

RESEARCH ARTICLE

Genomic analysis of *Leishmania turanica* strains from different regions of Central Asia

Tatiana S. Novozhilova¹, Daniil S. Chistyakov^{2,3}, Lyudmila V. Akhmadishina², Alexander N. Lukashev², Evgeny S. Gerasimov^{1*}, Vyacheslav Yurchenko^{4*}

1 Faculty of Biology, M.V. Lomonosov Moscow State University, Moscow, Russia, **2** Martsinovsky Institute of Medical Parasitology, Sechenov University, Moscow, Russia, **3** Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia, **4** Life Science Research Centre, Faculty of Science, University of Ostrava, Ostrava, Czech Republic

* jalgard@gmail.com (ESG); vyacheslav.yurchenko@osu.cz (VY)



Abstract

The evolution in *Leishmania* is governed by the opposite forces of clonality and sexual reproduction, with vicariance being an important factor. As such, *Leishmania* spp. populations may be monospecific or mixed. *Leishmania turanica* in Central Asia is a good model to compare these two types. In most areas, populations of *L. turanica* are mixed with *L. gerbilli* and *L. major*. Notably, co-infection with *L. turanica* in great gerbils helps *L. major* to withstand a break in the transmission cycle. Conversely, the populations of *L. turanica* in Mongolia are monospecific and geographically isolated. In this work, we compare genomes of several well-characterized strains of *L. turanica* originated from monospecific and mixed populations in Central Asia in order to shed light on genetic factors, which may drive evolution of these parasites in different settings. Our results illustrate that evolutionary differences between mixed and monospecific populations of *L. turanica* are not dramatic. On the level of large-scale genomic rearrangements, we confirmed that different genomic loci and different types of rearrangements may differentiate strains originated from mixed and monospecific populations, with genome translocations being the most prominent example. Our data suggests that *L. turanica* has a significantly higher level of chromosomal copy number variation between the strains compared to its sister species *L. major* with only one supernumerary chromosome. This suggests that *L. turanica* (in contrast to *L. major*) is in the active phase of evolutionary adaptation.

OPEN ACCESS

Citation: Novozhilova TS, Chistyakov DS, Akhmadishina LV, Lukashev AN, Gerasimov ES, Yurchenko V (2023) Genomic analysis of *Leishmania turanica* strains from different regions of Central Asia. *PLoS Negl Trop Dis* 17(3): e0011145. <https://doi.org/10.1371/journal.pntd.0011145>

Editor: Gabriele Schönián, Charité University Medicine Berlin, GERMANY

Received: October 13, 2022

Accepted: February 6, 2023

Published: March 6, 2023

Copyright: © 2023 Novozhilova et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The raw reads sequences were deposited to the NCBI database under BioProject accession number: PRJNA888552.

Funding: This work was funded by the Russian Science Foundation (19-15-00054) to V.Y., E.S.G., D.S.C., and L.V.A. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author summary

Leishmaniasis is one of the major neglected tropical diseases caused by members of the genus *Leishmania*. It can manifest in the symptoms ranging from rapidly self-healing ulcers in the cases of cutaneous leishmaniasis (CL) to systemic multiorgan failures in the cases of visceral leishmaniasis. The CL can be of anthroponotic (ACL, mainly caused by *L. tropica*) and zoonotic (ZCL, mainly caused by *L. major*) origin. In Central Asia, great gerbils are the main animal reservoirs for ZCL. These animals may be co-infected with *L. major*, *L. turanica*, and *L. gerbilli*. The evolution in *Leishmania* is driven by clonality and

Competing interests: The authors have declared that no competing interests exist.

genetic exchange, with geographical isolation playing an important role. In this sense, *Leishmania* spp. may either co-exist in the same area (mixed infections) or be geographically isolated without mixing-up with other species (monospecific infections). *Leishmania turanica* in Central Asia presents a good model to compare these two modes. Indeed, in most areas of great gerbils' distribution, populations of *Leishmania* spp. are mixed. Conversely, the populations of *L. turanica* in Mongolia are monospecific and geographically isolated. In this work, we compare genomes of several strains of *L. turanica* in order to shed light on genetic factors, which may drive evolution of these important parasites.

Introduction

Leishmaniasis is one of the major neglected tropical diseases in almost 100 countries. Its etiological agents are *Leishmania* spp. of the family Trypanosomatidae (Euglenozoa: Kinetoplastea) [1,2]. Well over 10 million people are infected worldwide with over one and a half million new cases being reported annually and 350 million people being at risk of infection [3]. Clinical manifestations of the disease range from often self-healing cutaneous ulcers in the cases of cutaneous leishmaniasis (CL) to systemic multiorgan failures in the cases of visceral leishmaniasis (VL) [4,5]. The CL can be of anthroponotic (ACL) and zoonotic (ZCL) origin mainly caused by *L. tropica* and *L. major*, respectively [6,7].

In Central Asia, great gerbils (*Rhombomys opimus*) serve as the main animal reservoirs for ZCL, although other animal species may also be involved [8–10]. These animals may be infected with three *Leishmania* spp.: i) *L. major* (pathogenic to humans, causing CL), ii) *L. turanica* (presumably, gerbil-restricted but causing mild self-healing symptoms in humans in experimental infections), and iii) *L. gerbilli* (strictly gerbil-restricted) [11–14]. In some instances, all three species may infect the same animal [15,16]. The main vectors in this area are sand flies *Phlebotomus papatasi*, *P. causicus*, *P. andrejevi*, *P. alexandri*, and *P. mongolensis* [10].

The human ZCL, which is caused by *L. major* in these areas, invariably happens in the context of *L. turanica* or, rarely, *L. gerbilli* infection. Importantly, co-infecting agents play an important functional role. For example, *L. turanica* was shown to aid *L. major* survival in gerbils during the 6–10 months gap in the transmission cycle of these flagellates [12]. The co-infections involving *L. turanica* and *L. major* in great gerbils are evolutionarily beneficial over the single-species infections. In the experimental settings, monospecific infections with *L. major* and *L. turanica* persisted for 7 and 15 months, respectively, while the co-infection of *L. major* and *L. turanica* lasted for over 25 months and, in contrast to self-healing monospecific infections, almost invariably resulted in a chronic disease [17].

The evolution in *Leishmania* appears to be driven by the opposite forces of clonality and sexual reproduction [18,19], with geographical isolation of populations (vicariance) playing an important role in any proposed model. Most of the intensely studied *Leishmania* spp. are not geographically isolated and often mixed [20–23]. *Leishmania turanica* in Central Asia presents a good and rather unique model to compare mixed and monospecific infections in natural populations. As mentioned above, in most areas of the *R. opimus* distribution, populations of *L. gerbilli*, *L. major*, and *L. turanica* are mixed with high prevalence of functionally important *L. major*–*L. turanica* co-infections [12,24]. Conversely, the populations of *L. turanica* in Mongolia are predominately monospecific (no documented *L. major* and very rare cases of *L. gerbilli* reported) and geographically isolated by Altai mountains [25,26].

