2017

# Natural selection and genetic variation in a promising Chagas disease drug target: Trypanosoma cruzi trans-sialidase

Joseph P. Gallant
*University of Vermont*

# NATURAL SELECTION AND GENETIC VARIATION IN A PROMISING CHAGAS DISEASE DRUG TARGET: *TRYPANOSOMA CRUZI* TRANS-SIALIDASE

A Thesis Presented

by

Joseph Gallant

to

The Faculty of the Graduate College

of

The University of Vermont

In Partial Fulfillment of the Requirements
for the Degree of Master of Science
Specializing in Pharmacology

October, 2017

Defense Date: July 27th, 2017
Thesis Examination Committee:

Lori Stevens Ph.D., Advisor
Bryan Ballif Ph.D., Chairperson
Karen Lounsbury, Ph.D.
Wolfgang Dostmann Ph.D., M.D.
Cynthia J. Forehand, Ph.D., Dean of the Graduate College

**ABSTRACT**

Rational drug design is a powerful method in which new and innovative therapeutics can be designed based on knowledge of the biological target aiming to provide more efficacious and responsible therapeutics. Understanding aspects of the targeted biological agent is important to optimize drug design and preemptively design to slow or avoid drug resistance. Chagas disease, an endemic disease for South and Central America and Mexico is caused by *Trypanosoma cruzi,* a protozoan parasite known to consist of six separate genetic clusters or DTUs (discrete typing units). Chagas disease therapeutics are problematic and a call for new therapeutics is widespread. Many researchers are working to use rational drug design for developing Chagas drugs and one potential target that receives a lot of attention is the *T. cruzi* trans-sialidase protein. Trans-sialidase is a nuclear gene that has been shown to be associated with virulence. In *T. cruzi,* trans-sialidase (TcTS) codes for a protein that catalyzes the transfer of sialic acid from a mammalian host coating the parasitic surface membrane to avoid immuno-detection. Variance in disease pathology depends somewhat on *T. cruzi* DTU, as well, there is considerable genetic variation within DTUs. However, the role of TcTS in pathology variance among and within DTU's is not well understood despite numerous studies of TcTS. These previous studies include determining the crystalline structure of TcTS as well as the TS protein structure in other trypanosomes where the enzyme is often inactive. However, no study has examined the role of natural selection in genetic variation in TcTS. In order to understand the role of natural selection in TcTS DNA sequence and protein variation, we sequenced 540 bp of the TcTS gene from 48 insect vectors. Because all 48 sequences had multiple polymorphic bases, we examined cloned sequences from two of the insect vectors. The data are analyzed to understand the role of natural selection in shaping genetic variation in TcTS and interpreted in light of the possible role of TcTS as a drug target.

**ACKNOWLEDGEMENTS**

First I would like to thank my family, who encouraged and supported me to pursue my passions. My parents who helped me wherever I needed, whether I needed them to be editors on a paper, an audience for a presentation, or an ear to hear out my stresses and calm me down. They were always there for me and willing to help. My brother, who has helped me more than he may realize and is always there for me. Without them none of this would be possible.

Next I would like to thank Dr. Lori Stevens and the Stevens lab. Lori took me into her lab as an undergraduate with no lab experience, and for that I will be forever lucky and grateful. Lori introduced me to research and now I can't imagine having ever wanted to do anything else. Thank you Lori for always being patient, encouraging and willing to help throughout my time at UVM. I will never forget the amazing experiences you made possible for me.

I would also like to thank all the other members of the Stevens lab who make it a fantastic place to come and work every day; Raquel Lima, I am lucky to have learned from such a fantastic teacher in the lab. Judith Keller, who helped me in and outside of lab, answering countless questions on "adulting". Silvia Justi, for bearing with me while I incorrectly pronounced the hundreds of Portuguese words she tried to teach me. Toni Viola, who with me ran countless experiments and problem solved all the questions we were too embarrassed to ask the grad students. Finally, all the wonderful undergraduates in lab, who help make the lab a great place for everyone.

Next I would like to thank my amazing committee members; Dr. Bryan Ballif, Dr. Wolfgang Dostmann, and Dr. Karen Lounsbury, who all took their time to give valuable advice and direction on this project.

As well a big thanks to Dr. Tony Morielli for helping me through this program. Dr. Morielli, no matter how many times I dropped in unexpectedly, always took time to help me through my questions and to see how I was doing.

Finally, I need to thank my family away from home, Cory Dawson, Jamie Thayer and Dan Hebert. Their friendship and the experiences we have shared is something that I will take with me forever. You guys will always be family to me.

## TABLE OF CONTENTS

Page

## LIST OF TABLES

## LIST OF FIGURES

## Chapter 1: Introduction to Chagas

### *A Brief History of Chagas Disease and Control Efforts*

Chagas disease was first described in 1909 by its namesake Dr. Carlos Chagas, and his description remains one of the most complete initial descriptions of any tropical disease [1]. Carlos Chagas was able to identify the infecting parasite, the vector mode of transmission, possible mammalian host species, and clinical symptoms. Despite the initial complete description, over a 100 years later Chagas disease remains one of the most impactful neglected diseases in Latin America [1]. As well, the burden of Chagas continues to grow and with the modern mobility of humans, Chagas exists globally, with an estimated 8 to 10 million people infected worldwide [2].

The protozoan family Trypanosoma contains pathogenic and non-pathogenic species [3] including *Trypanosoma cruzi*, the causative agent of Chagas Disease, and *Trypanosoma brucei*, the etiological agent of African Sleeping Sickness. The main transmission mode for Chagas is through insect vectors of the subfamily Triatominae [4]. The *T. cruzi* life cycle is often split into two phases: the insect vector cycle and the mammalian host cycle. The vector cycle includes the parasite being introduced to the vector through a blood meal of an infected mammal host, where the parasite is taken up as trypomastigotes [5]. Once inside the insect, the parasite progresses to the midgut as an epimastigote, where it proliferates, eventually progressing to the hindgut as metacyclic trypomastigote [5]. When the triatomine takes a subsequent blood meal, the parasite passes from the hindgut and is deposited in the feces onto the skin of the mammal blood source. This marks the transition from the insect vector and to the pathogenic human

cycle where the metacyclic trypomastigotes may enter the bite-wound site or through a mucous membrane into the host bloodstream. Once a mammal is infected the trypomastigotes enter cells, and transform into amastigotes which continue to proliferate within cells of infected tissues [5].

Understanding the parasite life-cycle allows for two distinct paths for managing the impact and reach of Chagas disease, one being through vector control. The second path, the topic of this thesis, is exploring possible pharmacological advancements specifically targeting the parasite within mammalian hosts through new drugs.

Vector control has been a primary method for reducing Chagas disease in many of the endemic countries in South and Central America and Mexico and has been especially effective in the South American countries where large multidisciplinary efforts were made. One large effort to reduce Chagas transmissions was the Southern Cone Initiative which was a coordinated effort between six South American countries (Argentina, Bolivia, Brazil, Chile, Paraguay, and Uruguay) by the World Health Organization (WHO) [6]. One of the main goals of the Southern Cone Initiative was to eradicate the main insect vector for the region, *Triatoma infestans*, largely through insecticide spraying of infested and at-risk dwellings [7]. The initiative was considered successful in most of the Southern Cone countries, drastically reducing populations of *T. infestans* and thus effectively reducing vector borne transmission and the associated economic and societal burdens of Chagas in these countries [8].

However, while successful, the Southern Cone Initiative does have several issues that provide reservations to its long term effect and future direction. One of the largest issues that needs to be addressed is possible long term effects on human or environmental health from the widespread insecticide use. Over 2 million houses were sprayed with insecticides ranging from chlorinated hydrocarbons to pyrethoids. Additionally in some regions insecticide laden paints were used outside and inside houses [7]. The most commonly used insecticide was lindane, a chlorinated hydrocarbon, the safety of which in agriculture and pharmacology has been widely debated. This has resulted in the chemical being banned, except for occasional pharmacological applications (e.g., third line treatment for lice), in over 50 countries including the United States [9]. Disregarding the possible benefits for reducing vector transmission of Chagas disease, the human toxicity and environmental pollution associated with lindane and other insecticides should be considered when evaluating their costs and benefit, in Chagas vector control.

Other potential challenges of the Southern Cone Initiative include the reemergence of *T. infestans* populations. In regions where *T.* infestans was originally endemic without recurrent insecticide spraying the vector is able to repopulate areas where it was previously removed [10]. This not only undermines the goal of the initiative, but it also may be a mechanism for the insects to develop insecticide resistance requiring the use of higher doses or different insecticides, potentially adding to human and environmental health impacts.

Vector control is often recognized as interrupting vector borne transmission rather than disease eradication or permanent reduction because of the potential for vector reemergence or ecological succession (i.e., other vector species replacing a previously common vector species) [8]. Vector control through insecticide use is challenging as a large-scale solution for the reasons mentioned above and because there simply is not one vector for Chagas transmission. There are over 140 species, all within the subfamily Triatominae. Epidemiologically important vector species are largely regional and vector control is complicated by many factors such as population density and the landscape of a region as both contribute to the diversity of vector species in an area.

The last pitfall of the Southern Cone Initiative to be addressed here is that although it reduced the rate and number of new infections, it did not address the population of already infected Chagas patients. Pharmacological advances would address both these issues and could be more stable long-term tools in fighting Chagas disease. New therapeutics aimed at Chagas need to be developed to improve the ability of clinicians to actually treat and possibly one-day cure patients of the disease. Research to develop Chagas drugs should be seen as complimenting the progress of vector control efforts and provide long term solutions to the existing Chagas patient population and offer better outlooks to new Chagas patients. The current pharmacological options available to treat Chagas revolve around two drugs Beznidadazole and Nifurtimox both of which face similar issues in their effectiveness and ability to manage Chagas long term [11].

*Human Pathology of Chagas Disease*

Although Chagas disease in humans occurs mainly through vector transmission, congenital infection is also possible as well as infecting through blood and organ donations [4]. Human Chagas disease has two main phases: acute and chronic. The short acute phase, which lasts from a few weeks to months is largely asymptomatic for many patients. For symptomatic cases, acute Chagas can present many symptoms including fever, edema, swollen lymph nodes (adenopathy) and localized swelling around the parasite entry point, also known as a Chagoma or Romaña's sign. In terms of acute Chagas Chagomas is considered the telltale symptom of *T. cruzi* infection [12]. The severity of the acute phase is contingent on many other factors including mode of parasite acquisition, patient age, patient overall health and the infecting DTU. The predominance of asymptomatic cases of Chagas presents a dilemma as while it is good that the patient is not dealing with potential ill health or life-impacting symptoms, it may mask the presence of the disease overall. This is problematic as the acute phase is the most efficacious time for the current Chagas drugs, Nifurtimox and Benznidazole [13].

