ABSTRACT

Title of Dissertation: MOLECULAR EVOLUTIONARY STUDIES ON *TRYPANOSOMA CRUZI*, THE AGENT OF CHAGAS DISEASE

Carlos A. Flores-López, Doctor of Philosophy, 2013

Dissertation directed by: Professor Carlos A. Machado
Department of Biology

The use of DNA sequences to address diverse evolutionary questions has increased steadily with the growing availability of genome sequence data. In this study, I make use of DNA sequence data to describe several evolutionary aspects of the protozoan parasite responsible for Chagas disease, *Trypanosoma cruzi*. Chagas is estimated to infect 7.7 million people and cause the deaths of approximately ten thousand people every year in Latin America. Just like many other parasitic diseases, Chagas does not have a vaccine or an effective drug treatment. In this body of work, I specifically: (1) describe the evolutionary history of the major strains of the parasite through the use of phylogenetic analyses of 32 loci and demonstrate that the parasite's original classification into two major evolutionary lineages does not reflect the evolutionary history of the parasite, (2) demonstrate that there is strong evidence for just one major recent hybridization event during the history of *T. cruzi* divergence and not two as previously suggested, (3) show

that all major extant *T. cruzi* lineages diverged recently (less than 3 million years ago), well before the arrival of humans in the Americas, (4) describe a new *T. cruzi* lineage that appears to have diverged in North America ("TcNA"), (5) show that a significantly larger fraction of protein-coding genes have experienced positive selection in *T. cruzi* than in *Leishmania spp.*, a pattern likely due to the greater versatility of *T. cruzi* in its host range, cell tropism and cell invasion mechanisms, and (6) illustrate a recent major expansion of a few surface protein families in *T.* cruzi that seem to be linked to the evolution of the parasite's ability to invade multiple cell tissues and multiple host species. These results demonstrate the applicability and power of molecular evolutionary analyses for understanding parasitic diseases.

MOLECULAR EVOLUTIONARY STUDIES ON *TRYPANOSOMA CRUZI*, THE AGENT OF CHAGAS DISEASE

by

Carlos A. Flores-López

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2013

Advisory Committee:
Professor Carlos A. Machado, Chair
Professor Michael P. Cummings
Professor Charles Delwiche
Professor Najib El-Sayed
Professor Thomas D. Kocher

© Copyright by

Carlos A. Flores-López

2013

DEDICATION

Esta tesis está dedicada a la persona que me inspiró en los momentos más difíciles a siempre mirar hacia delante. A mi hijo Ulises.

ACKNOWLEDGMENTS

I wouldn’t have been able to complete my PhD education without the wonderful support I have received from my parents Margarita López-Martínez and Carlos F. Flores-Luna. Their leadership by example, love and support since I was kid has always kept me motivated to pursue my goals. If I hadn't looked up to them I would have never been introduced to the many treasures that are connected with a life in science. I would also like to thank my advisor Dr. Carlos A. Machado for all the years I spent under his wing as his student. His support in all the definitions of the word (both work related and personal) has meant so much to me. This work wouldn’t have been possible without his help. I would also like to thank each member of my committee: Dr. Charles Delwiche, Dr. Michael Cummings, Dr. Thomas Kocher and Dr. Najib El-Sayed for the guidance, advice and support I received throughout my years at Maryland. This work would not of been possible without their crucial input. The many members of the Machado lab throughout the years that have been so supportive throughout all these years and who I owe so much, I say thank you.

Lastly I would like to thank all my friends and colleagues that have supported me throughout the years (some a long time ago when this adventure began at the University of Arizona, some more recently when the adventure moved to Maryland, and some all the

way back from México). Without their support and friendship in the stressful and hard times that come along with a PhD and the time spent away from my family and country for so many years, it is undisputable I would not been able to accomplish this work. Alejandra Velázquez, Gabi Flores, Javier and Liza Montenegro, Shane Bryan, Brian Dalton, Laura Rascón, Gladys Erives, Matt Palmer, Kawther Abdilleh, Kevin Nyberg, Bibek Yumman, Guillerhme Targino-Valente, Chris Arias, Matt Palmer, Miguel Carneiro, Rafael Tannus, Gonzálo de León-Girón, Alejandro González, Alonso Gutierrez, Ricardo Cruz, Ricardo Valencia, Sebastian Gómez, Enrique Gómez, Leonardo Carvalho and many others.

Chapter 2. I thank 4 reviewers for their comments, and Michel Tibayrenc and Christian Barnabé for kindly providing *T. cruzi* DNA

Chapter 3. I would like to thank all the collaborators of this project: Mitchell E., Reisenman C.E., and Williamson P. for their insightful comments on the manuscript. I would also like to thank Dr. Christian Barnabé for kindly donating *T. cruzi* reference strains (TcI-VI), to Teresa Gregory and William Savary for help collecting insects.

Chapter 4. I would like to thank Maximilian J. Telford for help on the translation/alignment analysis, and Kawther Abdilleh and Guilherme Targino Valente for constructive discussions.

TABLE OF CONTENTS

Dedication ii
Acknowledgments iii
Table of Contents v
List of Tables vii
List of Figures ix

Chapter 1: Introduction to Dissertation 1

Chapter 2: Analyses of 32 loci clarify phylogenetic relationships among *Trypanosoma cruzi* lineages and support a single hybridization prior to human contact 5
Abstract 6
Introduction 7
Materials and Methods 10
Results 16
Discussion 20
Conclusion 28
Tables 30
Figures 38

Chapter 3: Description of a new *Trypanosoma cruzi* lineage from the United States reveals an introduction into North America during the Pleistocene 41
Abstract 42
Introduction 42
Materials and Methods 46
Results 49
Discussion 52
Conclusions 55
Tables 56
Figures 60

Chapter 4: Positive selection has played a larger role in the evolution of *Trypanosoma cruzi* proteins than in the evolution of *Leishmania spp.* proteins 64
Abstract 65
Introduction 65
Materials and Methods 69
Results 74
Discussion 78
Tables 83
Figures 86

Chapter 5: Genomic changes associated with the evolution of multicellular invasion ability and adaptation to multiple hosts in a pathogen: Protein family expansions in the parasitic lifestyle of *Trypanosoma cruzi* …….. 91
Abstract …… 92
Introduction ……92
Materials and Methods ……...94
Results ……96
Discussion ……..99
Figures ……..105

Chapter 6. Overview and Conclusions to Dissertation ……...108
Overview ……..108
Future directions ……..111
Conclusion ……...113

Appendices ……..114

Bibliography ……...220

LIST OF TABLES

Table 2-1. The main *Trypanosoma cruzi* strains used in this study 30

Table 2-2. List of loci included in this study 31

Table 2-3. Results from the selection, distance and midpoint rooting analyses …….. 34

Table 2-4. Bayesian estimates of divergence time (in mya) for different *T. cruzi* lineages ……. 37

Table 3-1. Information for isolated samples from Texas and Arizona 56

Table 3-2 Bayesian estimates of divergence (time in millions of years) for the main *T. cruzi* lineages and TcNA …… 59

Table 4-1. The number of proteins under positive selection in *T. cruzi* and *Leishmania* spp ……..……..……..……..……..……..……..……..……..……..……..…….. 83

Table 4-2. Functional over representation of genes showing evidence of positive selection ……..……..……..……..……..……..……..……..……..……..……..…….. 84

Table 4-3. Statistically overrepresented hypothetical protein clusters under positive selection in *T. cruzi* and *Leishmania spp*. …….. 85

Table S2-1. Additional *Trypanosoma cruzi* strains used for some of the loci …… 114

Table S2-2. Amplified nuclear loci and PCR primers …… 115

Table S2-3. Results of the Shimodaira-Hasegawa tests …… 117

Table S2-4. LRT of Molecular clock on genes that had a homolog in *T. brucei* …… 118

Table S4-1. *L.major* gene codes for orthologous protein predicted to be under PS with models M8 versus M8a with 4 and 3 taxa …… 139

Table S4-2. *L.major* gene codes for orthologous protein predicted to be under PS with models M7 versus M8 with 4 and 3 taxa …… 155

Table S4-3. *T. cruzi* (Non-Esmeraldo) gene codes for orthologous proteins predicted to be under PS with models M8 versus M8a with 4 and 3 taxa ……. 173

Table S4-4. *T.cruzi* gene codes for orthologous protein predicted to be under PS with models M7 versus M8 with 4 and 3 taxa ……193