The main goal of this work was to compare genomes of several previously characterized strains of *L. turanica* originated from mixed (with *L. major*) and monospecific geographically

isolated populations in Central Asia in order to understand whether their distinct ecology is reflected at the genomic level. In addition, we aimed to determine the degree of differences between *L. turanica* strains from various geographic locations and compare them to those of well-studied *L. major*.

Results and discussion

Genome sequencing and variant calling in *L. turanica* strains

In total, eleven *Leishmania turanica* strains isolated between 1983 and 1995 from mixed and monospecific populations in Central Asia (Table 1) were sequenced at high depth using paired-end Illumina sequencing platform. All the analyzed strains were isolated from the same host species, *R. opimus*. Three Mongolian strains (hereafter referred as ‘MN’) and five strains from Turkmenistan (hereafter referred as ‘TM’) represent the monospecific and mixed populations, respectively. We also sequenced one strain of *L. turanica* from Uzbekistan and one strain of the same species from Kazakhstan to use them as an outgroup for more accurate measurement of genetic variations at species level for the analyzed groups. One of the strains (87568) turned out to be a mix of *L. major* and *L. turanica* and was excluded from further analysis.

The MultiQC analysis for duplicate reads confirmed that all samples were sequenced to sufficient depth (average genome coverage is about 65×). Read mapping statistics and initial SNP (Single Nucleotide Polymorphism) calling statistics for sequenced samples are summarized in Table 2. Two strains, namely KD51 and BK7, possess at least two times less homozygous and more heterozygous SNPs against the reference strain LEM423 (MMEL/SU/1979/MEL) than other analyzed strains of the MN and TM groups. Notably, the number of short indels is very similar in all the strains under analysis.

We further filtered the SNPs and produced a set 81,551 of confident homozygous SNPs in genomic positions well-covered in all sequenced samples of *L. turanica*. In total, about 14,000 homozygous SNPs are unique for only one strain, with MNP1 being the most divergent (about 4,000 unique SNPs) of all. Interestingly, two other MN and all TM strains are less divergent than strains from Kazakhstan and Uzbekistan (Fig 1A). Analysis of the SNP persistence among strains revealed that about 17,000 SNPs are shared by five strains (Fig 1B). Further analyses showed that this group of 5 strains share the same origin—Turkmenistan.

We also assembled the unmapped reads of all the studied strains *de novo* in order to detect genes possibly missing in the reference assembly of LEM423. We revealed 220–240 genes or

Table 1. *Leishmania turanica* strains analyzed in this study.

Strain ID	Code WHO	Year of isolation	Host	Country, place (GPS coordinates) of isolation
9104	MRHO/TM/91/Marz-9104	1991	<i>R. opimus</i>	Turkmenistan, Teze-Yël (37°29'55"N 60°21'4"E)
9105	MRHO/TM/91/Marz-9105	1991	<i>R. opimus</i>	Turkmenistan, Teze-Yël (37°29'55"N 60°21'4"E)
91014	MRHO/TM/91/Marz-91014	1991	<i>R. opimus</i>	Turkmenistan, Teze-Yël (37°29'55"N 60°21'4"E)
9562	MRHO/TM/95/Marz-9562	1995	<i>R. opimus</i>	Turkmenistan, Purnuar (38°51'55"N 56°19'26"E)
9563	MRHO/TM/95/Marz-9563	1995	<i>R. opimus</i>	Turkmenistan, Mekan obasy (Molochniy) (37°48'14"N 58°50'34"E)
MNR1	MRHO/MN/83/Marz-MNR1	1983	<i>R. opimus</i>	Mongolia, Aikhingol (Jinst) (45°24'33" N 100°35'26" E)
MNR4	MRHO/MN/83/Marz-MNR4	1983	<i>R. opimus</i>	Mongolia, Züünbayan (44°30'8"N 110°03'12"E)
MNR14	MRHO/MN/84/Marz-MNR14	1984	<i>R. opimus</i>	Mongolia, Züünbayan (44°30'8"N 110°03'12"E)
KD51	MRHO/UZ/83/Marz-KD51	1983	<i>R. opimus</i>	Uzbekistan, Mubarek (39°14'56"N 65°08'25"E)
BK7	MRHO/KZ/87/MARZ-BK7	1987	<i>R. opimus</i>	Kazakhstan, Bayrkum (42°07'04"N 68°08'43"E)
87568	MRHO/UZ/87/Marz-KD87568	1987	<i>R. opimus</i>	Uzbekistan, Qarshi (38°51'48"N 65°47'52"E)

<https://doi.org/10.1371/journal.pntd.0011145.t001>

Table 2. General mapping and SNP calling statistics for analyzed strains of *L. turanica*.

Strain	Mapped, %	SNPs	Insertions	Deletions	Ts/Tv	Heterozygous SNPs	Homozygous SNPs
9562	94.69	67,557	12,975	15,953	2.13	16,433	51,124
9563	93.3	67,967	13,151	15,996	2.11	16,833	51,134
9105	95.46	68,472	13,289	16,194	2.12	17,811	50,661
9104	96.54	68,557	13,038	16,096	2.12	19,523	49,034
91014	89.81	68,312	13,302	16,107	2.12	17,537	50,775
MNR1	94.87	75,769	14,030	17,391	2.11	37,946	37,823
MNR4	95.6	73,923	13,714	17,201	2.08	32,313	41,610
MNR14	96.43	73,528	13,822	17,312	2.1	31,996	41,532
BK7	96.17	83,694	14,483	17,263	2.07	58,002	25,692
KD51	95.56	83,641	14,699	17,653	2.13	57,591	26,050

Ts/Tv is transition/transversion ratio.

<https://doi.org/10.1371/journal.pntd.0011145.t002>

gene fragments, which share common annotations, in each of the strains. This implies that the reference assembly of LEM423 [27] lacks a small portion of *L. turanica* genes. Among these genes, the group of surface membrane glycoprotein GP46/M-2-like genes is always present on separate small scaffolds with coverage greater than average and the number of these genes vary between strains (from 2 to 10 copies). This suggests the presence of tandemly multiplied gene copies typical for trypanosomatid genomes and explains difficulties in assembling and annotating this locus.

