Parasitology Research
https://doi.org/10.1007/s00436-023-07835-1

RESEARCH

Genetic diversity and haplotype analysis of *Leishmania tropica* identified in sand fly vectors of the genera *Phlebotomus* and *Sergentomyia* using next-generation sequencing technology

Amer Al-Jawabreh[1,2] · Suheir Ereqat[3] · Hanan Al-Jawabreh[2,3] · Kamal Dumaidi[1] · Abedelmajeed Nasereddin[3]

Received: 5 January 2023 / Accepted: 30 March 2023

Abstract
Next-generation sequencing (NGS) was used to investigate the genetic diversity of *Leishmania tropica* in the sand fly vector, targeting the internal transcribed spacer 1 (ITS1) of the genus *Leishmania.* Bioinformatics analyses were conducted using Galaxy, MEGA version X, DnaSP ver. 6.12.03, and PopART 1.7 software for NGS analysis, phylogenetic tree, genetic diversity, and haplotype networking, respectively. A total of 307 engorged sand flies were trapped, with an overall *Leishmania* infection rate of 9.4 (29/307) and 6.8% by NGS and ITS1-PCR, respectively. Two *Leishmania*-infected sand fly genera were identified: *Phlebotomus* (10.2%, 26/254) and *Sergentomyia* (5.7% (3/53). The phylogenetic tree showed two clusters, cluster I included the four study sequences along with 25 GenBank-retrieved DNA sequences. Cluster II consisted of three sequences from Iran and Pakistan. The genetic diversity analysis for the 29 *L. tropica* sequences showed high haplotype (gene) diversity index (Hd) (0.62 ± 0.07) but low nucleotide diversity index (π) (0.04 ± 0.01). Tajima's D, a neutrality test, is more negative in cluster I ($D = -2.0$) than in total population ($D = -1.83$), but both are equally significant ($P < 0.001$), indicating that observed variation in cluster I and whole population is less frequent than expected. The median-joining haplotype network produced a total of 11 active haplotypes. In conclusion, *L. tropica* from sand flies in Palestine is monophyletic that assembled in one main phylogroup and one haplotype.

Keywords Genetic diversity · Haplotype analysis · Next-generation sequencing · *Leishmania tropica* · *Phlebotomus* · *Sergentomyia*

Introduction

Cutaneous leishmaniasis (CL) and visceral leishmaniasis (VL) are vector-borne diseases caused by haemoflagellate parasites of the genus *Leishmania*. Early in the twentieth century, sand flies of the genus *Phlebotomus* were incriminated as vectors of the leishmanial parasite by the Sergent brothers (1921) and Shortt and Swaminath (1928) (Dedet 2005; Shortt and Swaminath 1928). Sand flies belong to the order Diptera, suborder Nematocera, family Psychodidae, subfamily Phlebotominae, and the genera *Phlebotmus*, *Sergentomyia*, and *Chinius*, with the first containing 110 species (Killick-Kendrick 1990). Species of the genera *Phlebotomus* and *Sergentomyia* occur in the Mediterranean Region (Akhoundi et al. 2016). *Phlebotomus sergenti* is the main permissive vector of *Leishmania tropica* species in the region (Kamhawi 2006; Sawalha et al. 2017; Sawalha et al. 2003; Schnur et al. 1972). Strains of *L. tropica* are the main cause of CL in the Old World, being transmitted zoonotically and anthroponotically (Alvar et al. 2012; Ghatee et al. 2018; Jacobson 2003; Svobodova et al. 2006). The species *L. tropica* has been shown to be genetically diverse (Schonian et al. 2001; Schwenkenbecher et al. 2004; Schwenkenbecher et al. 2006) compared to other Old World species, e.g., *L. major* (Al-Jawabreh et al. 2008), *L. infantum* (*chagasi*) (Kuhls et al. 2011), and *L. donovani* (Alam et al. 2009;

Section Editor: Dorien Mabille

✉ Suheir Ereqat
sereqat@staff.alquds.edu

[1] Arab American University, Jenin, Palestine

[2] Leishmaniases Research Unit—Jericho, Jericho, Palestine

[3] Biochemistry and Molecular Biology Department, Faculty of Medicine, Al-Quds University, Abu Deis, East Jerusalem, Palestine

Published online: 10 April 2023

Srivastava et al. 2011). However, the genetic diversity seen among strains of *L. tropica* does not coincide with any tangible variables such as geographical origin and/or clinical picture (Schonian et al. 2001).

The methods used to demonstrate intra-species genetic variation in *Leishmania spp.* have developed extensively over time. It started by checking the antigenic variation seen among products secreted by the parasites termed "excreted factors" (Rioux et al. 1980; Schnur et al. 1972). Isoenzyme electrophoresis was developed to find intra- and inter-species differences in isozyme patterns using a set of known enzymes (Kreutzer and Christensen 1980). At the beginning of the millennium, molecular-based methods were introduced to genotype and screen for DNA variability of *L. tropica*. In addition to the internal transcribed spacer 1 (ITS1) region, methods like DNA fingerprinting, single-strand conformation polymorphism (SSCP), and multilocus microsatellite typing (MLMT) were used (Schonian et al. 2001; Schwenkenbecher et al. 2006). These molecular biological methods depended mainly on selected target DNA sequences in the genome, such as the internal transcriber spacer 1 (ITS1), and with aid of bioinformatics tools, genetic variation was discerned (Schwenkenbecher et al. 2006).

During the last decade or so, a new sequencing technology, called next-generation sequencing (NGS), was developed, leading to substantial improvement in the quality and yield of DNA sequences. This led to an advance in the diagnosis of infectious diseases and the investigation of genetic diversity of their causative agents, including leishmaniases (Carvalho et al. 2020; Chen et al. 2020; Kounosu et al. 2019). This novel method emerged in parallel to the abundance of genetic data provided by the availability of complete genomes of the *Leishmania spp*. Here, NGS was used to determine the natural rate of infection of *Leishmania spp.* in sand fly vectors as well as identify the suspected sand fly species and investigate the genetic diversity of *L. tropica* strains in sand fly vectors compared to that in different hosts.