Once the disease progresses past the acute phase the parasite enters a largely asymptomatic indeterminate phase which last 10-30 years after the initial infection. At least 30% of patients have the reemergence of symptoms in the chronic phase [12]. Progression of chronic disease within humans is somewhat associated with the particular genetic strain of the infecting parasite, but in many aspects is poorly understood [14]. Chagas disease that reaches the chronic phase mainly affects tissues of three organs; cardiac, esophageal, and intestinal. For example, the strain recognized as TcI often

progresses to Chagasic cardiomyopathy. While other strains have also been shown to cause Chagasic cardiomyopathy these strains have also been linked to megaesophagus and megacolon [14]. In this way diagnosing the infecting DTU becomes important to predict the prognosis and further development of the disease, and thus may be useful as a diagnostic tool.

While there is diversity in symptoms, severity, and progression of the disease, the therapeutic tools available lack diversity. Chagas disease as mentioned previously is treated predominantly with two drug options Nifurtimox and Benznidazole, both of which have toxic effects on the patient [15]. For example Benznidazole causes neutropenia and Nifurtimox can cause anorexia [16]. Overall, Benznidazole has less common but more debilitating side effects such as dermal hypersensitivity and polyneuropathy [11]. Another issue with the availability of only these two drugs is the possibility of *T. cruzi* evolving single drug resistance, and the more challenging possibility of cross-resistance between the drugs. One of the proposed contributing factors to the evolution of drug resistance is the inability to complete treatment regimens. A typical Nifurtimox course is 10mg/kg/day for 60 or 120 days, the long time frame, coupled with the side effects of the drug courses have the consequence that drug treatments are often abandoned early [16]. Compounding this issue, studies have shown cross resistance, the development of resistance to Nifurtimox by *T. cruzi* includes cross-resistance to Benznidazole, effectively removing both treatment options for those patients [16].

This introductory chapter has introduced some of the ecology, pharmacology and challenges for reducing Chagas disease. The next chapter explores genetic variation in a possible parasite protein trans-sialidase suggested as a possible drug target.

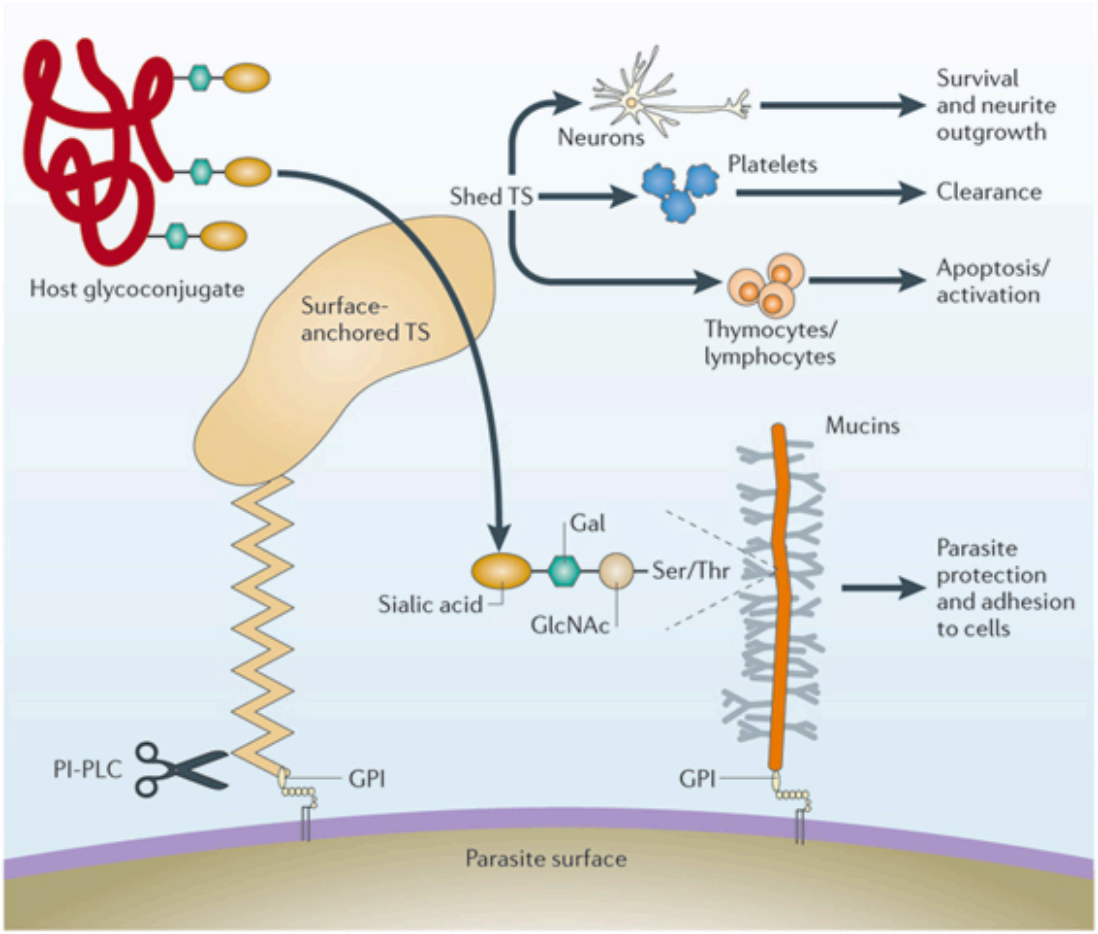





**Figure 1: Surface anchored trans-sialidase (TS) proteins on the parasite outer membrane mediate transfer of sialic acid from the host glycoconjugate to parasite surface mucins (sialylation) within a mammal host. The sialylated mucin provides parasite protection from the host immune system and helps in parasite adhesion to host cells. TS proteins shed into the host bloodsteam can interact with several cell types leading to multiple downstream biological effects. Used with permission from Nature Reviews: Microbiology (http://www.nature.com/nrmicro/journal/v4 /n3/box/nrmicro1351_BX1.html)**

## References


1. Coura JR, Borges-Pereira J. Chagas disease: 100 years after its discovery. A systemic review. Acta tropica. 2010;115(1):5-13.
2. Nunes MCP, Dones W, Morillo CA, Encina JJ, Ribeiro AL. Chagas Disease. Journal of the American College of Cardiology. 2013;62(9):767-76. doi: http://dx.doi.org/10.1016/j.jacc.2013.05.046.
3. Miles MA, Llewellyn MS, Lewis MD, Yeo M, Baleela R, Fitzpatrick S, et al. The molecular epidemiology and phylogeography of Trypanosoma cruzi and parallel research on Leishmania: looking back and to the future. Parasitology. 2009;136(12):1509-28. doi: 10.1017/S0031182009990977. PubMed PMID: 19691868.
4. Montgomery SP, Parise ME, Dotson EM, Bialek SR. What Do We Know about Chagas Disease in the United States? The American Journal of Tropical Medicine and Hygiene. 2016. doi: 10.4269/ajtmh.16-0213.
5. Rassi A, Rassi A, Marin-Neto JA. Chagas disease. The Lancet. 2010;375(9723):1388-402. doi: http://dx.doi.org/10.1016/S0140-6736(10)60061-X.
6. Dias JCP. Southern Cone Initiative for the elimination of domestic populations of Triatoma infestans and the interruption of transfusion Chagas disease: historical aspects, present situation, and perspectives. Memórias do Instituto Oswaldo Cruz. 2007;102:11-8.
7. Dias JCP, Silveira AC, Schofield CJ. The impact of Chagas disease control in Latin America: a review. Memórias do Instituto Oswaldo Cruz. 2002;97(5):603-12.
8. Moncayo A. Chagas disease: current epidemiological trends after the interruption of vectorial and transfusional transmission in the Southern Cone countries. Memórias do Instituto Oswaldo Cruz. 2003;98(5):577-91.
9. Nolan K, Kamrath J, Levitt J. Lindane Toxicity: A Comprehensive Review of the Medical Literature. Pediatric Dermatology. 2012;29(2):141-6. doi: 10.1111/j.1525-1470.2011.01519.x.
10. Delgado S, Ernst KC, Pumahuanca MLH, Yool SR, Comrie AC, Sterling CR, et al. A country bug in the city: urban infestation by the Chagas disease vector Triatoma infestans in Arequipa, Peru. International journal of health geographics. 2013;12(1):48.
11. Castro JA, deMecca MM, Bartel LC. Toxic Side Effects of Drugs Used to Treat Chagas' Disease (American Trypanosomiasis). Human & Experimental Toxicology. 2006;25(8):471-9. doi: 10.1191/0960327106het653oa.
12. Prata A. Clinical and epidemiological aspects of Chagas disease. The Lancet Infectious Diseases. 1(2):92-100. doi: 10.1016/S1473-3099(01)00065-2.
13. Bern C, Montgomery SP. An Estimate of the Burden of Chagas Disease in the United States. Clinical Infectious Diseases. 2009;49(5):e52-e4. doi: 10.1086/605091.
14. Zingales B, Miles MA, Moraes CB, Luquetti A, Guhl F, Schijman AG, et al. Drug discovery for Chagas disease should consider Trypanosoma cruzi strain diversity. Memórias do Instituto Oswaldo Cruz. 2014;109(6):828-33. doi: 10.1590/0074-0276140156.
15. Coura JR, De Castro SL. A critical review on Chagas disease chemotherapy. Memórias do Instituto Oswaldo Cruz. 2002;97(1):3-24.

16. Wilkinson SR, Taylor MC, Horn D, Kelly JM, Cheeseman I. A mechanism for cross-resistance to nifurtimox and benznidazole in trypanosomes. Proceedings of the National Academy of Sciences. 2008;105(13):5022-7.

**Chapter 2: The role of natural selection in shaping genetic variation in a promising Chagas Disease drug target: *Trypanosoma cruzi* trans-sialidase**

## Introduction

*Trypanosoma cruzi* is a human pathogen responsible for Chagas Disease, also known as American Trypanosomiasis. The disease is endemic in Mexico, Central and South America and is currently affecting an estimated 8-10 million people with another 70 million people at risk [1]. Primarily transmitted by insect vectors of the Triatominae sub-family (Hemiptera: Reduviidae), *T. cruzi* displays complex genetic diversity with resulting complexity in its pathogenic effects within and between mammalian species [2]. Various approaches have been used to characterize the diversity in *T. cruzi* with the general consensus of six major subdivisions known as known as Discrete Typing Units (DTUs) and named TcI – TcVI. A seventh DTU (TcBat), found within TcI has been brought to recent attention, as well as a lineage found only in bats, *T. cruzi marinkelli*, that forms a sister group to TcI-TcVI [3].

