, Hebert PDN (2005) Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. Philosophical Transactions of the Royal Society B: Biological Sciences 360: 1835–1845. https://doi.org/10.1098/rstb.2005.1715

Jumpato W, Tangkawanit U, Wongpakam K, Pramual P (2019) Molecular detection of Leucocytozoon (Apicomplexa: Haemosporida) in black flies (Diptera: Simuliidae) from Thailand. Acta Tropica 190: 228–234. https://doi.org/10.1016/j.actatropica.2018.11.024

Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A (2019) RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35: 4453–4455. https://doi.org/10.1093/bioinformatics/btz305

Kvist S, Oceguera-Figueroa A, Siddall ME, Erséus C (2010) Barcoding, types and the *Hirudo* files: Using information content to critically evaluate the identity of DNA barcodes. Mitochondrial DNA 21: 198–205. https://doi.org/10.3109/19401736.2010.529905

Letunic I, Bork P (2021) Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. Nucleic Acids Research 49: W293–W296. https://doi.org/10.1093/nar/gkab301

Low VL, Adler PH, Takaoka H, Ya'cob Z, Lim PE, Tan TK, Lim YAL, Chen CD, Norma-Rashid Y, Sofian-Azirun M (2014) Mitochondrial DNA markers reveal high genetic diversity but low genetic differentiation in the black fly *Simulium tani* Takaoka & Davies along an

elevational gradient in Malaysia. PLoS ONE 9: e100512. https://doi.org/10.1371/journal.pone.0100512

Low VL, Takaoka H, Adler PH, Ya'cob Z, Norma-Rashid Y, Chen CD, Sofian-Azirun M (2015) A multi-locus approach resolves the phylogenetic relationships of the *Simulium asakoae* and *Simulium ceylonicum* species groups in Malaysia: Evidence for distinct evolutionary lineages. Medical and Veterinary Entomology 29: 330–337. https://doi.org/10.1111/mve.12120

Low VL, Takaoka H, Pramual P, Adler PH, Ya'cob Z, Chen CD, Yotopranoto S, Zaid A, Hadi UK, Lardizabal ML, Nasruddin-Roshidi A, Sofian-Azirun M (2016a) Three taxa in one: cryptic diversity in the black fly *Simulium nobile* (Diptera: Simuliidae) in Southeast Asia. Journal of Medical Entomology 53: 972–976. https://doi.org/10.1093/jme/tjw058

Low VL, Takaoka H, Pramual P, Adler PH, Ya'cob Z, Huang Y-T, da Pham X, Ramli R, Chen CD, Wannaket A, Sofian-Azirun M (2016b) Delineating taxonomic boundaries in the largest species complex of black flies (Simuliidae) in the Oriental Region. Scientific Reports 6: 20346. https://doi.org/10.1038/srep20346

Meier R, Shiyang K, Vaidya G, Ng PKL (2006) DNA barcoding and taxonomy in Diptera: A tale of high intraspecific variability and low identification success. Systematic Biology 55: 715–728. https://doi.org/10.1080/10635150600969864

Meyer CP, Paulay G (2005) DNA barcoding: Error rates based on comprehensive sampling. PLoS Biology 3: e422. https://doi.org/10.1371/journal.pbio.0030422

Miller MA, Pfeiffer W, Schwartz T (2011) The CIPRES science gateway: A community resource for phylogenetic analyses. Proceedings of the TeraGrid 2011 Conference: Extreme Digital Discovery, TG'11: 41: 1–8. https://doi.org/10.1145/2016741.2016785

Pangjanda S, Pramual P (2017) Tests of conspecificity for closely related black fly (Diptera: Simuliidae) species of the *Simulium striatum* group in Thailand. Zootaxa 4231: 421–430. https://doi.org/10.11646/zootaxa.4231.3.8

Pramual P, Adler PH (2014) DNA barcoding of tropical black flies (Diptera: Simuliidae) of Thailand. Molecular Ecology Resources 14: 262–271. https://doi.org/10.1111/1755-0998.12174

Pramual P, Kuvangkadilok C (2012) Integrated cytogenetic, ecological, and DNA barcode study reveals cryptic diversity in *Simulium* (*Gomphostilbia*) *angulistylum* (Diptera: Simuliidae). Genome 55: 447–458. https://doi.org/10.1139/g2012-031

Pramual P, Jomkumsing P, Wongpakam K, Wongwian P (2021) DNA barcoding of tropical black flies (Diptera: Simuliidae) in Thailand: One decade of progress. Acta Tropica 224: 106116. https://doi.org/10.1016/j.actatropica.2021.106116

