

Original Article

Asian Pacific Journal of Tropical Medicine

doi: 10.4103/1995–7645.348157

Impact Factor: 3.041

Genetic variation of sand flies (Diptera: Psychodidae) in Gampaha and Kurunegala districts of Sri Lanka: Complementing the morphological identification

Tharaka Wijerathna¹, Nayana Gunathilaka^{1✉}, Wasana Rodrigo²¹Department of Parasitology, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka²Department of Zoology, Faculty of Natural Sciences, The Open University of Sri Lanka, Nawala, Nugegoda, Sri Lanka

ABSTRACT

Objective: To identify the variation of sand flies in the Gampaha and Kurunegala districts of Sri Lanka and to assess DNA barcoding as a complementing method for morphological identification.

Methods: A total of 38 441 sand flies were collected from selected localities in Gampaha and Kurunegala districts using standard entomological techniques from May 2017 to December 2018. Specimens were identified using morphological features and compared with mitochondrial cytochrome C oxidase subunit I gene-based DNA barcoding as an alternative tool.

Results: Morphological and molecular identification confirmed the presence of four species under two genera (*Phlebotomus* and *Sergentomyia*). *Phlebotomus argentipes* was the predominant species, followed by *Sergentomyia* (*S.*) *punjabensis*, *S. babu insularis*, and an unidentified *Sergentomyia* sp. *Phlebotomus argentipes* showed a clear genetic differentiation from other species. *S. babu insularis* and *S. punjabensis* showed a higher genetic affinity to each other than the unidentified species. The unidentified *Sergentomyia* species is morphologically similar to *S. zeylanica*, but differs only in clavate gonostyle.

Conclusions: DNA barcoding is an effective technique for the identification of sand flies. Further studies using molecular techniques will improve the knowledge of the cryptic diversity of Sri Lankan sand fly fauna. Establishing a reliable and standardized identification system for sand fly species in Sri Lanka is recommended.

KEYWORDS: Morphology; Sand fly; Identification; Barcoding; Sequencing

1. Introduction

Phlebotomine sand flies are of significant public health

importance, especially in tropical areas of the world due to the transmission of leishmaniasis[1]. The global sand fly fauna consists of over 800 species classified under five major genera: *Phlebotomus*, *Sergentomyia*, *Lutzomyia*, *Brumptomyia* and *Warileya*[2]. However, some species in genera *Phlebotomus* and *Lutzomyia* have been experimentally confirmed as vectors for leishmaniasis transmission[2,3]. In Sri Lanka, 22 species of sand flies have been identified so far[4]. However, only *Phlebotomus* (*P.*) *argentipes* has been characterized as the vector for disease transmission[5,6].

Accurate identification of sand flies in a disease endemic area is one of the main aspects that need to be addressed when defining the potential risk for disease transmission. Although morphology and morphometry based taxonomical keys are more useful for field surveillance, this has several limitations since it needs recognition and analysis of many structures, primarily in the

Significance

Twenty-two species of sand flies have been recorded in Sri Lanka. Most of these confirmations were mainly based on morphological characteristics. However, there could be some genetically variant species at different geographical locations with similar morphological features. This study documents the morphological and molecular level description of sand flies that were collected from different areas in the country and suggests that the confirmation of species identified by molecular-based approaches is of paramount importance to document the diversity.

✉To whom correspondence may be addressed. E-mail: n.gunathilaka@kln.ac.lk

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-Non Commercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

©2022 Asian Pacific Journal of Tropical Medicine Produced by Wolters Kluwer-Medknow.

How to cite this article: Wijerathna T, Gunathilaka N, Rodrigo W. Genetic variation of sand flies (Diptera: Psychodidae) in Gampaha and Kurunegala districts of Sri Lanka: Complementing the morphological identification. Asian Pac J Trop Med 2022; 15(7): 322-331.

Article history: Received 6 April 2022

Revision 14 June 2022

Accepted 21 June 2022

Available online 11 July 2022

head and genitalia[7]. On the other hand, species identification is complicated since it entails considerable skill and taxonomic expertise. Further, deterioration of samples during the collection, transport, preservation or improper mounting techniques hinders the morphological confirmation of species. If the specimen is undescribed, morphological keys will be of little help in detecting that. However, in such situations, the most promising approach for identification is DNA barcoding[7]. The most commonly used method is the amplification and sequencing of the sand fly mitochondrial cytochrome oxidase subunit I (*CO I*) gene, followed by the comparison with the NCBI database of known sequences[7].

Sri Lanka is a tropical island located in the Indian Ocean off the southern tip of the Indian peninsula. The island has an area of approximately 65 610 km² and has diverse habitats with a complex topography and variable rainfall patterns[8]. Sri Lanka is home to a rich diversity of fauna and flora with a high endemicity. Geographical separation and isolation of subpopulations can lead to speciation[9]. Distinct morphological characters can also evolve with time and such differentiation requires a long time since alterations in the morphological traits necessitate changes in multiple genes[10]. However, some genetically variant species could be at different geographical locations with similar morphological features. Therefore, confirmation of species identified by molecular-based approaches is of paramount importance in documenting the diversity of an organism. Despite the importance, the molecular characterization of sand flies is a poorly addressed area in Sri Lanka. Thus, the current study focused on the molecular level characterization and assessing the genetic diversity of phlebotomine sand flies in two endemic foci: Gampaha district and Kurunegala district in Sri Lanka.

2. Materials and methods

2.1. Ethical approval

Ethical clearance for the present study was obtained from the Ethics Review Committee, Faculty of Medicine, University of Kelaniya, Sri Lanka (Ref. No. P/204/12/2016).

2.2. Study areas

The Kurunegala district and Gampaha district showed a significant increase in leishmaniasis cases in the last five years[11]. Kurunegala district extends from the intermediate zone to the dry zone areas of Sri Lanka[12]. The average rainfall of the district ranged from 900 to 2 200 mm. The area receives rain from South Western (from May to August) and North Eastern Monsoons (October to January). The average annual temperature is 31.7 °C, while the relative humidity

ranges from 59% to 74% throughout the year, with a yearly average of 69.6%. The disease prevalence was considerably higher in some MOH areas, such as Polpithigama, Maho and Galamuwa MOH areas (Figure 1). These sites were selected for the field collection of sand flies.

