Potential DNA Barcoding of Local Stingless Honeybee (Tantadan) From Kiulu, Sabah using 28S Ribosomal DNA

A M S Abu Bakar1, A A M Faik2*, G Rajendraan2, J A Gansau2, and R Abdulla2
1Veterinary Diagnostics Laboratory, P.O. Box 59, 89457 Tanjung Aru, Kota Kinabalu, Sabah, Malaysia
2Faculty of Science and Natural Resources, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia
E-mail: ainol@ums.edu.my

Abstract. This study aims to use DNA sequence and perform classification of eight Tantadan honey bees species from Sabah that were morphologically identified through Pictorial Identification Guide and Composite Algorithm. The species involved were Tetragonula fuscobalteata, Tetragonula laeviceps, Tetragonula drescheri, Lepidotrigona cf. ventralis, Lepidotrigona terminata, Heterotrigona itama, and Tetragonula melanocephala. Normally, coI gene is used as it has elevated evolutionary rates higher than nuclear genes and is more specific in discriminating between closely related taxa. However, in this study, the nuclear gene was used as only this region could be amplified from this species. DNA barcoding is a method used for identifying unknown species and taxonomic classification at the molecular level using DNA fragments found in the genome of a species that is unique for the purpose of identification. However, morphological identification is also crucial as it provides the means of observing the differences between species based on the naked eyes. The methods used were DNA extraction, amplification of 28S nuclear DNA by PCR method, DNA sequencing and bioinformatics analysis of DNA to complete the identification process of these species based on their genetic material. Homology search of the 28S partial gene sequences revealed only three out of eight species were in agreement with taxonomic classification but five out of eight were not. Those five species are as follows; Tetragonula laeviceps identified as Tetragonula carbonaria, Tetragonula drescheri identified as Lepidotrigona terminata, Lepidotrigona cf. ventralis identified as Tetragonula carbonaria, Heterotrigona itama identified as Heterotrigona bakeri and Lepidotrigona melanocephala identified as Tetragonula carbonaria. The study revealed that 28S fragment of nuclear DNA is a suitable candidate to identify Tantadan honey bees species but it was not distinct enough for specific species identification, thus the result obtained here can be used for further study with primers targeting amplification of coI gene. In a nutshell, this study successfully demonstrated the use of DNA barcode using 28S rDNA in differentiating closely related species.

1. INTRODUCTION
Tantadan honeybees are stingless honeybees which are also widely known as ‘kelulut’, the term used by native people whom refer to this type of beehive, native to the tropical forests and jungles of Malaysia [1]. The nomenclature of the stingless bees was always changing due to diversified ecology or journals that are not found in all the libraries and not accessible to everyone. Also, most of the species normally needs more research on their natural variation, geographical range and inter-related boundaries. The molecular analysis using DNA barcoding is important in the sense that previous research has rarely been...
done in the region of Sabah particularly. This study is used as an attempt to find out the DNA barcode for the collected Tantadan honeybees and study their genetic differences by DNA barcoding. This study also reveals the importance of using DNA barcoding in parallel with morphological study in order to address the important issues on the genetic status of rare and possibly cryptic bee species due to the vast species of Tantadan honeybees.

The position of stingless bees comes under the kingdom of Animalia, phylum belonging to Arthropoda, class of Insecta. The bees, technically the Apiformes or Anthophila, are apparently a superfamily of the Apoidea. The bees are divided into several families, one of which is the Apidae, which includes the large subfamily Apinae, containing the bumblebees, honey bees, and the euglossines and within which is the tribe Meliponini, the stingless bees comes under this tribe [2]. Since the bees that are commonly known to the common public are the Apis honeybees as they cover almost all the region in the world and produce honey also, nevertheless, the stingless bees are tolerant to pests and diseases, smaller in size, do not sting and found in tropical regions compared to true honeybees that are susceptible to pests and diseases, bigger in size, stings and found in almost all the regions [3].