Phylogenetic inferences

For phylogenetic inference, we added the reference *L. turanica* LEM423 [27] and an outgroup species *L. major* Friedlin (MHOM/IL/81/Friedlin) [28]. In order to do that, we first performed orthologous genes clustering and discovered single copy orthologs between two *Leishmania* spp. Out of all genes, which included SNPs from confident set in *L. turanica* and had a single-copy ortholog determined in *L. major*, we randomly selected 2,000 genes to build a maximum-likelihood tree (Fig 2A). In agreement with previously reported data on other

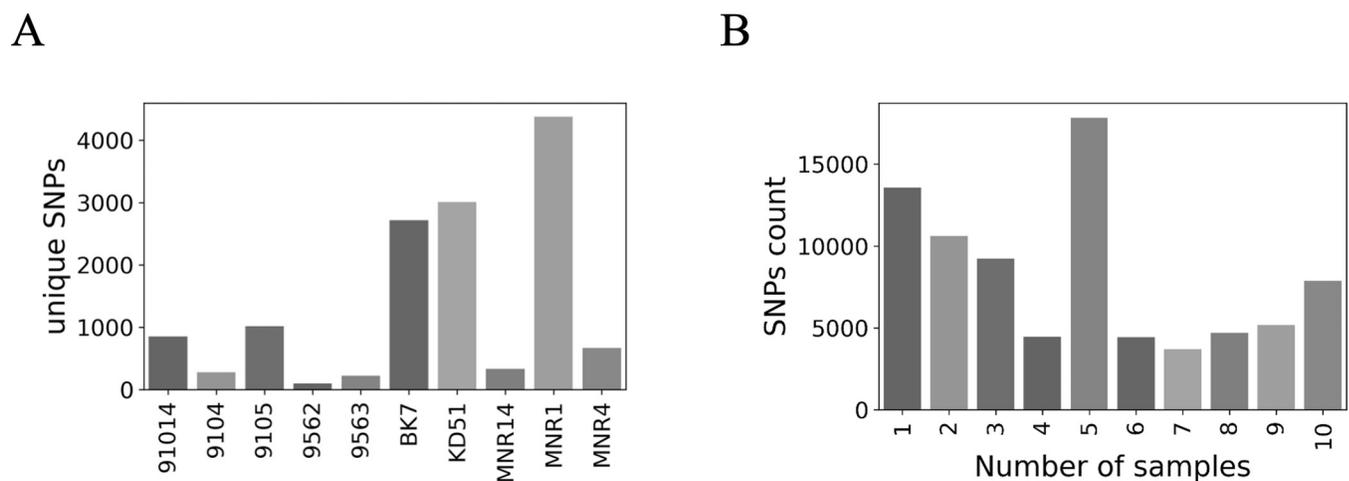


Fig 1. Single nucleotide polymorphisms analysis of *Leishmania turanica* strains. (A) Strain-specific SNPs. (B) Persistence of SNP in samples. The X axis shows the number of strains sharing SNPs. The shades of gray are only illustrative.

<https://doi.org/10.1371/journal.pntd.0011145.g001>

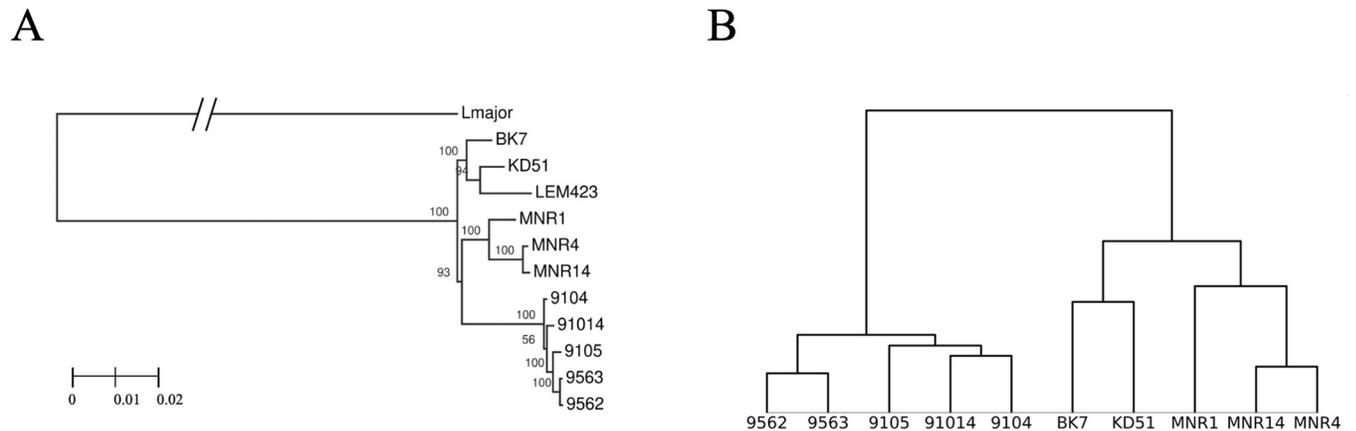


Fig 2. Phylogenomic analyses of *L. turanica* strains. (A) Maximum-likelihood phylogenomic tree of the investigated strains. *Leishmania major* is used as an outgroup. Bootstrap supports were determined using 1,000 replications. Branch length leading to *L. major* is manually scaled down 15× times as inter-species distance is much higher than that for strains of the same species. The scale bar represents substitutions per site. (B) Hierarchical clustering distance matrix output. All 81,551 confident homozygous SNPs were used to calculate the distance matrix.

<https://doi.org/10.1371/journal.pntd.0011145.g002>

Trypanosomatidae spp. [29], the inter-species distance between two *Leishmania* spp. is ~20× greater than average inter-strain distance. Phylogenetic analysis confirms that MN and TM strains are grouped into separate clades. Of note, BK7, KD51, and LEM423 also grouped together (hereafter, called LEM strains), making strains originated from mixed infections paraphyletic. The external branch lengths are in concert with unique SNPs counts (Fig 1A): MNR1, KD51, and BK7 strains are furthest from their last common ancestors.

As the phylogenetic tree in Fig 2A was constructed using only a subset of SNPs, we also performed a hierarchical clustering based on all confident SNPs (Fig 2B). The results of this analysis agree with the ML-tree. Five TM strains are placed in the more secluded clade and appear even closer to each other if the distance is measured only by homozygous SNPs.

Homozygous SNPs and codon usage in *Leishmania* spp

Next, we investigated the impact of homozygous SNPs on the codon usage in different strains of *Leishmania* sp. In all investigated strains, the general characteristics of SNPs are similar: i) about 60–65% of SNPs are intergenic; ii) on average, 4,200 genes contain at least one SNP; iii) the mean number of SNPs per gene is 2.0; and iv) the average synonymous to non-synonymous SNPs ratio is 1.2.

