
Article:
Mottram, Jeremy Charles orcid.org/0000-0001-5574-3766 (2024) LeishGEM:genome-wide deletion mutant fitness and protein localisations in Leishmania. Trends in parasitology. pp. 675-678. ISSN 1471-4922

https://doi.org/10.1016/j.pt.2024.06.003

Trends in
Parasitology

CellPress
OPEN ACCESS

Forum

LeishGEM: genome-wide deletion mutant fitness and protein localisations in *Leishmania*

The LeishGEM Team[1,*]

LeishGEM is a genome-wide functional annotation community resource for *Leishmania mexicana*, where deletion mutant growth *in vitro* and *in vivo* is measured and protein localisation is determined by endogenous tagging and LOPIT-DC (localisation of organelle proteins by isotope tagging with differential centrifugation) spatial proteomics. Data are being made available pre-publication via http://leishgem.org which allows data-driven identification of the mechanisms for *Leishmania* parasitism.

The power of genome-scale reverse genetics in *Leishmania* rapidly became apparent through development of high-throughput CRISPR/Cas9-assisted reverse genetics [1] and barcoded deletion mutants for parallel fitness phenotyping *in vitro* and *in vivo* [2–4]. This motivated development of the LeishGEM project to generate an open dataset for the parasitology community. The streamlined workflow designed for this project enables generation of deletion mutants at a rate of one 96-well plate per week (Figure 1, left). In parallel, using experience in high-throughput localisation [5], endogenously-tagged cell lines are generated and imaged at a rate of one 96-well plate every 2 weeks (Figure 1, right).

This project uses a Cas9 and T7 RNA polymerase-expressing *Leishmania mexicana* cell line [1], derived from strain MNYC/BZ/62/M379, competent for transmission through sand flies and for mammalian infection. All primers for PCR-based construct and single guide (sg)RNA generation for the deletion and tagging were designed based on the SNP- and indel-corrected genome sequence and refined gene model annotation (including novel small open reading frames with stable transcripts) of this cell line [6,7]. Over the course of this project, which started in 2021, all 8267 protein-coding genes, except those which cannot be targeted uniquely for technical reasons, will be targeted for deletion, and 2700 for endogenous tagging with a fluorescent protein. Each deletion mutant is checked by PCR for deletion of the target locus. To date, 77% of targeted loci yielded viable cell lines and ~51% (n = 2088) are confirmed as gene knockouts, with the remainder either technical genotyping failures or populations in which at least 1% of the population retains a copy of the gene.

Leishmania is transmitted by sand flies, growing as promastigote forms in the midgut before being transmitted by bite to a person or animal. There they infect macrophages and differentiate into the amastigote form. High-throughput phenotyping of *L. mexicana* mutants has been developed and has been used to assess parasite growth as promastigotes, differentiation to axenic amastigotes, ability to infect murine macrophages *in vitro*, and ability to infect mice. Figure 1 shows the workflow and multiple time points of this large-scale endeavour. For growth and fitness comparisons, approximately 300 barcoded deletion mutants and control cell lines are combined in pools. For measurement of promastigote growth in culture, the barcode proportions are assessed from DNA samples taken after 0 h, 24 h, 48 h, and 72 h of growth. Stationary-phase promastigote cultures are used to initiate cell differentiation assays and mouse infections. For *in vitro* assays, the library pools are grown either as axenic amastigotes or in murine bone-marrow-derived macrophages, for 24 and 72 h. For virulence and survival assays in the mouse, DNA samples are taken early post infection (at 72 h) to allow comparison with *in vitro* differentiation assays, while the later timepoints (3 weeks, 6 weeks) help to define the role of proteins involved in pathogenicity and establishing parasitism. In addition, a select number of deletion mutants will be phenotypically analysed in sandfly infections. Fitness is quantitatively assessed by barcode sequencing (bar-seq), with previously characterised deletion mutants and a dilution series of barcoded parental cell lines included in each mutant pool as controls and for pool-to-pool normalisation. This allows precise genome-scale quantitative phenotype mapping and identifies the proteins involved in virulence and longevity in all life cycle stages.