This heterogeneity in *T. cruzi* has been shown to be significant in human Chagas Disease as certain progressions of the disease have been linked to particular DTUs. For example, TcI has been linked to the development of Chagasic cardiomyopathies, while other DTUs have been shown to progress to more gastrointestinal symptoms such as megaesophagus [4].

Efforts to reduce autochronous Chagas transmission through vector control have been remarkably effective. However, in regions where sylvatic populations exist, vector control requires repeated use of insecticides around and within homes and such repeated

use of insecticides can have adverse ecological and human health effects as well as lead to insecticide resistance [5]. Further, although vector control methods have interrupted transmission significantly, they fail to address the growing patient population already infected. This combination of factors means that new therapeutics are needed.

The *T. cruzi* trans-sialidase (TcTS) receptor gene has been suggested as a biological target for a rationally designed Chagas drug [6-8,11-13,18]. This receptor is essential for parasite growth, because it is a receptor it can potentially be inhibited [6], and it has been implicated in virulence [7]. The TcTS receptor harvests sialic acid from the mammal host, a molecule critical for eukaryotic proliferation and survival that is produced by mammalian cells but not produced by *T. cruzi.* Among the Trypanosomatidae, which includes other human pathogens such as *Trypanosoma brucei* the etiological agent of African sleeping sickness, only *T. cruzi* is known to be unable to produce sialic acid and thus expresses high levels of the TcTS receptor in the trypomastigote stage (the stage infecting vertebrate blood) [7]. The TcTS receptor captures sialic acid from the blood of its mammalian host, transferring it to surface mucins coating its outer membrane. The sialidated molecule is important in evading immune detection [8]. These features of TcTS make it an excellent target for Chagas rational drug development [7].

Despite the high potential of the TcTS receptor as a target for drug design, high throughput screens for potential TcTS inhibitors have been relatively unsuccessful. One possible reason could be genetic diversity of the receptor [7].

The TcTS genes are part of a large superfamily of over 1400 genes containing the conserved VTVxNVxLYNR motif [9]. These genes form eight groups, of which only

group 1 codes for active trans-sialidases [10] with estimates of 1-32 enzymatically active TcTS gene copies per haploid genome [11]. Of the remaining genes about 700 are functional but produce inactive TS, and approximately 700 are pseudogenes [7].

The TcTS protein has an N-terminal catalytic domain (amino acids 1-371) that ends in a six bladed beta propeller, an alpha helix (amino acids 372-394) connects the catalytic domain to the C-terminal lectin-like domain (amino acids 395-632). Another alpha helix near the C-terminus (amino acids 614-626) is sometimes followed by a variable repetitive (100-500 amino acids) hydrophilic, 12 residue shed acute phase antigen (SAPA) motif [8].

Within the N-terminal catalytic domain, group I active TcTS differ from inactive forms by a single amino acid replacement, Tyr342His. The transfer of sialic acid from host to parasite glycoconjugates catalyzed by TcTS (Figure 2.1a) involves an additional seven amino acids known to bind to sialic acid. During the transfer reaction, Tyr342 is crucial in forming the covalent intermediate resulting from cleaving the sialic acid from the host glycoconjugate (Fig 2.1b) before it is transferred to a parasite glycoconjugate (Fig 2.1c) [12].

Identifying and further exploring the genetic diversity of TcTS is important towards drug development since the structure and function of proteins has been shown to be sensitive to amino acid changes anywhere in a peptide sequence, not just active domains (e.g.,[14]), genetic variation in TcTS may affect susceptibility to inhibitors, and may need to be considered when evaluating TcTS for rational drug design.

Rational drug design incorporates biological information and aims to create more effective and longer lasting therapies. Since drugs act as selective agents, it is

important to understand the evolutionary processes acting on a drug target [15]. There are several statistical tests to determine if DNA sequence variation results from neutral processes, or negative or positive selection [15, 16]. Negative, or purifying selection, is the removal of disadvantageous mutations. Positive selection, or adaptive evolution, includes the increase in frequency of advantageous mutations as well as types of selection important in host-parasite interactions such as balancing selection (also called heterozygote advantage or overdominance) and negative frequency-dependent selection (i.e., rare alleles are favored). These types of selection have different implications for a drug target. It has been suggested that positive selection is not beneficial for a drug target while negative selection does add to a targets justification [17].

We assessed DNA sequence and protein variability in an ~500 bp region of the TcTS gene of *T. cruzi* isolated from the abdomens of *Triatoma dimidiata* the major insect vector in Central America and Mexico and *Triatoma nitida,* a species with a more restricted geographic range that is sometimes found in sympatry with *T. dimidiata.* Our *T. dimidiata* specimens were from Guatemala and El Salvador, and the *T. nitida* from Guatemala. We did Sanger sequencing of 48 PCR products (hereafter referred to as PCRp). Because TcTS is a nuclear gene and insects could potentially be infected with one or more parasite strains and also because TcTS has 1-32 copies per haploid genome, for a subset of the insects we cloned the PCR products prior to sequencing (referred to as the 10 cloned sequences).

The region analyzed starts in the N terminal catalytic domain (amino acids 269 to 371), extends into the alpha helix (372-394) and ends in the C-terminal lectin-like domain (amino acids 373 to 434). This region includes four amino acids important in the binding

of sialic acid, Pro238, Trp312, Arg314, and Tyr342 (Table 2). The sequence data were used to identify the DTU of *T. cruzi* infecting the insects and verify we were examining sequence from the group I active TcTS subfamily. The sequence data were then analyzed for natural selection at both the level of the gene and the level of individual amino acids. First, the McDonald Kreitman (MK) test, which considers an entire gene or region of DNA, was used to test for selection within homogeneous regions (catalytic vs non-catalytic) [23,24]. Second, we used Selecton, an amino acid site-specific test [26,27]. The MK test was only appropriate for testing the cloned sequences, whereas both the cloned and 48 PCRp sequences were examined with Selecton.

## Results

The 58 new TcTS sequences from this study belong to group 1, the only one of the eight groups in the superfamily with active TS. Examination of the cloned TcTS sequences with the MK test indicated significant balancing selection in the catalytic-domain of the TcTS gene. The variation was not significantly different from a neutral model of evolution for the part of the C-terminal lectin-like domain region examined. In addition, for both the cloned and PCRp sequences, the test for selection at each individual amino acid site found evidence of positive and negative selection within the regions examined. Below we present the data supporting these results.

**Confirmation of TcTS sequences as TS group 1**

Phylogenetic analysis indicated that the sequences are from TcTS group 1 genes that code for active TS (Fig. 2.2) with strong statistical support (100% bootstrap support).

As expected, all 58 sequences from this study had Tyr342, the critical amino acid that distinguishes the active and inactive forms of TS.

**Tests for natural selection**

The multi-locus MK test indicated the two regions of the TcTS gene were heterogeneous (Table 1b, $\omega_{MK} = 2.077$, $\chi^2 = 6.205$, $p < 0.02$). We therefore analyzed each region separately with the standard MK test.

We detected statistically significant negative and/or balancing selection in the catalytic domain (Table 2, Neutrality index NI = 4.940, $\chi^2 = 14.348$, $p < 0.001$). Examination of the contingency table (Table 2) shows a higher number of non-synonymous polymorphisms than synonymous substitutions for this region (Pn = 15 > Ps = 8) suggesting balancing selection.

For the non-catalytic region that included the alpha helix and part of the C-terminal lectin-like domain, the pattern of variation is consistent with the neutral model of evolution (Table 3c, Neutrality index NI = 1.091, $\chi^2 = 0.047$, $p > 0.05$).

The test using Selecton that examined selection at each individual amino acid, showed evidence of both positive and negative selection over the region examined (Fig 2.3). Among the 166 amino acids studied, Selecton identified 59 sites of positive selection, including both directional and balancing selection, 81 sites of purifying selection and 26 neutral sites. Overall the model including positive selection is a better fit to the data than the null model that includes only purifying selection and neutral synonymous changes (log likelihood -1227.73, delta-log-likelihood -3.35, p-value < 0.02).

The conservation metric (Fig 2.3) provides an estimate of variation for a particular amino acid. Over the region examined, the conservation metric ranges from 0.5 to 1, with an average value of 0.962. The catalytic domain showed higher average conservation (0.982) than the non-catalytic region (0.930).

Selecton values span the possible range (1 indicates strongest positive, 4 is neutral, 7 is strongest purifying). The average selection value over the region examined was 4.361, the catalytic region was toward more purifying selection (average of 4.767) while the non-catalytic region indicated more positive selection (average of 3.398).

With respect to the amino acids essential to catalytic activity, sites 283, 342 and 357 showed negative selection, having both high conservation values and Selecton values. Site 314 showed evidence of positive or balancing selection indicated by high polymorphism (low conservation) and low Selecton scores, while site 312 was neutral.

Comparison of the Selecton values of the 10 clones with the 48 PCRp sequences shows the PCRp sequences show fewer sites of positive selection (Fig. 2.4). The selecton results of the PCRp sequences identified only 49 positively selected sites, 26 neutral sites, and 91 negatively selected sites. The average selecton value for the entire region examined was 4.6506, as well using the same division in the sequence as above, the catalytic region was more indicating of purifying selection (average of 4.9417) than the non-catalytic region (average of 4.1746). Similarly, the average conservation score for the entire region analyzed was 0.9564. While the catalytic domain region again showed higher conservation (0.9776) than the non-catalytic region (0.9216).

## Discussion

In this study analyzing how natural selection and genetic variation in the TcTS gene might affect its potential for rational drug design for our cloned sequences we found no DNA or amino acid variation at most of the sites essential to catalytic activity, however for one site there was significant diversifying or balancing selection. These results are explained in detail below.

In analyzing selection on our sequences, we first confirmed through phylogenetic analysis that our PCRp and cloned sequences represent TcTS protein group 1. This is important since TcTS group 1 is the only group within the trans-sialidase protein superfamily that produces active trans-sialidase.

The MK test, which was appropriate only for the cloned sequences, detected heterogeneity in selection between the catalytic domain, which showed signs of balancing and negative or purifying selection, and the non-catalytic portion of the protein which was not significantly different from expectations of a neutral model of evolution.