Recognizing the Meliponini is typically easy via their specific frame shape in comparison to other kinds of bees despite the fact that some may additionally seem comparable. A number of the bees are referred to as corbiculae bees due to the fact their females have a corbicula on each hind tibia. In non-corbiculae bees, hairs on the hind leg or on the bottom of the abdomen are used to lure pollen. Both stingless bees and honey bees have a pollen rake at the hind tibia, but whilst each group have a pollen press at the hind basitarsus, stingless bees have a noticeably weakly evolved pollen press whilst as compared to the accelerated pollen press in honeybees. Specific to stingless bees, they have a tuft of hairs on the joint of the tibia and basitarsus called the penicillum. An easy way of distinguishing stingless bees from different bees is through the placement of their hind legs while in flight. Other bees fold their hind legs under the abdomen during in flight whereas the hind legs of stingless bees dangle limply downwards. Additionally, other easy manner to differentiate the stingless bees is through their mandibles [4]. Although there are specific manners in distinguishing the stingless bees, but there are distinct species morphologically similar making them difficult to differentiate thus, molecular based phylogenetic analysis is conducted here to accurately classify them according to there genera.

A 658 bp region of the mitochondrial cytochrome c oxidase subunit I (known as col or cox1) gene has emerged as a proposed potential 'DNA barcode' tool [5]. However, in some studies, other genes have been used such as 16S rRNA and 23S rRNA regions of mitochondrial DNA. Specifically, mitochondrial DNA (mtDNA) seems to be prone specifically to interspecific introgression probably because of difference amongst sexes in mate-choice and dispersal [6]. This contrasts with the nuclear genome where it is approximately 100 000 times larger, where both males and females make contribution to the entire genomes to the gene pool and effective size is consequently proportional to 2 times the whole population size. According to Waikagul & Thaenkham (2014)[7], the genetic markers for molecular systematics are normally the gene or DNA regions of nuclear ribosomal rRNA and the genes of the mitochondrial genome. In their study, it was mentioned that the highly conserved regions of rDNA are useful for the studies of phylogenetic relationships among species from different genera to families.

DNA barcoding refers to the technique of sequencing a short fragment of the mitochondrial cytochrome c oxidase subunit I (col) gene, the "DNA barcode," from a taxonomically unknown specimen and performing comparisons with a reference library of barcodes of known species origin to establish a species-level identification,[8]. However, the DNA barcode fragment can also be from other gene regions such as 16S rRNA (16S ribosomal RNA) regions of mitochondrial DNA and 28S rRNA region of nuclear DNA. In this study, 28S region is used as the DNA barcode fragment for identification of the mentioned species.
DNA sequencing is used for determining the nucleotide sequence of an individual DNA or larger genetic regions of an organism and in sequencing RNA or proteins [9]. BLAST can be used to deduce useful and evolutionary relationships between sequences in addition to assist in identifying members of gene families. Generating a phylogenetic tree from DNA sequences derived from related species can also allow drawing inferences about how those species may be associated.

2. MATERIALS AND METHODS

2.1. Sample collection, identification and preparation
Tantadan honeybees were sampled from Kiulu, Tamparuli, Sabah. Eight species of honeybees were collected from the mentioned region. The samples were identified accordingly to their taxonomy which reprises mainly of their genera and species as the samples are all from the same tribe. The samples were examined under an Image Analyzer (Stereo) Leica M165C with software of Leica Application Suite Ver 4.4.0 and camera of Leica DFC495. Also, the taxonomy of each species was confirmed further by consulting with an entomologist, Dr. Suzan Benedick, Faculty of Sustainable Agriculture from Universiti Malaysia Sabah and by referring to Pictorial Identification Guide and Composite Algorithm compiled by Abu Hassan Jalil & Ibrahim Shuib [4]. There were three replicates for each species and were further stored at -20 °C.

2.2. Extraction of DNA from Tantadan species
The tissues were specifically obtained from the abdomens of each of the Tantadan honeybees. Three replicates of each sample were prepared and they were all labelled R1, R2 and R3 for each sample. DNA was extracted from the prepared samples using the Biolyse DNA Extraction Kit. The kit contains 10X Biolyse DNA lysis Buffer solution, ultrapure water and Proteinase K (20 mg/ml). Proteinase K with volume of 0.5 µL was added into a 1.5 mL centrifuge tube followed by 5 µL of 10X DNA Lysis Buffer solutions and 45 µL of ultrapure water were added along into the mixture [10]. The tissues were added prior to the mixture and were grinded using micro pestle to break the exoskeleton. After the lysis, the mixtures were then incubated at 55 °C for 10 minutes and were left at room temperature for 5 minutes. The lysate was then centrifuged at maximum speed, 13000 rpm for 10 minutes. Twenty microlitres of the lysate was transferred to a new microcentrifuge containing 80 µl of ultrapure water. The DNA samples were stored in -20°C for storage until further use.