The *Leishmania* cell is being mapped using microscopy (fluorescent protein tagging at the endogenous locus) and mass spectrometry (LOPIT-DC) [8]. The *Leishmania* cell is highly structured, meaning that subcellular microscopic localisation of a protein can often be mapped to a specific organelle [9]. Since TrypTag has already mapped subcellular protein localisation genome-wide in the related parasite *Trypanosoma brucei* [5], only proteins that lack or have low (<30%) sequence identity to a *T. brucei* ortholog were selected for tagging in *Leishmania* (approximately 30% of protein-coding genes). Over the course of this ongoing project, every candidate will be endogenously tagged at the N terminus and C terminus with mNeonGreen [10] and diffraction-limited widefield epifluorescence microscopy images of >200 cells at random cell cycle stages are captured per cell line, for both promastigotes and axenic amastigotes (Figure 1). Every image set is manually annotated, drawing on a

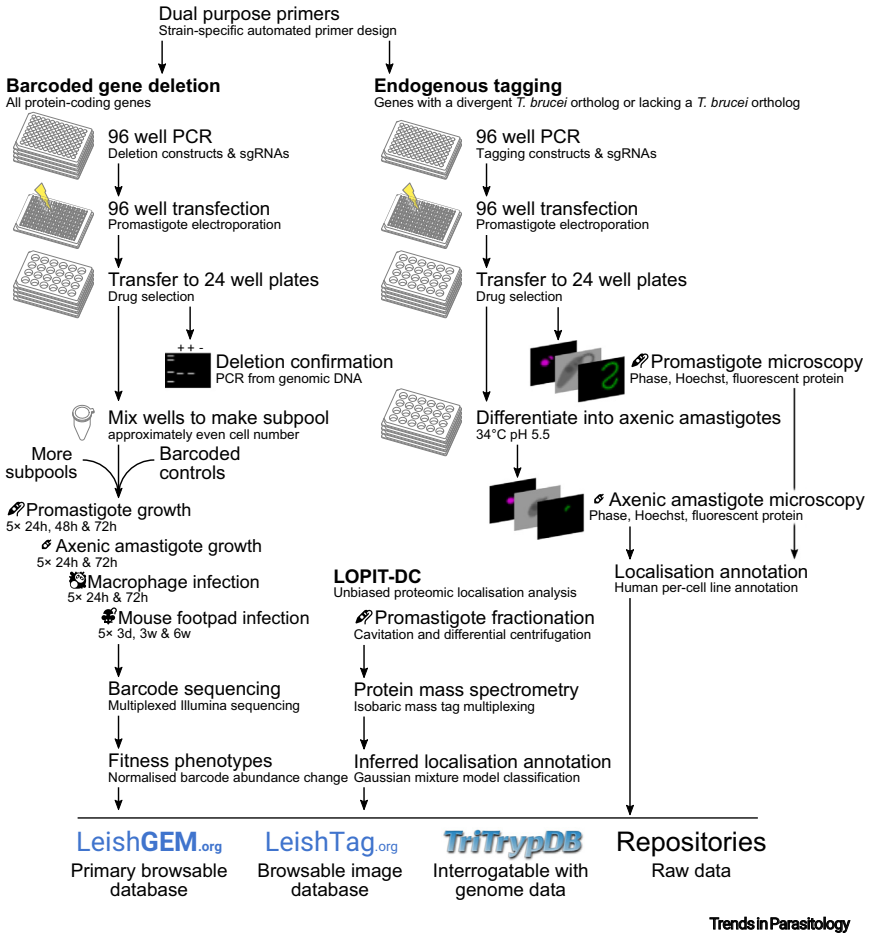

Figure 1. Workflow for high-throughput gene deletion, phenotyping and tagging. Schematic of the reverse genetics workflows, from automated primer design to data release routes, and integration of complementary localisation of organelle proteins by isotope tagging with differential centrifugation (LOPIT-DC) data. Abbreviations: sgRNA, single guide RNA; *T. brucei, Trypanosoma brucei.*

standardised ontology linked with gene ontology (GO) cellular component terms, and using a standardised modifier scheme to describe more complex localisation patterns. This maps localisation in the promastigote and any change of localisation or expression level in *in vitro* amastigotes. In addition to the tagging, proteome-wide LOPIT-DC data have been generated, with localisation annotated by Gaussian mixture model classification. This approach localises the endogenous protein and avoids possible tag interference with N- or C-terminal organelle targeting sequences or anchors. The LOPIT-DC approach also provides the information for non-microscopically analysable structures like ribosomes and proteins that are not included in the tagging set due to their high similarity to *T. brucei* proteins (discussed earlier). These complementary approaches will help to build a cellular map of a large majority of *Leishmania* proteins.

Our aim is to support the parasitology community by making available all the cell differentiation and fitness phenotyping data along with the complementary protein localisation. These data are accessible through two dedicated websites, in phased releases (Figure 1). Fitness phenotype data in promastigotes, axenic amastigotes, in macrophages, and in mice, localisation classifications from LOPIT-DC and annotated microscopic images from LeishTag are available on the LeishGEM Data Browser[i]. Microscopy images – example fields of view of N- and C-terminal tagging in promastigotes and axenic amastigotes – are available on LeishTag[ii]. Once data collection is complete, these data will be integrated into TriTrypDB [11] for access under the phenotype and localisation sections of the corresponding gene pages and integrated into the search tools for phenotype strength and localisation. Raw data will ultimately be available in open data repositories in standard data formats – sequencing read archive (SRA) for bar-seq, proteomics identifications database (PRIDE) for LOPIT-DC, and Zenodo for microscopy and supporting metadata. This will include the code necessary to replicate the quantitative analyses, and Python modules are planned to facilitate programmatic access to all data.