Although the MK test detected an overall signal of balancing or negative selection across the catalytic domain, detailed analysis of individual amino acids with Selecton confirmed the "negative" selection detected by the MK test to be a combination of negative and balancing selection. For the non-catalytic region of the protein, the MK test could not reject a neutral model of evolution.

With respect to the amino acids essential to catalytic activity, three of the five sites showed signs of negative selection, consistent with the explanation of low variation because most mutations are deleterious and recessive and are removed by selection or linger in the population at low frequency (e.g., sites 283, 342 and 357). However,

although most amino acids critical to protein function are under negative selection, one site 314 indicated balancing selection while site 312 has a neutral selecton value.

Analysis with Selecton shows quite a few amino acids under positive (balancing) selection in the C-terminal domain, 24 showed positive selection, 12 sites neutral and 27 sites showed negative selection.

For the case of the variation at Arg314, there is variation within the clonal sequences for one sample. However, the all other clonal samples and the 48 PCRp sequences are conserved for this site with no DNA variation.

Viewing the Selecton values on the 3D structure can help to identify how selection is distributed structurally. Projection of the selecton scores on the 3D structure of the protein (Fig. 2.4) reveals that the region of the catalytic domain sequenced in this study contains 60 of the overall 81 negatively selected sites indicating that there is strong constraint on the TcTS structure.

A multiple sequence alignment (MSA) was constructed with the PCRp and clone sequences combined with Genbank reference samples spanning all eight TS gene and pseudogene groups. The MSA facilitates examination of the amino acids essential to catalytic activity in the region sequenced in this study (S1). Among the 58 PCRp and Clone samples there is no DNA variation for three of the amino acids essential to TS transfer activity. In contrast, TS groups 2-8 show some significant variation at these sites. The amino acid variation at these sites was not conserved with any discernable pattern, that is TS groups 2 - 8 did not have distinct variation based on the respective TS group, and no pattern was discernable based on the taxa of the sequences.

How do these results inform rational drug design? Regions of purifying selection are considered strong targets. We detect strong negative selection for the amino acid that distinguishes active and inactive TcTS, Tyr342.

In contrast, regions of positive selection indicating adaptive evolution are not desirable drug targets. This is because positively selected regions show variability and the potential of the parasite to develop drug resistance. Drug targets should avoid positively selected amino acids such as Arg314, an amino acid important in the carboxylation step of TcTS activity.

Drug design should work to develop inhibitors acting specifically at sites such as Tyr342 that are crucial to the protein function and under negative (purifying) selection. Specificity ensures an inhibitor is acting on the desired amino acid residue and can be measured experimentally. A promising inhibitor believed to be specific for Tyr342 could tested by mutating or removing Tyr342 (e.g., through site-directed mutagenesis) and determining if the inhibitor exhibits lower affinity. For example, a recent study using drug repositioning to identify possible trypanocidal agents acting on TcTS [18] could also test the specificity of promising compounds with respect to identified critical and negatively selected amino acids. Trypanocidal agents specific to sites under negative selection would be preferential for long term drug efficacy than ones specific to sites under positive selection. Drug development for Chagas disease may consider not only anti-parasitic drugs but also anti-virulent drugs.

In summary, this study analyzing natural selection and genetic variation in the TcTS gene identified regions of purifying, balancing and positive selection in the TcTS

protein. Rational drug design should consider this variation to increase the likelihood of developing effective drugs with lower chances for the evolution of drug resistance.

## Methods

### Insect vectors, DNA isolation, PCR and sequencing

We examined parasites from two species of insect vectors, *Triatoma dimidiata* and *Triatoma nitida* (Table 1). The *T. dimidiata* were collected from three departments (Huehuetenango, Quiche and Jutiapa) in Guatemala and four in El Salvador (Morazan, Santa Ana, Ahuachapan, Sonsonate). The *T. nitida* were from two departments in Guatemala (Huehuetenango and Chiquimula).

The *T. cruzi* DNA was extracted from the last two segments of each insect abdomen using previously described methods [19]. The PCR amplification used primers (TS 31: TCACGCAGCGGTACGCATCCT, TS 51: GGAGGCTGTCGGCACGCTCTC) specific to group I TS genes that make active trans-sialidases [11]. The PCR conditions were as previously reported [11] and PCR results were confirmed via gel electrophoresis. Samples showing the appropriate size PCR product were sequenced by a commercial facility (Genewiz, South Plainfield, NJ).

The PCR reaction amplified 540 bp and we analyzed a fragment coding for 166 amino acids of the TcTS protein. The region analyzed includes five amino acid residues identified as important to the function and structure of the protein (Table 2): Pro238, Trp312, Arg314, Tyr342 and Glu357 [8] including the critical amino acid Tyr342His that distinguishes the active and inactive forms, but not including the conserved VTVxNVxLYNR motif [9].

The sequence files were trimmed to 498 bp removing the primer regions and edited using Sequencher (V5). Heterozygous peaks were evidence of polymorphism within an insect vector. Two insect vectors were selected for cloning to resolve the heterozygous peaks and identify haplotypes. We cloned fresh TcTS PCR product using the pGEM-T Easy cloning kit (Promega) following the manufacturer's instructions for both ligation and transformation, specifically using the pGEM-T Easy Vector, T4 DNA Ligase, and JM109 High Efficiency Competent Cells (Promega). In order to confirm the success of the transformation both ampicillin selection and a blue-white screen were used. Harvested white colonies were boiled to lyse the cells and extract DNA which was then PCR- amplified and sequenced as described above. The cloned sequences are hereafter referred to as such. The non-cloned, PCR product sequences are referred to as PCRp hereafter.

**Confirmation of TcTS sequences as TS group 1**

Our cloned and PCRp sequences were combined with GenBank TcTS sequences and other *Trypanosoma* species trans-sialidase sequences (Table 1) for a total of 102 sequences; 58 new from this study and 34 from Genbank. The 58 sequences from this study include 48 PCRp sequences and 10 cloned sequences (four from one and six from a second of the 48 PCRp sequences).

Phylogenetic analysis was used to confirm the TcTS group of the 58 sequences new from this study by comparison with reference sequences [9]. The sequences were aligned with ClustalW in MEGA (V7) and translated into amino acid sequences based on

GenBank reference sequences. Alignment and translation was performed with T-Coffee [20-22].

The sequences were analyzed in ProtTest (V2.4) to identify the best model parameters for construction of the phylogenetic tree. RaxML(V8.0.0) was used to construct the phylogeny within the CIPRES Science Gateway V3.3 server combined with bootstrapping to quantify branch support. The optimal tree output was then drawn in FigTree (V1.4.3).

**Tests for natural selection**

Two tests were done to determine the role of natural selection in TcTS DNA sequence variation and evolution, specifically, the McDonald-Kretiman (MK) test [23,24] which considers an entire gene (or portion) and Selecton, an amino acid site-specific test [26,27]. The 10 cloned sequences were examined with both the MK test and Selecton. Because of the ambiguous nucleotide sites, the 48 PCRp sequences could only be examined with Selecton.

To examine the role of neutral vs selective processes for the cloned sequences we used an online version for the MK test [23, 24]. *Trypanosoma brucei* group 1 TS (AF310232) was used as the outgroup [9]. The MK test assumes homogeneity across the gene, but because different regions of the TS gene may evolve differently, we used the multi-locus MK test to test for heterogeneity of part of the region with the N terminal catalytic domain (amino acids 269 to 371) compared to the region with the alpha helix (372-394) and part of the C-terminal lectin-like domain region (amino acids 373 to 434). Then we used the standard MK test to test for selection within homogeneous regions.

The MK test compares the within (polymorphic) *T. cruzi* non-synonymous to synonymous (neutral) changes (Pn/Ps [19]) to the between species (divergent) *T. cruzi* - *T. brucei* non-synonymous to synonymous changes (Dn/Ds) using a Chi-square test. The Neutrality Index (NI) indicates the strength of departure from the neutral model [25]: 1 indicates the data are consistent with a neutral model of evolution, > 1 indicates either Dn is lower than expected due to purifying selection against harmful mutations or Pn is higher than expected due to balancing selection, and < 1 indicates Dn is high due to an excess of fixation of non-neutral replacements from adaptive evolution or Ps is low [25].

In addition, selection at each individual amino acid site was examined and correlated with the protein structure using the Selecton 2007 online server (http://selecton.tau.ac.il/)[26, 27] . The test used Bayesian inference to calculate the dN/dS ratio and used a likelihood ratio test to compare a model with positive selection (their M8) with a null model (M8a) that assumes only purifying selection and neutral changes. The selection values for each amino acid ranged from 1 ( strong negative selection) to 7 (strong positive selection) and were projected on the 3D structure of the protein using the tool First Glance in Jmol (FGiJ, http://firstglance.jmol.org, PDB 1S0I) implemented in Selecton [26, 27].

In interpreting the results, for the MK test, NI < 1 ("positive" selection) refers to adaptive evolution through directional selection; whereas NI > 1 ("negative" selection) includes both purifying and balancing selection. For the Selecton test, negative selection refers to purifying selection, while positive selection refers to adaptive evolution through directional and balancing selection.

Finally, to compare the variation among the cloned sequences with the variation among the PCRp sequences, we calculated the conservation for each amino acid, on a scale of 0 to 1, where no variation = 1. Using the most common amino acid at each site as the reference, each amino acid was assigned a value of 1 if was identical to the reference, 0.5 if the variation did not change the amino acid class (e.g., polar to polar or negative charge to negative charge), and 0 if the variation caused a substitution of a different amino acid class (e.g., polar to negative charge). The sum for each site was divided by the number of total sequences (10 for clones and 48 for PCRp) creating a conservation score between 0 and 1 for each site.

**Acknowledgements**

This work was funded by NSF grant BCS-1216193 as part of the joint NSF-NIH-USDA Ecology and Evolution of Infectious Diseases program and by NIH grant R03AI26268/1-2. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding organizations.