Table S4-5. Nucleotide divergences among *T. cruzi* strains (haplotypes) …… 214

Table S4-6. Nucleotide divergences among *Leishmania spp.* species ……214

Table S4-7. The number of proteins under positive selection in *T. cruzi* and *Leishmania* spp. based on analyses of 3 taxa …… 215

Table S4-8. The number of putative false positives. False Positives (FP) …… 215

Table S4-9. Functional over representation of True Positives …….. 216

Table S5-1. Overrepresented chromosomal locations for the major protein family expansions …… 217

LIST OF FIGURES

Figure 2-1. Phylogenetic topologies obtained from the 32 analyzed loci …… 38

Figure 2-2. Maximum likelihood tree of concatenated data set …… 39

Figure 2-3. Divergence times for main DTU clades of *T. cruzi* using nuclear loci with the relaxed clock model. ……. 40

Figure 3-1. Map of sample location in the USA ……..…….. 60

Figure 3-2. Bayesian mitochondrial phylogeny ……. 61

Figure 3-3. Bayesian nuclear phylogeny ……..…….. 62

Figure 3-4. Ancestral TcI and TcNA entries from South America into North America ……..……..……..……..……..……..……..……..…… 63

Figure 4-1. Phylogenies of *Leishmania* species and *T. cruzi* strains used in this study ……..……..……..……..……..……..……..……. 86

Figure 4-2. Difference in the number of proteins showing evidence of positive selection within each taxa ……..…… 87

Figure 4-3A. dN/dS for sites predicted to be evolving under positive selection ……. 88

Figure 4-3B. Proportion of sites predicted to be under positive selection …… 89

Figure 4-4. Comparison of dN/dS values for all protein coding genes …… 90

Figure 5-1. Evolutionary relationships among kinetoplastids ……..…… 105

Figure 5-2. Age distribution of duplicate pairs of genes ……..…… 106

Figure 5-3. Age distribution of duplicate pairs of genes in the surface protein families …….. 107

Figure S2-1. Individual Maximum likelihood trees of each amplified locus ……. 120

Figure S2-2. Divergence times for main DTU clades of *T. cruzi* using nuclear loci with the strict clock model …… 135

Figure S3-1. Maximum likelihood phylogenetic tree of Tc00.1047053506529.310 nuclear loci …… 136

Figure S3-2. Maximum likelihood phylogenetic tree of DHFR nuclear loci …… 137

Figure S3-3. Maximum likelihood phylogenetic tree of MSH2 nuclear loci …… 138

Figure S5-1. Neighbor Joining tree of TS's found in both family expansions …….. 219

Chapter 1: Introduction to Dissertation

Chagas disease

Trypanosoma cruzi is the etiological agent of American Trypanosomiasis, also known as Chagas disease. Carlos Chagas, a Brazilian physician, first described the parasite in 1907. Since its original description, *T. cruzi* has been isolated from more than 100 species of mammals and more than 130 species of Triatomine bugs (Coura and Dias 2009). It is mostly found in South America, Central America and Mexico, although the parasite has been also isolated from sylvatic mammals and triatomines in the Southern United States (Roellig, et al. 2013). *T. cruzi*, like many other closely related parasites, has a heteroxenous lifestyle, thus requiring an insect vector and a mammalian host to complete its life cycle (Barrett, et al. 2003). The parasite has four distinct developmental stages, two in the insect vector (epimastigotes and metacyclic trypomastigotes) and two in the vertebrate host (amastigotes and bloodstream trypomastigotes) (Coura and Dias 2009). The parasite is transmitted to humans through contaminated Triatomine feces and urine. Triatomine insects are hematophagous Hemiptera ("true bugs") of the family Reduviidae, a family of true bugs that are mostly predatory on other insects. The fossil record indicates that the earliest evidence for hematophagy in the Triatominae is approximately 5 million years ago (Gorla, et al. 1997). Once a Triatomine bug feeds on a mammalian host, the bug usually defecates. The flagellated metacyclic trypomastigotes, which are found in the feces and urine, then proceed to enter the vertebrate host through the open skin wound the insect created or through any mucosal membrane (the eye is a very common route of parasite entry) (WHO 2011). The metacyclic trypomastigote attaches to a diverse array of host cells and is internalized by the host cells in a membrane-bound

compartment identified as the parasitophorous vacuole (Fernandes and Andrews 2012). The parasites then ensue to escape the vacuole and transform into intracellular amastigotes. The intracellular amastigotes go through nine cycles of binary fission and then proceed to burst out of the infected host cell (Dvorak and Hyde 1973). The amastigotes can then follow two paths: (1) differentiate into bloodstream trypomastigotes, which can then be picked up by a feeding Triatomine bug, or (2) re-invade an uninfected host cell. Bloodstream trypomastigotes that enter a Triatomine bug reach the midgut and transform into epimastigotes. The epimastigotes replicate within the midgut of the insect. Epimastigotes transform into metacyclic trypomastigotes at the distal region of the intestine, which can then complete the life cycle by entering a new mammal host (WHO 2011).

The first contact of *T. cruzi* with humans is thought to have occurred when settlers first arrived to the Americas, between 10-30 thousand years ago. The infection in humans is known as American Trypanosomiasis or Chagas disease. Chagas has a diverse array of symptoms. There are two phases of the disease. The first 4-8 weeks of infection is known as the acute phase, where the symptoms can include swelling at infection site, fever, fatigue, rash, headaches, loss of appetite, nausea, swollen glands, and enlargement of liver or spleen. Most infected patients survive the acute stage; only rarely children or individuals with weakened immune systems will die due to complications in this stage. Fatalities associated with the acute stage usually involve severe cases of myocarditis or meningoencephalitis. The more lasting phase of the infection is known as the chronic stage. About 20-30% of infected patients present complications during this stage 10-30 years after the initial infection. During the chronic stage the most common tissues affected are the heart, esophagus and colon. Most of the fatalities (~70%) are related with

cardiac damage, the other fatalities are usually associated with enlarged esophagus (megaesophagus) or colons (megacolon) (Barrett, et al. 2003; CDC 2003; WHO 2011).

T. cruzi strains are currently separated into six distinct evolutionary lineages, termed TcI-VI (Zingales, et al. 2012). However, the full genetic diversity of the parasite is still being discovered. For instance, a recent study described a new lineage (TcBAT) isolated from bats in Central America (Pinto, et al. 2012), and in this work we describe a new lineage endemic to North America (TcNA, Chapter 3). The diverse array of symptoms that are observed in Chagas disease are suspected to be associated with the parasite's genetic diversity (Di Noia, et al. 2002; Freitas, et al. 2005), underscoring the need to have a good understanding of the full genetic diversity and evolutionary history of the parasite (Chapters 2 and 3).

Molecular evolution and the study of parasitic disease

Molecular evolutionary studies can have a very direct impact in our lives. There are many examples of cases where molecular evolutionary studies have been used to determine vital features of human diseases. In bacteria, whole genome sequencing has facilitated the development of novel interventions and therapeutic targets (Wilson 2012). A molecular evolutionary approach is usually common practice when a virus outbreak occurs. When the swine-influenza outbreak occurred in 2009, evolutionary analyses were applied to resolve how long the newly found virus had been circulating among humans (Garten, et al. 2009). Phylogenetic analyses are used to predict the influenza strain that might be responsible for the next flu season, and consequently contribute to the information used for the selection of the next strain used for a vaccine (Bush, et al. 1999).

With the surge of DNA sequence data, the number of evolutionary studies has seen a dramatic increase. In the year 2012, a pub-med search with the keywords "evolution", "DNA" and "sequence" produced 8,252 publication records, while the same search in the year 2002 produced only 3,885 results. While this is a crude way of estimating research effort and output, it sheds light on the impact genome data has had on the field of molecular evolution in the past 10 years. With the availability of genome data, many questions that just a few decades ago would not have been feasible to answer are now a becoming a possibility. The genome of *T. cruzi* was sequenced in 2005, in conjunction with two other kinetoplastid genomes of human importance (*T. brucei* and *Leishmania major*) (Berriman, et al. 2005; El-Sayed, et al. 2005a; Ivens, et al. 2005). These data have opened the door to conducting evolutionary studies on these important human pathogens, either directly by using the genome data (Chapters 4 and 5) or indirectly by using the genome data to design PCR primers to amplify phylogenetic useful loci in a large range of isolates collected from the wild (Chapters 1 and 2).

Chapter 2

Analyses of 32 loci clarify phylogenetic relationships among *Trypanosoma cruzi* lineages and support a single hybridization prior to human contact

This chapter is published as:

Flores-Lopez CA, Machado CA (2011) Analyses of 32 loci clarify phylogenetic relationships among *Trypanosoma cruzi* lineages and support a single hybridization prior to human contact. PLoS Negl Trop Dis 5: e1272.

