Novel Computational Methods for the Comparison of *Leishmania* Genomes and Transcriptomes

D. Klimes, T. Belew, N. El-Sayed

Abstract

Over 20 species of *Leishmania* protozoa cause 2 million cases of Leishmaniasis each year. Current anti-leishmanial drugs are toxic, and the significant diversity between *Leishmania* species has impeded progress towards an effective vaccine. Many *Leishmania* have been rigorously investigated individually via genome and transcriptome analysis. However, previous methods could only compare less than 70% of genes between species. Here, we describe a novel computational tool that effectively compares 87% of genes across 8 *Leishmania* species. With this ability, we provide new insights into the shared and unique *Leishmania* infection pathways.

Objectives

- Develop a new method to compare the genomes across *Leishmania* species
- Identify trends in genetic variants
- Describe functional patterns in non-conserved gene groups
- Characterize differences in gene expression
- Understand the shared infection pathways
- Classify transcriptome differences between different clinical manifestations of Leishmaniasis

Methods

Pseudogenomes were made for 7 *Leishmania* species, each with their respective species' sequence and the annotations of *L. major*. Genetic variants were classified for each pseudogenome and annotated according to their position relative to the nearest gene. Functional trends were determined for regions with low conservation 142 RNAseq samples from each species were then aligned against the respective pseudogenomes for transcriptome comparison. Differential expression analysis was performed between the pre-infectious (promastigote) and infectious (amastigote) *Leishmania* stages.

1 Generating a Pseudogenome while Maintaining the Feature Annotations of the Reference Genome

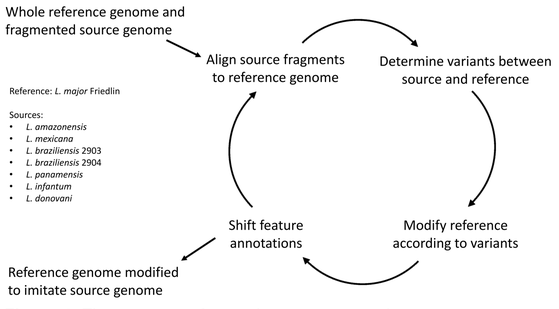


Figure 1. The process of pseudogenome creation

Results

2 Pseudogenomes are Iteratively Built through Application of Fragment Variants

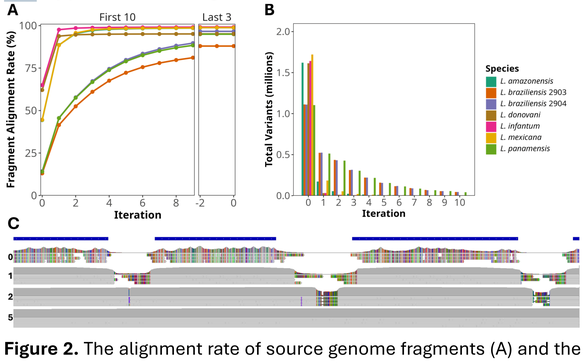


Figure 2. The alignment rate of source genome fragments (A) and the total variants called in each iteration (B) are shown. A representative view of the alignment of fragments in subsequent iterations is provided(C).

Results

3 Variant Frequencies Show Distinct Distribution Patterns between Coding Sequences and UTRs

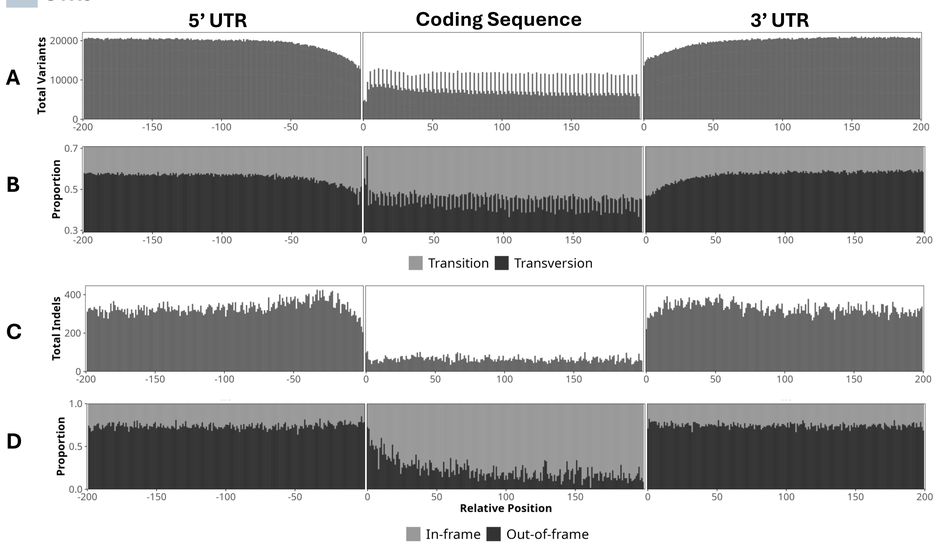


Figure 3. All genomic variants were classified by their relative positions within genes (A). SNPs were additionally classified as transitions and transversions (B). Total indel frequency (C) and the frequency of frameshifts (D) are shown. Features were limited to 200 bases each for visualization.