2.3. DNA amplification using Polymerase Chain Reaction (PCR), Gel electrophoresis, Quantification and DNA Sequencing of Purified PCR products
In this study, specific set of forward and reverse primers were used, 28SF_D2 as the forward primer and 28SR_D3 as the reverse primer and the sequences are 28SF_D2: 5’-AAG AGA GAG TTC AAG AGT ACG TG-3’ and 28SR_D3: 5’-TAG TTC ACC ATC TTT CGG GTC CC-3’ [12]. Approximately a 600bp PCR product was amplified under the following conditions: 5.0 µl DNA template, 0.25 U Taq Polymerase, 1X PCR Buffer, 1.5 mM MgCl2, 0.2 mM dNTPs and 0.2 mM of forward and reverse primers respectively. The whole PCR was conducted using the MJ Research PTC-200 Thermal Cycler with the initial denaturation at 94°C for 5 minutes, followed by denaturation at 94°C for 30 seconds, primer annealing at 55°C for 30 seconds, extension at 72°C for 30 seconds. The final extension was at 72°C for 5 minutes. The cycle was repeated for 40 times. The PCR products were run under 2% TAE agarose gel. Positive PCR products were carefully excised using a clean razor blade. Extraction of DNA from agarose was carried out using NORGEN DNA gel extraction kit according to manufacturer’s manual. The PCR products were then quantified to determine the concentration, for sample with low DNA concentration, the PCR products were analysed by running under 2% TAE gel electrophoresis. For DNA sequencing purposes, the sample were sent to First Base Laboratories Sdn. Bhd., Malaysia.
2.4. Sequence analysis
Homology search on GenBank for each sequence was identified using BLAST. The most similar sequence was used for determination of the most probable species. DNA sequence of each sample was then inputted into the MEGA6 software for performing analysis between species. Multiple sequence alignment (MSA) was carried out using ClustalW that functions to align sequences based on their homology and provided score. The p-distance was calculated using pairwise distance model to compute the number of base differences between the Tantadan species. The phylogenetic tree was constructed using the Neighbour Joining (NJ) method of Kimura-2 Parameter model to see the proximity of closeness between the Tantadan species.

3. RESULTS AND DISCUSSION

3.1. Individual sequence analysis
Since each species generates approximately different lengths of 28S fragment, within 558-719 bp, with the shortest size was 558 bp for species G.1. Even though the sequence length was shorter than 650 bp, this short sequence was sufficient enough for assigning the sequence to specific species. The different lengths of sequence generate different similarity scores for each replicate although their BLAST results are the same for that species. All the sequences matched with the 28S sequence in the NCBI database with 99% similarity (Table 1). Each honeybee was identified based on their morphology (Figure 1).

From the BLAST result, it can be seen that only three out of eight species, namely G.1, G.4 and G.6 were in agreement with morphological identification by comparison with the low e-value and high sequence similarity scores of the sequences of the species that was submitted to the GenBank. The remaining five species i.e., G2, G3, G5, G7 and G8 were not in agreement with taxonomic identification, as summarized in Table 1. The three replicates for each species show identical results with low e-value but differently similarity scores between them due to the different sequence lengths in base pairs (bp).

Table 1. BLAST result of Tantadan honeybee samples with NCBI database.

| Species                   | BLAST query result                 | E-value | Identities          |
|---------------------------|------------------------------------|---------|---------------------|
| G.1 Tetragonula fuscobalteata | Tetragonula fuscobalteata          | 0.0     | 542/543(99%)       |
| G.2 Tetragonula laeviceps  | Tetragonula carbonaria             | 0.0     | 547/548(99%)       |
| G.3 Tetragonula drescheri  | Lepidotrigona terminata            | 0.0     | 684/692(99%)       |
| G.4 Tetragonula fuscobalteata | Tetragonula fuscobalteata          | 0.0     | 549/550(99%)       |
| G.5 Lepidotrigona cf. ventralis | Tetragonula carbonaria             | 0.0     | 688/693(99%)       |
| G.6 Lepidotrigona terminate | Lepidotrigona terminata            | 0.0     | 690/695(99%)       |
| G.7 Heterotrigona itama    | Heterotrigona bakeri               | 0.0     | 647/655(99%)       |
| G.8 Tetragonula melanocephala | Tetragonula carbonaria             | 0.0     | 691/699(99%)       |
3.2. *p*-distance Between Sequences

Pairwise alignment was done to compare all eight species using MEGA 6 software to observe the degree of divergence between any two sequences by determining the number of base substitution per site from between 24 sequences. The alignment was done using *p*-distance model in which the divergence of two sequences are determined using the ratio of sequence transition and transversion [6]. Table 2 shows the sequence divergence for all the replicates of eight species with the outgroup, *Aphis mellifera ligustica*. From the pairwise alignment shown, the sequence divergence of the 28S sequence of all eight species can be analyzed closely as two 28S sequence were compared and the sequence divergence are being measured by measuring the rate of nucleotide substitution. From the result in Table 2, it can be seen that the sequence divergence is quite high between all the species. However, some replicates from the species, *Lepidotrigona terminata*, *Tetragonula fuscobalteata* and *Tetragonula carbonaria*, showed low sequence divergence.