Pramual P, Thaijarern J, Wongpakam K (2016) DNA barcoding of human-biting black flies (Diptera: Simuliidae) in Thailand. Acta Tropica 164: 33–40. https://doi.org/10.1016/j.actatropica.2016.08.016

Pramual P, Thaijarern J, Tangkawanit U, Wongpakam K (2020) Molecular identification of blood meal sources in black flies (Diptera: Simuliidae) suspected as Leucocytozoon vectors. Acta Tropica 205: 105383. https://doi.org/10.1016/j.actatropica.2020.105383

Pramual P, Wongpakam K, Kuvangkadilok C (2008) Cytogenetics of the black fly *Simulium aureohirtum* Brunetti from Thailand. Cytologia 73: 293–304. https://doi.org/10.1508/cytologia.73.293

Puillandre N, Brouillet S, Achaz G (2021) ASAP: assemble species by automatic partitioning. Molecular Ecology Resources 21: 609–620. https://doi.org/10.1111/1755-0998.1328

Putt QY, Ya'cob Z, Adler PH, Chen CD, Hew YX, Izwan-Anas N, Lau KW, Sofian-Azirun M, Pham XD, Takaoka H, Low VL (2023) From bites to barcodes: Uncovering the hidden diversity of black flies (Diptera: Simuliidae) in Vietnam. Parasites Vectors 16: 266. https://doi.org/10.1186/s13071-023-05892-0

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systematic Biology 67: 901–904. https://doi.org/10.1093/sysbio/syy032

Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). Molecular Ecology Notes 7: 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Rivera J, Currie DC (2009) Identification of Nearctic black flies using DNA barcodes (Diptera: Simuliidae). Molecular Ecology Resources 9: 224–236. https://doi.org/10.1111/j.1755-0998.2009.02648.x

Robinson E, Blagoev G, Hebert P, Adamowicz S (2009) Prospects for using DNA barcoding to identify spiders in species-rich genera. ZooKeys 16: 27–46. https://doi.org/10.3897/zookeys.16.239

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) Mrbayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61: 539–542. https://doi.org/10.1093/sysbio/sys029

Ruiz-Arrondo I, Hernández-Triana LM, Ignjatović-Ćupina A, Nikolova N, Garza-Hernández JA, Rodríguez-Pérez MA, Oteo JA, Fooks AR, Lucientes Curdi J (2018) DNA barcoding of blackflies (Diptera: Simuliidae) as a tool for species identification and detection of hidden diversity in the eastern regions of Spain. Parasites Vectors 11: 463. https://doi.org/10.1186/s13071-018-3046-7

Saeung A, Srisuka W, Aupalee K, Fukuda M, Otsuka Y, Taai K, Maleewong W, Takaoka H (2020) Natural infections with larvae of *Onchocerca* species type I in the human-biting black fly, *Simulium nigrogilvum* (Diptera: Simuliidae), in western Thailand. Acta Tropica 204: 105344. https://doi.org/10.1016/j.actatropica.2020.105344

Sayers EW, Cavanaugh M, Clark K, Ostell J, Pruitt KD, Karsch-Mizrachi I (2020) GenBank. Nucleic Acids Research 48: D84–D86. https://doi.org/10.1093/nar/gkz956

Sriphirom P, Sopaladawan PN, Wongpakam K, Pramual P (2014) Molecular phylogeny of black flies in the *Simulium tuberosum* (Diptera: Simuliidae) species group in Thailand. Genome 57: 45–55. https://doi.org/10.1139/gen-2013-0145

Srisuka W, Takaoka H, Aupalee K, Saeung A (2022) Discovery of a new black fly species in the *Simulium* (*Simulium*) *striatum* species-group (Diptera: Simuliidae) from northeastern Thailand and its phylogenetic relationships with seven Thai related species. Acta Tropica 238: 106790. https://doi.org/10.1016/j.actatropica.2022.106790

Takaoka H (2003) The Black Flies (Diptera: Simuliidae) of Sulawesi, Maluku and Irian Jaya. Kyushu University Press, Fukuoka, Japan, 581 pp.

Takaoka H (2012) Morphotaxonomic revision of *Simulium* (*Gomphostilbia*) (Diptera: Simuliidae) in the Oriental Region. Zootaxa 3577: 1–42. https://doi.org/10.11646/zootaxa.3577.1.1

Takaoka H (2017) Morphotaxonomic revision of species-groups of *Simulium* (*Simulium*) (Diptera: Simuliidae) in the Oriental Region. Zootaxa 4353: 425–446. https://doi.org/10.11646/zootaxa.4353.3.2

Takaoka H, Douglas M. Davies (1995) The Black Flies (Diptera: Simuliidae) of West Malaysia. Kyushu University Press.

