Title: Evolution of Variation: What contributes to mitochondrial DNA sequence variation in the Chagas disease parasite, *Trypanosoma cruzi*?

Abstract:

The protozoan parasite *Trypanosoma cruzi* is the etiological agent of Chagas disease, a neglected tropical disease causing severe digestive and cardiac issues. Most Chagas parasite and insect-vector research concerns South America, with few studies regarding the wider geographic distribution of North, Central, and South American landscapes. We focus on detecting and analyzing the impact of natural selection, demographic processes, and mutation on the genetic variation in the *T. cruzi* mitochondrial COII gene for Central and South American locations. By understanding the parasite's genetic variation, we aim to increase the knowledge applicable to new vector control strategies, thereby decreasing Chagas incidence.

Description of the Project:

Within Dr. S's research group, I will focus my project on detecting and analyzing the impact of natural selection, demographic (random genetic drift, migration), mutation or neutral processes on the genetic variation in a specific gene of *T. cruzi's,* mitochondrial Cytochrome Oxidase subunit II (COII) gene. We ask, does the Chagas Disease parasite genetics vary depending on location, insect vector species, or between vector species and between locations? The analysis of these signatures of different types of selection models will compare parasites from multiple geographic regions and from two insect vector species (*Triatoma dimidiata* and *Triatoma nitida).* Specific population genetics tests to be used

include F_{ST} (Holsinger and Weir 2009), the Hudson Kreitman Agaudé (HKA) test, the MK test (McDonald and Kreitman 1991), Tajima's D (Tajima 1991), and others. These directions for the project are influenced by Gallant et al. (2018)'s selection analysis of the *T. cruzi* trans-sialidase protein (TcTS), which is responsible for sialic acid transfer from the host to the parasite and is a protein implicated in Chagas disease parasite immune evasion in Central American and Mexican geographic regions. Results revealed negative selection on five drug-candidate amino acids in this gene, especially that of Tyr342.

Because the Cytochrome Oxidase gene is also implicated in virulence, evasion of the host immune system, and basic metabolism (Pentinsari et al., 2016), understanding its evolution will provide important information for understanding transmission dynamics and developing treatments.

Previous Work:

A. Phylogenetic analysis:

Organisms, such as the *T. cruzi* parasite, contain unique variations in the genetic material encoding proteins and thus life. Pech-May et al. (2019), described a scope of analysis for *T. dimidiata* diversity and DTU infection frequencies, with an overarching goal to provide a larger spanning investigation of genetic diversity across domestic, anthropogenically modified landscapes and sylvatic environments (essentially sub-regional, regional, and global Neotropical realms). This transition from sub-regional, regional, and then to global landscapes results in different perspectives on the conventions of phylogenetic arrangement.

Moreover, Dorn et al. (2017) investigated *T. cruzi* as it specifically pertains to the insect-vector, *Triatoma dimidiata* from Central American locations spanning from Mexico to Columbia. The phylogenetic strain typing for the parasite is a point of great dissension. The current paradigm of strain phylogenetics describes six monophyletic clades based on nuclear genes, named discrete typing units (DTUs TcI-VI) (Barnabé 2016, Dorn 2017, Brenière 2016, Cosentino 2012, Messenger et al. 2015). Recently a seventh DTU, initially observed in bats, has been proposed (Cosentino 2012). Nevertheless, identifying the parasite's strains and the diversity the strains comprise is an ongoing process. Within the same span of time, Barnabé et al. (2016) subverted the conventional nuclear strain model for a proposal of three mitochondrial clades. In analyzing open-sourced sequences on the *Genbank* database for the entire geographic distribution of Chagas disease, Barnabé et al. discussed the necessity for standardizing the genetic markers for phylogenetic structuring across larger distributions and the potential for new clusterings from these re-considered clades, mtTcI, mtTcII, and mtTcIII.

