



OPEN

Identification of a unique conserved region from a kinetoplastid genome of *Leishmania orientalis* (formerly named *Leishmania siamensis*) strain PCM2 in Thailand

Pornchai Anuntasomboon^{1,2}, Suradej Siripattanapipong³, Sasimanas Unajak⁴, Kiattawee Choowongkamon⁴, Richard Burchmore⁵, Saovanee Leelayoova⁶, Mathirut Mungthin⁶ & Teerasak E-kobon^{1,2}✉

Mitochondrial DNAs (mtDNAs) appear in almost all eukaryotic species and are useful molecular markers for phylogenetic studies and species identification. Kinetoplast DNAs (kDNAs) are structurally complex circular mtDNA networks in kinetoplastids, divided into maxicircles and minicircles. Despite several kDNAs of many *Leishmania* species being examined, the kDNAs of the new species, *Leishmania orientalis* (formerly named *Leishmania siamensis*) strain PCM2, have not been explored. This study aimed to investigate the maxicircle and minicircle DNAs of *L. orientalis* strain PCM2 using hybrid genome sequencing technologies and bioinformatic analyses. The kDNA sequences were isolated and assembled using the SPAdes hybrid assembler from the Illumina short-read and PacBio long-read data. Circular contigs of the maxicircle and minicircle DNAs were reconstructed and confirmed by BLASTn and rKOMICs programs. The kDNA genome was annotated by BLASTn before the genome comparison and phylogenetic analysis by progressiveMauve, MAFFT, and MEGA programs. The maxicircle of *L. orientalis* strain PCM2 (18,215 bp) showed 99.92% similarity and gene arrangement to *Leishmania enriettii* strain LEM3045 maxicircle with variation in the *12s rRNA* gene and divergent region. Phylogenetics of the whole sequence, coding regions, divergent regions, and *12s rRNA* gene also confirmed this relationship and subgenera separation. The identified 105 classes of minicircles (402–1177 bp) were clustered monophyletically and related to the *Leishmania donovani* minicircles. The kinetoplast maxicircle and minicircle DNAs of *L. orientalis* strain PCM2 contained a unique conserved region potentially useful for specific diagnosis of *L. orientalis* and further exploration of this parasite population genetics in Thailand and related regions.

The genus *Leishmania* is a member of the order Trypanosomatidae and one of the most important infectious parasitic groups^{1,2}. One characteristic of the Trypanosomatids is the presence of a vast DNA network known as kinetoplast DNAs (kDNAs) within the mitochondrion. This DNA network contains dozens of relaxed maxicircles (20–40 kb) encoded for mitochondrial rRNAs and protein subunits of the electron transport complexes, similar to other eukaryotes, and thousands of minicircles (0.5–2 kb) per species which encode most of the guide RNAs (gRNAs) involved in RNA editing (insertion or deletion of uridine) of some transcripts from the maxicircles^{3–7}. The RNA editing process of *Leishmania tarentolae* has been widely studied as the best model among several trypanosomatids^{8–10}. A single mitochondrion of *L. tarentolae* has approximately 5000–10,000 minicircles and 20–50 maxicircles¹¹. The *L. tarentolae* maxicircle contains genes encoding for two small mitochondrial rRNAs (9S and 12S rRNAs), several proteins responsible for the electron transport chain, including cytochrome b (Cyb),

¹Department of Genetics, Faculty of Science, Kasetsart University, Bangkok, Thailand. ²Omic Center for Agriculture, Bioresources, Food, and Health, Kasetsart University (OmiKU), Bangkok, Thailand. ³Department of Microbiology, Faculty of Science, Mahidol University, Bangkok, Thailand. ⁴Department of Biochemistry, Faculty of Science, Kasetsart University, Bangkok, Thailand. ⁵Glasgow Polyomics, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK. ⁶Department of Parasitology, Phramongkutklao College of Medicine, Bangkok, Thailand. ✉email: teerasak.e@ku.th

cytochrome oxidase subunits I, II, and III (COI, COII, and COIII), NADH dehydrogenase subunits 1, 2, 3, 4, 5, 7, 8, and 9 (ND subunits), a ribosomal protein (RpS12), and four uncharacterized genes i.e., MURF2, MURF5 or uS3m (a component of the mitoribosome), G3 and G4 G-rich genes (components of the respiratory complex I)^{12–14}. Some genes of the trypanosomes (COIII, Cytosine-rich Region 4 (CR4), ND9, and RpS12) were dual-coded, resulting in different proteins^{15–19}. The transcripts derived from the leishmanial genes required editing, i.e., the high- or pan-editing of ND8, ND9, G3, G4, ND3, and RpS12 genes before the mRNA translation^{20–22}. The large divergent region (DR) of the *L. tarentolae* maxicircle (approximately 12 kb in length) consists of numerous tandem repeats and is hypothesized to be a functionally unknown noncoding region^{7,8}. Interestingly, another study in *Leishmania major* displayed different variations in the DR region compared between the promastigote and amastigote stages²³.

Several previous studies uncovered the maxicircle DNA fragments from different *Leishmania* species. These findings included the discovery of the 8.4-kb maxicircle fragment from *L. major* strain MHOM/SU/73/5ASKH in 2008²⁴, an extensive RNA editing of genes encoded from the maxicircle of *Leishmania donovani* in 2009²⁵, and the identification of the 6.5-kb and 4.3-kb maxicircle fragments from *Leishmania braziliensis* strain MHOM/BR/75/M2904 in 2011²⁶. Advancement of next-generation sequencing technologies allowed quick and high-throughput determination of the nuclear genomic sequences of several *Leishmania* species and strains^{27–30}. Only a few mitochondrial genomes were characterized in *Leishmania*^{31–33}. In 2019, Camocho et al. used a BLAST search to extract sequence contigs specific to the maxicircles and employed a circularity test to confirm the minicircles of three *Leishmania* species (*L. major*, *Leishmania infantum*, and *L. braziliensis*) from the genomic sequence contigs³¹. Comparison of the maxicircles from these three species showed a conserved structure and gene order, while the number of minicircle classes varied: 97 classes for *L. major*, 49 classes for *L. infantum*, and three classes for *L. braziliensis*. Most identified minicircles shared conserved sequence blocks called CSBs. Another study identified 114 minicircle classes from *L. tarentolae* strain LEM125 and 24 classes from the UC strain using the PacBio sequencing platform¹³. A comparative genome study of seven *L. infantum* isolates in 2020 showed intra-specific conservation of the maxicircle and 59 distinct minicircle classes³³. These authors found that drug-resistant strains had unique minicircle class patterns and tended to harbour an increased copy number of certain minicircle classes. Structural variation within the maxicircle of *Leishmania panamensis* strains was also observed, for example, within the ND5 and *rpl12* genes, particularly in the DR region³². These genetic variations in the maxicircle and minicircle DNAs would benefit phylogenetic interpretation and understanding of the virulence and resistance of *Leishmania* species and strains.

Leishmania orientalis, formerly named *Leishmania siamensis*, is a new *Leishmania* species associated with cutaneous and visceral leishmaniasis observed in immunocompromised and immunocompetent patients, raising concerns about its resistance and pathogenesis owing to limited population genetic and genomic information³⁴. Recently, the nuclear genomes of two *L. orientalis* strains (PCM2 from the southern province and LSCM4 from the northern province of Thailand) were published, and comparative analysis showed intra-specific genomic similarity³⁵. However, the maxicircle and minicircle DNAs of *L. orientalis* remain unknown. Here, we aimed to identify the mitochondrial DNAs of *L. orientalis* strain PCM2 from next-generation and third-generation sequencing data and compare them with previously published maxicircle and minicircle DNAs of other *Leishmania* species. As a result, we presented the first maxicircle sequences and different minicircle classes of *L. orientalis* strain PCM2. These are useful for understanding parasite virulence, the RNA editing process, as well as further investigating the *L. orientalis* population genetics and classification.