Gampaha district is located in the wet zone of Sri Lanka[13]. The mean annual rainfall stands at 1 423 mm with a mean annual temperature of 27.3 °C (22.7-34.3 °C). This district is a new focus of leishmaniasis in Sri Lanka with a considerable increase in the patients during the last two years[14]. Therefore, considering the prevalence of disease incidence and occurrence of sand flies, Mirigama, Ragama and Aththanagalla MOH areas were selected for field surveillance (Figure 1).

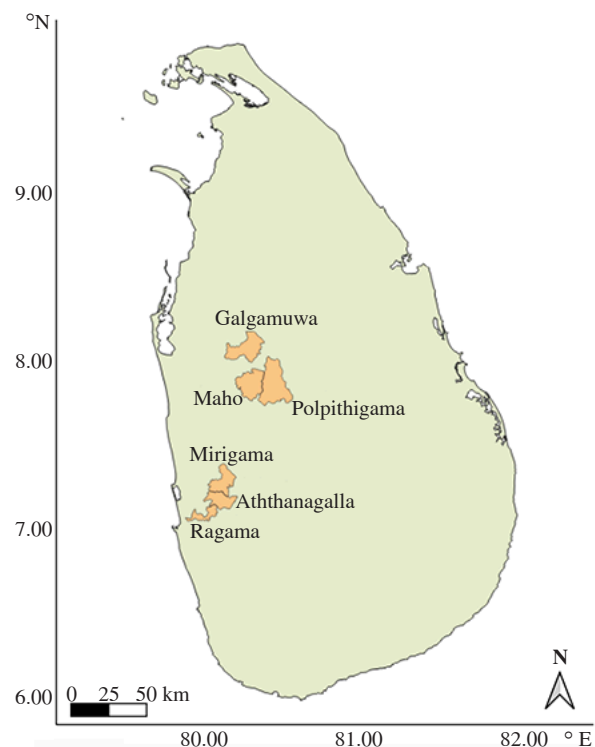


Figure 1. Locations of the sand fly collection sites Sri Lanka.

2.3. Collection of sand flies

Field entomological surveillance was carried out monthly from May 2017 to December 2018. Sand flies were collected from selected localities in each district using four different field collection techniques: hand collection (HC), light traps (LT), sticky traps (ST) and cattle baited net trap (CBNT) according to the entomological field techniques specified by the World Health Organization[15].

2.4. Specimen processing and morphological identification

The specimens were dissected on a glass slide with sterile saline using sterile dissecting needles separating the head, wings and the

distal end of the abdomen from the rest of the body. These parts were mounted in Hoyer's medium and used for morphological identification. The remaining parts of the specimens, including the thorax with legs and proximal parts of the abdomen were transferred individually to sterilized 1.5 mL microcentrifuge tubes and used for DNA extraction.

Mounted sand fly specimens were examined using a dissecting microscope (Labomed CZM4 Binocular Zoom Microscope, Lobo American Inc, USA, 10× to 40×) and separated according to sex. Specimens were dissected on a glass slide with saline separating the terminal part of the abdomen, wings and the entire head using a fine needle and mounted with Hoyers medium for later identification. The sand flies were identified based on morphometric and meristic characters[16]. All morphometric measurements were taken using an ocular micrometre mounted to the binocular microscope. Ten groups of sand flies were selected depending on the species and locality of collection. This includes two species (*S. punjabensis* and *P. argentipes*) collected from all 3 sites of Kurunegala District (6 groups) and 2 species (Unidentified *Sergentomyia* sp. and *S. babu insularis*) from Mirigama (2 groups), one species from Ragama (1 group) and one species (Unidentified *Sergentomyia* sp.) from Aththanagalla (1 group). Three randomly chosen specimens (only 2 if not available) from each group were used for barcoding analysis.

2.5. DNA extraction

MightyPrep reagent for DNA (Takara Bio Inc, Japan) was used for DNA extraction according to the manufacturer's guide with some minor modifications. A volume of 100 µL from MightyPrep reagent was added into each microcentrifuge tube with the thorax and the proximal parts of the abdomen of the sand fly. The specimens were crushed individually using a sterile 200 µL pipette tip. New pipette tips were used every time to prevent cross-contamination. The samples were subjected to a hard vortex for 10 seconds, followed by incubation at 95 °C for 20 minutes. The lysates were allowed to cool down to room temperature and subjected to another hard vortex for 10 seconds. Finally, the DNA extracts were centrifuged at 8000×g for 10 minutes and stored at 20 °C until use for the PCR assays.

2.6. Amplification of the genomic DNA of the field caught sand flies

The sand fly DNA was amplified using standard insect DNA barcoding primers LCO 1490 (Forward primer: 5'-GGTCAACAAATCATAAAGATATTGG-3') and HCO 2198 (Reverse primer: 5'-AAACTTCAGGGTGACCAAAAATCA-3')[17] targeting mitochondrial *CO I* of the sand fly (at least 650 bp). The amplification was carried out using MightyAmp DNA polymerase version 3 kit (Takara, Japan) according to the manufacturer's instructions, subjected to minor modifications.

The reactions were carried out in a volume of 20 µL of the solution, with 1 µL of DNA extract as the template, 10 µL of MightyAmp Buffer V3 containing 4 mM MgCl₂, dNTP (600 µM of each), 2 µL of 10× additives, 2 µL of Rediload dye (Invitrogen, USA), 0.12 µL of 50 µM of each primer (forward and reverse), 0.4 µL of 2.5 U MightyAmp DNA polymerase and 4.36 µL of nuclease-free water.

Amplification was performed in a thermal cycler (SimpliAmp™, Applied Biosystems) programmed for an initial heating step at 98 °C for 2 minutes, 40 cycles of denaturation at 98 °C for 10 seconds, annealing at 55 °C for 15 seconds, extension at 68 °C for 40 seconds, and a final cooling step at 4 °C for 5 minutes.