Leishmania sp. are deadly human parasites which evolved vertebrate parasitism independently of the *Trypanosoma* lineages [12]. Despite recent advances – such as Alphafold structural prediction and the TrypTag resource – an enormous portion of the *Leishmania* genome lacks primary functional annotation. LeishGEM data will comprehensively map the most important systems for infectivity to a mammalian host and map the cellular structures in which they reside (Figure 2). Moreover, the similarity of gene content in the related human-infective *Leishmania* species [13] means that LeishGEM is relevant to research on all these parasites. As of April 2024, phenotype data for 2305 deletion mutants, microscopy subcellular localisations for 1209 proteins, and LOPIT localisation classifications for 2406 proteins are available. Given current throughput, the complete fitness and localisation data are expected to be available by 2025,

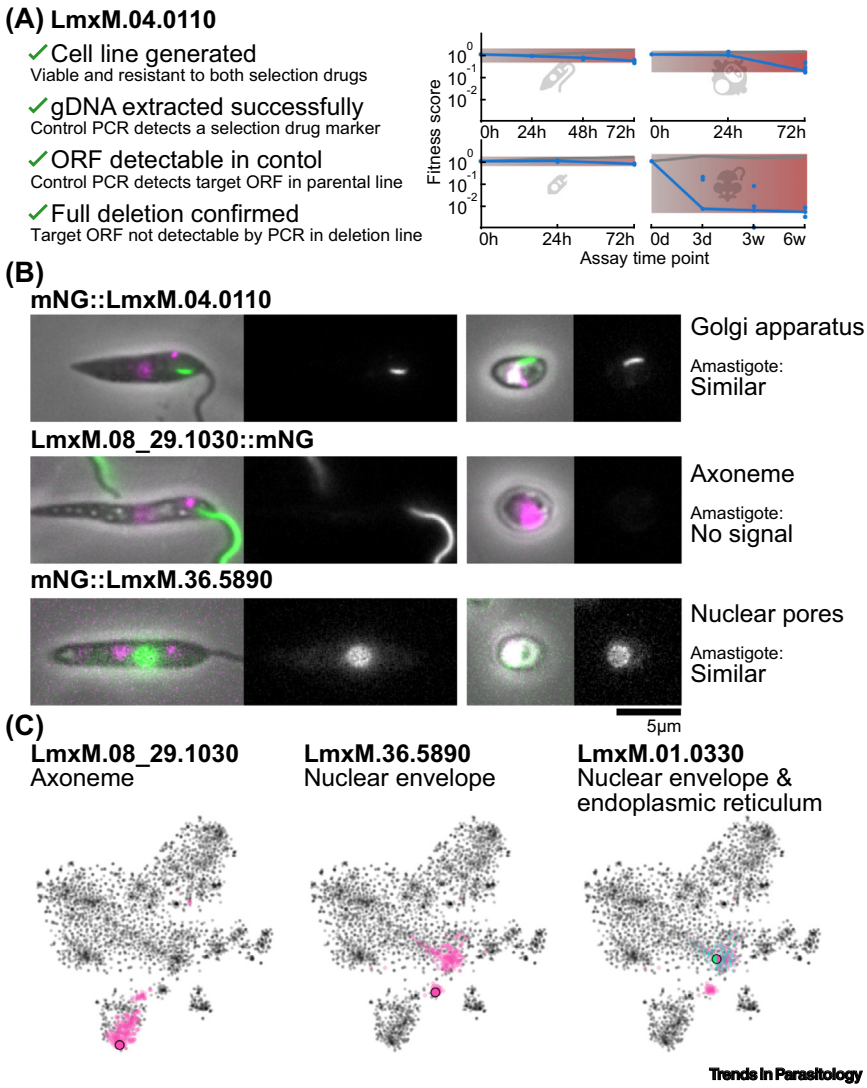

Figure 2. Examples of fitness and localisation data. Illustrative data for genes encoding conserved hypothetical proteins. (A) Detailed summary of deletion mutant and fitness phenotyping data for an example gene. Graphs show the fitness (blue) relative to parental cell line controls (grey). Shades of red indicate a statistically significant deleterious fitness phenotype. (B) Three examples of endogenous tagging and annotated protein subcellular localisation in promastigotes and after differentiation to amastigotes. Showing the protein from (A) and two more examples. (C) Three example t-SNE plots of localisation of organelle proteins by isotope tagging (LOPIT) data, highlighting the group (pink) or groups (pink and cyan) classified by Gaussian mixture model and the protein (circled). Showing two proteins from (B), and one not scheduled for tagging. Abbreviations: gDNA, genomic DNA, ORF, open reading frame.

