

This is a repository copy of *Genome sequence of Leishmania mexicana MNYC/BZ/62/M379 expressing Cas9 and T7 RNA polymerase.*

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/208220/>

Version: Published Version

---

**Article:**

Mottram, Jeremy Charles orcid.org/0000-0001-5574-3766 (2023) Genome sequence of *Leishmania mexicana* MNYC/BZ/62/M379 expressing Cas9 and T7 RNA polymerase. Wellcome Open Research. 294. ISSN 2398-502X

<https://doi.org/10.12688/wellcomeopenres.18575.2>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



DATA NOTE

**REVISED** **Genome sequence of *Leishmania mexicana***

**MNYC/BZ/62/M379 expressing Cas9 and T7 RNA polymerase**

[version 2; peer review: 3 approved]

Tom Beneke<sup>1,2</sup>, Ulrich Dobramysl<sup>3</sup>, Carolina Moura Costa Catta-Preta<sup>4</sup>,  
Jeremy Charles Mottram<sup>4</sup>, Eva Gluenz<sup>1,5</sup>, Richard J. Wheeler<sup>3</sup>

<sup>1</sup>Sir William Dunn School of Pathology, University of Oxford, Oxford, UK

<sup>2</sup>Department of Cell and Developmental Biology, Biocentre, University of Würzburg, Würzburg, 97074, Germany

<sup>3</sup>Peter Medawar Building for Pathogen Research, Nuffield Department of Medicine, University of Oxford, Oxford, UK

<sup>4</sup>York Biomedical Research Institute, Department of Biology, University of York, York, UK

<sup>5</sup>Institute of Cell Biology, University of Bern, Bern, Switzerland

**v2** **First published:** 05 Dec 2022, 7:294  
<https://doi.org/10.12688/wellcomeopenres.18575.1>  
**Latest published:** 23 Feb 2023, 7:294  
<https://doi.org/10.12688/wellcomeopenres.18575.2>

### Abstract

We present the genome sequence of *Leishmania mexicana* MNYC/BZ/62/M379 modified to express Cas9 and T7 RNA-polymerase, revealing high similarity to the reference genome (MHOM/GT2001/U1103). Through RNAseq-based annotation of coding sequences and untranslated regions, we provide primer sequences for construct and sgRNA template generation for CRISPR-assisted gene deletion and endogenous tagging.


### Keywords

*Leishmania mexicana*, Genome, Pilon polish

### Open Peer Review

**Approval Status** ✓✓✓

	1	2	3
<b>version 2</b> (revision) 23 Feb 2023		✓ view	
<b>version 1</b> 05 Dec 2022	✓ view	↑ ?	✓ view

1. **José María Requena** , Universidad Autónoma de Madrid, Madrid, Spain

2. **Richard McCulloch** , University of Glasgow, Glasgow, UK

3. **Stephen M. Beverley**, Washington University School of Medicine, St. Louis, USA

Any reports and responses or comments on the article can be found at the end of the article.

**Corresponding author:** Richard J. Wheeler ([richard.wheeler@ndm.ox.ac.uk](mailto:richard.wheeler@ndm.ox.ac.uk))

**Author roles:** **Beneke T:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Resources, Software, Writing – Review & Editing; **Dobramysl U:** Data Curation, Formal Analysis, Software, Validation, Writing – Review & Editing; **Catta-Preta CMC:** Investigation, Resources; **Mottram JC:** Conceptualization, Project Administration, Supervision, Writing – Review & Editing; **Gluezn E:** Funding Acquisition, Investigation, Project Administration, Supervision, Writing – Review & Editing; **Wheeler RJ:** Data Curation, Formal Analysis, Project Administration, Software, Supervision, Validation, Writing – Original Draft Preparation, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** This work was supported by Wellcome 221944, [<https://doi.org/10.35802/221944>]; 211075, [<https://doi.org/10.35802/211075>]; 200807, [<https://doi.org/10.35802/200807>] and an MRC PhD studentship [15/16\_MSD\_836338]. Eva Gluezn was supported by the Royal Society.