1. WHO (2015) Chagas disease in Latin America: an epidemiological update based on 2010 estimates. (0049-8114 (Print)).
2. Brenière SF, Waleckx E, Barnabé C. Over six thousand Trypanosoma cruzi strains classified into Discrete Typing Units (DTUs): Attempt at an inventory. PLoS neglected tropical diseases. 2016;10(8):e0004792.
3. Zingales B, Miles MA, Campbell DA, Tibayrenc M, Macedo AM, Teixeira MMG, et al. The revised Trypanosoma cruzi subspecific nomenclature: Rationale, epidemiological relevance and research applications. Infection, Genetics and Evolution. 2012;12(2):240-53. doi: http://dx.doi.org/10.1016/j.meegid.2011.12.009.
4. Zingales B, Miles MA, Moraes CB, Luquetti A, Guhl F, Schijman AG, et al. Drug discovery for Chagas disease should consider Trypanosoma cruzi strain diversity. Memórias do Instituto Oswaldo Cruz. 2014;109(6):828-33. doi: 10.1590/0074-0276140156.
5. Moncayo A. Chagas disease: current epidemiological trends after the interruption of vectorial and transfusional transmission in the Southern Cone countries. Memórias do Instituto Oswaldo Cruz. 2003;98(5):577-91.
6. Buscaglia CA, Campo VA, Frasch ACC, Di Noia JM. Trypanosoma cruzi surface mucins: host-dependent coat diversity. Nat Rev Micro. 2006;4(3):229-36.
7. Neres J, Bryce RA, Douglas KT. Rational drug design in parasitology: trans-sialidase as a case study for Chagas disease. Drug Discovery Today. 2008;13(3–4):110-7. doi: http://dx.doi.org/10.1016/j.drudis.2007.12.004.
8. Buschiazzo A, Amaya MaF, Cremona MaL, Frasch AC, Alzari PM. The Crystal Structure and Mode of Action of Trans-Sialidase, a Key Enzyme in Trypanosoma cruzi Pathogenesis. Molecular Cell. 2002;10(4):757-68. doi: http://dx.doi.org/10.1016/S1097-2765(02)00680-9.
9. Chiurillo MA, Cortez DR, Lima FM, Cortez C, Ramírez JL, Martins AG, et al. The diversity and expansion of the trans-sialidase gene family is a common feature in Trypanosoma cruzi clade members. Infection, Genetics and Evolution. 2016;37:266-74. doi: http://dx.doi.org/10.1016/j.meegid.2015.11.024.
10. Freitas LM, Dos Santos SL, Rodrigues-Luiz GF, Mendes TAO, Rodrigues TS, Gazzinelli RT, et al. Genomic analyses, gene expression and antigenic profile of the trans-sialidase superfamily of Trypanosoma cruzi reveal an undetected level of complexity. PLoS One. 2011;6(10):e25914.
11. Burgos JM, Risso MG, Brenière SF, Barnabé C, Campetella O, Leguizamón MS. Differential Distribution of Genes Encoding the Virulence Factor Trans-Sialidase along Trypanosoma cruzi Discrete Typing Units. PLOS ONE. 2013;8(3):e58967. doi: 10.1371/journal.pone.0058967.
12. Watts AG, Damager I, Amaya ML, Buschiazzo A, Alzari P, Frasch AC, et al. Trypanosoma c ruzi Trans-sialidase Operates through a Covalent Sialyl− Enzyme Intermediate: Tyrosine Is the Catalytic Nucleophile. Journal of the American Chemical Society. 2003;125(25):7532-3.
13. De Celis SSCR. Surface Topology Evolution of Trypanosoma Trans-Sialidase. In: Santos ALS, Branquinha MH, d'Avila-Levy CM, Kneipp LF, Sodré CL, editors. Proteins and Proteomics of Leishmania and Trypanosoma. Dordrecht: Springer Netherlands; 2014. p. 203-16.

14. Dong Y, Somero GN. Temperature adaptation of cytosolic malate dehydrogenases of limpets (genus <em>Lottia</em>): differences in stability and function due to minor changes in sequence correlate with biogeographic and vertical distributions. Journal of Experimental Biology. 2008;212(2):169.
15. Suzuki Y. Natural Selection on the Influenza Virus Genome. Molecular Biology and Evolution. 2006;23(10):1902-11. doi: 10.1093/molbev/msl050.
16. Suzuki Y. Detection of positive selection eliminating effects of structural constraints in hemagglutinin of H3N2 human influenza A virus. Infection, Genetics and Evolution. 2013;16:93-8. doi: https://doi.org/10.1016/j.meegid.2013.01.017.
17. Searls DB. Pharmacophylogenomics: genes, evolution and drug targets. Nature Reviews Drug Discovery. 2003;2(8):613-23.
18. Lara-Ramirez EE, López-Cedillo JC, Nogueda-Torres B, Kashif M, Garcia-Perez C, Bocanegra-Garcia V, et al. An in vitro and in vivo evaluation of new potential trans-sialidase inhibitors of Trypanosoma cruzi predicted by a computational drug repositioning method. European Journal of Medicinal Chemistry. 2017;132:249-61. doi: http://dx.doi.org/10.1016/j.ejmech.2017.03.063.
19. Pizarro JC, Stevens L. A new method for forensic DNA analysis of the blood meal in Chagas disease vectors demonstrated using Triatoma infestans from Chuquisaca, Bolivia. PLoS One. 2008;3(10):e3585.
20. Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. Journal of molecular biology. 2000;302(1):205-17.
21. Li W, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, et al. The EMBL-EBI bioinformatics web and programmatic tools framework. Nucleic acids research. 2015;43(W1):W580-W4.
22. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, et al. Analysis tool web services from the EMBL-EBI. Nucleic acids research. 2013;41(W1):W597-W600.
23. Egea R, Casillas S, Barbadilla A. Standard and generalized McDonald–Kreitman test: a website to detect selection by comparing different classes of DNA sites. Nucleic acids research. 2008;36(suppl_2):W157-W62.
24. McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in Drosophila. Nature. 1991;351(6328):652.
25. Rand DM, Kann LM. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from Drosophila, mice, and humans. Molecular biology and evolution. 1996;13(6):735-48.
26. Doron-Faigenboim A, Stern A, Mayrose I, Bacharach E, Pupko T. Selecton: a server for detecting evolutionary forces at a single amino-acid site. Bioinformatics. 2005;21(9):2101-3.
27. Stern A, Doron-Faigenboim A, Erez E, Martz E, Bacharach E, Pupko T. Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. Nucleic acids research. 2007;35(suppl_2):W506-W11.

Table 1: Comprehensive list of newly sequenced samples and Genbank reference samples used in this study.