Materials and methods

The study area

This study was conducted in the village of Tayasir (32°20′26″, N 35°23′49″E) on the hills overlying the Jordan Valley as part of the Tubas district north of the West Bank, Palestine (Palestinian-Central-Bureau-of-Statistics 2020). The area was reported in the last couple of years as an active focus of *Leishmania tropica* (Al-Jawabreh et al. 2017).

Sand fly trapping and identification

In 2019, engorged sand flies that have higher probability of harboring *Leishmania* were collected during biting season (June to end of October) in and next to rock hyrax active dens on the outskirts of the study village. In each of the selected dens, one CDC miniature light trap (John W. Hock Co., Gainesville, FL, USA) was placed inside the den to attract sand flies during night foraging, and two castor oil–soaked papers (21.3 29.5 cm) stapled vertically on wooden stakes at a height of 20 cm were placed 2–4 m around the entrance of the den to intercept sand flies during entering and exiting (Alexander 2000; Killick-Kendrick 1987). Traps were set just before sunset and collected the next morning, just before sunrise. Sandfly trapping sessions were carried out 13 times with a frequency of one trapping session every 5–7 days in all nine sites. The use of two types of traps was to increase the yield of sand flies collected, and both of these types are suited for the area of survey around rock hyrax (*Procavia capnesis*) dens.

Collected engorged sand flies were removed from traps, washed in detergent solution, labeled, and stored in 70% alcohol for each trap separately. Then engorged female sand flies were dissected, in which head and the terminal abdominal segments mounted in Berlese's medium for sand fly classification based on taxonomic keys. The remaining parts, thorax and abdomen, were kept separately in 70% ethanol for determining *Leishmania* infection in sand flies by polymerase chain reaction (PCR) (Lane 1986; Lewis and Buttiker 1982; Lewis 1982; Perffl'ev 1966).

Extraction of DNA from sand flies

The DNA was extracted from the thorax and abdomen of the engorged sand flies individually as described by others with modifications in the preparatory pre-extraction step (Caligiuri et al. 2019; Casaril et al. 2017). In brief, the 70% ethanol-preserved thorax and abdomen of the engorged sandflies were transferred into a 1.5-mL microtube filled with 250 µL TENCa buffer (30 mM Tris-HCl pH 8, 10 mM EDTA, 50 mM NaCl, and 5 mM $CaCl_2$). Wearing a head loupe, the sand fly parts were ground with fine glass rod against the walls of the microtube. Glass disruptor beads (0.5 mm, Scientific Industries, USA) were added to ¼ tube full and beaten at 2800 rpm for 5 min using a cell disruptor (Disruptor Genie, Scientific Industries, USA). The microtube was centrifuged for 30 s to settle down the sandfly parts in lysis buffer. Then, 20 µL (final conc. 400 µg = 1.6 µg/µL) of proteinase K (10 mg/mL) and 2.5 µL of Triton X-100 were added. Again, the microtube was vortexed for 10 s and centrifuged for 30 s to settle down the sand fly parts in lysis buffer. Then, the tube was incubated at 56 °C with

continuous mixing at 500 rpm using thermal block (Eppendorf thermomixer R) overnight or until the sand fly parts disappear indicating complete digestion. The DNA chloroform extraction, precipitation and elution, and storage were performed as described in the references.

ITS1 library preparation and amplicon based-next generation sequencing

Amplification of ITS1 partial regions

Amplification was run as described elsewhere (Nasereddin et al. 2022). Briefly, *Leishmania* DNA was detected using 2 primers: forward (ITS1NGSF:5′-AGCTGGATCATTTTCCGATG) and reverse (ITS1NGSR: 5′-ATCGCGACACGTTATGTGAG). Both primers were modified by adding the Illumina overhang adapter sequences at the 5′ ends of the forward (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG) and reverse primers (GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG), producing 343bp-segment of the ITS1 region of the genus *Leishmania*. Nuclease-free water was used as negative control, and a reference *Leishmania* strain (*MHOM/PS/2001/ISL590*) was used as positive control. Gel electrophoresis was run, visualized, and captured by gel documentation system. DNA library was prepared by using 20 µL of PCR product as described earlier (Nasereddin et al. 2022).

Identification of Leishmania-positive sand flies by NGS targeting a partial sequence of 18S rRNA

NGS was used to identify the sand fly DNA using the modified primers SFNGSF: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGCGGTTAAAACGTTCGTAG and SFNGSR: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGACCGGTAAAACATCCGTCAC primers, targeting a 230 bp-fragment from the 18S rRNA gene of Phlebotominae subfamily as described elsewhere (Giantsis et al. 2017; Nasereddin et al. 2022). Controls and gel electrophoresis were implemented as described above.

The PCR product was cleaned and underwent a second amplification round to include unique index sequences (N7XX and S5XX) for barcoding of each sample using Nextera XT Index Kit (Illumina, San Diego, CA, USA). Then followed by pooling, purification, normalization to 4 nM, and then sequenced as described previously from the forward read direction (Nasereddin et al. 2022).

Bioinformatics analysis and ITS-1 Sanger sequencing

Free online Galaxy analysis pipeline (https://usegalaxy.org/) was used for bioinformatics analysis as described previously (Nasereddin et al. 2022). The ITS1 region was chosen over others due to being one of the most commonly used targets for laboratory diagnosis of leishmaniasis, and, unlike kDNA, ITS1 has enough genetic variation to distinguish between almost all Leishmania species in the world and all those in the study area, *L. major*, *L. tropica*, and *L. infantum*. Moreover, polymorphic sequences in ITS1 region for *Leishmana* along with 18S rDNA for sand fly were used to create virtual specific probes that can be used in our workflow bioinformatics analysis using Galaxy.

A receiver operating characteristic curve, or ROC curve, was used to determine the threshold cut-off value for number of NGS reads based on plotting the true positive rate (sensitivity) against false positive rate (1-specificity) at different threshold values.