B. Selection analyses:

Detection and analysis of evolutionary mechanisms—such as the impact of natural selection, demographic processes (random genetic drift, migration, mating), and mutation—are instrumental to contextualizing the genetic variation one observes in *T. cruzi.* For example, Gallant et al. (2018) observed negative selection, or selective removal of deleterious alleles, through analysis of the *T. cruzi* trans-sialidase protein (TcTS) implicated in Chagas disease parasite immune

evasion in Central American and Mexican geographic regions. Other literature agrees with this trend of negative selection, using a plethora of different statistical approaches, including F_{ST} (Holsinger and Weir 2009), the Hudson Kreitman Agaudé (HKA) test, the MK test (McDonald and Kreitman 1991), Tajima's D (Tajima 1991), and others. These tests reveal how evolutionary forces—natural selection, mutation, migration, and random genetic drift—may contribute to genetic variation.

C. Functional analysis:

Functional investigations concerning the COI gene illustrate the utility of broader interpretations of sequence data. A massive study on cytochrome c oxidase subunit 1 (COI) sequences included measurements of variation at amino acid sites, mapping conserved and variable sites on the COI protein structure, and potential functional analysis of DNA barcode variation (Ruiz-Lopez 2012). In essence, these investigations are a stepping stone in providing strong evidence for specific biological processes. Any potential variants in the *T. cruzi* parasite's DNA may result in significant, higher-order differences in the proteins and, as a result, differences in the biology of the organism.

Significance:

Most of the Chagas disease research on insect-vector and parasite dynamics has been focused on South America, with few studies concerning wider-spanning geographic distribution across North, Central, and South American landscapes. My project focuses on a wider geographic range, analyzing the impact of natural selection, demographic processes (random genetic drift,

migration, mating), and mutation on the genetic variation in the *T. cruzi* mitochondrial COII gene for locations across Central and South American countries. This wider geographic sampling provides the opportunity to contribute to the current body of literature comprising analyses conducted in other, often under-represented countries. My work is significant for advancing the knowledge of the dynamics of Chagas disease transmission and allow us to increase our understanding of parasite-vector interactions across different vector species and locations. Specifically, an understanding of differential selective pressures on the gene of interest enables a greater understanding of transmission and thus management of Chagas disease.

Previous work has initiated investigation of *T. cruzi* genes for developing drugs and treating Chagas disease with the use of evolutionary mechanistic analysis. My work provides a complementary understanding for analysis of these evolutionary processes using a gene more characterized by metabolic processes. Moreover, the mitochondrial approach for analysis complements the more nuclear-centric analysis currently in the literature. Knowledge of a single or even a handful of genes is not sufficient for developing drug treatments and understanding transmission dynamics. Mitochondrial and nuclear genes are known to evolve differently, and pathogens are known to evolve resistance to single-target drug approaches. Such evolutionary mechanisms of natural selection, demographic processes (random genetic drift, migration, mating), and mutation are instrumental to both contextualizing the genetic variation one observes in *T. cruzi* and developing more complex targeted drug approaches.

Finally, my work is directly positioned within the works of previous projects as well as ongoing projects. For example, one of my laboratory group's collaborators works closely with indigenous populations in Guatemala and directly uses such information in her control efforts (Bustamante et al., 2015). To this end, by understanding Chagas disease, its genetic variation, and how genetic variation in the host and parasite influence disease transmission and severity, our research group's goal is to add to the database of knowledge applicable to new vector control strategies needed to decrease vector-human contact and Chagas incidence.

Proposed Methodology:

Parasite DNA will be extracted from the abdomens of multiple species of infected insect vectors, *Triatoma sanguisuga, Triatoma rubida, Triatoma dimidiate,* and *Triatoma nitida,* collected from locations across Central and South America. I have already sequenced the COII gene from 55 samples from *T. dimidiata* and *T. nitida* from Guatemala. Grant funding I have received for the semester will enable me to examine an additional 96 samples (already available in Dr. S's lab) from additional species and locations. I will be working with graduate student R.L. and Professor S. Analyzing selection and the other factors require DNA extraction, amplification of the COII gene using the polymerase chain reaction followed by Sanger sequencing by a commercial facility, and statistical analysis of the DNA sequence data.

