Systematic Entomology

Systematic Entomology (2015), **40**, 733–744 DOI: 10.1111/syen.12135

A phylogeny of sand flies (Diptera: Psychodidae: Phlebotominae), using recent Ethiopian collections and a broad selection of publicly available DNA sequence data

DANIELLE M. GRACE-LEMA[1], SOLOMON YARED[2,‡], ANDREW QUITADAMO[3], DANIEL A. JANIES[3], WARD C. WHEELER[1], MESHESHA BALKEW[2], ASRAT HAILU[4], ALON WARBURG[5] and RONALD M. CLOUSE[1,3]

[1]The American Museum of Natural History, New York, NY, U.S.A., [2]Aklilu Lemma Institute of Pathobiology, Addis Ababa University, Addis Ababa, Ethiopia, [3]Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, NC, U.S.A., [4]Department of Microbiology, Immunology and Parasitology, Faculty of Medicine, Addis Ababa University, Addis Ababa, Ethiopia and [5]Department of Microbiology and Molecular Genetics, The Institute of Medical Research Israel-Canada, The Kuvin Center for the Study of Infectious and Tropical Diseases, Faculty of Medicine, The Hebrew University, Hadassah Medical School, Jerusalem, Israel

Abstract. Sand flies in the psychodid subfamily Phlebotominae carry important human pathogens in the trypanosomatid protozoan genus *Leishmania* (Cupolillo). Despite the fact that hundreds of sequences for this group are now publicly available, they constitute different sets of taxa and genetic markers. Integrating these data to construct a molecular phylogeny of the family is a significant bioinformatics challenge. We used sequences of eight markers obtained from freshly collected sand flies from Ethiopia and combined them with over 1300 publicly available sequences, performing a combined analysis after generating single terminal sequences from ancestral reconstructions for some individual markers. The resulting phylogeny had 113 terminals and recovered Phlebotominae and certain species groups as monopheletic. Although the 20 outgroups in Psychodinae were recovered as a well-resolved clade with bootstrap support for many internal clades, Phlebotominae was recovered as several lineages with unclear relationships among them. However, phlebotomines clustered by geographic region, the most notable being all the New World species except *Brumptomyia* (Galati), which were recovered as monophyletic. Our phylogeny suggests a Sub-Saharan African or South Asian origin for the subfamily, which subsequently expanded to the north and west, and eventually to the New World. Supported species groups are often composed of widespread species with overlapping ranges. This result highlights the need for a large increase in the amount and diversity of molecular sequence data, and a broad selection of terminals, to test taxonomic hypotheses and examine speciation processes in this important group of flies.

Correspondence: Danielle M. Grace-Lema, Department of Biology, The City College of New York, Marshak Science Building, Room 526, 160 Convent Avenue, New York, NY 10031, U.S.A. E-mail: dgrace000@citymail.cuny.edu

‡Present address: Department of Biology, College of Natural Science, Jigjiga University, Jigjiga, Ethiopia.

Introduction

Haematophagous insects of the family Psychodidae are mostly found in the aptly named subfamily Phlebotominae. This subfamily contains several hundred species, and many of them transmit protozoan pathogens in the genus *Leishmania* (Cupolillo) to humans and other mammals. Victims of the resulting infectious disease, leishmaniasis, can suffer disfigurement or death. Leishmaniasis is estimated to be responsible for more than 20 000 deaths/year worldwide (Alvar *et al.*, 2012). The deadliest form of the disease, visceral leishmaniasis, occurs mainly in the Asian subcontinent, East Africa and Brazil (Alvar *et al.*, 2012). Cutaneous leishmaniasis occurs in Afghanistan, Algeria, Bolivia, Brazil, Colombia, Ethiopia, Iran, Peru, Saudi Arabia and Syria (WHO, 2014).

The subfamily Phlebotiminae has had a simple internal classification for much of its history, with only six genera – *Warileya* (Galati), *Brumptomyia* (Galati) and *Lutzomyia* (Galati) in the New World (North, Central and South America), and *Sergentomyia* (Galati), *Phlebotomus* (Galati) and *Chinius* (Galati) in the Old World (Europe, Africa, and Asia) (Lewis, 1982; Lane, 1993). However, recent advances in phlebotomine taxonomy (Galati, 1995, 2010, 2014) have resulted in many more genera, especially among the New World species, and their recombination into new subtribes. The subtribe Sergentomyiina now includes those species formerly in *Sergentomyia* but also some former members of *Lutzomyia*, thus giving it a global range.

A phylogenetic hypothesis of the relationships of phlebotomine flies is important for several reasons: (i) to test the current taxonomy and understand which morphological characters are convergent, derived, or ancestral; (ii) to reconstruct the geographic origin and routes of dispersal for its different lineages; (iii) to track the evolution of host–vector–pathogen relationships that underlie disease transmission; (iv) to identify possible cases of ongoing or historical genetic hybridization; and (v) to estimate the timing of major historical events in the group. However, despite the importance of understanding Phlebotominae and the diseases they carry, the subfamily has not been the subject of comprehensive phylogenetic analyses, especially using molecular data.

Phlebotominae systematics has drawn the attention of such luminaries as Graham Fairchild (Fairchild, 1955) and Willi Hennig (Hennig, 1972), and in recent years they have been the subject of some molecular studies (Depaquit *et al.*, 1998, 2000a,2000b, 2004; Esseghir & Ready, 2000; Aransay *et al.*, 2000b; Beati *et al.*, 2004; reviewed by Depaquit, 2014). Morphological data continue to be used and are the basis of the most inclusive phylogenies, and such analyses still tend to use terminal sets that are restricted to the Old or New World (Galati, 1995, 2010, 2014; Rispail & Léger, 1998a,1998b; Ilango, 2004; Pinto *et al.*, 2010). Thus relationships within the family as a whole remain in question.

Phlebotomine DNA sequence data are increasingly common, although they are usually restricted to particular species or species groups (Di Muccio *et al.*, 2000; Soto *et al.*, 2001; Depaquit *et al.*, 2002, 2008; Aransay *et al.*, 2003; Mazzoni *et al.*, 2008; Franco *et al.*, 2010; Scarpassa & Alencar, 2013) or regions (Di Muccio *et al.*, 2000; Aransay *et al.*, 2003; Lins *et al.*, 2008; Franco *et al.*, 2010; Gutiérrez *et al.*, 2014) and are often focused on population genetics (Soto *et al.*, 2001; Depaquit *et al.*, 2008) or the identification of informative markers (Peixoto *et al.*, 2001; Lins *et al.*, 2008). Most importantly, it is rare for individual specimens to be sequenced for multiple markers. A notable exception is an analysis using ten loci by Mazzoni *et al.* (2008); although narrow in taxonomic sampling, comparing only two very closely related species, it nevertheless reaffirmed the usefulness of incorporating multiple loci for inferring relationships. Thus the universe of available molecular data for the subfamily consists of isolated data sets that resist combination and thus a comprehensive analysis of its internal relationships. A thorough survey of all molecular data studies of phlebotomines has recently been done by Depaquit (2014), and confirms the initial observations that motivated our study here: although becoming voluminous, phlebotomine sequence data have come mostly from studies of populations and small species groups, and among them, mostly those of medical importance. In short, the sequence data have not been collected with the aim of resolving long-standing questions in morphology-based taxonomy.

Furthermore, due to disease impact, the majority of molecular studies are aimed at rapid, high-throughput polymerase chain reaction identification techniques, and microsatellite (MLMT) or enzyme analysis (e.g. enzyme-linked immunosorbent assay, multilocus enzyme electrophoresis, or restriction fragment length polymorphism) for clinical implications, epidemiology, exploring vector–host relationships, and pathogen control of the *Leishmania* protozoa (Aransay *et al.*, 1999, 2000a; Jorquera *et al.*, 2005; Martin-Sanchez *et al.*, 2006; Kato *et al.*, 2007; Rassi *et al.*, 2008; Gebre-Michael *et al.*, 2010; Hamarsheh, 2011; Alam *et al.*, 2012; Berdjane-Brouk *et al.*, 2012). Many of these studies continue to use morphology as the basis for sand fly identification. Morphological identifications require extensive labour and expertise and can be prone to mis-identifications, which can confound epidemiological studies. DNA sequencing also requires labour and expertise, but it can also provide large numbers of characters for analysis. As different flies carry various parasites leading to different diseases, studies involving well-identified phlebotomine species are crucial for advances in the clinical constraint of the disease. A comprehensive Phlebotominae phylogeny is needed to validate morphology-based species delineations and allow exploration of inter- and intraspecific vector relationships. Finally, a Phlebotominae phylogeny will provide a knowledge base for deciphering sand fly biology, ecology, and the spread of diseases.

