May 31, 2021

Genome assemblies across the diverse evolutionary spectrum of *Leishmania* protozoan parasites

Wesley C. Warren[1], Natalia S. Akopyants[2], Deborah E. Dobson[2], Christiane Hertz-Fowler[3], Lon-Fye Lye[2], Peter J. Myler[4,5], Gowthaman Ramasamy[4], Achchuthan Shanmugasundram[3], Fatima Silva-Franco[3], Sascha Steinbiss[3], Chad Tomlinson[6], Richard K. Wilson[7], and Stephen M. Beverley[2,*]

[1] Department of Animal Sciences, Institute of Data Science and Informatics, Bond Life Sciences Center, University of Missouri, Columbia, MO, USA.

[2] Department of Molecular Microbiology, Washington University School of Medicine, St Louis, MO; Washington University School of Medicine, St Louis, MO.

[3] Institute of Integrative Biology, Biosciences Building, University of Liverpool, Crown Street, Liverpool L69 7ZB, UK.

[4] Center for Global Infectious Disease Research, Seattle Children's Research Institute, Seattle, WA, USA.

[5] Department of Pediatrics, Department of Biomedical Informatics and Medical Education and Department of Global Health, University of Washington, Seattle WA, USA

[6] McDonnell Genome Institute, Washington University School of Medicine, St Louis, MO; USA

[7] The Steve and Cindy Rasmussen Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, Ohio, USA.

*Corresponding author: Stephen M. Beverley, Molecular Microbiology, Washington University School of Medicine, St Louis, MO 63110, USA. Telephone 1-314-747-2630; email stephen.beverley@wustl.edu.

Keywords: *Leishmania* species, comparative genomics, evolution of parasitism, Kinetoplastida, Trypanosomatidae.

Running title: Genome assemblies across *Leishmania* phylogeny

Abstract

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

Species of the protozoan *Leishmania* 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. It is estimated that 12 million people show active disease with more than 50,000 fatalities recorded annually (1-4). The genus comprises more than 50 species, which have undergone multiple taxonomic classifications (not all of which are universally accepted), although at the molecular level there is widespread consensus of relationships (5). Species within the subgenus *Leishmania* comprise the most widespread group causing disease in humans, ranging from mild cutaneous lesions to more disseminated disease, to fatal visceralizing disease. Species of the subgenus *Viannia* are found primarily in South America and additionally are associated with disfiguring mucocutaneous disease, potentially exacerbated by the presence of widespread RNA viruses (6). Recently a new subgenus *Mundinia* was proposed, comprising a diverse collection of parasites isolated from kangaroos or humans from southeast Asia and the Americas (5). While most *Leishmania* species have been shown or presumed to be transmitted by phlebotomine sand flies, *Mundinia* are distinct in employing chironomid midge hosts (5). Another well-defined molecular clade contains an assemblage of species variously termed *Endotrypanum* or *Leishmania*, primarily found in sloths, but also reported in human infections (5).

While basic principles of *Leishmania* vertebrate parasitism have been intensively studied, our understanding of the species-specific factors that enable mammalian or insect host infections and survival are less well understood. To provide a broad phylogenetic snapshot of *Leishmania* evolution, we selected a spectrum of species and strains across the Leishmaniinae sub-family, targeting those lineages for which relatively little information was available at the time these studies were initiated. These include new members of the *Leishmania* subgenus (*L. aethiopica, L. arabica, L. gerbilli, L. tropica, L. turanica,* and two isolates of *L. major*), two species of *Mundinia* (*L. martiniquensis and L. enriettii),* two species of *Viannia* (*L. braziliensis*, *L. panamensis*), as well as the outgroup species *Endotrypanum monterogeii*, and *Crithidia fasciculata*. The specific strains used, including WHO reference identifiers, provenance, and taxonomic classification, are provided in Table 1.

Parasites were cultivated in M199 or Schneider's medium (7) and grown to late log phase, before harvesting, lysis, and DNA purification by phenol chloroform extraction and/or banding in CsCl gradients (the later step was selected to minimize the contribution of the kinetoplast (mitochondrial) maxi- or minicircle DNA). Sequencing was performed on either a Roche 454 Titanium, or Illumina HiSeq2000 instrument, except for *Crithidia* which additionally utilized PacBio generated sequence data (8). *De novo* assemblies were performed using Newbler (Roche) or ALLPATHS (9). Contigs and scaffolds were organized into pseudochromosomes using ABACAS2, a currently unpublished successor to ABACAS 9 (http://abacas.sourceforge.net/index.html), by alignment to the *Leishmania major* Friedlin genome, with the exception of *L. braziliensis* M2903 and *L. panamensis*, which were aligned to the *L. braziliensis* M2904 genome. The estimated haploid genome sizes ranged from 30.4 to 41.3 megabases, typical for *Leishmania* (10), with assembly parameters summarized in Table 1.

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 (11). The number of protein-coding genes predicted ranged from 8,285 to 9,619, a range typical of *Leishmania* genomes (10). Full annotations, as well as a variety of tools for the visualization or analysis of these genomes, are available on the TriTrypDB website (www.tritrypdb.org),

Public availability and accession number(s). Assemblies are deposited in the NCBI GenBank Repository under the Bioproject numbers in Table 1, including links to the primary data and annotations (PRJNA50303, PRJNA50301, PRJNA192717, PRJNA192712, PRJNA192710, PRJNA169676, PRJNA169673, PRJNA165955, PRJNA165959, PRJNA192711, PRJNA192703, PRJNA165953, PRJNA165885). Chromosome builds are available through the TriTrypDB portal (http://tritrypdb.org/tritrypdb).

Acknowledgements.

