

RESEARCH

Open Access



Improving the COI DNA barcoding library for Neotropical phlebotomine sand flies (Diptera: Psychodidae)

Laura Posada-López^{1,2*} , Bruno Leite Rodrigues² , Ivan Dario Velez¹  and Sandra Uribe³ 

Abstract

Sand fly species are traditionally identified using morphological traits, though this method is hampered by the presence of cryptic species. DNA barcoding is a widely used tool in the case of insects of medical importance, where it is necessary to know quickly which species are present in a transmission area. Here, we assess the usefulness of mitochondrial cytochrome c oxidase subunit I (COI) DNA barcoding as a practical tool for species identification, correct assignment of isomorphic females, and to evaluate the detection of cryptic diversity that occurs in the same species. A fragment of the COI gene was used to generate 156 new barcode sequences for sand flies from different countries of the Neotropical region, mainly Colombia, which had been identified morphologically as 43 species. The sequencing of the COI gene allowed the detection of cryptic diversity within species and correctly associated isomorphic females with males identified by morphology. The maximum intraspecific genetic distances ranged from 0 to 8.32% and 0 to 8.92% using uncorrected *p* distances and the Kimura 2-parameter (K2P) model, respectively. The minimum interspecific distance (nearest neighbor) for each species ranged from 1.5 to 14.14% and 1.51 to 15.7% using *p* and K2P distances, respectively. Three species had more than 3% maximum intraspecific distance: *Psychodopygus panamensis*, *Micropygomyia cayennensis cayennensis*, and *Pintomyia evansi*. They also were split into at least two molecular operational taxonomic units (MOTUs) each, using different species delimitation algorithms. Regarding interspecific genetic distances, the species of the genera *Nyssomyia* and *Trichophoromyia* generated values lower than 3% (except *Nyssomyia ylephiletor* and *Ny. trapidoi*). However, the maximum intraspecific distances did not exceed these values, indicating the presence of a barcode gap despite their proximity. Also, nine sand fly species were DNA barcoded for the first time: *Evandromyia georgii*, *Lutzomyia sherlocki*, *Ny. ylephiletor*, *Ny. yuilli pajoti*, *Psathyromyia punctigeniculata*, *Sciopemyia preclara*, *Trichopygomyia triramula*, *Trichophoromyia howardi*, and *Th. velezbernali*. The COI DNA barcode analysis enabled the correct delimitation of several Neotropical sand fly species from South and Central America and raised questions about the presence of cryptic species for some taxa, which should be further assessed.

Keywords Phlebotominae, Barcoding, Single-locus, Species delimitation, Molecular taxonomy, Molecular systematics

*Correspondence:

Laura Posada-López
lcposada@usp.br

¹ PECET (Programa de Estudio y Control de Enfermedades Tropicales), Universidad de Antioquia, Medellín, Colombia

² Programa de Pós-graduação em Saúde Pública, Faculdade de Saúde Pública (FSP/USP), São Paulo, SP, Brasil

³ Grupo de Investigación en Sistemática Molecular, Universidad Nacional de Colombia, Campus, Medellín, Colombia

Introduction

The subfamily Phlebotominae Rondani and Berté, in Rondani 1840, comprises about 1060 species worldwide, including 555 in the Neotropical region [1-3], distributed in 23 genera. Within the subfamily, there are species of particular interest due to their role as vectors of *Leishmania* protozoans, the causative agents of leishmaniasis, which is why the taxonomy and



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

systematics of this group are crucial to understanding the biological and ecological aspects that determine patterns and dynamics of the diseases they transmit. Entomological surveillance and species identification are necessary to predict possible risk areas of disease transmission and to adopt more efficient control measures in endemic localities [4].

The species-level identification of sand flies is based on morphological characteristics; however, there are limitations, such as phenotypic plasticity within the same species, the presence of cryptic species, isomorphic females, inappropriate mounting techniques, and specimens damaged during collection and transport. Additionally, a high degree of skill and taxonomic expertise is required to carry out the identification of some taxa. These aspects point to the need for integrative approaches which address morphological, molecular, behavioral, and ecological data to better understand the taxonomic status of this subfamily [5, 6].

The DNA barcoding initiative has become an attractive tool in the case of insects of medical importance, where it is necessary to know quickly and accurately which species are present in a transmission area [7–10]. This tool has been well received due to the connectivity and common language of DNA sequences, which allows researchers from different parts of the world to advance in taxonomic and systematic studies of various groups of organisms, including the vectors of diseases [11–13].

Some sand fly species have already been processed for the cytochrome *c* oxidase subunit I (*COI*) gene and are available in genetic databases for molecular identification of these taxa (e.g., NCBI [National Center for Biotechnology Information] GenBank and BOLD [Barcode of Life Data] Systems). However, there is a knowledge gap in *COI* barcode sequences for some groups, as only a quarter of current species have been sequenced for this marker [13]. Thus, several efforts are underway to make new sequences available, in addition to evaluating their usefulness in the species delimitation within this subfamily [10, 14–20]. The *COI* barcode sequences have revealed the presence of cryptic diversity within sand fly species and enabled the correct association of male–female specimens, but may fail to recognize species previously delimited by morphological characters, especially in the case of recent species, which increases the relevance of studies that evaluate the use of this tool [13]. Here, we aim to assess the usefulness of *COI* DNA barcoding as a practical tool for species identification and correct assignment of isomorphic females, and for evaluating the detection of cryptic diversity within species.

Methods

Sand fly sampling and morphological identification

The sand flies were collected according to the parameters of Colombian Decree 1376, which regulates the permits for specimen collection of biologically diverse wild species for non-commercial research.

The collections were carried out between 2013 and 2016 in nine locations belonging to five departments of Colombia: Amazonas Department: (1) Leticia (69°56'35" S; 4°12'29" W) and (2) Puerto Nariño (70°22'59" S; 3°46'13" W); Antioquia Department: (3) Apartadó (76°37'55" S; 7°53'0.9" E) and (4) Remedios (74°41'38" S; 7°1'39" E); Caldas Department: (5) Norcasia (74°53'20" S; 5°34'27" E), (6) Samaná (74°59'34" S; 5°24'47" E) and (7) Victoria (74°54'45" S; 5°18'59" E); Magdalena Department: (8) Santa Marta (74°11'56" S; 11°14'26" E); and Sucre Department: (9) Ovejas (75°13'37" S; 9°31'32" E). The study locations were selected based on epidemiological studies of leishmaniasis transmission previously carried out by the Program for the Study and Control of Tropical Diseases (PECET). The specimens collected in four Central American countries between 2010 and 2012 were also included; the collections were made as part of a training program carried out by PECET researchers and in association with PAHO/WHO (Pan American Health Organization/Organización Panamericana de la Salud). The locations were as follows: Nicaragua: (10) León/Rota (85° 3'1.7" S; 12°32'53" E); Costa Rica: (11) Limón/San Vicente (84°2'51" S; 9°57'36" E) and (12) Limón/Sibuju (84°2'51" S; 9°57'36" E); Panama: (13) Panama Oeste/Capira-Ollas Arriba (79°54'32" S; 8°48'30" E); and Honduras: (14) Valle/Amapala-El Caracol (87°39'14" S; 13°17'31" E). Figure 1 shows a map of the 14 locations. The collections carried out on private property received verbal permission from the landowners before the sampling.

Sand flies were collected using Centers for Disease Control and Prevention (CDC) light traps located in peridomiciliary environments, preferably near domestic animal shelters, where insects are more abundant, and forested fragments. The traps operated overnight, and were installed at 17:00 and withdrawn at 7:00 the next day.

The insects were killed by freezing at -20°C , and then stored in 70% alcohol. Subsequently, the insects were processed in the Medical and Molecular Entomology Laboratory of PECET at the University of Antioquia. The thorax, legs, and wings were dissected and stored dry at -20°C until they were processed using molecular techniques, while the head and abdomen were slide-mounted in Canadian balsam medium for morphological identification following Galati [1]. The generic abbreviations followed the proposal by Marcondes [21].