Materials and methods

Culture of *Leishmania orientalis* strain PCM2

Leishmania orientalis strain PCM2 was maintained and provided by the Department of Parasitology, Phramongkutklao College of Medicine, Bangkok, Thailand. The promastigotes were grown at 26 °C in RPMI 1640-modified with 13.3 mM glutamine, 2.5 mM arginine, 0.3 mM cysteine, 1.7 mM glutamate, 62.1 mM proline, 0.6 mM ornithine, 3.8 mM glucose, 2.2 mM fructose, 5.1 mM malate, 2.8 mM α -ketoglutarate, 0.5 mM fumarate, 0.5 mM succinate, 25 mM HEPES, 50 μ g/mL gentamicin, 2X MEM vitamins (Gibco, US), and 20% heat-inactivated fetal bovine serum (HIFBS, Gibco, US).

Genomic DNA preparation

Genomic DNAs of *L. orientalis* were prepared from the late logarithmic phase of the promastigotes. The promastigote pellets (approximately 5×10^7 parasites/mL) were collected and incubated in the lysis buffer (10 mM Tris, 10 mM KCl, 10 mM MgCl₂, 0.5 M NaCl, 2 mM EDTA, 0.5% SDS, and 10 μ L of 10 mg/mL Proteinase K solution) at 60 °C for 1 h until the pellet disappeared. After incubation, the samples were cooled to room temperature before a 1:1 ratio of chloroform-isoamyl alcohol (24:1, v/v) was added and gently vortexed for 30 s. The samples were centrifuged for 15 min at 10,000 rpm, and 300 μ L of the aqueous phases were transferred to 1.5 mL tubes. Ten microliters of RNase (20 mg/mL) were added, and the samples were incubated at room temperature for 3 min. The chloroform-isoamyl alcohol step was repeated to remove RNase. 750 μ L absolute ethanol and 100 μ L of 7.5 M ammonium acetate were added, and the samples were gently mixed until the solution was homogenous before incubating at -70 °C overnight. The samples were centrifuged for 15 min at 10,000 rpm. After discarding the supernatant, the DNA pellet was washed twice with 70% ethanol. Residual ethanol was carefully removed, and the pellet was air-dried at room temperature. Next, 100–200 μ L Tris-EDTA buffer (pH 8.0) was added to the sample tubes and incubated at 37 °C, allowing the DNA pellet to dissolve. All samples were stored at -70 °C until further use. The DNA concentration was measured using the NanoDrop spectrophotometer (Thermo Fisher Scientific, MA, US), and the genomic integrity was analyzed by 1% agarose gel electrophoresis. The samples

with the following criteria: total yield of 100–150 ng/μL, A260/A280 ratio of 1.80 ± 0.05 , and A260/A230 ratio of 1.5 ± 0.05 , were proceeded through the genome sequencing steps.

Isolation of maxicircle and minicircle reads from whole genome sequence data

For short-read sequencing, a 101-bp pair-end read library was constructed for whole-genome sequencing using the Illumina HiSeq2000 platform (Illumina, US). For long-read sequencing, a 20-kb PacBio library was constructed using the PacBio standard library preparation methods (Pacific Biosciences, CA) and the SMRTbell Express Template Preparation Kit 1.0. The sequence raw reads were quality-checked by FastQC program version 0.11.9 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>)³⁶ and processed through the filtering and trimming steps with a cut-off value of 30 using BBDuk program in the BBtools pipeline (sourceforge.net/projects/bbmap/). The reads from both platforms were used to generate hybrid assemblies of the circular DNA using the SPAdes assembler version 3.15.3^{37,38}, and the parameters were set as `-pacbio -plasmid` to assemble only the reads that could be constructed into the circular chromosome. Contigs were verified by a BLASTn search against the NCBI database³⁹. The maxicircle and minicircle data, including the coding regions, divergent regions, and *12s rRNA* gene from the NCBI database, were downloaded using an in-house Python script. The contigs that shared sequence homology corresponding to the maxicircle of *Leishmania* were selected for further analyses. The rKOMICS package extracted minicircle sequences and polished the contigs from the SPAdes assembly output⁴⁰. This package performed the minicircle sequence cluster (MSC) analysis, which automated the assembly and circularization of mitochondrial genomes. The process of this toolkit was divided into three main steps: (1) assembling the contigs using MEGAHIT before extracting the minicircles using conserved sequences, (2) identifying the start and end of the minicircle for circularization using BLAST, and (3) correcting the direction of the CSB3 motifs for the polishing step. The polishing process generated a minicircle alignment based on the CSB3 sequence (GGG GTTG[G/A]TGTA) to rearrange the wrong sequence orientation in the minicircle contigs. After the assembly and identification, rKOMICS circularized the assembled minicircles using the BLASTn strategy to detect and remove duplicated sequences. Subsequently, rKOMICS polished the circularized minicircle DNAs to ensure the correct orientation of the CSB3 motifs, created minicircle alignments with the CSB1 motif at the starting position, and clustered the contigs based on the minimum percentage of identity using VSEARCH program⁴¹.

Annotation and visualization of the maxicircle and minicircle sequences

The extracted maxicircle contigs were annotated manually using the kDNA of *Leishmania enriettii* strain LEM3045 (GenBank accession number: BK010880.1) as a reference. The maxicircle genome was visualized and explored using SnapGene[®] software version 6.1.1 (<https://www.snapgene.com/>). The circularized and extended minicircle genomes were annotated using the BLASTn program³⁹. Gene annotation and orientation of the maxicircle sequence were compared between different *Leishmania* species and visualized using the progressiveMauve alignment program⁴². The *L. orientalis* maxicircle scaffold sequences were validated by realigning with long-read sequence data using the minimap2 program version 2.14^{43,44} before polishing with the short-read data using the Pilon program version 1.24⁴⁵. Quality and precision of the alignment results were assessed based on the read depth coverage using the samtools program version 1.9⁴⁶ and visualized by the SeqMonk program version 1.48.1 (<https://www.bioinformatics.babraham.ac.uk/projects/seqmonk/>).

Phylogenetic analysis of the whole sequences and specific regions of the *L. orientalis* maxicircle and minicircle DNAs

Phylogenetic trees were constructed using the whole sequence, coding regions, divergent regions, and *12s rRNA* gene of the maxicircle DNA and conserved sequence blocks (CSBs) regions from minicircles of *L. orientalis* strain PCM2 compared with those of other *Leishmania* species. The MAFFT program version 7⁴⁷ on the Unipro UGENE software suit version 44.0⁴⁸ was used to align the maxicircle DNAs using the DNA gap open penalty of 1.53, the DNA gap extension penalty of 0.123, and the omitted manual refinements. The pairwise comparison of the maxicircle sequences was conducted using the EMBOSS Water program with the following parameters: EDNAFULL for matrix, the gap penalty of 10.0, and the extended penalty of 0.5⁴⁹. Phylogenetic relationships were inferred using the Randomized Axelerated Maximum Likelihood (RaxML) program version 8⁵⁰ with the bootstrap values of 10,000 replicates for the complete maxicircle, coding regions, divergent regions, and the *12s rRNA* gene as well as those of minicircles. For the phylogenetic analysis, the FastTree program version 2.1⁵¹ and Interactive Tree Of Life (iTOL) version 5⁵² were used to construct a phylogenetic tree, and the tree labels were adjusted using the MEGA 11 program⁵³.