In exploring Phlebotominae phylogenetics, Ethiopia is of special interest to us for the relative abundance of host–vector–pathogen interactions seen in north-western Ethiopia, and because its location is central to the ranges of several Old World groups. It highlights the need we have for taxonomic clarity and information about diversification dynamics on this group. *Phlebotomus (Lar.) orientalis* is responsible for the transmission of *Leishmania donovani* parasites

causing visceral leishmaniasis in eastern Sudan (Elnaiem *et al.*, 1998). This species is the predominant sand fly and the most likely visceral leishmaniasis vector in north-west Ethiopia (Gebre-Michael *et al.*, 2010). *Phlebotomus (Lar.) longipes* and *Phlebotomus (Lar.) pedifer* have been commonly identified as proven vectors of *Leishmania aethiopica*, the principal cause of cutaneous leishmaniasis in Ethiopia, found throughout different parts of the country and widespread in the highlands (Bsrat *et al.*, 2015). *Phlebotomus (Phl.) papatasi* is a known vector of *Leishmania major*, causing cutaneous leishmaniasis in northern Africa, including parts of Sudan (Ibrahim, 1990; El Sayed *et al.*, 1991), but to date has not been found as a vector of any *Leishmania* parasite in Ethiopia. So far the only Ethiopian collections of *P. (Phl.) papatasi* have been in the north-west, where *L. major* has not yet been observed (Gebre-Michael *et al.*, 2010). Recently, though, *L. major* infections were identified in humans by real-time PCR in a cohort study in Tahtay Adiabo, northern Ethiopia (Abbasi *et al.*, 2013), although *P. (Phl.) h. papatasi* has rarely been found in this region. *Phlebotomus (Phl.) bergeroti* is another species found in north-west Ethiopia; its medical importance in the country is unknown. On the other hand, *Phlebotomus (Phl.) duboscqi* is a known vector of *L. major* in southern Ethiopia (Gebre-Michael *et al.*, 1993). Additionally, *L. major* and *Leishmania tropica* have rarely been implicated in cutaneous leishmaniasis in areas of Ethiopia where *Phlebotomus (Par.) sergenti* and *Phlebotomus (Pab.) saevus* have been found infected with the parasite (Hailu & Formmel, 1993; Hailu *et al.*, 2006). The medical importance of *Phlebotomus lesleyae* is still unknown, and its taxonomic position remains unresolved. Lewis & Kirk (1946) and Ashford (1974) described and named the group in the *Phlebotomus* subgenus *Phlebotomus (Phl.) lesleyae*, although others include it under the subgenus *Parvidens* (Theodor & Mesghali, 1964) or in the genus *Sergentomyia* (Davidson, 1982). In this report we try to resolve some of these taxonomic ambiguities pertaining to *P. lesleyae*, as well as provide a phylogenetic backbone for investigating vector–host relationships.

Using fresh specimens collected in Ethiopia, we attempt to tie together the various and discordant publicly available data by sequencing several specimens for several markers that overlap with the public data. In addition to this method, we develop a novel way to summarize the public data in such a way as to create single terminals with several markers for individual species. Using this combined data set, we ask whether there is molecular evidence for the genera and subgenera of Phlebotominae, and whether any particular biogeographic history can be inferred. We do not test the monophyly of the subfamily, as it requires sequence data from a much larger selection of psychodids than are currently available. We aim to produce a road map for issues in phlebotomine systematics, identifying those relationships most in need of re-examination and possible revision, and forwarding a testable hypothesis when more comprehensive molecular data for this group are available. In addition, we test the ability of using ancestral sequence reconstructions as a useful method for summarizing molecular data for species when numerous individuals are sequenced.

Materials and methods

Specimen acquisition

Phlebotomine sand flies were collected in Kafta Humera district, the main focus of visceral leishmaniasis in north-west Ethiopia. The study area is located at altitude ranges of 500–600 m above sea level (*c.* 14°04.23′N, 36°33.62′E). Collections of sand flies were made using standard CDC miniature light traps (John W. Hock Co., Gainesville, FL, U.S.A.) during 2011 and 2012. Sand flies were caught in both farm fields and villages. The trapped sand flies were anaesthetized with chloroform and sorted under a dissecting microscope by preliminary morphological identification into *Phlebotomus* and *Sergentomyia*. The head and abdominal tips of each sand fly were removed and mounted on slides on a drop of Hoyer mountant for species identification; the remaining body parts were preserved in absolute alcohol for phylogenetic analysis. Sand flies were identified to species level based upon cibarial and pharyngeal armature, as well as spermatheca of females and external genitalia of males using morphological keys (Quate, 1964; Lewis, 1982). Specimens of *P. (Lar.) orientalis*, *P. (Phl.) bergeroti*, *P. (Phl.) duboscqi*, *P. (Phl.) papatasi*, *P. (Phl.) lesleyae*, and *Sergentomyia* spp. were used for phylogenetic analyses.

DNA extraction, amplification, and sequencing

Total DNA was extracted from sand fly abdomens using DNEasy® tissue kit (Qiagen, Valencia, CA, U.S.A.). Using primers and protocols known to work on sand flies and other insects (Table S1), we then amplified three mitochondrial and six nuclear loci. Due to the nature and disparity of publicly available sequence data, and limited published molecular marker data for this group, we decided to include in our dataset markers not commonly used for Phlebotominae, but used in analyses of other insects, for instance, long-wavelength rhodopsin (Ward & Downie, 2005). Some markers were those developed for use in Phlebotominae, homologous to common markers used for studies of *Drosophila* (Lins *et al.*, 2008; Mazzoni *et al.*, 2008). Temperature profiles followed those accompanying published primers (Table S1). For the nuclear gene *zetacop*, we only obtained two sequences, so it was not used in phylogenetic analysis. All sequences more than 200 bp are deposited in GenBank under accession numbers KR020546–KR020683. Shorter sequences and all data used are available in the Supporting Information.

We downloaded from GenBank 1363 phlebotomine and outgroup sequences for each of the markers we sequenced (if available). For population studies, in which hundreds of individuals from the same location were sequenced for the same marker, we took a random sample of the sequences. We did not use the sequences of the marker 18S rDNA obtained by Aransay *et al.* (2000b), as they represented only 14 species and, being clones, had sequence variation within individuals (as part of the nuclear ribosomal array, 18S is represented in the genome by a large array of concerted but not identical paralogues) (Hillis &

Dixon, 1991); we did not sequence 18S from our specimens. The original identifications of specimens on GenBank were updated using the taxonomy of Galati (2003), as updated by Galati (2014), and we use that taxonomy throughout. In figures and supplemental material, we use the generic and subgeneric abbreviations proposed by Marcondes (2007). The recently synonymized species *Phlebotomus (Par.) riouxi* and *Phlebotomus (Par.) chabaudi* (Tabbabi *et al.,* 2014) were treated as different species here as a test of that synonymization.

All the sequences, our own and those from GenBank, were subjected to multiple sequence alignment in MAFFT (Katoh *et al.*, 2002) using the command 'adjustdirectionaccurately', which allows sequences to be considered in reverse complement, as some GenBank entries were as such. For all sequences except ITS2, the gap-opening cost was set at 1.0, and the gap-extension cost set at 0.05. For ITS2, these costs were 3.0 and 0, respectively, due to its large length variation. Besides ITS2, only *cacophony* and *paralytic* had insertion-deletion (indel) events, but all markers were aligned to check for likely contaminant sequences and the location of missing data due to laboratory artifacts, recorded as '?' in sequences. All terminals and sequences obtained are detailed in Tables S3 and S4.

Ancestral sequence reconstructions

The primary challenge in performing a phylogenetic reconstruction of Phlebotominae was the creation of terminals with multiple markers representing a wide variety of taxa. For the specimens we sequenced, more than one marker was sequenced per specimen, thus allowing them to connect all the terminals in the analysis, despite the large amount of missing data. However, we had no representatives of certain genera and wanted to include them by using the disparate sets of publicly available sequence data. Phlebotomine molecular studies have traditionally focused on single markers, and when the same species have been sequenced in different studies, different specimens were used for each marker.

One option was to randomly choose a representative sequence for each marker and species. This ran the risk of generating chimeric terminals that could contain highly derived or rare versions of particular markers, thus misinforming the analysis. Another option was to make consensus sequences of those from the same species, but this would assume the correct identification and monophyly of those specimens. A third option, which we chose, was to use a reconstruction of the ancestral sequence from the base of the clade containing most specimens from that species in single-marker phylogenies. Limitations of this method are that the sequences used are ancestral reconstructions and not direct observations, and reconstructions are dependent on the algorithms used by the phylogenetic program and phylogenetic context of any particular tree search (terminal selection, outgroups, cost schemes). However, we considered this option better informed than selecting or collapsing sequences from supposed conspecifics outside of any phylogenetic analysis.

