Journal of Widlife and Parks, 35: 73-91 (2020)

TAXONOMIC ASSIGNMENT OF DIETARY ARTHROPODS IN MALAYSIAN SWIFTLETS (*Aerodramus* sp.) BASED ON DNA METABARCODING

Chan Kok Sim, Tan Ji & *Goh Wei Lim

Faculty of Science, Universiti Tunku Abdul Rahman, Jalan Universiti, Bandar Barat, 31900, Kampar, Perak, Malaysia.

*Corresponding author's email: wlgoh@utar.edu.my

ABSTRACT

Accuracy in species identification based on the DNA sequences is an important aspect of diet profiling. This study aims to compare the taxonomy assignment by BOLD Systems and BLAST for nucleotides (BLASTn) and to recommend general criteria that are crucial for taxonomic assignment in molecular identification of the dietary arthropods. We have obtained 27 DNA haplotypes of the partial mitochondrial cytochrome *c* oxidase I (COI) region for the dietary arthropods of house-farm swiftlets (Apodidae, *Aerodramus*). Searches were performed using the Barcodes of Life Data (BOLD) Systems (Public Record Barcode Database and All Barcode Record in BOLD) and Basic Local Alignment Search Tool for nucleotide (BLASTn) search. High level of "No match" cases (ca. 78%) in BOLD Systems and discrepancy cases in the search results based on different databases (ca. 22%) were observed in this study. We demonstrated how we had identified the COI sequences to specific-level (ca. 7%), generic-level (4%), familial-level (63%) or order-level (ca. 19%). We recommend three criteria that are crucial in molecular identification using a relatively short-length metabarcode sequence of the dietary arthropods: (1) the query sequence search should be performed in more than one database; (2) the query sequence and the sequences of its BLASTn top hit should be subject to phylogenetic analysis; and (3) identification at low taxonomic levels (generic- and specific-levels) should be verified with the geographical distribution records.

Keywords: Barcode of Life Data, GenBank, mitochondrial cytochrome *c* oxidase I, taxonomic assignment, tree-based identification.

Received (23-July-2019); Accepted (10-March-2020); Available online (25-September-2020)

Citation: Chan, K.S., Tan, J. & Goh, W.L. (2020). Taxonomic assignment of dietary arthropods in Malaysian swiftlets (*Aerodramus* sp.) based on DNA metabarcoding. *Journal of Wildlife and National Parks,* **35**: *73-91*

INTRODUCTION

The past few decades saw an increase in the prevalence of molecular approaches used in identifying species within a mixture of bio-samples. The standard operation generally involves extraction of DNA mixture from an environment sample (e.g. faeces, food boluses, soil etc.), amplification of the desired DNA region using polymerase chain reaction (PCR), molecular cloning into a plasmid vector, and DNA sequencing of the cloning vector. Recent advances in the next-generation sequencing (NGS) technology have paved way for better metagenomic profiling of the environmental samples as a large volume of data can be obtained (Meusnier *et al.,* 2008; Bohmann *et al.,* 2011; Hajibabaei *et al.,* 2011; Razgour *et al.,* 2011; Zeale *et al.,* 2011).

The NGS technology has also benefited the research on diet profiling of insectivores, with the development of high-throughput methodologies for various arthropod taxa such as Hymenoptera, Lepidoptera, Coleoptera and other insect taxa (Hajibabaei *et al.,* 2006, 2011; Meusnier *et al.,* 2008; Zeale *et al.,* 2011; Yu *et al.,* 2012). However, the relatively short read-length requirement of the next-generation sequencing platform, e.g. Illumina platform often limits the size of the DNA metabarcodes to less than 300-bp. Although numerous genetic markers were designed over time for the identification of arthropods (Folmer *et al.,* 1994; Hebert *et al.,* 2003, Hajibabaei *et al.,* 2006; 2011; Meusnier *et al.,* 2008; Zeale *et al.,* 2011), two sets of metabarcoding primers were noteworthy in terms of detection rate, namely mlCOIintF/HCO2198 (Leray *et al.,* 2013) and LepF1/MLepF1-Rev (Brandon-Mong *et al.,* 2015), both targeting specific, partial regions of the mitochondrial cytochrome *c* oxidase I (COI). These primers were later used in several high-throughput diet profiling of insectivorous bats and birds (Bohmann *et al.,* 2011; Razgour *et al.,* 2011; Jedlicka *et al.,* 2013; 2016; Vesterinen *et al.,* 2013; Burgar *et al.,* 2014; Hope *et al.,* 2014; Piñol *et al.,* 2015; Crisol-Martínez *et al.,* 2016; Mansor *et al.,* 2018).

