

 Open access • Posted Content • DOI:10.1101/2021.05.29.446295

Genome assemblies across the diverse evolutionary spectrum of Leishmania protozoan parasites — [Source link](#)

Wesley C. Warren, Natalia S. Akopyants, Deborah E. Dobson, Christiane Hertz-Fowler ...+10 more authors

Institutions: University of Missouri, Washington University in St. Louis, University of Liverpool, University of Washington ...+3 more institutions

Published on: 29 May 2021 - bioRxiv (Cold Spring Harbor Laboratory)

Topics: Crithidia

Related papers:

- [Genome Assemblies across the Diverse Evolutionary Spectrum of Leishmania Protozoan Parasites.](#)
- [Genome Analysis of Endotrypanum and Porcisia spp., Closest Phylogenetic Relatives of Leishmania, Highlights the Role of Amastins in Shaping Pathogenicity.](#)
- [Comparative genomics of Leishmania \(Mundinia\).](#)
- [Genome of Leptomonas pyrrhocoris: a high-quality reference for monoxenous trypanosomatids and new insights into evolution of Leishmania](#)
- [Genetic Data Showing Evolutionary Links between Leishmania and Endotrypanum](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/genome-assemblies-across-the-diverse-evolutionary-spectrum-43sg9u3mlh>

1 July 29, 2021

revised

2

3

4 **Genome assemblies across the diverse evolutionary spectrum**

5 **of *Leishmania* protozoan parasites**

6

7 Wesley C. Warren¹, Natalia S. Akopyants², Deborah E. Dobson², Christiane Hertz-Fowler³,

8 Lon-Fye Lye², Peter J. Myler^{4,5}, Gowthaman Ramasamy⁴, Achchuthan Shanmugasundram³,

9 Fatima Silva-Franco³, Sascha Steinbiss⁶, Chad Tomlinson⁷, Richard K. Wilson⁸,

10 and Stephen M. Beverley^{2,*}

11

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

14 ² Department of Molecular Microbiology, Washington University School of Medicine, St Louis,
15 MO; Washington University School of Medicine, St Louis, MO.

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

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

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

22 ⁶ Wellcome Sanger Institute, Hinxton, Cambridgeshire, UK, CB10 1SA

23 ⁷ McDonnell Genome Institute, Washington University School of Medicine, St Louis, MO; USA

24 ⁸ The Steve and Cindy Rasmussen Institute for Genomic Medicine, Nationwide Children's
25 Hospital, Columbus, Ohio, USA.

26

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

29 stephen.beverley@wustl.edu.

30

31 **Keywords:** *Leishmania*, *Viannia*, *Mundinia*, comparative genomics, evolution of parasitism,
32 Kinetoplastida, Trypanosomatidae.

33

34 Running title: Genome assemblies across *Leishmania* phylogeny

35

36 **Abstract**

37

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

42

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

52 Parasites were cultivated in M199 or Schneider's medium (6) and grown to late log
53 phase, before harvesting, lysis, and DNA purification by phenol/chloroform extraction and/or
54 banding in CsCl gradients (to remove mitochondrial maxi- or minicircle DNA). Sequencing
55 libraries were generated using the Illumina paired-end DNA sample preparation kit (PE-102-
56 1001) according to the manufacturer's directions. Fragment libraries of 3 and 8kb were prepared
57 using protocols for 454 sequencing (Roche Life Sciences). Sequencing was performed on either a
58 454 GS FLX Titanium (average read length of 305bp; Roche 454 Life Sciences), or Illumina
59 GAIIX and HiSeq2000 instruments (paired end 100bp read length), except for *Crithidia* which
60 additionally utilized long-reads generated on a RSII instrument (P5/C3 chemistry; Pacific
61 BioSciences) (7). Total sequence genome coverage on the Illumina GAIIX or HiSeq 2000
62 instrument was on average 105x with tiered library insert sizes (50x fragments, 45x 3kb, and 10x
63 8kb and 0.05x 40kb). For all Illumina sequences we used the read processing steps within the
64 ALLPATHS-LG (8) software prior to *de novo* assembly, which incorporates read error correction
65 methods described by Pevzner *et al* (9). Genome assemblies were performed with default
66 parameters using Newbler v2.0.1 (10) for 454 reads, ALLPATHS-LG (8) for Illumina reads, or
67 HGAP version 3 (11) for long-reads (Table 1). Contigs and scaffolds were organized into
68 pseudochromosomes using ABACAS2 (<https://github.com/satta/ABACAS2>), a successor to
69 ABACAS (12), by alignment to the *Leishmania major* Friedlin genome, with the exception of *L.*

70 *braziliensis* M2903 and *L. panamensis*, which were aligned to the *L. braziliensis* M2904 genome.