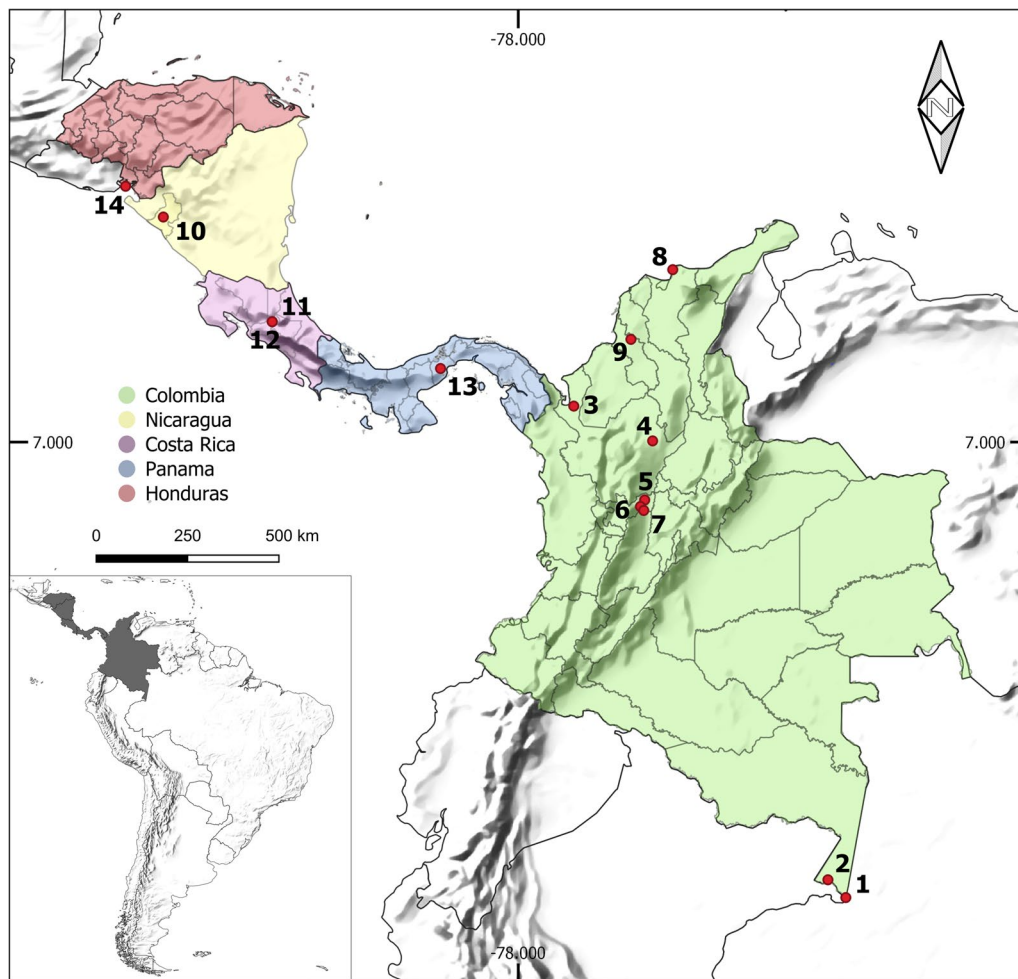


Fig. 1 Map showing the sampling sites of sand flies species used in this study. Colombia: (1) Leticia (Amazonas), (2) Puerto Nariño (Amazonas), (3) Apartadó (Antioquia), (4) Remedios (Antioquia), (5) Norcasia (Caldas), (6) Samaná (Caldas), (7) Victoria (Caldas), (8) Santa Marta (Magdalena), (9) Ovejas (Sucre); Nicaragua: (10) León/Rota; Costa Rica: (11) Limón/San Vicente, (12) Limón/Sibuju; Panama: (13) Panama Oeste/Capira-Ollas Arriba; Honduras: (14) Valle/Amapala-El Caracol

DNA extraction, polymerase chain reaction (PCR), and sequencing

Total DNA from each specimen was extracted from the remaining parts of sand flies (thorax, legs, and wings) using the high salt concentration protocol described by Porter and Collins [22]. The *COI* gene fragment was amplified using the primers LCO1490 (5' GGT CAA CAA ATC ATA AAG ATA TTG G 3') and HCO2198 (5' TAA ACT TCA GGG TGA CCA AAA AAT CA 3') [23]. The PCR products were visualized on electrophoresis using 1% agarose gel and sequenced in both chain directions by Macrogen, Inc. (Korea).

Sequence analysis

The obtained chromatograms were edited using BioEdit v7.0.9 software [24] to generate a consensus sequence

for each specimen. All sequences were aligned using the ClustalW algorithm and then visually examined to ensure there were no stop codons, pseudogenes, or nuclear copies of mitochondrial origin (NUMTs) using MEGA [Molecular Evolutionary Genetics Analysis] v7 software [25]. The *COI* sequences were then submitted to the BOLD Systems database [26] and are available in the “CLBAR—Improving the DNA barcoding library for Neotropical sand flies” project and NCBI GenBank [27] database, being assigned accession numbers OP964207–OP964362.

The sequence alignment was done using MUSCLE (Multiple Sequence Comparison by Log-Expectation) [28] implemented in MEGA v7. Pairwise genetic distances for both maximum intraspecific and minimum interspecific (nearest neighbor, NN) distances were

generated in the BOLD Systems environment using the Barcode Gap Analysis tool with uncorrected (p distances) or the Kimura 2-parameter (K2P) models. The consensus alignment was then used to generate a phenogram using the neighbor-joining (NJ) method with pairwise genetic distances and 1000 bootstrap pseudoreplicates in the software MEGA v7. Also, a phylogenetic gene tree was generated using the maximum likelihood (ML) method in the software RAxML v8 [29] and its graphical user interface, raxmlGUI v2.0 [30]. For the ML tree, the GTR+G+I substitution model was used as suggested by jModelTest v2 [31], and the data were partitioned according to codon position. A sequence of *Sycorax konopiki* (KT946601.1) was included as an outgroup to root the ML tree.

The DNA barcode sequences were also identified at the molecular operational taxonomic unit (MOTU) level, which include groups of specimens based on their molecular similarity at a given molecular marker [32]. Several algorithms were designed to sort barcode sequences of a given dataset into MOTUs without a priori information (i.e., discovery approaches [33]). Therefore, several single-locus species delimitation methods were used to associate morphologically distinct species with MOTUs, evaluating the usefulness of *COI* DNA barcodes for the taxonomy of different sand flies from the Neotropical region. For this, the following methods were employed: (i) automatic barcode gap discovery (ABGD) [34]; (ii) refined single linkage (RESL) [35]; (iii) TCS haplotype networks using statistical parsimony [36, 37]; and (iv) Poisson tree processes (PTP) [38]. The ABGD analysis (available at <https://bioinfo.mnhn.fr/abi/public/abgd/abgdweb.html>) clusters sequences according to their similarity in a given genetic pairwise distance matrix according to inferred barcode gaps in the dataset and then recursively applies this procedure to the obtained MOTUs to obtain finer partitions. Two different ABGD analyses were run using uncorrected p distances and the K2P model, and the parameters $P_{min}=0.005$, $P_{max}=0.1$, and $X=1.0$. For ABGD, it was considered the recursive partitions generated with a range of prior intraspecific divergence between 1% and 2.5% [17]. The RESL algorithm was designed to deal with large amounts of DNA barcode sequences in the BOLD Systems (<https://boldsystems.org/>) and operates by linking similar sequences and then optimizing the MOTU delimitation with a graphic analytical approach using Markov clustering (MCL). RESL analysis was run inside the BOLD environment using the 'cluster sequences' tool and default parameter. The software TCS v1.21 infers haplotype networks using the statistical parsimony method and can be used for species delimitation. This approach can generate disconnected networks for different morphospecies

while analyzing sand fly DNA barcode datasets [39, 40]. The networks inferred by TCS were visualized and edited using the tcsBU web server [41]. Lastly, the PTP algorithm is a coalescent-derived method that seeks to differentiate stochastic population processes from speciation events in a phylogenetic gene tree. Therefore, the MOTU delimitation by PTP was conducted by submitting the ML gene tree (after pruning the outgroup) to the web server (available at <https://species.h-its.org/ptp/>) and using its default settings, except for the number of Markov chain Monte Carlo (MCMC) generations, which was changed to 500,000.

Also, we produced an NJ dendrogram and TCS haplotype networks from an alignment containing only species from the genera *Trichophomyia* and *Nyssomyia*, due to the phylogenetic proximity of these two genera and because they have some different morphospecies that were merged into the same MOTU (see below).