The DNA sequencing, either using the conventional molecular cloning or advanced high-throughput sequencing approach, is generally followed by molecular identification based on reference sequences in DNA databases. The two commonly used tools to achieve this are the Basic Local Alignment Search Tool (BLAST) search using GenBank database (Altschul *et al.,* 1990) and searches against Barcode of Life Data (BOLD) Systems (Ratnasingham & Hebert, 2007). For the BLAST users, many used a threshold of 98% similarity to

represent species-level identification as suggested by Clare *et al.* (2009, 2011) and Mansor *et al.* (2018). In the BOLD Systems search, the query sequence will be assigned a "Best ID" upon a comprehensive tree-based identification method (Wilson *et al.,* 2011) based on four criteria: (1) liberal; (2) strict; (3) liberal and exclusive; and (4) strict and exclusive. These criteria are based on the monophyly and the exclusivity of the query and number of highly similar sequences in the search. It was suggested that the conservative "strict" criterion should be employed for large-scale species identifications. With this criterion, the query will be assigned to a taxon when it is nested within a clade formed by the members of the taxon, although some other members of this correct taxon can also be found elsewhere on the tree (Wilson *et al.,* 2011). If the query sequence is not assigned to any "Best ID", "No match" appears in the search results and it should be regarded as of uncertain taxon.

However, species identification based on the similarity value in the BLAST search received criticism from Munch *et al.* (2008) because: (1) the genetic variation across populations and closely related species are ignored, and (2) the measure of confidence only reflects the local sequence similarity and not the significance of the species assignment. On the other hand, query sequences with high intraspecific divergences cannot be identified in BOLD Systems due to the conservative nature of the "strict" criterion in its tree-based identification, as there are no reference sequences with less than 1% divergence (Sauer & Hausdorf, 2012).

The present study aims to develop a taxonomic assignment procedure of the DNA metabarcodes obtained from the diet profiling of the farmed Malaysian swiftlets *(Aerodramus* sp.), using both BOLD Systems and GenBank identification systems. The specific objectives of this study are: (1) to compare the taxonomy assignment by BOLD Systems and BLAST for nucleotides (BLASTn); and (2) to recommend a general criterion that is crucial for taxonomic assignment in molecular identification for the dietary arthropods.

METHODOLOGY

Sample Collection

Fresh faecal samples were collected from multiple individuals from seven swiftlet house-farms in Perak, Malaysia, during the month of October in 2017. Each swiftlet house-farm had an approximate colony size of 400 birds (personal observation). Faecal collection was done in the evening by placing the plastic-enclosed cardboards (Chan *et al.*, 2019) on the ground in areas with around ten swiftlet nests. The faecal samples were collected the next morning from each swiftlet house-farm and stored separately in sterile containers before being stored at -20ºC.

Total DNA Extraction

A total of 100 mg faecal sample was ground using liquid nitrogen. The total DNA was extracted using a PowerFecal Extraction kit (Qiagen, Germany) following the manufacturer's instructions. A final elution of 30 µL was stored at -20 ºC until use. The purity and quantity of extracted DNA were assessed using a Nanodrop 2000 Spectrophotometer (Thermo Fisher Scientific, USA).

Polymerase Chain Reaction (PCR) and Cloning

The present study used the LepF1 (5'-ATTCAACCAATCATAAAGATATTGG-3') primer and MLepF1-Rev primer (5'-CGTGGAAAWGCTATATCWGGTG-3') developed for the high-throughput metabarcoding analysis of insects by Brandon-Mong *et al.* (2015). These primers code for 218-bp sized of the partial COI region (hereafter referred to as Lep region). It was designed in such a length to fit the capacity of the Illumina platform in NGS. This region has been widely used in various insect metagenomic study (Beng *et al*., 2016; Jusino *et al*., 2019; Piñol *et al*., 2018). PCR was performed using a 1× GoTaq® Green Master Mix (Promega, USA) with 0.5 µM of forward and reverse primers each. The thermocycling conditions were: an initial denaturation of 95ºC for 2 min; 50 cycles of 95ºC for 45 sec, 53ºC for 1 min and 72ºC for 3 min; and a final extension of 72ºC for 10 min.

Amplicons were purified using a Nucleospin Gel and PCR Clean-Up Kit (Macherey-Nagel, Germany) following instructions by the manufacturer. The purified samples with an estimated concentration of 50ng/µL were subsequently quantified using Nanodrop 2000 Spectrophotometer prior to ligation using a pGEM®-T Easy Vector (Promega, USA) and transformation into JM109 competent cells. Colony PCR was performed for 64 white colonies. The desired amplicons were purified before being commercially sequenced by Apical Scientific Sdn. Bhd. (Malaysia).

Data Analyses

The DNA sequences obtained for the 64 Lep clones were analysed using DNaSP version 5.0 (Librado & Rozas, 2009) to identify unique haplotypes. Each haplotype was then queried for identification in BOLD Systems against: (1) All Barcode Records on BOLD and (2) Public Record Barcode Database. Each haplotype was also queried in BLASTn and the top hits were recorded.

A neighbor-joining (NJ) analysis was performed to show the genetic relatedness between each of the haplotypes and their respective first top hits from BLASTn.

The DNA sequences were aligned using ClustalX version 2.1 (Larkin *et al.,* 2007) and manually trimmed in BioEdit version 7.0.5 (Hall, 1999). NJ tree reconstruction with 1,000 bootstrap replicates was performed in Molecular Evolutionary Genetics Analysis (MEGA) version 7.0 (Kumar *et al.,* 2016). Psocoptera (GenBank accession number HQ658137) was set as the outgroup as it was shown to be relatively distant to all other arthropod orders based on past phylogenomic studies (Misof *et al.,* 2014; Kjer *et al.,* 2016). All DNA sequences obtained in this study were deposited in GenBank (Table 1).