Previously, we demonstrated that genomes of two *Crithidia* spp. greatly varied between strains collected in different geographic locations in terms of SNPs, yet their codon usage remained strictly species-specific [29]. We performed similar analysis for *Leishmania* spp. and demonstrated that all *L. turanica* strains belong to one species (the lines representing LEM423 and two other strains are fused together indicating zero overall difference in codon usage) (Fig 3). Interestingly, *L. gerbilli* and *L. major*, which are considered phylogenetically close [30], have similar codon usage pattern, but still detectable differences (for example, triplets GCC, GCG, AGC, or TCT).

Chromosome copy number variation and genome-wide rearrangements in *L. turanica* strains

Chromosome copy number variation (CCNV) analyzed with GIP revealed that in all analyzed strains, the chromosome 31 has a ploidy of 4, which is typical for other *Leishmania* spp.

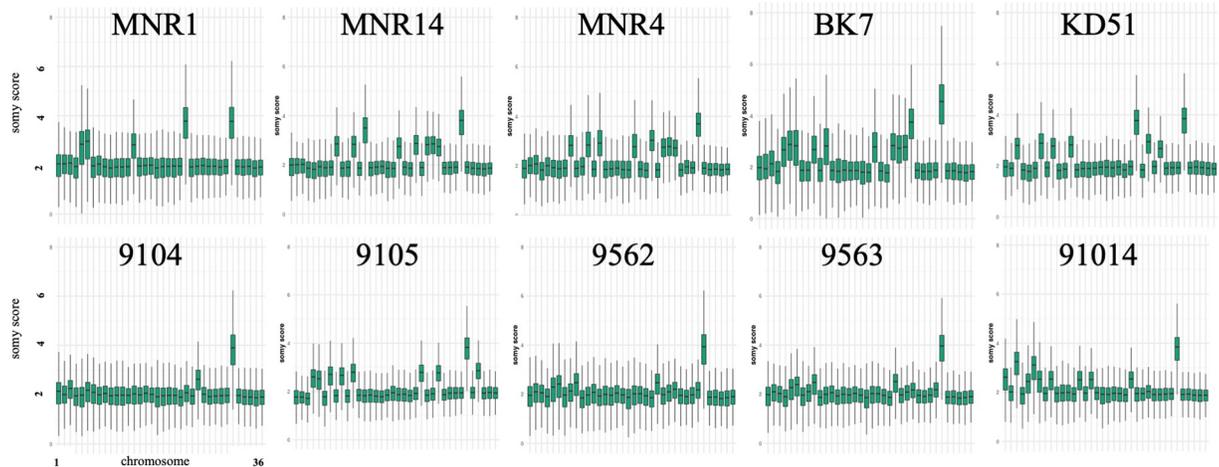


Fig 4. Chromosome copy number variations in sequenced strains of *L. turanica*.

<https://doi.org/10.1371/journal.pntd.0011145.g004>

translocations persist in all analyzed strains and involve repeat and telomeric/sub-telomeric regions of chromosomes, therefore, we excluded them from further analysis, as they might be caused by repetitive loci read multimapping artifacts. For other types of large-scale variations, we identified the events common to specific clades in Fig 2. These variations are presented in Fig 5 for representatives of the LEM (mixed), MN (monospecific), and TM (mixed) groups. We conclude that at the level of large-scale genomic variations these groups have different evolutionary patterns and different chromosomes appear to be involved in these events. For example, translocations (black lines) are more common for the strains from mixed infections of both LEM (9 translocations) and TM (35 translocations) groups than for their MN single-infection kin (2 translocations).

To understand the possible evolutionary impact of described genome variations, we determined, what genes are involved in these changes (S1 Table). For this, we analyzed sets of amplified genes (present in more than 2 copies relative to the diploid LEM423 genome) and genes with confident homozygous SNPs (Tables 2 and S1) in detail. Five hundred eighty-six, 345, and 249 genes were documented amplified for LEM, MN, and TM strains, respectively. Out of these, 212 genes (out of 626 in total) were amplified in all three groups, and large proportions of genes were amplified in two groups (Fig 6A). Similarly, we analyzed the genes that were significantly depleted with reads (gene deletions), and found very few such cases (less than 8 genes per strain). Therefore, we conclude that gene duplication events are prevalent in evolution of *Leishmania* genomes. Next, we compared sets of genes with confident homozygous SNPs and found that a significantly smaller proportion of them (230 out of 4,449 in total) is shared by the MN, LEM, and TM groups (Fig 6B). Out of these, only 2 genes were commonly shared by all three groups in categories amplified and confident homozygous SNPs-containing. The combination of the following two factors may explain the observed effect: i) the amplified copies remain under strong selection pressure and do not multiply SNPs, because each copy defines precise gene dosage [38,39], and/or ii) the amplification is recent.

Finally, we analyzed a set of positively selected genes in *L. turanica* strains (1,398 genes in total) and overlapped this dataset with that for genes with confident homozygous SNPs (Fig 6B). There was no clear prevalence of strain-specific selection acting on the group of genes in any strain, 91 genes out of 230 shared by all groups were under positive selection. We also conducted Gene Ontology enrichment studies for the genes under positive selection and found no obvious functional enrichment. To assess the question of selection on the branches leading to