Results

The morphological identification of specimens revealed the presence of 43 species, of which 41 were identified at the species level while the other two were assigned only at subgenus, *Lutzomyia* (*Tricholateralis*) Galati, 2003, and series level, *Psychodopygus* Guyanensis series Barretto, 1962, due to the poor visibility of the morphological characters. They could not be molecularly associated with other sequences from our database or GenBank records, so they were not assigned to any specific taxa. Regarding the species-level identification, the taxa belong to the genera *Brumptomyia*, *Evandromyia*, *Lutzomyia*, *Micropygomyia*, *Nyssomyia*, *Pintomyia*, *Pressatia*, *Psathyromyia*, *Psychodopygus*, *Sciopemyia*, *Trichophomyia*, and *Trichopygomyia*. Of these, 156 new *COI* barcode sequences were generated for phlebotomine sand flies from different countries of the Neotropical region, mainly from Colombia (Table 1). The complete information on the analyzed species and sample locations is listed in Table 1 and Additional file 2: Table S1.

The sequencing resulted in a consensus alignment of 601 base pairs (bp) of the standard *COI* DNA barcode fragment described by Folmer (1994) [23]. The visual inspection of the alignment indicates the absence of stop codons in the middle of sequences, pseudogenes, and/or NUMTs.

The number of barcoded specimens per species ranged from 1 to 21. The maximum intraspecific genetic distances ranged from 0 to 8.32% and 0 to 8.92% using uncorrected p distances and the K2P model, respectively (Table 1). The minimum interspecific distance (NN) for each species ranged from 1.5 to 14.14% and 1.51 to 15.7% using p and K2P distances, respectively (Table 1). Three species had more than

Table 1 Nominal species, collection sites, maximum intraspecific genetic divergence, and the minimum distance to the nearest neighbor of sand fly species from the Neotropical region analyzed in this study

Species	Collection sites ^b	n	Max. intraspecific distance (mean)		Nearest neighbor (NN) species	Distance to the NN	
			p	K2P		p	K2P
<i>Brumptomyia leopoldoi</i> (Rodríguez, 1953)	5	1	0 (N/A)	0 (N/A)	<i>Mi. atroclavata</i>	11.98	13.09
<i>Brumptomyia mesai</i> Sherlock, 1962	2	2	1.16 (1.16)	1.18 (1.18)	<i>Ps. panamensis</i>	12.81	14.06
<i>Evandromyia dubitans</i> (Sherlock, 1962)	13	1	0 (N/A)	0 (N/A)	<i>Ny. umbratilis</i>	11.65	12.66
<i>Evandromyia georgii</i> (Freitas & Barrett, 2002) ^a	1	1	0 (N/A)	0 (N/A)	<i>Pi. evansi</i>	14.14	15.7
<i>Evandromyia saulensis</i> (Floch & Abonnenc, 1944)	2	1	0 (N/A)	0 (N/A)	<i>Mi. micropyga</i>	13.81	15.3
<i>Evandromyia walkeri</i> (Newstead, 1914)	1, 2	6	1.16 (0.64)	1.18 (0.65)	<i>Ev. dubitans</i>	11.81	12.9
<i>Lutzomyia bifoliata</i> Osorno-Mesa, Morales, Osorno & Hoyos, 1970	4, 7	3	0.83 (0.55)	0.84 (0.56)	<i>Mi. atroclavata</i>	11.15	12.07
<i>Lutzomyia gomezi</i> (Nitzulescu, 1931)	4, 6, 8, 9	6	2.5 (1.61)	2.55 (1.64)	<i>Lu. sherlocki</i>	10.82	11.71
<i>Lutzomyia hartmanni</i> (Fairchild & Hertig, 1957)	6, 7	3	0 (0)	0 (0)	<i>Pi. evansi</i>	12.31	13.45
<i>Lutzomyia lichyi</i> (Floch & Abonnenc, 1950)	4	2	0.17 (0.17)	0.17 (0.17)	<i>Lu. bifoliata</i>	11.81	12.87
<i>Lutzomyia longipalpis</i> (Lutz & Neiva, 1912)	14	2	0.5 (0.5)	0.5 (0.5)	<i>Ny. fraihai</i>	10.48	11.3
<i>Lutzomyia sherlocki</i> Martins, Silva & Falcão 1971 ^a	1	4	0.67 (0.33)	0.67 (0.33)	<i>Lu. gomezi</i>	10.82	11.71
<i>Lutzomyia (Tricholateralis) sp.</i> Galati, 2003	1, 2	2	0.33 (0.33)	0.33 (0.33)	<i>Mi. micropyga</i>	13.14	14.45
<i>Micropygomyia atroclavata</i> (Knab, 1913)	9	4	1.33 (0.67)	1.35 (0.67)	<i>Ny. fraihai</i>	10.65	11.49
<i>Micropygomyia cayennensis cayennensis</i> (Floch & Abonnenc, 1941)	9, 13	5	4.49 (1.83)	4.68 (1.9)	<i>Mi. micropyga</i>	11.15	12.1
<i>Micropygomyia chassigneti</i> (Floch & Abonnenc, 1944)	1, 2	3	1.33 (0.89)	1.35 (0.9)	<i>Mi. atroclavata</i>	10.98	11.87
<i>Micropygomyia micropyga</i> (Mangabeira, 1942)	9	2	0.17 (0.17)	0.17 (0.17)	<i>Lu. bifoliata</i>	11.15	12.08
<i>Micropygomyia trinidadensis</i> (Newstead, 1922)	3, 9	4	2.5 (1.28)	2.56 (1.3)	<i>Mi. atroclavata</i>	12.98	14.25
<i>Nyssomyia antunesi</i> (Coutinho, 1939)	2	9	0.67 (0.29)	0.67 (0.29)	<i>Ny. yuilli pajoti</i>	1.5	1.52
<i>Nyssomyia fraihai</i> (Martins, Falcão & Silva, 1979) ^a	1, 2	5	1 (0.5)	1.01 (0.5)	<i>Ny. yuilli yuilli</i>	2.83	2.91
<i>Nyssomyia trapidoi</i> (Fairchild & Hertig, 1952)	4, 6, 12	6	2.66 (1.76)	2.72 (1.79)	<i>Ny. ylephiletor</i>	8.99	9.64
<i>Nyssomyia umbratilis</i> (Ward & Fraiha, 1977)	1	7	0.83 (0.44)	0.84 (0.45)	<i>Ny. yuilli yuilli</i>	2.33	2.37
<i>Nyssomyia ylephiletor</i> (Fairchild & Hertig, 1952) ^a	12	1	0 (N/A)	0 (N/A)	<i>Ny. trapidoi</i>	8.99	9.64
<i>Nyssomyia yuilli yuilli</i> (Young & Porter, 1972)	4, 5, 6	3	0.33 (0.22)	0.33 (0.22)	<i>Ny. umbratilis</i>	2.33	2.37
<i>Nyssomyia yuilli pajoti</i> (Abonnenc, Léger & Fauran 1979) ^a	2	3	0.67 (0.44)	0.67 (0.44)	<i>Ny. antunesi</i>	1.5	1.52
<i>Pintomyia evansi</i> (Nuñez-Tovar, 1924)	9, 10, 14	21	8.32 (2.71)	8.92 (2.83)	<i>Pr. choti</i>	11.48	12.51
<i>Pintomyia rangeliana</i> (Ortiz, 1953)	9	1	0 (N/A)	0 (N/A)	<i>Pi. evansi</i>	12.65	13.92
<i>Pressatia camposi</i> (Rodríguez, 1950)	2	2	1.33 (1.33)	1.35 (1.35)	<i>Pr. choti</i>	5.99	6.3
<i>Pressatia choti</i> (Floch & Abonnenc, 1941)	3, 6	8	0.33 (0.08)	0.33 (0.08)	<i>Pr. camposi</i>	5.99	6.3
<i>Psathyromyia aragaoi</i> (Costa Lima, 1932)	1, 2	2	0.5 (0.5)	0.5 (0.5)	<i>Pa. punctigeniculata</i>	10.82	11.68
<i>Psathyromyia carpenteri</i> (Fairchild & Hertig, 1953)	5	1	0 (N/A)	0 (N/A)	<i>Pa. aragaoi</i>	11.48	12.48
<i>Psathyromyia dendrophyla</i> (Mangabeira, 1942)	2	1	0 (N/A)	0 (N/A)	<i>Pa. shannoni</i>	10.98	11.91
<i>Psathyromyia punctigeniculata</i> (Floch & Abonnenc, 1944) ^a	1	3	0.33 (0.22)	0.33 (0.22)	<i>Ps. panamensis</i>	10.15	10.91
<i>Psathyromyia shannoni</i> (Dyar, 1929)	4	2	1.33 (1.33)	1.35 (1.35)	<i>Pa. punctigeniculata</i>	11.98	13.06
<i>Psychodopygus Guyanensis</i> series Barretto, 1962	2	1	0 (N/A)	0 (N/A)	<i>Ps. panamensis</i>	10.15	10.91
<i>Psychodopygus ayrozai</i> (Barretto & Coutinho, 1940)	6	3	0.17 (0.11)	0.17 (0.11)	<i>Pa. punctigeniculata</i>	11.31	12.32
<i>Psychodopygus hirsutus</i> (Mangabeira, 1942)	2	1	0 (N/A)	0 (N/A)	<i>Ps. panamensis</i>	10.65	11.49
<i>Psychodopygus panamensis</i> (Shannon, 1926)	6, 7, 11, 13	5	3.33 (2.33)	3.42 (2.38)	<i>Pa. punctigeniculata</i>	10.15	10.91
<i>Sciopemyia preclara</i> (Young & Arias, 1984) ^a	2	5	0 (0)	0 (0)	<i>Sc. sordellii</i>	13.14	14.51
<i>Sciopemyia sordellii</i> (Shannon & Del Ponte, 1927)	1	3	0.67 (0.44)	0.67 (0.45)	<i>Mi. atroclavata</i>	10.82	11.68
<i>Trichopygomyia tiramula</i> (Fairchild & Hertig, 1952) ^a	6	2	2.16 (2.16)	2.2 (2.2)	<i>Pi. evansi</i>	11.65	12.68
<i>Trichophoromyia howardi</i> (Young, 1979) ^a	1	2	0 (0)	0 (0)	<i>Th. velezbernali</i>	1.5	1.51
<i>Trichophoromyia velezbernali</i> Posada-López, Galvis & Galati, 2017 ^a	2	3	0.67 (0.44)	0.67 (0.45)	<i>Th. howardi</i>	1.5	1.51