ABSTRACT

Trypanosoma cruzi is the protozoan parasite that causes Chagas disease, a major health problem in Latin America. The genetic diversity of this parasite has been traditionally divided in two major groups: *T. cruzi I* and *II*, which can be further divided in six major genetic subdivisions (subgroups TcI-TcVI). *T. cruzi I* and *II* seem to differ in important biological characteristics, and are thought to represent a natural division relevant for epidemiological studies and development of prophylaxis. Having a correct reconstruction of the evolutionary history of *T. cruzi* is essential for understanding the potential connection between the genetic and phenotypic variability of *T. cruzi* with the different manifestations of Chagas disease. Here we present results from a comprehensive phylogenetic analysis of *T. cruzi* using more than 26 Kb of aligned sequence data. We show strong evidence that *T. cruzi* II (TcII-VI) is not a natural evolutionary group but a paraphyletic lineage and that all major lineages of *T. cruzi* evolved recently (<3 million years ago (mya)). Furthermore, the sequence data is consistent with one major hybridization event having occurred in this species recently (< 1 mya) but well before *T. cruzi* entered in contact with humans in South America.

INTRODUCTION

Trypanosoma cruzi is the etiological agent of American Trypanosomiasis, also known as Chagas disease. Recent estimates suggest that about 7.7 million people in Latin America are infected with this parasite, and 10 thousand people die every year of the disease (WHO 2011). In nature, the parasite has two different cycles: a sylvatic cycle in which *T. cruzi* cycles between triatomines and wild mammalian reservoirs (e.g. opossums, raccoons, armadillos), and a domestic cycle in which *T. cruzi* infects humans through domiciliated triatomines (Barrett, et al. 2003; Miles, et al. 2003).

Since the 1980's the genetic variability and population structure of *T. cruzi* have been extensively characterized with a wide array of genetic markers (Brisse, et al. 2000; Brisse, et al. 2000b; de Freitas, et al. 2006; Llewellyn, et al. 2009b; Machado and Ayala 2001; Miles, et al. 1978; Souto and Zingales 1993; Telleria, et al. 2010; Tibayrenc 2010; Tibayrenc and Ayala 1988; Tibayrenc, et al 1986). Three main conclusions have been drawn from these studies: 1) *T. cruzi* has a mainly clonal mode of reproduction (Llewellyn, et al. 2009; Tibayrenc and Ayala 1988; Tibayrenc, et al. 1986), although historical and experimental evidence of sporadic genetic exchange has been uncovered (Bogliolo, et al. 1996; Brisse, et al. 1998, Brisse, et al 2000b; Brisse, et al. 2003; Carrasco, et al. 1996; Gaunt, et al. 2003; Higo, et al. 2004; Machado and Ayala 2001; Machado and Ayala 2002; Ocana-Mayorga, et al. 2010; Sturm, et al. 2003; Westenberger, et al. 2005). 2) The genetic variability of *T. cruzi* can be divided in two major groups (Nunes, et al. 1997; Souto, et al. 1996; Souto and Zingales 1993; Tibayrenc 1995; Tibayrenc, et al. 1993; Zingales, et al. 1999), originally termed *T. cruzi* I and *T. cruzi* II (Anon 1999). *T. cruzi* II was additionally divided in 5 distinct subgroups or stable discrete typing units (DTUs IIa-IIe) (Barnabe, et al. 2000; Brisse, et al. 2000). 3) DTUs IId and IIe are hybrids, the

result of recent genetic exchange between ancestors of lineages IIb and IIc (Brisse, et al. 2003; Machado and Ayala 2001). Although a new intraspecific nomenclature was recently proposed (Zingales, et al. 2009), renaming the six major *T. cruzi* DTUs (I, IIa-IIe) as TcI-TcVI, no changes in the inferred division of *T. cruzi* in the two major evolutionary groups *T. cruzi* I (DTU TcI) and *T. cruzi* II (DTUs TcII-VI)) were implied or proposed.

The two major groups of *T. cruzi* seem to differ in important biological characteristics (e.g. pathogenicity in mice, doubling time of epimastigotes in vivo, susceptibility to drugs), and thus are thought to represent a natural division relevant for epidemiological studies and development of prophylaxis (Andrade 1997; Laurent, et al. 1997; Revollo, et al. 1998). For instance, in the southern region of South America, where Chagas disease is most devastating, it has been observed that *T. cruzi* II strains (TcII-VI) are usually responsible for human infections, whereas *T. cruzi* I strains (TcI) are usually associated with the sylvatic cycle (Breniere, et al. 1998; Coura, et al. 2002; Di Noia, et al. 2002; Luquetti, et al. 1986; Yeo, et al. 2005; Zingales 1998). Further, in regions north of the Amazon basin *T. cruzi* I strains are the main cause of Chagas disease, although the most acute manifestations of the disease are seemingly less common than in the southern cone of South America where most research on the disease has been conducted (Higo, et al. 2004; Montilla, et al. 2002; Zingales 1998). Thus, the current consensus is that *T. cruzi* II strains (TcII-VI) are more pathogenic to humans than *T. cruzi* I strains (TcI), although at least one author has clearly stated that the six DTUs (TcI-VI) should be considered the only relevant units of analyses for epidemiology and clinical studies (Tibayrenc 2010).

Although the division of *T. cruzi* in two major evolutionary lineages has become deeply rooted in the literature, even leading to a recent suggestion that they correspond to two different

species (Tomazi, et al. 2009), there are strong reasons to doubt that this classification truly reflects the evolutionary history of this parasite. First, this classification is mostly based on codominant molecular markers (e.g. allozymes, microsatellites, RAPDs), which are not as phylogenetically informative as nucleotide sequences. Second, most studies that have used nucleotide sequences have not used an outgroup species in the phylogenetic reconstruction (Augusto-Pinto, et al. 2003; de Freitas, et al. 2006; Robello, et al. 2000; Westenberger, et al. 2005). That is a critical issue since the lack of outgroups does not allow for proper rooting of the tree and may lead to artificial evolutionary groupings. Further, with two exceptions (Machado and Ayala 2001; Subileau, et al. 2009), the studies that have included outgroup sequences have failed to interpret the observed phylogenies in the context of the proposed division of *T. cruzi* in two major evolutionary groups. Third, in each of the few studies where outgroup sequences have been included, the two expected major monophyletic lineages corresponding to *T. cruzi* I (TcI) and II (TcII-VI) are not observed (Brisse, et al. 2003; Broutin, et al. 2006; Kawashita, et al. 2001; Llewellyn, et al. 2009a; Llewellyn, et al. 2009b; Machado and Ayala 2001; Machado and Ayala 2002; Subileau, et al. 2009); instead, the evidence suggests that *T. cruzi* II (TcII-VI) is not a natural group since it appears to be paraphyletic.

To understand the diverse phenotypic differences among different *T. cruzi* strains and the potential connection between that variability and different manifestations of Chagas disease, it is essential to have a correct reconstruction of the evolutionary history of *T. cruzi*. A classification that represents evolutionary relationships is highly desirable because it may play an important role in strategic decisions about control and prophylaxis of Chagas disease. Here we present results from the largest sequence-based phylogenetic study of *T. cruzi* to date. We describe separate and combined phylogenetic analyses of nucleotide sequences from 31 nuclear genes and

1 mitochondrial region and provide estimates of the time of divergence of the main lineages of *T. cruzi*. We show that there is overwhelming evidence that *T. cruzi* II (TcII-VI) is not a natural evolutionary group but a paraphyletic lineage, and we provide a clear hypothesis of relationships among the six major DTUs of this parasite. Further, we estimate the time of diversification of *T. cruzi* strains and assess whether the sequence data is consistent with the two hybridization events that have been proposed for this species.

MATERIALS AND METHODS

Samples: For every locus we collected sequences from *Trypanosoma cruzi* strains representing five of the six principal subgroups or discrete typing units (DTUs) of *T. cruzi*: TcI (I), TcIV (IIa), TcII (IIb), TcIII (IIc) and TcV (IId) (Table 2-1) (Brisse, et al. 2000). Data from the sixth subgroup, TcVI (IIe), was already available as part of the *T. cruzi* genome sequence (www.genedb.org) (El-Sayed, et al. 2005a). Additional *T. cruzi* strains were sequenced in 9 of the 29 newly amplified loci (Tables 2-2 & S2-1). Sequences were also collected from two closely related bat trypanosomes, *T. cruzi marinkellei* (Strain N6) and *T. vespertilionis* (Strain 593), which were used as outgroups. All the strains used in this study have been widely characterized with a diverse array of genetic markers (Brisse, et al. 2000a; Brisse, et al. 2000b; Machado and Ayala 2001; Machado and Ayala 2002; Tibayrenc and Ayala 1988). Purified DNA samples for all strains sequenced were provided by Michel Tibayrenc and Christian Barnabé from the Centre d'Etudes sur le Polymorphisme des Microorganismes (CEPM), CNRS (Montpellier, France).