2.7. Genomic DNA sequencing

The PCR amplicons were purified using QIAquick PCR Purification Kit (Qiagen) and those samples were sent to Macrogen, South Korea (Macrogen Inc., 1001, 254 Beotkkot-ro, Geumcheon-gu, Seoul, South Korea) for *CO I* partial gene sequencing with the same universal primers (LCO 1490 and HCO 2198) and sequenced using the Sanger method.

2.8. Nucleotide sequence analysis

The consensus sequences were generated for each specimen by editing the obtained chromatograms to remove overlapping peaks at the two ends using BioEdit 7.2 (Informer Technologies, Inc.). The DNA sequences were aligned with Clustal W tool[18] of Molecular Evolutionary Genetics Analysis 10.0 (Pennsylvania State University)[19]. Each sequence was compared with other mitochondrial genomes using the BLAST (NCBI, USA) programme to confirm the species or genus level identity.

The DnaSP6 programme[20] was used to obtain descriptive information on the sequences, such as the number of polymorphic sites and haplotypes per species. The model suitable for the evolutionary relationship analysis was statistically selected using the maximum likelihood method in Molecular Evolutionary Genetics Analysis 10.0. The lowest values of Bayesian Information Criterion and Akaike Information Criterion, corrected were observed for Tamura-Nei (TN93) model[21], incorporating the shape parameter of the gamma distribution (G=0.3034). Intra and interspecies divergences of nucleotide sequence and composition were calculated using the selected model as well as Kimura 2-Parameter (K2P) model[22], which is the traditionally employed model in DNA barcoding result analysis. A dendrogram was constructed by the neighbour-joining (NJ) method according to the genetic distances calculated via different evolutionary models. Branch support for NJ was calculated using the bootstrapping method with 1000 replicates. The topology of the resulting phylogenetic tree was used to evaluate the species monophyly.

Table 1. Species of sand flies in genera *Phlebotomus* and *Sergentomyia* characterized by DNA barcoding.

Species	Locality	N	H	K2P distance	T93 distance
<i>Phlebotomus argentipes</i>	Polpithigama, Maho, Galgamuwa, Ragama	13	5	0.00	0.00
<i>Sergentomyia punjabensis</i>	Polpithigama, Maho	5	5	0.02	0.03
<i>Sergentomyia babu insularis</i>	Mirigama	2	2	0.00	0.00
<i>Sergentomyia</i> sp. (unidentified)	Mirigama, Bemulla	5	4	0.00	0.01

N=number of specimens; H=number of haplotypes.

Table 2. Minimum and maximum interspecific genetic distances according to two models used: K2P and TN93.

Species	Minimum		Maximum	
	K2P	TN93	K2P	TN93
<i>Phlebotomus argentipes</i>	0.184	0.187	0.211	0.215
(<i>Sergentomyia babu insularis</i>)	(<i>Sergentomyia babu insularis</i>)	(unidentified <i>Sergentomyia</i> sp.)	(unidentified <i>Sergentomyia</i> sp.)	
<i>Sergentomyia punjabensis</i>	0.129	0.131	0.206	0.210
(<i>Sergentomyia babu insularis</i>)	(<i>Sergentomyia babu insularis</i>)	(<i>Phlebotomus argentipes</i>)	(<i>Phlebotomus argentipes</i>)	
<i>Sergentomyia babu insularis</i>	0.129	0.131	0.184	0.187
(<i>Sergentomyia punjabensis</i>)	(<i>Sergentomyia punjabensis</i>)	(<i>Phlebotomus argentipes</i>)	(<i>Phlebotomus argentipes</i>)	
<i>Sergentomyia</i> sp. (unidentified)	0.147	0.151	0.211	0.215
(<i>Sergentomyia babu insularis</i>)	(<i>Sergentomyia babu insularis</i>)	(<i>Phlebotomus argentipes</i>)	(<i>Phlebotomus argentipes</i>)	

3. Results

3.1. Morphological characterization

A total of 38441 sand flies were collected during the entomological surveillance. According to the morphological characterization, this collection consisted of four species under two genera, i.e., *Phlebotomus* and *Sergentomyia*. The most common species was *P. argentipes* (n=38 218, 99.42%), followed by *Sergentomyia* (*S.*) *punjabensis* (n=192, 0.50%), *Sergentomyia* sp. which was not able to characterise to the species level by morphology (n=28, 0.07%) and *S. babu insularis* (n=3, 0.01%). The female to male ratios were 1: 6.25, 1:2.62 and 1:8.33 for *P. argentipes*, *S. punjabensis* and *Sergentomyia* sp., respectively. The *S. babu insularis* collection consisted of only male specimens.

Female *Phlebotomus* was characterised by the wings being broader and asymmetrical, top to bottom. Also, the cibarium is unarmed or with scattered spicules without a pigment patch (Figure 2A). The sand flies of the genus *Sergentomyia* were characterised by having narrow and lanceolate wings (when the wing folded in half from

top to bottom, the two halves would fit on top of each other) and cibarium having one or more rows of teeth (Figure 2B-2D). The pigment patch is usually presented in *Sergentomyia* sp.

The style had five spines in male *Phlebotomus* specimens (only *P. argentipes*) and subterminal spines were presented (Figure 3A). In all the collected *Sergentomyia* spp., the number of spines of the style was four. If all the spines are not terminal, two spines are terminal and two are sub-terminal and arranged in pairs (Figure 3B-3D). The morphological characterisation of the encountered species is described below.

3.1.1. *Phlebotomus argentipes* female

Head of the spermatheca does not consist of a neck. Spermatheca has around 16 segments and the apical segment is enlarged (Figure 4). Both the dorsal and lateral sides of the thorax are blackish. Wing overlap (R1 overlap with R2/complete length of R2) is approximately 0.2. The wing index (R2/R2+3) is around 2. The ratio between ascoid lengths to antennal flagellomere is approximately 0.4-0.5.