NIH grant AI29646 to Stephen M. Beverley provided support for NSA, DED and L-FL, NIH-NHGRI grant HG00307907 to Richard K. Wilson supported WCW, and CT, and AI103858 to Peter J Myler supported GA. Funding to Christiane Hertz-Fowler from Wellcome Trust grants WT099198MA and WT108443MA provided support for FS-F, AS and SS. We thank Patrick Minx for assembly curation. We thank the colleagues listed in Table 1 for generously providing strains and Brian Brunk, David Roos, Thomas Otto and Matt Berriman for discussions.

References.

1. 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.
2. Alvar J, Velez ID, Bern C, Herrero M, Desjeux P, Cano J, Jannin J, den Boer M. 2012. Leishmaniasis worldwide and global estimates of its incidence. PLoS One 7:e35671.
3. 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 17:1451-61.
4. Singh OP, Hasker E, Sacks D, Boelaert M, Sundar S. 2014. Asymptomatic *Leishmania* infection: a new challenge for Leishmania control. Clin Infect Dis 58:1424-9.
5. 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.
6. Hartley MA, Drexler S, Ronet C, Beverley SM, Fasel N. 2014. The immunological, environmental, and phylogenetic perpetrators of metastatic leishmaniasis. Trends Parasitol 30:412-22.
7. Lye LF, Owens K, Shi H, Murta SM, 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.
8. 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.
9. 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 108:1513-8.

10. Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, Sisk E, Rajandream MA, Adlem E, Aert R, Anupama A, Apostolou Z, Attipoe P, Bason N, Bauser C, Beck A, Beverley SM, Bianchettin G, Borzym K, Bothe G, Bruschi CV, Collins M, Cadag E, Ciarloni L, Clayton C, Coulson RM, 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 309:436-42.
11. 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:W29-34.
12. Iantorno SA, Durrant C, Khan A, Sanders MJ, Beverley SM, Warren WC, Berriman M, Sacks DL, Cotton JA, Grigg ME. 2017. Gene Expression in *Leishmania* Is Regulated Predominantly by Gene Dosage. mBio 8.
13. Albanaz ATS, Gerasimov ES, Shaw JJ, Sadlova J, Lukes J, Volf P, Opperdoes FR, Kostygov AY, Butenko A, Yurchenko V. 2021. Genome Analysis of Endotrypanum and Porcisia spp., Closest Phylogenetic Relatives of *Leishmania*, Highlights the Role of Amastins in Shaping Pathogenicity. Genes (Basel) 12.
14. Butenko A, Kostygov AY, Sadlova J, Kleschenko Y, Becvar T, Podesvova L, Macedo DH, Zihala D, Lukes J, Bates PA, Volf P, Opperdoes FR, Yurchenko V. 2019. Comparative genomics of *Leishmania* (Mundinia). BMC Genomics 20:726.

Table 1. Description of *Leishmania* species and strains including assembly parameters and links.

Subgenus	Species	strain ID	WHO Code	Source	Provenance	NCBI GenBank Bioproject	Assembly (bp)	pseudo-chromo-somes	N50 ctg length	protein coding genes	other assemblies
L. Leishmania	*L. major*	SD75.1 (clone)	MHOM/SN/74/SD	human, cutaneous	D. Sacks, Bethesda USA	PRJNA50303	31,727,271	36	95,380	8,818	
		LV39 clone5	MHOM/Sv/59/P	human, cutaneous	R. Titus, Boston, USA	PRJNA50301	31,961,985	36	71,452	8,971	
	L. gerbili	LEM452	MRHO/CN/60/GERBILLI	gerbil	P. Bastien, J-P Dedet, Montpellier, France	PRJNA192717	30,822,621	36	57,008	8,599	
	L. turanica	LEM563	MMEL/SU/79/MEL	gerbil	P. Bastien, J-P Dedet, Montpellier, France	PRJNA192712	30,876,294	36	39,210	8,608	
	L. arabica	LEM1108	MPSA/SA/83/JISH220	*Psammomys*	P. Bastien, J-P Dedet, Montpellier, France	PRJNA192710	30,774,332	36	52,119	8,646	
	L. tropica	L590	MHOM/IL/1990/P283	human, cutaneous	C. Jaffe, Jerusalem, Israel	PRJNA169676	31,326,083	36	32,739	8,824	(12)
	L. aethiopica	L147	MHOM/ET/1972/L100	human, diffuse cutaneous, relapsing	C. Jaffe, Jerusalem, Israel	PRJNA169673	31,026,739	36	38,498	8,722	
L. Viannia	*L.braziliensis*	M2903	MHOM/BR/72/M2903	human, cutaneous	J. Shaw, Brasil	PRJNA165955	32,590,753	35	61,918	9,269	
	L. panamensis	L13	MHOM/COL/81/L13	human, mucosal	N. Saravia, Cali, Colombia	PRJNA165959	31,108,242	35	22,576	8,665	
L. Mundinia	*L. enriettii*	LEM3045	MCAV/BR/95/CUR3	*Cavia porcellus* guinea pig	P. Bastien, J-P Dedet, Montpellier, France	PRJNA192711	30,427,298	36	102,666	8,731	(14)
	L. martiniquensis	LEM2494	MHOM/MQ/92/MAR1	human, diffuse cutaneous, HIV	P. Bastien, J-P Dedet, Montpellier, France	PRJNA192703	30,528,357	36	147,290	8,483	(14)
Endotrypanum	*Endotrypanum monterogei*	LV88	none	*Choloepus hoffmani* sloth	Michael Chance, Liverpool UK	PRJNA165953	32,086,870	36	33,059	8,285	(13)
Crithidia	*Crithidia fasciculata*	Cf-C1 (clone)	none	mosquito	Larry Simpson, Los Angeles, USA	PRJNA165885	41,297,378	30	778,443	9,619	(8)