Results

This study reports the kinetoplast genome assembly of *L. orientalis* strain PCM2 from an integrative analysis of short-read and long-read whole-genome sequencing data. The maxicircle scaffolds of *L. orientalis* strain PCM2 were validated by the alignment of the PacBio sequencing reads back to the scaffold, and a high level of read coverage (99.52%) was achieved throughout the scaffold length. This output highlighted the robustness and fidelity of the mapping procedure, indicating a comprehensive selection of the maxicircle sequences mixed within the whole genomic read data (Fig. 1a). The de novo assembly of the maxicircle genome using SPAdes resulted in a circular DNA contig of 18,215 base pairs, similar to the complete kinetoplast of *L. enriettii* strain LEM3045 (GenBank accession number: BK010880.1) with a length of 17,999 base pairs, e-value of 0.0, 100% coverage, and 99.92% identity. The *L. enriettii* LEM3045 kinetoplast sequence was shorter because of the different lengths of the divergent region (Fig. 1b). Pairwise alignment of the maxicircle genomes of *L. orientalis* strain PCM2 and *L. enriettii* strain LEM3045 confirmed the high level of relatedness of the subgenus *Mundinia* members, with 98.6% identity, 98.6% similarity, and 1.2% gaps.



Figure 1. Validation and comparative analysis of the *L. orientalis* maxicircle DNAs with other *Leishmania* species. The long-read data were re-mapped to the assembled maxicircle DNA of *L. orientalis* strain PCM2 to validate the assembly confidence (a). The percentage of coverage is shown in the blue bar graph. The maxicircles of *L. enriettii* strain LEM3045, *L. orientalis* strain PCM2, and *L. tarentolae* were structurally compared using the progressiveMauve alignment program (b). Distinct coloured blocks denoted regions of RNA genes (red), protein-coding genes (white), and the divergent region (pink). The red bar graph represents the percentage of similarity. The maxicircle gene annotation and arrangement of these three *Leishmania* species included the protein-coding genes (purple), rRNA genes (light green), and divergent regions (blue–green) (c). Arrows indicate gene transcription directions.

To investigate the maxicircle structure, the progressiveMauve alignment program compared the annotated maxicircles of three *Leishmania* species, including *L. tarentolae* (20,992 bp), *L. enriettii* strain LEM3045 (17,999 bp), and *L. orientalis* strain PCM2 (18,215 bp), and showed a more similar genome structure between *L. orientalis* and *L. enriettii* (Fig. 1b). Manual annotation of the *L. orientalis* strain PCM2 maxicircle was conducted by multiple sequence alignment. The annotation revealed the conserved location and orientation of genes and regions closer to the annotated maxicircle of *L. enriettii* strain LEM3045 than those of *L. tarentolae* (Fig. 1c). Similar gene arrangements of the maxicircles of *L. orientalis* strain PCM2, *L. enriettii* strain LEM3045, and *L. tarentolae* were in the following order: 12s rRNA, 9s rRNA, ND8, ND9, MURF5, ND7, COIII, Cyb, ATPase 6, ND2, G3, ND1, COII, MURF2, COI, G4, ND4, ND3, Rps12, ND5, and DR regions (Fig. 1c). The DR region of *L. tarentolae* was much longer than those of the other two *Leishmania* species.

To investigate the evolution and similarity of *Leishmania* maxicircles, the complete maxicircle sequence of *L. orientalis* PCM2 was compared with that of other *Leishmania* species. The results showed that the maxicircle could be used to separate different subgenera of *Leishmania* (Fig. 2a). The close relationship between the maxicircles of *L. orientalis* strain PCM2 and *L. enriettii* strain LEM3045 was supported by the phylogenetic results of the coding and divergent regions (Figs. 2b and 2c). Besides the highly conserved structure of trypanosomatid maxicircles^{54,55}, the DR region, which consists of repeating sequences and displays variations across species⁵⁶ was also selected and examined. The phylogenetic tree of the *Leishmania* DR regions showed the separation of the subgenus *Mundinia* from other *Leishmania* subgenera, as well as the split of the DR regions of the subgenera *Leishmania* and *Viannia* into two clusters, similar to the multiple-group separation of *Trypanosoma* (Fig. 2c).

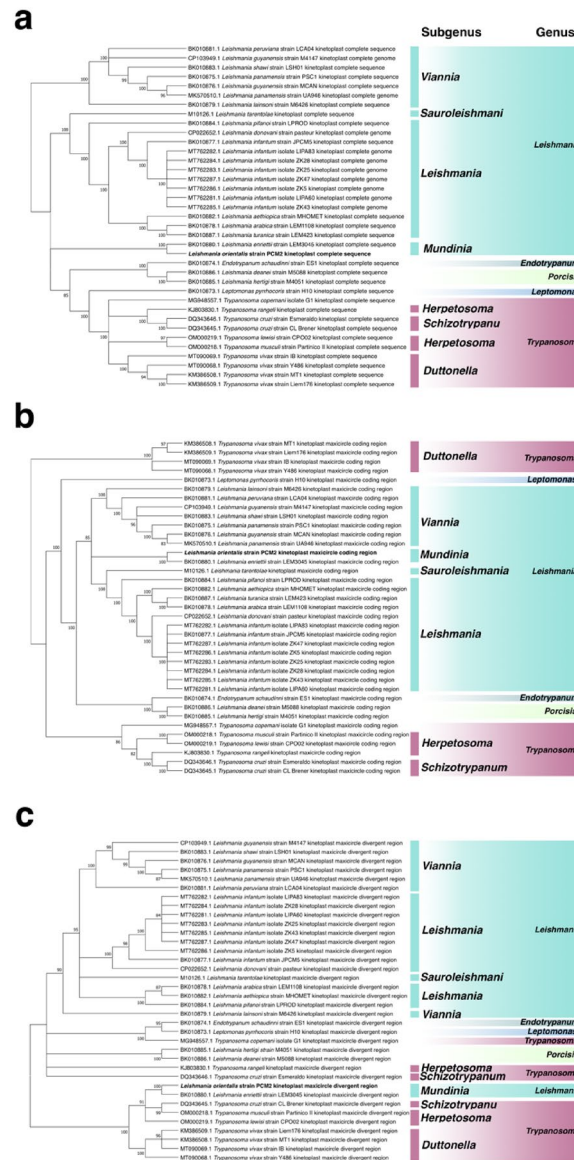


Figure 2. Phylogenetic tree of 22 maxicircles of *Leishmania* and outgroups (*Endotrypanum schaudinni*, *Trypanosoma*, and *Leptomonas pyrrochocis*) obtained from the NCBI database compared with those of *L. orientalis* strain PCM2 and *L. enriettii* strain LEM3045. The whole sequence of maxicircles (a), coding regions (b), and divergent region (c) were analyzed using the Randomized Axelerated Maximum Likelihood method with the GTR+I+G4, GTR+I+G4, and TPM3uf+I+G4 models, respectively. Bootstrap values were derived from 10,000 repetitions. The numbers lower than 80 were not presented. Highlighted colours showed different genera and sub-genera of *Leishmania*.

Therefore, the DR region may still be used to distinguish the subgenus *Mundinia* from other *Leishmania* sub-genera (Fig. 2c).

A 0.02% difference between *L. orientalis* strain PCM2 and *L. enriettii* strain LEM3045 was observed not only in the DR region but also in the *12s rRNA* gene region between positions 1270 and 1505 (Figs. 3a and 3b). Eleven SNPs and two deletions were detected when comparing the *12s rRNA* gene of these two *Leishmania* species. A phylogenetic tree of the *12s rRNA* genes from different *Leishmania* species was constructed. It showed that the variants within this gene could be used to distinguish *L. orientalis* strain PCM2 from *L. enriettii* strain LEM3045 and to discriminate subgenus *Mundinia* from other subgenera and outgroups (Fig. 3c).

