AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Chromosome-Level Genome Sequence of *Leishmania* (*Leishmania*) *tropica* Strain CDC216-162, Isolated from an Afghanistan Clinical Case

Yvette Unoarumhi,[e,f] Dhwani Batra,[a] Mili Sheth,[a] Vidhya Narayanan,[b,c] Wuling Lin,[b,c] Yueli Zheng,[b,d] Lori A. Rowe,[a] Jan Pohl,[a] Marcos de Almeida[b]

[a]Biotechnology Core Facility Branch, Division of Scientific Resources, Centers for Disease Control and Prevention, Atlanta, Georgia, USA
[b]Reference Diagnostic Laboratory, Center for Global Health, Division of Parasitic Diseases and Malaria, Centers for Disease Control and Prevention, Atlanta, Georgia, USA
[c]IHRC, Inc., Atlanta, Georgia, USA
[d]Eagle Global Scientific, San Antonio, Texas, USA
[e]Oak Ridge Institute for Science and Education (ORISE), Oak Ridge, Tennessee, USA
[f]Association of Public Health Laboratories (APHL), Silver Spring, Maryland, USA

ABSTRACT PacBio and Illumina MiSeq platforms were used for genomic sequencing of a *Leishmania (Leishmania) tropica* strain isolated from a patient infected in Pakistan. PacBio assemblies were generated using Flye v2.4 and polished with MiSeq data. The results represent a considerable improvement of the currently available genome sequences in the GenBank database.

Leishmaniasis is a vector-born disease caused by >20 species of parasites in the genus *Leishmania* which affects millions of people worldwide and causes thousands of deaths yearly (1, 2). The spectrum of the disease includes the following three main clinical manifestations, depending on the *Leishmania* species: (i) cutaneous leishmaniasis (CL), the most common clinical manifestation, characterized by development of ulcerative cutaneous lesions; (ii) mucocutaneous leishmaniasis (MCL), which can cause permanent disfiguration of oral and nasopharyngeal mucosa or death; and (iii) visceral leishmaniasis (VL) that is fatal in most untreated cases (3–5). Recently, genome sequence data have been used to determine phylogenetic relationships between *Leishmania* parasites, providing insights into parasite classification, genetic polymorphism, virulence, and drug resistance and also supporting the development of specific diagnostics for preventing the morbidity and mortality of the disease (17–20).

In this study, we present a chromosome-level genome sequence of *Leishmania* (*Leishmania*) *tropica*, an etiological agent of CL endemic in the Middle East, northern Africa at the Mediterranean Sea, and some parts of Asia. A skin biopsy specimen obtained from an American traveler in Afghanistan infected with *Leishmania* (*L.*) *tropica* was cultured in 10% fetal bovine serum (FBS)-RPMI axenic medium (Life Technologies, CA). PacBio and Illumina MiSeq libraries were prepared with DNA from cultured parasites using the MagAttract high-molecular-weight (HMW) DNA kit (Qiagen, MD) as previously described (6–8). PacBio libraries were prepared using the PacBio 20-kb kit (Pacific Biosciences, CA), size selected with BluePippin (Sage Science, MA), and sequenced with C4v2 chemistry for 360-minute movies on the RS II instrument (Pacific Biosciences). Dual-indexed libraries were prepared with the NEBNext Ultra library prep kit (New England BioLabs, MA) and sequenced using the MiSeq 2 × 250-cycle sequencing kit (Illumina, CA). PacBio filtered reads (minlength = 1,000; number of reads = 293,219; average length = 9,974) were *de novo* assembled using Flye v2.4 (9–13) (-g 32m).

The *de novo* assembly comprised 88 fragments which were scaffolded using Companion v1.0.1 (14) (default parameters), using *Leishmania* (*L.*) *major* as the reference genome. The

Citation Unoarumhi Y, Batra D, Sheth M, Narayanan V, Lin W, Zheng Y, Rowe LA, Pohl J, de Almeida M. 2021. Chromosome-level genome sequence of *Leishmania* (*Leishmania*) *tropica* strain CDC216-162, isolated from an Afghanistan clinical case. Microbiol Resour Announc 10:e00842-20. https://doi.org/10.1128/MRA.00842-20.