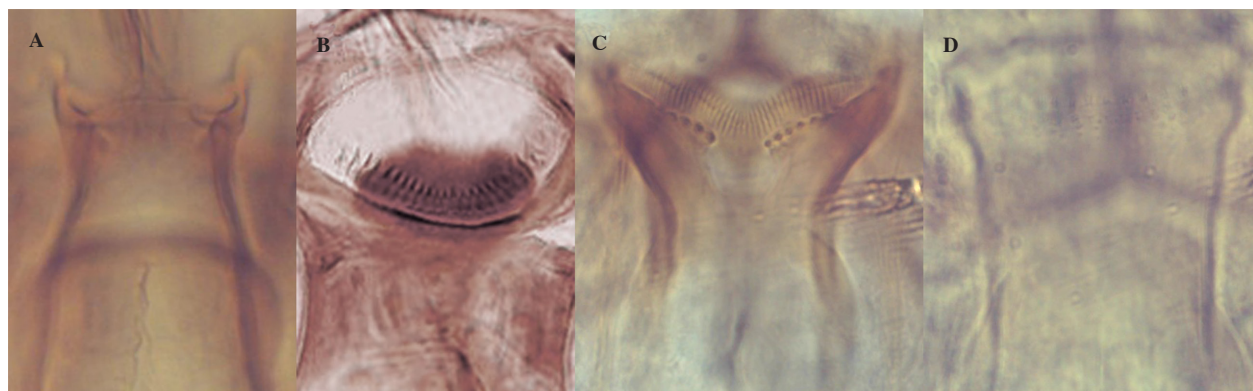


Figure 2. Cephalic structures of collected sand fly species ($\times 1000$). A. *Phlebotomus argentipes*; B. *Sergentomyia punjabensis*; C. *Sergentomyia babu insularis*; D. *Sergentomyia* sp.

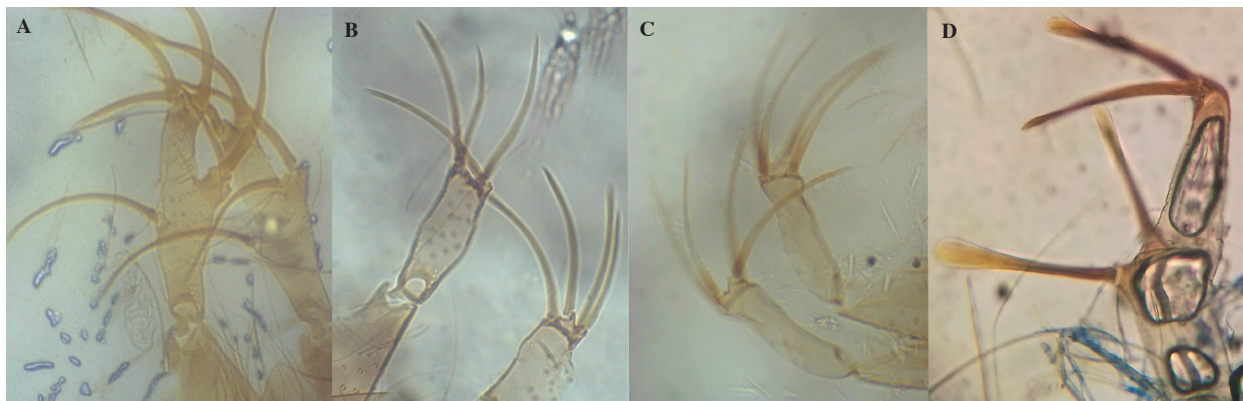


Figure 3. Terminal structures of collected sand fly species (×400). A. *Phlebotomus argentipes*; B. *Sergentomyia punjabensis*; C. *Sergentomyia babu insularis*; D. *Sergentomyia* sp.

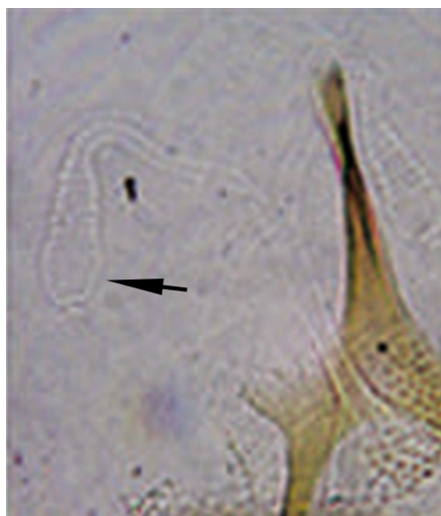


Figure 4. Spermatheca of *Phlebotomus argentipes* (×1000). The arrow indicates enlarged apical segment.

3.1.2. *Phlebotomus argentipes* male

The paramere consists of 2 ventral processes. The style has five long spines, two of which are terminal and three are subterminal (Figure 3A). The ratio between gonocoxite and gonostyle lengths ranges from 1.72 to 1.74.

3.1.3. *Sergentomyia punjabensis* female

The spermatheca is tubular and does not have a capsule or striation. The pharynx is broad at the posterior end and deep constriction at the base. Cibarium consists of approximately 29–30 teeth (Figure 2B).

3.1.4. *Sergentomyia punjabensis* male

The *Aedeagus* is thick and finger-shaped (Figure 5) and all the cibarial teeth are arranged in one line.

3.1.5. *Sergentomyia babu insularis* female

Spermatheca consists of a smooth capsule. The pharynx has distinctly pointed teeth. Cibarium has a deep notch in the hind end of the ventral plate and cibarium with more than 45 teeth (Figure 2C).

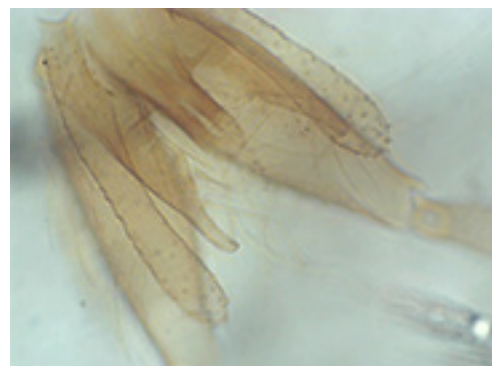


Figure 5. Terminalia of *Sergentomyia punjabensis* (×400).