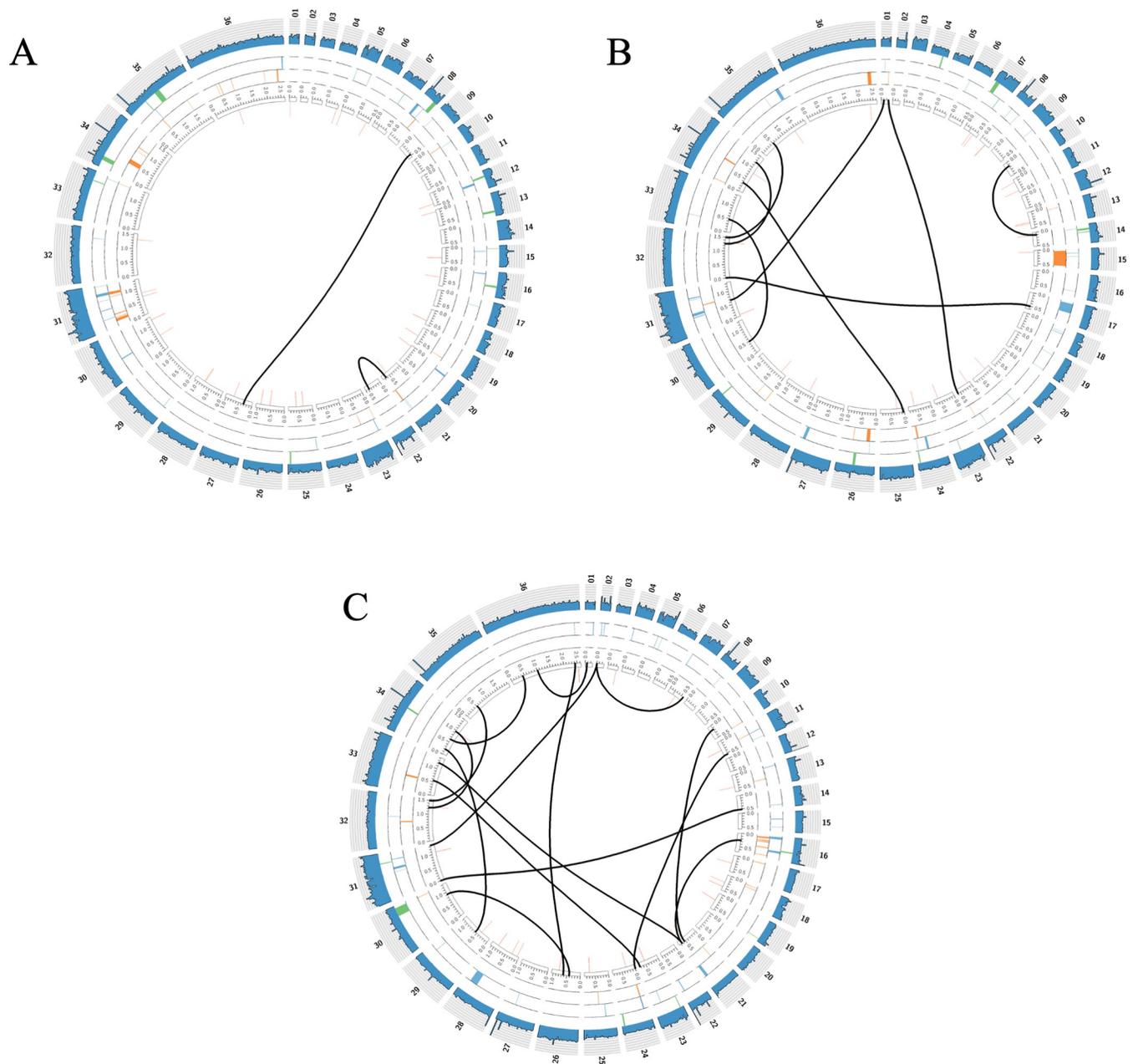


Fig 5. Whole-genome-scale structural variations specific for *L. turanica* strains. (A) Single infections MN strains (coverage track exemplified by MNR4). (B) Mixed infections LEM strains (coverage track exemplified by KD51). (C) Mixed infections TM strains (coverage track exemplified by 9105). Outer histograms (blue) show chromosome read coverage, inner tracks with highlights show the inversions (green), deletions (blue), duplications (orange), and insertions (red). Black lines show inter-chromosome translocations.

<https://doi.org/10.1371/journal.pntd.0011145.g005>

MN and TM clades, we used branch models from PAML and found 19 and 46 genes under positive selection on these branches, respectively. These sets have no overlap.

Conclusions

Taken together, our results illustrate that evolutionary differences between strains of *L. turanica* originated from mixed and monospecific infections are not dramatic in terms of codon

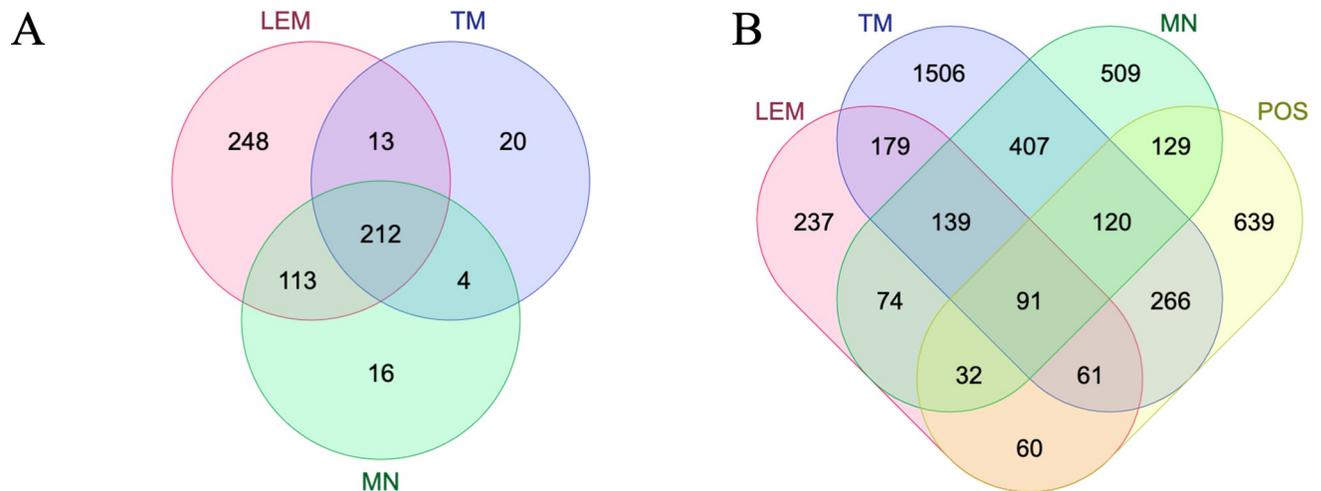


Fig 6. Venn diagrams showing the intersections of the gene sets under investigation. (A) Intersection between amplified (> 2n) genes for MN, LEM, and TM groups. (B) Intersection between sets of genes with confident homozygous SNPs and genes under positive selection (POS) for MN, LEM, and TM groups. Genes IDs are provided after assembly of *L. turanica* LEM423.

<https://doi.org/10.1371/journal.pntd.0011145.g006>

usage (Fig 3) or CCNV (Fig 4). Instead, we documented very targeted evolutionary changes, focused on groups of genes (S1 Table and Fig 6). Conversely, on the level of large-scale genomic rearrangements, we confirmed that different genomic loci and different types of rearrangements were implicated in groups of species under analysis (for example, strains from monospecific infections have fewer translocation events than their counterparts from mixed infections, Fig 5). We conclude that large-scale genomic rearrangements may precede “fine tuning” of the speciation process in *Leishmania turanica*.