71 The estimated haploid genome sizes ranged from 30.4 to 41.3 megabases (13).

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

77

78 **Public availability and accession number(s).** Assemblies are deposited in the NCBI GenBank
79 Repository under the Bioproject numbers in Table 1, including links to the primary data and
80 annotations ([PRJNA50303](https://ncbi.nlm.nih.gov/bioproject/PRJNA50303), [PRJNA50301](https://ncbi.nlm.nih.gov/bioproject/PRJNA50301), [PRJNA192717](https://ncbi.nlm.nih.gov/bioproject/PRJNA192717), [PRJNA192712](https://ncbi.nlm.nih.gov/bioproject/PRJNA192712), [PRJNA192710](https://ncbi.nlm.nih.gov/bioproject/PRJNA192710),
81 [PRJNA169676](https://ncbi.nlm.nih.gov/bioproject/PRJNA169676), [PRJNA169673](https://ncbi.nlm.nih.gov/bioproject/PRJNA169673), [PRJNA165955](https://ncbi.nlm.nih.gov/bioproject/PRJNA165955), [PRJNA165959](https://ncbi.nlm.nih.gov/bioproject/PRJNA165959), [PRJNA192711](https://ncbi.nlm.nih.gov/bioproject/PRJNA192711), [PRJNA192703](https://ncbi.nlm.nih.gov/bioproject/PRJNA192703),
82 [PRJNA165953](https://ncbi.nlm.nih.gov/bioproject/PRJNA165953), [PRJNA165885](https://ncbi.nlm.nih.gov/bioproject/PRJNA165885)). Chromosome builds are available through the TriTrypDB
83 portal (<http://tritrypdb.org/tritrypdb>).

84

85 **Acknowledgements.**

86 NIH grant AI29646 to Stephen M. Beverley provided support for NSA, DED and L-FL, NIH-
87 NHGRI grant HG00307907 to Richard K. Wilson supported WCW, and CT, and AI103858 to
88 Peter J Myler supported GA. Funding to Christiane Hertz-Fowler from Wellcome Trust grants
89 WT099198MA and WT108443MA provided support for FS-F, AS and SS. We thank Patrick
90 Minx for assembly curation. We thank the colleagues listed in Table 1 for generously providing
91 strains and Brian Brunk, David Roos, Thomas Otto and Matt Berriman for discussions. The WU
92 Institutional Biosafety Committee reviewed and approved the parasite work reported here (01-
93 015).

94

95 **References.**

- 96
- 97 1. Pigott DM, Bhatt S, Golding N, Duda KA, Battle KE, Brady OJ, Messina JP, Balard
98 Y, Bastien P, Pratlong F, Brownstein JS, Freifeld CC, Mekaru SR, Gething PW,
99 George DB, Myers MF, Reithinger R, Hay SI. 2014. Global distribution maps of the
100 leishmaniases. *Elife* 3:e28051.
- 101 2. Alvar J, Velez ID, Bern C, Herrero M, Desjeux P, Cano J, Jannin J, den Boer M.
102 2012. Leishmaniasis worldwide and global estimates of its incidence. *PLoS One*
103 7:e35671.
- 104 3. Banuls AL, Bastien P, Pomares C, Arevalo J, Fisa R, Hide M. 2011. Clinical
105 pleiomorphism in human leishmaniases, with special mention of asymptomatic
106 infection. *Clin Microbiol Infect* 17:1451-61.
- 107 4. Singh OP, Hasker E, Sacks D, Boelaert M, Sundar S. 2014. Asymptomatic
108 Leishmania infection: a new challenge for *Leishmania* control. *Clin Infect Dis*
109 58:1424-9.
- 110 5. Espinosa OA, Serrano MG, Camargo EP, Teixeira MMG, Shaw JJ. 2018. An
111 appraisal of the taxonomy and nomenclature of trypanosomatids presently classified
112 as *Leishmania* and *Endotrypanum*. *Parasitology* 145:430-442.
- 113 6. Lye LF, Owens K, Shi H, Murta SM, Vieira AC, Turco SJ, Tschudi C, Ullu E,
114 Beverley SM. 2010. Retention and loss of RNA interference pathways in
115 trypanosomatid protozoans. *PLoS Pathog* 6:e1001161.
- 116 7. Filosa JN, Berry CT, Ruthel G, Beverley SM, Warren WC, Tomlinson C, Myler PJ,
117 Dudkin EA, Povelones ML, Povelones M. 2019. Dramatic changes in gene
118 expression in different forms of *Crithidia fasciculata* reveal potential mechanisms
119 for insect-specific adhesion in kinetoplastid parasites. *PLoS Negl Trop Dis*
120 13:e0007570.
- 121 8. Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T,
122 Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L,
123 Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB. 2011. High-quality draft