Table 1 List of accessions of the haplotypes with their respective collection localities and GenBank accession numbers.

COI Haplotype	Collection localities	GenBank accession numbers
1	Beruas, Perak, Peninsular Malaysia	MT410749
2	Ber uas, Perak, Peninsular Malaysia	MT410760
3	Sitiawan, Perak, Peninsular Malaysia	MT410741
4	Beruas, Perak, Peninsular Malaysia	MT410754
5	Beruas, Perak, Peninsular Malaysia	MT410743
6	Beruas, Perak, Peninsular Malaysia	MT410752
7	Beruas, Perak, Peninsular Malaysia	MT410765
8	Beruas, Perak, Peninsular Malaysia	MT410758
9	Beruas, Perak, Peninsular Malaysia	MT410755
10	Sitiawan, Perak, Peninsular Malaysia	MT410766
11	Sitiawan, Perak, Peninsular Malaysia	MT410751
12	Sitiawan, Perak, Peninsular Malaysia	MT410759
13	Sitiawan, Perak, Peninsular Malaysia	MT410767
14	Sitiawan, Perak, Peninsular Malaysia	MT410742
15	Sitiawan, Perak, Peninsular Malaysia	MT410748
16	Beruas, Perak, Peninsular Malaysia	MT410745
17	Beruas, Perak, Peninsular Malaysia	MT410746
18	Beruas, Perak, Peninsular Malaysia	MT410753
19	Beruas, Perak, Peninsular Malaysia	MT410750
20	Ipoh, Perak, Peninsular Malaysia	MT410761
21	Gopeng, Perak, Peninsular Malaysia	MT410762
22	Gopeng, Perak, Peninsular Malaysia	MT410747
23	Gopeng, Perak, Peninsular Malaysia	MT410763
24	Pantai Remis, Perak, Peninsular Malaysia	MT410764
25	Pantai Remis, Perak, Peninsular Malaysia	MT410756
26	Pantai Remis, Perak, Peninsular Malaysia	MT410744
27	Pantai Remis, Perak, Peninsular Malaysia	MT410757

RESULTS AND DISCUSSION

BOLD Systems Search

The 64 clones sequenced for the Lep region contained 27 unique haplotypes. This accounted for approximately 10% of the arthropod operational taxonomic units (OTU) sequenced using NGS-Targeted Amplicon Sequencing in a separate study (Chan *et al.*, 2019). Using the Public Record Barcode Database in BOLD, our searches recorded 78% (21 out of 27 haplotypes) as "No match" while the rest of the haplotypes (Haplotypes 4, 8, 11, 12, 14 and 20) were successfully assigned to genus or species (Table 2). When the searches were carried out using All Barcode Records Database, the "No match" cases were reduced to 37% (10 out of 27 haplotypes). Fourteen haplotypes were assigned the arthropod orders and three haplotypes were identified to specific-level (Table 2).

However, three discrepancies were observed in the taxonomic assignments based on the two databases. Firstly, Haplotype 8 was assigned as *Deronectes platynotus* (Coleoptera) in the Public Record Barcode Database but assigned as Diptera when queried against All Barcode Records on BOLD. Next, Haplotypes 11 and 12 were identified as *Galathea* (Decapoda) and *Deronectes platynotus* (Coleoptera), respectively, in the Public Record Barcode Database, but identified as Diptera when queried against All Barcode Records on BOLD (Table 2).

BLASTn Search and Neighbor-Joining (NJ) Analysis

The top BLASTn hits for each haplotype are also listed in Table 2. Discrepancies in the taxonomic assignment by BOLD Systems and BLASTn were observed for six haplotypes, four (Haplotypes 8, 11, 12 and 16) of which at order-level and two (Haplotypes 18 and 19) at species-level (Table 2).

In the NJ tree (Figure 1), Haplotype 23 was clustered with members of Hemiptera (87%) even though it was identified as *Euplatypus* (Coleoptera) when queried against GenBank. For the rest of the haplotypes, they formed clusters with their top hits, even though some clusters were not well supported. Most clusters appear to be indicative of the arthropod relationships at family-level.

Table 2 Search results of the 27 haplotypes based on BOLD and top hits based on BLASTn. Cases of discrepancies were shaded.

Haplotype	Best ID in Public Record Barcode Database	Best ID in All Barcode Records in BOLD	GenBank BLASTn			Decision (Order, family or species)
			Top hit (Acc. number)	Similarity %	Clustering in NJ tree	
1	No match	No match	Lepidoptera, *Urania leilus* (KX781989.1)	94	Yes	Lepidoptera, Uranidae
2	No match	Diptera sp.	Diptera, *Nephrotoma alterna,* (MF838043.1)	93	Yes	Diptera, Tipulinae
3	No match	Hymenoptera sp.	Hymenoptera, *Pachycondyla* sp. (MF673717.1)	92	Yes	Hymenoptera, Formicidae
4	Coleoptera, *Carpophilus marginellus*	Coleoptera, *Carpophilus marginellus*	Coleoptera, *Carpophilus marginellus* (KU914959.1)	99	Yes	[1]*Carpophilus marginellus*
5	No match	Hymenoptera, *Hypoponera*	Hymenoptera, *Hypoponera* sp. (KY845694.1)	99	Yes	Hymenoptera, Formicidae
6	No match	No match	Diptera, Limoniidae sp. (KX053827.1)	94	Yes, but also with other family	Diptera, Unidentified families
7	No match	Hemiptera sp.	Hemiptera, Delphacidae sp. (KR578572.1)	87	Yes	Hemiptera, Delphacidae
8	Coleoptera, *Deronectes platynotus*	Diptera sp.	Diptera, *Nemorimyza* sp. (MF641766.1)	94	Yes, but also with other families	Diptera, Unidentified families
9	No match	No match	Coleoptera, *Blemus discus* (KU919098.1)	92	Yes	Coleoptera, Carabidae
10	No match	No match	Hemiptera, *Eysarcoris* sp. (KY847240.1)	96	Yes	Hemiptera, Pentatomidae
11	Decapoda, *Galathea* sp.	Diptera sp.	Diptera, *Tricimba* sp. (KR639430.1)	96	Yes	Diptera, Chloropidae

