

# Improved assembly of the *Leishmania panamensis* strain PSC-1 genome

Aldair Fabrega,<sup>1,2</sup> Ana Cristina Ortega,<sup>1,3</sup> Letizia Pardo,<sup>2</sup> Carlos M. Restrepo,<sup>1,4</sup> Alejandro Llanes<sup>1,4</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 2.

**ABSTRACT** *Leishmania panamensis* is a causative agent of leishmaniasis in Latin America, with significant incidence in Panama and Colombia. The first annotated genome draft for this parasite remained incomplete due to its highly repetitive nature. Here, we combined short- and long-read sequencing data to improve this genome draft.

**KEYWORDS** *Leishmania panamensis*, genome announcements, trypanosomatids, intracellular parasites

A draft of the *Leishmania panamensis* strain PSC-1 genome, released in 2015 (1), remained the only annotated genome available for this species in the GenBank and RefSeq databases at the time of this study. This draft was assembled using 454 pyrosequencing and Illumina reads, resulting in several unfinished regions mainly due to tandem gene arrays. Such arrays arise from the amplification of specific genes, a mechanism allowing *Leishmania* and other trypanosomatid parasites to modulate gene dosage in absence of transcriptional regulation of gene expression (2). The abundance of gene arrays and other repetitive elements often leads to highly fragmented assemblies when sequencing trypanosomatid genomes (3). With the advent of long-read platforms, such as PacBio and Oxford Nanopore (ONT), several reference genomes have been re-assembled to solve issues related to fragmentation (4–6). Here, we used a combination of ONT and Illumina reads to improve the *L. panamensis* strain PSC-1 genome draft.

Strain MHOM/PA/94/PSC-1 was isolated in 1994 from a skin lesion of a human patient from Panama. Cryopreserved parasites were cultured in serum-supplemented Schneider's insect medium incubated at 25°C. Genomic DNA was extracted using the Wizard Genomic DNA Purification Kit (Promega). The ONT library was prepared with the Ligation Sequencing Kit SQK-LSK110, without shearing or size selection. This library was sequenced using an R9.4.1 flow cell in a PromethION instrument, generating approximately 581,000 reads ( $N_{50} = 17$  kb). Guppy v.3.2.4 (HAC model) and Porechop v.0.2.4 were used for base calling and adapter trimming, respectively. The Illumina library was prepared with the TruSeq DNA HT Sample Prep Kit and sequenced in a HiSeq 2000 instrument, yielding 29.4 million 100 bp paired-end reads. FastQC v.0.11.19 and Trimmomatic v.0.39 (7) were used for quality control and trimming, respectively. Canu v.2.2 (8) was used to assemble the ONT reads *de novo* with the "genomeSize=34m" option, resulting in a set of 35 non-redundant contigs ( $N_{50} = 1,281$  kb), consistent with the 35 *L. panamensis* chromosomes. Illumina reads were then used for error correction with Pilon v.1.24 (9) in seven iterations. Illumina and ONT reads were aligned to the polished assembly with BWA-MEM v.0.7.17 (10) and Minimap2 v.2.24 (11), respectively. These alignments were visually inspected in Artemis v.17.0.1 (12) to detect possible misassemblies in regions that were not supported by the reads, all of which were manually corrected. Default parameters for each type of read were used for all software except otherwise noted.

**Editor** Jason E. Stajich, University of California Riverside, Riverside, California, USA

Address correspondence to Alejandro Llanes, allanes@indicat.org.pa.

The authors declare no conflict of interest.

See the funding table on p. 2.

**Received** 10 June 2025

**Accepted** 3 July 2025

**Published** 21 July 2025

Copyright © 2025 Fabrega et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

TABLE 1 Global statistics of the first two versions of the *Leishmania panamensis* strain PSC-1 genome

Genome statistic	Version 1 (2015) <sup>a</sup>	Version 2 (2025)
Total size (Mb)	30.7	33.9
Sequencing coverage	100×	100×
Number of chromosomes	35	35
Chromosome length range (Mb)	0.3–2.6	0.3–3.1
GC content (%)	57.6	58.2
Number of gaps	553	0
N content (%)	2.3	0
Protein-coding genes	7,748	8,621
Pseudogenes	185	219

<sup>a</sup>Data taken from Llanes et al. (1), provided here only for comparative purposes.

Annotation of the improved assembly was performed with RATT v.1.0 (13) and the Companion server v.2.2.11 (14) to transfer previously annotated genes and to identify new ones, respectively. When compared with the 2015 genome draft, the improved assembly contains more than 3 Mb of additional sequence (Table 1). Newly assembled regions putatively encode more than 800 protein-coding genes, many of which were confirmed to be located within tandem gene arrays. Assessment of genome completeness with BUSCO v.5.4.4 (15) identified 100% of the genes in the “euglenozoa\_odb10” lineage data set. The presence of *Leishmania* telomeric repeats (16, 17) was also confirmed near chromosome ends.

## ACKNOWLEDGMENTS

This work was supported by Sistema Nacional de Investigación (SNI) with contract number SNI-043-2023, granted to Alejandro Llanes.

We thank Ms. Ana Lucia Altieri for her invaluable help during genome annotation.