^a Species that had their COI barcode fragment sequenced for the first time^b The locality codes correspond to those listed on the map of the collection sites (Fig. 1)

3% of maximum intraspecific distance: *Psychodopygus panamensis* (3.33% using *p* distances; 3.42% with the K2P model), *Micropygomyia cayennensis cayennensis* (4.49; 4.68), and *Pintomyia evansi* (8.32; 8.92), but their distances to the NNs were 10.15/10.91, 11.15/12.1, and 11.48/12.51, respectively (Table 1). Regarding interspecific genetic distances, the species of the genera *Nyssomyia* and *Trichophoromyia* showed values lower than 3% (except for *Nyssomyia ylephiletor* and *Ny. trapidoi*) for both *p* and K2P distances (Table 1). However, the maximum intraspecific distances did not exceed these values, indicating the presence of a barcode gap despite their proximity.

The NJ phenogram and ML phylogenetic tree grouped conspecific sequences into well-supported clusters/clades for all the analyzed species, sometimes splitting them into more than one group (Fig. 2 and Additional file 1: Fig. S1). Some closely related species of the genera *Nyssomyia* and *Trichophoromyia* formed different clusters comprising only one nominal species each. However, the association of species-level identified males with females of the two *Trichophoromyia* species was somewhat hampered due to the lack of diagnostic characters, which resulted in the absence of a clear clustering pattern of female specimens of this genus, so the four female specimens of *Trichophoromyia* were not assigned to the species level. (Fig. 3). All analyses separated the species *Ny. yuilli yuilli* from *Ny. fraihai*, the latter only collected in the Colombian Amazon. Further, the ML tree indicates that the species pair *Ny. yuilli pajoti*/*Ny. antunesi* may show a paraphyletic pattern (Additional file 1: Fig. S1). On the other hand, NJ analysis split *Ps. panamensis*, *Micropygomyia trinidadensis*, *Pi. evansi*, *Mi. cayennensis cayennensis*, and *Lutzomyia gomezi* into at least two well-supported clades which agree with the samples' geographical location, except for *Lu. gomezi* (Fig. 2).

Regarding species delimitation analyses, the algorithms ABGD (using *p* and K2P distances), RESL, TCS, and PTP clustered the barcode sequences of the 43 morphospecies into 40, 41, 47, 47, and 40 MOTUs, respectively (Fig. 2). ABGD and PTP generated the most conservative partition since, in both cases, five of the nominal species within the *Nyssomyia* genus merged into the same MOTU, and the same happened for *Trichophoromyia howardi* and *Th. velezbernali* (Fig. 2). The algorithms RESL and TCS correctly partitioned the barcode sequences into MOTUs according to the morphospecies but also merged the pairs *Ny. yuilli pajoti*/*Ny. antunesi* (only for TCS), and *Th. howardi*/*Th. velezbernali* (Fig. 2). In contrast, some or all algorithms split *Ps. panamensis*, *Mi. trinidadensis*, *Pi. evansi*, *Mi. cayennensis cayennensis*, and *Lu. gomezi* into at least two well-supported clades

which agree with the well-supported clusters/clades of the NJ and ML analysis (Fig. 2 and Additional file 1: S1).

Discussion

This study helps improve the digital repository of barcode sequences, the Barcode of Life Data Systems (BOLD), which was designed to assemble and organize all barcode sequence records and provide tools for analyzing these sequences [26]. New *COI* barcode sequences were generated, and some of these for sand fly species that had not been previously processed. Furthermore, our results indicate that *COI* DNA barcoding is a useful tool to delimit and identify different sand fly species from the Neotropical region. Different single-locus species delimitation methods were employed to analyze nucleotide divergences from different perspectives, and in almost all cases, the nominal species were assigned as belonging to at least one MOTU.

This study generated the first *COI* sequences for nine sand fly species: *Evandromyia georgii*, *Lutzomyia sherlocki*, *Ny. ylephiletor*, *Ny. yuilli pajoti*, *Psathyromyia punctigeniculata*, *Sciopemyia preclara*, *Trichopygomyia triramula*, *Th. howardi*, and *Th. velezbernali*. All these cases were compared to related species within the same genus, and all seemed to have unique barcode sequences—except for some cases of *Nyssomyia* and *Trichophoromyia* genera—which can be used for future molecular identification of these taxa. Some other studies have evaluated the usefulness of *COI* barcodes in the identification of sand flies from Central America [8, 42] and Colombia [10, 18, 43, 44]. Moreover, in Colombia, the sampling efforts were carried out mainly in the Caribbean and Andean regions of the country, which can have different sand fly fauna compared with the southeast region [45]. Consequently, our study surveyed sand flies from the Amazon region of Colombia, in the municipalities of Puerto Nariño and Leticia, close to the borders of Peru and Brazil. Most of the new DNA barcode records comprise specimens from these locations (Table 1, Fig. 1).

For this new dataset, various algorithms were employed for species delimitation to associate morphologically distinct species with MOTUs. In general, both the NJ/ML analysis and the species delimitation made by ABGD/RESL/TCS/PTP provide consistent results concerning the nominal species (Fig. 2). However, the distance and tree-based methods—ABGD and PTP—were more conservative and merged some closely related species, while RESL and TCS worked well for our dataset, providing more reliable partitions with the sampled species. None of these methods on their own are capable of accurately delimiting evolutionary