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2023 Beneke T *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Beneke T, Dobramysl U, Catta-Preta CMC *et al.* **Genome sequence of *Leishmania mexicana* MNYC/BZ/62/M379 expressing Cas9 and T7 RNA polymerase [version 2; peer review: 3 approved]** Wellcome Open Research 2023, **7**:294 <https://doi.org/10.12688/wellcomeopenres.18575.2>

**First published:** 05 Dec 2022, **7**:294 <https://doi.org/10.12688/wellcomeopenres.18575.1>

**REVISED Amendments from Version 1**

Changes made in this version in response to referee comments:

- We corrected a typo in the number of read base pairs we obtained during sequencing of the before-passage sample (Methods section, first paragraph).
- We explained what we mean by “small insertions” and “small deletions” detected by the Pilon polishing run (Methods section, third paragraph).
- We clarified that we used Pilon to polish the reference genome with our new reads from the Cas9/T7 strain (Methods section, third paragraph).
- We made clearer what we mean by “sequencing read coverage” (Methods section, fourth paragraph).

**Any further responses from the reviewers can be found at the end of the article**

**Introduction**

*Leishmania mexicana* is a human-infective unicellular eukaryote and one of the species which cause leishmaniasis. It is commonly used as a model *Leishmania* species for molecular cell biology due to its lower virulence (causing cutaneous rather than visceral leishmaniasis) and its ability to readily differentiate into the amastigote form in appropriate axenic culture. We have previously described the generation of a genetically modified *L. mexicana* MNYC/BZ/62/M379 expressing Cas9 and T7 RNA polymerase as a strain enabling for rapid reverse genetic modifications<sup>1</sup>. As this is not the reference genome strain (which is MHOM/GT/2001/U1103)<sup>2</sup> and may have accumulated mutations during laboratory culture and/or selection pressures of Cas9 or T7 expression, we sequenced the genome of this widely used strain as a high-quality reference for design of reverse genetic strategies.

**Methods**

We have previously confirmed that these promastigotes are infectious to the sandfly vector<sup>3</sup>. To ensure that the line was infectious to mammals, we infected an eight-week-old female BALB/c mouse footpad with stationary phase promastigotes ( $2.0 \times 10^6$ ); after four weeks we purified amastigotes from the excised resulting lesion, which were then back-transformed to promastigotes in axenic culture in M199 supplemented with 20% FCS and 50 µg/ml gentamycin (Roche) and grown for seven passages. This gave rise to the cell line *L. mex* Cas9 T7 M. Genomic DNA from before and after mouse passage was extracted using phenol-chloroform DNA extraction as previously described<sup>4</sup>. For Illumina sequencing, isolated DNA was diluted in 300 µL resuspension buffer (from Illumina TruSeq Nano DNA Library kit) and sonicated for nine (9) minutes using a Bioruptor 300 (Diagenode, set to low [0.5 interval]). The resulting 600 bp DNA fragments were processed for library construction using the TruSeq Nano DNA Library kit (Illumina). Sequencing followed on an Illumina NextSeq550 in paired-end mode (2×150nt) using a NextSeq 500/550 Mid Output 300 Cycles v2 Kit (Illumina, v2 kits now discontinued). 24,656,567 (7.4 Gb) and 26,015,449 reads (7.8 Gb) from

before and after passage were obtained respectively. For Nanopore sequencing, a library was constructed from the after-passage sample (not additionally sonicated) using the 1D<sup>2</sup> (SQK-LSK309) kit and sequenced on a MinION (FLO-MIN106 flow cell), obtaining 329,692 1D and 48,711 1D<sup>2</sup> reads (total, 2.3 Gb; mean length, 6.1 kb; read  $N_{50}$ , 8.0 kb; longest read, 448 kb).

We generated a trial *de novo* MNYC/BZ/62/M379 Cas9/T7 assembly using the Nanopore reads. Following adapter trimming using Porechop v0.2.4 we used a minimap2/racon/miniasm pipeline (v2.17-r974, v1.4.20 and v0.3-r179 respectively)<sup>5-7</sup>, then polished the assembly by mapping the adapter-trimmed (using TrimGalore! v0.6.0)<sup>8</sup> Illumina reads using BWA-MEM v0.7.17<sup>9</sup> and 10 rounds of polishing using Pilon v1.23<sup>10</sup> (total length, 31.4 Mb; contigs, 109; mean length, 291 kb;  $N_{50}$ , 640 kb; longest contig, 2.87 Mb). Synteny inspection using SyMAP<sup>11</sup> in comparison to the MHOM/GT/2001/U1103 reference genome<sup>2</sup> showed one (1) or two (2) contigs per chromosome (except for chromosomes 34, four (4) contigs; 19, three (3) contigs; 10, three (3) contigs; 8, four (4) contigs) and no evidence for chromosomal segmental deletion or duplication.