Conventional PCR was used to amplify the ribosomal the full internal transcribed spacer 1 (ITS1) region (343 bp) separating the genes coding for ssu rRNA and L5.8S rRNA as previously described (Al-Jawabreh et al. 2004; Al-Jawabreh et al. 2006; Schonian et al. 2003), to confirm the NGS results. The reaction was conducted using master mix kit PCR-Ready from Syntezza (Syntezza Bioscience Ltd., Jerusalem). The ITS1-positive samples were commercially Sanger sequenced.

Leishmania tropica genetic diversity analysis

The leishmanial DNA sequences, obtained by Sanger and NGS, along with sequences from the GenBank, were aligned and analyzed by neighbor-joining phylogenetic 1000-iteration bootstrap consensus tree using the MEGA version X (Kumar et al. 2018).

DnaSP ver. 6.12.03 was used to calculate the genetic diversity indices. The indices included haplotype (gene) diversity (Hd), calculated as the probability that two randomly chosen haplotypes from a given population were different, and nucleotide diversity (π) as the average number of nucleotide differences per site between two DNA sequences in all possible sample pairs in the population. Other calculated genetic diversity indices included (K) as average number of nucleotide differences (per DNA sequence) between any two given DNA sequences and (S) as the number of variable sites in a DNA sequence.

Furthermore, the genetic differentiation statistics Fst and Nm as the average level of gene flow based on allele frequencies under the infinite-site model, as well as neutrality tests including Tajima's *D* as the difference between the mean pairwise nucleotide differences (π) and the number of segregating sites (S) relative to their standard error, and Fu Li's *F* test as number of derived nucleotide variants observed only once in a sample with the mean pairwise difference between sequences, were calculated using DnaSP ver. 6.12.03 (Rozas et al. 2017). The analyses were conducted separately for each

cluster and for all clusters grouped together. Haplotype analysis network based on single nucleotide variation (SNV) was built based on median-joining method using PopART 1.7 software (Bandelt et al. 1999) with zero epsilon as a default parameter. Nexus format file generated by DnaSP version 6.12.03 was used as an input file for PopART 1.7 software. Coloring of haplotype network was based on country of origin of leishmanial DNA sequences.

Results

Sand fly populations and rate of infection

The total number of engorged sand flies trapped and tested during the 12-week study period was 307 (17%) out of a total of 1786 collected sand flies. The ROC curve showed that the cut-off value was 12,500 reads, which is considered positive for *Leishmania*. Table 1 shows that the overall infection rate of *Leishmania* in sand fly vector was 9.4% (29/307) as revealed by NGS compared to 6.8% by ITS1-PCR. Of the two genera, *Phlebotomus* and *Sergentomyia*, seven species of *Phlebotomus* were trapped. The infection rate of *Leishmania* in *Phlebotomus* based on NGS was 10.2% (26/254), while it was 5.7% (3/53) in *Sergentomyia*. Of the *Phlebotomus* genus, approximately 44% (136/307) were *P. sergenti* followed by 16.2% (50/307) *P. major complex*. The *Leishmania* infection rate in *P. sergenti* was 13.8% (19/136). One *P. sergenti* sand fly harbored *L. major* the others ($n = 15$) have *L. tropica* (Table 1). The genus *Sergentomyia* was present (17%) with one harboring *L. tropica*. All ITS1-PCR *leishmania*-positive sand flies were positive by NGS, with the latter showing superiority of 2.6% ($n = 7$) more positive cases.

Of the 29 *Leishmania* positive strains, five were ITS-1 sanger-sequenced, and 17 were Ilumina-ITS-1 NGS sequenced. The twenty-two sequences produced by the study

Table 2 *Leishmania* strains deposited by this study in the GenBank with their corresponding sand fly vector

Accession no.	Sand fly	*Leishmania*	Type of sequence
MT966013	*P. sergenti*	*L. tropica*	Sanger
MT966014	*P. sergenti*	*L. tropica*	Sanger
MT966015	*P. sergenti*	*L. tropica*	Sanger
MT966016	*Sergentomyia spp.*	*Leishmania spp.*	Sanger
MT966017	*P. sergenti*	*Leishmania spp.*	Sanger
MW111284	*P. sergenti*	*L. tropica*	Ilumina NGS
MW111285	*P. sergenti*	*L. tropica*	Ilumina NGS
MW111286	*P. sergenti*	*L. tropica*	Ilumina NGS
MW111287	*P. sergenti*	*L. tropica*	Ilumina NGS
MW111288	*P. sergenti*	*L. major*	Ilumina NGS
MW111289	*P. sergenti*	*L. tropica*	Ilumina NGS
MW111290	*P. sergenti*	*L. tropica*	Ilumina NGS
MW111291	*P. sergenti*	*L. tropica*	Ilumina NGS
MW111292	*P. sergenti*	*L. tropica*	Ilumina NGS
MW111293	*P. sergenti*	*L. tropica*	Ilumina NGS
MW111294	*P. sergenti*	*L. tropica*	Ilumina NGS
MW111295	*P. sergenti*	*L. tropica*	Ilumina NGS
MW111296	*P. sergenti*	*L. tropica*	Ilumina NGS
MW111297	*P. sergenti*	*L. tropica*	Ilumina NGS
MW111298	*P. sergenti*	*L. tropica*	Ilumina NGS
MW111299	*P. sergenti*	*L. tropica*	Ilumina NGS
MW111302	*Sergentomyia dentata*	*L. tropica*	Ilumina NGS

Underlined accession numbers were selected for phylogenetic analyses

Table 1 Sand fly populations trapped in the study area in 2019 and the corresponding *Leishmania* infection rate per species as revealed by ITS1-PCR and NGS

Species	ITS1-PCR		NGS			
	No. (%)	No. infected (%)	Leish (%)	Tro	Maj	*Spp.*
P. sergenti	136 (44)	14 (10.3)	19 (13.8)	15	1	3
Sergentomyia spp.	53 (17)	2 (3.8)	3 (5.7)	1	0	2
P. major complex	50 (16)	5 (10)	6 (12)	6	0	0
P. tobbi	35 (11)	0 (0)	0	0	0	0
P. perfiliewi	11 (3.5)	0 (0)	0	0	0	0
P. papatasi	7 (2.2)	0 (0)	0	0	0	0
P. kazeruni	4 (1.3)	0 (0)	0	0	0	0
P. arabicus	1 (0.3)	0 (0)	0	0	0	0
Phlebotomus spp.	5 (1.6)	0 (0)	0	0	0	0
Undetermined	5 (1.6)	0 (0)	1 (20)	0	0	0
Total	307 (100)	21 (6.8)	29 (9.4)	22	1	5

Leish, *Leishmania*; Tro, *L. tropica*; Maj, *L. major*; ITS1-PCR, internal transcribed spacer 1-polymerase chain reaction; NGS, new-generation sequencing

were deposited in the GenBank (Table 2). Most of the strains (82%) were *L. tropica*, one was *L. major* found in *P. sergenti*, and one *L. tropica* found in *Sergentomyia dentata*.