Table 2 (continued)

Haplotype	Best ID in Public Record Barcode Database	Best ID in All Barcode Records in BOLD	GenBank BLASTn			Decision (Order, family or species)
			Top hit (Acc. number)	Similarity %	Clustering in NJ tree	
12	Coleoptera, *Deronectes platynotus*	Diptera sp.	Diptera, *Nemorimyza* sp. (MF641766.1)	94	Yes, but also with other families	Diptera, Unidentified families
13	No match	No match	Psocoptera, *Liposcelis entomophila* (HQ658137.1)	100	Yes	Psocoptera, Liposcelidae
14	Hymenoptera, *Odontomachus simillimus*	Hymenoptera, *Odontomachus simillimus*	Hymenoptera, *Odontomachus simillimus* (KU504909.1)	100	Yes	[2]*Odontomachus simillimus*
15	No match	Coleoptera sp.	Coleoptera, *Stelidota geminate* (KM444965.1)	91	Yes	Coleoptera, Nitidulidae
16	No match	Coleoptera sp.	Diptera, *Spilogona* sp. (KR438577.1)	90	Yes, but also with other families	Uncertain order
17	No match	No match	Diptera, Cecidomyiidae (KM626905.1)	92	Yes, but also with other families	Diptera, Unidentified families
18	No match	Coleoptera, *Epuraea luteolus*	Coleoptera, *Epuraea signata* (KM442541.1)	94	Yes	Coleoptera, Nitidulidae
19	No match	Coleoptera, *Epuraea luteolus*	Coleoptera, *Epuraea signata* (KM442541.1)	92	Yes	Coleoptera, Nitidulidae
20	Diptera *Chironomus circumdatus*	Diptera *Chironomus circumdatus*	Diptera, *Chironomus circumdatus* (KJ530965.1)	99	Yes	[3]*Chironomus*

Table 2 (continued)

Haplotype	Best ID in Public Record Barcode Database	Best ID in All Barcode Records in BOLD	GenBank BLASTn			Decision (Order, family or species)
			Top hit (Acc. number)	Similarity %	Clustering in NJ tree	
21	No match	Hemiptera sp.	Hemiptera, *Trienopa* sp. (KX702955.1)	91	Yes	Hemiptera,Tropiduchidae
22	No match	Hymenoptera sp.	Hymenoptera, *Crematogaster* sp. (KC501979.1)	88	Yes	Hymenoptera,Formicidae
23	No match	No match	Coleoptera, *Euplatypus* sp. (MF804642.1)	92	Clustered with Hemiptera	Uncertain order
24	No match	No match	Hymenoptera, *Dolichoris* sp. (JQ256562.1)	88	Yes	Hymenoptera, Agaonidae
25	No match	No match	Diptera, *Simulium chromatinum* (KM497573.1)	92	Yes	Diptera, Simuliidae
26	No match	Hymenoptera sp.	Hymenoptera, *Philidris* sp. (MF804755.1)	99	Yes	Hymenoptera, Formicidae
27	No match	No match	Diptera, *Drosophila* (DQ471549.1)	93	Yes, but also with other families	Diptera, Unidentified families

Note:

[1]*Carpophilus marginellus* was reported from Peninsular Malaysia by Hill (2008) and Nor-Atikah *et al.* (2019).

[2]*Odontomachus simillimus* was reported from Peninsular Malaysia by Hashim *et al.* (2009) and Noor-Izwan & Amirrudin (2014).

[3]*Chironomus* was reported from Peninsular Malaysia by Al-Shami *et al.* (2010) and Al-Shami *et al.* (2011).