To simplify genome annotation, we therefore opted to polish the MHOM/GT/2001/U1103 genome (NCBI Genome Assembly GCA\_000234665.4) with the Illumina reads to generate a MNYC/BZ/62/M379 Cas9/T7 genome instead of using the *de novo* assembly. Following adapter trimming using TrimGalore! V0.6.0 (default settings) and removal of unfixable reads, the genome was polished by mapping the Illumina reads to the genome using BWA-MEM v0.7.17<sup>9</sup> and one (1) round of polishing using Pilon v1.23 fixing SNPs and indels<sup>10</sup>, identifying 21500 SNPs, 3828 small insertions and 4878 small deletions (as defined by Pilon<sup>10</sup>; mean insertion length 2.25 bp, mean deletion length 3.01 bp). The resulting updated and annotated genome sequence is available as a supplement<sup>12</sup>. Pilon, run in changes mode, identified only 193 SNPs and no changes in coding sequences following mouse passage. Note that neither T7 nor Cas9 are present in this polished genome as Cas9 is not chromosomally integrated (instead expressed from an episome<sup>1</sup>) and T7 is integrated into the highly repetitive 18S rRNA array which is collapsed in the reference genome.

Aneuploidy is known to be common among *Leishmania*<sup>2</sup>. Indeed, sequencing read coverage (how many times each nucleotide was sequenced) was not uniform across all chromosomes. Average read coverage throughout most chromosome was  $144 \pm 5$  (mean  $\pm$  sd.; diploid) excluding three outlier chromosomes: chromosome 3 (coverage, 218; triploid), 16 (coverage, 214; triploid) and 30 (coverage, 277; tetraploid).

Updated MHOM/GT/2001/U1103 genome annotations were prepared from existing resources, then transferred to MNYC/BZ/62/M379 Cas9/T7 accounting for coordinate changes due to indels. MHOM/GT/2001/U1103 ORFs and non-coding RNAs from TriTrypDB v50<sup>13</sup> were taken as the start

set. Previous RNAseq analysis<sup>14</sup> (BioProject accession number PRJEB8829) mapped spliced leader acceptor sites (SLASs, the site of trans splicing of a leader sequence common to all processed mRNAs) and polyadenylation sites (PASs) which define the bounds of the mRNA, from which suggested gene extensions and truncations were listed. We included these changes when a valid ORF (with a start and stop codon and no internal stops) was retained and mapped the 5' and 3' UTR based on the most commonly observed SLAS and PAS for each gene respectively. Previously identified novel genes with a valid ORF and evidence for expression as a polyadenylated transcript<sup>13</sup> were also included. To distinguish these gene models from the reference genome annotation we prefixed the gene names with “LmxM379c” indicating the strain name and its expression of Cas9 and T7.

The Cas9 enables CRISPR-assisted genome editing, while the T7 RNA polymerase allows use of sgDNA encoding a T7 promoter and sgRNA to program Cas9 activity. Using our previously published ‘LeishGEdit’ pipeline<sup>1</sup> and our updated primer design software<sup>15</sup> that is based on the CCTop CRISPR/Cas9 Target Prediction Tool<sup>16</sup>, we designed primers for: 1) PCR-based generation of constructs for endogenous protein tagging (uf/ur primers for N terminal tagging or df/dr primers for C terminal tagging, using the pPOT, pLPOT and pPLOT series of plasmids) and gene deletion (uf/dr primers, using the pT series). 2) PCR-based generation of sgRNA templates for tagging and deletion (5g/3g primers for deletion, 5g primer for N terminal tagging and 3g primer for C). 3) Primers within each protein-coding gene ORF validation of gene deletion by diagnostic PCR (vf/vr primers) (based on the Primer3 primer design software<sup>17</sup>). We also designed uf primers carrying a unique 17-nt DNA barcode<sup>15</sup> for generating barcoded pools of deletion mutants.

As this set of primers accounts for strain-specific SNPs and indels we recommended them as a standardised ‘first attempt’ for tagging and deletion genes in *L. mex* Cas9 T7 M, and we will be using them for future high-throughput reverse genetic analyses.