- 124 assemblies of mammalian genomes from massively parallel sequence data. Proc Natl
125 Acad Sci U S A 108:1513-8.
- 126 9. Pevzner PA, Tang H, Waterman MS. 2001. An Eulerian path approach to DNA
127 fragment assembly. Proc Natl Acad Sci U S A 98:9748-53.
- 128 10. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J,
129 Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV,
130 Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie
131 TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li
132 J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW,
133 Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons
134 JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, et
135 al. 2005. Genome sequencing in microfabricated high-density picolitre reactors.
136 Nature 437:376-80.
- 137 11. Chaisson MJ, Huddleston J, Dennis MY, Sudmant PH, Malig M, Hormozdiari F,
138 Antonacci F, Surti U, Sandstrom R, Boitano M, Landolin JM, Stamatoyannopoulos
139 JA, Hunkapiller MW, Korlach J, Eichler EE. 2015. Resolving the complexity of the
140 human genome using single-molecule sequencing. Nature 517:608-11.
- 141 12. Assefa S, Keane TM, Otto TD, Newbold C, Berriman M. 2009. ABACAS:
142 algorithm-based automatic contiguation of assembled sequences. Bioinformatics
143 25:1968-9.
- 144 13. Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, Sisk E,
145 Rajandream MA, Adlem E, Aert R, Anupama A, Apostolou Z, Attipoe P, Bason N,
146 Bauser C, Beck A, Beverley SM, Bianchettin G, Borzym K, Bothe G, Bruschi CV,
147 Collins M, Cadag E, Ciarloni L, Clayton C, Coulson RM, Cronin A, Cruz AK,
148 Davies RM, De Gaudenzi J, Dobson DE, Duesterhoeft A, Fazelina G, Fosker N,
149 Frasc AC, Fraser A, Fuchs M, Gabel C, Goble A, Goffeau A, Harris D, Hertz-
150 Fowler C, Hilbert H, Horn D, Huang Y, Klages S, Knights A, Kube M, Larke N,
151 Litvin L, et al. 2005. The genome of the kinetoplastid parasite, *Leishmania major*.
152 Science 309:436-42.

- 153 14. Steinbiss S, Silva-Franco F, Brunk B, Foth B, Hertz-Fowler C, Berriman M, Otto
154 TD. 2016. Companion: a web server for annotation and analysis of parasite
155 genomes. *Nucleic Acids Res* 44:W29-34.

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

Subgenus	Species	strain ID	WHO Code	Source	Provenance	NCBI GenBank Bioproject ¹	Sequencing Platforms	Assembler	Contigs	N50 contig length	Assembly size (bp)	%GC	Pseudo-chromosomes	Protein coding genes	Other assemblies
<i>L. Leishmania</i>	<i>L. major</i>	SD75.1 (clone)	MHOM/SN/74/SD	human, cutaneous	D. Sacks, Bethesda USA	PRJNA50303	454, Illumina	Newbler	891	95,380	31,727,271	59.3	36	8,818	
		LV39 (clone5)	MHOM/Sv/59/P	human, cutaneous	R. Titus, Boston, USA	PRJNA50301	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	PRJNA192717	454, Illumina	Allpaths	1,248	57,008	30,822,621	59.6	36	8,599	
	<i>L. turanica</i>	LEM563	MMEL/SU/79/MEL	gerbil	P. Bastien, J-P Dedet, Montpellier, France	PRJNA192712	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> human, cutaneous	P. Bastien, J-P Dedet, Montpellier, France	PRJNA192710	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, diffuse cutaneous	C. Jaffe, Jerusalem, Israel	PRJNA169676	454, Illumina	Allpaths	1,938	32,739	31,326,083	59.6	36	8,824	(12)	
<i>L. aethiopica</i>	L147	MHOM/ET/1972/L100	relapsing cutaneous, human	C. Jaffe, Jerusalem, Israel	PRJNA169673	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/72/M2903	human, cutaneous	J. Shaw, Brasil	PRJNA165955	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	PRJNA165959	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	PRJNA192711	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	PRJNA192703	454, Illumina	Allpaths	628	147,290	30,528,357	59.6	36	8,483	(14)
<i>Endotrypanum</i>	<i>Endotrypanum monterogeii</i>	LV88	none	<i>Chaloepeus hoffmani</i> sloth	Michael Chance, Liverpool UK	PRJNA165953	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, USA	PRJNA165885	454, Illumina, PacBio	HGAP	494	778,443	41,297,378	57.0	30	9,619	(8)

Footnote 1 Each bioproject link contains links to the current assembly, primary data sets, and other relevant information.