Editor Jason E. Stajich, University of California, Riverside

Address correspondence to Marcos de Almeida, MdeAlmeida@cdc.gov.

Received 11 August 2020
Accepted 12 April 2021
Published 20 May 2021

TABLE 1 Comparison of proposed assemblies of *Leishmania* (*L.*) *tropica*

L. (*L.*) *tropica* isolate	Read type	Total genome length (bp)	No. of scaffolds	No. of contigs	N_{50} length (bp)	L_{50} count	G+C content (%)	Coverage (×)	GenBank accession no.
L590	Illumina MiSeq	32,989,014	448	1,938	32,739	275	59.30	253	GCA_000410715.1
ATCC 50129	Illumina MiSeq	30,870,161	1,928	1,928	32,161	274	59.90	150	GCA_011316065.1
WHO2017-IK[a]	Illumina MiSeq	32,139,927	9,499	20,898	3,299	20,898	57.20	150	GCA_003067545.1
WHO2015-IK[b]	Illumina HiSeq	32,280,712	17,013	18,935	6,327	1,404	59.60	150	GCA_003352575.1
CDC216-162	PacBio, Illumina MiSeq	32,702,166	36 + bin	36 + bin	1,070,514	11	59.61	75 (PacBio RS II CLR), 100 (Illumina)	GCA_014139745.1

[a] MHOM/LB/2017/IK.
[b] MHOM/LB/2015/IK.

resultant scaffolds were examined for correct orderings and orientation using Mauve, and the misassembled contigs were manually curated and fixed. To improve the quality and contiguity of scaffolds, PacBio reads were mapped to chromosome scaffolds generated by Companion, mapped reads were binned by chromosomes, and each bin was assembled independently using Flye v 2.4 (-g 2m), resulting in 36 scaffolds. These scaffolds were processed again using Companion as described above, producing 36 contigs representing each of the expected 36 chromosomes. Additionally, one round of polishing with Illumina reads was performed using POLCA (15) to correct inherent PacBio sequencing errors. The final assembly was evaluated for completeness by BUSCO 4.0.6 (16) against euglenozoa_odb10 (completeness [C], 100.0%; complete and duplicated [D], 0.0%; complete and single-copy [S], 100.0%; fragmented [F], 0.0%; missing [M], 0.0%; number of genes used [n], 130). Default software parameters were used in this study, except where otherwise noted.

High-quality contiguous genomic sequences are crucial for more robust background and investigations in the leishmaniasis field, which are essential for the development of more species-specific diagnostic tests and clinical management. Therefore, as shown in Table 1, the proposed assembly with fully resolved *Leishmania* (*L.*) *tropica* chromosomes presented here represents a considerable improvement over the existing assemblies.

Data availability. This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number JACCHM000000000 and under BioProject number PRJNA484340. The version described in this paper is the first version, JACCHM010000000. The accession numbers for raw reads are SRR7867286, SRR7867287, SRR7867288, SRR7867289, SRR7867290, SRR7867291, and SRR10771526.

ACKNOWLEDGMENTS

This work was made possible through support from the CDC Advanced Molecular Detection (AMD) program, Association of Public Health Laboratories (APHL), and CDC Division of Scientific Resources.

This study was reviewed by the internal CDC board. The clinical specimen was used in accordance with a CDC-approved protocol for use of residual diagnostic specimens from humans for laboratory methods via research protocol number 6756.