#### Ethics statement

All experiments were conducted according to the Animals (Scientific Procedures) Act of 1986, United Kingdom, and had approval from the University of York Animal Welfare

and Ethical Review Body (AWERB) committee. All efforts were undertaken to minimise the suffering of animals.

#### Data availability

##### Underlying data

BioProject: *Leishmania mexicana* MNYC/BZ/62/M379 Cas9/T7 whole genome sequencing; Accession number: PRJNA853937, <https://identifiers.org/bioproject:PRJNA853937><sup>18</sup>

Sequence Read Archive: Next generation sequencing of *Leishmania mexicana* Cas9/T7 strain: after mouse passage (SRR19895146). Accession number: SRR19895146; <https://identifiers.org/insdc.sra:SRR19895146><sup>19</sup>

Sequence Read Archive: Nanopore sequencing of *Leishmania mexicana* Cas9/T7 strain (SRR20123517). Accession number: SRR20123517; <https://identifiers.org/insdc.sra:SRR20123517><sup>20</sup>

##### Extended data

Zenodo: Gene tagging and gene deletion resources for *Leishmania mexicana* MNYC/BZ/62/M379 Cas9/T7 strain, <https://doi.org/10.5281/zenodo.7313190><sup>12</sup>

This project contains the primer sequences, barcodes and the GFF file containing the sequence and the annotations of the *L. mexicana* MNYC/BZ/62/M379 T7/Cas9.

#### Analysis code

All code for genome assembly, polishing, annotation updates and annotation transfer are available from GitHub: <https://github.com/Wheeler-Lab/genome-lmexcas9t7/tree/v1.0.1>; and archived in Zenodo: <https://doi.org/10.5281/zenodo.7357174><sup>21</sup>

License: [GNU GPL-3.0](https://www.gnu.org/licenses/gpl-3.0.html)

#### Reporting guidelines

Zenodo: ARRIVE E-10 Checklist for “Genome sequence of *Leishmania mexicana* MNYC/BZ/62/M379 expressing Cas9 and T7 RNA polymerase”, <https://doi.org/10.5281/zenodo.7330926><sup>22</sup>.

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/) (CC-BY 4.0).

## References

1. Beneke T, Madden R, Makin L, et al.: **A CRISPR Cas9 high-throughput genome editing toolkit for kinetoplastids.** *R Soc Open Sci.* 2017; 4(5): 170095. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
2. Rogers MB, Hillel JD, Dickens NJ, et al.: **Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*.** *Genome Res.* 2011; 21(12): 2129–42. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
3. Beneke T, Demay F, Hookway E, et al.: **Genetic dissection of a *Leishmania* flagellar proteome demonstrates requirement for directional motility in sand fly infections.** *PLoS Pathog.* 2019; 15(6): e1007828. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. Sambrook J, Russell DW: **Purification of nucleic acids by extraction with phenol:chloroform.** *CSH Protoc.* 2006; 2006(1): pdb.prot4455. [PubMed Abstract](#) | [Publisher Full Text](#)