**Figure 1.** Tantadan honeybee species identified based on taxonomic identification; a) G.1_ *Tetragonula fuscobalteata*, b) G.2_ *Tetragonula laeviceps*, c) G.3_ *Tetragonula drescheri*, d) G.4_ *Tetragonula fuscobalteata*, e) G.5_ *Lepidotrigona cf. ventralis*, f) G.6_ *Lepidotrigona terminata*, g) G.7_ *Heterotrigona itama* and h) G.8_ *Tetragonula melanocephala*. 
Table 2. The sequence divergence between three replicates of eight species computed using the p-distance method

|                | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  | 13  | 14  | 15  | 16  | 17  | 18  | 19  | 20  | 21  | 22  | 23  | 24  | 25  |
|----------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| G.1_Tetragonula_fuscobalteata_285 | 0.002 | 0.004 | 0.002 | 0.002 | 0.002 | 0.007 | 0.004 | 0.004 | 0.001 | 0.009 | 0.004 | 0.007 | 0.006 | 0.010 | 0.006 | 0.005 | 0.006 | 0.007 | 0.008 | 0.009 | 0.010 | 0.008 | 0.006 | 0.005 | 0.004 | 0.003 |
| G.1_Tetragonula_fuscobalteata_285 | 0.002 | 0.004 | 0.002 | 0.002 | 0.002 | 0.007 | 0.004 | 0.004 | 0.001 | 0.009 | 0.004 | 0.007 | 0.006 | 0.010 | 0.006 | 0.005 | 0.006 | 0.007 | 0.008 | 0.009 | 0.010 | 0.008 | 0.006 | 0.005 | 0.004 | 0.003 |
| G.1_Tetragonula_fuscobalteata_285 | 0.002 | 0.004 | 0.002 | 0.002 | 0.002 | 0.007 | 0.004 | 0.004 | 0.001 | 0.009 | 0.004 | 0.007 | 0.006 | 0.010 | 0.006 | 0.005 | 0.006 | 0.007 | 0.008 | 0.009 | 0.010 | 0.008 | 0.006 | 0.005 | 0.004 | 0.003 |

3.3. Phylogenetic Tree of Nucleotide Sequences

The evolutionary history was inferred by constructing a Neighbour-Joining tree by taking the barcode fragment as the molecular marker [6]. Kimura-2 parameter is used as the substitution model to observe the rate of nucleotide substitution between all eight species. From the phylogenetic tree in Figure 2, it can be observed that the 28S sequence for each Tantadan species were clearly discriminated from each other and the outgroup, Apis mellifera ligustica is located far away from the other eight species indicating that there is high sequence divergence between this species and the other eight given sample. The sequences that formed a clade showed a low sequence divergence between those sequences whereas the sequence that is in a different node indicated a high sequence divergence between those sequences. Lepidotrigona terminata and Heterotrigona bakeri forms a clade together showing low sequence divergence and these two species is closer to the outgroup compared to the other two species in different node showing high sequence divergence between the species [11]. Tetragonula carbonaria and Tetragonula fuscobalteata falls in the same clade but the branches show high sequence divergence between these species despite being in the same genus.

The discrepancy between taxonomy and molecular phylogeny is suggestive of either incorrect species identification, DNA contamination, or error in uploading sequence to the GeneBank database [2]. For example, taxonomically identified as G.1_Tetragonula fuscobalteata and G.8_Tetragonula melanoccephala are identified as Tetragonula fuscobalteata BLAST query, although these species are otherwise genetically and morphologically disparate. Also the NCBI taxonomic database are not always updated with the latest classifications which results in this study to be similar despite the differences morphologically. This could be prevented by doing a manually accurate sequence records as reflected in the study of [2].
According to [3], the stingless honeybees are suitable for comparative studies between species of the various adaptations in highly eucosial bees due to the diversity of their evolutionary responses to specific ecological challenges. The importance of this study is because colonies tend to adapt to their environment to maximize the food intake into the nest by adjusting their foraging strategy according to the respective environmental situation, including climatic factors, presence or absence of competitors, identity of competitors, and resource abundance and quality. However, this works differently in different species, thus, some species are able to collect resources under weather conditions that prohibit the activity of others, thereby avoiding interference competition at the food sources [11]. This difference of behaviour can be further identified through evolutionary relationship between species by increasing both taxonomic sampling and the number of genes sampled for phylogenetic analysis to obtain a more accurate reading and avoiding descrepancy between taxonomy and molecular phylogeny [12].

