

COMPUTATIONAL GENE PREDICTION IN *LEISHMANIA MAJOR*

*Allison Griggs and Shuba Gopal**, Department of Biological Sciences, adg9398@rit.edu, sxgsbi@rit.edu

The degree of evolutionary divergence between *L. major* and other species with annotated genomes precludes the viability of traditional gene prediction methods. *L. major*, while eukaryotic, does not contain introns in its genome, so it is still possible to obtain putative coding regions by identifying open reading frames (ORFs). To ascertain which subset of these ORFs are likely to be biologically relevant, an aspect of trypanosome biology was exploited. The trypanosome family possesses a unique signal that alters the nucleotide frequencies between coding and non-coding regions. We exploited this feature to create a gene verification algorithm for ORFs that classifies them as coding or non-coding.

Current efforts at gene identification often combine several methods and can have as high as a 50% false negative rate. Our method has a distinct advantage over current models: it uses only one algorithm for all data sets. The method compares dinucleotide transition probabilities of coding and non-coding regions using linear discriminant analysis, generating a model that accurately identifies 90% of coding regions when tested on known genes, with a 10% false negative rate. This approach improves significantly on existing methods and has now been used to annotate the entire *L. major* genome.

Allison would like to present this work as a poster at the Symposium.