with subsequent papers describing and analysing the dataset. In the meantime, it is anticipated this research will underpin fundamentally new types of analyses, from the evolution of systems necessary for vertebrate parasitism in *Leishmania* to identification of novel organelle subdomains or subcompartments necessary for their function. For large-scale analyses of pre-publication data made available through LeishGEM, the authors are keen to establish collaborations, while for smaller-scale analyses, like fitness phenotype or localisation of one or two proteins, researchers may do so by citing this publication and the LeishGEM or LeishTag website, respectively.

Acknowledgments

The LeishGEM project is supported by a Wellcome Trust collaborative award (221944/A/20/Z).

Declaration of interests

The authors declare no competing interests.

Resources

[i]http://browse.leishgem.org

[ii]http://leishtag.org

[1]Affiliations: University of Bern, Bern, Switzerland, Oxford Brookes University, Oxford, United Kingdom, University of Glasgow, Glasgow, United Kingdom, University of Oxford, Oxford, United Kingdom, University of York, York, United Kingdom, Curtin University, Perth, Australia, Charles University, Prague, Czechia

The LeishGEM Team (in alphabetical order); Sidonie Aellig, Karen Billington, Jeziel D. Damasceno, Laura Davidson, Ulrich Dobramysl, Ruth Etzensperger, Eden Ramalho Ferreira, Eva Gluenz, Jeremy C. Mottram, Rachel Neish, Raquel Pereira, James Smith, Jack D. Sunter, Petr Volf, Richard J. Wheeler, Matthew Young

*Correspondence:
eva.gluenz@unibe.ch (E. Gluenz),
jeremy.mottram@york.ac.uk (J.C. Mottram),
jsunter@brookes.ac.uk (J.D. Sunter), and
richard.wheeler@ndm.ox.ac.uk (R.J. Wheeler).

https://doi.org/10.1016/j.pt.2024.06.003

References

1. Beneke, T. *et al.* (2017) A CRISPR Cas9 high-throughput genome editing toolkit for kinetoplastids. *R. Soc. Open Sci.* 4, 170095
2. Beneke, T. *et al.* (2019) Genetic dissection of a *Leishmania* flagellar proteome demonstrates requirement for directional motility in sand fly infections. *PLoS Pathog.* 15, e1007828
3. Beneke, T. and Gluenz, E. (2020) Bar-seq strategies for the LeishGEdit toolbox. *Mol. Biochem. Parasitol.* 239, 111295
4. Baker, N. *et al.* (2021) Systematic functional analysis of Leishmania protein kinases identifies regulators of differentiation or survival. *Nat. Commun.* 12, 1244
5. Billington, K. *et al.* (2023) Genome-wide subcellular protein map for the flagellate parasite Trypanosoma brucei. *Nat. Microbiol.* 8, 533–547

6. Beneke, T. *et al.* (2022) Genome sequence of *Leishmania mexicana* MNYC/BZ/62/M379 expressing Cas9 and T7 RNA polymerase. *Wellcome Open Res.* 7, 294
7. Fiebig, M. *et al.* (2015) Comparative lifecycle transcriptomics revises *Leishmania mexicana* genome annotation and links a chromosome duplication with parasitism of vertebrates. *PLoS Pathog.* 11, e1005186
8. Geladaki, A. *et al.* (2019) Combining LOPIT with differential ultracentrifugation for high-resolution spatial proteomics. *Nat. Commun.* 10, 331
9. Halliday, C. *et al.* (2019) Cellular landmarks of *Trypanosoma brucei* and *Leishmania mexicana*. *Mol. Biochem. Parasitol.* 230, 24–36
10. Shaner, N.C. *et al.* (2013) A bright monomeric green fluorescent protein derived from *Branchiostoma lanceolatum*. *Nat. Methods* 10, 407–409
11. Amos, B. *et al.* (2021) VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center. *Nucleic Acids Res.* 50, D898–D911
12. Lukes, J. *et al.* (2014) Evolution of parasitism in kinetoplastid flagellates. *Mol. Biochem. Parasitol.* 195, 115–122
13. Rogers, M.B. *et al.* (2011) Chromosome and gene copy number variation allow major structural change between species and strains of *Leishmania*. *Genome Res.* 21, 2129–2142