4 Pseudogenome Writing Compares 87% of 8 *Leishmania* Genomes. Non-conserved Regions are Consistent Between Species and Show Functional Trends

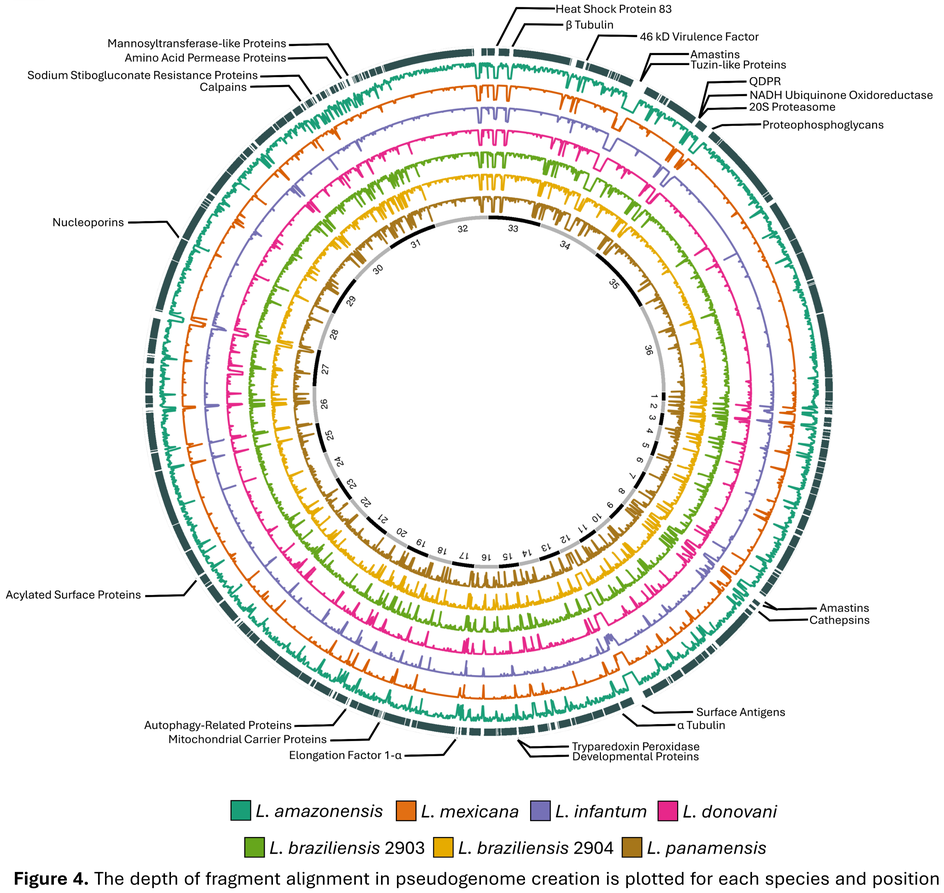


Figure 4. The depth of fragment alignment in pseudogenome creation is plotted for each species and position within the genome (chromosomes labeled). Low alignment depth indicates non-conserved regions too genetically different to compare. Low-depth regions were labeled by the functional trends of genes within them.

5 *Leishmania* Transcriptomes Separate by Life Cycle Stage and Secondarily by Species. Amastigote Samples Show Greater Species Similarity and Separate by Time of Infection

