Chromosome-Scale Assembly of the Complete Genome Sequence of *Leishmania* (*Mundinia*) sp. Ghana, Isolate GH5, Strain LV757

Hatim Almutairi,[a,b] Michael D. Urbaniak,[a] Michelle D. Bates,[a] Godwin Kwakye-Nuako,[c] Waleed S. Al-Salem,[b] Rod J. Dillon,[a] Paul A. Bates,[a] Derek Gatherer[a]

[a]Division of Biomedical & Life Sciences, Faculty of Health & Medicine, Lancaster University, Lancaster, United Kingdom
[b]Ministry of Health, Riyadh, Saudi Arabia
[c]Department of Biomedical Sciences, School of Allied Health Sciences, College of Health & Allied Sciences, University of Cape Coast, Cape Coast, Ghana

ABSTRACT *Leishmania* (*Mundinia*) sp. Ghana is a kinetoplastid parasite isolated in 2015 in Ghana. We report the complete genome sequence of *L.* (*M.*) sp. Ghana, sequenced using combined short-read and long-read technologies. This will facilitate greater understanding of this novel pathogen and its relationships within the subgenus *Mundinia*.

In 2015, a hitherto unknown parasite of the genus *Leishmania* was detected in a case of human cutaneous leishmaniasis in Ghana (1). This putative new species has not yet been formally named but was classified as the fifth species in the recently established subgenus *Mundinia* (2, 3), which also includes *Leishmania enriettii* (2), *Leishmania macropodum* (4), *Leishmania orientalis* (5), and *Leishmania martiniquensis* (6). Phylogenetic analyses indicate that the subgenus *Mundinia* is the sister group to the other *Leishmania* subgenera (1, 7). Furthermore, *Mundinia* species are found on every continent except Antarctica (8), supporting the hypothesis of evolution from a common ancestor prior to the division of the Gondwana supercontinent (9). We report the complete genome assembly and annotation of *L.* (*M.*) sp. Ghana, isolate GH5, strain LV757 (WHO code MHOM/GH/2012/GH5;LV757). This will contribute to research on the origins and expansion of *Mundinia*.

Parasites were grown using an *in vitro* culture system previously developed for *L.* (*M.*) *orientalis* axenic amastigotes (10) in Schneider's insect medium at 26°C as promastigotes, then in M199 medium supplemented with 10% fetal calf serum (FCS), 2% stable human urine, 1% basal medium Eagle vitamins, and 25 μg/ml gentamicin sulfate, with subpassage to fresh medium every 4 days to sustain parasite growth and viability. DNA was extracted and purified using a Qiagen DNeasy blood and tissue kit using the spin column protocol, according to the manufacturer's instructions. The extracted DNA concentration was assessed using a Qubit fluorometer, microplate reader, and agarose gel electrophoresis. All sequencing libraries were based on the same extracted DNA sample to avoid any inconsistency.

Short-read library construction and sequencing were contracted to (i) BGI (Shenzhen, China), who constructed DNBSEQ libraries producing paired-end reads (270 bp and 500 bp) using the Illumina HiSeq platform, and (ii) Aberystwyth University (Aberystwyth, UK), who constructed TruSeq Nano DNA libraries producing paired-end reads (300 bp) using the Illumina MiSeq platform. We performed long-read library preparation and sequencing according to the Nanopore protocol (SQK-LSK109) on R9 flow cells (FLO-MIN106). The read quality was assessed using MultiQC (11), incorporating the use of FastQC for the Illumina short reads and pycoQC for the Nanopore long reads.

We assembled the long reads using Flye (12), with default parameters, to generate chromosome-scale scaffolds. Then, using Minimap2 (13) and SAMtools (14), we mapped the short reads onto the assembled scaffolds to compensate for erroneous bases within the long reads and create consensus sequences. After polishing the assembly using Pilon (15), another