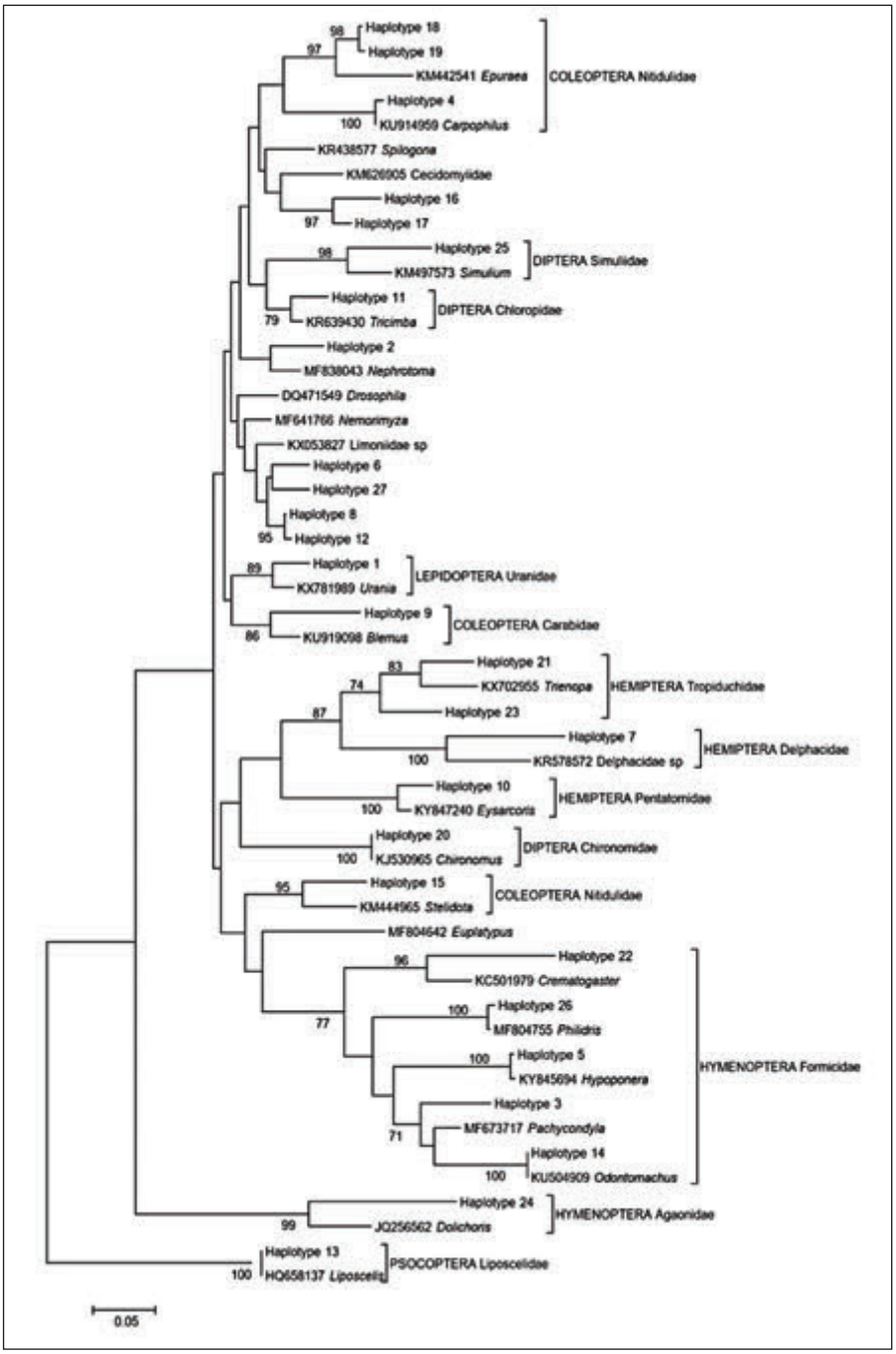

Figure 1 NJ tree reconstruction based on the Lep region. Nodal supports indicate NJ bootstrap supports of > 70%.

Decisions on Taxonomic Assignment

As summarised in Table 2, a total of 11.11% (3 out of 27) of the haplotypes (Haplotypes 8, 11 and 12) showed taxonomic assignment discrepancies based on the two databases, Public Record Barcode Database and All Barcode Records on BOLD of the BOLD Systems. Meanwhile, a total of 22.22% (6 out of 27) of the haplotypes (Haplotypes 8, 11, 12, 16, 18 and 19) showed discrepancies in taxonomic assignment by BOLD Systems and BLASTn. In NJ tree analysis, only 3.7% (1 out 27) haplotypes (Haplotype 23) showed discrepancy with its identified order through BLASTn.

Considering the cases of discrepancy as mentioned above, we apply the following rules to the taxonomic assignment:

(1) No Match is regarded as a search limitation, not a conflict to any other result.

(2) When there is a discrepancy at the order-level, we follow the majority consensus. Decisions for Haplotypes 8, 10 and 12 were made based on this rule.

(3) If there is a discrepancy at the order-level without majority consensus, a sample will not be assigned any order. This explains the uncertain order for Haplotype 16.

(4) If a haplotype does not cluster with its top hit sequence, but clusters with the sequences of other orders in the phylogenetic tree, it will not be assigned any order. This explains the uncertain order for Haplotype 23.

(5) Arthropod family assignment was done for 15 haplotypes as they form clusters exclusively with members of the same family.

(6) The taxonomic assignment at generic- and species-levels requires that the resulting ID be the same from both BLASTn and BOLD Systems databases, followed by validation based on geographical distribution records. We apply this rule to make decisions for Haplotypes 4, 14 and 20. By applying this rule, Haplotype 4 was assigned as *Carpophilus marginellus*, a sap beetle commonly found in oil palm plantations of Peninsular Malaysia (Hill, 2008; Nor-Atikah *et al*., 2019). On the other hand, Haplotype 14 was assigned as *Odontomachus simillimus*, an ant species inhabiting the mangrove forests and oil palm plantations in Peninsular Malaysia (Hashim *et al*., 2009; Noor-Izwan & Amirrudin, 2014). Haplotype 20 was assigned to *Chironomus*, a genus of nonbiting midgets that was reported in the river system of Peninsular Malaysia (Al-Shami *et al.*, 2010; 2011).