One hundred and five minicircle classes of *L. orientalis* strain PCM2 were identified and compared. Three conserved sequence blocks (CSBs)⁵⁷ were detected in the conserved regions of these minicircles, representing the minicircle signature (Fig. 4a). The identification of CSB-1 (AgGGGCGTTC), CSB-2 (cCCCGTNC), and CSB-3 (GGGGTTGGTGTA) confirmed the accuracy of the minicircle contigs of *L. orientalis* strain PCM2 (Figs. 4b and 4c). Another conserved region at position 1888–1915 was identified by multiple sequence comparisons between these *L. orientalis* minicircles, indicating the within-species similarity of these minicircle DNAs (Fig. 4d).

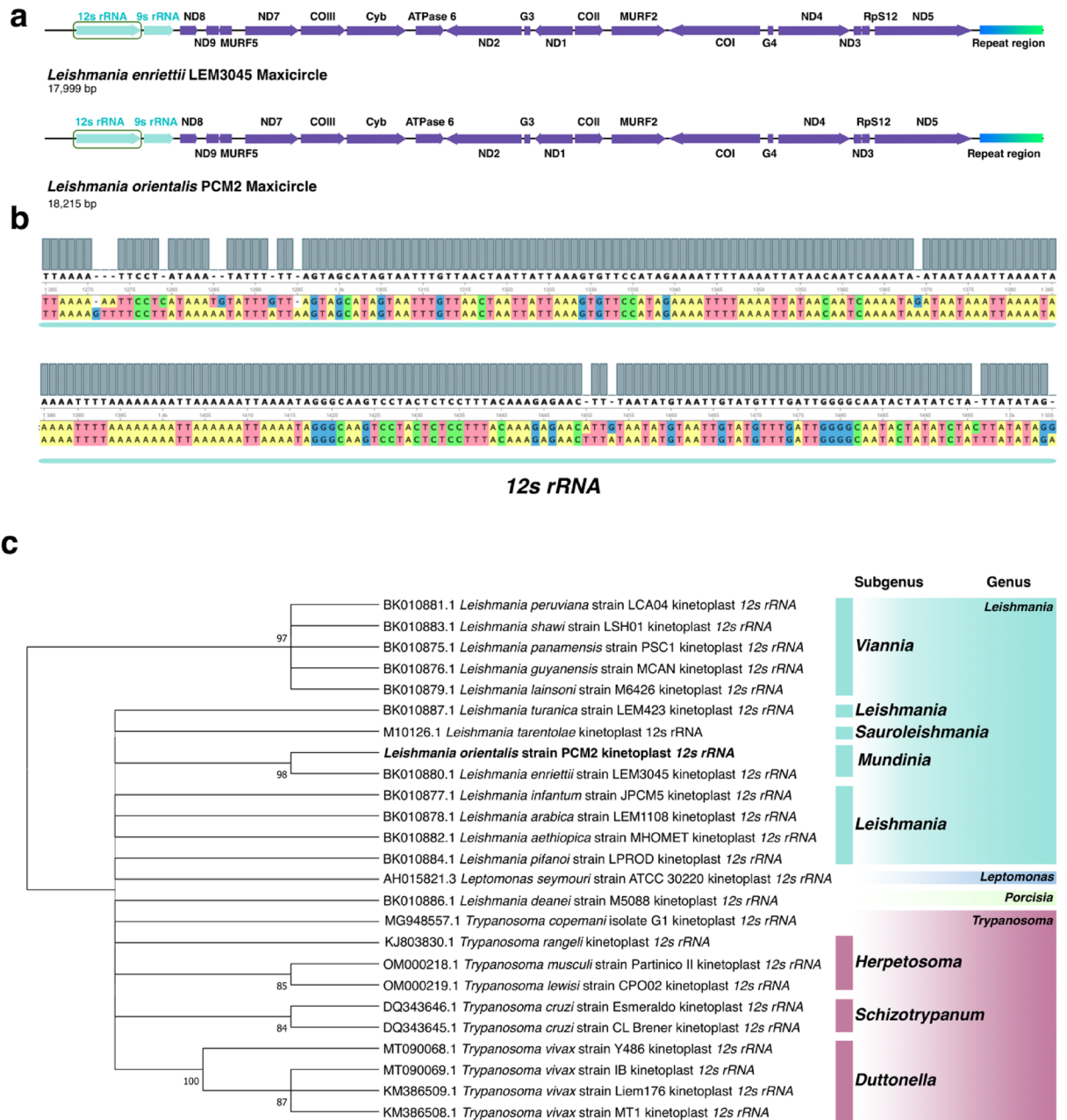


Figure 3. Comparative analysis of *12s rRNA* gene on the *Leishmania* maxicircle DNAs. (a) Position of the *12s rRNA* gene (marked by the green box) on the maxicircle DNA of *L. enriettii* strain LEM3045 and *L. orientalis* strain PCM2, and their nucleotide sequence variations were represented in (b). The ‘-’ in the row of coloured nucleotides indicated the indel position. (c) Phylogenetic tree of *12s rRNA* gene on the maxicircles of 14 *Leishmania* species with those of *Trypanosoma* and *Leptomonas seymouri* as outgroups obtained from the NCBI database. The Randomized Axelerated Maximum Likelihood method and the TrN + G4 model were used for the phylogenetic analysis. Bootstrap values were derived from 10,000 repetitions. The numbers lower than 80 were not presented. Highlighted colours showed different sub-genera of *Leishmania*.

Comparison of 105 minicircle sequences of *L. orientalis* strain PCM2 with 2111 minicircle DNAs of other *Leishmania* species from the NCBI database showed that most of the *L. orientalis* strain PCM2 minicircles were monophyletically arranged within the same clade as some of the *L. donovani* minicircles. In contrast, subgenera *Leishmania*, *Viannia*, and *Sauroleishmania* minicircles were split into multiple groups, indicating the genetic diversification of these minicircles within the same *Leishmania* species (Fig. 5).

Discussion

This study successfully identified the mitochondrial genome, including a maxicircle and several minicircles, of *L. orientalis* strain PCM2 to complement its nuclear genome, which was previously published³⁵. A new molecular-integrated taxonomic scheme for *Leishmania* suggested dividing *Leishmania* species into two major evolutionary