3.1.6. Unidentified *Sergentomyia* female

Spermatheca is tubular and consists of striations. Cibarium has three rows of fore-teeth (Figure 2D). Cibarium does not have a deep notch but the pigment patch is prominent.

3.1.7. Unidentified *Sergentomyia* male

Aedeagus gradually tapers to the end. Paramere has hairy tubercles on the ventral side. Genital filaments have narrow ends. Gonostyle has four spines in pairs: one is terminal and the other is subterminal. These specimens are similar to *S. zeylanica* from the *Aedeagus* shape, paramere features, and the arrangement of the gonostyles. However, the shape of the spines differ, which are clavate in the new specimens (Figure 3D), whereas it was pointed in *S. zeylanica*.

3.2. Molecular characterization

A total of 25 sequences of 681 nucleotide pairs of the sand fly *CO I* gene belonged to four species under two genera. One species from the genus *Phlebotomus* and three species in the genus *Sergentomyia* were among the recorded species. The most common species was *P. argentipes*, the known vector for *Leishmania donovani* in the South-Asian region. The three species of genus *Sergentomyia* were *S. babu insularis*, *S. punjabensis* and an unidentified *Sergentomyia* sp., which could be closely related to *S.*

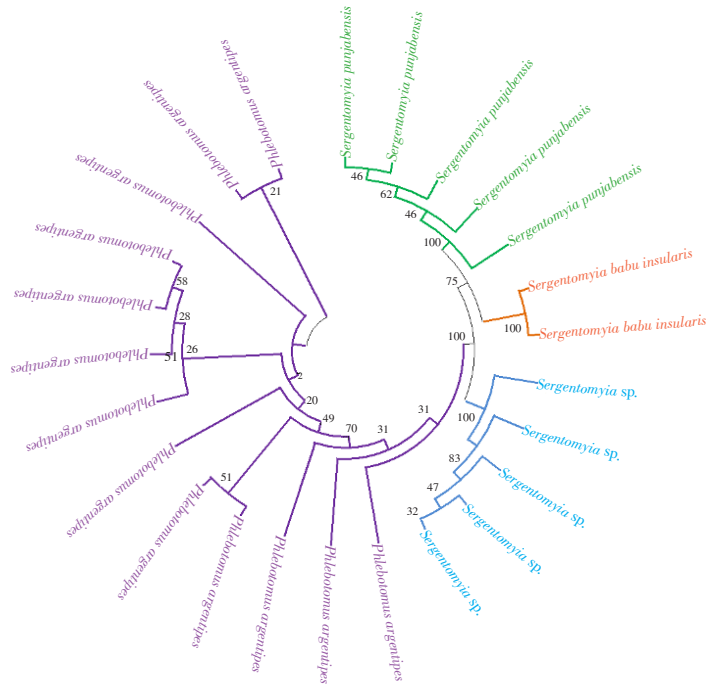


Figure 6. Evolutionary relationships of the sand flies inferred by Neighbour-Joining method using evolutionary distances computed according to the Tamura-Nei (TN93) model.

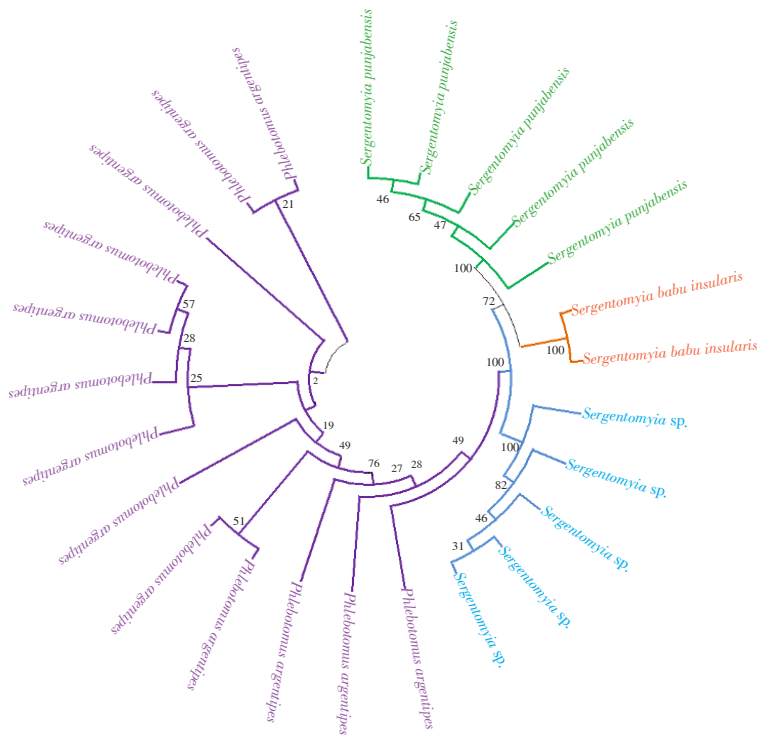


Figure 7. Evolutionary relationships of sand flies inferred by the Neighbour-Joining method using evolution.

koloshanensis of the subgenus *Neophlebotomus* with an approximate 88% similarity (Table 1). Analysis of the nucleotide sequences of DNA fragments provided a coverage of 100% and a similarity of 98%-100% with the *CO I* of Old World sand flies characterized in previous studies[23].

3.3. Nucleotide composition of sand fly *COI* gene

The majority of the nucleotide composition in the *CO I* gene of the sand flies was made up of adenine (A), which contributed to 37.71% of the total nucleotides. About 28.63% of the nucleotide has been made up of thymine, while guanine and cytosine contributed to 18.11% and 15.54% of the total nucleotides, respectively.

3.4. Interspecific and intraspecific variations

The multiple alignments of nucleotides indicated the presence of 166 polymorphic sites, 10 of them were singleton sites and 156 corresponded to phylogenetically informative sites with up to three variants distributed along the alignment. Analysis of intraspecific variations indicated that haplotypes per species varied from 2 to 5, where the highest values were observed for *P. argentipes* and *S. punjabensis*, followed by unidentified *Sergentomyia* sp. and *S. babu insularis* (Table 1).