| Species | Sample or Isolate ID | Vector Species | TS Protein Group | Country collected | Department | Latitude (N) | Longitude (W) | Accession Number |
|---|---|---|---|---|---|---|---|---|
| *Trypanosoma cruzi* | A10050-TS | *Triatoma dimidiata* | TS group 1 ( *T. cruzi clade*) | Guatemala | Huehuetenango | 15° 21´30.815¨ | 91° 27´5.12¨ | |
| *T. cruzi* | A10051-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | Guatemala | Huehuetenango | 15° 21´30.815¨ | 91° 27´5.12¨ | |
| *T. cruzi* | A10052-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | Guatemala | Huehuetenango | 15° 21´30.815¨ | 91° 27´5.12¨ | |
| *T. cruzi* | A10053-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | Guatemala | Huehuetenango | 15° 21´30.815¨ | 91° 27´5.12¨ | |
| *T. cruzi* | A10054-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | Guatemala | Huehuetenango | | | |
| *T. cruzi* | A10055-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | Guatemala | Huehuetenango | | | |
| *T. cruzi* | A10062-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | Guatemala | Huehuetenango | 15°24'00" | 91°58'00" | |
| *T. cruzi* | A10063-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | Guatemala | Huehuetenango | 15°24'00" | 91°58'00" | |
| *T. cruzi* | A10067-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | Guatemala | Huehuetenango | 15° 40´46.271¨ | 91° 49´7.159¨ | |
| *T. cruzi* | A10071-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | Guatemala | Huehuetenango | 15° 21´30.815¨ | 91° 27´5.12¨ | |
| *T. cruzi* | A10083-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | Guatemala | Huehuetenango | | | |
| *T. cruzi* | A10227-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | Guatemala | Huehuetenango | | | |
| *T. cruzi* | A10230-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | Guatemala | Huehuetenango | 15° 39´ 5.85¨ | 91° 46´16.903¨ | |
| *T. cruzi* | A10231-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | Guatemala | Huehuetenango | 15° 39´ 5.85¨ | 91° 46´16.903¨ | |
| *T. cruzi* | A10232-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | Guatemala | Huehuetenango | 15° 39´ 5.85¨ | 91° 46´16.903¨ | |
| *T. cruzi* | A10250-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | Guatemala | Huehuetenango | 15°20'35" | 91°18'42" | |
| *T. cruzi* | A10304-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | Guatemala | Huehuetenango | | | |
| *T. cruzi* | A10305-TS | *Triatoma nitida* | TS group 1 ( *T. cruzi clade*) | Guatemala | Huehuetenango | 15° 19´10.709¨ | 91° 29´ 32.358¨ | |
| *T. cruzi* | A10308-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | Guatemala | Huehuetenango | | | |
| *T. cruzi* | A10311-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | Guatemala | Huehuetenango | | | |
| *T. cruzi* | A10313-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | Guatemala | Huehuetenango | | | |
| *T. cruzi* | A8974-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | Guatemala | Quiche | 14° 55´23.401¨ | 90° 39´45.567¨ | |
| *T. cruzi* | A9948-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | Guatemala | Jutiapa | 14° 16´44.899¨ | 89° 52´31.972¨ | |
| *T. cruzi* | A9950-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | Guatemala | Jutiapa | 14° 16´44.899¨ | 89° 52´31.972¨ | |
| *T. cruzi* | FER530-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | ElSalvador | Morazan | 13 55' 46" N | 88 11' 12" W | |
| *T. cruzi* | FER535-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | ElSalvador | Morazan | 14°04'49.0" | 89°32'09.9" | |
| *T. cruzi* | S234-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | ElSalvador | SantaAna | 14°03'54.6" | 89°31'46.3" | |
| *T. cruzi* | S236-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | ElSalvador | SantaAna | 14°03'54.6" | 89°31'46.3" | |
| *T. cruzi* | S238-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | ElSalvador | SantaAna | 14°03'54.6" | 89°31'46.3" | |
| *T. cruzi* | S328-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | ElSalvador | Ahuachapan | 13° 58' 0" | 89° 49' 0" | |
| *T. cruzi* | S354-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | ElSalvador | Sonsonate | 13° 42' 0" | 89° 37' 0" | |
| *T. cruzi* | S384-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | ElSalvador | Ahuachapan | 13° 58' 0" | 89° 49' 0" | |
| *T. cruzi* | S445-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | ElSalvador | Sonsonate | 13° 45' 43" | 89° 31' 36" | |
| *T. cruzi* | S448-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | ElSalvador | Sonsonate | 13° 45' 43" | 89° 31' 36" | |
| *T. cruzi* | S455-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | ElSalvador | Sonsonate | 13° 45' 43" | 89° 31' 36" | |
| *T. cruzi* | S478-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | ElSalvador | SantaAna | 13° 52´ 59¨ | 89° 28´ 59¨ | |
| *T. cruzi* | S486-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | ElSalvador | SantaAna | 13° 52´ 59¨ | 89° 28´ 59¨ | |
| *T. cruzi* | S507b-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | ElSalvador | SantaAna | 13° 52´ 59¨ | 89° 28´ 59¨ | |
| *T. cruzi* | TPG1019-TS | *T. nitida* | TS group 1 ( *T. cruzi clade*) | Guatemala | Chiquimula | 14° 41´36¨ | 89° 22´23.4¨ | |
| *T. cruzi* | TPG891-TS | *T. nitida* | TS group 1 ( *T. cruzi clade*) | Guatemala | Chiquimula | 14° 41´36¨ | 89° 22´23.4¨ | |
| *T. cruzi* | TPG892-TS | *T. nitida* | TS group 1 ( *T. cruzi clade*) | Guatemala | Chiquimula | 14° 41´36¨ | 89° 22´23.4¨ | |
| *T. cruzi* | TPG1239-TS | *T. nitida* | TS group 1 ( *T. cruzi clade*) | Guatemala | Chiquimula | 14° 41´36¨ | 89° 22´23.4¨ | |
| *T. cruzi* | TPG1247-TS | *T. nitida* | TS group 1 ( *T. cruzi clade*) | Guatemala | Chiquimula | 14° 41´36¨ | 89° 22´23.4¨ | |
| *T. cruzi* | TPG761-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | Guatemala | Chiquimula | 14° 43´18.36¨ | 89° 16´31.32¨ | |
| *T. cruzi* | TPS053-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | ElSalvador | SantaAna | 14°03'54.6" | 89°31'46.3" | |
| *T. cruzi* | TPS180-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | ElSalvador | SantaAna | 14° 6' 0" | 89° 27' 0" | |
| *T. cruzi* | TPS400-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | ElSalvador | SantaAna | 14° 6' 0" | 89° 27' 0" | |
| *T. cruzi* | TPS68-TS | *T. dimidiata* | TS group 1 ( *T. cruzi clade*) | ElSalvador | SantaAna | 14°4'45" | 89°31'54" | |
| Control Samples | | | | | | | | |
| *T. cruzi* | TcCM187-TS | | TS group 1 ( *T. cruzi clade*) | | | | | |
| *T. cruzi* | TcI-TS | | TS group 1 ( *T. cruzi clade*) | | | | | |
| *T. cruzi* | TcV-TS | | TS group 1 ( *T. cruzi clade*) | | | | | |
| *T. cruzi* | TcVI-TS | | TS group 1 ( *T. cruzi clade*) | | | | | |
| Clonal Samples | | | | | | | | |
| *T. cruzi* | A9950_RE1 | | TS group 1 ( *T. cruzi clade*) | | | | | |
| *T. cruzi* | A9950_RE2 | | TS group 1 ( *T. cruzi clade*) | | | | | |
| *T. cruzi* | A9950_RE3 | | TS group 1 ( *T. cruzi clade*) | | | | | |
| *T. cruzi* | A9950_RE7 | | TS group 1 ( *T. cruzi clade*) | | | | | |
| *T. cruzi* | A9950_RE8 | | TS group 1 ( *T. cruzi clade*) | | | | | |
| *T. cruzi* | A9950_RE12 | | TS group 1 ( *T. cruzi clade*) | | | | | |
| *T. cruzi* | A10055_RE2 | | TS group 1 ( *T. cruzi clade*) | | | | | |
| *T. cruzi* | A10055_RE4 | | TS group 1 ( *T. cruzi clade*) | | | | | |
| *T. cruzi* | A10055_RE5 | | TS group 1 ( *T. cruzi clade*) | | | | | |
| *T. cruzi* | A10055_RE7 | | TS group 1 ( *T. cruzi clade*) | | | | | |
| Genbank Samples | | | | | | | | Accession number |
| T. cruzi | G/TCC30 | *Opossum* | TS group 2 | Brazil | | | | EF154827 |
| T. cruzi | G/TCC31 | *Opossum* | TS group 2 | Brazil | | | | AF426132 |
| T. cruzi | CA-I | *H. sapiens* | TS group 1 | Argentina | | | | X57235 |
| T. cruzi | CL | *Triatoma infestans* | TS group 7 | Brazil | | | | X70948 |
| T. cruzi | CL | *T. infestans* | TS group 2 | Brazil | | | | AF128843 |
| T. cruzi | CL Brener | *T. infestans* | TS group 2 | Brazil | | | | XM_799078 |
| T. cruzi | CL Brener | *T. infestans* | TS group 5 | Brazil | | | | XM_800086 |
| T. cruzi | CL Brener | *T. infestans* | TS group 2 | Brazil | | | | XM_800788 |
| T. cruzi | CL Brener | *T. infestans* | TS group 1 | Brazil | | | | XM_802406 |
| T. cruzi | CL Brener | *T. infestans* | TS group 3 | Brazil | | | | XM_802711 |
| T. cruzi | CL Brener | *T. infestans* | TS group 4 | Brazil | | | | XM_803072 |
| T. cruzi | CL Brener | *T. infestans* | TS group 6 | Brazil | | | | XM_803086 |
| T. cruzi | CL Brener | *T. infestans* | TS group 5 | Brazil | | | | XM_803518 |
| T. cruzi | CL Brener | *T. infestans* | TS group 2 | Brazil | | | | XM_805296 |
| T. cruzi | CL Brener | *T. infestans* | TS group 7 | Brazil | | | | XM_805583 |
| T. cruzi | CL Brener | *T. infestans* | TS group 4 | Brazil | | | | XM_806976 |
| T. cruzi | CL Brener | *T. infestans* | TS group 8 | Brazil | | | | XM_807627 |
| T. cruzi | CL Brener | *T. infestans* | TS group 1 | Brazil | | | | XM_808522 |
| T. cruzi | CL Brener | *T. infestans* | TS group 8 | Brazil | | | | XM_808523 |
| T. cruzi | CL Brener | *T. infestans* | TS group 6 | Brazil | | | | XM_809532 |
| T. cruzi | CL Brener | *T. infestans* | TS group 4 | Brazil | | | | XM_810613 |
| T. cruzi | CL Brener | *T. infestans* | TS group 2 | Brazil | | | | XM_812072 |
| T. cruzi | CL Brener | *T. infestans* | TS group 2 | Brazil | | | | XM_811278 |
| T. cruzi | CL Brener | *T. infestans* | TS group 2 | Brazil | | | | XM_811657 |
| T. cruzi | CL Brener | *T. infestans* | TS group 3 | Brazil | | | | XM_814626 |
| T. rangeli | DOG82 | *Dog* | TS group 2 | Venezuela | | | | FJ404802 |
| T. rangeli | DOG83 | *Dog* | TS group 2 | Venezuela | | | | FJ404803 |
| T. rangeli | DOG84 | *Dog* | TS group 2 | Venezuela | | | | AF426022 |
| T. rangeli | Choachi | *R. prolixus* | TS group 8 | Colombia | | | | KC544946 |
| T. rangeli | Choachi | *R. prolixus* | TS group 5 | Colombia | | | | KC544956 |
| T. Brucei | TREU927 | *Glossina sp* | TS group 1 | Kenya | | | | XM_842470 |
| T. Brucei | EATRO 427 | *Sheep* | TS group 1 | Uganda | | | | AF310232 |
| T. carassii | | *cyprinid fish* | TS group 1 | | | | | AY142111 |
| T. grayi | ANR4 | *G. palpalis gambiensis* | TS group 1 | Gambia | | | | XM_009317479 |

Table 2: Genetic variation and selection for the TcTS amino acids essential to catalytic activity in the region sequenced in this study.

| Residue | Possible Role | Clone Variation | Clone Selection | PCRp Variation | PCRp Selection |
|---|---|---|---|---|---|
| Pro 283 | Transglycosylation/Hydroxylation | None | Purifying | None | Purifying |
| Trp 312 | Aromatic Sandwich/Hydrogen Bonding | None | Positive | None | Purifying |
| Arg 314 | Carboxylate Fixation | Non-synonymous | Neutral | None | Purifying |
| Tyr 342 | Enzymatic Nucleophile | None | Purifying | None | Purifying |
| Glu 357 | Catalysis | None | Purifying | None | Purifying |

Table 3: Results of multi-locus McDonald Kreitman test for the N-terminal catalytic domain as region 1 and the alpha helix domain and C-terminal lectin like domain as region 2

McDonald Kretiman Test

Multi-locus MK Results

| | | Polymorphism | Divergence | Total | Neutrality Index | $\chi^2$ | p-value |
|---|---|---|---|---|---|---|---|
| Region 1 | Neutral | 8 | 209.8 | 217.8 | | | |
| | Non-neutral | 15 | 79.62 | 94.62 | 4.94 | 14.348 | <0.000 |
| | Total | 23 | 289.42 | 312.42 | | | |

| | | Polymorphism | Divergence | Total | Neutrality Index | $\chi^2$ | p-value |
|---|---|---|---|---|---|---|---|
| Region 2 | Neutral | 13 | 32.76 | 45.76 | | | |
| | Non-neutral | 26 | 60.04 | 86.04 | 1.091 | 0.047 | 0.828 |
| | Total | 39 | 92.8 | 131.8 | | | |

Multi-Locus Mantel-Haenszel

| | $\chi^2$ | p-value |
|---|---|---|
| Mantel-Haenszel Test of homogeneity | 6.205 | 0.012 |

| | Neutrality Index | $\chi^2$ | p-value |
|---|---|---|---|
| Mantel-Haenszel Estimator | 2.077 | 6.047 | 0.013 |