Materials and methods

Sample collection, DNA preparation, and whole-genome sequencing

Strains of *Leishmania turanica* from the collection of Martsinovskiy Institute of Medical Parasitology, (Table 1) were cultivated as described previously [40,41]. These strains were isolated between 1983 and 1995 in Kazakhstan, Mongolia, Turkmenistan, and Uzbekistan from great gerbils (*Rhombomys opimus*). The strain identity was confirmed as in [42]. Total genomic DNA was isolated from 10 ml of the log-phase cultures using DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer’s instructions. Samples were sequenced using HiSeq 4000 platform (Illumina, San Diego, USA) in PE150 mode at SkolTech Genomics core facility (Moscow, Russia) to the depth of approximately 7 million read pairs per sample, giving mean *Leishmania* genome coverage of 65x. The obtained data were deposited to GenBank (BioProject PRJNA888552). One of the strains (87568) was a mix of *L. major* and *L. turanica* and was excluded from further analysis. In total, our dataset included 3 strains from Mongolia (labeled MN), 5 strains from Turkmenistan (labeled TM), and 3 strains (from Uzbekistan and Kazakhstan, 1 reference, 2 sequenced in the frame of this work) used as an outgroup.

Read mapping and processing

The paired-end Illumina reads of *L. turanica* strains were trimmed for quality and sequencing adaptors with Trimmomatic v. 0.36 [43] and quality controlled using FastQC v. 0.11.8 [44] and MultiQC v. 1.13 [45]. Each sample was mapped to the reference genome of *L. turanica*

LEM423 [27] using the BWA MEM v. 0.7.17 [46]. Read processing was done with SAMtools v. 1.9 [47] using ‘fixmate’, ‘view’, ‘sort’ and ‘markdup’ commands. Sorted and deduplicated bam files were used for all further processing.

De novo assembly of unmapped reads

Reads, not mapped on the nuclear genome, were collected with SAMtools using a custom Python script (read pair considered unmapped if any read of the pair was unmapped). Unmapped read pairs of each strain were assembled *de novo* with SPAdes v. 3.13.0 with default settings [48]. Genes on assembled scaffolds were predicted with Augustus v. 3.3.2 [49] with species model for *Leishmania tarentolae*. Predicted proteins were BLASTed against the local copy of UniProt database [50] using BLASTp.

Variant calling and processing

Initial variant calling was done with Freebayes v. 0.9.21 [51] with the following options “—read-indel-limit 1—read-mismatch-limit 3—read-snp-limit 3—min-alternate-fraction 0.05—min-base-quality 10—min-alternate-count 2—pooled-continuous -p 2”. Collected variants were filtered using custom Python script, filtering was done on all analyzed samples simultaneously. Only homozygous SNPs (single nucleotide polymorphisms) with coverage of at least 15 reads in all sequenced samples and with calling quality over 20 were gathered for further analyzes. These SNPs were inserted into the genome sequence to obtain strain-specific gene sequences using VCF-consensus tool from VCFtools v. 0.1.16 [52]. Allele frequency histogram and allele depth of each SNP were analyzed with a custom Python script and plotted using seaborn Python library.

Analysis of gene orthologs

Orthologous gene sets between *L. major* and *L. turanica* were determined by OrthoFinder v. 2.5.4 [53] on annotated protein-coding genes of *L. major* strain Friedlin and *L. turanica* strain LEM423. Single-copy orthologs were used to determine 1-to-1 correspondence between genes of two species. Sets of *L. turanica* genes not included into an orthologous group with any *L. major* gene were considered *L. turanica*-specific.

Phylogenetic analysis

Gene sequences of *L. major* Friedlin [31] were used for an outgroup. A set of confident homozygous SNPs determined in *L. turanica* strains and a set of 1-to-1 orthologs between two *Leishmania* spp. were used to randomly select 2,000 genes for multiple sequence alignment. Each selected gene included at least one SNP from the confident SNPs set. The sequence alignments were done with MAFFT v. 7.475 [54] and combined into a single multiple sequence alignment file with custom Python script. The tree was inferred in RAxML-NG v. 1.0.2 program [55] with 1,000 bootstrap replicates under GTR + G model and visualized with Toytrees Python library [56].

Gene variation and selection

Group-specific SNPs were extracted from the set of confident SNPs using a custom Python script. Genes with group-specific SNPs were extracted using the combination of BEDtools v. 2.30.0 [57] ‘intersect’ and ‘getfasta’ programs. General SNP effect statistics was collected using SnpEff v. 5.1 [58].

The site and branch tests (CODEML program from the PAML package v. 4.9 [59]) were used to test for selection. Positively-selected sites were chosen by comparing the likelihood of the M7/M8 models for each gene. Branch tests were performed for i) a branch leading to the common ancestor of MNR1, MNR4, and MNR14 as a foreground branch or ii) a branch leading to the common ancestor of 9104, 9105, 91014, 9562, and 9563 strains as a foreground branch to find the genes under positive selection during evolution of Mongolian and Turkmen strains, respectively. All other branches were considered a background branches, free-ratio, one-ratio, and two-ratios models were compared with the likelihood-ratio test, LRT.

Genome-level structural variations analysis

The structural variations on the whole-genome level, including translocations, long inversions, tandem duplications, long deletions, and locus copy number variations (including both gene and chromosome copy number variations) were detected with the GIP pipeline v. 1.0.9 [60]. The output of the DELLY v. 1.1.5 [61] was re-analyzed with a custom Python script to detect specific variations in MN and TM groups. Data were visualized using modified scripts from the GIP pipeline.

Supporting information

S1 Table. Datasets of amplified genes and genes with confident homozygous SNPs in analyzed groups of *L. turanica*.
(XLSX)

Acknowledgments

We thank Dr. M. Strelkova for assistance with the collection of *Leishmania* spp. at the Martshinovskiy Institute of Medical Parasitology. The computational resources were partially funded by the European Regional Funds (grant CZ.02.1.01/16_019/ 0000759).

Author Contributions

Conceptualization: Vyacheslav Yurchenko.

Data curation: Alexander N. Lukashev, Evgeny S. Gerasimov, Vyacheslav Yurchenko.

Formal analysis: Tatiana S. Novozhilova, Daniil S. Chistyakov, Lyudmila V. Akhmadishina, Alexander N. Lukashev, Evgeny S. Gerasimov, Vyacheslav Yurchenko.

Funding acquisition: Vyacheslav Yurchenko.

Investigation: Tatiana S. Novozhilova, Daniil S. Chistyakov, Lyudmila V. Akhmadishina, Evgeny S. Gerasimov, Vyacheslav Yurchenko.

Project administration: Evgeny S. Gerasimov, Vyacheslav Yurchenko.

Resources: Vyacheslav Yurchenko.

Supervision: Alexander N. Lukashev, Evgeny S. Gerasimov, Vyacheslav Yurchenko.

Validation: Tatiana S. Novozhilova, Evgeny S. Gerasimov, Vyacheslav Yurchenko.

Visualization: Tatiana S. Novozhilova, Evgeny S. Gerasimov.

Writing – original draft: Evgeny S. Gerasimov, Vyacheslav Yurchenko.