The high degree of polymorphism observed in the nucleotide sequence remained in amino acid sequences. Multiple alignments of 215 amino acids indicated the presence of 108 variable sites, where 102 of them are parsimoniously informative sites with up to six variants and six singleton sites. Nevertheless, no stop codons, insertions or deletions were encountered in the amino acid sequence.

3.5. Phylogenetic analysis

The intraspecific genetic distance with the K2P model was 0.02 for *S. punjabensis*, while it was zero for all three other species (Table 1). The calculation of intraspecific genetic distance by the TN93 model showed 0.03 for *S. punjabensis* and 0.01 for the unidentified *Sergentomyia* sp. As observed for the K2P model, the genetic distance was zero for the other species. The interspecific genetic distance calculated using the K2P model ranged from 0.129 to 0.211. The genetic distance between species was ranged from 0.131 to 0.215 according to the TN93 model (Table 2).

The dendrograms generated by the neighbour-joining (NJ) method with TN93 (Figure 6) and K2P (Figure 7) indicate that the two genera clustered separately and specimens from the same species clustered together. However, within the genus *Sergentomyia*, *S. punjabensis* and *S. babu insularis* are more closely related than the other unidentified *Sergentomyia* sp. All the species level separations were confirmed with 100% bootstrap values.

4. Discussion

Morphological identification of sand flies is somewhat complicated due to laborious microscopic examinations and dissections of very tiny specimens[24]. These procedures are highly time-consuming and require considerable skill and taxonomic expertise. Furthermore, morphological identification is limited as it is not ideal for damaged samples and identification may be difficult due to improper mounting of specimens. Therefore, molecular level identification has received wider attention for insect authentication[7].

In general, the use of a conserved region in the genome such as *CO I* and internal transcribed spacer 2, which are used for the identification of various medically important insects[23,25]. The *CO I* is considered the ideal primers for dipteran insects due to internal transcribed spacer 2 sequences that do not discriminate against some closely related species and sibling species[26]. Primers used in the present study targeting the *CO I* region have been used for mosquitoes and sand flies[23,27]. These primers (LCO 1490/HCO2198) have been used in many sand fly barcoding studies[7,23,28]. The amplification of *CO I* nuclear pseudogenes of mitochondrial origin is one of the major issues encountered in DNA barcoding restriction profiles and sequence ambiguities in polymorphic sites. Furthermore, in conditions where both template strands are sequenced, the presence of frameshift mutations, stop codons, and unexpected phylogenetic relationships can also be seen[29]. Therefore, such potential can cause false interpretations. In the present study, no stop codons, insertions or deletions were encountered. Hence, the probability of co-amplification of nuclear pseudogenes of mitochondrial origin as well as pseudogenes can be excluded.

The nucleotide sequence length in the present investigation was 681 bp, in agreement with previous studies[29]. The relatively longer sequences ensure the accuracy in sequence-based identification of sand fly species. However, this is not consistent across other studies, as some investigations have reported shorter fragments, such as 549 bp and 548 bp[28,30], or longer lengths of sequences with 700 bp and 689 bp[7,23].

In the sequence of sand fly DNA, the content of A and T bases was higher (66.34%) than G and C (33.66%). These average values of A+T percentage observed during the current study were closer to the values reported previously for Dipterans[31–33]. The A+T values observed in this study were similar to the values observed for *Lutzomyia* species in South America[7,30] as well as *Sergentomyia* and *Phlebotomus* species from South Asia and the Mediterranean islands[23,34,35].

Codon usage indicates that the A+T at the third codon position is more common. This is reported widely for other sand flies[7,23,35] and supported by the fact that the nucleotide composition of insect mitochondrial DNA is correlated with codon usage because genes encoding mitochondrial proteins exhibit a preference for

the use of codons rich in A+T[36,37]. This has been observed in other gene fragments of sand fly genome, such as NADH-ubiquinone oxidoreductase chain 4 (A+T=72.5%), cytochrome b (A+T=73.09%), serine transfer RNA (A+T=81.20%)[38–40].

The intraspecific genetic divergence from species in the same genus is usually associated with the geographical scale of the sampling site distribution[41,42]. The highest percentage of haplotypes were observed for *S. babu insularis* and *S. punjabensis*. However, the sample number was very low for both species due to the lower abundance. This prevents the comparison of intraspecific genetic diversity between species collected during the study. The interspecific genetic divergence was lower between the unidentified *Sergentomyia* sp. and *S. babu insularis*, which were collected from closer proximities. Therefore, it is apparent that these results align with the previous findings that reported the association of the geographical scale of sampling with the intraspecific genetic divergence.

The substitutions in variable positions were largely synonymous and of the transitional type. This means a purine base substitute a purine base and a pyrimidine base are substituted by a pyrimidine base at the third position of the codon. Due to the changes in the third position are not chemically significant and the lack of tRNA specificity at wobble position, it does not change the amino acid that codon represents at the protein synthesis complex. Thus, does not affect the amino acid composition of the protein. However, the interspecific level sequence analyses indicated a gradual increase in substitution rates. Thus, the nucleotide substitution patterns could be considered a reliable diagnostic feature for species differentiation[23].

It is important to note that the species denoted as the unidentified *Sergentomyia* sp. was not morphologically similar to any of the *Sergentomyia* species reported previously from Sri Lanka. Similar morphological characters shared by members of species complexes make the identification a difficult task based on morphological taxonomic keys. One of the other limitations in the morphological characterization is that some specimens may classify to the species level that shares similarities to most of the features in the existing identification key. Therefore, deviations in the morphology may not be considered and some new or unrecorded species may be neglected in the surveys. Hence, it is a need for an alternative, universally-applicable method to support the existing sand fly identification methods, which cannot be resolved through morphological characterization.

