





# Genome Assemblies across the Diverse Evolutionary Spectrum of *Leishmania* Protozoan Parasites

Wesley C. Warren,<sup>a</sup> Natalia S. Akopyants,<sup>b</sup> Deborah E. Dobson,<sup>b</sup> Christiane Hertz-Fowler,<sup>c</sup> Lon-Fye Lye,<sup>b</sup>  Peter J. Myler,<sup>d,e,f,g</sup> Gowthaman Ramasamy,<sup>d</sup> Achchuthan Shanmugasundram,<sup>c</sup> Fatima Silva-Franco,<sup>c</sup> Sascha Steinbiss,<sup>h</sup> Chad Tomlinson,<sup>i</sup> Richard K. Wilson,<sup>j</sup>  Stephen M. Beverley<sup>b</sup>

<sup>a</sup>Department of Animal Sciences, Institute of Data Science and Informatics, Bond Life Sciences Center, University of Missouri, Columbia, Missouri, USA

<sup>b</sup>Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, Missouri, USA

<sup>c</sup>Institute of Integrative Biology, University of Liverpool, Liverpool, United Kingdom

<sup>d</sup>Center for Global Infectious Disease Research, Seattle Children's Research Institute, Seattle, Washington, USA

<sup>e</sup>Department of Pediatrics, University of Washington, Seattle, Washington, USA

<sup>f</sup>Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, Washington, USA

<sup>g</sup>Department of Global Health, University of Washington, Seattle, Washington, USA

<sup>h</sup>Wellcome Sanger Institute, Hinxton, Cambridgeshire, United Kingdom

<sup>i</sup>McDonnell Genome Institute, Washington University School of Medicine, St. Louis, Missouri, USA

<sup>j</sup>The Steve and Cindy Rasmussen Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, Ohio, USA

**ABSTRACT** We report the high-quality draft assemblies and gene annotations for 13 species and/or strains of the protozoan parasite genera *Leishmania*, *Endotrypanum*, and *Crithidia*, which span the phylogenetic diversity of the subfamily Leishmaniinae within the kinetoplastid order of the phylum Euglenozoa. These resources will support studies on the origins of parasitism.

*Leishmania* species are widespread parasites of mammals transmitted by biting insects. Over 1.7 billion people worldwide are at risk, with hundreds of millions of people infected (1–4). The genus comprises more than 50 species, which are primarily zoonotic but in humans cause disease ranging from mild cutaneous lesions to more disseminated forms to fatal visceralizing disease (5). While parasitism by *Leishmania* has been intensively studied, the species-specific factors that enable mammalian or insect host infections are less well understood. To provide a broad phylogenetic snapshot, we selected a spectrum of species and strains across the subfamily Leishmaniinae (5), targeting lineages within the subgenera *Leishmania*, *Viannia*, and *Mundinia*, as well as the allied genus *Endotrypanum* and the outgroup *Crithidia fasciculata* (Table 1). The WU Institutional Biosafety Committee reviewed and approved the parasite work reported here (01-015).

Parasites were cultivated in M199 or Schneider's medium (6) and grown to late log phase before harvesting, lysis, and DNA purification by phenol-chloroform extraction and/or banding in CsCl gradients (to remove mitochondrial maxi- or minicircle DNA). Sequencing libraries were generated using the Illumina paired-end DNA sample preparation kit (PE-102-1001) according to the manufacturer's directions. Fragment libraries of 3 and 8 kb were prepared using protocols for 454 sequencing (Roche Life Sciences). Sequencing was performed on either a 454 GS FLX Titanium (average read length, 305 bp; Roche 454 Life Sciences) or Illumina Genome Analyzer IIx (GAIIx) and HiSeq 2000 instruments (paired-end 100-bp format), except for *Crithidia*, which additionally utilized long reads generated on an RS II instrument (P5/C3 chemistry; Pacific Biosciences) (7). The total sequence genome coverage on the Illumina GAIIx and HiSeq 2000 instruments was on average 105× with tiered library insert sizes (50× fragments;

**Citation** Warren WC, Akopyants NS, Dobson DE, Hertz-Fowler C, Lye L-F, Myler PJ, Ramasamy G, Shanmugasundram A, Silva-Franco F, Steinbiss S, Tomlinson C, Wilson RK, Beverley SM. 2021. Genome assemblies across the diverse evolutionary spectrum of *Leishmania* protozoan parasites. *Microbiol Resour Annot* 10:e00545-21. <https://doi.org/10.1128/MRA.00545-21>.