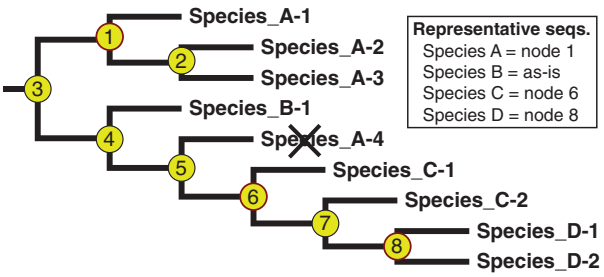

Fig. 1. Hypothetical relationship among nine specimens identified to four species, illustrating the rules by which ancestral sequences were used to represent species in the final analysis.

This was done in the phylogenetic program POY v. 5.0.0 beta (Varón *et al.*, 2009), using the commands 'graphdiagnosis' and 'diagnosis.' POY is a program that performs progressive sequence alignment and tree-search simultaneously under cost schemes chosen by the user (Wheeler, 1996). POY's efficacy versus other mainstream parsimony methods has been tested (Kjer *et al*., 2007; Ogden & Rosenberg, 2007; Lehtonen, 2007; Liu *et al*., 2009) and reviewed (Blair & Murphy, 2010). It was shown that direct optimization methods implemented through POY can be very accurate in building alignments and phylogenetic hypotheses, as compared with traditional methods. The 'graphdiagnosis' command returns a plot of the best tree with all nodes numbered. For each node, the 'diagnosis' command returns three ancestral reconstructed sequences: (i) *Preliminary*, which results from simply the down-pass (determining the most parsimonious reconstruction working from the terminals down to the root); (ii) *Final*, which results from both the down-pass and a subsequent up-pass; and (iii) *Single*, which is the selection of a single possible sequence for that node, essentially one of the many possible *Final* sequences suggested by ambiguities at various positions. For example, for one node in the tree made using *paralytic*, a segment of the sequence was reconstructed in *Preliminary* form as KA8AAAMVKCCC, *Final* as TAAAAA--STCC, and *Single* as TAAAAA--CTCC. Ambiguity codes with numeric symbols indicate the possibility of a gap as well as various combinations of nucleotides at that position. The *Single* reconstructions were the ones used in the phylogenetic analyses here.

We did tree searches for each marker using ten different transformation cost schemes, from all transformations being equal to gaps, to 1 in which gaps cost 16× that of transitions. We reconstructed sequences at each node of each resulting phylogeny, and we chose the *Single* reconstructions for the cost scheme that returned *Final* reconstructions with the fewest ambiguities (Fig. 1). This was found to be when gaps cost 4, transversions cost 4, and transitions cost 1 ('141'; see Giribet, 2003 – the first number in cost scheme monikers is the ratio of gap costs to transversion costs). We then did final tree searches and ancestral sequence reconstructions under this cost scheme for each marker. All tree searches conducted for acquiring ancestral sequences were done with unaligned data, allowing POY to find the optimal sequence alignment and tree simultaneously. After the tree searches were done to obtain

Single ancestral sequences for terminals from GenBank, our Ethiopian samples were added to generate the final data set.

When the ancestral reconstructions were obtained, they were used to replace multiple sequences from different specimens for each marker. In some cases, species were not recovered as monophyletic, so it was assumed that small numbers of individuals found outside the largest clade for that species were misidentifications (Fig. 2). Although all specimens from GenBank were used, unidentified specimens that did not place within a large clade of identified ones were excluded from the final alignment. In cases where one species was recovered as rendering another paraphyletic, the ancestral sequence for the latter was simply the sequence at the ancestor of both species. Species that had only one representative for a single marker were taken as-is to represent the species for that marker.

Phylogenetic analysis

Initial analyses were rooted with several terminals from the families Scatopsidae, Tipulidae and Culicidae found in GenBank to have some of the markers that overlap with our data, but these rendered Psychodidae polyphyletic. We considered this outcome to be the result of a lack of conserved genes needed to resolve familial relationships, a query beyond the scope of this study. Thus, for outgroups we used 20 nonphlebotomine psychodids, all in the subfamily Psychodinae, each of which had cytochrome *c* oxidase I (COI) sequence, and some of which also had Cytochrome B (CytB) sequence. We rooted tree searches by one outgroup terminal and did not constrain the monophyly of either subfamily. Our final combined phylogeny included all those terminals that had either COI or NADH dehydrogenase subunit 4 (ND4) (101 and 44 terminals, respectively, in the final alignment), as these were the markers most available for the largest number and widest diversity of our specimens. The mitochondrial marker CytB appeared to be available for a large number of terminals as well (68), but actually most of the New World specimens (including all of the *Brumptomyia*, 13 terminals in total) had only the final ∼250 base pairs sequenced and did not overlap by more than a few bases with the fragment we sequenced from our Ethiopian specimens. We treated CytB in our final analyses as two different partitions.

Tree searches were conducted under maximum likelihood (ML) with a general time-reversible model plus substitution rate heterogeneity (GTR + gamma) in the program RAXML (Stamatakis *et al.*, 2008) on the CIPRES (Miller *et al.*, 2010) computing cluster for the terminal set consisting of all markers with either COI or ND4. RAXML defaults to this well-known model on the computing cluster and has only one other available ('GTR-CAT'), which appears to have a main advantage, which is to lower computing costs. With the same settings, we also conducted tree searches in RAXML for each marker independently.

The same combined terminal set (having either COI or ND4) was also used to conduct tree searches under parsimony in POY using the data as either unaligned or aligned, and the cost transformation scheme as either '141' (the scheme under which ancestral sequences were obtained) or '111' (all transformations equal). Searching was done on 25 processors of the local computing cluster using timed searches of 12 h for the unaligned data and 1 h for aligned data. Resulting trees were compared to the tree recovered under ML using the program CLADESCAN 1.0 (Sanders, 2009).

All sequence alignments, which include raw data from our specimens and ancestral reconstructions from groups on conspecifics from GenBank, are available in the Supporting Information.

Results

The terminal set of those species and specimens that had either COI or ND4 consisted of 113 members, and the most likely phylogeny recovered under ML (the 'final combined phylogeny') had a 1nL of −35 323.538996. All the phlebotomine terminals in the final combined phylogeny were in the tribe Phlebotomini, although a member of Hertigiini (*Chinius* sp.) was included in the single-gene phylogeny for CytB. In the final combined phylogeny, the only taxonomic groups with more than one terminal which were recovered as monophyletic were as follows: Psychodinae and Phlebotiminae; the subtribe Brumptomyiina and the genus *Brumptomyia*; the genera *Psychodopygus*, *Evandromyia* and *Pressatia* (two terminals each); and the *Phlebotomus* subgenera *Euphlebotomus*, *Phlebotomus*, *Synphlebotomus* and *Transphlebotomus*. The clade [*Phlebotomus* (*Anaphlebotomus*) *rodhaini* + *Phlebotomus* (*Euphlebotomus*)] was recovered with high bootstrap support as monophyletic and with moderate support as sister to the remaining phlebotomines. The Malagasy species *Phlebotomus berentiensis*, which was recently recombined in the subgenus *Anaphlebotomus* (Depaquit *et al.,* 2004), was not recovered with *P. (Ana.) rodhaini*, but these two terminals also did not have overlapping markers in the final alignment. The species *P. (Par.) riouxi* and *P. (Par.) chabaudi* were recovered as sister species with 100% likelihood bootstrap support and under parsimony using all parameters, supporting their synonomization (Tabbabi *et al*., 2014); although the branch lengths between them were quite long, this could have been caused by the effects of missing data on branch length calculations.