## AUTHOR AFFILIATIONS

<sup>1</sup>División de Salud Humana y Enfermedades, Instituto de Investigaciones Científicas y Servicios de Alta Tecnología AIP (INDICASAT-AIP), Panama City, Panama

<sup>2</sup>Facultad de Ciencias de la Salud, Universidad Latina de Panamá, Panama City, Panama

<sup>3</sup>Facultad de Ciencias y Tecnología, Universidad Tecnológica de Panamá (UTP), Panama City, Panama

<sup>4</sup>Sistema Nacional de Investigación (SNI), Panama City, Panama

## AUTHOR ORCIDS

Alejandro Llanes  <http://orcid.org/0000-0002-6949-2782>

## FUNDING

Funder	Grant(s)	Author(s)
Sistema Nacional de Investigación (SNI), Panama	SNI-043-2023	Alejandro Llanes

## AUTHOR CONTRIBUTIONS

Aldair Fabrega, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft | Ana Cristina Ortega, Data curation | Letizia Pardo, Data curation | Carlos M. Restrepo, Formal analysis, Investigation, Methodology, Resources | Alejandro Llanes, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Writing – original draft

## DATA AVAILABILITY

The sequence reads, assembly, and annotation reported here are available in GenBank under BioProject [PRJNA235344](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA235344). Raw sequence reads have been deposited in the Sequence Read Archive (SRA) under accession numbers [SRX681913](https://www.ncbi.nlm.nih.gov/sra/SRX681913) for Illumina reads and [SRX28833427](https://www.ncbi.nlm.nih.gov/sra/SRX28833427) for ONT reads. Annotated chromosomes were deposited in GenBank with accession numbers [CP009370.2](https://www.ncbi.nlm.nih.gov/nuccore/CP009370.2) to [CP009404.2](https://www.ncbi.nlm.nih.gov/nuccore/CP009404.2). The version described in this paper is the second version.

## REFERENCES

- Llanes A, Restrepo CM, Del Vecchio G, Anguizola FJ, Leonart R. 2015. The genome of *Leishmania panamensis*: insights into genomics of the *L. (Viannia)* subgenus. *Sci Rep* 5:8550. <https://doi.org/10.1038/srep08550>
- Rogers MB, Hilley JD, Dickens NJ, Wilkes J, Bates PA, Depledge DP, Harris D, Her Y, Herzyk P, Imamura H, Otto TD, Sanders M, Seeger K, Dujardin JC, Berriman M, Smith DF, Hertz-Fowler C, Mottram JC. 2011. Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*. *Genome Res* 21:2129–2142. <https://doi.org/10.1101/gr.122945.111>
- Pita S, Díaz-Viraqué F, Iraola G, Robello C. 2019. The tritryps comparative repeatome: insights on repetitive element evolution in trypanosomatid pathogens. *Genome Biol Evol* 11:546–551. <https://doi.org/10.1093/gbe/evz017>
- González-de la Fuente S, Peiró-Pastor R, Rastrojo A, Moreno J, Carrasco-Ramiro F, Requena JM, Aguado B. 2017. Resequencing of the *Leishmania infantum* (strain JPCM5) genome and *de novo* assembly into 36 contigs. *Sci Rep* 7:18050. <https://doi.org/10.1038/s41598-017-18374-y>
- Díaz-Viraqué F, Pita S, Greif G, CM, Iraola G, Robello C. 2019. Nanopore sequencing significantly improves genome assembly of the protozoan parasite *Trypanosoma cruzi*. *Genome Biol Evol* 11:1952–1957. <https://doi.org/10.1093/gbe/evz129>
- Lypaczewski P, Hoshizaki J, Zhang W-W, McCall L-I, Torcivia-Rodriguez J, Simonyan V, Kaur A, Dewar K, Matlashewski G. 2018. A complete *Leishmania donovani* reference genome identifies novel genetic variations associated with virulence. *Sci Rep* 8:16549. <https://doi.org/10.1038/s41598-018-34812-x>
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res* 27:722–736. <https://doi.org/10.1101/gr.215087.116>
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>
- Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26:589–595. <https://doi.org/10.1093/bioinformatics/btp698>
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28:464–469. <https://doi.org/10.1093/bioinformatics/btr703>
- Otto TD, Dillon GP, Degraeve WS, Berriman M. 2011. RATT: rapid annotation transfer tool. *Nucleic Acids Res* 39:e57–e57. <https://doi.org/10.1093/nar/gkq1268>
- Haese-Hill W, Crouch K, Otto TD. 2024. Annotation and visualization of parasite, fungi and arthropod genomes with Companion. *Nucleic Acids Res* 52:W39–W44. <https://doi.org/10.1093/nar/gkae378>
- Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2018. BUSCO Applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol* 35:543–548. <https://doi.org/10.1093/molbev/msx319>
- Fu G, Barker DC. 1998. Characterisation of *Leishmania* telomeres reveals unusual telomeric repeats and conserved telomere-associated sequence. *Nucleic Acids Res* 26:2161–2167. <https://doi.org/10.1093/nar/26.9.2161>
- Dossin F de M, Dufour A, Dusch E, Siqueira-Neto JL, Moraes CB, Yang GS, Cano MI, Genovesio A, Freitas-Junior LH. 2008. Automated nuclear analysis of *Leishmania* major telomeric clusters reveals changes in their organization during the parasite's life cycle. *PLoS ONE* 3:e2313. <https://doi.org/10.1371/journal.pone.0002313>