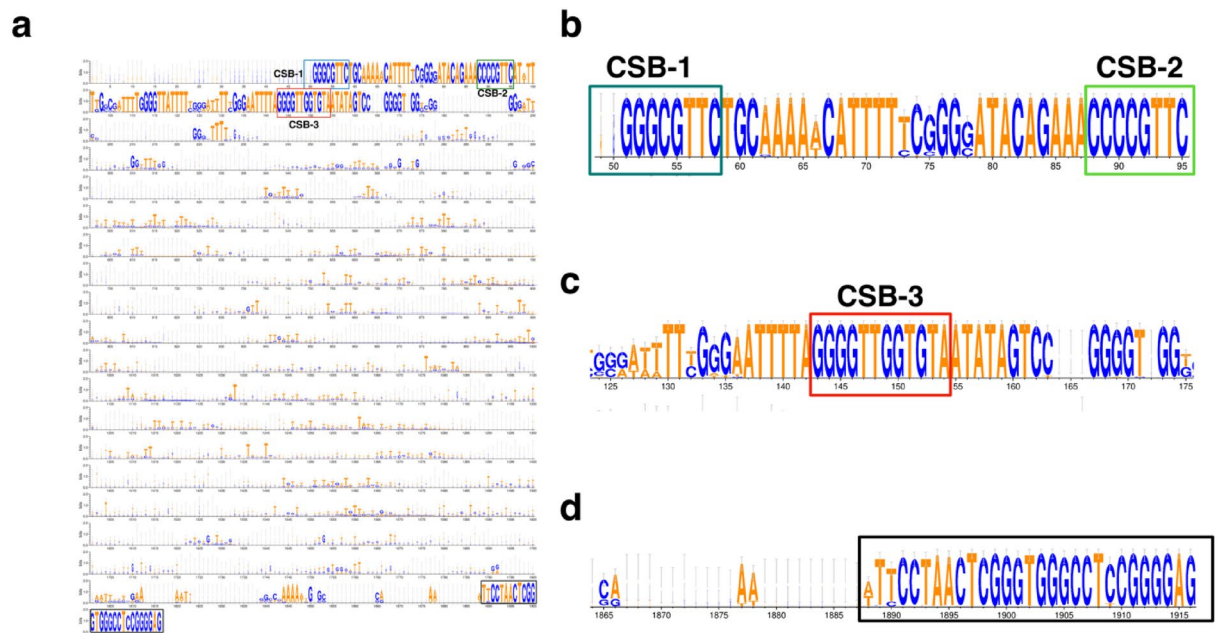


Figure 4. Nucleotide sequence alignment of minicircle DNAs found in *Leishmania orientalis* strain PCM2. The alphabet plot was used to visualize the full-length sequence alignment of 105 minicircle classes by the MAFFT and Weblogo3 programs (a). The first conserved region included three conserved motifs: CSB-1, CSB-2, and CSB-3 (b and c), and the second conserved region was unique to the minicircle of *L. orientalis* strain PCM2 (d).

lineages, *Euleishmania* and *Paraleishmania*⁵⁸. Since 1999, *L. martiniquensis* has been discovered in many CL and VL cases in Thailand and is genetically linked to *L. enriettii*^{59,60}. The subgenus *Mundinia* was established in 2018 and included several *Leishmania* species responsible for human and animal illnesses, *L. (Mundinia) enriettii*, *L. (Mundinia) martiniquensis*, formerly assigned to the subgenus *Leishmania*, and *L. (Mundinia) orientalis*, formerly named *L. siamensis*, a recently discovered species responsible for CL in Thailand^{59,61}. Our previous study illustrated the genomic differences between the two strains of *L. orientalis* in Thailand³⁵. The findings raised several questions regarding the genetics and population diversity of *L. orientalis* in Thailand and neighbouring countries. The hybrid assembly of the short- and long-read sequence data allowed the successful discovery of a complete maxicircle strand, which was similar to that of *L. enriettii* in terms of sequence and structural arrangement (Fig. 1), and unique classes of 105 minicircles that shared conserved motifs with those of other *Leishmania* species (Figs. 4 and 5).

As the maxicircle of *Leishmania* was previously reported to be useful for inferring phylogenetic relationships⁶¹, different formats of maxicircle DNAs, i.e. complete sequence, coding regions, divergent regions, and specific genes (*12s rRNA*) were explored in this study. Despite the high similarity (99.98%) between the maxicircles of *L. orientalis* strain PCM2 and *L. enriettii* strain LEM3045, subtle differences were exclusively observed in the *12s rRNA* gene and the highly variable or divergent regions. Phylogenetic analysis of these sequences consistently confirmed the close genetic relationship between *L. orientalis* and *L. enriettii* within the subgenus *Mundinia*, which was nicely separated from other subgenera (Figs. 2 and 3), supporting the proposed usage of the maxicircle for evolutionary examination within the population or the same subgenus of *Leishmania*.

The divergent region is one of the problematic parts of maxicircle assembly because of the repetitive sequences in this region, and its function remains to be investigated^{8,62,63}, unlike the coding region (CR), which contains fewer repeats. This study employed a hybrid assembly approach that combined short- and long-read sequencing platforms to enhance the coverage of the divergent region with minimized errors compared to short-read data alone, which was not long enough to cover the repeat area. Interestingly, the observed variation within this divergent region was also closely shared among members of the subgenus *Mundinia* (Fig. 2c), consistent with a previous study on the divergent region in Trypanosomatidae, which could be used to distinguish between different *Leishmania* species⁵⁶. This finding was supported by a study by Flegontov et al., who identified conserved repeat patterns in the divergent region of six *Leishmania* species, including *L. tarentolae*, *L. amazonensis*, *L. mexicana*, *L. major*, *L. turanica*, and *L. chagasi*, using PCR amplification and Sanger sequencing^{7,23}. Therefore, the uniqueness of the divergent region of *L. orientalis* maxicircle could benefit further investigation of *L. orientalis* population genetics and diversity in this Southeast Asian area.

In contrast, subtle variations were detected only in the coding region of the *12s rRNA* gene of the maxicircles of *L. orientalis* and *L. enriettii* (Fig. 3). This gene is a mitochondrial RNA gene that plays a role in the translation process and is recognized as a highly conserved gene in the Trypanosomatidae family⁵⁶. The *12s rRNA* variation was also observed in other *Leishmania* and trypanosome species, suggesting that this gene variation was perhaps the first event that occurred during the divergence of these two members in the subgenus *Mundinia*. The trypanosomal mitochondrial rRNAs, including 12S and 9S rRNAs, are substantially smaller than the eubacterial 23S