A clade containing specimens of what was originally described by Lewis & Kirk (1946) as *P. lesleyae*, but which has been considered more recently by some as belonging to *Sergentomyia* (Duckhouse & Lewis, 1980; Davidson, 1982), was recovered outside of *Sergentomyia* and sister to a clade consisting of the *Phlebotomus* subgenera *Paraphlebotomus*, *Synphlebotomus* and *Phlebotomus*. This is the first molecular analysis of this enigmatic taxon, which has morphological characteristics similar to both *Phlebotomus* and *Sergentomyia* (Lewis & Kirk, 1946; Ashford, 1974; Davidson, 1982). Although our final combined phylogeny suggests the species belongs to *Phlebotomus*, we note that analyses of separate genes of these specimens reveals conflict at a molecular level as well: ITS2, ND4, and COI phylogenies place our *P. lesleyae* specimens well away from GenBank specimens identified as *Sergentomyia*, but our CytB analysis indicates a very close

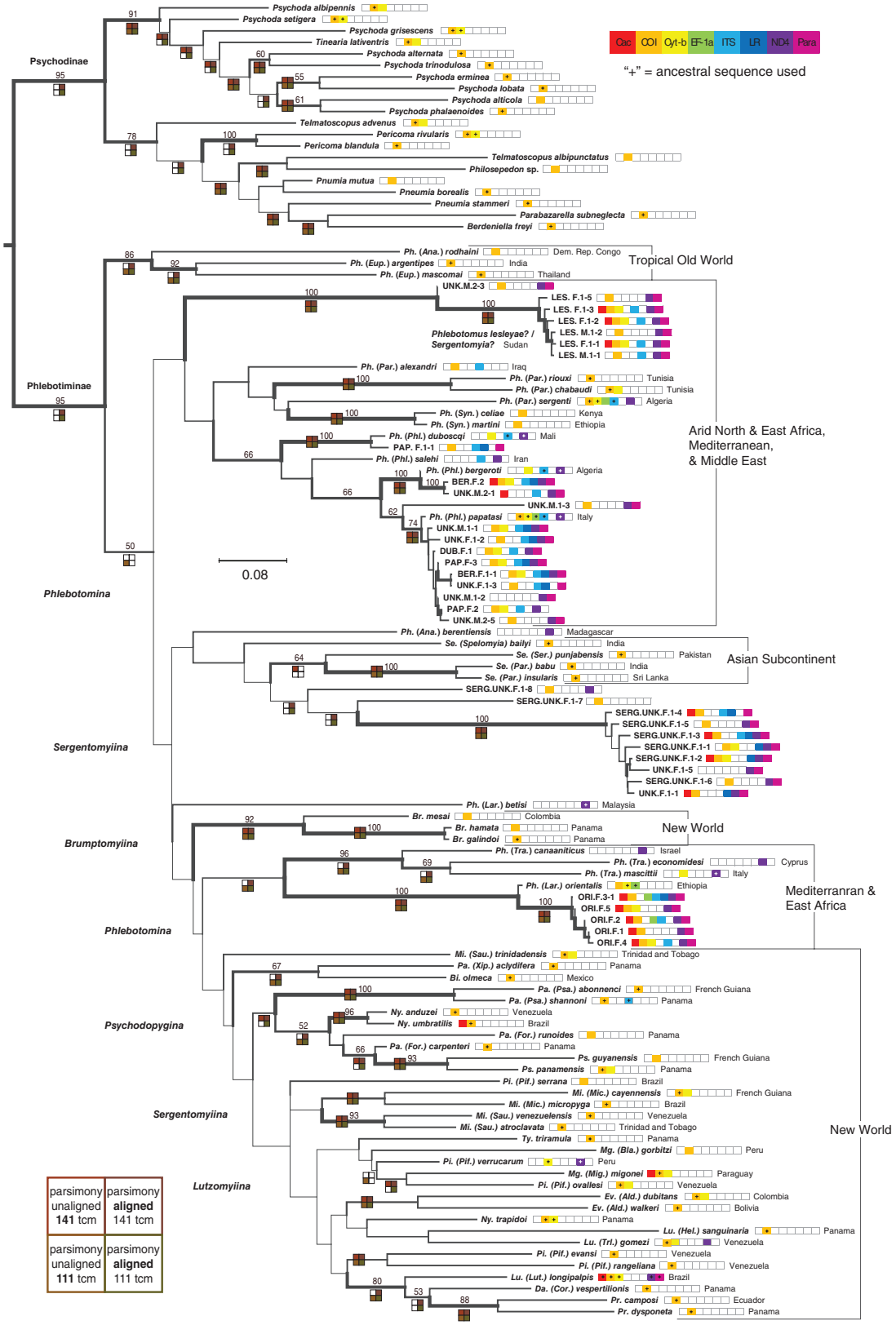

Fig. 2. Legend on Next page.

relationship between them and certain Old World *Sergentomyia* species (and as well *Phlebotomus (Lar.) perfiliewi*).

Specimens of *P. (Lar.) orientalis* collected in north-west Ethiopia were recovered in a well-supported clade with exemplars from GenBank and indicate strong genetic support for their morphological identification. The molecular identification of this species is useful for its medical importance and implications in visceral leishmaniasis disease epidemiology. It was not found to be sister to *Phlebotomus (Lar.) betisi*, and COI sequences for *P. (Lar.) orientalis* placed it as highly derived and away from other *Phlebotomus* species, in a supported clade with members of Lutzomyiina, Sergentomyiina and Psychodopygina.

Our specimens of *P. (Phl.) bergeroti*, *P. (Phl.) papatasi*, and *P. (Phl.) duboscqi*, did not group clearly with the sequences of these species downloaded from GenBank. The GenBank specimens formed a clade with *Phlebotomus (Phl.) salehi*, but intermingled with them were our identified and unidentified specimens. In addition, one of our male specimens, 'UNK.M.1-3,' placed among *Phlebotomus* specimens when using ND4 (Figure S2) but among *Sergentomyia* specimens with COI (Figure S1).

Sequences from New World specimens that had previously been identified as *Lutzomyia* (i.e. all New World specimens except the Brumptomyiina) were recovered as monophyletic. These species are today split into several genera and subtribes, which we recovered in close association. With three exceptions, *Micropygomyia (Sauromyia) trinidadensis*, *Pintomyia (Pifanomyia)* and *Nyssomyia trapidoi*, the subtribes of this clade were arranged as follows: (Psychodopygina (Psychodopygina (New World Sergentomyiina + Lutzomyiina))).

Relationships deeper than the levels of small species groups and above the level of the subfamily and its earliest diverging lineage had nearly no resampling support. That is, despite the recovery of certain genera as monophyletic in the optimal tree, such relationships had little redundant signal in our data set. In contrast, relationships among the psychodinine outgroups were well resolved and usually well supported and stable.

Many of the relationships that were recovered under ML were also recovered under parsimony, using different alignment and transformation cost schemes. In the program POY, we analysed under parsimony the same terminal set and alignment (both as aligned and unaligned) used under likelihood. However, the only major taxonomic groups recovered as monophyletic under parsimony were Phlebotominae and Brumptomyiina, as well as the *Phlebotomus* subgenera *Euphlebotomus* and *Transphlebotomus*. One shortest tree was recovered using parsimony with each of the cost and alignment schema. Under the 141 cost scheme, the tree recovered using unaligned data was 19 193 steps, and that recovered using aligned data was 21 240 steps. Using equal costs (111), the tree recovered using unaligned data was 7256 steps, and that recovered using aligned data was 7786 steps.

Discussion

Biogeographic signal

The most striking aspect of our final combined phylogeny is that species cluster in greater agreement with biogeographic region than with higher-level taxonomy. The clearest example of this is with the subtribe Sergentomyiia, the New World members of which (*Micropygomyia*) were never recovered in close association with Old World Sergentomyiina in our final combined phylogeny or the separate gene trees, and in fact formed a clade with the New World subtribes Psychodopygina and Lutzomyiina. This clade and the Brumptomyiina seem to be independent colonizations of the New World. In general, our phylogeny implies a Tropical Old World origin for the phlebotomines and the subsequent diversification north and into Asia, and later into the New World.

Despite some species with large ranges, phlebotomine flies do appear to have dispersal constraints that cause several closely related species to have limited, overlapping distributions (Table S2). It is striking that some species with extremely large ranges can be both closely related and overlap, such as *Lutzomyia (Lut.) longipalpis*, *Dampfomyia (Coromyia) vespertilionis*, *Pressatia dysponeta*, and *Pressatia camposi* in Mesoamerica and large portions of South America. These forms are either species that have arisen sympatrically through ecological isolation, morphological variants within single species, or once-separate species experiencing extensive hybridization (a phenomenon not unknown in phlebotomines Mazzoni *et al*., 2008). In fact, the latter mechanism could be responsible for the considerable geographic signal we recovered in our tree, for if unrelated lineages in a region hybridize, they could appear to be more closely related over time in combined analyses of different molecular markers.