**Editor** Vincent Bruno, University of Maryland School of Medicine

**Copyright** © 2021 Warren 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/).

Address correspondence to Stephen M. Beverley, [stephen.beverley@wustl.edu](mailto:stephen.beverley@wustl.edu).

**Received** 2 June 2021

**Accepted** 10 August 2021

**Published** 2 September 2021

**TABLE 1** Description of *Leishmania* species and strains, including assembly parameters and links

Genus or subgenus	Species	Strain	WHO code	Source	Provenance	BioProject accession no. <sup>a</sup>	Sequencing platforms	Assembler	No. of contigs	N <sub>50</sub> (bp)	Assembly size (bp)	% GC	No. of pseudo-chromosomes	No. of protein-coding genes	Reference for other assembly
<i>L. Leishmania</i>	<i>L. major</i>	SD75.1 (clone)	MHOM/SN/74/SD	Human, cutaneous	D. Sacks, Bethesda, MD	<a href="#">PRINA50303</a>	454, Illumina	Newbler	891	95,380	31,727,271	59.3	36	8,818	
		LV39 (clone 5)	MHOM/Sv/59/P	Human, cutaneous	R. Titus, Boston, MA	<a href="#">PRINA50301</a>	454, Illumina	Newbler	1,667	71,452	31,961,985	59.5	36	8,971	
	<i>L. gerbilli</i>	LEM452	MRHO/CN/60/GERBILLI	Gerbil	P. Bastien, J.-P. Dedet, Montpellier, France	<a href="#">PRINA192717</a>	454, Illumina	AllPaths	1,248	57,008	30,822,621	59.6	36	8,599	
	<i>L. turanica</i>	LEM423	MRHO/SU/65/VL	Gerbil	P. Bastien, J.-P. Dedet, Montpellier, France	<a href="#">PRINA192712</a>	454, Illumina	AllPaths	1,669	39,210	30,876,294	59.5	36	8,608	
	<i>L. arabica</i>	LEM1108	MPSA/SA/83/JISH220	<i>Psammomys</i> sp.	P. Bastien, J.-P. Dedet, Montpellier, France	<a href="#">PRINA192710</a>	454, Illumina	AllPaths	1,530	52,119	30,774,332	59.2	36	8,646	
	<i>L. tropica</i>	L590	MHOM/IL/1990/P283	Human, cutaneous	C. Jaffe, Jerusalem, Israel	<a href="#">PRINA169676</a>	454, Illumina	AllPaths	1,938	32,739	31,326,083	59.6	36	8,824	12
	<i>L. aethiopia</i>	L147	MHOM/ET/1972/L100	Human, diffuse cutaneous, relapsing	C. Jaffe, Jerusalem, Israel	<a href="#">PRINA169673</a>	454, Illumina	AllPaths	1,758	38,498	31,026,739	60.1	36	8,722	
<i>L. Viannia</i>	<i>L. braziliensis</i>	M2903	MHOM/BR/75/M2903	Human, cutaneous	J. Shaw, Brazil	<a href="#">PRINA165955</a>	454, Illumina	Newbler	3,934	61,918	32,590,753	57.4	35	9,269	
	<i>L. panamensis</i>	L13	MHOM/COL/81/L13	Human, mucosal	N. Saravia, Cali, Colombia	<a href="#">PRINA165959</a>	454, Illumina	AllPaths	3,163	22,576	31,108,242	57.4	35	8,665	
<i>L. Mundinia</i>	<i>L. enriettii</i>	LEM3045	MCAV/BR/95/CUR3	<i>Cavia porcellus</i> (guinea pig)	P. Bastien, J.-P. Dedet, Montpellier, France	<a href="#">PRINA192711</a>	454, Illumina	AllPaths	1,171	102,666	30,427,298	59.3	36	8,731	14
	<i>L. martiniquensis</i>	LEM2494	MHOM/MQ/92/MAR1	Human, diffuse cutaneous, HIV	P. Bastien, J.-P. Dedet, Montpellier, France	<a href="#">PRINA192703</a>	454, Illumina	AllPaths	628	147,290	30,528,357	59.6	36	8,483	14
<i>Endotrypanum</i>	<i>Endotrypanum montergei</i>	LV88	None	<i>Choloepus hoffmanni</i> (sloth)	Michael Chance, Liverpool, UK	<a href="#">PRINA165953</a>	454, Illumina	Newbler	3,517	33,059	32,086,870	52.5	36	8,285	13
<i>Crithidia</i>	<i>Crithidia fasciculata</i>	CF-C1 (clone)	None	Mosquito	Larry Simpson, Los Angeles, CA	<a href="#">PRINA165885</a>	454, Illumina, PacBio	HGAP	494	778,443	41,297,378	57.0	30	9,619	8