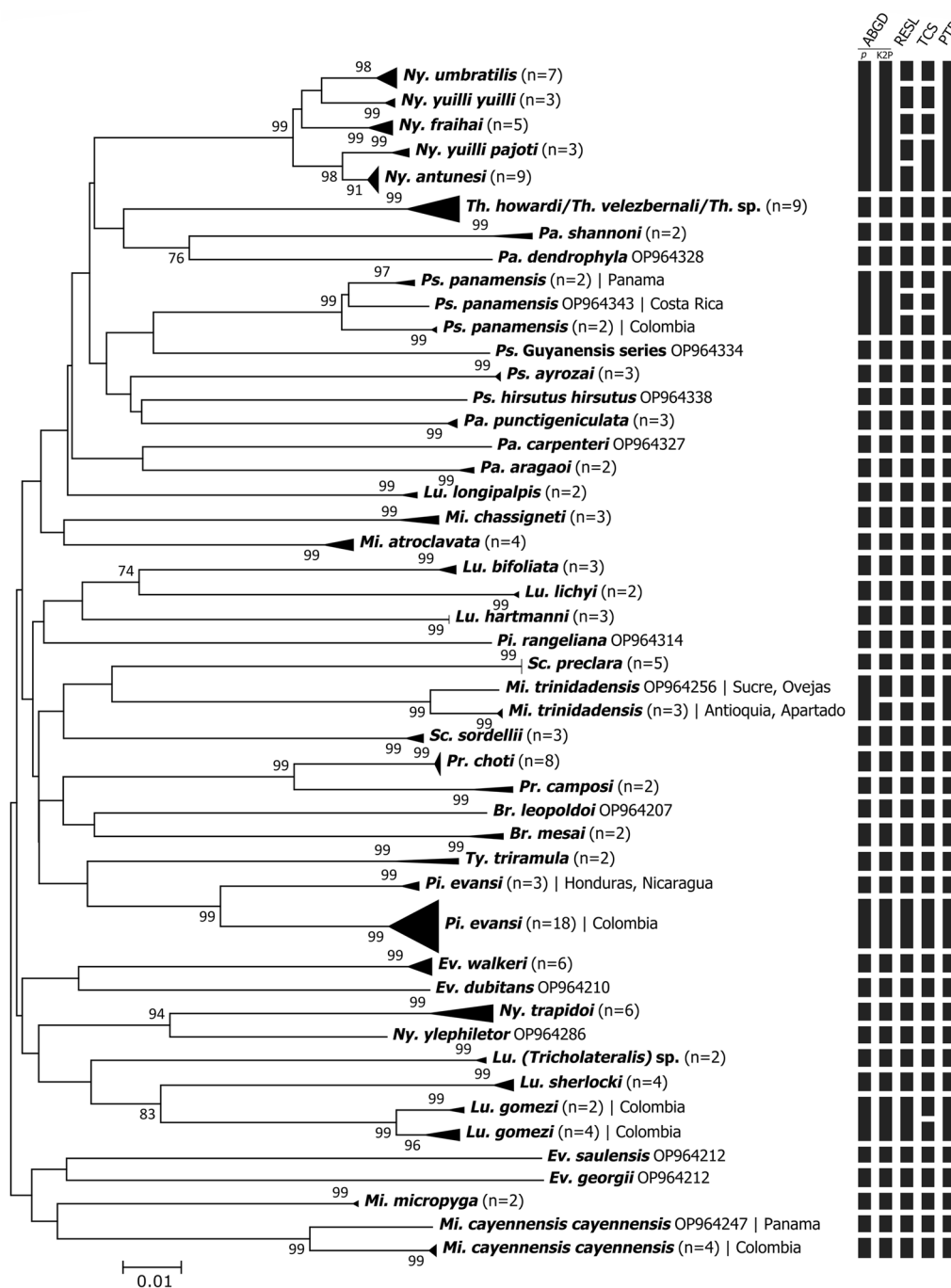


Fig. 2 Neighbor-joining phenogram of *COI* sequences of sand fly species from the Neotropical Region. Numbers near nodes indicate bootstrap values above 70. Lateral black bars indicate the MOTU species delimitation partitions made by the algorithms ABGD, RESL, TCS, and PTP

lineages due to their limitations when analyzing single-locus DNA barcode data without a priori information on the species boundaries, therefore, it is preferable to use a set of these methods and other lines of evidence to propose putative species [33]. The algorithms used here see the data in different ways, so the congruence

between them may indicate that the resulting delimitation is probably correct, but the disagreements should be analyzed with caution [46]. Although some algorithms merged some species within the *Nyssomyia* and *Trichophoromyia* genera, it does not mean that the delimitation of these taxa by morphology is incorrect.

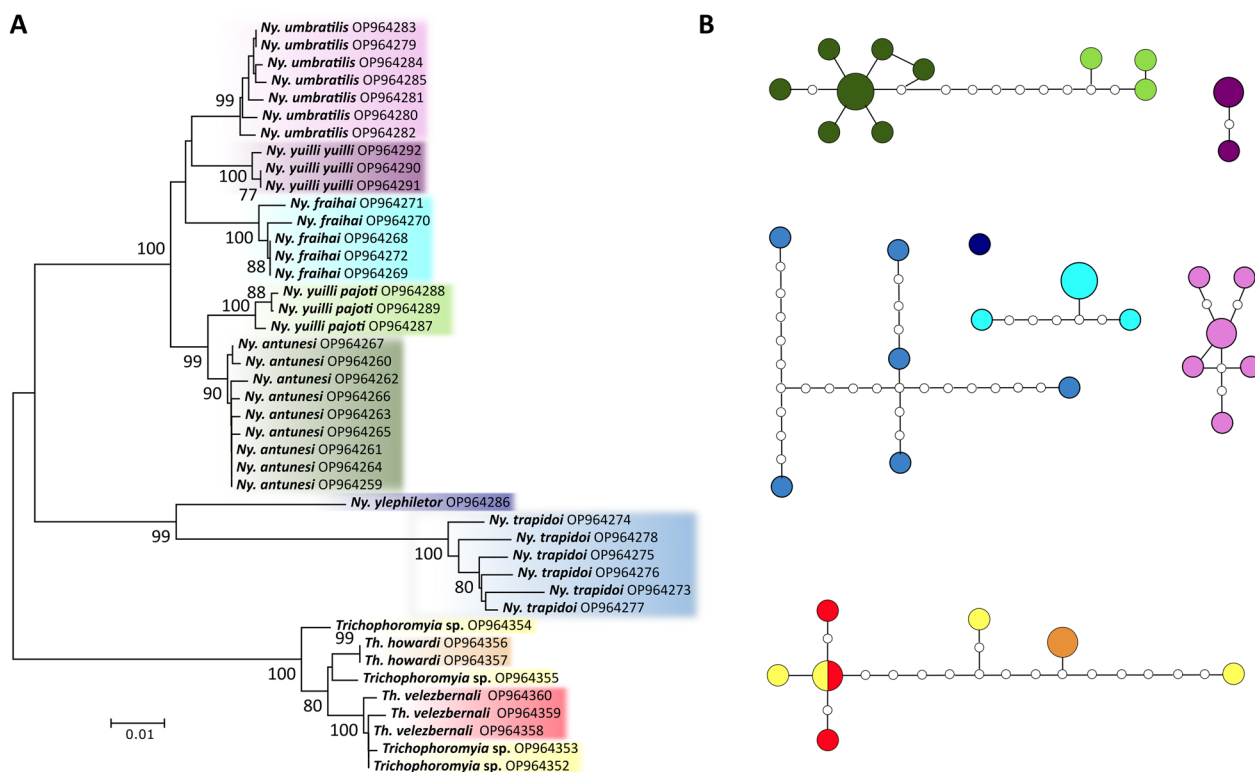


Fig. 3 **A** Neighbor-joining phenogram of *COI* sequences of the sand fly genera *Nyssomyia* and *Trichophoromyia*. Numbers near nodes indicate bootstrap values above 70. Clusters are colored according to the nominal species. **B** TCS haplotype network analysis of *COI* sequences of the sand fly genera *Nyssomyia* and *Trichophoromyia*. Unconnected networks are delimited MOTUs and are colored according to nominal species and NJ patterns. Each circle represents a unique haplotype, with its size proportional to the number of individuals, and the small white circles represent inferred unobserved haplotypes

The pairwise genetic distances—whether uncorrected or using the K2P model—indicate the presence of a “barcode gap” within species and their NN (Table 1). This pattern is usually used to define the *COI* barcode as an excellent molecular marker for species identification [47], including sand flies [10, 16], but the clear distinction between these two classes of distances (intra- and interspecific) may overlap while analyzing larger datasets with closely related taxa [17, 20, 40, 48, 49]. Indeed, it is impossible to establish a generalized standard limit between intra- and interspecies barcode divergence to many groups of organisms. This seems to be true even when analyzing species within the same sand fly subgenus (e.g., *Evandromyia* (*Aldamyia*), [40]). The extent of the overlap and the absence of the so-called barcode gap should not interfere with the rate of successful identifications by DNA barcodes of insects [49].