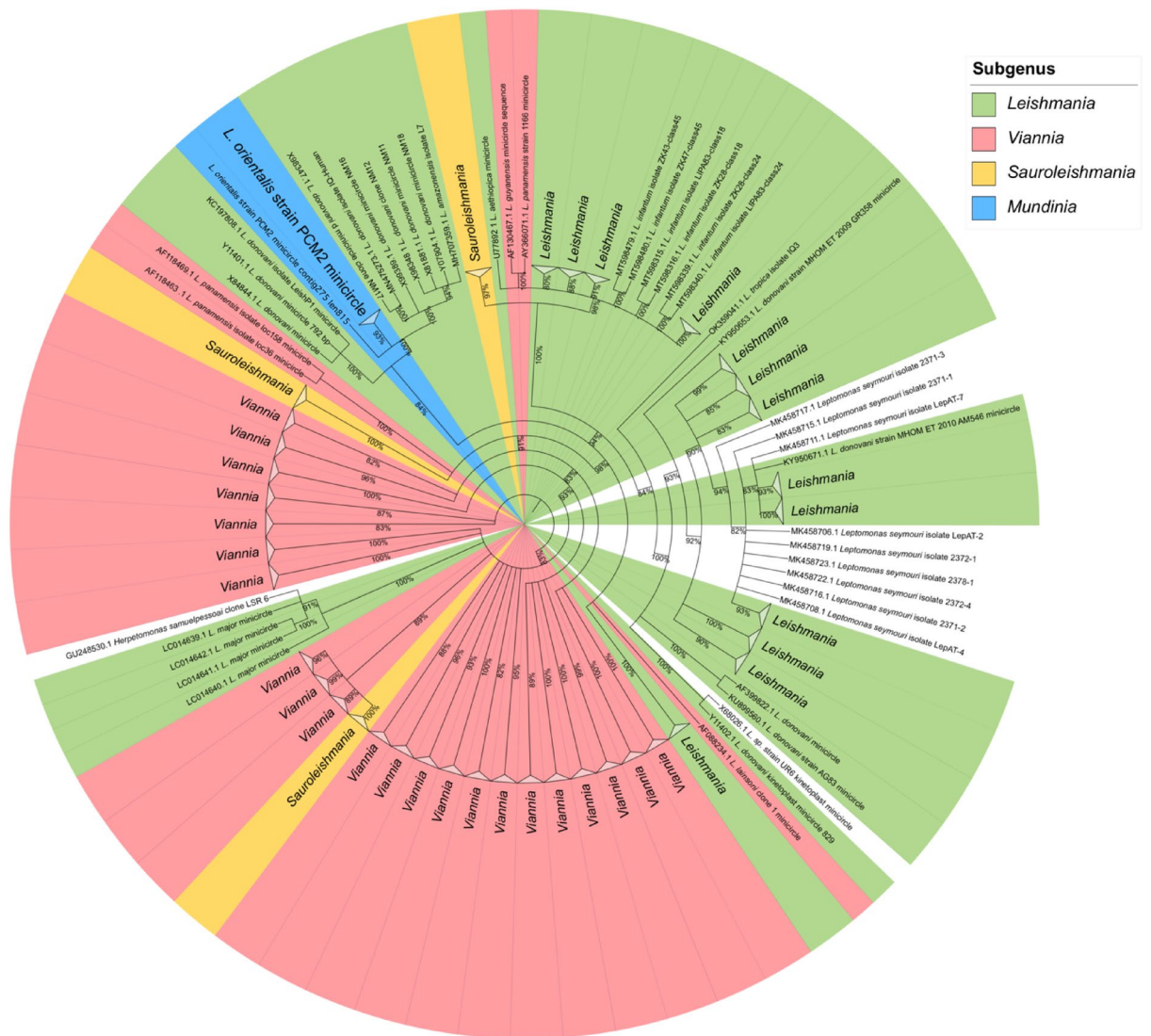


Figure 5. The Randomized Axelerated Maximum Likelihood (RaxML) phylogenetic tree of 105 classes of minicircle DNAs of *Leishmania orientalis* strain PCM2 and 2111 sequences from other *Leishmania* species. The minicircles of *Leptomonas seymouri* and *Herpetomonas samuelpessoai* were defined as outgroups. The local bootstrap support values showed at the branch point, and some branches of the *L. orientalis* strain PCM2 minicircles were collapsed for clear visualization. The concealed clades were represented as triangles at the end of branches. Sequences with values less than 80 were not displayed. Different sub-genera of *Leishmania* were highlighted using distinct colours.

and 16S rRNAs and also smaller than the mammalian mitochondrial 16S and 12S rRNAs due to the missing stem-loop structure in the secondary structure^{64,65}. This variation may affect rRNA stability, secondary structure, and translation process⁶⁶. The variable positions of the *12s rRNA* gene were thus helpful for distinguishing between *L. orientalis* and *L. enriettii* and for further improvement of species-specific diagnostic methods.

The second part of the *Leishmania* kDNA network contains several thousands of minicircles ranging from 0.5 to 2 kbp, which transcribe guide RNAs required for the RNA editing process of maxicircle gene transcripts³. These guide RNAs bind to mitochondrial mRNA and edit the targeted mRNA sequences^{5,67,68}. In *Trypanosoma brucei*, transcript editing accounts for over half of mature mRNAs⁶⁹. According to the large number of minicircles (approximately 10,000 molecules classified into different classes) per cell, these minicircle DNAs were previously used as another suitable target for PCR-based *Leishmania* diagnosis^{70–72}. The present and previous evolutionary studies have consistently distinguished the Old World (*L. donovani*, *L. infantum*, *L. chagasi*, *L. tropica*, *L. major*, and *L. aethiopica*) and New World (*L. mexicana*, *L. amazonensis*, *L. venezuelensis*, *L. braziliensis*, *L. guyanensis*, *L. panamensis*, and *L. peruviana*) *Leishmania* species using minicircle DNA analysis^{73–75}.

Interestingly, in the present study, different classes of minicircles were detected in *L. orientalis* strain PCM2, which could be classified in the same clade and were closely related to *L. donovani* minicircles (Fig. 5). This study also identified an *L. orientalis*-specific conserved region in all samples between positions 1888 and 1916 (A/TTC CCCTAACTCGGGTGGGCTCCGGGGAG) of their minicircle DNAs, which was not related to any known

minicircle DNAs in the NCBI nucleotide database (Fig. 4d). However, minicircle data of the subgenus *Mundinia* remain limited, and the inter- and intra-specific relationships within this genus could be further expanded using minicircle DNAs⁷⁶. Considering the kDNA function, maxicircle DNAs function as mtDNA of other organisms, while minicircle DNAs play crucial roles in RNA editing⁷⁷. The high genetic heterogeneity of these minicircles was previously reported and showed that different minicircular sequences could drive similar editing functions, but perhaps on different groups of genes^{78–80}. The minicircular variation could also be divided into more than 200 classes within the same species, suggesting unexplained adaptable and pathogenic mechanisms to be further investigated⁸¹ as well as possible different functions of these different guide RNA genes⁸². Therefore, the unique sequences observed in the maxicircle and minicircle DNAs of *L. orientalis* strain PCM2 may be useful for a better specific diagnosis and identification of *L. orientalis* from more samples in the future.

Conclusion

This study was the first report to identify the mitochondrial genome or kDNA of *L. orientalis* strain PCM2, comprising a maxicircle and 105 minicircle classes, using hybrid whole-genome sequencing data. The maxicircle displayed a conserved genomic structure and gene arrangement in closely related *Leishmania* species. Variations in maxicircle and minicircle DNAs would benefit species identification and population genetic studies. Classes and sequence variations of minicircle DNAs could affect parasite adaptation and pathogenesis via RNA editing. Further investigation of the maxicircle and minicircle of more *L. orientalis* isolates would allow monitoring of this parasite for better medical treatment and healthcare.

Data availability

The data supporting this study's findings are submitted to the NCBI GenBank database, accession numbers OP159070-OP159175.