<sup>a</sup> Each BioProject link contains links to the current assembly, primary data sets, and other relevant information.

45 × 3 kb, 10 × 8 kb, and 0.05 × 40 kb). For all Illumina sequences, we used the read processing steps within the AllPaths-LG (8) software prior to *de novo* assembly, which incorporates read error correction methods described by Pevzner et al. (9). Genome assemblies were conducted with default parameters using Newbler v2.0.1 (10) for 454 reads, AllPaths-LG (8) for Illumina reads, and HGAP v3 (11) for long reads (Table 1). Contigs and scaffolds were organized into pseudochromosomes using ABACAS2 (<https://github.com/satta/ABACAS2>), a successor to ABACAS (12), by alignment with the *Leishmania major* Friedlin genome sequence, with the exception of *Leishmania braziliensis* M2903 and *Leishmania panamensis*, which were aligned to the *L. braziliensis* M2904 genome. The estimated haploid genome sizes ranged from 30.4 to 41.3 megabases (13).

Gene annotations were performed using the comprehensive Companion tool, which incorporates a variety of *de novo* prediction criteria, as well as information from closely related genomes when available (14). The number of protein-coding genes predicted ranged from 8,285 to 9,619, typical of other *Leishmania* species (13). Full annotations, as well as a variety of tools for the visualization or analysis of these genomes, are available from TriTrypDB ([www.tritrypdb.org](http://tritrypdb.org)).

**Data availability.** The assemblies have been deposited in the NCBI GenBank repository under the BioProject accession numbers in Table 1, including links to the primary data and annotations (PRJNA50303, PRJNA50301, PRJNA192717, PRJNA192712, PRJNA192710, PRJNA169676, PRJNA169673, PRJNA165955, PRJNA165959, PRJNA192711, PRJNA192703, PRJNA165953, and PRJNA165885). The chromosome builds are available through the TriTrypDB portal (<http://tritrypdb.org/tritrypdb>).

## ACKNOWLEDGMENTS

NIH grant AI29646 to S.M.B. provided support for N.S.A., D.E.D., and L.-F.L.; NIH-NHGRI grant HG00307907 to R.K.W. supported W.C.W. and C.T.; and AI103858 to P.J.M. supported G.R. Funding to C.H.-F. from Wellcome Trust grants WT099198MA and WT108443MA provided support for F.S.-F., A.S., and S.S.

We thank Patrick Minx for assembly curation. We thank the colleagues listed in Table 1 for generously providing the strains and Brian Brunk, David Roos, Thomas Otto, and Matt Berriman for discussions.

This research was funded in whole, or in part, by the Wellcome Trust (WT099198MA and WT108443MA). For the purpose of open access, the author has applied a CC BY public copyright license to any author accepted manuscript version arising from this submission.