![Figure 2](image.png)

**Figure 2.** The evolutionary history of Tantadan species was inferred using the Neighbor-Joining method. The evolutionary distances were computed using the Kimura 2-parameter method and are in the units of the number of base substitutions per site. The analysis involved 25 nucleotide sequences.

### 4. SUMMARY

Three out of eight species of Tantadan honeybees from the region Kiulu, Tamparuli, Sabah was identified similarly by morphological and molecular identification. However, morphological identification of the other five species was not similar to the BLAST results. The difference was possibly due to incorrect species identification, DNA contamination, or error in uploading sequence to the GeneBank database. Also this could happen as stingless honeybees might have been identified as a different species by the person who submitted their sequence into the database and as further
recommendation, classification into subspecies are required as these species are very different morphologically.

From this study, DNA barcoding is indeed helpful to recognize a species and assign them according to their taxonomy just with a short fragment of 28S fragment. The 28S region was successfully extracted and amplified by PCR. Also, the phylogenetic tree showed the species were classified under a total of seven clades. However, this study can be further analyzed to obtain a more accurate identification of each species by finding the accurate barcode sequence for specific species identification especially with coI gene. Thus, from the preliminary results that were collected from this study; morphological identification, molecular identification and phylogenetic analysis, DNA barcoding with coI gene can be done for more accurate species identification.

5. References
[1] Mohd Faik AA, Lee JW, Lee PC, Abu Bakar AMS, Matthew Huan YM, Lee YK 2018 Application of DNA Barcoding in Species Identification of Pieridae Family from Entopia Penang Butterfly Farm. ASM Sci J. 11 31–9.
[2] Hedtke SM, Patiny S and Danforth BN 2013. The bee tree of life: a supermatrix approach to apoid phylogeny and biogeography. BMC Evolutionary Biology 13 138.
[3] Hrncir M, Jarau S and Barth, FG 2016 Singless bee (Meliponini): sense and behaviour. Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioural Physiology. 202 597-601.
[4] Jalil AH and Ibrahim S 2012 Indo Malayan Stingless Bee Pictorial Identification Guide and Composite Algorithm.
[5] Kerr WE and Maule V 1964 Geographic distribution of stingless bees and its implications (Hymenoptera: Apidae). Journal of the New York Entomological Society, 72 2-18.
[6] Kumar GR, Reybroeck W, Veen VJW and Gupta, A 2014 Beekeeping for poverty alleviation and livelihood security, vol. 1: Technological aspects of beekeeping.
[7] Waikagul, J & Thaenkham, U 2014 Methods of Molecular Study. In Approaches to Research on the Systematics of Fish-Borne Trematodes, Elsevier, pp. 77-90.
[8] Noll FB 2002 Behavioral phylogeny of corbiculate Apidae (Hymenoptera; Apinae), with special reference to social behaviour. Cladistics, 18 137-153.
[9] Roubik DW 2013 Why do they keep changing the names of our stingless bees (Hymenoptera: Apidae; Meliponini)? A little history and guide to taxonomy Smithsonian Tropical Research Institute, Balboa, Ancon, Republic of Panama, 1-7.
[10] Sambrook J, Fritsch EF and Maniatis, T 1989 Molecular cloning: A laboratory manual. Second edition. Volumes 1, 2, and 3. Current protocols in molecular biology. Volumes 1 and 2. Cold Spring Harbor, New York. Cold Spring Harbor Laboratory Press, 16.
[11] Schwarz H 1937 Results of the Oxford University Sarawak (Borneo) expidition: Borneone stingless bees the genus of Trigona. Bulletin of the American Museum of Natrual History, 73 281-329.
[12] Tamura K, Stecher G, Peterson D, Filipski A and Kumar S 2013 MEGA 6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular Biology and Evolution, 30 2725-2729.