5. Li H: **Minimap and miniasm: fast mapping and *de novo* assembly for noisy long sequences.** *Bioinformatics*. 2016; **32**(14): 2103–10.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
6. Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics*. 2018; **34**(18): 3094–100.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
7. Vaser R, Sovic I, Nagarajan N, *et al.*: **Fast and accurate *de novo* genome assembly from long uncorrected reads.** *Genome Res*. 2017; **27**(5): 737–46.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
8. Krueger F: **Trim Galore.** 2021.  
[Reference Source](#)
9. Li H: **Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.** arxiv: 13033997 [q-bio]. 2013.  
[Publisher Full Text](#)
10. Walker BJ, Abeel T, Shea T, *et al.*: **Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement.** *PLoS One*. 2014; **9**(11): e112963.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
11. Soderlund C, Bomhoff M, Nelson WM: **SyMAP v3.4: a turnkey synteny system with application to plant genomes.** *Nucleic Acids Res*. 2011; **39**(10): e68.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
12. Beneke T, Dobramysl U, Catta-Preta CMC, *et al.*: **Gene tagging and gene deletion resources for *Leishmania mexicana* MNYC/BZ/62/M379 Cas9/T7 strain.** *Zenodo*. 2022.  
<http://www.doi.org/10.5281/zenodo.6832399>
13. Amos B, Aurrecochea C, Barba M, *et al.*: **VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center.** *Nucleic Acids Res*. 2022; **50**(D1): D898–D911.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
14. Fiebig M, Kelly S, Gluenz E: **Comparative Life Cycle Transcriptomics Reveals *Leishmania mexicana* Genome Annotation and Links a Chromosome Duplication with Parasitism of Vertebrates.** *PLoS Pathog*. 2015; **11**(10): e1005186.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
15. Beneke T, Gluenz E: **Bar-seq strategies for the LeishGEdit toolbox.** *Mol Biochem Parasitol*. 2020; **239**: 111295.  
[PubMed Abstract](#) | [Publisher Full Text](#)
16. Stemmer M, Thumberger T, Del Sol Keyer M, *et al.*: **CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool.** *PLoS One*. 2015; **10**(4): e0124633.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
17. Untergasser A, Cutcutache I, Koressaar T, *et al.*: **Primer3—new capabilities and interfaces.** *Nucleic Acids Res*. 2012; **40**(15): e115.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
18. Beneke T, Dobramysl U, Catta-Preta CMC, *et al.*: ***Leishmania mexicana* MNYC/BZ/62/M379 Cas9/T7 whole genome sequencing [Data set].** NCBI BioProject PRJNA853937. 2022.
19. Beneke T, Dobramysl U, Catta-Preta CMC, *et al.*: **Next generation sequencing of *Leishmania mexicana* Cas9/T7 strain: after mouse passage (SRR19895146).** [Data set]. Sequence Read Archive; Accession number: SRR19895146. 2022.
20. Beneke T, Dobramysl U, Catta-Preta CMC, *et al.*: **Nanopore sequencing of *Leishmania mexicana* Cas9/T7 strain (SRR20123517).** [Data set]. Sequence Read Archive; Accession number: SRR19895146. 2022.
21. Dobramysl U, Wheeler RJ: **Pilon polish code for the *Leishmania mexicana* MNYC/BZ/62/M379 Cas9/T7 genome.** *Zenodo*. 2022.  
<http://www.doi.org/10.5281/zenodo.7357174>
22. Beneke T, Dobramysl U, Catta-Preta CMC, *et al.*: **ARRIVE E-10 Checklist for “Genome sequence of *Leishmania mexicana* MNYC/BZ/62/M379 expressing Cas9 and T7 RNA polymerase”.** *Zenodo*. 2022.  
<http://www.doi.org/10.5281/zenodo.7330925>

# Open Peer Review

Current Peer Review Status:   

---

## Version 2

Reviewer Report 23 February 2023

<https://doi.org/10.21956/wellcomeopenres.21134.r55188>

© 2023 McCulloch R. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Richard McCulloch** 

Wellcome Centre for Integrative Parasitology, School of Infection and Immunity, University of Glasgow, Glasgow, UK

The authors have more than adequately addressed the relatively minor issues raised.

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Genome biology of trypanosomatid parasites

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

---

## Version 1

Reviewer Report 12 January 2023

<https://doi.org/10.21956/wellcomeopenres.20597.r53618>

© 2023 Beverley S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Stephen M. Beverley**

Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, MO, USA

This is a fairly straightforward work providing the genome sequence of a workhorse *Leishmania mexicana* strain in common use for CRISPR genome editing studies. It will be a useful resource to

many laboratories.

The authors might comment upon the 'clonality' of this line - in past work from these investigators the parasites were not always of clonal origin and hence some heterogeneity may exist which could confound the analysis. Were the parent lines clonal, and since their origin, how many cell doublings have occurred that might further introduce heterogeneity?

Secondly, the use of triploid or tetraploid for specific chromosomes is incorrect - it should be trisomic or tetrasomic.

**Is the rationale for creating the dataset(s) clearly described?**

Yes

**Are the protocols appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and materials provided to allow replication by others?**

Yes

**Are the datasets clearly presented in a useable and accessible format?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Leishmania molecular genetics

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 04 January 2023

<https://doi.org/10.21956/wellcomeopenres.20597.r53621>

© 2023 McCulloch R. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

? **Richard McCulloch** 

Wellcome Centre for Integrative Parasitology, School of Infection and Immunity, University of Glasgow, Glasgow, UK

In this article, Beneke and colleagues provide an extremely useful resource: the genome sequence of a *Leishmania mexicana* strain expressing Cas9 and T7 polymerase, which is widely used to allow for CRISPR genome editing. The main focus of the article is in understanding if these modifications of an existing and sequenced 'reference' *L. mexicana* strain (to allow CRISPR editing) have resulted in undetected genome changes, including sequence variations that might reduce the effectiveness



of primers based on the reference genome. The data suggest no large scale genome changes have occurred, but that some SNPs should be accounted for in designing primers for CRISPR-based experiments. As such, the study and new genome are clearly of considerable value. Nonetheless, some methods could be more clearly explained.