Regarding the closely related species *Ny. yuilli yuilli* and *Ny. fraihai*, the interspecific pairwise genetic distance (2.83/2.91 for *p* and K2P distances) and NJ analysis indicate the separation into two clusters. These two nominal species have isomorphic females, but *Ny. yuilli*

yuilli is restricted to the Andean and trans-Andean regions of Colombia, while *Ny. fraihai* is widely distributed in Brazil’s Amazon and Atlantic Forest regions [50]. Here, we correctly associated the isomorphic females of the Andean region with males morphologically identified as *Ny. yuilli yuilli*, but the same was not possible for Amazonian specimens, since it was only possible to collect females in these locations. However, these Amazonian females were assigned as *Ny. fraihai*, as the collection sites border Brazilian states located in the Amazon region, in which this species has already been reported [51]. A unique *COI* barcode sequence of *Ny. fraihai* (GenBank accession: KP112771) was previously generated from a male specimen collected near a type location in the state of Bahia, in the Atlantic Forest region of Brazil [17, 50], but the nucleotide distance analyses indicated that this individual could represent a different species from those analyzed in this study (data not shown). Therefore, future studies that analyze a greater number of *Ny. fraihai* specimens of different sexes from both biomes, Atlantic Forest, and Amazon, should assess the presence of possible new species of these taxa.

Here, the taxa *Ny. antunesi*, *Ny. fraihai*, *Ny. umbratilis*, *Ny. yuilli yuilli*, *Ny. yuilli pajoti*, *Th. howardi*, and *Th. velezbernali* did not reach even 3% minimum divergence from the NN, which is usually a very low interspecific value, and can be seen in other sand fly species as an intraspecific divergence (Table 1). Nevertheless, the low values did not prevent the formation of individual genetic clusters for all nominal species in the NJ analysis (Fig. 3). However, some species delimitation algorithms merged the above-mentioned *Nyssomyia* and *Trichophoromyia* species into a single MOTU each, and for the species pair *Ny. antunesi*/*Ny. yuilli pajoti*, it was not possible to form monophyletic clades in the ML gene tree (Figure S1). According to the phylogenetic systematic analysis of morphological data, these two genera belong to the Psychodopygina subtribe and are considered the most derived groups [51]. Other studies that analyzed *COI* barcodes of *Nyssomyia* spp. from Brazil indicate that the nucleotide divergence between species in this genus is low and may differ from other sand fly taxa [15, 17, 20]. Despite this, *Ny. trapidoi* and *Ny. ylephiletor* achieved a reasonable degree of interspecific genetic divergence, being correctly delimited in all species delimitation algorithms.

Regarding the molecular taxonomy status of the genus *Trichophoromyia*, little information is found in the literature, and *COI* barcode sequences are publicly available for only three species: *Th. reburra*, *Th. ininii*, and *Th. viannamartinsi*, which were sequenced and analyzed in different studies [10, 17]. Considering the incredible richness of this genus and the fact that most females are isomorphic [51], this knowledge gap must be filled. In the present study, more than one species of *Trichophoromyia* were analyzed for the first time, which appear to have similar, or even smaller, nucleotide distances than those of the genus *Nyssomyia* (Table 1). In fact, all species delimitation algorithms merged *Th. howardi* and *Th. velezbernali* into a single MOTU, but the NJ and TCS analysis indicates the absence of shared haplotypes, at least for male specimens (Fig. 3). Further, some female specimens—which are isomorphic for these two species—could not be correctly associated with males due to a lack of informative characters (Fig. 3). Our sampling effort was not satisfactory for this taxa, and future studies may elucidate the actual taxonomic status of these and other species of *Trichophoromyia* using multilocus efforts, which are highly recommended when there are poly and paraphyletic patterns in the genealogy of the alleles studied due to recent speciation processes [52].

One of the main benefits of integrating molecular data into insect taxonomy is the association of immature life stages of holometabolous taxa with adults and isomorphic females with males identified by morphology [53,

54]. The present study focused only on generating *COI* sequences for adult specimens. Some species of our dataset have females that are indistinguishable using morphological characters, and it was only possible to correctly associate male and female specimens of the taxa *Mi. cayennensis cayennensis*, *Mi. chassigneti*, *Lutzomyia hartmanni*, *Pressatia choti*, *Pr. camposi*, and *Ty. triramura*. These findings are increasingly relevant because the entomological monitoring of sand flies and leishmaniasis in endemic areas is based on species-level identification, especially female specimens, which actively participate in the transmission of pathogens. Highlighting the usefulness of *COI* barcodes in identifying these species may contribute to the use of this tool for monitoring these insects, which can also help identify vector species in studies that assess vector competence/capacity and natural infection by *Leishmania*. In addition, this correct association indicates that the morphological identification of vouchers can be examined in more detail to assess the existence of other morphological characteristics for identifying the sexes that are considered isomorphic until then. Other sand fly DNA barcoding efforts established this molecular marker as relevant for this type of association in the genera *Brumptomyia* [17], *Psychodopygus* [20], and *Phlebotomus* [39]. Therefore, the sequencing of complex groups in which several females are isomorphic, such as the genera *Brumptomyia*, *Lutzomyia*, *Pintomyia* Townsendi series, *Psychodopygus* Chagasi series, *Pressatia*, *Trichopygomyia*, and *Trichophoromyia* should continue to be evaluated.

The sequencing of the *COI* gene allowed the detection of cryptic diversity within species. The species delimitation analysis of RESL and TCS split into at least two MOTUs the species *Ps. panamensis*, *Mi. trinidadensis*, *Mi. cayennensis cayennensis*, *Pi. evansi*, and *Lu. gomezi*, which also achieved high rates of maximum intraspecific pairwise distances (Table 1) and were grouped in well-supported NJ clusters (Fig. 2). In the first four cases, the detection of these genetic lineages may be associated with microevolutionary processes due to isolation by distance and geographic barriers (e.g., Andean region) since these four species presented clusters related to the geographic locations where they were sampled (Fig. 2). This pattern has been seen in studies with a wide geographic distribution of the analyzed species [17, 20, 40, 55, 56] or when clear geographic barriers are assessed, such as Amazonian riverbanks [15, 57] and caves in Thailand [19]. In the present study, a remarkable case comprising specimens of *Pi. evansi*, which has more than 8% of intraspecific genetic distance, were split into two MOTUs, the first comprising sequences from the Colombian department of Sucre and the other with sequences of specimens from Nicaragua and Honduras (Figs. 1

and 2). Also, the comparison of different populations of *Mi. cayennensis cayennensis* and *Ps. panamensis* reinforces the possible isolation of these insects between the countries of South and Central America, since there is an evident structuring mainly between populations of Colombia and Panama. On the other hand, specimens from *Lu. gomezi*, despite the low values of divergence, may represent two sympatric lineages, since different clades were formed with samples from the same locality in both. These results raise the hypothesis that these populations represent distinct species, especially in the case of *Pi. evansi* due to the high genetic divergence, but the findings should be validated using integrative approaches to elucidate the actual taxonomic status of this species. Regardless of whether *Pi. evansi* represents different species, these molecular lineages may be taken into account from an epidemiological point of view because there may be variations regarding the ecological aspects and the vector–parasite interactions [58], and this should also be considered for the species *Ps. panamensis*, *Mi. trinidadensis*, and *Lu. gomezi* due to their vectorial role in transmitting *Leishmania* pathogens in the Neotropical region [59].

The phylogenetic analysis of the *COI* gene allowed the formation of well-supported clades for the nominal species but failed to recover the evolutionary relationships of larger groups. Molecular markers of the mitochondrial DNA (mtDNA) have a relatively high mutation rate, which is appropriate for identifying species and population structures but can fail in phylogenetic reconstructions of supraspecific relationships [60, 61]. Indeed, the DNA barcoding approaches should not claim to establish evolutionary relationships based on a single rapidly evolving molecular marker, and other conserved genes such as ribosomal RNA (rRNA) 28S should be used for this purpose. Beyond that, even for conserved genes, multiple markers must be evaluated to generate species trees rather than gene trees [62]. However, some assumptions regarding *COI* barcode phylogenies can be raised when using appropriate methods of phylogenetic inference, such as Bayesian inference and ML. In the present work, some relationships between species were observed (Additional file 1: Fig. S1), and a well-supported clade was reconstructed for (i) all species of the *Lutzomyia* (*Tricholateralis*) subgenus; (ii) two representatives of the *Brumptomyia* subtribe and *Brumptomyia* genus; (iii) the closely related species *Lutzomyia lichyi* and *Lu. bifoliata*, both of the *Lutzomyia* (*Lutzomyia*) subgenus; (iv) *Pr. camposi* and *Pr. choti*; (v) two species of the *Psathyromyia* (*Forattiniella*) subgenus, *Pa. carpenter* and *Pa. aragaoi*; (vi) *Trichophoromyia howardi* and *Th. velezbernali*;

and (vii) two well-supported clades within the *Nyssomyia* genus that are not necessarily related because of the low support value, the first containing the species *Ny. ylephiletor* and *Ny. trapidoi*, and the other comprising *Ny. fraihai*, *Ny. yuilli yuilli*, *Ny. yuilli pajoti*, *Ny. umbratilis*, and *Ny. antunesi*. In these last two cases, the clades were formed by the species that showed the lowest pairwise divergences in our dataset. There is no morphological evidence that the sampled *Nyssomyia* species are polyphyletic, so other approaches should be used to assess the natural relationships between these taxa.