Molecular methods: New sequence data was collected for 29 nuclear loci (Table 2-2). In addition, previously published data sets from one mitochondrial region (COII-ND1) and two

nuclear genes (DHFR-TS, TR) (Machado and Ayala 2001; Machado and Ayala 2002) were also included in the analyses, for a total of 32 loci. PCR primers were designed for 28 of the nuclear loci using Primer3 (Table S2-2) (Rozen and Skaletsky 2000); primers for the intergenic region of Hsp70 were previously published (Sturm, et al. 2003). Loci were selected using the published genome sequence of the CL Brener strain of *Trypanosoma cruzi* (El-Sayed, et al. 2005a). Annotated loci were randomly selected from the genome based on two criteria: 1) lack of paralogous copies in the genome to avoid amplification of non-orthologous genes, 2) presence of conserved regions between both CL Brener haplotypes (if present) that would allow the design of conserved primers. The nuclear loci are located in 19 of the 41 predicted chromosomes of *T. cruzi* based on a recent genome assembly (Weatherly, et al. 2009) (Table 2-2). Six of the 32 loci did not have a putative homolog in *T. brucei*. Putative function information for each locus was obtained from GeneDB and by conducting a blastp search on the *T. brucei* predicted protein database in GeneDB.

Conditions for the PCR amplifications were: 35 cycles of a 30 second denaturation step at 94°C, annealing at 56-60°C for 30 seconds, and extension at 72°C for 1 minute. PCR primers were used for bidirectional sequencing on a 3730xl DNA Analyzer (Applied Biosystems). Sequences were edited using Sequencher (GeneCodes). In cases where sequences had polymorphic nucleotides (determined by the presence of multiple double peaks in the chromatogram), PCR fragments were cloned using the TA cloning kit (Invitrogen) and three to five cloned PCR fragments were sequenced to identify both haplotypes. Singleton mutations that were observed only in the sequences from cloned fragments and not in the sequences from the PCR products were not included in the final sequence of each haplotype used in the analyses. Sequences have been deposited in GenBank (Accession Numbers HQ859465- HQ859886).

Phylogenetic analyses: Sequences were manually aligned using SE-AL version 2.0 (Rambaut 2002). A Neighbor Joining (NJ) tree was reconstructed for each data set and each topology was used to estimate maximum likelihood parameters for different models of nucleotide substitution. The most appropriate nucleotide substitution model to analyze each locus was chosen using Modeltest 3.7 (Posada and Crandall 1998). Maximum likelihood (ML) trees were individually obtained for each locus using ML heuristic searches in PAUP* 4.0b10 (Swofford 1998) using the tree bisection-reconnection (TBR) branch swapping algorithm. Bootstrap support values were obtained by ML analyses of 100 pseudoreplicates of each dataset.

MrBayes 3.1.2 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003) was used to conduct Bayesian analyses using the substitution models chosen by Modeltest 3.7 (Posada and Crandall 1998). We ran two independent simultaneous Markov Chain Mote Carlo runs with four chains each for 100,000 generations and sampled trees every 10 generations. If the standard deviation of split frequencies were not below 0.01 after analyses were done, the analyses were ran for an additional 100,000 generations and were stopped after convergence (i.e. standard deviation of split frequencies ≤ 0.01). Parameters and corresponding trees were summarized after discarding the initial 25% of each chain as burnin.

Data from the 32 loci were concatenated (26,329 nucleotides per strain) to reconstruct a consensus phylogenetic tree. Nuclear loci from the hybrid strains of *T. cruzi*, TcV (IId) and TcVI (IIe), usually have two different haplotypes, one of which groups with TcII (IIb) and the other with TcIII (IIc) (Machado and Ayala 2001; Machado and Ayala 2002). To analyze the concatenated data using haplotypes from the two hybrid strains included (SO3 cl5, CL Brener), we sorted each haplotype accordingly depending on the results from the ML and Bayesian phylogenetic analyses, concatenating haplotypes that had the same phylogenetic position (i.e.

that grouped with the same "parental" clade). The concatenated alignment was analyzed using ML methods as described above. Bayesian analyses were performed in MrBayes 3.1.2 as described above, for 100 million generations with two parallel searches, with a burnin of 10% of the generations (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003).

To test the topological congruence among the gene trees, we used PAUP* 4.0b10 (Swofford 1998) to perform the incongruence length difference test (ILD) among all data sets (Farris, et al. 1994). In addition, the Shimodaira-Hasegawa congruency test (Shimodaira and Hasegawa 1999) was performed on each dataset as well as in the concatenated dataset in order to compare the likelihood of the phylogeny obtained by ML and the likelihood of the tree when *T. cruzi* I and II (TcI and TcII-VI) are enforced to be monophyletic (see Topology H, Figure 2-1). This was done in order to assess the support of the current division of *T. cruzi* in two major phylogenetic groups.

Tests of selection: Non-neutral evolutionary patterns can affect inferences of phylogenetic relationships (e.g. (Castoe, et al. 2009)). Therefore each locus was examined for evidence of positive selection acting across the complete sequence and among codon sites using the codeml application from the PAML package (Yang 2007). Pairs of nested models were compared using a likelihood ratio test (LRT) under the assumption that the LRT statistic follows a chi-square distribution with the number of degrees of freedom dependent on the estimated number of parameters differentiating the nested models. We compared three pairs of nested site models: 1) M1 (neutral) versus M2 (selection); 2) M7 (beta) versus M8 (beta & ω); 3) M8 versus M8a (beta & $\omega = 1$) (Swanson, et al. 2003; Yang and Swanson 2002). Significance of the LRT of M1 vs M2 and M7 vs M8 was determined using 2 degrees of freedom. Since M8a is not fully nested on M8, a strict LRT for these two models is not possible. However, it has been suggested that

significance of the LRT can be determined by halving the p value from a chi-square test with 1 degree of freedom (Yang 2007).

Divergence time estimates: Likelihood Ratio Tests (LRT) were performed to evaluate the null hypothesis that each locus of the concatenated dataset evolved under a molecular clock (Felsenstein 1988). The molecular clock was rejected in only 3 genes (DHFR-TS, Tc00.1047053504059.20, Tc00.1047053509561.20) (Table S2-4). The remaining 22 loci in which the molecular clock was not rejected, and that had a homolog in *T. brucei*, were concatenated for these analyses. Divergence dates were estimated using Bayesian analysis in BEAST v1.5.3 (Drummond and Rambaut 2007). Both the strict and relaxed Lognormal clock models were used to estimate divergence times on the mitochondrial and the concatenated nuclear loci data sets. Analyses were run separately for nuclear and mitochondrial sequences since previous analyses gave very different estimates for each type of data (Machado and Ayala 2001). All analyses were conducted without any topological constraints using the HKY substitution model with the gamma plus invariant sites as the site heterogeneity model, with 4 gamma categories, as well as partitioning of codons into 3 positions. All priors were set to default values, except for the divergence estimate between *T. cruzi* and *T. brucei*, which was set to 100 million years ago (mya) under a normal distribution with 10 mya as the standard deviation. This date (100 mya) is a conservative estimate of the time to the last common ancestor of *T. cruzi* and *T. brucei* using the time of separation of Africa and South America (Hay, et al. 1999). Times of divergence were obtained by converging 10 independent Markov Chain Monte Carlo (MCMC) runs in Tracer v1.5 (Drummond and Rambaut 2007) in order to ensure convergence between the runs. Burnin of 20% of the samples was used. Each run had a chain length of 10 million, with sampling every 1000 chains. Although the mitochondrial data had been previously

analyzed using a simpler method (Machado and Ayala 2001), we decided to reanalyze them with the Bayesian framework described above to compare previous estimates with the new Bayesian estimates.