1. Are the authors correct when they state that they generated 734 Gb and 7.8 Gb of sequence from before and after passage? This difference appears at odds with the very similar number of reads recovered from the two samples: 24,656,567 and 26,015,449.
2. In the article, the authors suggest that they used Pilon v1.23 to 'polish the [reference] MHOM/GT/2001/U1103 genome'. I would suggest this phrasing is perhaps confusing and could be improved. Pilon is an algorithm that can be used to improve an existing genome assembly or identify differences between related assemblies. In the context of the de novo assembly of the MNYC/BZ/62/M379 Cas9/T7 genome, 'polishing' appears appropriate, since the Illumina reads improve the Nanopore assembly. However, it appears that the approach taken when comparing the MNYC/BZ/62/M379 Cas9/T7 Illumina reads to the MHOM/GT/2001/U1103 genome assembly has not 'polished' (i.e. altered) the latter genome, but instead identified differences between them.
3. In describing the differences between the two genomes, what do the authors mean by 'small' insertions and deletions; can they specify the size range?
4. The following sentence is somewhat confusing:  
'Indeed, chromosomal coverage was not uniform, coverage was  $144 \pm 5$  (mean  $\pm$  sd.) excluding three outliers: chromosome 3 (coverage, 218; triploid), 16 (coverage, 214; triploid) and 30 (coverage, 277; tetraploid).' What is the 'coverage' they refer to; and do they mean that coverage only deviated from relatively uniform levels in the three chromosomes mentioned, or are they suggesting there may be further ploidy variation (not discussed)?

**Is the rationale for creating the dataset(s) clearly described?**

Yes

**Are the protocols appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and materials provided to allow replication by others?**

Partly

**Are the datasets clearly presented in a useable and accessible format?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Genome biology of trypanosomatid parasites

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have**

**significant reservations, as outlined above.**

Reviewer Report 23 December 2022

<https://doi.org/10.21956/wellcomeopenres.20597.r53617>

© 2022 Requena J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**José María Requena** 

Centro de Biología Molecular Severo Ochoa (CSIC-UAM), Departamento de Biología Molecular, Instituto Universitario de Biología Molecular (IUBM), Universidad Autónoma de Madrid, Madrid, Spain

In this article, Beneke *et al.* inform on the genomic sequencing and genome assembly for the *Leishmania mexicana* MNYC/BZ/62/M379 strain, which was manipulated to express CRISPR-associated protein 9 (Cas9) and T7 RNA polymerase. The creation of this cell line represents the starting point (and essential step) to apply the successful methodology of CRISPR-assisted gene deletion and endogenous tagging that was developed previously by these authors. Although the genome sequence for other *L. mexicana* strain (MHOM/GT/2001/U1103) is publically available, it is necessary to know the exact nucleotide sequence for the genes in the particular Leishmania line in which CRISPR methodology will be used. As mentioned by the authors the *L. mexicana* MNYC/BZ/62/M379 strain is widely used in different laboratories.

The methodological procedures used for generating the raw sequencing data are described in a detailed manner. Also, the bioinformatics procedures used for genome assembly and sequence polishing are clearly described. Apart from generating the genome sequence, they conducted genome annotations and transcripts delineation (including UTRs). As another valuable resource, the authors have designed oligonucleotide sequences to accomplish gene tagging or deletion of every gene in *L. mexicana* MNYC/BZ/62/M379.

In sum, this article contains valuable datasets. In this regard, this reviewer acknowledges the effort made by the authors to generate these valuable resources and encourages users to inform about possible inconsistencies in the information provided, as an effective way to improve and curate genome annotations between all.

**Is the rationale for creating the dataset(s) clearly described?**

Yes

**Are the protocols appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and materials provided to allow replication by others?**

Yes

**Are the datasets clearly presented in a useable and accessible format?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Molecular Biology, Leishmania, NGS, genomics, transcriptomics

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

---