Phylogenetic analysis

Neighbor-joining phylogenetic tree of the ITS1 noncoding region of *Leishmania* genome was constructed from the four selected *L. tropica* GenBank-deposited sequences of the current study (Table 2) along with 25 sequences retrieved from the GenBank representing different countries and different reservoirs. Two of the four were Sanger-sequenced, while the other two were NGS-sequenced, with the latter representing the 16 identical NGS sequences. The phylogenetic tree showed two clusters: the first (cluster I) is a major cluster that contained most sequences, including the study sequences, while the second one (cluster II) consisted of three sequences from Iran and Pakistan (Fig. 1). The two clusters were strongly supported by bootstrap values. The length of the two Sanger sequences included in the phylogenetic tree is approximately 272 bp, compared to 151 bp for the NGS sequences.

Genetic differentiation and diversity

Based on the two clusters generated by the phylogenetic tree, genetic diversity in the form of haplotype diversity (Hd), nucleotide diversity (π) indices, and neutrality tests were calculated (Table 3).

The genetic diversity analysis for the 29 *L. tropica* sequences showed the haplotype (gene) diversity index (Hd) was high (0.62 ± 0.07) compared to low nucleotide diversity index (π) (0.04 ± 0.01). The nucleotide diversity index (π) for clusters I and II was as low as 0.01 ± 0.004 and 0.10 ± 0.03, respectively. Conversely, the haplotype (gene) diversity index (Hd) for same clusters was as high as 0.53 ± 0.01 and 1.0 ± 0.07, respectively (Table 3). With cluster I containing more sequences (22) than cluster II (3), the DnaSP ver. 6.12.03 estimated the total number of haplotypes for clusters I and II at 8 and 3, respectively (Table 3). However, haplotype diversity index (Hd), nucleotide diversity index (π), as well as the average number of nucleotide differences between any two sequences (k) and number of segregating (polymorphic) sites (S) were higher in cluster II than cluster I, which subsequently reflected high genetic diversity in cluster II. The 18 Palestinian *L. tropica* isolates

Fig. 1 Consensus neighbor-joining phylogenetic tree (1000 replicates) of *L. tropica* based on the ITS1 non-coding region with concomitant mapping of vector and host reservoirs by country of origin. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The branch labels in red represent the study sequences that originated from sand fly vectors based on Sanger sequencing method, while the blue labels represent the sequences from sand fly vectors based on 151-bp Illumina NGS, which represent other 14 identical sequences depicted in blue triangle. All study sequences are mapped as 18 in red circle. The black labels are sequences from different hosts retrieved from the GenBank shown in the map as numbers in black circles. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 44 nucleotide sequences conducted in MEGA X (Kumar et al. 2018; Tamura and Nei 1993)

Table 3 Haplotype/nucleotide diversity and neutrality tests of the two *L. tropica* clusters based in ITS1 region

Cluster	Haplotype-nucleotide diversity							Neutrality tests	
	n	*h*	*h:n*	Hd ± SD	$\pi \pm$ SD	*K*	*S*	Tajima's *D*	Fu-Li's *F*
Cluster-I	26	8	0.3:1	0.53 ± 0.01	0.01 ± 0.004	0.75	8	− 2.00*	− 1.01
Cluster-II	3	3	1:1	1.0 ± 0.07	0.10 ± 0.03	6.00	9	ND	ND
Total	29	11	0.4:1	0.62 ± 0.10	0.04 ± 0.01	2.42	18	− 1.83*	− 1.11

n, number of sequences; *h*, number of haplotypes; *Hd*, haplotype (gene) diversity; π, nucleotide diversity (per site) (Nei 1987); *K*, average number of nucleotide differences between two randomly chosen sequences from within in the population (Tajima 1983); *S*, number of variable/segregating sites. *, $P < 0.01$

were confined to cluster I. Tajima's *D* and Fu-Li's *F* tests were negative for cluster I and the total population, while it was not done for cluster II as it contained fewer samples to be valid for calculations. Tajima's *D* is more negative in cluster I ($D = -2.0$) than in total population ($D = -1.83$), but both are equally significant ($P < 0.001$), indicting that observed variation in cluster I and whole population is less frequent than expected in the Wright–Fisher model. Moreover, cluster I and the whole population had negative values for Fu'Li's *F* statistics, however, did not depart significantly from neutrality ($P > 0.01$) (Table 3).

Inter-population pairwise genetic differentiation index (Fst), also called fixation index, between the two *L. tropica* populations (I and II), and its related analog estimate Gst, showed high values (0.64 and 0.11, respectively), with Nm value of 0.14 (Table 4), indicating genetic significant differentiation (> 0.25) and minimal migration and gene flow (Nm) between the two populations. Inter-population nucleotide difference (Kxy), the net nucleotide substitutions per site (Da), and the average number of nucleotide substitutions per site between cluster I and II (Dxy) showed high values (9.25, 0.16, and 0.11, respectively), indicating population differentiation (Table 4).