The Relaxed Lognormal Clock model allows assessing how clock-like the data are (i.e. whether there is large rate heterogeneity among lineages), by using the estimate of the ucld.stedv parameter. A value of 0 means that the data is reasonably clock-like, whereas a value much greater than 1 indicates that the data has considerable rate heterogeneity among lineages (Drummond, et al. 2007). The nuclear data set had a ucld.stdev of 0.392, while the mitochondrial data set had a higher ucld.stdev value (0.701), indicating higher rate heterogeneity among lineages. However, the Relaxed Lognormal Clock model for the mitochondrial data set did not converge even after combining 10 independent runs in Tracer. Therefore the estimates of the mitochondrial data with this model were not reliable and are not presented.

In addition, we analyzed *T. cruzi* genome sequence data (El-Sayed, et al. 2005a) to obtain synonymous substitution (Ks) values for all annotated genes that had a single copy of each Esmeraldo-like (TcII (IIb)) and non-Esmeraldo-like (TcIII (IIc)) ortholog in the genome sequence. Our phylogenetic analyses (see below) show that nucleotide distances between Esmeraldo-like and non-Esmeraldo-like alleles from the heterozygous genome strain represent maximum distances within *T. cruzi*. Thus, those distances can be used to estimate the time to the most recent common ancestor of the major extant lineages of the parasite. A list of 4,568 Esmeraldo-like and non-Esmeraldo-like orthologs was obtained from Table S2-1 of El-Sayed et al (El-Sayed, et al. 2005b) and sequences were downloaded from TriTrypDB (tritrypdb.org). The orthologous sequences were pairwise-aligned using ClustalW (Thompson, et al. 1994) and the resulting alignments were passed to PAML for estimation of Ks using the codeml program with

the pairwise distance estimation option (runmode = -2) (Yang 2007). The average Ks value (0.0404) was used to estimate the time back to the most recent common ancestor of extant *T. cruzi* lineages using an estimate of the mutation rate for *T. brucei* (Lynch 2010; Valdes, et al. 1996) (see Discussion).

RESULTS

Phylogenetic analyses: The predominant clonal mode of propagation of *T.cruzi* and lack of evidence of intragenic recombination in the data (not shown) allow the use of nuclear gene sequences for reconstructing intraspecific phylogenies. The 31 nuclear loci we analyzed are randomly distributed in the genome. They are located in 19 of the 41 predicted chromosomes of *T. cruzi*, and when located on the same chromosome the loci are at least 30 Kb apart (in most cases >100 Kb apart) (Table 2-2). The ML and Bayesian phylogenetic analyses of each one of the 32 individual loci (Figure S1) produced seven different topologies (Figure 2-1). The ILD partition test confirmed that at least one of these trees was significantly different from the others (p=0.01). All 32 loci confirm the paraphyletic nature of *T. cruzi* II. Analyses of 24 of the 32 loci produced individual phylogenetic trees with the same topology (topology A), including the three genes that we previously analyzed (Machado and Ayala 2001; Machado and Ayala 2002) (Table 2-2). Sequences from *T. cruzi* II strains were never monophyletic in any of the genes surveyed (represented by Topology H). Topology A is consistent with a history of divergence in which *T. cruzi* II strains are paraphyletic.

To test the validity of the division of *T. cruzi* in two major groups, we performed the Shimodaira-Hasegawa test on each gene tree (Shimodaira and Hasegawa 1999). The test was

conducted to determine if a constrained topology representing the division of *T. cruzi* into two different reciprocally monophyletic lineages, *T. cruzi* I (TcI) and *T. cruzi* II (TcII-VI), was as good an explanation of the data as the ML trees obtained for each gene. For every gene the constrained topologies in which *T. cruzi* I and *T. cruzi* II were reciprocally monophyletic were significantly worse than the ML phylogenies (Table S2-3), rejecting the prevalent idea that *T. cruzi* is divided in the two major evolutionary lineages *T. cruzi* I (TcI) and *T. cruzi* II (TcII-VI).

ML and Bayesian phylogenetic trees reconstructed with the concatenated multilocus dataset (Figure 2-2) were also congruent with the ubiquitous topology A found on the majority of analyses of individual loci (Figure 2-1, Figure S2-1). All internal nodes in this topology are strongly supported either by ML or Bayesian analyses (Figure 2-2). Moreover, a constrained phylogeny consistent with the current division of *T. cruzi* in two major reciprocally monophyletic groups is significantly worse than the best ML tree from the multilocus concatenated dataset ($p < 0.0001$). This result provides further evidence that the current division of *T. cruzi* in two major evolutionary lineages (Anon 1999) is a classification that does not reflect evolutionary relationships among strains of *T. cruzi*.

The basic relationships suggested by our analyses show that there are two major clades in the phylogeny of *T. cruzi.* The first clade, which harbors the most genetic diversity, includes DTUs TcI (I), TcIV (IIa), TcIII (IIc), and one haplotype from each of the two hybrid DTUs TcV (IId) and TcVI (IIe). The second lineage includes DTU TcII (IIb) and the other haplotype from each of the two hybrid DTUs TcV and TcVI. In 26 of the 32 nuclear loci analyzed we observed divergent allele sequences in members of both hybrid DTUs (TcV, TcVI) (Figure 2-1: Topologies A,C), in 4 loci both hybrid DTUs were homozygous or had barely divergent alleles (Figure 2-1: Topologies B,D,G), and in 2 loci one of the hybrid DTUs was homozygous while

the other still had divergent alleles (Figure 2-1: Topologies E,F). Consistent with previous analyses (Brisse, et al. 2003; Machado and Ayala 2001), we only observe evidence of one major hybridization event during the history of *T. cruzi*: between the ancestors of DTUs TcII and TcIII to generate DTUs TcV and TcVI (see Discussion).

Selection tests: Only 8 of the 32 genes show evidence that some of their nucleotide sites have been under positive selection (Table 2-3). However, of these eight genes only four were highly significant in all three tests (M1 vs M2, M7 vs M8, M8 vs M8a). Three of the genes were only significant at the 5% level, but not at the 1% level, and only significant when M8a was compared to M8. The reconstructed phylogeny from 2 of the 8 genes that showed evidence of selection was different from the main topology A (Tc00.1047053506529.310: Topology C; Tc00.1047053510765.50: Topology C), but in none of those two cases sequences from all *T. cruzi* II strains were monophyletic. The other six genes that showed evidence of positive selection produced topology A. These results show that the loci used in this study are mostly evolving neutrally (24 out of 32 loci) and that phylogenetic analyses from 75% of the neutrally evolving loci (16 of 24) rendered the most common topology A (Figures 2-1 and 2-2), suggesting that results from the phylogenetic analyses have not been biased by loci that have been under positive selection.

Estimates of divergence time: The molecular clock was rejected on the concatenated dataset ($p<0.001$). Therefore, each individual locus was tested for the molecular clock and loci for which a homolog could be confidently identified in *T. brucei* and for which the Likelihood Ratio Test could not reject the Molecular clock (21 loci, Table S2-4) were chosen to become part of a concatenated dataset suitable to run the Bayesian divergence time analyses. The divergence estimates from the mitochondrial dataset differ significantly from the nuclear loci estimates

(Table 2-4). The estimated time to the most recent common ancestor (tMRCA) using mitochondrial data suggest that *T. cruzi*'s major lineages diverged during the Miocene (tMRCA = 11.0 (7.0-15.2) mya), estimates that are similar to those presented by Machado and Ayala (Machado and Ayala 2001) using less sophisticated methods. On the other hand, the dates estimated with the concatenated data from 20 nuclear loci point towards a Pleistocene origin of *T. cruzi* (tMRCA = 1.36 (1.0-1.7) mya (strict); tMRCA = 2.18 (0.9-3.7) mya (relaxed)) (Table 2-4, Figures 2-3 & S2-2). Those dates are more recent than previously estimated divergence times using a single locus (TR: tMRCA = 3.91 mya) (Machado and Ayala 2001). We also obtained very similar divergence estimates from the concatenated data set of all nuclear loci that had a homolog in *T. brucei* (24 loci, Table S2-4) including genes that rejected the molecular clock hypothesis (not shown).

The discrepancy between the dates estimated with the mitochondrial and nuclear loci is likely the result of saturation of substitutions between the mitochondrial sequences of *T. cruzi* and the *T. brucei* outgroup used for the time calibration. Within *T. cruzi* the largest distance at silent sites (Ks) in the mitochondrial genes used is at least 6 times larger than that of any nuclear gene (Table 2-3), but most importantly substitutions at silent sites between *T. cruzi* and *T. brucei* are overly saturated (Ks = 77.32). This observation is not surprising given the large divergence time between the two species, but leads to an overestimation of divergence times in more recently diverged lineages. For that reason we will not discuss the mitochondrial estimates any further.