Furthermore, an interesting grouping pattern was observed for the Psychodopygina group, as all species of this subtribe were grouped into a single clade, despite the low support value. The monophyly of the subtribe Psychodopygina appears to be consistent and has already been demonstrated in studies using multiple genetic markers [63] and analyzing a fragment of the rRNA 28S gene [64]. However, regarding the Phlebotominae subfamily, the efforts to elucidate evolutionary relationships, propose new phylogenetic classifications, or corroborate the existing ones are hampered by the low sampling of molecular markers for different sand fly species, especially the most conserved genes [13].

Conclusion

In summary, the sequencing and analysis of the *COI* DNA barcoding fragment enabled the correct delimitation of several Neotropical sand fly species from South and Central America. New important sequences of sand flies that had not been previously processed for this molecular marker were generated, which increases the relevance of DNA repositories so that more accurate identification of sand flies is possible using integrative tools. The findings of cryptic diversity within *Ps. panamensis*, *Mi. trinidadensis*, *Pi. evansi*, *Mi. cayennensis cayennensis*, and *Lu. gomezi* should be further evaluated to elucidate the possible presence of cryptic species, mainly considering the wide geographic distribution and the epidemiological importance of these species in transmitting pathogens to humans and other vertebrate hosts.

Abbreviations

<i>COI</i>	Cytochrome c oxidase subunit I
mtDNA	Mitochondrial DNA
MOTU	Molecular operational taxonomic unit
ABGD	Automatic barcode gap discovery
RESL	Refined single linkage
PTP	Poisson tree processes
ML	Maximum likelihood
NJ	Neighbor-joining

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13071-023-05807-z>.

Additional file 1: Figure S1. Phylogenetic gene tree based on *COI* DNA barcode sequences of Neotropical sand flies. Numbers near nodes indicate bootstrap values above 70, except the clade comprising *Psychodopygina* species, which is highlighted in bold.

Additional file 2: Table S1. Sample IDs, BOLD processes IDs, GenBank accession numbers, nominal species, sex of the specimens, and collection sites of sand fly species from the Neotropical region analyzed in this study.

Acknowledgements

We would like to offer our special thanks to the field team that carried out the collection, the PAHO for making the trip to the Central American countries possible, and the ecoepidemiology and Medical Entomology Units of PECET. We would also like to thank Andres Velez, Horacio Cadena, Rafael Vivero, and Luz Adriana Agudelo. Finally, we would like to thank PECET for funding this project at all stages and Dra Eunice Galati from FSP/USP Brazil for revising the manuscript.

Author contributions

Study design: LCPL, SIU. Funding acquisition: SIU, IDV. Specimen collection and identification: LCPL, SIU; Sample processing: LCPL. Formal analysis and data interpretation: LCPL, BLR. Original draft preparation: LCPL, BLR. All authors read and approved the final manuscript.

Funding

The study was funded by the Program for the Study and Control of Tropical Diseases (PECET) of the University of Antioquia, Colombia. The findings and conclusions in this report are those of the author(s) and do not necessarily represent the views of PECET.

Availability of data and materials

All sequences obtained from the study were deposited in the GenBank database under the accession numbers: OP964207–OP964362.

Declarations

Ethics approval and consent to participate

Local or regional ethics committee approval was not required for the insect collections.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 31 January 2023 Accepted: 10 May 2023

Published online: 12 June 2023

References

- Galati EAB, Rodrigues BL. Phlebotominae (Diptera, Psychodidae) Classificação, Morfologia, Terminologia e Identificação de Adultos [Phlebotominae (Diptera, Psychodidae) classification, morphology, terminology and adult identification]. *Apostila Disciplina HEP 5752 Bioecologia e Identificação de Phlebotominae*. I. São Paulo: Brasil. Press, Departamento de Epidemiologia Faculdade de Saúde Pública Universidade de São Paulo; 2021. p.126.
- Shimabukuro PHF, Andrade AJ, Galati EAB. Checklist of American sand flies (Diptera, Psychodidae, Phlebotominae): genera, species, and their distribution. *ZooKeys*. 2017;660:67–106. <https://doi.org/10.3897/zookeys.660.10508>.
- Galati EAB, Rodrigues BL. A Review of Historical Phlebotominae Taxonomy (Diptera: Psychodidae) Neotropical Entomology. 2023. <https://doi.org/10.1007/s13744-023-01030-8>
- Bates PA, Depaquit J, Galati EA, Kamhawi S, Maroli M, McDowell MA, et al. Recent advances in phlebotomine sand fly research related to leishmaniasis control. *Parasit Vectors*. 2015;8:1–8.
- Caterino MS, Cho S, Sperling FA. The current state of insect molecular systematics: a thriving Tower of Babel. *Annu Rev Entomol*. 2000;45:1–54.
- Padial JM, Miralles A, De la Riva I, Vences M. The integrative future of taxonomy. *Front Zool*. 2010;7:16. <https://doi.org/10.1186/1742-9994-7-16>.
- Besansky NJ, Severson DW, Ferdig MT. DNA barcoding of parasites and invertebrate disease vectors: what you don't know can hurt you. *Trends Parasitol*. 2003;12:546.
- Azpuruja 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:e627.
- Cohnstaedt LW, Beati L, Caceres A, Ferro C, Munstermann LE. Phylogenetics of the Phlebotomine Sand Fly Group Verrucarum (Diptera: Psychodidae: Lutzomyia). *Am J Trop Med Hyg*. 2011;84:913–22.
- Contreras Gutierrez MA, Vivero RJ, Velez ID, Porter CH, Uribe S. DNA barcoding for the identification of sand fly species (Diptera, Psychodidae, Phlebotominae) in Colombia. *PLoS ONE*. 2014;9:e85496.
- Hebert PD, Gregory TR. The promise of DNA barcoding for taxonomy. *Syst Biol*. 2005;54:852–9.
- Depaquit J. Molecular systematics applied to Phlebotomine sandflies: Review and perspectives. *Infect Genet Evol*. 2014. <https://doi.org/10.1016/j.meegid.2014.10.027>.
- Rodrigues BL, Galati EA. Molecular taxonomy of phlebotomine sand flies (Diptera, Psychodidae) with emphasis on DNA barcoding: a review. *Acta Trop*. 2023;238:106778.
- Kumar NP, Srinivasan R, Jambulingam P. DNA barcoding for identification of sand flies (Diptera: Psychodidae) in India. *Mol Ecol Resour*. 2012;12:414–20. <https://doi.org/10.1111/j.1755-0998.2012.03117>.
- Scarpassa VM, Alencar RB. Molecular taxonomy of the two *Leishmania* vectors *Lutzomyia umbratilis* and *Lutzomyia anduzei* (Diptera: Psychodidae) from the Brazilian Amazon. *Parasit Vectors*. 2013;6:1–1.
- Nzelu CO, Cáceres AG, Arrunátegui-Jiménez MJ, Lañas-Rosas MF, Yañez-Trujillano HH, Luna-Caipe DV, et al. DNA barcoding for identification of sand fly species (Diptera: Psychodidae) from leishmaniasis-endemic areas of Peru. *Acta Trop*. 2015;145–51.
- Pinto IS, Chagas BD, Rodrigues AAF, Ferreira AL, Rezende HR, Bruno RV, et al. DNA barcoding of Neotropical sandflies (Diptera, Psychodidae, Phlebotominae): species identification and discovery within Brazil. *PLoS ONE*. 2015;10:e0140636.
- Romero-Ricardo L, Lastre-Meza N, Perez-Doria A, Bejarano EE. DNA barcoding to identify species of phlebotomine sand fly (Diptera: Psychodidae) in the mixed leishmaniasis focus of the Colombian Caribbean. *Acta Trop*. 2016;1:125–31.
- Sukantamala J, Sing KW, Jaturas N, Polseela R, Wilson JJ. Unexpected diversity of sandflies (Diptera: Psychodidae) in tourist caves in Northern Thailand. *Mitochondrial DNA Part A*. 2016;28:949–55.
- Rodrigues BL, Carvalho-Costa LF, Pinto ID, Rebêlo JM. DNA barcoding reveals hidden diversity of sand flies (Diptera: Psychodidae) at fine and broad spatial scales in Brazilian endemic regions for leishmaniasis. *J Med Entomol*. 2018;55:893–901.
- Marcondes CB. A proposal of generic and subgeneric abbreviations for Phlebotomine sandflies (Diptera: Psychodidae: Phlebotominae) of the world. *Entomol News*. 2007;118:351–6.
- Porter C, Collins F. Species-diagnostic differences in a ribosomal DNA internal transcribed spacer from the sibling species *Anopheles freeborni* and *Anopheles hermsi* (Diptera: Culicidae). *Am J Trop Med Hyg*. 1991;45:271–9.
- Folmer O, Black M, Hoeh WRL, Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol*. 1994;3:294–9.
- Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *In Nucleic acids symposium series*. 1999;41:95–8.

25. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016;33:1870–4.
26. Ratnasingham S, Hebert PD. BOLD: the barcode of life data system. *Mol Ecol Notes.* 2007;7:355–64.
27. Sayers EW, Cavanaugh M, Clark K, Pruitt KD, Schoch CL, Sherry ST, et al. GenBank. *Nucleic Acids Res.* 2021;49:D92–6.
28. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32:1792–7.
29. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30:1312–3.
30. Edler D, Klein J, Antonelli A, Silvestro D. raxmlGUI 20: a graphical interface and toolkit for phylogenetic analyses using RAxML. *Methods Ecol Evol.* 2021;12:373–7.
31. Darrriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods.* 2012;9:772.
32. Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, Floyd R, et al. Defining operational taxonomic units using DNA barcode data. *Phil Trans Royal Soc B Biol Sci.* 2005;360:1935–43.
33. Carstens BC, Pelletier TA, Reid NM, Satler JD. How to fail at species delimitation. *Mol Ecol.* 2013;22:4369–83.
34. Puillandre N, Lambert A, Brouillet S, Achaz G. ABGD, automatic barcode gap discovery for primary species delimitation. *Mol Ecol.* 2012;21:1864–77.
35. Ratnasingham S, Hebert PD. A DNA-based registry for all animal species: the barcode index number (BIN) system. *PLoS ONE.* 2013;8:e66213.
36. Templeton AR, Crandall KA, Sing CF. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III Cladogram Estimat Genet. 1992;132:619–33.
37. Clement M, Posada DCKA, Crandall KA. TCS: a computer program to estimate gene genealogies. *Mol Ecol.* 2000;9:1657–9.
38. Zhang J, Kapli P, Pavlidis P, Stamatakis A. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics.* 2013;29:2869–76.
39. Kasap OE, Linton YM, Karakus M, Ozbek Y, Alten B. Revision of the species composition and distribution of Turkish sand flies using DNA barcodes. *Parasit Vectors.* 2019;12:1–20.
40. Rodrigues BL, Baton LA, Shimabukuro PH. Single-locus DNA barcoding and species delimitation of the sandfly subgenus *Evandromyia* (*Aldomyia*). *Med Vet Entomol.* 2020;34:420–31.
41. Santos AM, Cabezas MP, Tavares AI, Xavier R, Branco M. tcsBU: a tool to extend TCS network layout and visualization. *Bioinformatics.* 2016;32:627–8.
42. Mejía Á, Matamoros G, Fontecha G, Sosa-Ochoa W. Bionomic aspects of *Lutzomyia evansi* and *Lutzomyia longipalpis*, proven vectors of *Leishmania infantum* in an endemic area of non-ulcerative cutaneous leishmaniasis in Honduras. *Parasit Vectors.* 2018;11:1–1.
43. Vivero RJ, Bejarano EE, Estrada LG, Florez F, Uribe Soto S, Edgar Y, Ortega Gómez E, Torres Gutiérrez C, Muskus López C. DNA barcode for identification of immature stages of sand flies (Diptera: Psychodidae) collected from natural breeding sites. *Artículo de revista.* 2019 Sep 25.
44. Carrero-Sarmiento D, Hoyos-López R. Molecular identification and genetic diversity of *Lutzomyia gomezi* (Diptera: Psychodidae) using DNA barcodes in Córdoba. *Colombia Tropical Biomed.* 2018;35:100–10.
45. Ferro C, López M, Fuya P, Lugo L, Cordovez JM, González C. Spatial distribution of sand fly vectors and eco-epidemiology of cutaneous leishmaniasis transmission in Colombia. *PLoS ONE.* 2015;10:e0139391.
46. Dellicour S, Flot JF. The hitchhiker's guide to single-locus species delimitation. *Mol Ecol Resour.* 2018;18:1234–46.
47. Hebert PD, Stoeckle MY, Zemlak TS, Francis CM, Godfray C. Identification of birds through DNA barcodes. *PLoS Biol.* 2004;2:e312.
48. Wiemers M, Fiedler K. Does the DNA barcoding gap exist?—a case study in blue butterflies (Lepidoptera: Lycaenidae). *Front Zool.* 2007;4:1–6.
49. Virgilio M, Backeljau T, Nevado B, De Meyer M. Comparative performances of DNA barcoding across insect orders. *BMC Bioinform.* 2010;11:1.
50. Godoy RE, Galati EAB. Revalidation of *Nyssomyia fraihai* (Martins, Falcão & Silva 1979) (Diptera: Psychodidae). *J Med Entomol.* 2016;53:1303–11. <https://doi.org/10.1093/jme/tjw108>.
51. Galati EAB. Morphology and Terminology of Adults and Identification of the American Taxa. In: Rangel EF, Shaw JJ, editors. *Brazilian Sand Flies*. Cham: Springer; 2018. p. 251–98.
52. Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C. Coalescent-based species delimitation in an integrative taxonomy. *Trends Ecol Evol.* 2012;27:480–8.
53. Jinbo U, Kato T, Ito M. Current progress in DNA barcoding and future implications for entomology. *Entomol Sci.* 2011;14:107–24.
54. Kjørandsen J. Current State of DNA barcoding of Sciaroidea (Diptera)—highlighting the need to build the reference library. *Insects.* 2022;13:147.
55. Florin DA, Rebollar-Téllez EA. Divergence of *Lutzomyia* (*Psathyromyia*) *shannoni* (Diptera: Psychodidae: Phlebotominae) is indicated by morphometric and molecular analyses when examined between taxa from the southeastern United States and southern Mexico. *J Med Entomol.* 2013;50:1324–9.
56. Contreras Gutierrez MA, Lopez RO, Ramos AT, Vélez ID, Gomez RV, Arrivillaga-Henríquez J, et al. DNA barcoding of *Lutzomyia longipalpis* species complex (Diptera: Psychodidae), suggests the existence of 8 candidate species. *Acta Trop.* 2021;1:105983.
57. Scarpassa VM, Cunha-Machado AS, Alencar RB. Multiple evolutionary lineages for the main vector of *Leishmania guyanensis*, *Lutzomyia umbratilis* (Diptera: Psychodidae), in the Brazilian Amazon. *Sci Rep.* 2021;11:1–7.
58. Sousa-Paula LD, Pessoa FA, Otranto D, Dantas-Torres F. Beyond taxonomy: species complexes in New World phlebotomine sand flies. *Med Vet Entomol.* 2021;35:267–83.
59. Maroli M, Feliciangeli MD, Bichaud L, Charrel RN, Gradoni L. Phlebotomine sandflies and the spreading of leishmaniasis and other diseases of public health concern. *Med Vet Entomol.* 2013;27:123–47.
60. Avise JC. *Phylogeography: the history and formation of species*. Cambridge: Harvard University Press; 2000.
61. Kaur R, Singh D. Molecular markers a valuable tool for species identification of insects: a review. *Ann Entomol.* 2020;38:1–20.
62. Maddison WP. Gene trees in species trees. *Syst Biol.* 1997;46:523–36.
63. Grace-Lema DM, Yared S, Quitadamo A, Janies DA, Wheeler WC, Balkew M, et al. A phylogeny of sand flies (Diptera: Psychodidae: Phlebotominae), using recent Ethiopian collections and a broad selection of publicly available DNA sequence data. *Syst Entomol.* 2015;40:733–44.
64. Beati L, Cáceres AG, Lee JA, Munstermann LE. Systematic relationships among *Lutzomyia* sand flies (Diptera: Psychodidae) of Peru and Colombia based on the analysis of 12S and 28S ribosomal DNA sequences. *Int J Parasitol.* 2004;34:225–34.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