A. Number of chromosomes and unplaced contigs

B. N50 and length of longest contig/scaffold

C. Genome length

- L. (M.) sp. Ghana LV757 (GCA_017918215.1- this paper)
- L. (M.) enriettii LEM3045 (GCA_000410755.2)
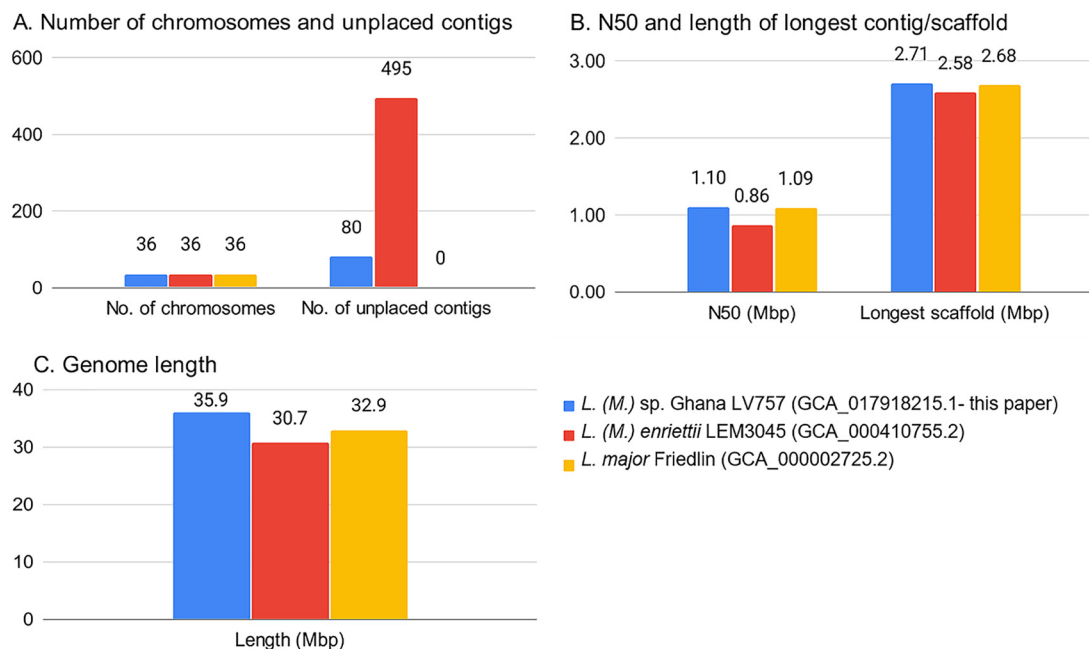- L. major Friedlin (GCA_000002725.2)

FIG 1 Assembly comparison of *L. (M.) sp.* Ghana LV757 with *L. (M.) enriettii* LEM3045 and *L. major* Friedlin.

round of consensus short-read mapping was performed. Then, we removed duplicated contigs and sorted the remainder according to length using Funannotate (16). Finally, we separated the chimeric sequences and performed scaffolding using RaGOO (17), with the *L. major* Friedlin strain genome (GenBank accession number GCA_000002725.2) (18) as a reference guide, aligning all 36 chromosomes for our assembly, thereby also determining the chromosome ends to be complete, with the exception of 80 unplaced contigs totaling 1,077,537 bp.

The analysis workflow for assembly and annotation was performed using Snakemake (19) and is available online for reproducibility purposes (https://github.com/hatimalmutairi/LGAAP), including the software versions and parameters used (20). Figure 1 compares our assembly with other complete genome sequences.

We assessed the assembly completeness using BUSCO (21), with the lineage data set for the phylum *Euglenozoa*, containing 130 single-copy orthologs from 31 species, and found 123 of these to be present (94.6% completeness). We carried out functional

TABLE 1 Detailed summary metrics of the genome sequencing, assembly, and annotation for *L. (M.) sp.* Ghana LV757

Feature(s)	Metric(s)
Total no. of reads	49,308,106
No. of MiSeq reads	5,195,324
No. of HiSeq reads	43,244,422
MinION reads (bp)	868,360
MinION read N_{50} (bp)	19,170
Bases (Gb)	26.93
Genome coverage ($\times$)	371.2
Total no. of scaffolds	116
Genome size (bp)	35,953,538
N_{50} (bp)	1,100,365
% GC content	59.70
No. of Ns (% of genome)	481 (0.001)
No. of genes	8,119
Gene density (Mb)	225.8
No. of exons	8,119
Mean gene length (bp)	1,838
Total length of CDSs[a] (Mb) (% of genome)	14.92 (41.51)

[a]CDSs, coding DNA sequences.

annotation and prediction using the MAKER2 (22) annotation pipeline in combination with AUGUSTUS (23) gene prediction software. Table 1 shows further summary metrics for the sequencing, assembly, and annotation.