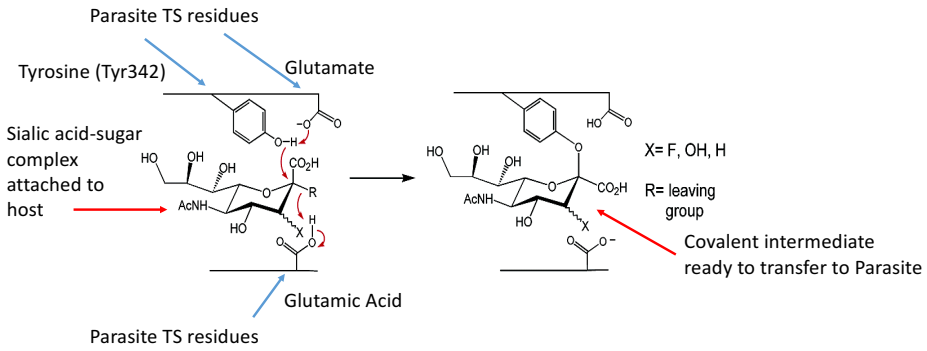
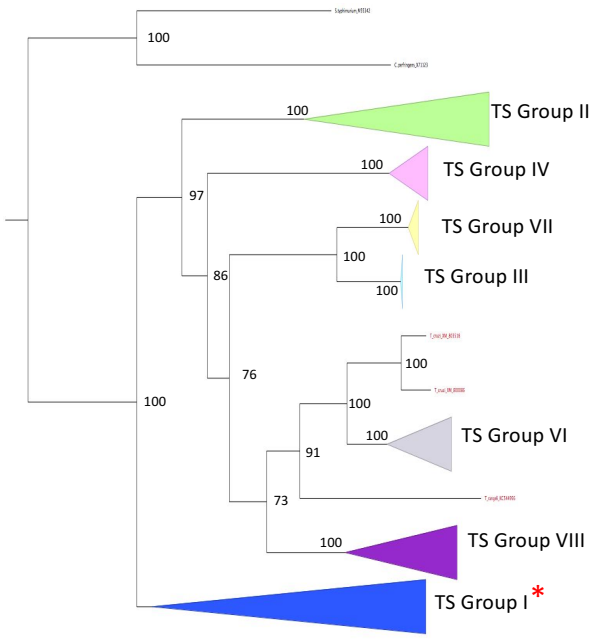



**Fig 2.1** Chemical reaction catalyzed by *Trypanosoma cruzi* trans sialidase (TcTS).. Mechanism of removal of sialic acid from host, creating a stable covalent intermediate. Used with permission from: http://pubs.rsc.org/en/content/articlehtml/2011/OB/C0OB00826E#imgfig1




* All samples unique to this study resolve in TS Group 1. See Table 1 for list of samples.

**Fig 2.2** Best Maximum Likelihood tree reconstruction for the trans-sialidase (TS) protein family constructed from *T. cruzi* TS samples and GenBank reference samples for each representative group for the TS protein family. Boostrap values (0-100) are indicated at branch nodes.

**Fig 2.3**

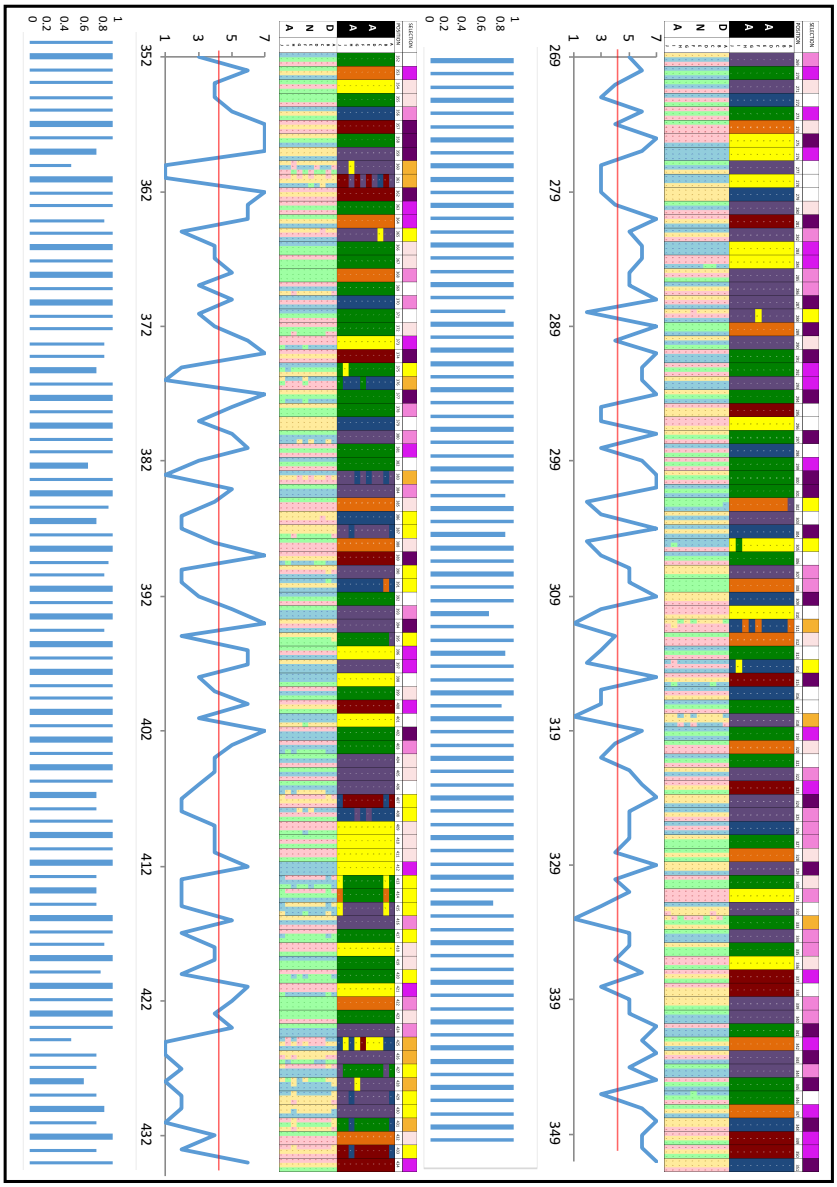


**Fig 2.3** Summary of results, showing DNA and amino acid sequence variation, and results of the analysis of Selecton test for natural selection. Based on analysis of the 10 clone sequences. Rows (identified with ALL CAPITAL LETTERS) indicate: SELECTON, the Selecton score (Orange is strong positive selection, light pink is neutral and purple is strong negative selection); POSITION, the amino acid position in the TcTS protein (our sequences cover amino acids 269-434); AA, the amino acid for each of the 10 clones; and DNA, the corresponding nucleotides for each amino acid. The blue line shows the Selecton score for each amino acid with the red horizontal line indicating neutral evolution or no selection. The blue histograms show the conservation of each amino acid site (1 = no variation, 0 = all sites unique).

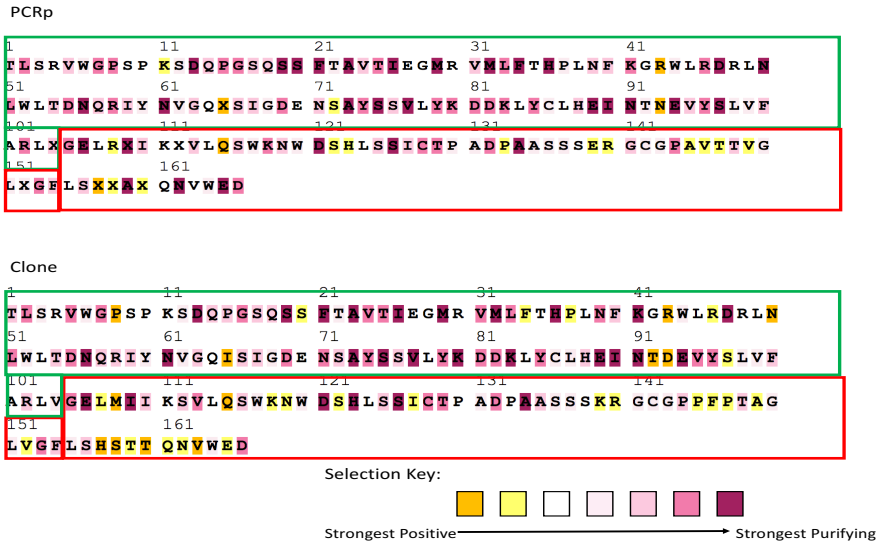



**Fig. 2.4** Comparison of Selection values between the clone and 48 PCRp sequences. Orange indicates strong positive selection, yellow is neutral and magenta is strong negative selection.The catalytic region is outlined in green and the non-catalytic region is outlined by red.

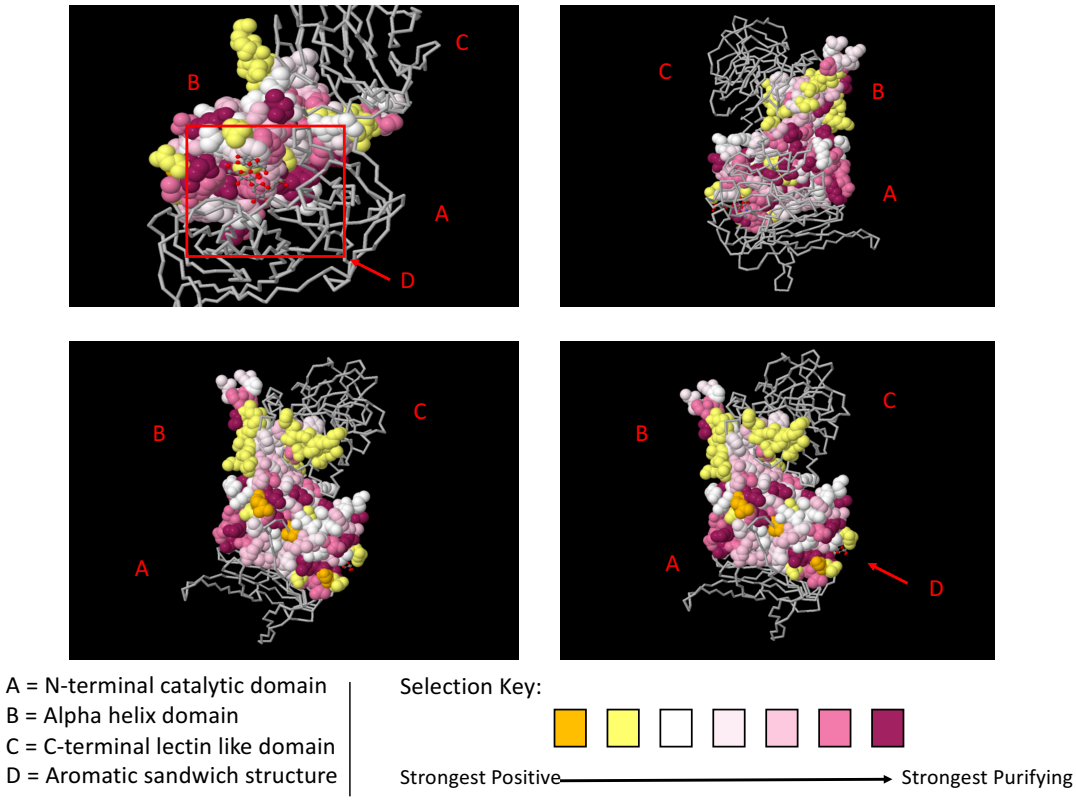



**Fig. 2.5** Three-dimension projection of *Trypanosoma cruzi* trans-sialidase (TcTS) molecule with amino acids shaded to indicate Selection values. A. N-terminal catalytic domain. B. alpha-helix domain, C. C-terminal lectin-like domain. D. Highlight of aromatic sandwich (Trp312) of the catalytic domain.