The data allowed estimating the age of the major hybridization event in the history of *T. cruzi*: the generation of DTU's TcV and TcVI (IId and IIe) by hybridization of DTUs TcII and TcIII (IIc and IIb). The time of this event was estimated using the observed divergences between

alleles from the putative parental and hybrid lineages (i.e. TcII vs TcV-TcVI and TcIII vs TcV-TcVI). This hybridization event occurred <1 mya, well before *T. cruzi* entered in contact with humans in South America, and the two independent estimates of the event are remarkably similar although the estimates from the strict clock model (tMRCA = 0.49 (0.3-0.6) mya, 0.49 (0.3-0.6) mya) are more recent than the estimates from the relaxed lognormal clock model (tMRCA = 0.8 (0.3-1.4) mya, 0.73 (0.3-1.3) mya) (Table 2-4, Figures 2-3 & S2-2).

DISCUSSION

The evolutionary history of *Trypanosoma cruzi*: From the early 1990's *T. cruzi* was divided in two major groups, *T. cruzi* I and *T. cruzi* II (Anon 1999; Nunes, et al. 1997; Souto, et al. 1996; Souto and Zingales 1993; Tibayrenc, et al. 1993; Zingales, et al. 1999). One of the groups, *T. cruzi* II, was further divided into 5 stable Discrete Typing Units (DTUs TcI-TcVI) based on additional genetic data (Barnabe, et al. 2000; Brisse, et al. 2000a; Zingales, et al. 2009). Our study aims to clarify the phylogenetic relationships among the currently defined six major DTUs and represents a comprehensive molecular phylogenetic analysis of the largest nucleotide sequence dataset collected for this parasite (26,329 nucleotides per strain). Although we focused the sequencing on the seven strains listed in Table 2-1, for 10 of the 32 loci we obtained sequences from 20-48 strains (Tables 2-2 and S2-1). Results from the more deeply sampled loci are consistent with the overall results, and in particular there is no evidence of additional recombination/hybridization events (see below). The predominantly clonal population structure of *T. cruzi* (Llewellyn, et al. 2009; Tibayrenc and Ayala 1988; Tibayrenc, et al. 1986) justifies sampling a limited number of strains representing the six major lineages of this parasite. The strains that constitute the core of the data presented here are widely studied standard laboratory

strains which have been consistently used to make inferences about genetic and biological variability in *T. cruzi*. There is no indication that those strains represent outliers within *T. cruzi* and as such they are useful for making inferences about major evolutionary events in this parasite.

The concatenated phylogeny (Figure 2-2) is well supported and its topology is consistent with results from previous analyses of smaller sequence datasets that used outgroup sequences (Machado and Ayala 2001; Subileau, et al. 2009). Furthermore, it corresponds to the most commonly reconstructed topology using single loci (Topology A, Figure 2-1). This phylogeny shows that *T. cruzi* is divided in two clearly defined clades that do not correspond to the two originally defined major lineages *T. cruzi* I and *T. cruzi* II. Results from Shimodaira-Hasegawa tests applied to every locus (Table S2-3) provide strong evidence that the previously defined lineage *T. cruzi* II is paraphyletic and therefore does not represent a natural evolutionary lineage. One of the clades of the concatenated phylogeny includes TcI, TcIII, TcIV and one of the haplotypes from each of the two hybrid lineages TcV and TcVI. The other clade includes TcII and the alternative haplotypes from hybrid lineages TcV and TcVI. The phylogenetic placement of DTU TcIV (IIa) is less well resolved than the position of the other lineages. Although the bootstrap support of the branch separating TcIV from the TcI-TcIII-TcV-TcVI clade is 72% in the concatenated tree, the phylogenetic position of TcIV is quite variable in the individual trees (Figure S2-1). In 11 of the 24 trees consistent with Topology A (Figure 2-1) the placement of TcIV is the same as in the concatenated phylogeny and is supported with bootstrap values >55% (>80% in 5 trees). It is likely that the most sensible approach to attain full resolution of the phylogenetic position of TcIV is to increase the number of loci sampled. The availability of

genome sequences of additional *T. cruzi* strains (e.g. (Franzen, et al. 2011)) should help resolve this issue.

Our results show that the classification of *T. cruzi* in two major evolutionary lineages (Anon 1999), which has become deeply rooted in the literature, does not reflect the evolutionary history of this species. This classification arose from analyses of codominant molecular markers (e.g. allozymes, microsatellites, RAPDs) and PCR fragment sizes of different regions of rRNA genes and a mini-exon (Nunes, et al. 1997; Souto, et al. 1996; Souto and Zingales 1993; Tibayrenc 1995; Tibayrenc, et al. 1993; Zingales, et al. 1999), and appeared to be consistent with results from phylogenetic analyses of small nucleotide sequence datasets (Augusto-Pinto, et al. 2003; de Freitas, et al. 2006; Robello, et al. 2000; Westenberger, et al. 2005). However, none of those analyses included data from outgroups, a critical issue since lack of data from outgroup taxa does not allow for proper rooting of phylogenies and can generate artificial evolutionary groupings. Data from outgroups allow differentiating between derived (apomorphic) and ancestral (plesiomorphic) characters, which is fundamental for conducting proper phylogenetic analyses (Maddison 1984).

In our locus by locus analyses using outgroup data we never obtained topology H (Figure 2-1, Table 2-2), which corresponds to the phylogeny in which all *T. cruzi* II strains (TcII-VI) are monophyletic as suggested by the two group classification of *T. cruzi*. However, when we conducted the same analyses for every locus removing the outgroup sequences and rooting the tree at the longest internal branch (midpoint rooting), topology H was reconstructed 4 times (Table 2-3). Furthermore, in those analyses without outgroup we observed a different tree reconstructed in 15 of the 32 genes analyzed (Table 2-3). Those results suggest that the lack of

outgroups in previous phylogenetic analyses of *T. cruzi* could be partially responsible for the original partition of the genetic diversity of this species in two major lineages.

The observation of distinct PCR fragment sizes in different regions of rRNA genes or mini-exon sequences (Nunes, et al. 1997; Souto, et al. 1996; Souto and Zingales 1993; Zingales, et al. 1999) was instrumental for the original division of *T. cruzi* in two major groups. Our phylogenetic results show that those studies simply uncovered derived character states in *T. cruzi* I (TcI) strains for the molecular traits studied, but the uncovered similarities in traits across strains do not correspond to actual evolutionary relationships among the strains. Presence-absence morphological or molecular characters can be useful for finding similarities among organisms but their utility for inferring evolutionary relationships is limited when the number of characters is very small and there is no additional supporting information. Without the context of a supported phylogeny it is not possible to determine if the observed character similarity truly reflects shared ancestry or homoplasy, as evidenced by the spurious relationships first described for *T. cruzi*.

The age of *Trypanosoma cruzi*: Our calculations point towards a Pleistocene origin of the extant lineages of *T. cruzi* (tMRCA = 1.36 (1.0-1.7) mya (strict); tMRCA = 2.18 (0.9-3.7) mya (relaxed)) (Table 2-4, Figures 2-3 & S2-2). Furthermore, the major hybridization event that led to the origin of DTU's TcV and TcVI (IId and IIe) by hybridization of DTUs TcII and TcIII (IIc and IIb) occurred <1 mya, well before *T. cruzi* entered in contact with humans in South America. Estimated divergence times are dependent on the available calibration point(s), which in this study was the estimated separation time of Africa and South America (~100 mya) based on geological evidence (Hay, et al. 1999). That date is thought to be the last time *T. cruzi* and *T. brucei* shared a common ancestor (Lake, et al. 1988; Stevens, et al. 1999). Older divergence

estimates for all the clades in the phylogeny can be obtained if older separation dates of Africa and South America are considered. However, obtaining estimates of *T. cruzi* divergence time as old as those suggested in other studies (e.g. 37-88 mya) (Briones, et al. 1999; Gaunt and Miles 2000) requires using unrealistic calibration dates.

Even if there are uncertainties about the calibration point, the estimated recent divergence of *T. cruzi* is consistent with the small nucleotide divergences observed among the different lineages (Table 2-3) and leads to reasonable estimates of substitution rates in *T. cruzi*. The estimated silent site substitution rates per year ($8.4\text{-}5.2 \times 10^{-9}$) based on the average silent site divergence in *T. cruzi* (Ks = 0.0228) and the estimated divergence times using nuclear loci (Table 2-4) fall within the range of silent site substitution rates estimated for other organisms (Li 1997; Lynch 2010). Further, independent estimates of the age of divergence of *T. cruzi* can be obtained using estimates of the nucleotide substitution rate per million year (my) and the observed average divergence at silent sites (Nei 1987). Using the estimated mutation rate in *T. brucei* (1.65×10^{-9} per generation) (Lynch 2010) and its generation time (7-10 generations/year) (Koffi, et al. 2009), we obtain an estimate of the neutral mutation rate of 0.0115-0.0165 per my. Using that substitution rate and the observed average silent site divergence for 4569 single copy heterozygous genes from the *T. cruzi* genome (Ks = 0.0404), the tMRCA of *T. cruzi* is estimated to be 1.73-1.21 mya, consistent with the phylogeny-based estimates obtained using BEAST (Table 2-4).