Using the rules above, we manage to identify six arthropod orders (Diptera, Coleoptera, Hemiptera, Hymenoptera, Lepidoptera and Psocoptera) in the diet of the house-farm swiftlets (Table 3). For most of these orders, one to four families could be determined. Three species were identified for Diptera, Coleoptera and Hymenoptera, respectively. The proportions of the insect orders consumed by the house-farm swiftlets were largely consistent with results based on the high-throughput amplicon sequencing approach in Chan *et al*. (2019).

Table 3 Summary of taxonomic assignment of the 27 haplotypes obtained in this study.

Order	**Family**	**Species**	**Number of haplotypes**
Diptera	Chironomidae	*Chironomus* spp.	1
	Chloropidae		1
	Simuliidae		1
	Tipulinae		1
	Unidentified		5
Coleoptera	Carabidae		1
	Nitidulidae	*Carpophilus marginellus*	1
	Nitidulidae	Unidentified species	3
Hemiptera	Delphacidae		1
	Pentatomidae		1
	Tropiduchidae		1
Hymenoptera	Formicidae	*Odontomachus simillimus*	1
	Formicidae	Unidentified species	4
	Agaonidae		1
Lepidoptera	Uranidae		1
Psocoptera	Liposcelidae		1
Uncertain order			2
Total			**27**

General Recommendation for Molecular Identification of Arthropods

From our results, we suggest that BLASTn could be more informative compared to BOLD Systems search, judging from the 'No Match' cases and discrepancies in the BOLD Systems search results. However, we have also demonstrated that BLASTn does not always give correct results, so the researchers should exercise extra-caution by cross-checking the query sequence in other databases. Our results showed that a phylogenetic analysis on the sequences obtained from BLASTn is necessary to decide which taxonomic rank that the query sequence should be assigned. All assignments at specific-levels (i.e., Haplotype 4 *C. marginellus*, Haplotype 14 *O. simillimus* and Haplotype 20 *Chironomus circumdatus*) have 99–100 % of similarity with their top hits in BLASTn. However, no all haplotypes that achieved 99-100 % of similarity can be directly assigned to their respective top hits, suggesting that there is no specific threshold of similarity that can indicate the taxonomic-level of identification. The haplotypes that were assigned to their orders or families showed 87–96 % similarity index. The haplotypes of uncertain order (because of the conflicting results in other database or conflicting phylogenetic placement in NJ tree) scored 90–92 % of similarity with their top hits.

We, therefore, recommend the researchers to consider the following criteria before making a taxonomic assignment:
1. The query sequence search should be performed in more than one database, e.g., (a) Public Record Database in BOLD, (b) All Record Barcode Database in BOLD, and (c) BLASTn.
2. The query sequence and the sequences of its BLASTn top hit should be subject to phylogenetic analysis (which includes selected reference taxa) to confirm if their similarity is also supported in the dendrogram.
3. Identification at low taxonomic levels (generic- and specific-levels) should be verified with the geographical distribution records.

Limitations of the Present Study

Kvist (2013) reported that out of 1,242,040 recognised arthropod COI sequences used in his study, only 149,997 sequences (12.08%) matched the sequences in BOLD Systems whereas only 69,123 (5.56%) matched those in GenBank, thereby indicating the incompleteness of these databases at present. The accuracy of taxonomic assignment could possibly be improved if the existing COI records in the BOLD Systems and GenBank databases are better represented by an increase in taxon density.

Considering the general lack of entomological records of insects in Malaysia (Cheng & Kirton, 2007), insect taxa identified using molecular methods should also be cross-checked with their known records of geographical distribution. These limitations point toward the urgent need for a taxonomic catalogue of

insects in Malaysia which would undoubtedly benefit a wide range of research including the diet profiling of insectivores within the country. For the time being, online databases established by Singapore, e.g. Digital Reference Collection for Singapore's Biodiversity (https://singapore.biodiversity.online/) by Ng *et al.* (2011) may be a useful reference for the insect species that are possibly found in Peninsular Malaysia.

ACKNOWLEDGEMENTS

This project was funded by Universiti Tunku Abdul Rahman Research Fund (IPSR/RMC/UTARRF/2016-C2/G04). K.S. Chan was supported by UTAR Research Scholarship Scheme under vote number 6220/C16. We thank the swiftlet farm owners Mr Tan Yoke Tian, Mr Lai, Mr Ho, Mr Yeong, Mr Zhou and Mr Zhong for granting permission to access and collect faecal samples from their farms. We also thank Y.Q. Lan, J.L. Lee, M.S. Tan and T.J. Tham for their assistance in the laboratory work and Prof. Dr A.C. Peter Ooi, Dr G. Khoo and Dr K.W. Loo for their useful comments and suggestions in the preparation of this manuscript.