## Comprehensive Bibliography


Bern C, Montgomery SP. An Estimate of the Burden of Chagas Disease in the United States. Clinical Infectious Diseases. 2009;49(5):e52-e4. doi: 10.1086/605091.

Brenière SF, Waleckx E, Barnabé C. Over six thousand Trypanosoma cruzi strains classified into Discrete Typing Units (DTUs): Attempt at an inventory. PLoS neglected tropical diseases. 2016;10(8):e0004792.

Burgos JM, Risso MG, Brenière SF, Barnabé C, Campetella O, Leguizamón MS. Differential Distribution of Genes Encoding the Virulence Factor Trans-Sialidase along Trypanosoma cruzi Discrete Typing Units. PLOS ONE. 2013;8(3):e58967. doi: 10.1371/journal.pone.0058967.

Buscaglia CA, Campo VA, Frasch ACC, Di Noia JM. Trypanosoma cruzi surface mucins: host-dependent coat diversity. Nat Rev Micro. 2006;4(3):229-36.

Buschiazzo A, Amaya MaF, Cremona MaL, Frasch AC, Alzari PM. The Crystal Structure and Mode of Action of Trans-Sialidase, a Key Enzyme in Trypanosoma cruzi Pathogenesis. Molecular Cell. 2002;10(4):757-68. doi: http://dx.doi.org/10.1016/S1097-2765(02)00680-9.

Castro JA, deMecca MM, Bartel LC. Toxic Side Effects of Drugs Used to Treat Chagas' Disease (American Trypanosomiasis). Human & Experimental Toxicology. 2006;25(8):471-9. doi: 10.1191/0960327106het653oa.

Chiurillo MA, Cortez DR, Lima FM, Cortez C, Ramírez JL, Martins AG, et al. The diversity and expansion of the trans-sialidase gene family is a common feature in Trypanosoma cruzi clade members. Infection, Genetics and Evolution. 2016;37:266-74. doi: http://dx.doi.org/10.1016/j.meegid.2015.11.024.

Coura JR, Borges-Pereira J. Chagas disease: 100 years after its discovery. A systemic review. Acta tropica. 2010;115(1):5-13.

Coura JR, De Castro SL. A critical review on Chagas disease chemotherapy. Memórias do Instituto Oswaldo Cruz. 2002;97(1):3-24.

De Celis SSCR. Surface Topology Evolution of Trypanosoma Trans-Sialidase. In: Santos ALS, Branquinha MH, d'Avila-Levy CM, Kneipp LF, Sodré CL, editors. Proteins and Proteomics of Leishmania and Trypanosoma. Dordrecht: Springer Netherlands; 2014. p. 203-16.

Delgado S, Ernst KC, Pumahuanca MLH, Yool SR, Comrie AC, Sterling CR, et al. A country bug in the city: urban infestation by the Chagas disease vector Triatoma infestans in Arequipa, Peru. International journal of health geographics. 2013;12(1):48.

Dias JCP. Southern Cone Initiative for the elimination of domestic populations of Triatoma infestans and the interruption of transfusion Chagas disease: historical aspects, present situation, and perspectives. Memórias do Instituto Oswaldo Cruz. 2007;102:11-8.

Dias JCP, Silveira AC, Schofield CJ. The impact of Chagas disease control in Latin America: a review. Memórias do Instituto Oswaldo Cruz. 2002;97(5):603-12.

Dong Y, Somero GN. Temperature adaptation of cytosolic malate dehydrogenases of

limpets (genus <em>Lottia</em>): differences in stability and function due to minor changes in sequence correlate with biogeographic and vertical distributions. Journal of Experimental Biology. 2008;212(2):169.
Doron-Faigenboim A, Stern A, Mayrose I, Bacharach E, Pupko T. Selecton: a server for detecting evolutionary forces at a single amino-acid site. Bioinformatics. 2005;21(9):2101-3.
Egea R, Casillas S, Barbadilla A. Standard and generalized McDonald–Kreitman test: a website to detect selection by comparing different classes of DNA sites. Nucleic acids research. 2008;36(suppl_2):W157-W62.
Freitas LM, Dos Santos SL, Rodrigues-Luiz GF, Mendes TAO, Rodrigues TS, Gazzinelli RT, et al. Genomic analyses, gene expression and antigenic profile of the trans-sialidase superfamily of Trypanosoma cruzi reveal an undetected level of complexity. PLoS One. 2011;6(10):e25914.
Lara-Ramirez EE, López-Cedillo JC, Nogueda-Torres B, Kashif M, Garcia-Perez C, Bocanegra-Garcia V, et al. An in vitro and in vivo evaluation of new potential trans-sialidase inhibitors of Trypanosoma cruzi predicted by a computational drug repositioning method. European Journal of Medicinal Chemistry. 2017;132:249-61. doi: http://dx.doi.org/10.1016/j.ejmech.2017.03.063.
Li W, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, et al. The EMBL-EBI bioinformatics web and programmatic tools framework. Nucleic acids research. 2015;43(W1):W580-W4.
McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in Drosophila. Nature. 1991;351(6328):652.
McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, et al. Analysis tool web services from the EMBL-EBI. Nucleic acids research. 2013;41(W1):W597-W600.
Miles MA, Llewellyn MS, Lewis MD, Yeo M, Baleela R, Fitzpatrick S, et al. The molecular epidemiology and phylogeography of Trypanosoma cruzi and parallel research on Leishmania: looking back and to the future. Parasitology. 2009;136(12):1509-28. doi: 10.1017/S0031182009990977. PubMed PMID: 19691868.
Moncayo A. Chagas disease: current epidemiological trends after the interruption of vectorial and transfusional transmission in the Southern Cone countries. Memórias do Instituto Oswaldo Cruz. 2003;98(5):577-91.
Montgomery SP, Parise ME, Dotson EM, Bialek SR. What Do We Know about Chagas Disease in the United States? The American Journal of Tropical Medicine and Hygiene. 2016. doi: 10.4269/ajtmh.16-0213.
Neres J, Bryce RA, Douglas KT. Rational drug design in parasitology: trans-sialidase as a case study for Chagas disease. Drug Discovery Today. 2008;13(3–4):110-7. doi: http://dx.doi.org/10.1016/j.drudis.2007.12.004.
Nolan K, Kamrath J, Levitt J. Lindane Toxicity: A Comprehensive Review of the Medical Literature. Pediatric Dermatology. 2012;29(2):141-6. doi: 10.1111/j.1525-1470.2011.01519.x.
Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. Journal of molecular biology. 2000;302(1):205-17.
Nunes MCP, Dones W, Morillo CA, Encina JJ, Ribeiro AL. Chagas Disease. Journal of

the American College of Cardiology. 2013;62(9):767-76. doi: http://dx.doi.org/10.1016/j.jacc.2013.05.046.
Pizarro JC, Stevens L. A new method for forensic DNA analysis of the blood meal in Chagas disease vectors demonstrated using Triatoma infestans from Chuquisaca, Bolivia. PLoS One. 2008;3(10):e3585.
Prata A. Clinical and epidemiological aspects of Chagas disease. The Lancet Infectious Diseases. 1(2):92-100. doi: 10.1016/S1473-3099(01)00065-2.
Rand DM, Kann LM. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from Drosophila, mice, and humans. Molecular biology and evolution. 1996;13(6):735-48.
Rassi A, Rassi A, Marin-Neto JA. Chagas disease. The Lancet. 2010;375(9723):1388-402. doi: http://dx.doi.org/10.1016/S0140-6736(10)60061-X.
Searls DB. Pharmacophylogenomics: genes, evolution and drug targets. Nature Reviews Drug Discovery. 2003;2(8):613-23.
Stern A, Doron-Faigenboim A, Erez E, Martz E, Bacharach E, Pupko T. Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. Nucleic acids research. 2007;35(suppl_2):W506-W11.
Suzuki Y. Natural Selection on the Influenza Virus Genome. Molecular Biology and Evolution. 2006;23(10):1902-11. doi: 10.1093/molbev/msl050.
Suzuki Y. Detection of positive selection eliminating effects of structural constraints in hemagglutinin of H3N2 human influenza A virus. Infection, Genetics and Evolution. 2013;16:93-8. doi: https://doi.org/10.1016/j.meegid.2013.01.017.
Watts AG, Damager I, Amaya ML, Buschiazzo A, Alzari P, Frasch AC, et al. Trypanosoma cruzi Trans-sialidase Operates through a Covalent Sialyl− Enzyme Intermediate: Tyrosine Is the Catalytic Nucleophile. Journal of the American Chemical Society. 2003;125(25):7532-3.
Wilkinson SR, Taylor MC, Horn D, Kelly JM, Cheeseman I. A mechanism for cross-resistance to nifurtimox and benznidazole in trypanosomes. Proceedings of the National Academy of Sciences. 2008;105(13):5022-7.
WHO (2015) Chagas disease in Latin America: an epidemiological update based on 2010 estimates. (0049-8114 (Print)).
Zingales B, Miles MA, Campbell DA, Tibayrenc M, Macedo AM, Teixeira MMG, et al. The revised Trypanosoma cruzi subspecific nomenclature: Rationale, epidemiological relevance and research applications. Infection, Genetics and Evolution. 2012;12(2):240-53. doi: http://dx.doi.org/10.1016/j.meegid.2011.12.009.
Zingales B, Miles MA, Moraes CB, Luquetti A, Guhl F, Schijman AG, et al. Drug discovery for Chagas disease should consider Trypanosoma cruzi strain diversity. Memórias do Instituto Oswaldo Cruz. 2014;109(6):828-33. doi: 10.1590/0074-0276140156.