Received: 27 February 2023; Accepted: 3 November 2023

Published online: 10 November 2023

References

- Cecilio, P. *et al.* Deception and manipulation: The arms of *Leishmania*, a successful parasite. *Front. Immunol.* **5**, 480 (2014).
- Akhoundi, M. *et al.* A historical overview of the classification, evolution, and dispersion of *Leishmania* parasites and sandflies. *PLoS Negl. Trop. Dis.* **10**(3), e0004349 (2016).
- Jensen, R. E. & Englund, P. T. Network news: The replication of kinetoplast DNA. *Annu. Rev. Microbiol.* **66**, 473–491 (2012).
- Cavalcanti, D. P. & de Souza, W. The kinetoplast of trypanosomatids: From early studies of electron microscopy to recent advances in atomic force microscopy. *Scanning* **2018**, 9603051 (2018).
- Blum, B., Bakalara, N. & Simpson, L. A model for RNA editing in kinetoplastid mitochondria: RNA molecules transcribed from maxicircle DNA provide the edited information. *Cell* **60**(2), 189–198 (1990).
- Felsenstein, J. Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evolut.* **17**(6), 368–376 (1981).
- Flegontov, P. N., Guo, Q., Ren, L., Strelkova, M. V. & Kolesnikov, A. A. Conserved repeats in the kinetoplast maxicircle divergent region of *Leishmania* sp. and *Leptomonas seymouri*. *Mol. Genet. Genom.* **276**(4), 322–333 (2006).
- Muhich, M. L., Neckelmann, N. & Simpson, L. The divergent region of the *Leishmania tarentolae* kinetoplast maxicircle DNA contains a diverse set of repetitive sequences. *Nucleic Acids Res.* **13**(9), 3241–3260 (1985).
- Alfonzo, J. D., Blanc, V., Estevez, A. M., Rubio, M. A. T. & Simpson, L. C to U editing of the anticodon of imported mitochondrial tRNA^{Trp} allows decoding of the UGA stop codon in *Leishmania tarentolae*. *EMBO J.* **18**(24), 7056–7062 (1999).
- Sturm, N. R., Maslov, D. A., Blum, B. & Simpson, L. Generation of unexpected editing patterns in *Leishmania tarentolae* mitochondrial mRNAs: Misediting produced by misguiding. *Cell* **70**(3), 469–476 (1992).
- Simpson, L. The mitochondrial genome of kinetoplastid protozoa: Genomic organization, transcription, replication, and evolution. *Annu. Rev. Microbiol.* **41**(1), 363–380 (1987).
- Ramrath, D. J. *et al.* Evolutionary shift toward protein-based architecture in trypanosomal mitochondrial ribosomes. *Science* **362**(6413), eaau7735 (2018).
- Simpson, L., Douglass, S. M., Lake, J. A., Pellegrini, M. & Li, F. Comparison of the mitochondrial genomes and steady state transcriptomes of two strains of the trypanosomatid parasite, *Leishmania tarentolae*. *PLoS Negl. Trop. Dis.* **9**(7), e0003841 (2015).
- Duarte, M. & Tomás, A. M. The mitochondrial complex I of trypanosomatids—an overview of current knowledge. *J. Bioenerg. Biomembr.* **46**(4), 299–311 (2014).
- Kirby, L. E. & Koslowsky, D. Mitochondrial dual-coding genes in *Trypanosoma brucei*. *PLoS Negl. Trop. Dis.* **11**(10), e0005989 (2017).
- Ochsenreiter, T. & Hajduk, S. L. Alternative editing of cytochrome c oxidase III mRNA in trypanosome mitochondria generates protein diversity. *EMBO Rep.* **7**(11), 1128–1133 (2006).
- Sykes, S. E. & Hajduk, S. L. Dual functions of α -ketoglutarate dehydrogenase E2 in the Krebs cycle and mitochondrial DNA inheritance in *Trypanosoma brucei*. *Eukaryot. Cell* **12**(1), 78–90 (2013).
- Sykes, S., Szempruch, A. & Hajduk, S. The krebs cycle enzyme α -ketoglutarate decarboxylase is an essential glycosomal protein in bloodstream African trypanosomes. *Eukaryot. Cell* **14**(3), 206–215 (2015).
- Ochsenreiter, T., Cipriano, M. & Hajduk, S. L. Alternative mRNA editing in trypanosomes is extensive and may contribute to mitochondrial protein diversity. *PLoS One* **3**(2), e1566 (2008).
- Maslov, D. A. Complete set of mitochondrial pan-edited mRNAs in *Leishmania mexicana amazonensis* LV78. *Mol. Biochem. Parasitol.* **173**(2), 107–114 (2010).
- Simpson, L. & Shaw, J. RNA editing and the mitochondrial cryptogenes of kinetoplastid protozoa. *Cell* **57**(3), 355–366 (1989).
- Thiemann, O. H., Maslov, D. A. & Simpson, L. Disruption of RNA editing in *Leishmania tarentolae* by the loss of minicircle-encoded guide RNA genes. *EMBO J.* **13**(23), 5689–5700 (1994).
- Flegontov, P. N., Strelkova, M. V. & Kolesnikov, A. A. The *Leishmania* major maxicircle divergent region is variable in different isolates and cell types. *Mol. Biochem. Parasitol.* **146**(2), 173–179 (2006).
- Yatawara, L., Le, T. H., Wickramasinghe, S. & Agatsuma, T. Maxicircle (mitochondrial) genome sequence (partial) of *Leishmania* major: Gene content, arrangement and composition compared with *Leishmania tarentolae*. *Gene* **424**(1–2), 80–86 (2008).
- Neboháčová, M., Kim, C. E., Simpson, L. & Maslov, D. A. RNA editing and mitochondrial activity in promastigotes and amastigotes of *Leishmania donovani*. *Int. J. Parasitol.* **39**(6), 635–644 (2009).

26. Nocua, P., Ramírez, C., Requena, J. M. & Puerta, C. J. Secuencia parcial del genoma del maxicirculo de *Leishmania braziliensis*, comparación con otros tripanosomátidos. *Univ. Sci.* **16**(1), 29–50 (2011).
27. Rogers, M. B. *et al.* Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*. *Genome Res.* **21**(12), 2129–2142 (2011).
28. Singh, N., Chikara, S. & Sundar, S. SOLiD™ sequencing of genomes of clinical isolates of *Leishmania donovani* from India confirm leptomonas co-infection and raise some key questions. *Plos one.* **8**(2), e55738 (2013).
29. Real, F. *et al.* The genome sequence of *Leishmania* (*Leishmania*) amazonensis: Functional annotation and extended analysis of gene models. *DNA Res.* **20**(6), 567–581 (2013).
30. Llanes, A., Restrepo, C. M., Vecchio, G. D., Anguizola, F. J. & Lleontart, R. The genome of *Leishmania panamensis*: Insights into genomics of the *L.* (*Viannia*) subgenus. *Sci. Rep.* **5**(1), 1–10 (2015).
31. Camacho, E. *et al.* *Leishmania* mitochondrial genomes: Maxicircle structure and heterogeneity of minicircles. *Genes* **10**(10), 758 (2019).
32. Urrea, D. A., Triana-Chavez, O. & Alzate, J. F. Mitochondrial genomics of human pathogenic parasite *Leishmania* (*Viannia*) panamensis. *PeerJ.* **7**, e7235 (2019).
33. Bussotti, G. *et al.* Nuclear and mitochondrial genome sequencing of North-African *Leishmania infantum* isolates from cured and relapsed visceral leishmaniasis patients reveals variations correlating with geography and phenotype. *Microb. Genom.* **6**(10), mgen000444 (2020).
34. Bualert, L. *et al.* Case Report: Autochthonous disseminated dermal and visceral leishmaniasis in an AIDS patient, southern Thailand, caused by *Leishmania siamensis*. *Am. J. Trop. Med. Hyg.* **86**(5), 821 (2012).
35. Anuntasomboon, P. *et al.* Comparative draft genomes of *Leishmania orientalis* Isolate PCM2 (formerly named *Leishmania siamensis*) and *Leishmania martiniquensis* Isolate PCM3 from the Southern Province of Thailand. *Biology* **11**(4), 515 (2022).
36. Andrews, S. FastQC: A quality control tool for high throughput sequence data. In *Babraham Bioinformatics* (Babraham Institute, Cambridge, 2010).
37. Antipov, D., Korobeynikov, A., McLean, J. S. & Pevzner, P. A. hybridSPAdes: An algorithm for hybrid assembly of short and long reads. *Bioinformatics* **32**(7), 1009–1015 (2016).
38. Antipov, D. *et al.* plasmidSPAdes: Assembling plasmids from whole genome sequencing data. *Bioinformatics* **32**(22), 3380–3387 (2016).
39. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**(3), 403–410 (1990).
40. Geerts, M., Schnauffer, A. & Van den Broeck, F. rKOMICS: An R package for processing mitochondrial minicircle assemblies in population-scale genome projects. *BMC Bioinform.* **22**(1), 1–14 (2021).
41. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: A versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).
42. Darling, A. C., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **14**(7), 1394–1403 (2004).
43. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**(18), 3094–3100 (2018).
44. Li, H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **37**(23), 4572–4574 (2021).
45. Walker, B. J. *et al.* Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**(11), e112963 (2014).
46. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**(16), 2078–2079 (2009).
47. Katoh, K., Misawa, K., Kuma, K. I. & Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**(14), 3059–3066 (2002).
48. Okonechnikov, K., Golosova, O., Fursov, M., & Team, U. Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics.* **28**(8), 1166–1167 (2012).
49. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**(22), 4673–4680 (1994).
50. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**(9), 1312–1313 (2014).
51. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evolut.* **26**(7), 1641–1650 (2009).
52. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**(W1), W293–W296 (2021).
53. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evolut.* **35**(6), 1547 (2018).
54. Callejas-Hernández, F., Herreros-Cabello, A., del Moral-Salmoral, J., Fresno, M., & Gironès, N. The complete mitochondrial DNA of *Trypanosoma cruzi*: Maxicircles and minicircles. *Front. Cell. Infect. Microbiol.* **556** (2021).
55. Berná, L. *et al.* Maxicircle architecture and evolutionary insights into *Trypanosoma cruzi* complex. *PLoS Negl. Trop. Dis.* **15**(8), e0009719 (2021).
56. Gerasimov, E. S. *et al.* Common structural patterns in the maxicircle divergent region of Trypanosomatidae. *Pathogens* **9**(2), 100 (2020).
57. Ray, D. S. Conserved sequence blocks in kinetoplast minicircles from diverse species of trypanosomes. *Mol. Cell. Biol.* **9**(3), 1365–1367 (1989).
58. Killick-Kendrick, R., Lainson, R., Rioux, J.-A., & Sarjanova, V. The taxonomy of *Leishmania*-like parasites of reptiles. pp. 143–8 (1986).
59. Espinosa, O., Serrano, M., Camargo, E., Teixeira, M. & Shaw, J. An appraisal of the taxonomy and nomenclature of trypanosomatids presently classified as *Leishmania* and *Endotrypanum*. *Parasitology* **145**(4), 430–442 (2018).
60. Pothirat, T. *et al.* First isolation of *Leishmania* from Northern Thailand: Case report, identification as *Leishmania martiniquensis* and phylogenetic position within the *Leishmania enriettii* complex. *PLoS Negl. Trop. Dis.* **8**(12), e3339 (2014).
61. Jariyapan, N. *et al.* *Leishmania* (*Mundinia*) *orientalis* n. sp. (Trypanosomatidae), a parasite from Thailand responsible for localised cutaneous leishmaniasis. *Parasites Vectors* **11**(1), 1–9 (2018).
62. Sloof, P. *et al.* The nucleotide sequence of the variable region in *Trypanosoma brucei* completes the sequence analysis of the maxicircle component of mitochondrial kinetoplast DNA. *Mol. Biochem. Parasitol.* **56**(2), 289–299 (1992).
63. Lee, S. T., Liu, H. Y., Chu, T. & Lin, S. Y. Specific A+ T-rich repetitive DNA-sequences in maxicircles from wildtype *Leishmania mexicana* amazonensis and variants with DNA amplification. *Exp. Parasitol.* **79**(1), 29–40 (1994).
64. De la Cruz, V., Neckelmann, N. & Simpson, L. Sequences of six genes and several open reading frames in the kinetoplast maxicircle DNA of *Leishmania tarentolae*. *J. Biol. Chem.* **259**(24), 15136–15147 (1984).
65. Eperon, I., Janssen, J., Hoesjmakers, J. H. J. & Borst, P. The major transcripts of the kinetoplast DNA of *Trypanosoma brucei* are very small ribosomal RNAs. *Nucleic Acids Res.* **11**(1), 105–125 (1983).
66. Michelotti, E. F., Harris, M. E., Adler, B., Torri, A. F. & Hajduk, S. L. *Trypanosoma brucei* mitochondrial ribosomal RNA synthesis, processing and developmentally regulated expression. *Mol. Biochem. Parasitol.* **54**(1), 31–41 (1992).