REFERENCES

Al-Shami, S., Rawi, C.S.M., Nor, S.A.M., Ahmad, A.H. & Ali, A. (2010). Morphological deformities in *Chironomus* spp. (Diptera: Chironomidae) larvae as a tool for impact assessment of anthropogenic and environmental stresses on three rivers in the Juru River System, Penang, Malaysia. *Environmental Entomology*, **39**: 210-222.

Al-Shami, S.A., Salmah, M.R.C., Hassan, A.A. & Azizah, M.N.S. (2011). Fluctuating asymmetry of *Chironomus* spp. (Diptera: Chironomidae) larvae in association with water quality and metal pollution in Permatang Rawa River in the Juru River Basin, Penang, Malaysia. *Water, Air, and Soil Pollution*, **216**: 203-216.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, **215**(3): 403-410.

Beng, K.C., Tomlinson, K.W., Shen, X.H., Surget-Groba, Y., Hughes, A.C., Corlett, R.T. & Slik, J.W.F. (2016). The utility of DNA metabarcoding for studying the response of arthropod diversity and composition to land-use change in the tropics. *Scientific Reports*, **6**: 1-13.

Bohmann, K., Monadjem, A., Noer, C.L., Rasmussen, M., Zeale, M.R.K., Clare, E., Jones, G., Willerslev, E. & Gilbert, M.T.P. (2011). Molecular diet analysis of two African free-tailed bats (Molossidae) using high throughput sequencing. *PLoS ONE*, **6**(6): e21441.

Brandon-Mong, G.J., Gan, H.M., Sing, K.W., Lee, P.S., Lim, P.E. & Wilson, J.J. (2015). DNA metabarcoding of insects and allies: an evaluation of primers and pipelines. *Bulletin of Entomological Research*, **105**(6): 717-727.

Burgar, J.M., Murray, D.C., Craig, M.D., Haile, J., Houston, J., Stokes, V. & Bunce, M. (2014). Who's for dinner? High-throughput sequencing reveals bat dietary differentiation in a biodiversity hotspot where prey taxonomy is largely undescribed. *Molecular Ecology*, **23**(15): 3605-3617.

Chan, K.S., Tan, J. & Goh, W.L. (2019). Diet profiling of house-farm swiftlets (Aves, Apodidae, *Aerodramus* sp.) in three landscapes in Perak, Malaysia, using high-throughput sequencing. *Tropical Ecology*, **60**(3): 379-388.

Cheng, S. & Kirton, L.G. (2007). Overview of insect biodiversity research in Peninsular Malaysia. In *Status of biological diversity in Malaysia and threat assessment of plant species in Malaysia* (Chua, L.S.L, Kirton, L.G. & Saw L.G., eds.), pp; 121-128, Proceeding of Seminar and Workshop, Kuala Lumpur, Malaysia: Forest Research Institute of Malaysia.

Clare, E.L., Fraser, E.E., Braid, H.E., Fenton, M.B. & Hebert, P.D. (2009). Species on the menu of a generalist predator, the eastern red bat (*Lasiurus borealis*): using a molecular approach to detect arthropod prey. *Molecular Ecology*, **18**(11): 2532-2542.

Clare, E.L., Barber, B.R., Sweeney, B.W., Hebert, P.D.N. & Fenton, M.B. (2011). Eating local: influences of habitat on the diet of little brown bats (*Myotis lucifugus*). *Molecular Ecology,* **20**(8): 1772-1780.

Crisol-Martínez, E., Moreno-Moyano, L.T., Wormington, K.R., Brown, P.H. & Stanley, D. (2016). Using next-generation sequencing to contrast the diet and explore pest-reduction services of sympatric bird species in macadamia orchards in Australia. *PloS ONE*, **11**(3): e0150159.

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**(5): 294-299.

Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G.A.C. & Baird, D.J. (2011). Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS ONE*, **6**(4): e17497.

Hajibabaei, M., Smith, M.A., Janzen, D.H., Rodriguez, J.J., Whitfield, J.B. & Hebert, P.D. (2006). A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes*, **6**(4): 959-964.

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**: 95-98.

Hashim, N.R., Wan-Jusoh, W.F.A. & Mohd-Nasir, M.N.S. (2009). Ant diversity in a Peninsular Malaysian mangrove forest and oil palm plantation. *Asian Myrmecology*, **3**: 5-8.

Hebert, P.D., Cywinska, A., Ball, S.L. & Dewaard, J.R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **270**(1512): 313-321.

Hill, D.S. (2008). *Pests of crops in warmer climates and their control*. India: Springer Science and Business Media.

Hope, P.R., Bohmann, K., Gilbert, M.T.P., Zepeda-Mendoza, M.L., Razgour, O. & Jones, G. (2014). Second generation sequencing and morphological faecal analysis reveal unexpected foraging behaviour by *Myotis nattereri* (Chiroptera, Vespertilionidae) in winter. *Frontiers in Zoology*, **11**(1): 39.

Jedlicka, J.A., Sharma, A.M. & Almeida, R.P.P. (2013). Molecular tools reveal diets of insectivorous birds from predator fecal matter. *Conservation Genetics Resources*, **5**(3): 879-885.

Jedlicka, J.A., Vo, A.T.E. & Almeida, R.P. (2016). Molecular scatology and high-throughput sequencing reveal predominately herbivorous insects in the diets of adult and nestling Western Bluebirds (*Sialia mexicana*) in California vineyards. *The Auk: Ornithological Advances*, **134**(1): 116-127.