Haplotype network analysis

The median-joining haplotype network constructed by PopART 1.7 using the 29 taxa produced a total of 11 active haplotypes, with haplotypes 1 and 3 one consisting of 18 and 2 sequences, respectively (Fig. 2). All study and GenBank-retrieved *L. tropica* sequences isolated from vectors were included in haplotype 1. In addition, *L. tropica* from lizards (*Eremias velox borowskii* and *Phrynocephalus axillaris*) were contained in haplotype 1. Haplotype 1 piled sequence from six countries in Asia, Europe, and Africa (Fig 2). Peripheral haplotypes 3–6 had single nucleotide variation (SNV) from central haplotype 1 (Fig. 2), indicating closely related haplotypes, while haplotype 8 from rock hyrax (*P. capensis*), as well as 9, 10, and 11 from human hosts in Iran and Pakistan, were distanced by multiple SNVs from the haplotype 1. The mutation difference between haplotype 1, the main haplotype, and the other haplotypes ranged between 1 and 12 SNVs.

Discussion

Leishmania infection rate in sand fly vectors varies across species and fauna. Using NGS method, the present study showed *Leishmania* infection rate in *Phlebotomus spp*, the mainly incriminated Old World *Leishmania* vector, to be as high as 10.2 and 14% in *P. sergenti*. This was several folds higher than in the Galilee (Al-Galeel) focus, 50 km north of our study area, 2 decades ago, where it was 2.7% in *Phlebotomus spp.* and 1.2% in *P. sergenti* (Jacobson 2003), during which the insensitive classical sand fly dissection method was compared to the extremely sensitive NGS method that is capable of detecting one parasite or traces of its DNA (Nasereddin et al. 2022). Nonetheless, it should be mentioned that classical sand fly dissection detects the presence of the active parasite with transmission capacity, thus contributing to the actual prevalence of leishmaniasis. In one study in Morocco, the natural infection rate in *P. sergenti* was 3.24% using nested kDNA PCR.

Table 4 Inter-population differentiation indices between the two *L. tropica* clusters estimated from ITS1 region

Pop 1	Pop 2	Fst	Nm	Kxy	Dxy	Gst	Da
Cluster-I	Cluster-II	0.64	0.14	9.25	0.16	0.11	0.11

Fst, Wright's Fst (fixation index), measure of genetic differentiation between populations (range 0–1) (Wright 1951), *Nm*, gene flow and population migration among populations; Nm = (1-Fst) / 4Fst (diploid) (Hudson et al. 1992; Wright 1951); *Kxy*, average proportion of nucleotide differences between populations. *Dxy*, the average number of nucleotide substitutions per site between populations (Nei 1987); *Da*, the number of net nucleotide substitutions per site between populations (Nei 1987); *Gst*. genetic differentiation index based on the frequency of haplotypes (Nei 1973)

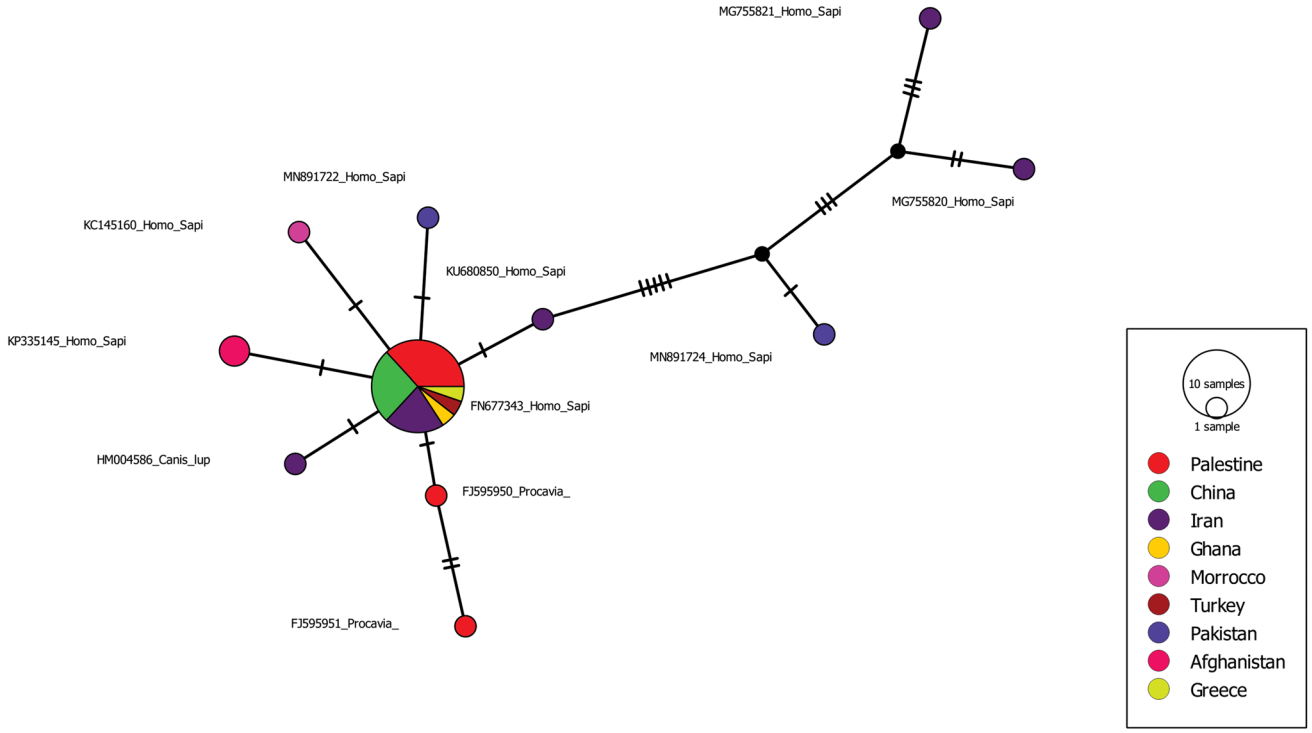

Fig. 2 Median-joining haplotype network inferred from ITS1 region in *L. tropica* constructed using PopART 1.7 with 1000 iterations. The network analysis shows 11 *L. tropica* haplotypes. Each circle represents a unique haplotype, color represents country of genome origin, and the size of the circle is proportional to number of isolates included. The lines (hatch marks) on the branches represent the mutated position with one line per mutation. Black circle represent hypothetical haplotypes