67. Pollard, V. W., Rohrer, S. P., Michelotti, E. F., Hancock, K. & Hajduk, S. L. Organization of minicircle genes for guide RNAs in *Trypanosoma brucei*. *Cell* **63**(4), 783–790 (1990).
68. Koslowsky, D. J., Riley, G. R., Feagin, J. E. & Stuart, K. Guide RNAs for transcripts with developmentally regulated RNA editing are present in both life cycle stages of *Trypanosoma brucei*. *Mol. Cell. Biol.* **12**(5), 2043–2049 (1992).
69. Read, L. K., Myler, P. J. & Stuart, K. Extensive editing of both processed and preprocessed maxicircle CR6 transcripts in *Trypanosoma brucei*. *J. Biol. Chem.* **267**(2), 1123–1128 (1992).
70. Selvapandiyar, A. *et al.* A *Leishmania* minicircle DNA footprint assay for sensitive detection and rapid speciation of clinical isolates. *Transfusion* **48**(9), 1787–1798 (2008).
71. Noyes, H. A., Reyburn, H., Bailey, J. W. & Smith, D. A nested-PCR-based schizodeme method for identifying *Leishmania* kinetoplast minicircle classes directly from clinical samples and its application to the study of the epidemiology of *Leishmania tropica* in Pakistan. *J. Clin. Microbiol.* **36**(10), 2877–2881 (1998).
72. Aransay, A. M., Scoulica, E. & Tselentis, Y. Detection and identification of *Leishmania* DNA within naturally infected sand flies by seminested PCR on minicircle kinetoplastic DNA. *Appl. Environ. Microbiol.* **66**(5), 1933–1938 (2000).
73. Nicolas, L., Milon, G. & Prina, E. Rapid differentiation of Old World *Leishmania* species by LightCycler polymerase chain reaction and melting curve analysis. *J. Microbiol. Methods* **51**(3), 295–299 (2002).
74. de Monbrison, F., Mihoubi, I. & Picot, S. Real-time PCR assay for the identification of cutaneous *Leishmania* parasite species in Constantine region of Algeria. *Acta Trop.* **102**(2), 79–83 (2007).
75. de Paiva Cavalcanti, M. *et al.* Quantitative real time PCR assays for the detection of *Leishmania* (*Viannia*) *braziliensis* in animals and humans. *Mol. Cell. Probes* **27**(3–4), 122–128 (2013).
76. Junqueira, A. C., Degraive, W. & Brandão, A. Minicircle organization and diversity in *Trypanosoma cruzi* populations. *Trends Parasitol.* **21**(6), 270–272 (2005).
77. Stuart, K. & Feagin, J. E. Mitochondrial DNA of kinetoplastids. *Int. Rev. Cytol.* **141**, 65–88 (1992).
78. Morel, C. *et al.* Strains and clones of *Trypanosoma cruzi* can be characterized by pattern of restriction endonuclease products of kinetoplast DNA minicircles. *Proc. Natl. Acad. Sci.* **77**(11), 6810–6814 (1980).
79. Macedo, A. M., Martins, M. S., Chiari, E. & Pena, S. D. DNA fingerprinting of *Trypanosoma cruzi*: A new tool for characterization of strains and clones. *Mol. Biochem. Parasitol.* **55**(1–2), 147–153 (1992).
80. Hanke, J. *et al.* Mapping the *Trypanosoma cruzi* genome: Analyses of representative cosmid libraries. *Biotechniques* **21**(4), 686–693 (1996).
81. Hong, M. & Simpson, L. Genomic organization of *Trypanosoma brucei* kinetoplast DNA minicircles. *Protist* **154**(2), 265–279 (2003).
82. Rusman, F., Florida-Yapur, N., Tomasini, N. & Diosque, P. Guide RNA repertoires in the main lineages of *Trypanosoma cruzi*: High diversity and variable redundancy among strains. *Front. Cell. Infect. Microbiol.* **11**, 663416 (2021).

Acknowledgements

The authors would like to thank the Department of Parasitology, Phramongkutkloa College of Medicine; the Department of Microbiology, Faculty of Science, Mahidol University; the Department of Genetics, Faculty of Science, Kasetsart University for supporting this project; and the Faculty of Science, Kasetsart University for providing high-performance computing facilities.

Author contributions

Conceptualisation, T.E., M.M., S.L., and R.B.; methodology, T.E., P.A., and S.S.; software, T.E. and P.A.; validation, T.E. and P.A.; investigation, T.E. and P.A.; resources, K.C. and S.U.; data curation, T.E.; writing—original draft preparation, T.E. and P.A.; writing—review and editing, M.M., S.L., and T.E.; supervision, M.M., S.L., R.B., and T.E.; project administration, T.E.; funding acquisition, T.E. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by Kasetsart University Research and Development Institute (KURDI) (Grant Number FF(KU)6.65).

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to T.E.

Reprints and permissions information is available at www.nature.com/reprints.

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



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

© The Author(s) 2023