The recent divergence dates are also consistent with the idea that the diversification of *T. cruzi* was linked to the origin of its blood-sucking triatomine vectors, which occurred in the last 5 my (Gorla, et al. 1997; Schofield 2000). Molecular clock calibrations using *cytochrome b* sequences suggest a Pleistocene origin of *Rhodnius prolixus* and *R. robustus* (Monteiro, et al.

2003), and the observation of almost identical transposable elements in *R. prolixus* and opossums and squirrel monkeys suggest a very recent association of vector and hosts (Gilbert, et al. 2010).

The evidence for hybridization events during *T. cruzi* divergence*:* Previous studies have established that hybridization events have played an important role during the diversification of this parasite (Brisse, et al. 2003; de Freitas, et al. 2006; Machado and Ayala 2001; Sturm and Campbell 2010; Westenberger, et al. 2005). Two different scenarios involving hybridization events have been proposed to explain the current genetic structure of *T. cruzi*. The first scenario proposes that a recent single hybridization event took place between the ancestors of DTU's TcII (IIb) and TcIII (IIc), which generated hybrid DTUs TcV (IId) and TcVI (IIe) (de Freitas, et al. 2006; Machado and Ayala 2001). The second scenario proposes that in addition to the recent hybridization event responsible for hybrid DTUs TcV and TcVI, there was an ancestral hybridization event between the ancestors of DTUs TcI (I) and TcII that gave rise to the ancestors of DTUs TcIV (IIa) and TcIII (Sturm and Campbell 2010; Westenberger, et al. 2005).

Our results provide additional evidence supporting the single recent hybridization event leading to the evolution of hybrid DTUs TcV (IId) and TcVI (IIe) (Brisse, et al. 2003; de Freitas, et al. 2006; Machado and Ayala 2001). The main evidence is the presence of multiple heterozygous loci with divergent alleles, where the alleles have close genetic distances to alleles from the putative parental lineages TcII (IIb) and TcIII (IIc). This pattern was first observed in several nuclear genes (Brisse, et al. 2003; Machado and Ayala 2001) and later observed across thousands of genes in the genome sequence of *T. cruzi* strain CL Brener (TcVI) (El-Sayed, et al. 2005a). In this study we observed this pattern in 26 out of the 32 nuclear loci analyzed (Figure 2-1, Topologies A and C). More importantly, we did not observe any additional putative hybridization events that could be identified from loci with multiple polymorphic nucleotide sites.

Our estimates of the age of the hybridization event suggest that this hybridization occurred less than 1 mya (Table 2-4, Figures 2-3 & S2-2), consistent with the observation that the alleles from the hybrid lineages have few nucleotide differences with the alleles from the putative parental lineages.

The ancestral hybridization event previously proposed (Sturm and Campbell 2010; Westenberger, et al. 2005) requires the heterozygosity from the ancestral hybrid lineage to be lost through genome-wide homogenization by homologous recombination or gene conversion, given that the extant DTUs TcIV (IIa) and TcIII (IIc) show widespread homozygosity. This scenario suggests that the homogenization process should have left clear signals of the ancestral hybridization in patterns of SNP variation, which should show mixed signals of phylogenetic affinity to either one of the parental lineages. Unfortunately, two missing key factors in the original phylogenetic analyses conducted to support the ancestral hybridization event (Westenberger, et al. 2005) have likely contributed to misinterpreting the data. The first and most important factor is the lack of outgroup sequences in the phylogenetic analyses. Our study shows that failure to include outgroup sequences can alter phylogenetic reconstruction in *T. cruzi* (Table 2-3). The second factor is the lack of bootstrap support values on key nodes of the trees that support the ancestral hybridization scenario.

We question the evidence for the ancestral hybridization scenario on three grounds. First, the origin of DTUs TcIII (IIc) and TcIV (IIa) is fairly recent, only about twice as old as the recent hybridization event leading to the origin of hybrid DTUs TcV (IId) and TcVI (IIe) (Table 2-4, Figures 2-3 & S2-2). It is therefore difficult to explain why there is still so much widespread allelic heterozygosity left in the hybrid DTUs TcV and TcVI, while there is (potentially) none left in DTUs TcIII and TcIV. For instance, the sequence of the genome strain CL Brener (TcVI)

contains over 30 Mb of combined contig size in non-repetitive heterozygous regions and only 2 Mb in homozygous regions (see Table S2-2 from (El-Sayed, et al. 2005a)). Given that pattern, it is clear that the proposed homogenization process that led to widespread loss of heterozygosity in the ancestor of DTUs TcIII and TcIV needs to be very different (at least in speed) than the process currently occurring in the recent hybrid strains. Second, the suggestion that DTUs TcIII and TcIV show mosaic sequences with SNPs that match DTUs TcI (I) or TcII (IIb) (Sturm and Campbell 2010; Westenberger, et al. 2005) is hard to reconcile with patterns observed in our data, in data from a recent study (Subileau, et al. 2009), and in the sequenced strain of *T. cruzi*. To our knowledge there are no examples of obvious mosaic sequences in CL Brener, and, more importantly, the presence of interspersed SNPs matching either of the putative parental lines in small sequenced regions (~1-2Kb) will require fairly high rates of recombination which are not consistent with what is observed in the genome strain or in sequences from the hybrid strains. Third, a prediction of the ancestral hybridization scenario is that one should observe mixed phylogenetic signals across different loci (Westenberger, et al. 2005): in some loci, alleles from DTUs TcIV and TcIII will show strong phylogenetic affinities with alleles from DTU TcI, and in other loci with alleles from DTU TcII; other loci would show little phylogenetic resolution if they are mosaics from both ancestral parental lineages. Here, we have shown that there is overwhelming support (i.e. strong phylogenetic signal) linking alleles from DTUs TcIII and TcIV with alleles from DTU TcI (Figure 2-1, topology A), and in no case did we observe strong support for a link of DTUs TcIII and TcIV with alleles from DTU TcII (Figure 2-1, topology H; Table S2-3). To explain this pattern under the ancestral hybridization scenario one would also need to propose an additional mechanism whereby during homogenization there was gene conversion biased towards the allele from DTU TcI. Interestingly, the genome sequence of *T.*

cruzi shows an excess of TcII-like homozygous regions relative to TcI-like homozygous regions (see Table S2-2 from (El-Sayed, et al. 2005a)), contrary to the biased gene conversion towards TcI alleles required to explain our data under the ancestral hybridization scenario.

As the most appropriate explanation should be the most parsimonious, we suggest that the scenario requiring a single hybridization event leading to the generation of the extant hybrids DTUs TcV (IId) and TcVI (IIe) is the only one that is currently strongly supported by data. The analysis of complete genome sequences from multiple lineages of *T. cruzi* should provide a definitive test of the ancestral hybridization scenario, but it is telling that analyses of the large number of randomly selected loci presented here are not consistent with predictions from that hypothesis.

CONCLUSION

We have reconstructed the evolutionary history of the major lineages of the human parasite *Trypanosoma cruzi* using nucleotide sequences from one mitochondrial region and 31 unlinked nuclear loci. Our results show that the original classification of *T. cruzi* in two major groups, *T. cruzi I* (TcI) and *T. cruzi II* (TcII-VI), does not reflect the evolutionary history of the parasite, that its diversification into the current extant lineages was recent (<1-3 mya), and that there is only strong evidence for one major hybridization event that occurred <1 mya, well before *T. cruzi* entered in contact with humans in South America. It is possible that by sampling a small number of strains one could miss detecting rare recombination or hybridization events (although we did not see this in loci that were more deeply sampled). Thus, future multilocus phylogenetic studies should also attempt conducting more in-depth sampling of strains. Based on our results

we suggest that it is important to reconsider conclusions from previous studies that have attempted to uncover important biological differences between the two originally defined major lineages of *T. cruzi*. Conclusions from studies that report results of analyses from one or few strains that do not encompass all the genetic variability of the artificial group "*T. cruzi* II" should be carefully dissected to determine if the findings do in fact reflect fundamental biological differences between the natural group "T. cruzi I" and the artificial group "T. cruzi II" or simply reflect differences among the specific DTUs studied. A thorough review of the literature suggests that many of the studies that report differences, or lack thereof, between the two originally defined lineages of this parasite are typically based on observations from very few strains (Flores-López and Machado, in prep.). Future work should focus on trying to determine if, as previously suggested (Tibayrenc 2010), the currently defined six major lineages of this parasite (TcI-TcVI), for which we now have well supported evolutionary relationships, do indeed represent independent relevant groups for epidemiological studies and development of prophylaxis.