However, the study used pooled samples from unengorged sand fly females, unlike our study, which used engorged individual sand flies, leading to such discrepancy in prevalence (Mhaidi et al. 2018). This study showed that NGS was superior to ITS1-PCR as the former estimated the rate of infection to be 9.4% compared to a lower rate of 6.8% in the latter, which is in complete congruence with a previous study (Sawalha et al. 2022). Several methods have been used to estimate the *Leishmania* rate of infection in the sand fly vector, ranging from dissection method and culture to PCR targeting kinetoplast DNA (kDNA), ITS1, 18sRNA, miniexon-derived RNA gene, rRNA gene, repeated genomic sequences, and to real-time PCR (qPCR) as a more sensitive method. In addition to the high throughput, comprehensive genomic coverage, and quicker turnaround time, NGS is ultrasensitive with the ability to detect small amount of variation as it sequences millions of the same DNA fragment in parallel per run (depth of coverage), compared to Sanger sequence, which is based on the sequencing of single DNA fragment (Rivas et al. 2011; Schuster 2008; Shendure and Ji 2008). The millions of reads or copies of the same DNA fragment eliminate the errors and mutations by ignoring the very small percentage of the incorrectly sequenced DNA fragment and adopting the mainstream consensus DNA fragment. This, in addition to gene diversity and confined study area, may explain why all 16 *L. tropica* NGS sequences from sand fly vector were identical.

The two statistically supported genetic clusters produced by the neighbor-joining phylogenetic tree of the ITS1 region were supported by high genetic differentiation statistic (Fst) along with low Nm, indicating very low transmission of *L. tropica* between the two clusters. Thus affirming clusters and inter-population estimates (Kxy, Dxy, and Da) also confirmed high genetic population differentiation between the two clusters, I and II (Table 4). However, the high haplotype diversity (Hd) and low genetic diversity (π) designate closely-related haplotypes that may reflect recent expansion of population after bottleneck event during which many alleles were lost and sharp reduction in the size of population occurred, inbreeding, or genetic drift. Another explanation would be the selective sweep in which mutations increased and became fixed in a population, consequently decreasing genetic variation. The overall significant negative bias from neutrality of Tajima's *D* value supported this finding. Nonetheless, Fu-Li's *F* insignificant negative value did not substantiate this. Cluster I has more haplotypes ($h = 8$) than cluster II ($h = 3$), albeit the haplotype diversity (Hd) and nucleotide diversity (π) were lower than cluster

II, indicating more genetic diversity in cluster II. These findings are in congruence with similar studies involving *L. tropica*, except for the association of geographic origin with clusters (El Hamouchi et al. 2019).

The clustering was not related to geography, reservoir, host, or vector. Cluster I, as the major phylogroup, extended over the four continents of the Old World that included all study sequences from Palestine, regardless of the type of sequence, Sanger or NGS. Contrarily, other studies that depended on microsatellite analysis showed that *L. tropica* has higher genetic variation than the other *Leishmania spp*. However, the number of phylogroups in other local and global studies that used kDNA with RFLP (restriction fragment length polymorphism), ITS1, and MLMT (multi-locus microsatellite typing) was restricted to 1–3, which is in congruence with our study (Fakhar et al. 2016; Ghatee et al. 2018; Krayter et al. 2014). Krayter et al. concluded correlation between genetic diversity and geographic origin of strains; however, other explanations were put forward, such as the migration of the reservoir (Krayter et al. 2014). In the present study, the two clusters did not run parallel with the geographical origin or type of host or vector. In another MLMT study from Palestine, three clusters were demonstrated, with one pure cluster from Palestine and vicinity while the other two were either intermixed with strains from Africa or Asia (Azmi et al. 2017).

Haplotype network representation showed one distinct major haplotype (Hap 1) originating from six countries with six different haplotypes (2–7) star-like radiating out of it in one SNV difference. Although the network methods do not rely on evolutionary models, however, when intertwined with NJ phylogenetic tree and genetic variation analyses, the indication of common origin of haplotypes 1–8 becomes evident. At the same time, haplotypes 9–11, which appear as cluster II in the NJ tree, are distanced from each other with multiple SNVs, forming a scattered bunch of haplotypes. The appearance of hypothetical haplotypes (black circles) in the middle of the scattered haplotypes (9–11) may indicate insufficient amount of strains to build up a well-distinct group of haplotypes as in haplotype I group. A Moroccan study revealed 13 haplotypes of *L. tropica* from human hosts, with one haplotype containing 82% of the isolates (El Hamouchi et al. 2019).

The study was limited to ITS1 region only. Multiple targets such as cytochrome b mitochondrial DNA (mtDNA) and kinetoplast DNA (kDNA) in addition to ITS1 will provide a more comprehensive view of the genetic variation. Encompassing more samples from different locations in the study country and vicinity will create clearer haplotypes with less hypothetical ones and provide clues on relationship between clustering and origin of samples, whether by country or isolation source.

Conclusion

The *L. tropica* isolated from sand flies in Palestine is monophyletic that assembled in one main phylogroup and one haplotype using ITS1 region. Our study revealed one highly genetically diverse population (cluster II) and one relatively low genetically diverse population (cluster I), regardless of origin of samples, geographic, or isolation. Infection rate of *L. tropica* in *Phlebotomus spp.* is higher than expected. Epidemiological surveys of *Leishmania* infection in sand fly vectors are essential to assess control strategies and populations at risk.

Author contributions Conceptualization, A. J, S.E, and A.N; methodology, S. E, A.N, K. D, and H.J; software, A. N, and A.J; validation, S. E, K.D, and H.J; formal analysis, A. N, S.E, K. D, and AJ; investigation, S. E, A.J, and K.D; resources, H.J; data curation, S. E and AN; writing—original draft preparation, S.E; writing—review and editing, A. J and A.N; visualization, K.D; supervision, A.N; project administration, A.J; funding acquisition, A.J.

Funding This work was supported by USAID (M33-07).

Data availability All data are available in the manuscript, including the accession numbers of 22 DNA sequences that have been deposited in the NCBI GenBank. However, the datasets analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication The authors declare that they are aware of the publication of this study.