Writing – review & editing: Evgeny S. Gerasimov, Vyacheslav Yurchenko.

References

1. Kostygov AY, Karnkowska A, Votýpka J, Tashyreva D, Maciszewski K, Yurchenko V, et al. Euglenozoa: taxonomy, diversity and ecology, symbioses and viruses. *Open Biol.* 2021; 11: 200407. <https://doi.org/10.1098/rsob.200407> PMID: 33715388
2. Stuart K, Brun R, Croft S, Fairlamb A, Gürtler RE, McKerrow J, et al. Kinetoplastids: related protozoan pathogens, different diseases. *J Clin Invest.* 2008; 118: 1301–1310. <https://doi.org/10.1172/JCI33945> PMID: 18382742
3. WHO (2022) Leishmaniasis. <https://www.who.int/en/news-room/fact-sheets/detail/leishmaniasis>.
4. Bruschi F, Gradoni L (2018) The leishmaniasis: old neglected tropical diseases. Cham, Switzerland: Springer. 245 pp. p.
5. Mann S, Frasca K, Scherrer S, Henao-Martinez AF, Newman S, Ramanan P, et al. A review of leishmaniasis: current knowledge and future directions. *Curr Trop Med Rep.* 2021; 8: 121–132. <https://doi.org/10.1007/s40475-021-00232-7> PMID: 33747716
6. Ghatee MA, Taylor WR, Karamian M The geographical distribution of cutaneous leishmaniasis causative agents in Iran and its neighboring countries, a review. *Front Public Health.* 2020; 8: 11. <https://doi.org/10.3389/fpubh.2020.00011> PMID: 32133334
7. Akilov OE, Khachemoune A, Hasan T Clinical manifestations and classification of Old World cutaneous leishmaniasis. *Int J Dermatol.* 2007; 46: 132–142. <https://doi.org/10.1111/j.1365-4632.2007.03154.x> PMID: 17269962
8. Elfari M, Schnur LF, Strelkova MV, Eisenberger CL, Jacobson RL, Greenblatt CL, et al. Genetic and biological diversity among populations of *Leishmania major* from Central Asia, the Middle East and Africa. *Microbes Infect.* 2005; 7: 93–103.
9. Akhavan AA, Yaghoobi-Ershadi MR, Khamesipour A, Mirhendi H, Alimohammadian MH, Rassi Y, et al. Dynamics of *Leishmania* infection rates in *Rhombomys opimus* (Rodentia: Gerbillinae) population of an endemic focus of zoonotic cutaneous leishmaniasis in Iran. *Bull Soc Pathol Exot.* 2010; 103: 84–89.
10. Yurchenko V, Chistyakov DS, Akhmadishina LV, Lukashev AN, Sádlová J, Strelkova MV Revisiting epidemiology of leishmaniasis in Central Asia: lessons learnt. *Parasitology.* 2023; 150(2): 129–136. <https://doi.org/10.1017/S0031182022001640> PMID: 36453145
11. Akhavan AA, Mirhendi H, Khamesipour A, Alimohammadian MH, Rassi Y, Bates P, et al. *Leishmania* species: detection and identification by nested PCR assay from skin samples of rodent reservoirs. *Exp Parasitol.* 2010; 126: 552–556.
12. Strelkova MV, Eliseev LN, Ponirovsky EN, Dergacheva TI, Annacharyeva DK, Erokhin PI, et al. Mixed leishmanial infections in *Rhombomys opimus*: a key to the persistence of *Leishmania major* from one transmission season to the next. *Ann Trop Med Parasitol.* 2001; 95: 811–819.
13. Strelkova MV, Shurkhal AV, Kellina OI, Eliseev LN, Evans DA, Peters W, et al. A new species of *Leishmania* isolated from the great gerbil *Rhombomys opimus*. *Parasitology.* 1990; 101: 327–335.
14. Guan LR, Yang YQ, Qu JQ, Shen WX Discovery and study of *Leishmania turanica* for the first time in China. *Bull World Health Organ.* 1995; 73: 667–672.
15. Strelkova MV [The isoenzyme identification and pathogenic characteristics of clones of *Leishmania major*, *L. sp. nov.* and *L. gerbilli*]. *Med Parazitol (Mosk).* 1990: 9–13. (in Russian).
16. Strelkova MV, Shurkhal AV, Eliseev LN, Kellina OI, Rakitskaia TA, Zviagintseva TV, et al. [The isoenzyme identification and pathogenic characteristics of the *Leishmania* isolated in natural foci of cutaneous leishmaniasis in the USSR]. *Med Parazitol (Mosk).* 1990: 43–48. (in Russian).
17. Strelkova MV [Susceptibility to and the characteristics of the course of experimental leishmaniasis in different species of mammals infected with *Leishmania major*, *L. turanica* and *L. gerbilli*]. *Med Parazitol (Mosk).* 1991: 35–39. (in Russian).
18. Tibayrenc M, Ayala FJ *Leishmania* and the model of predominant clonal evolution. *Microorganisms.* 2021; 9: 2409.
19. Rougeron V, De Meeûs T, Bañuls AL A primer for *Leishmania* population genetic studies. *Trends Parasitol.* 2015; 31: 52–59.
20. Mendoza-Roldan JA, Zatelli A, Latrofa MS, Iatta R, Bezerra-Santos MA, Annoscia G, et al. *Leishmania (Sauroleishmania) tarentolae* isolation and sympatric occurrence with *Leishmania (Leishmania) infantum* in geckoes, dogs and sand flies. *PLoS Negl Trop Dis.* 2022; 16: e0010650.
21. Cortes S, Esteves C, Maurício I, Maia C, Cristovão JM, Miles M, et al. *In vitro* and *in vivo* behaviour of sympatric *Leishmania (V.) braziliensis*, *L. (V.) peruviana* and their hybrids. *Parasitology.* 2012; 139: 191–199.
22. Akhoundi M, Downing T, Votýpka J, Kuhls K, Lukeš J, Cannet A, et al. *Leishmania* infections: molecular targets and diagnosis. *Mol Aspects Med.* 2017; 57: 1–29.