Phylogenetics and systematics, history and implications

Our phylogeny is only a preliminary assessment of where the needs are in phlebotominae systematics, and its large amount of missing data had important effects on the resampling support

Fig. 2. Most likely phylogeny of the terminal set consisting of all those with either cytochrome *c* oxidase I (COI) or ND4, with type localities to the right of each terminal. Terminals are coloured by subtribe, which are indicated to the left of the tree. The markers available for each terminal are shown in the multicoloured bar to the right of each terminal, the key for which is in the upper right. A plus sign (+) in any marker box indicates the use of an ancestral reconstruction from many specimens for that particular marker. Branch thickness indicates bootstrap support (thick branches receiving higher support). Grids below support values indicate recovery of the same clades under parsimony and various data treatments, as summarized in the lower left. Coloured blocks denote recovery under those conditions, and grids are not shown for clades of closely related specimens separated by short branches (e.g. among Ethiopian specimens of *P. (Lar.) orientalis*) Data treatments under parsimony varied between being unaligned (where tree-searching and the multiple sequence alignment are done simultaneously) and aligned (the same alignment as was used under maximum likelihood), and between a transformation cost matrix that penalized gaps and transversions four steps (141) and one in which gaps, transversions and transitions all costs the same (111).

and probably even the topology and branch lengths (Lemmon *et al.,* 2009; Roure *et al*., 2013). However, the amount of missing data was not a good predictor of ambiguous or controversial placement in our phylogeny, and our single-gene mitochondrial phylogenies showed little or no support for mid-level relationships, indicating a special inadequacy of mitochondrial markers in Phlebotominae. Among our Psychodinae outgroups, COI and CytB alone seem to be sufficient for resolution, stability and support. [However, we note that Espíndola *et al*. (2012) did not recover expected tribal relationships in Psychodinae using only mitochondrial markers from 88 specimens, so the utility of the mitochondrial DNA markers could be highly dependent on taxon sampling.] Curler & Moulton (2012) had high bootstrap support within Phlebotominae in a phylogeny of psychodid subfamilies using the nuclear markers 18S and *Peregrin*, and we suggest these data set be expanded across phlebotominae genera; in addition, this may resolve the problem of using outgroups from other families, which we found to be impractical with our data set.

The enormity of Phlebotominae, which has over 800 described species (Young & Duncan, 1994; Marcondes, 2007; Bisby *et al.*, 2010, Curler & Moulton, 2012; Galati 2014), makes elucidating its phylogeny a considerable undertaking. It has been suggested that the recent increase in genera in Phlebotominae is impractical to medical parasitologists, and that epidemiological issues should inform their taxonomy (Ready, 2011). This would make sand flies exceptional in the task of understanding the history of life and capturing it in the Linnaean taxonomic system, and we would expect such a large group to have numerous cases of morphological convergence, host–parasite shifts, range expansions and contractions, and other deviations from a neat historical story; thus the importance of molecular data for providing a large number of characters that can be selected, so as to minimize the influences of convergence and hybridization.

Previously, systematic classification has been largely based on morphology of the New World (Young & Duncan, 1994; Galati, 1995) and Old World species (Rispail & Léger, 1998a,1998b), and molecular analyses of Phlebotominae as a whole are generally lacking. Although several comparative molecular analyses exist, many are restricted to specific subgenera (Di Muccio *et al.*, 2000; Esseghir & Ready, 2000; Pinto *et al.*, 2010; Cohnstaedt *et al.*, 2011). Nonetheless, our phylogeny here does allow for limited comparisons against previous hypotheses, some of which receive additional support here.

Our combined phylogeny is consistent with certain subgeneric relationships supported by female spermathecal and male genitalic characters (Ilango, 2004): (*Anaphlebotomus* + *Euphlebotomus*), (*Transphlebotomus* + *Larroussius*) and (*Phlebotomus* + *Paraphlebotomus* + *Synphlebotomus*). The latter relationship is further supported by the fact that they also are transmitters of *L. major* and *L. tropica*, found across the drier areas of North Africa, Arabia, and Central Asia, as well as *L. donovani* in East Africa (Ilango, 2004).

One of the earliest molecular phylogenies of the phlebotomines, (Depaquit *et al.*, 1998), a parsimony analysis of 455 bp of the nuclear ribosomal marker 28S (the D2 domain) from nine species in three genera, inferred monophyly of the genus *Lutzomyia*. *Sergentomyia*, which did not cluster with *Phlebotomus*, was recovered as sister to *Lutzomyia*. *Phlebotomus* was recovered as paraphyletic, with closely related clades of the subgenera *Phlebotomus* and *Paraphlebotomus* basal to *Larroussius*. This result was reaffirmed (Depaquit *et al.*, 2000a) with a similar sequence analysis – 28S rDNA from the subgenus *Paraphlebotomus* – and in this latter analysis the nuclear ribosomal region ITS2 was tested as a potential useful marker in conjunction with the 28S rDNA.

Depaquit *et al.* (1998, 2000a) recovered the subgenera *Phlebotomus* and *Paraphlebotomus* as sister taxa, and we found them, with *Synphlebotomus* to form a clade; and they recovered most *Phlebotomus* subgenera as para- or polyphyletic, as we have here. They inferred *Lutzomyia* and *Sergentomyia* as sister genera, which was not found here, but they also found that *Lutzomyia* received little bootstrap support, as here. Both our results and their results showed *Lutzomyia* (shown on our tree by its current taxonomy, split into three subtribes and several genera) as monophyletic, and apomorphic relative to the other genera. They had low bootstrap support for *Sergentomyia* and *Phlebotomus (Larroussius)*, and we also had low support for the Old World *Sergentomyia* and did not recover *P. (Lar.) orientalis* as sister to *P. (Lar.) betisi*. *Phlebotomus berentiensis*, which had been originally described in *Sergentomyia* (Léger & Rodhain, 1978), was later placed within the subgenus *Anaphlebotomus* by Depaquit *et al*. (2004), but this was not supported by our analysis.

Di Muccio *et al.* (2000) have investigated the interspecies relationships within the *Phlebotomus* subgenus *Larroussius*. Conducting tree-searches under parsimony and using neighbor-joining tree-building techniques to analyse sequences of ITS2 rRNA, they recovered monophyly of the subgenus. This is also supported by morphological characters and geographical distributions. Esseghir & Ready (2000) recovered very similar trees of *Larroussius*, inferred from the mitochondrial marker CytB and the nuclear marker EF1-alpha. This contrasts with our analysis, which did not recover a sister relationship between the two *Larroussius* species included.

For the close to 400 species formerly in the genus *Lutzomyia*, molecular analyses have generally been restricted to certain subgroups. Cohnstaedt *et al.* (2011) performed a phylogenetic analysis on the *Lutzomyia verrucarum* [now *Pintomyia (Pifanomyia) verrucarum*] species group using COI, and they recovered a phylogeny that matched the morphological taxonomy of Galati (1995) and Young & Duncan (1994). Their analysis also suggested that COI alone, while adequate and informative in their analysis at the species level among recently evolved lineages, is not a useful marker for phylogenetic reconstructions at the interspecific level in sand flies, and they suggest using one or more nuclear gene fragments as well. Using the mitochondrial ribosomal marker 12S and the nuclear ribosomal marker 28S, Beati *et al.* (2004), conducted a phylogenetic analysis of seven *Lutzomyia* subgenera. This was the first phylogeny of New World Phlebotominae that included more than one subgroup or subgenus. They also found good support for the morphological classifications of both Galati (1995) and Young & Duncan (1994). While sampling for Beati *et al.* (2004) analysis was limited, they recovered three clades that match the subgenera *Helcocyrtomyia*

and *Lutzomyia* and the combined species groups *migonei* and *verrucarum*. Surprisingly, since that time, there have been no other analyses of *Lutzomyia* that include multiple subgenera, species groups, and markers until now.

Conclusions

The phylogeny presented here not only recovers intriguing relationships and biogeographic patterns, but it also highlights specific needs for future molecular work on Phlebotiminae. Our difficulty rooting the subfamily using outgroups from other families indicates a lack of overlapping conserved genes which can inform the resolution of deeper nodes, as well as a lack of overlapping sequence data from psychodids outside of Phlebotiminae. The lack of bootstrap support among phlebotomine lineages and the observation of conflict in the relationships recovered using different markers separately both indicate the need for a significant increase in the amount of data from each terminal. Our results here as well as previous studies suggest this need for more data is especially acute in Phlebotiminae, as many morphological interpretations are still debated, and our psychodinae outgroups resolved much more easily in our analysis. Gathering more data could take the path of a single study using next-generation sequencing across a wide range of taxa, and this would be ideal. However, outside of this scale of study, which would require considerable resources, we suggest sequencing of the nuclear ribosomal array (18S, ITS and 28S) from exemplar specimens during population and regional studies, both for future use and also to investigate hybridization. Markers such as *paralytic* and *cacophony* appear to be informative but too rarely sequenced to be of general use. With mitochondrial markers, caution should be taken to sequence the complete fragment obtainable for CytB (not one end or the other), and more than COI is needed to resolve relationships within this subfamily.