Competing interest The authors declare no competing interests.

References

Akhoundi M et al (2016) A historical overview of the classification, evolution, and dispersion of Leishmania parasites and sandflies. PLoS Negl Trop Dis 10(3):e0004349. https://doi.org/10.1371/journal.pntd.0004349

Al-Jawabreh A et al (2008) Identification of geographically distributed sub-populations of Leishmania (Leishmania) major by microsatellite analysis. BMC Evol Biol 8:183. https://doi.org/10.1186/1471-2148-8-183

Al-Jawabreh A et al (2017) Molecular epidemiology of human cutaneous leishmaniasis in Jericho and its vicinity in Palestine from 1994 to 2015. Infect Genet Evol 50:95–101. https://doi.org/10.1016/j.meegid.2016.06.007

Al-Jawabreh A et al (2004) The recent emergence of Leishmania tropica in Jericho (A'riha) and its environs, a classical focus of L. major. Trop Med Int Health 9(7):812–816. https://doi.org/10.1111/j.1365-3156.2004.01268.x

Al-Jawabreh A, Schoenian G, Hamarsheh O, Presber W (2006) Clinical diagnosis of cutaneous leishmaniasis: a comparison study between standardized graded direct microscopy and ITS1-PCR

of Giemsa-stained smears. Acta Trop 99(1):55–61. https://doi.org/10.1016/j.actatropica.2006.07.001

Alam MZ et al (2009) Multilocus microsatellite typing (MLMT) reveals genetic homogeneity of Leishmania donovani strains in the Indian subcontinent. Infect Genet Evol 9(1):24–31. https://doi.org/10.1016/j.meegid.2008.09.005

Alexander B (2000) Sampling methods for phlebotomine sandflies. Med Vet Entomol 14(2):109–122. https://doi.org/10.1046/j.1365-2915.2000.00237.x

Alvar J et al (2012) Leishmaniasis worldwide and global estimates of its incidence. PloS One 7(5):e35671. https://doi.org/10.1371/journal.pone.0035671

Azmi K et al (2017) Increased prevalence of human cutaneous leishmaniasis in Israel and the Palestinian Authority caused by the recent emergence of a population of genetically similar strains of Leishmania tropica. Infect Genet Evol 50:102–109. https://doi.org/10.1016/j.meegid.2016.07.035

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16(1):37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036

Caligiuri LG et al (2019) Optimization of DNA Extraction from individual sand flies for PCR amplification. Methods Protoc 2(2). https://doi.org/10.3390/mps2020036

Carvalho KSS et al (2020) Application of next generation sequencing (NGS) for descriptive analysis of 30 genomes of Leishmania infantum isolates in Middle-North Brazil. Sci Rep 10(1):12321. https://doi.org/10.1038/s41598-020-68953-9

Casaril AE et al (2017) Standardization of DNA extraction from sand flies: application to genotyping by next generation sequencing. Exp Parasitol 177:66–72. https://doi.org/10.1016/j.exppara.2017.04.010

Chen H et al (2020) leishmaniasis diagnosis via metagenomic next-generation sequencing. Front Cell Infect Microbiol 10:528884. https://doi.org/10.3389/fcimb.2020.528884

Dedet JP (2005) Stages in the identification of phlebotomine sandflies as vectors of Leishmaniases and other tropical diseases. Parassitologia 47(3-4):291–295

El Hamouchi A, Ajaoud M, Arroub H, Charrel R, Lemrani M (2019) Genetic diversity of Leishmania tropica in Morocco: does the dominance of one haplotype signify its fitness in both predominantly anthropophilic Phlebotomus sergenti and human beings? Transbound Emerg Dis 66(1):373–380. https://doi.org/10.1111/tbed.13031

Fakhar M et al (2016) Genetic diversity of Leishmania tropica strains isolated from clinical forms of cutaneous leishmaniasis in rural districts of Herat province, Western Afghanistan, based on ITS1-rDNA. Infect Genet Evol 41:120–127. https://doi.org/10.1016/j.meegid.2016.03.031

Ghatee MA, Mirhendi H, Marashifard M, Kanannejad Z, Taylor WR, Sharifi I (2018) Population structure of leishmania tropica causing anthroponotic cutaneous leishmaniasis in Southern Iran by PCR-RFLP of kinetoplastid DNA. Biomed Res Int 2018:6049198. https://doi.org/10.1155/2018/6049198

Giantsis IA, Chaskopoulou A, Claude Bon M (2017) Direct multiplex PCR (dmPCR) for the Identification of six phlebotomine sand fly species (Diptera: Psychodidae), including major Leishmania vectors of the Mediterranean. J Econ Entomol 110(1):245–249. https://doi.org/10.1093/jee/tow269

Hudson RR, Slatkin M, Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. Genetics 132(2):583–589

Jacobson RL, et al. (2003) Outbreak of cutaneous leishmaniasis in northern Israel. J Infect Dis 188(7):1065-1073. https://doi.org/10.1086/378204

Kamhawi S (2006) Phlebotomine sand flies and Leishmania parasites: friends or foes? Trends Parasitol 22(9):439–445. https://doi.org/10.1016/j.pt.2006.06.012

Killick-Kendrick R (1987) Methods for the study of phlebotomine sandflies. In: Peters W, Killick-Kendrick R (eds) The leishmaniases in biology and medicine. Academic Press, London, pp 473–497

Killick-Kendrick R (1990) Phlebotomine vectors of the Leishmaniases: a review. Med Vet Entomol 4(1):1–24. https://doi.org/10.1111/j.1365-2915.1990.tb00255.x

Kounosu A, Murase K, Yoshida A, Maruyama H, Kikuchi T (2019) Improved 18S and 28S rDNA primer sets for NGS-based parasite detection. Sci Rep 9(1):15789. https://doi.org/10.1038/s41598-019-52422-z

Krayter L et al (2014) Multilocus microsatellite typing reveals a genetic relationship but, also, genetic differences between Indian strains of Leishmania tropica causing cutaneous leishmaniasis and those causing visceral leishmaniasis. Parasit Vectors 7:123. https://doi.org/10.1186/1756-3305-7-123