Jusino, M.A., Banik, M.T., Palmer, J.M., Wray, A.K., Xiao, L., Pelton, E., Barber, J.R., Kawahara, A.Y., Gratton, C., Peery, M.Z. & Lindner, D.L. (2019). An improved method for utilizing high-throughput amplicon sequencing to determine the diets of insectivorous animals. *Molecular Ecology Resources*, 19(1): 176-190.

Kjer, K.M., Simon, C., Yavorskaya, M. & Beutel, R.G. (2016). Progress, pitfalls and parallel universes: a history of insect phylogenetics. *Journal of the Royal Society Interface*, **13**(121): 20160363.

Kumar, S., Stecher, G. & Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, **33**(7): 1870-1874.

Kvist, S. (2013). Barcoding in the dark? a critical view of the sufficiency of zoological DNA barcoding databases and a plea for broader integration of taxonomic knowledge. *Molecular Phylogenetics and Evolution*, **69**(1): 39-45.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R. & Thompson, J.D., Gibson, T.J. & Higgins, D.G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**(21): 2947-2948.

Leray, M., Yang, J.Y., Meyer, C.P., Mills, S.C., Agudelo, N., Ranwez, V., Boehm, J.T. & Machida, R.J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in Zoology*, **10**(1): 34.

Librado, P. & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**(11): 1451-1452.

Mansor, M.S., Nor, S.M. & Ramli, R. (2018). Assessing diet of the Rufous-Winged Philentoma (*Philentoma pyrhoptera*) in lowland tropical forest using Next-Generation Sequencing. *Sains Malaysiana*, **47**(5): 1045-1050.

Meusnier, I., Singer, G.A.C., Landry, J.F., Hickey, D.A., Hebert, P.D.N. & Hajibabaei, M. (2008). A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics*, **9**(1): 214-217.

Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., Beutel, R.G., Niehuis, O. *et al.* (2014). Phylogenomics resolves the timing and pattern of insect evolution. *Science*, **346**(6210): 763-767.

Munch, K., Boomsma, W., Huelsenbeck, J.P., Willerslev, E. & Nielsen, R. (2008). Statistical assignment of DNA sequences using Bayesian phylogenetics. *Systematic Biology*, **57**(5): 750-757.

Ng, P.K.L., Corlett, R. & Tan, H.T.W. (2011). *Singapore biodiversity: an encyclopedia of the natural environment and sustainable development.* Singapore: Editions Didier Millet.

Noor-Izwan, A. & Amirrudin, B.A. (2014). Diversity of ants (Hymenoptera: Formicidae) at Kuala Lompat, Krau Wildlife Reserve, Pahang, Malaysia. *Journal of Wildlife and Parks*. **28**: 31-39.

Nor-Atikah, A.R., Halim, M., Syarifah-Zulaikha, S.A. & Yaakop, S. (2019). Molecular identification and first documentation of seven species of Carpophilus Stephens (Nitidulidae: Carpophilinae) in oil palm ecosystem, Peninsular Malaysia. *Journal of Asia-Pacific Entomology*, **22**(2): 619-624.

Piñol, J., Mir, G., Gomez-Polo, P. & Agustí, N. (2015). Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Molecular Ecology Resources*, **15**(4): 819-830.

Piñol, J., Senar, M.A. & Symondson, W.O.C. (2018). The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative. *Molecular Ecology*, **28**(2): 407-419.

Ratnasingham, S. & Hebert, P.D.N. (2007). BOLD: The Barcode of Life Data System (http://www. barcodinglife. org). *Molecular Ecology Notes*, **7**(3): 355-364.

Razgour, O., Clare, E.L., Zeale, M.R.K., Hanmer, J., Schnell, I.B., Rasmussen, M., Gilbert, T.P. & Jones, G. (2011). High-throughput sequencing offers insight into mechanisms of resource partitioning in cryptic bat species. *Ecology and Evolution*, **1**(4): 556-570.

Sauer, J. & Hausdorf, B. (2012). A comparison of DNA-based methods for delimiting species in a Cretan land snail radiation reveals shortcomings of exclusively molecular taxonomy. *Cladistics*, **28**(3): 300-316.

Vesterinen, E.J., Lilley, T., Laine, V.N. & Wahlberg, N. (2013). Next generation sequencing of fecal DNA reveals the dietary diversity of the widespread insectivorous predator Daubenton's bat (*Myotis daubentonii*) in Southwestern Finland. *PLoS ONE*, **8**(11): e82168.

Wilson, J.J., Rougerie, R., Schonfeld, J., Janzen, D.H., Hallwachs, W., Hajibabaei, M., Kitching, I.J., Haxaire, J. & Hebert, P.D. (2011). When species matches are unavailable are DNA barcodes correctly assigned to higher taxa? An assessment using sphingid moths. *BMC Ecology*, **11**(1): 11-18.

Yu, D.W., Ji, Y., Emerson, B.C., Wang, X., Ye, C., Yang, C. & Ding, Z. (2012). Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution*, **3**(4): 613-623.

Zeale, M.R.K., Butlin, R.K., Barker, G.L.A., Lees, D.C. & Jones, G. (2011). Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces. *Molecular Ecology Resources*, **11**(2): 236-244.