Takaoka H, Choochote W, Aoki C, Fukuda M, Bain O (2003) Black flies (Diptera: Simuliidae) attracted to humans and water buffalos and natural infections with filarial larvae, probably *Onchocerca* sp., in northern Thailand. Parasite 10: 3–8. https://doi.org/10.1051/parasite/2003101p3

Takaoka H, Low VL, Srisuka W, Ya'cob Z, Saeung A (2018c) Morphological and molecular evidence for multiple taxa in *Simulium parahiyangum* (Diptera: Simuliidae), a geographic generalist black fly in Southeast Asia. Journal of Medical Entomology 55: 1453–1463. https://doi.org/10.1093/jme/tjy109

Takaoka H, Sofian-Azirun M, Ya'cob Z, Chen CD, Lau KW, Pham XDA (2015) The black flies (Diptera: Simuliidae) from Thua Thien Hue and Lam Dong Provinces, Vietnam. Zootaxa 3961: 1–96. https://doi.org/10.11646/zootaxa.3961.1.1

Takaoka H, Sofian-Azirun M, Ya'cob Z, Chen CD, Lau KW, Low VL, Suana IW (2017b) The black flies (Diptera: Simuliidae) of the Lesser Sunda Archipelago, Indonesia. Acta Tropica 169: 170–186. https://doi.org/10.1016/j.actatropica.2017.01.016

Takaoka H, Sofian-Azirun M, Ya'cob Z, Chen CD, Lau KW, Low VL, da Pham X, Adler PH (2017a) The black flies (Diptera: Simuliidae) of Vietnam. Zootaxa 4261: 1–165. https://doi.org/10.11646/zootaxa.4261.1.1

Takaoka H, Srisuka W, Saeung A (2017c) A new human-biting black fly species of *Simulium* (*Simulium*) (Diptera: Simuliidae) from Thailand. Journal of Medical Entomology 54: 945–948. https://doi.org/10.1093/jme/tjx064

Takaoka H, Srisuka W, Saeung A (2019) Checklist and keys for the black flies (Diptera: Simuliidae) of Thailand. Medical Entomology and Zoology 70: 53–77. https://doi.org/10.7601/mez.70.53

Takaoka H, Srisuka W, Maleewong W, Saeung A (2018a) New records of 13 species of black flies (Diptera: Simuliidae) from Myanmar. Acta Tropica 178: 281–289. https://doi.org/10.1016/j.actatropica.2017.12.009

Takaoka H, Ya'cob Z, Sofian-Azirun M (2018b) Classification, annotated list and keys for the black flies (Diptera: Simuliidae) of Peninsular Malaysia. Zootaxa 4498: 1–65. https://doi.org/10.11646/zootaxa.4498.1.1

Tamura K, Stecher G, Kumar S (2021) MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Molecular Biology and Evolution 38: 3022–3027. https://doi.org/10.1093/molbev/msab120

Thaijarern J, Sopaladawan PN, Wongpakam K, Pramual P (2014) Phylogeography of the black fly *Simulium aureohirtum* (Diptera: Simuliidae) in Thailand. Genome 57: 411–417. https://doi.org/10.1139/gen-2014-0103

Thaijarern J, Tangkawanit U, Wongpakam K, Pramual P (2019) Molecular detection of *Trypanosoma* (Kinetoplastida: Trypanosomatidae) in black flies (Diptera: Simuliidae) from Thailand. Acta Tropica 200: 105196. https://doi.org/10.1016/j.actatropica.2019.105196

World Health Organization (WHO) (2022) Onchocerciasis. Available from: https://www.who.int/news-room/fact-sheets/detail/onchocerciasis (October 10, 2022).

Ya'cob Z, Takaoka H, Low VL, Sofian-Azirun M (2017a) Uncovering the mask of the *Simulium feuerborni* complex (Diptera: Simuliidae): Description of a new pseudocryptic species *Simulium pairoti* from Malaysia. Acta Tropica 169: 133–141. https://doi.org/10.1016/j.actatropica.2017.02.005

Ya'cob Z, Takaoka H, Low VL, Sofian-Azirun M (2017b) First description of a new cryptic species, *Simulium vanluni* from Peninsular Malaysia: An integrated morpho-taxonomical and genetic approach for naming cryptic species in the family Simuliidae. Acta Tropica 167: 31–39. https://doi.org/10.1016/j.actatropica.2016.12.009

Ya'cob Z, Takaoka H, Pramual P, Low VL, Sofian-Azirun M (2016) Distribution pattern of black fly (Diptera: Simuliidae) assemblages along an altitudinal gradient in Peninsular Malaysia. Parasites Vectors 9: 1–16. https://doi.org/10.1186/s13071-016-1492-7