## REFERENCES

- Pigott DM, Bhatt S, Golding N, Duda KA, Battle KE, Brady OJ, Messina JP, Balard Y, Bastien P, Pratlong F, Brownstein JS, Freifeld CC, Mekaru SR, Gething PW, George DB, Myers MF, Reithinger R, Hay SI. 2014. Global distribution maps of the leishmaniases. *Elife* 3:e02851. <https://doi.org/10.7554/eLife.02851>.
- Alvar J, Velez ID, Bern C, Herrero M, Desjeux P, Cano J, Jannin J, den Boer M, WHO Leishmaniasis Control Team. 2012. Leishmaniasis worldwide and global estimates of its incidence. *PLoS One* 7:e35671. <https://doi.org/10.1371/journal.pone.0035671>.
- Banuls AL, Bastien P, Pomares C, Arevalo J, Fisa R, Hide M. 2011. Clinical pleiomorphism in human leishmaniases, with special mention of asymptomatic infection. *Clin Microbiol Infect* 17:1451–1461. <https://doi.org/10.1111/j.1469-0691.2011.03640.x>.
- Singh OP, Hasker E, Sacks D, Boelaert M, Sundar S. 2014. Asymptomatic *Leishmania* infection: a new challenge for *Leishmania* control. *Clin Infect Dis* 58:1424–1429. <https://doi.org/10.1093/cid/ciu102>.
- Espinosa OA, Serrano MG, Camargo EP, Teixeira MMG, Shaw JJ. 2018. An appraisal of the taxonomy and nomenclature of trypanosomatids presently classified as *Leishmania* and *Endotrypanum*. *Parasitology* 145: 430–442. <https://doi.org/10.1017/S0031182016002092>.
- Lye L-F, Owens K, Shi H, Murta SMF, Vieira AC, Turco SJ, Tschudi C, Ullu E, Beverley SM. 2010. Retention and loss of RNA interference pathways in trypanosomatid protozoans. *PLoS Pathog* 6:e1001161. <https://doi.org/10.1371/journal.ppat.1001161>.
- Filosa JN, Berry CT, Ruthel G, Beverley SM, Warren WC, Tomlinson C, Myler PJ, Dudkin EA, Povelones ML, Povelones M. 2019. Dramatic changes in gene expression in different forms of *Crithidia fasciculata* reveal potential mechanisms for insect-specific adhesion in kinetoplastid parasites. *PLoS Negl Trop Dis* 13:e0007570. <https://doi.org/10.1371/journal.pntd.0007570>.
- Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci U S A* 108:1513–1518. <https://doi.org/10.1073/pnas.1017351108>.
- Pevzner PA, Tang H, Waterman MS. 2001. An Eulerian path approach to DNA fragment assembly. *Proc Natl Acad Sci U S A* 98:9748–9753. <https://doi.org/10.1073/pnas.171285098>.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jiracek KB, Kim J-B, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, et al. 2005. Genome sequencing in microfabricated high-density

- picolitre reactors. *Nature* 437:376–380. <https://doi.org/10.1038/nature03959>.
11. Chaisson MJ, Huddleston J, Dennis MY, Sudmant PH, Malig M, Hormozdiari F, Antonacci F, Surti U, Sandstrom R, Boitano M, Landolin JM, Stamatoyannopoulos JA, Hunkapiller MW, Korlach J, Eichler EE. 2015. Resolving the complexity of the human genome using single-molecule sequencing. *Nature* 517:608–611. <https://doi.org/10.1038/nature13907>.
  12. Assefa S, Keane TM, Otto TD, Newbold C, Berriman M. 2009. ABACAS: algorithm-based automatic contiguation of assembled sequences. *Bioinformatics* 25:1968–1969. <https://doi.org/10.1093/bioinformatics/btp347>.
  13. Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, Sisk E, Rajandream M-A, Adlem E, Aert R, Anupama A, Apostolou Z, Attipoe P, Bason N, Bauser C, Beck A, Beverley SM, Bianchetti G, Borzym K, Bothe G, Bruschi CV, Collins M, Cadag E, Ciarloni L, Clayton C, Coulson RMR, Cronin A, Cruz AK, Davies RM, De Gaudenzi J, Dobson DE, Duesterhoeft A, Fazelina G, Fosker N, Frasch AC, Fraser A, Fuchs M, Gabel C, Goble A, Goffeau A, Harris D, Hertz-Fowler C, Hilbert H, Horn D, Huang Y, Klages S, Knights A, Kube M, Larke N, Litvin L, et al. 2005. The genome of the kinetoplastid parasite, *Leishmania major*. *Science* 309:436–442. <https://doi.org/10.1126/science.1112680>.
  14. Steinbiss S, Silva-Franco F, Brunk B, Foth B, Hertz-Fowler C, Berriman M, Otto TD. 2016. Companion: a Web server for annotation and analysis of parasite genomes. *Nucleic Acids Res* 44:W29–W34. <https://doi.org/10.1093/nar/gkw292>.