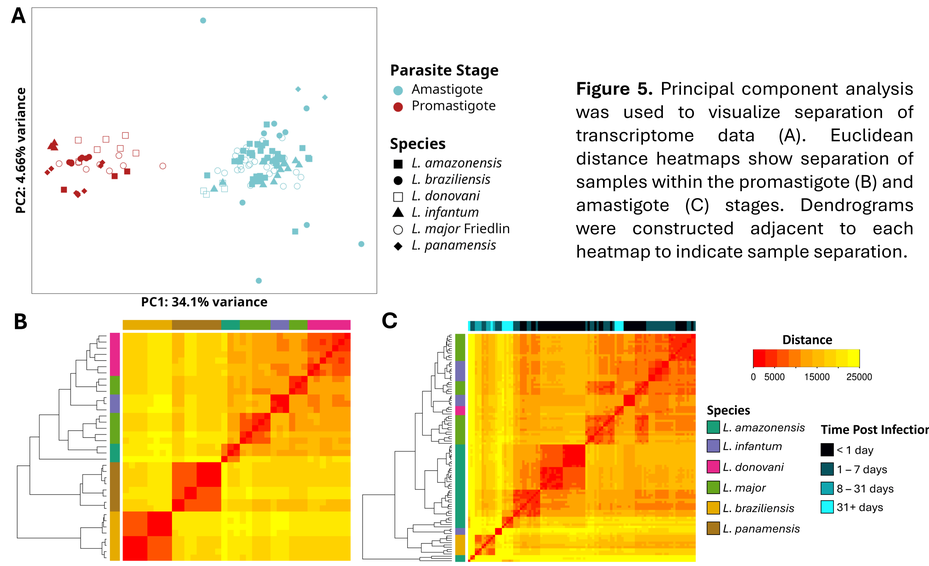
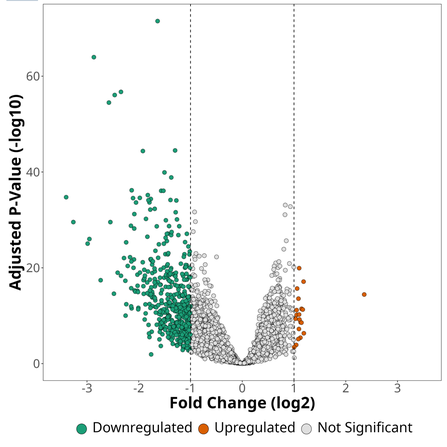


Figure 5. Principal component analysis was used to visualize separation of transcriptome data (A). Euclidean distance heatmaps show separation of samples within the promastigote (B) and amastigote (C) stages. Dendrograms were constructed adjacent to each heatmap to indicate sample separation.

6 Differential Expression Analysis Across all Species Reveals Pathways for Infection and Survival Conserved Across Species



Gene Function	FC (log2)
Amastin protein	2.36
Ubiquitin-activating enzyme	1.14
Arginine N-methyltransferase	1.10
Phosphoglycan galactosyltransferase	1.12
Acetylornithine deacetylase	1.04
Anti-apoptosis inhibitor	1.03

Gene Function	FC (log2)
Paraflagellar rod protein	-3.27
Amastin protein	-2.48
PAP2 superfamily member	-2.29
Mitogen-activated kinase	-2.21
Cysteine peptidase	-1.79
Flagellar member	-1.43

Figure 6. Differential expression analysis between the amastigote and promastigote stages was performed on 142 RNAseq samples spanning 6 *Leishmania* species. Genes were classified as being significantly up or downregulated in the transition to the infectious stage (amastigote).

7 Unique Clinical Manifestations of Leishmaniasis Show Significant Differences in Gene Expression and Infection Pathways

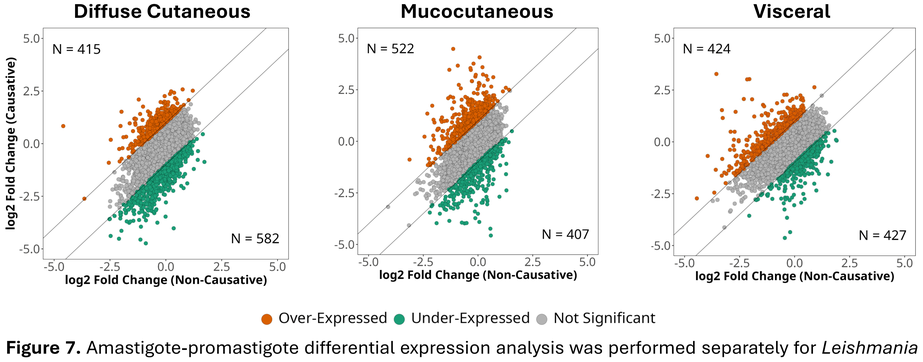


Figure 7. Amastigote-promastigote differential expression analysis was performed separately for *Leishmania* that could and could not cause unique clinical manifestations of Leishmaniasis. The relationship between resultant log2 fold changes is shown for each gene. Expression was deemed significantly different (colored) if the log2 fold changes differed by more than 1