According to the *CO I* gene sequencing, the closest one was *S. koloshanensis* of the subgenus *Neophlebotomus* with an approximate 88% similarity. This may be because the *Sergentomyia* species observed in the present investigation is a new species or the *CO I* gene has not been sequenced yet; thus, absent in the NCBI

database. The identification of these specimens based on the morphology may lead to a wrong confirmation as *S. zeylanica* when considering the features of spermatheca and cibarium and the arrangement of gonostyles. However, these specimens differ from *S. zeylanica* in their characteristic shape of the gonostyle. None of the other described sand fly species is known to have clavate gonostyles. The similarity in morphology to *S. zeylanica* also depicts that this species may be a morphological variant of *S. zeylanica* or limited sequence availability as references for comparison have restricted its usage on species identification. These new sequences would also contribute to the ongoing global effort to standardize DNA barcoding as a molecular means of species identification by the Consortium for the Barcode of Life[23].

In summary, the findings of the present study suggested that *CO I* based DNA barcode can effectively be used when morphological traits of certain sand fly species do not clearly distinguish one species from another. Therefore, DNA barcodes also facilitate the taxonomists to re-confirm the reference voucher specimens. Hence, *CO I* -based molecular characterization would be a useful complementary tool for the identification of sand fly species.

In this study, the applicability of *CO I* -based DNA barcoding in the identification of sand fly species in Sri Lanka was evaluated and the genetic variation of sand flies collected from two leishmaniasis endemic districts of Sri Lanka was characterized. Identification based on morphology and *CO I* -based barcoding confirmed four sand fly species encountered in study areas belonging to 2 genera. The most common species was *P. argentipes*, and the other three species were *S. babu insularis*, *S. punjabensis* and an unidentified *Sergentomyia* sp. Intraspecific genetic variation was higher for *P. argentipes* and *S. punjabensis*, collected from sites spread over a broader geographical scale than other species. *P. argentipes* show a clear genetic differentiation from other species. In the genus *Sergentomyia*, *S. babu insularis* and *S. punjabensis* showed a higher genetic affinity to each other than the unidentified species. Therefore, it is recommended to establish a reliable and standardized identification system for sand fly species found in Sri Lanka.

In the collection of sand flies, some of the species had a smaller sample size due to the lower abundance in the environment. Repetitive collections did not include any new specimens from that species. Furthermore, the sample size had to be limited due to funding constraints. The sampling was done only in two districts considering the disease prevalence and the logistic issues. Expanding this study into other areas would be a needed next step. However, despite the limitations, this provides valuable information to conduct more studies highlighting the importance of further studies focused towards cryptic diversity of sand flies in Sri Lanka.

Conflict of interest statement

The authors declare that they have no competing interests.

Acknowledgments

National Research Council, Sri Lanka and Research Council, University of Kelaniya, Sri Lanka are acknowledged for supporting the research activities.

Funding

This study was funded by the National Research Council, Sri Lanka (Grant No. NRC 16-142) and University of Kelaniya, Research Council (Grant No. RC/SROG/2021/03).

Authors' contributions

TW conducted the laboratory investigations, analysed the data and compiled the manuscript. WR processed sequencing data and supervised the study. NG supervised the research work, reviewed the manuscript and made critical revisions. All authors read and approved the final version of the manuscript.

References

- [1] Killick-Kendrick R. The biology and control of Phlebotomine sand flies. *Clin Dermatol* 1999; **17**(3): 279-289.
- [2] Ready PD. Biology of Phlebotomine sand flies as vectors of disease agents. *Annu Rev Entomol* 2013; **58**(1): 227-250.
- [3] World Health Organisation. *Control of the leishmaniasis: Report of a meeting of the WHO Expert Committee on the Control of Leishmaniasis*. Geneva: World Health Organization; 2010.
- [4] Wijerathna T, Gunathilaka N. Phlebotomine sand flies (Psychodidae: Diptera) of Sri Lanka: A review on diversity, biology and bionomics. *J Insect Biodivers* 2019; **11**(2): 41-58.
- [5] Wijerathna T, Gunathilaka N, Gunawardena K, Fujii Y, Gunasekara D. Detection of *Leishmania donovani* DNA within field-caught Phlebotomine sand flies (Diptera: Psychodidae) in three cutaneous leishmaniasis endemic foci of Kurunegala District, Sri Lanka. *J Trop Med* 2021. doi: <https://doi.org/10.1155/2021/6650388>.
- [6] Senanayake SASC, Abeyewicreme W, Dotson EM, Karunaweera ND. Characteristics of Phlebotomine sandflies in selected areas of Sri Lanka. *Southeast Asian J Trop Med Public Health* 2015; **46**(6): 994-1004.
- [7] Gutiérrez MAC, Vivero RJ, Vélez ID, Porter CH, Uribe S. DNA barcoding for the identification of sand fly species (Diptera, Psychodidae, Phlebotominae) in Colombia. *PLoS One* 2014; **9**(1): e85496.
- [8] Pathirana HDNC. Geology of Sri Lanka in relation to Plate Tectonic. *J Natl Sci Found* 1980; **8**(1): 75-85.
- [9] Avise JC. *Phylogeography: The history and formation of species*. 1st ed. Cambridge: Harvard University Press; 2000.
- [10] Ridley Mark. *Evolution*. 3rd ed. New Jersey: Wiley-Blackwell; 2004.
- [11] Epidemiology Unit, Ministry of Health, Sri Lanka. *Weekly epidemiology reports*. [Online]. Available from: http://www.epid.gov.lk/web/index.php?option=com_content&view=article&id=148&Itemid=449&lang=en. [Accessed on 9 February 2022].
- [12] Department of Meteorology, Sri Lanka. *Climate of Sri Lanka*. [Online]. Available from: http://www.meteo.gov.lk/index.php?option=com_content&view=article&id=94&Itemid=310&lang=en&lang=en. [Accessed on 9 February 2022].
- [13] Climate Change Secretariat, Sri Lanka. *Sri Lanka climate profile*. [Online]. Available from: http://www.climatechange.lk/Climate_Profile.html. [Accessed on 9 February 2022].
- [14] Mallawarachchi CH, Chandrasena N, Wijerathna T, Dalpadado R, Mallawarachchi MSMNS, Gunarathna DGAM, et al. An investigation of a new cutaneous leishmaniasis endemic area in Western Sri Lanka. *Trans R Soc Trop Med Hyg* 2021; **115**(11): 1288-1297.
- [15] Kalra NL, Bang YH. *Manual on entomology in visceral leishmaniasis*. 1st ed. New Delhi: World Health Organization; 1988.
- [16] Wijerathna T, Gunathilaka N. Morphological identification keys for adults of sand flies (Diptera: Psychodidae) in Sri Lanka. *Parasit Vectors* 2020; **13**: 450.
- [17] Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome C oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol* 1994; **3**(5): 294-299.
- [18] Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 1997; **25**(24): 4876-4882.
- [19] Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 2018; **35**(6): 1547-1549.
- [20] Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol* 2017; **34**(12): 3299-3302.
- [21] Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 1993; **10**(3): 512-526.
- [22] Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 1980; **16**(2): 111-120.
- [23] Kumar NP, Srinivasan R, Jambulingam P. DNA barcoding for identification of sand flies (Diptera: Psychodidae) in India. *Mol Ecol Resour* 2012; **12**(3): 414-420.
- [24] Galati EAB, Galvis-Ovallos F, Lawyer P, Léger N, Depaquit J. An illustrated guide for characters and terminology used in descriptions of