Our specimens collected in Ethiopia demonstrate the difficulties sand flies present, even at the species level, in understanding taxonomy and speciation dynamics. For example, the lack of resolution between *P. (Phl.) bergeroti*, *P. (Phl.) papatasi* and *P. (Phl.) duboscqi* is incongruent with generally accepted taxonomy, as these species in particular have been shown to be morphologically and molecularly distinct in a previous study (Khalid *et al.*, 2010) and even to show genetic differentiation at the population level (Khalid *et al.*, 2012). This may be a result of preliminary morphological mis-identification, as these three species occur sympatrically, and females are not easily distinguished morphologically. However, our samples were identified by the third and fourth antennae ascoid and pharyngeal armature for females (considered the most informative characters, Khalid *et al.*, 2010), and males were easily identified by observing the terminal genitalia. Another possibility is that this was a result of the inability of the markers we chose to resolve the relationships of these very closely related species, and indeed in the phylogenies made from individual markers (Figures S1–S4), these species are not always monophyletic. In addition, missing data and using ancestral reconstructions to represent whole species, both of which apply among these terminals, may have masked the small number of molecular synapomorphies defining these species. Another explanation may be hybridization among these species, which has the ability to confound combined mitochondrial and nuclear molecular analyses (Skála & Zrzavý, 1994; Mazzoni *et al.*, 2008).

Another example of a species that placed strangely in our phylogeny was the medically important species *P. (Lar.) orientalis*. It and other species from East Africa and the Mediterranean formed a clade that was recovered among the New World species. This was not an artifact of our use of ancestral reconstructed sequences for data from GenBank, but rather the behaviour of the COI fragment for this species. In single gene phylogenies of COI, *P. (Lar.) orientalis* placed among New World species, whether using ancestral reconstructed sequences (Figure S1) or just the raw data for all specimens (not shown).

The classification, morphological characters and internal relationships of Phlebotominae are long debated, and there is a clear need for a large analysis that overcomes several of the limitations posed by previous small analyses. Phylogenies of particular genera or species groups assume the monophyly of those groups, and phylogenies based on single markers are unable to resolve deep and shallow nodes simultaneously. Here we attempt to provide an initial groundwork for comprehensive phylogenies of Phlebotominae by combining all known molecular data in a single phylogenetic analysis. With a large amount of missing data, resampling support was not expected to be high, and indeed bootstrap values for most nodes were low, but nonetheless, certain key groups identified by morphology, such as *Brumptomyia*, the Old World Sergentomyiina, and certain subgenera, were recovered in our phylogeny, suggesting they are true phlebotomine lineages. Other groups were not recovered, but their constituent lineages instead grouped by geographic region, most notably the New World Sergentomyiina, implying morphological convergence or regional hybridization.

Supporting Information

Additional Supporting Information may be found in the online version of this article under the DOI reference: 10.1111/syen.12135

Figure S1. Phylogeny recovered under maximum likelihood using the marker COI. Thicker branches received bootstrap support greater than 50%. Terminals are coloured according to subtribe.

Figure S2. Phylogeny recovered under maximum likelihood using the marker ND4. Thicker branches received bootstrap support greater than 50%. Terminals are coloured according to subtribe.

Figure S3. Phylogeny recovered under maximum likelihood using the marker CytB. Thicker branches received bootstrap support greater than 50%. Terminals are coloured according to subtribe, and asterisks denote those that did not meet the criterion for inclusion in the final tree (COI or ND4 available).

Figure S4. Phylogeny recovered under maximum likelihood using the marker ITS2. Thicker branches received bootstrap support greater than 50%. Terminals are coloured according to subtribe, and asterisks denote those that did not meet the criterion for inclusion in the final tree (COI or ND4 available).

Table S1. Markers (abbreviated), fragment lengths and primers used in this study.

Table S2. Species groups recovered with bootstrap support in the optimal phylogeny using all terminals with COI or ND4. Author and distribution data from the BioSystematic Database of World Diptera (Evenhuis *et al.*, 2010).

Table S3. GenBank accession numbers for all specimens and markers used in this study.

Table S4. Taxonomy of all terminals, with DNA sequence data available for each. "Original" sequence data were generated during this study and used in phylogenetic analyses as-is. "Ancestral" sequence data were generated from phylogenetic analyses of single markers.

Acknowledgements

This study was supported by Bill and Melinda Gates Foundation Global Health Program (grant number OPPGH5336) and by a grant from the Defense Advanced Research Projects Agency (DARPA, W911NF-05-1-0271). We would like to thank Wondwossen A. Gebreyes of The Ohio State University for coordinating the meeting of US and Ethiopian researchers.

References

Abbasi, I., Aramin, S., Hailu, A. *et al.* (2013) Evaluation of PCR procedures for detecting and quantifying *Leishmania donovani* DNA in large numbers of dried human blood samples from a visceral leishmaniasis focus in Northern Ethiopia. *BMC Infectious Diseases*, **13**, 153.

Alam, M.S., Kato, H., Fukushige, M., Wagatsuma, Y. & Itoh, M. (2012) Application of RFLP-PCR-Based identification for sand fly surveillance in an area endemic for Kala-Azar in Mymensingh, Bangladesh. *Journal of Parasitology Research*, **2012**, 1–4.

Alvar, J., Vélez, I.D., Bern, C. *et al.* (2012) Leishmaniasis worldwide and global estimates of its incidence. *PLoS ONE*, **7**, e35671.

Aransay, A.M., Scoulica, E., Chaniotis, B. & Tselentis, Y. (1999) Typing of sandflies from Greece and Cyprus by DNA polymorphism of 18S rRNA gene. *Insect Molecular Biology*, **8**, 179–184.

Aransay, A.M., Scoulica, E. & Tselentis, Y. (2000a) Detection and identification of *Leishmania* DNA within naturally infected sand flies by seminested PCR on minicircle kinetoplastic DNA. *Applied and Environmental Microbiology*, **66**, 1933–1938.

Aransay, A.M., Scoulica, E., Tselentis, Y. & Ready, P.D. (2000b) Phylogenetic relationships of phlebotomine sandflies inferred from small subunit nuclear ribosomal DNA. *Insect Molecular Biology*, **9**, 157–168.

Aransay, A.M., Ready, P.D. & Morillas-Marquez, F. (2003) Population differentiation of *Phlebotomus perniciosus* in Spain following post-glacial dispersal. *Heredity*, **90**, 316–325.

Ashford, T.W. (1974) Sandflies (Diptera, Phlebotomidae) from Ethiopia: taxonomic and biological notes. *Journal of Medical Entomology*, **2**, 605–615.

Beati, L., Caceres, A.G., Lee, J.A. & Munstermann, L.E. (2004) Systematic relationships among *Lutzomyia* sand flies (Diptera: Psychodidae) of Peru and Colombia based on the analysis of 12S and 28S ribosomal DNA sequences. *International Journal for Parasitology*, **34**, 225–234.

Berdjane-Brouk, Z., Kone, A.K., Djimde, A.A. *et al.* (2012) First detection of *Leishmania major* DNA in *Sergentomyia (Spelaeomyia) darlingi* from cutaneous leishmaniasis foci in Mali. *PLoS ONE*, **7**, e28266.

Bisby, F.A., Roskov, Y.R., Orrell, T.M., Nicolson, D., Paglinawan, L.E., Bailly, N., Kirk, P.M., Bourgoin, T. & Baillargeon, G. (2010) Species 2000 & ITIS Catalogue of Life: 2010 Annual Checklist. Species 2000: Reading. [WWW document]. URL Digital resource at http://www.catalogueoflife.org/annual-checklist/2010 [accessed on 26 October 2014].

Blair, C. & Murphy, R.W. (2010) Recent trends in molecular phylogenetic analysis: where to next? *Journal of Heredity*, **102**, 130–138.

Bsrat, A., Berhe, N., Balkew, M. *et al.* (2015) Epidemiological study of cutaneous leishmaniasis in Saesie Tsaeda-emba district, eastern Tigray, northern Ethiopia. *Parasites & Vectors*, **8**, 758.

Cohnstaedt, L.W., Beati, L., Caceres, A.G., Ferro, C. & Munstermann, L.E. (2011) Phylogenetics of the Phlebotomine sand fly group *Verrucarum* (Diptera: Psychodidae: *Lutzomyia*). *American Journal of Tropical Medicine and Hygiene*, **84**, 913–922.

Collins, F.H. & Paskewitz, S.M. (1996) A review of the use of ribosomal DNA (rDNA) to differentiate among cryptic *Anopheles* species. *Insect Molecular Biology*, **5**, 1–9.

Curler, G.R. & Moulton, J.K. (2012) Phylogeny of psychodid subfamilies (Diptera: Psychodidae) inferred from nuclear DNA sequences with a review of morphological evidence for relationships. *Systematic Entomology*, **37**, 603–616.

Davidson, I.H. (1982) The taxon *Parvidens* (Diptera: Psychodidae: Phlebotominae) with description of a new species from Namibia. *Journal of the Entomological Society of South Africa*, **45**, 105–108.