REFERENCES

1. Alvar J, Velez ID, Bern C, Herrero M, Desjeux P, Cano J, Jannin J, den Boer M, Team 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.
2. 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.
3. David CV, Craft N. 2009. Cutaneous and mucocutaneous leishmaniasis. Dermatol Ther 22:491–502. https://doi.org/10.1111/j.1529-8019.2009.01272.x.
4. Diniz JL, Costa MO, Goncalves DU. 2011. Mucocutaneous leishmaniasis: clinical markers in presumptive diagnosis. Braz J Otorhinolaryngol 77:380–384. https://doi.org/10.1590/S1808-86942011000300018.
5. Ronet C, Beverley SM, Fasel N. 2011. Muco-cutaneous leishmaniasis in the New World: the ultimate subversion. Virulence 2:547–552. https://doi.org/10.4161/viru.2.6.17839.
6. Batra D, Lin W, Narayanan V, Rowe LA, Sheth M, Zheng Y, Loparev V, de Almeida M. 2019. Draft genome sequences of Leishmania (Leishmania) amazonensis, Leishmania (Leishmania) mexicana, and Leishmania (Leishmania) aethiopica, potential etiological agents of diffuse cutaneous leishmaniasis. Microbiol Resour Announc 8:e00269-19. https://doi.org/10.1128/MRA.00269-19.
7. Batra D, Lin W, Rowe LA, Sheth M, Zheng Y, Loparev V, de Almeida M. 2018. Draft genome sequence of French Guiana Leishmania (Viannia) guyanensis strain 204–365, assembled using long reads. Microbiol Resour Announc 7:e01421-18. https://doi.org/10.1128/MRA.01421-18.
8. Lin W, Batra D, Narayanan V, Rowe LA, Sheth M, Zheng Y, Juieng P, Loparev V, de Almeida M. 2019. First draft genome sequence of Leishmania

(Viannia) lainsoni strain 216–34, isolated from a Peruvian clinical case. Microbiol Resour Announc 8:e01524-18. https://doi.org/10.1128/MRA.01524-18.
9. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8.
10. Lin Y, Yuan J, Kolmogorov M, Shen MW, Chaisson M, Pevzner PA. 2016. Assembly of long error-prone reads using de Bruijn graphs. Proc Natl Acad Sci U S A 113:E8396–E8405. https://doi.org/10.1073/pnas.1604560113.
11. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191.
12. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
13. Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27:2987–2993. https://doi.org/10.1093/bioinformatics/btr509.
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.
15. Zimin AV, Puiu D, Luo MC, Zhu T, Koren S, Yorke JA, Dvorak J, Salzberg S. 2017. Hybrid assembly of the large and highly repetitive genome of Aegilops tauschii, a progenitor of bread wheat, with the mega-reads algorithm. Genome Res 27:787–792. https://doi.org/10.1101/gr.213405.116.
16. Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness. In Kollmar M (ed), Gene Prediction. Methods in molecular biology, vol 1962. Humana, New York, NY.
17. Bennis I, Belaid L, De Brouwere V, Filali H, Sahibi H, Boelaert M. 2017. "The mosquitoes that destroy your face". Social impact of cutaneous leishmaniasis in south-eastern Morocco, a qualitative study. PLoS One 12:e0189906. https://doi.org/10.1371/journal.pone.0189906.
18. Bennis I, Thys S, Filali H, De Brouwere V, Sahibi H, Boelaert M. 2017. Psychosocial impact of scars due to cutaneous leishmaniasis on high school students in Errachidia province, Morocco. Infect Dis Poverty 6:46. https://doi.org/10.1186/s40249-017-0267-5.
19. Aronson N, Herwaldt BL, Libman M, Pearson R, Lopez-Velez R, Weina P, Carvalho E, Ephros M, Jeronimo S, Magill A. 2017. Diagnosis and treatment of leishmaniasis: clinical practice guidelines by the Infectious Diseases Society of America (IDSA) and the American Society of Tropical Medicine and Hygiene (ASTMH). Am J Trop Med Hyg 96:24–45. https://doi.org/10.4269/ajtmh.16-84256.
20. Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res 27:737–746. https://doi.org/10.1101/gr.214270.116.