- Phlebotominae (Diptera, Psychodidae). *Parasite* 2017; **24**: 26.
- [25]Rivera J, Currie DC. Identification of nearctic black flies using DNA barcodes (Diptera: Simuliidae). *Mol Ecol Resour* 2009; **9** (Suppl 1): 224-236.
- [26]Moin-Vaziri V, Oshaghi MA, Yaghoobi-Ershadi MR, Derakhshandeh-Peykar P, Abaei MR, Mohtarami F, et al. ITS2-rDNA sequence variation of *Phlebotomus sergenti* s.l. (Dip: Psychodidae) populations in Iran. *J Arthropod Borne Dis* 2016; **10**(4): 462.
- [27]López-Rubio A, Suaza-Vasco J, Marcet PL, Ruíz-Molina N, Cáceres L, Porter C. Use of DNA barcoding to distinguish the malaria vector *Anopheles neivai* in Colombia. *Zootaxa* 2016; **4175**(4): 377-389.
- [28]Hoyos RL, Uribe SS, Velez I. Tipificación molecular de especímenes colombianos de *Lutzomyia longipalpis* (Diptera: Psychodidae) mediante DNA Barcodes. *Rev Colomb Entomol* 2012; **38**(1): 134-140.
- [29]Cohnstaedt LW, Beati L, Cáceres AG, Ferro C, Munstermann LE. Phylogenetics of the phlebotomine sand fly group Verrucarum (Diptera: Psychodidae: Lutzomyia). *Am J Trop Med Hyg* 2011; **84**(6): 913-922.
- [30]Azpurua J, De La Cruz D, Valderama A, Windsor D. Lutzomyia sand fly diversity and rates of infection by *Wolbachia* and an exotic *Leishmania* species on Barro Colorado Island, Panama. *PLoS Negl Trop Dis* 2010; **4**(3): e627. doi: <https://doi.org/10.1371/journal.pntd.0000627>.
- [31]Clary DO, Wolstenholme DR. The mitochondrial DNA molecular of *Drosophila yakuba*: Nucleotide sequence, gene organization, and genetic code. *J Mol Evol* 1985; **22**(3): 252-271.
- [32]Lunt D, Zhang D, Szymura J, Hewitt G. The insects and *CO I* gene: Evolutionary patterns and conserved primers for phylogenetic studies. *Insect Mol Biol* 1996; **5**: 153-165.
- [33]Beard CB, Hamm DM, Collins FH. The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. *Insect Mol Biol* 1993; **2**(2): 103-124.
- [34]Srinivasan R, Jambulingam P, Kumar NP. *Sergentomyia (Neophlebotomus) monticola*, a new species of sand fly (Diptera: Psychodidae) from the Western Ghats, Thiruvananthapuram District, Kerala, India. *Acta Trop* 2014; **137**: 74-79.
- [35]Dokianakis E, Tsirigotakis N, Christodoulou V, Poulakakis N, Antoniou M. Identification of wild-caught phlebotomine sand flies from Crete and Cyprus using DNA barcoding. *Parasit Vectors* 2018; **11**(1): 94.
- [36]Crozier RH, Crozier YC. The mitochondrial genome of the honeybee *Apis mellifera*: Complete sequence and genome organization. *Genetics* 1993; **133**(1): 97-117.
- [37]Herback JT, November J. Codon usage patterns in cytochrome oxidase I across multiple insect orders. *J Mol Evol* 2003; **56**(1): 691-701.
- [38]Vivero RJ, Contreras-Gutiérrez MA, Bejarano EE. Cambios en el extremo carboxilo terminal de citocromo b como carácter taxonómico en *Lutzomyia* (Diptera: Psychodidae). *Rev Colomb* 2009; **35**(1): 83-88.
- [39]Vivero RJ, Contreras-Gutiérrez MA, Bejarano EE. Análisis de la estructura primaria y secundaria del ARN de transferencia mitocondrial para serina en siete especies de *Lutzomyia*. *Biomédica* 2007; **27**(3): 429.
- [40]Soto SIU, Lehmann T, Rowton ED, Vélez BID, Porter CH. Speciation and population structure in the morphospecies *Lutzomyia longipalpis* (Lutz & Neiva) as derived from the mitochondrial *ND4* gene. *Mol Phylogenet Evol* 2001; **18**(1): 84-93.
- [41]Wright S. Isolation by distance. *Genetics* 1943; **28**(2): 114-138.
- [42]Bergsten J, Bilton DT, Fujisawa T, Elliott M, Monaghan MT, Balke M, et al. The effect of geographical scale of sampling on DNA barcoding. *Syst Biol* 2012; **61**(5): 851-869.