Specific tests to be used include F_{ST} (Holsinger and Weir 2009), the Hudson Kreitman Agaudé (HKA) test, the MK test (McDonald and Kreitman

1991), Tajima's D (Tajima 1991), and others (Calvo-Martín et al. 2017). These statistical tests rely on the null hypothesis that genetic variation in the parasite population is neutral with respect to factors that can change the frequencies of DNA variants. Alternatively, genetic variation may be the result of one of four forces: natural selection, mutation, migration and random genetic drift. Each factor can involve multiple processes; for example, examining random genetic drift includes testing for increasing, stable, and decreasing population size.

My analysis will depend on the potential mechanisms of evolution I distinguish for the samples. For example, if I detect natural selection on the parasite, it would lead to further determining if the selection was from the insect, the mammal host, or some other factor. Evidence of population structure and migration would indicate that the parasite distribution is non-random over the landscape and that disease control efforts need to consider the parasite population structure. If the genetic variation is associated with differences in virulence, one might see different disease dynamics between locations or depending on the vector species. With the expanded sampling, I plan to write the results up for publication in a peer-reviewed journal, such as the *Journal of Medical Entomology*. To this end, my thesis will reveal deeper insights into the selective pressures on *T. cruzi* parasite causing Chagas disease across the broader geographic range of Central and South America. By extension, these understandings of the parasite's population genetics inform the evolutionary history and dynamics required to develop strategies for disease transmission control.

Timeline for the stages of the project:

	September	October	November	December	January	February	March	April	May
1. prepare/send remaining samples for sequencing	X								
2. compile/curate database of samples for analysis	X	X							
3. phylogenetic reconstruction		X	X						
4. selection analysis		X	X	X					
5. writing/drafting thesis	X	X	X	X	X	X	X	X	
6. defense									X

References:

Brenière, S. F., Waleckx, E., & Barnabé, C. (2016). Over Six Thousand Trypanosoma cruzi Strains Classified into Discrete Typing Units (DTUs): Attempt at an Inventory. *PLOS Neglected Tropical Diseases*, *10*(8), e0004792. https://doi.org/10.1371/journal.pntd.0004792

Cosentino, R. O., & Agüero, F. (2012). A Simple Strain Typing Assay for Trypanosoma cruzi: Discrimination of Major Evolutionary Lineages from a Single Amplification Product. *PLOS Neglected Tropical Diseases*, *6*(7), e1777. https://doi.org/10.1371/journal.pntd.0001777

Gallant, J. P., Lima-Cordón, R. A., Justi, S. A., Monroy, M. C., Viola, T., & Stevens, L. (2018). The role of natural selection in shaping genetic variation in a promising Chagas disease drug target: Trypanosoma cruzi trans-sialidase. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *62*, 151–159. https://doi.org/10.1016/j.meegid.2018.04.025

Messenger, L. A., Garcia, L., Vanhove, M., Huaranca, C., Bustamante, M., Torrico, M., … Llewellyn, M. S. (2015). Ecological host fitting of Trypanosoma cruzi TcI in Bolivia: Mosaic population structure, hybridization and a role for humans in Andean parasite dispersal. *Molecular Ecology*, *24*(10), 2406–2422. https://doi.org/10.1111/mec.13186

Messenger, L. A., Llewellyn, M. S., Bhattacharyya, T., Franzén, O., Lewis, M. D., Ramírez, J. D., … Miles, M. A. (2012). Multiple Mitochondrial Introgression Events and Heteroplasmy in Trypanosoma cruzi Revealed by Maxicircle MLST and Next Generation Sequencing. *PLOS Neglected Tropical Diseases*, *6*(4), e1584. https://doi.org/10.1371/journal.pntd.0001584

Messenger, L. A., Yeo, M., Lewis, M. D., Llewellyn, M. S., & Miles, M. A. (2015). Molecular Genotyping of Trypanosoma cruzi for Lineage Assignment and Population Genetics. In C. Peacock (Ed.), *Parasite Genomics Protocols* (Vol. 1201, pp. 297–337). https://doi.org/10.1007/978-1-4939-1438-8_19

Pentinsaari, M., Salmela, H., Mutanen, M., & Roslin, T. (2016). Molecular evolution of a widely-adopted taxonomic marker (COI) across the animal tree of life. *Scientific Reports*, *6*. https://doi.org/10.1038/srep35275