Depaquit, J. (2014) Molecular systematics applied to Phlebotomine sandflies: review and perspectives. *Infection, Genetics and Evolution*, **28**, 744–756.

Depaquit, J., Perroteya, S., Lecointreb, G. *et al.* (1998) Molecular systematics of Phlebotominae: a pilot study. Paraphyly of the genus *Phlebotomus*. *Comptes Rendus de l'Académie des Sciences. Série III*, **321**, 849–855.

Depaquit, J., Ferte, H. & Leger, N. (2000a) The subgenus *Paraphlebotomus (Phlebotomus* Phlebotominae – Psychodidae – Diptera): a review. Morphological and molecular studies. *Annales Pharmaceutiques Françaises*, **58**, 333–340.

Depaquit, J., Ferté, H., Léger, N. *et al.* (2000b) Molecular systematics of the Phlebotomine sandflies of the subgenus *Paraphlebotomus* (Diptera, Psychodidae, *Phlebotomus*) based on ITS2 rDNA sequences. Hypotheses of dispersion and speciation. *Insect Molecular Biology*, **9**, 293–300.

Depaquit, J., Ferte, H., Leger, N. *et al.* (2002) ITS2 sequences heterogeneity in *Phlebotomus sergenti* and *Phlebotomus similis* (Diptera, Psychodidae): possible consequences in their ability to transmit *Leishmania tropica*. *International Journal for Parasitology*, **32**, 1123–1131.

Depaquit, J., Léger, N., Ferté, H. & Robert, V. (2004) Les Phlébotomes de Madagascar (Diptera: Psychodidae)-II-description de la femelle de *Phlebotomus (Anaphlebotomus) fertei* Depaquit, Léger & Robert, 2002; description du mâle et redescription de la femelle de *Phlebotomus (Anaphlebotomus) berentiensis* (Léger & Rodhain, 1978) comb. nov. *Parasite*, **11**, 201–209.

Depaquit, J., Lienard, E., Verzeaux-Griffon, A. *et al.* (2008) Molecular homogeneity in diverse geographical populations of *Phlebotomus papatasi* (Diptera, Psychodidae) inferred from ND4 mtDNA and ITS2 rDNA epidemiological consequences. *Infection, Genetics and Evolution*, **8**, 159–170.

Di Muccio, T., Marinucci, M., Frusteri, L., Maroli, M., Pesson, B. & Gramiccia, M. (2000) Phylogenetic analysis of *Phlebotomus* species belonging to the subgenus *Larroussius* (Diptera, Psychodidae) by ITS2 rDNA sequences. *Insect Biochemistry and Molecular Biology*, **30**, 387–393.

Duckhouse, D.A. & Lewis, D.J. (1980) Family Psychodidae. *Catalogue of the Diptera of the Afrotropical Region* (ed. by R.W. Crosskey), pp. 93–105. British Museum (Natural History), London.

Elnaiem, D.A., Hassan, K.H., Ward, R.D., Miles, M.A. & Frame, I.A. (1998) Infection rates of *Leishmania donovani* in *Phlebotomus orientalis* from visceral leishmaniasis focus Eastern Sudan. *Annals of Tropical Medicine and Parasitology*, **92**, 229–232.

El Sayed, S.M., El Raaba, P.M. & El Nur, A. (1991) Daily and seasonal activities of some sandflies from surrugia village, Khartoum, Sudan. *Parassitologia*, **33**, 205–215.

Espíndola, A., Buerki, S., Jacquier, A., Ježek, J. & Alvarez, N. (2012) Phylogenetic relationships in the subfamily Psychodinae (Diptera, Psychodidae). *Zoologica Scripta*, **41**, 489–498.

Esseghir, S. & Ready, P.D. (2000) Speciation of *Phlebotomus* sandflies of the subgenus *Larroussius* coincided with the late Miocene-Pliocene aridification of the Mediterranean subregion. *Biological Journal of the Linnean Society*, **70**, 189–219.

Evenhuis, N.L., Pape, T., Pont, A.C. & Thompson, F.C. (2010) *Biosystematic Database of World Diptera* [WWW document]. URL http://www.diptera.org [accessed on 2 May 2014].

Fairchild, G.B. (1955) The relationships and classification of the Phlebotominae (Diptera: Psychodidae). *Annals of the Entomological Society of America*, **48**, 182–196.

Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R.C. (1994) DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294–299.

Franco, F.A.L., Morillas-Marquez, F., Baron, S.D. *et al.* (2010) Genetic structure of *Phlebotomus (Larroussius) ariasi* populations, the vector of *Leishmania infantum* in the western Mediterranean: epidemiological implications. *International Journal for Parasitology*, **40**, 1335–1346.

Galati, B.E.A. (1995) Phylogenetic systematics of Phlebotominae (Diptera, Psychodidae) with emphasis on American groups. *Boletín de la Dirección de Malariología y Saneamiento Ambiental*, **35**, 133–142.

Galati, B.E.A. (2010) Bioecologia e Identificação de Phlebotominae. Atualização do capítulo 2 –Morfologia e Taxonomia: 2,1, Classificação de Phlebotominae. *Flebotomíneos do Brasil* (ed. by E.F. Rangel and R. Lainson) (org.), pp. 23–51 e 2.2. Morfologia, Terminologia de Adultos e Identificação dos táxons da América, p. 53–75. FIOCRUZ, Rio de Janeiro 2003 367 p.

Galati, B.E.A. (2014) 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, Vol. **I**, p. 127 p. Press, Departamento de Epidemiologia Faculdade de Saúde Pública Universidade de São Paulo, São Paulo.

Gebre-Michael, T., Pratlong, F. & Lane, P.R. (1993) *Phlebotomus (Phlebotomus) duboscqi* (Diptera: Phlebotominae) naturally infected with *Leishmania major* in Southern Ethiopia. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **87**, 10–11.

Gebre-Michael, T., Balkew, M., Berhe, N., Hailu, A. & Mekonnen, Y. (2010) Further studies on the phlebotomine sandflies of the kala-azar endemic lowlands of Humera-Metema (North-West Ethiopia) with observations on their natural blood meal sources. *Parasites & Vectors*, **3**, 6.

Giribet, G. (2003) Stability in phylogenetic formulations and its relationship to nodal support. *Systematic Biology*, **52**, 554–564.

Gutiérrez, M.A.C., Vivero, R.J., Vélez, I.D., Porter, C.H. & Uribe, S. (2014) DNA barcoding for the identification of sand fly species (Diptera, Psychodidae, Phlebotominae) in Colombia. *PLoS ONE*, **9**, e85496.

Hailu, A. & Formmel, D. (1993) Leishmaniasis in Ethiopia. *The Ecology of Health and Disease in Ethiopia* (ed. by H. Kloos and Z.A. Zein), pp. 375–388. West View Press, Boulder, Colorado.

Hailu, A., Di Muccio, T., Abebe, T., Hunegnaw, M., Kager, P.A. & Gramiccia, M. (2006) Isolation of *Leishmania tropica* from an Ethiopian cutaneous leishmaniasis patien**t**. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **100**, 53–58.

Hamarsheh, O. (2011) Distribution of *Leishmania major* zymodemes in relation to populations of *Phlebotomus papatasi* sand flies. *Parasites & Vectors*, **4**, 1–6.

Hennig, W. (1972) Beiträge zur Kenntnis der rezenten und fossilen Carnidae, mit besonderer Berücksichtigung einer neuen Gattung aus Chile (Diptera: Cyclorrhapha). *Stuttgarter Beiträge zur Naturkunde*, **240**, 1–20.

Hillis, D.M. & Dixon, M.T. (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. *The Quarterly Review of Biology*, **66**, 411–453.

Ibrahim, E.A. (1990) *Phlebotomus papatasi*, a possible vector of cutaneous leishmaniasis in Khartoum, Sudan. *Journal of the Egyptian Society of Parasitology (Cairo)*, **20**, 349–350.

Ilango, K. (2004) Phylogeny of the Old World Phlebotomine sandflies (Diptera: Psychodidae) with special reference to structural diversity of female spermathecae. *Oriental Insects*, **38**, 419–462.

Jorquera, A., Gonzalez, R., Marchan-Marcano, E., Oviedo, M. & Matos, M. (2005) Multiplex-PCR for detection of natural *Leishmania* infection in *Lutzomyia* spp. captured in an endemic region for cutaneous leishmaniasis in state of Sucre, Venezuela. *Memórias do Instituto Oswaldo Cruz*, **100**, 45–48.

Kato, H., Uezato, H., Gomez, E.A., Terayama, Y., Calvopina, M., Iwata, H. & Hashiguchi, Y. (2007) Establishment of a mass screening method of sand fly vectors for *Leishmania* infection by molecular biological methods. *American Journal of Tropical Medicine and Hygiene*, **77**, 324–329.