Kreutzer RD, Christensen HA (1980) Characterization of Leishmania spp. by isozyme electrophoresis. Am J Trop Med Hyg 29(2):199–208. https://doi.org/10.4269/ajtmh.1980.29.199

Kuhls K et al (2011) Comparative microsatellite typing of new world leishmania infantum reveals low heterogeneity among populations and its recent old world origin. PLoS Negl Trop Dis 5(6):e1155. https://doi.org/10.1371/journal.pntd.0001155

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35(6):1547–1549. https://doi.org/10.1093/molbev/msy096

Lane RP (1986) The sand flies of Egypt (Diptera: Phlebotominae). Bull Br Mus Nat Hist Entomol 52:1–35

Lewis D, Buttiker W (1982) Insects of Saudi Arabia: the taxonomy and distribution of Saudi Arabian Phlebotomus sandflies (Diptera: Psychodidae). Fauna of Saudi Arabia 4:353–383

Lewis DJ (1982) A taxonomic review of the genus Phlebotomus (Diptera: Psychodidae). Bulletin of the British museum (natural history) Entomology series 45:121–209

Mhaidi I et al (2018) Molecular identification of Leishmania infection in the most relevant sand fly species and in patient skin samples from a cutaneous leishmaniasis focus, in Morocco. PLoS Negl Trop Dis 12(3):e0006315. https://doi.org/10.1371/journal.pntd.0006315

Nasereddin A et al (2022) Concurrent molecular characterization of sand flies and Leishmania parasites by amplicon-based next-generation sequencing. Parasit Vectors 15(1):262. https://doi.org/10.1186/s13071-022-05388-3

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci U S A 70(12):3321–3323. https://doi.org/10.1073/pnas.70.12.3321

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Palestinian-Central-Bureau-of-Statistics (2020) Projected mid -year population for Tubas & Northern Governorate by Locality 2017-2021. http://www.pcbs.gov.ps/Portals/_Rainbow/Documents/TubasE.html

Perffl'ev PP (1966) Sandflies (family Phlebotomidae) [In Russian] fauna of the USSR new series, 93. vol 3 (2). Academy of sciences, Moscow, p 382

Rioux JA, Lanotte G, Maazoun R, Perello R, Pratlong F (1980) Leishmania infantum Nicolle, 1908, the agent of the autochthonous oriental sore. Apropos of the biochemical identification of 2 strains isolated in the eastern Pyrenees. C R Seances Acad Sci D 291(8):701–703

Rivas MA et al (2011) Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nat Genet 43(11):1066–1073. https://doi.org/10.1038/ng.952

Rozas J et al (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol 34(12):3299–3302. https://doi.org/10.1093/molbev/msx248

Sawalha S et al (2022) Effectiveness of insecticide thermal fogging in hyrax dens in the control of leishmaniasis vectors in rural Palestine: a prospective study. PLoS Negl Trop Dis 16(9):e0010628. https://doi.org/10.1371/journal.pntd.0010628

Sawalha SS, Ramlawi A, Sansur RM, Salem IM, Amr ZS (2017) Diversity, ecology, and seasonality of sand flies (Diptera: Psychodidae) of the Jenin District (Palestinian Territories). J Vector Ecol 42(1):120–129. https://doi.org/10.1111/jvec.12246

Sawalha SS, Shtayeh MS, Khanfar HM, Warburg A, Abdeen ZA (2003) Phlebotomine sand flies (Diptera: Psychodidae) of the Palestinian West Bank: potential vectors of leishmaniasis. J Med Entomol 40(3):321–328. https://doi.org/10.1603/0022-2585-40.3.321

Schnur LF, Zuckerman A, Greenblatt CL (1972) Leishmanial serotypes as distinguished by the gel diffusion of factors excreted in vitro and in vivo. Isr J Med Sci 8:932–942

Schonian G et al (2003) PCR diagnosis and characterization of Leishmania in local and imported clinical samples. Diagn Microbiol Infect Dis 47(1):349–358. https://doi.org/10.1016/s0732-8893(03)00093-2

Schonian G et al (2001) Genetic heterogeneity in the species Leishmania tropica revealed by different PCR-based methods. Trans R Soc Trop Med Hyg 95(2):217–224. https://doi.org/10.1016/s0035-9203(01)90173-7

Schuster SC (2008) Next-generation sequencing transforms today's biology. Nat Methods 5(1):16–18. https://doi.org/10.1038/nmeth1156

Schwenkenbecher JM, Frohlich C, Gehre F, Schnur LF, Schonian G (2004) Evolution and conservation of microsatellite markers for Leishmania tropica. Infect Genet Evol 4(2):99–105. https://doi.org/10.1016/j.meegid.2004.01.005

Schwenkenbecher JM et al (2006) Microsatellite analysis reveals genetic structure of Leishmania tropica. Int J Parasitol 36(2):237–246. https://doi.org/10.1016/j.ijpara.2005.09.010

Shendure J, Ji H (2008) Next-generation DNA sequencing. Nat Biotechnol 26(10):1135–1145. https://doi.org/10.1038/nbt1486

Shortt HE, Swaminath CS (1928) The method of feeding of Phlebotomus argipentes with relation to its bearing on the transmission of kala-azar. Indian J Med Res 15:827–836

Srivastava P, Singh T, Sundar S (2011) Genetic heterogeneity in clinical isolates of Leishmania donovani from India. J Clin Microbiol 49(10):3687–3690. https://doi.org/10.1128/JCM.00729-11

Svobodova M et al (2006) Distinct transmission cycles of Leishmania tropica in 2 adjacent foci, northern Israel. Emerg Infect Dis 12(12):1860–1868. https://doi.org/10.3201/eid1212.060497

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. Genetics 105(2):437–460

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10(3):512–526. https://doi.org/10.1093/oxfordjournals.molbev.a040023

Wright S (1951) The genetical structure of populations. Ann Eugen 15(4):323–354. https://doi.org/10.1111/j.1469-1809.1949.tb02451.x