23. Dujardin JC, Banuls AL, Dujardin JP, Arevalo J, Tibayrenc M, Le Ray D Comparison of chromosome and isoenzyme polymorphism in geographical populations of *Leishmania* (*Viannia*) *peruviana*. *Parasitology*. 1998; 117: 547–554. <https://doi.org/10.1017/s0031182098003357> PMID: 9881379
24. Strelkova MV, Shendrik AG, El Fari M, Schönian G [Ecology and the genetic structure of sympatric *Leishmania* species circulating in the intra-continental deserts of the south Palaeartic region]. *Med Parazitol (Mosk)*. 2003; 12–18. (in Russian).
25. Shurkhal AV, Strelkova MV, Passova OM, Rakitskaia TA, Podogas AV [Genetic characteristics of *Leishmania* strains isolated from gerbils in the Mongolian People's Republic]. *Med Parazitol (Mosk)*. 1985; 38–44. (in Russian).
26. Neronov VM, Strelkova MV, Shurkhal AA, Luschekina AA, Artemyev MM Natural focalities of zoonotic cutaneous leishmaniasis in the Mongolian People's Republic; results and objectives of integrated research. *Folia Parasitol*. 1987; 34: 1–9.
27. Warren WC, Akopyants NS, Dobson DE, Hertz-Fowler C, Lye LF, Myler PJ, et al. Genome assemblies across the diverse evolutionary spectrum of *Leishmania* protozoan parasites. *Microbiol Resour Announc*. 2021; 10: e0054521.
28. Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, et al. The genome of the kinetoplastid parasite, *Leishmania major*. *Science*. 2005; 309: 436–442.
29. Gerasimov E, Zemp N, Schmid-Hempel R, Schmid-Hempel P, Yurchenko V Genomic variation among strains of *Crithidia bombi* and *C. expoeki*. *mSphere*. 2019; 4: e00482–00419.
30. Butenko A, Kostygov AY, Sádlová J, Kleschenko Y, Bečvář T, Podešvová L, et al. Comparative genomics of *Leishmania* (*Mundinia*). *BMC Genomics*. 2019; 20: 726.
31. Rogers MB, Hilley JD, Dickens NJ, Wilkes J, Bates PA, Depledge DP, et al. Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*. *Genome Res*. 2011; 21: 2129–2142.
32. Dumetz F, Imamura H, Sanders M, Seblová V, Myšková J, Pescher P, et al. Modulation of aneuploidy in *Leishmania donovani* during adaptation to different *in vitro* and *in vivo* environments and its impact on gene expression. *mBio*. 2017; 8: e00599–00517.
33. Sterkers Y, Lachaud L, Bourgeois N, Crobu L, Bastien P, Pagès M Novel insights into genome plasticity in Eukaryotes: mosaic aneuploidy in *Leishmania*. *Mol Microbiol*. 2012; 86: 15–23.
34. Mannaert A, Downing T, Imamura H, Dujardin JC Adaptive mechanisms in pathogens: universal aneuploidy in *Leishmania*. *Trends Parasitol*. 2012; 28: 370–376.
35. Leprohon P, Légaré D, Raymond F, Madore E, Hardiman G, Corbeil J, et al. Gene expression modulation is associated with gene amplification, supernumerary chromosomes and chromosome loss in antimony-resistant *Leishmania infantum*. *Nucleic Acids Res*. 2009; 37: 1387–1399.
36. Ubeda JM, Légaré D, Raymond F, Ouameur AA, Boisvert S, Rigault P, et al. Modulation of gene expression in drug resistant *Leishmania* is associated with gene amplification, gene deletion and chromosome aneuploidy. *Genome Biol*. 2008; 9: R115.
37. Lachaud L, Bourgeois N, Kuk N, Morelle C, Crobu L, Merlin G, et al. Constitutive mosaic aneuploidy is a unique genetic feature widespread in the *Leishmania* genus. *Microbes Infect*. 2014; 16: 61–66.
38. Iantorno SA, Durrant C, Khan A, Sanders MJ, Beverley SM, Warren WC, et al. Gene expression in *Leishmania* is regulated predominantly by gene dosage. *mBio*. 2017; 8: e01393–01317.
39. Rogozin IB, Charyyeva A, Sidorenko IA, Babenko VN, Yurchenko V Frequent recombination events in *Leishmania donovani*: mining population data. *Pathogens*. 2020; 9: 572.
40. Kleschenko Y, Grybchuk D, Matveeva NS, Macedo DH, Ponirovsky EN, Lukashev AN, et al. Molecular characterization of *Leishmania RNA virus 2* in *Leishmania major* from Uzbekistan. *Genes*. 2019; 10: e830.
41. Kostygov AY, Grybchuk D, Kleschenko Y, Chistyakov DS, Lukashev AN, Gerasimov ES, et al. Analyses of *Leishmania*-LRV co-phylogenetic patterns and evolutionary variability of viral proteins. *Viruses*. 2021; 13: 2305.
42. Yurchenko V, Lukeš J, Xu X, Maslov DA An integrated morphological and molecular approach to a new species description in the Trypanosomatidae: the case of *Leptomonas podlipaevi* n. sp., a parasite of *Boisea rubrolineata* (Hemiptera: Rhopalidae). *J Eukaryot Microbiol*. 2006; 53: 103–111.
43. Bolger AM, Lohse M, Usadel B Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014; 30: 2114–2120.
44. Andrews S (2019) FastQC: a quality control tool for high throughput sequence data.
45. Ewels P, Magnusson M, Lundin S, Käller M MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016; 32: 3047–3048.

46. Li H, Durbin R Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25: 1754–1760.
47. Ramirez-Gonzalez RH, Bonnal R, Caccamo M, Maclean D Bio-SAMtools: Ruby bindings for SAMtools, a library for accessing BAM files containing high-throughput sequence alignments. *Source Code Biol Med*. 2012; 7: 6.
48. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012; 19: 455–477. <https://doi.org/10.1089/cmb.2012.0021> PMID: 22506599
49. Stanke M, Diekhans M, Baertsch R, Haussler D Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics*. 2008; 24: 637–644.
50. Consortium TU UniProt: the universal protein knowledgebase. *Nucleic Acids Res*. 2018; 46: 2699.
51. Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing.
52. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011; 27: 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330> PMID: 21653522
53. Emms DM, Kelly S OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019; 20: 238. <https://doi.org/10.1186/s13059-019-1832-y> PMID: 31727128
54. Katoh K, Standley DM MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013; 30: 772–780. <https://doi.org/10.1093/molbev/mst010> PMID: 23329690
55. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*. 2019; 35: 4453–4455.
56. Eaton DAR Toytree: a minimalist tree visualization and manipulation library for Python. *Methods Ecol Evol*. 2019; 11: 187–191.
57. Quinlan AR BEDTools: the swiss-army tool for genome feature analysis. *Curr Protoc Bioinformatics*. 2014; 47: 11.12.11–11.12.34.
58. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*. 2012; 6: 80–92.
59. Yang Z PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007; 24: 1586–1591.
60. Späth GF, Bussotti G GIP: an open-source computational pipeline for mapping genomic instability from protists to cancer cells. *Nucleic Acids Res*. 2022; 50: e36. <https://doi.org/10.1093/nar/gkab1237> PMID: 34928370
61. Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbelt JO DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*. 2012; 28: i333–i339.