ACKNOWLEDGMENT

REFERENCES

1. Kwakye-Nuako G, Mosore M-T, Duplessis C, Bates MD, Puplampu N, Mensah-Attipoe I, Desewu K, Afegbe G, Asmah RH, Jamjoom MB, Ayeh-Kumi PF, Boakye DA, Bates PA. 2015. First isolation of a new species of *Leishmania* responsible for human cutaneous leishmaniasis in Ghana and classification in the *Leishmania enriettii* complex. Int J Parasitol 45:679–684. https://doi.org/10.1016/j.ijpara.2015.05.001.
2. Muniz J, Medina H. 1948. Cutaneous leishmaniasis of the guinea pig, *Leishmania enriettii* n. sp. Hospital (Rio J) 33:7–25. (In Portuguese.)
3. 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.
4. Rose K, Curtis J, Baldwin T, Mathis A, Kumar B, Sakthianandeswaren A, Spurck T, Choy JL, Handman E. 2004. Cutaneous leishmaniasis in red kangaroos: isolation and characterisation of the causative organisms. Int J Parasitol 34:655–664. https://doi.org/10.1016/j.ijpara.2004.03.001.
5. Jariyapan N, Daroontum T, Jaiwong K, Chanmol W, Intakhan N, Sor-Suwan S, Siriyasatien P, Somboon P, Bates MD, Bates PA. 2018. *Leishmania* (*Mundinia*) *orientalis* n. sp. (Trypanosomatidae), a parasite from Thailand responsible for localised cutaneous leishmaniasis. Parasit Vectors 11:351. https://doi.org/10.1186/s13071-018-2908-3.
6. Desbois N, Pratlong F, Quist D, Dedet J-P. 2014. *Leishmania* (*Leishmania*) *martiniquensis* n. sp. (Kinetoplastida: Trypanosomatidae), description of the parasite responsible for cutaneous leishmaniasis in Martinique Island (French West Indies). Parasite 21:12. https://doi.org/10.1051/parasite/2014011.
7. Butenko A, Kostygov AY, Sádlová J, Kleschenko Y, Bečvář T, Podešvová L, Macedo DH, Žihala D, Lukeš J, Bates PA, Volf P, Opperdoes FR, Yurchenko V. 2019. Comparative genomics of *Leishmania* (*Mundinia*). BMC Genomics 20:726. https://doi.org/10.1186/s12864-019-6126-y.
8. Becvar T, Siriyasatien P, Bates P, Volf P, Sadlova J. 2020. Development of *Leishmania* (*Mundinia*) in guinea pigs. Parasit Vectors 13:181. https://doi.org/10.1186/s13071-020-04039-9.
9. Barratt J, Kaufer A, Peters B, Craig D, Lawrence A, Roberts T, Lee R, McAuliffe G, Stark D, Ellis J. 2017. Isolation of novel trypanosomatid, *Zelonia australiensis* sp. nov. (Kinetoplastida: Trypanosomatidae) provides support for a Gondwanan origin of dixenous parasitism in the *Leishmaniinae*. PLoS Negl Trop Dis 11:e0005215. https://doi.org/10.1371/journal.pntd.0005215.
10. Chanmol W, Jariyapan N, Somboon P, Bates MD, Bates PA. 2019. Axenic amastigote cultivation and *in vitro* development of *Leishmania orientalis*. Parasitol Res 118:1885–1897. https://doi.org/10.1007/s00436-019-06311-z.
11. Ewels P, Magnusson M, Lundin S, Kaller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32:3047–3048. https://doi.org/10.1093/bioinformatics/btw354.
12. 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.

13. Li H. 2016. Minimap and miniasm: fast mapping and *de novo* assembly for noisy long sequences. Bioinformatics 32:2103–2110. https://doi.org/10.1093/bioinformatics/btw152.
14. 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 format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
15. 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.
16. Li W-C, Wang T-F. 2021. PacBio long-read sequencing, assembly, and Funannotate reannotation of the complete genome of *Trichoderma reesei* QM6a. Methods Mol Biol 2234:311–329. https://doi.org/10.1007/978-1-0716-1048-0_21.
17. Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, Lippman ZB, Schatz MC. 2019. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol 20:224. https://doi.org/10.1186/s13059-019-1829-6.
18. 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, Bianchettin 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.
19. Mölder F, Jablonski KP, Letcher B, Hall MB, Tomkins-Tinch CH, Sochat V, Forster J, Lee S, Twardziok SO, Kanitz A, Wilm A, Holtgrewe M, Rahmann S, Nahnsen S, Köster J. 2021. Sustainable data analysis with Snakemake. F1000Res 10:33. https://doi.org/10.12688/f1000research.29032.2.
20. Almutairi H, Urbaniak MD, Bates MD, Jariyapan N, Kwakye-Nuako G, Thomaz-Soccol V, Al-Salem WS, Dillon RJ, Bates PA, Gatherer D. 2021. LGAAP: *Leishmaniinae* Genome Assembly and Annotation Pipeline. Microbiol Resour Announc 10:e00439-21. https://doi.org/10.1128/MRA.00439-21.
21. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351.
22. Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12:491. https://doi.org/10.1186/1471-2105-12-491.
23. Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a Web server for gene finding in eukaryotes. Nucleic Acids Res 32:W309–W312. https://doi.org/10.1093/nar/gkh379.