Katoh, K., Misawa, K., Kuma, K. & Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, **30**, 3059–3066.

Khalid, N., Elnaiem, D., Aboud, M., Al Rabba, F. & Tripet, F. (2010) Morphometric and molecular differentiation of *Phlebotomus* (*Phlebotomus*) sandflies. *Medical and Veterinary Entomology*, **24**, 352–360.

Khalid, N.M., Aboud, M.A., Alrabba, F.M., Elnaiem, D.E.A. & Tripet, F. (2012) Evidence for genetic differentiation at the microgeographic scale in *Phlebotomus papatasi* populations from Sudan. *Parasites & Vectors*, **5**, 249.

Kjer, K.M., Gillespie, J.J. & Ober, K.A. (2007) Opinions on multiple sequence alignment, and an empirical comparison of repeatability and accuracy between POY and structural alignment. *Systematic Biology*, **56**, 133–146.

Lane, R.P. (1993) Sandflies (Phlebotominae). In *Medical Insects and Arachnids* (Lane, R.P. and Crosskey, R.W., eds.), pp. 92–96. Chapman and Hall, London.

Léger, N. & Rodhain, F. (1978) *Sergentomyia berentiensis* n. sp. (Diptera, Psychodidae). Description à partir d'un exemplaire femelle

récolté à Madagascar. *Bulletin de la Societe de Pathologie Exotique*, **71**, 476–479.

Lehtonen, S. (2007) Phylogeny estimation and alignment via POY versus Clustal + PAUP*: a response to Ogden and Rosenberg. *Systematic Biology*, **57**, 653–657.

Lemmon, A.R., Brown, J.M., Stanger-Hall, K. & Lemmon, E.M. (2009) The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and Bayesian inference. *Systematic Biology*, **58**, 130–145.

Lewis, D.J. (1982) A taxonomic review of the genus *Phlebotomus* (Diptera: Psychodidae). *Bulletin of the British Museum of Natural History (Entomology)*, **45**, 121–209.

Lewis, D.J. & Kirk, R. (1946) Five new species of *Phlebotomus* (Diptera, Psychodidae) from the Sudan. *Proceedings of the Royal Entomological Society of London, Series B*, **15**, 55–62.

Lins, R., Souza, N. & Peixoto, A. (2008) Genetic divergence between two sympatric species of the *Lutzomyia longipalpis* complex in the paralytic gene, a locus associated with insecticide resistance and lovesong production. *Memórias do Instituto Oswaldo Cruz*, **103**, 736–740.

Liu, K., Nelesen, S., Raghavan, S., Linder, C.R. & Warnow, T. (2009) Barking up the wrong treelength: the impact of gap penalty on alignment tree accuracy. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **6**, 7–21.

Marcondes, C.B. (2007) A proposal of generic and subgeneric abbreviations for phlebotomine sandflies (Diptera: Psychodidae: Phlebotominae) of the world. *Entomological News*, **118**, 351–356.

Martin-Sanchez, J., Gallego, M., Baron, S., Castillejo, S. & Morillas-Marquez, F. (2006) Pool screen PCR for estimating the prevalence of *Leishmania infantum* infections in sandflies (Dipter: Nematocera, Phlebotomidae). *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **100**, 527–532.

Mazzoni, C.J., Araki, A.S., Ferreira, G.E., Azevedo, R.V., Barbujani, G. & Peixoto, A.A. (2008) Multilocus analysis of introgression between two sand fly vectors of leishmaniasis. *BMC Evolutionary Biology*, **8**, 141.

Miller, M.A., Pfeiffer, W. & Schwartz, T. (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Gateway Computing Environments Workshop (GCE).* pp. 1–8.

Ogden, T.H. & Rosenberg, M.S. (2007) Alignment and topological accuracy of the direct optimization approach via POY and traditional phylogenetics via ClustalW+PAUP*. *Systematic Biology*, **56**, 182–193.

Peixoto, A.A., Gomes, C.A., De Amoretty, P.R., Lins, R.M., Meireles-Filho, A.C., De Souza, N.A. & Kyriacou, C.P. (2001) New molecular markers for phlebotomine sand flies. *International Journal for Parasitology*, **31**, 635–639.

Pinto, I.S., Andrade Filho, J.D., Santos, C.B., Falqueto, A. & Leite, Y.L.R. (2010) Phylogenetic relationships among species of *Lutzomyia*, subgenus *Lutzomyia* (Diptera: Psychodidae). *Journal of Medical Entomology*, **47**, 16–21.

Quate, L.W. (1964) *Phlebotomus* sandflies of the Paloich Area in the Sudan (Diptera, Psychodidae). *Journal of Medical Entomology*, **1**, 213–268.

Rassi, Y., Abai, M.R., Javadian, E. *et al.* (2008) Molecular data on vectors and reservoir hosts of zoonotic cutaneous leishmaniasis in central Iran. *Bulletin de la Societe de Pathologie Exotique (Paris)*, **101**, 425–428.

Ready, P.D. (2011) Should sand fly taxonomy predict vectorial and ecological traits? *Journal of Vector Ecology*, **36**, S17–S22.

Rispail, P. & Léger, N. (1998a) Numerical taxonomy of Old World Phlebotominae (Diptera: Psychodidae). 2. Restatement of classification upon subgeneric morphological characters. *Memórias do Instituto Oswaldo Cruz*, **93**, 787–793.

Rispail, P. & Léger, N. (1998b) Numerical taxonomy of the Old World Phlebotominae (Diptera: Psychodidae). 1. Considerations of morphological characters in the genus *Phlebotomus* Rondani & Berté 1840. *Memórias do Instituto Oswaldo Cruz*, **93**, 773–785.

Roure, B., Baurain, D. & Philippe, H. (2013) Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. *Molecular Biology and Evolution*, **30**, 197–214.

Sanders, J. (2009) Cladescan, a program for automated phylogenetic sensitivity analysis. *Cladistics*, **26**, 114–116.

Scarpassa, V.M. & Alencar, R.B. (2013) Molecular taxonomy of the two *Leishmania* vectors *Lutzomyia umbratilis* and *Lutzomyia anduzei* (Diptera: Psychodidae) from the Brazilian Amazon. *Parasites & Vectors*, **6**, 258.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. & Flook, P. (1994) Evolution weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, **87**, 651–701.

Skála, Z. & Zrzavý, J. (1994) Phylogenetic reticulations and cladistics: discussion of methodological concepts. *Cladistics*, **10**, 305–313.

Soto, S.I., Lehmann, T., Rowton, E.D., Velez, B.I. & Porter, C.H. (2001) Speciation and population structure in the morphospecies *Lutzomyia longipalpis* (Lutz & Neiva) as derived from the mitochondrial ND4 gene. *Molecular Phylogenetics and Evolution*, **18**, 84–93.

Stamatakis, A., Hoover, P. & Rougemont, J. (2008) A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology*, **57**, 758–771.

Tabbabi, A., Rhim, A., Ghrab, J., Martin, O., Aoun, K., Bouratbine, A. & Ready, P.D. (2014) Phlebotomus (Paraphlebotomus) riouxi: a synonym of *Phlebotomus chabaudi* without any proven vectorial role in Tunisia and Algeria. *Medical and Veterinary Entomology*, **28**, 51–59.

Theodor, O. & Mesghali, A. (1964) On the Phlebotominae of Iran. *Journal of Medical Entomology*, **1**, 285–300.

Varón, A., Sy Vinh, L. & Wheeler, W.C. (2009) POY version 4: phylogenetic analysis using dynamic homologies. *Cladistics*, **26**, 72–85.

Ward, P.S. & Downie, D.A. (2005) The ant subfamily Pseudomyrmecinae (Hymenoptera: Formicidae): phylogeny and evolution of big-eyed arboreal ants. *Systematic Entomology*, **30**, 310–335.

Wheeler, W. (1996) Optimization alignment: the end of multiple sequence alignment in phylogenetics? *Cladistics*, **12**, 1–9.

Whiting, M.F. (2002) Mecoptera is paraphyletic: multiple genes and phylogeny of Mecoptera and Siphonaptera. *Zoologica Scripta*, **31**, 93–104.

WHO (2014) International travel and health: Leishmaniasis (cutaneous, mucosal and visceral forms) [WWW document]. URL http://www.who.int/ith/diseases/leishmaniasis/en/ [accessed on 23 September 2014].

Young, D.G. & Duncan, M.A. (1994) Guide to the identification and geographic distribution of *Lutzomyia* sand flies in Mexico, the West Indies, Central and South America (Diptera: Psychodidae). *Memoirs of the American Entomological Institute*, **54**, 1–881.

Accepted 27 April 2015
First published online 29 July 2015