Tables

Table 2-1. The main *Trypanosoma cruzi* strains used in this study. [a] Discrete typing unit (DTU) (Brisse, et al. 2000); [b](Miles, et al. 1978); [c](Tibayrenc and Breniere 1988); [d] (1999); [e](Zingales, et al. 2009).

Strain	DTU [a]	Zymodeme [b]	Isoenzyme type [c]	1999 nomenclature [d]	New nomenclature [e]
SO34 cl4	I	Z1	20	Tc I	TcI
SC13	I	Z1	?	Tc I	TcI
EP 255	IIa	Z3	nd(27)	Tc II	TcIV
CBB cl3	IIb	Z2	32	Tc II	TcII
M6241 cl6	IIc	Z2	35	Tc II	TcIII
SO3 cl5	IId	Z2	39	Tc II	TcV
CL Brener (CL F11F5)	IIe	Z2	43	Tc II	TcVI

Table 2-2. List of loci included in this study.

[a] N: number of haplotypes sequenced. [b] bp: sequenced region (in base pairs). [c] (Weatherly, et al. 2009). [d] See Figure 2-1.

Locus ID	N [a]	bp [b]	Chromosome [c]	Gene location in Chr. (strand), gene length [c]	Predicted function	Topology [d]
COII-ND1	48	1226	Maxicircle (mtDNA)	N/A	Cytochrome oxidase subunit II-NADH dehydrogenase subunit 1	A
Tc00.1047053503555.30	11	1290	Chr 37	713055 - 714533 (-), 1479 bp	Trypanothione reductase (TR)	A
Tc00.1047053509153.90	40	1473	Chr 27	718463 - 720028 (+), 1566 bp	Dihydrofolate reductase-thymidylate synthase (DHFR-TS)	A
HSP70	11	508	Chr 32	699686 - 700540 (-), 855 bp	Intergenic region	A
Tc00.1047053503885.80	12	946	Chr 26	163788 - 164993 (+), 1206 bp	Hypothetical protein, conserved	A
Tc00.1047053503891.50	10	813	Chr 20	75320 - 76489 (-), 1170 bp	Hypothetical protein, conserved	A
Tc00.1047053503909.76	12	614		556434 - 557156 (+), 723 bp	Ferric reductase transmembrane protein, putative	B
Tc00.1047053504013.40	25	805	Chr 34	465693 - 466718 (-), 1026 bp	Serine acetyltransferase, putative	A
Tc00.1047053504045.100	10	886	Chr 40	1854961 - 1856415 (-), 1455 bp	Hypothetical protein, conserved	A
Tc00.1047053504057.80	10	858	Chr 34	417310 - 418677 (-), 1368 bp	Hypothetical protein, conserved	D
Tc00.1047053504059.20	13	896	Chr 14	465730 - 467526 (-), 1797 bp	Endomembrane protein, putative	A
Tc00.1047053506247.200	10	920	Chr 37	133811 - 136708 (+), 2898 bp	Beta-adaptin, putative	A

Table 2-2 continued

Tc00.1047053506525.150	13	821	Chr 40	593462 - 594415 (+), 954 bp	Hypothetical protein, conserved	A
Tc00.1047053506529.310	27	727	Chr 6	97318 - 98676 (-), 1359 bp	Hypothetical protein, conserved	C
Tc00.1047053506739.20	11	810	Chr 3	25655 - 27589 (-), 1935 bp	Hypothetical protein, conserved	F
Tc00.1047053507801.70	11	677	Chr 23	535126 - 535959 (+), 834 bp	Protein kinase, putative	A
Tc00.1047053508153.540	13	774	Chr 36	699363 - 700391 (+), 1029 bp	Hypothetical protein, conserved	A
Tc00.1047053508461.80	20	838	Chr 39	1187987 - 1189126 (-), 1140 bp	Prostaglandin F2alpha synthase	G
Tc00.1047053508719.70	24	709	Chr 37	375185 - 376402 (+), 1218 bp	Hypothetical protein, conserved	A
Tc00.1047053509007.30	13	815	Chr 31	573767 - 574690 (+), 924 bp	Hypothetical protein, conserved	E
Tc00.1047053509105.70	24	897	Chr 37	769449 - 770786 (-), 1338 bp	Thiol-dependent reductase 1, putative	A
Tc00.1047053509561.20	23	880	Chr 12	285842 - 287581 (-), 1740 bp	Flagellum-adhesion glycoprotein, putative	A
Tc00.1047053509967.50	11	595	Chr 10	184622 - 185329 (+), 708 bp	Hypothetical protein, conserved	A
Tc00.1047053510101.480	12	829	Chr 27	190063 - 191427 (-), 1365 bp	Hypothetical protein, conserved	B
Tc00.1047053510123.24	12	880	Chr 20	372476 - 373429 (+), 954 bp	Hypothetical protein, conserved	A
Tc00.1047053510131.90	12	936	Chr 30	340360 - 342003 (+), 1644 bp	Hypothetical protein, conserved	A
Tc00.1047053510765.50	13	817	Chr 39	1780396 - 1781763 (+), 1368 bp	Hypothetical protein, conserved	C
Tc00.1047053510877.190	8	453	Chr 34	493531 - 494328 (-), 798 bp	Hypothetical protein, conserved	A

Table 2-2 continued

Tc00.1047053510889.210	25	693	Chr 6	154383 - 156290 (-), 1908 bp	Hypothetical protein, conserved	A
Tc00.1047053510889.310	23	763	Chr 6	193929 - 196025 (+), 2097 bp	Hypothetical protein, conserved	A
Tc00.1047053511153.124	12	513	Chr 27	412720 - 413271 (+), 552 bp	Hypothetical protein, conserved	A
Tc00.1047053511529.200	13	667	Chr 35	170438 - 171232 (-), 795 bp	Hypothetical protein, conserved	A

Table 2-3. Results from the selection, distance and midpoint rooting analyses. N: number of strains (haplotypes) sequenced. bp: number of aligned nucleotides used in the PAML analyses. [a] The percent of sites with ω = dN/dS>1. ω estimated from the M8 model implemented in PAML. NA: Non-applicable, non-coding intergenic region. [b] Average ω (dN/dS) for sites with dN/dS>1. NA: Non-applicable, since no sites had dN/dS>1. [c] dN/dS estimated for the two haplotypes of CL Brener using PAML's codeml program with the pairwise distance estimation option (runmode = -2). [d] Estimate of the % corrected distance for all sites (Kimura 2-parameter) or for synonymous sites only (Ks) between a strain of TcI (SC13) and a strain of TcII (CBB cl3), corresponding to the largest genetic distance within *T. cruzi*. NA: Non-applicable, non-coding intergenic region.[e] The topology obtained with midpoint rooting (See Figure 2-1 for Topology definitions). ND (Non Described topology): the topology obtained was different from the topologies described in Figure 2-1. [f] HSP70 is an intergenic region, thus selection tests were not conducted. [g] The midpoint rooting topology was different from the topology reconstructed with an outgroup (Table 2-2). * Only significant for M8 vs M8a ($p \leq 0.05$). ** Significant for M8 vs M8a ($p \leq 0.01$), M1 vs M2 ($p \leq 0.05$), and M7 vs M8 ($p \leq 0.05$).***p value ≤0.0001 in all three tests (M1 vs M2, M7 vs M8, M8 vs M8a)

Locus ID	**N**	**bp**	**% sites ω>1 [a]**	**ω [b]**	**CL Brener ω [c]**	**Kimura [d]**	**Ks [d]**	**Midpoint rooting [e]**
COII-ND1	48	1226	0	NA	Only 1 copy	0.095	0.3677	A
Tc00.1047053503555.30	11	1290	0	NA	0.08	0.016	0.0520	A
Tc00.1047053509153.90	40	1473	0.07	1.09	0.04	0.014	0.0403	A
Hsp70 [f]	11	508	NA	NA	NA	0.053	NA	A
Tc00.1047053503885.80	12	945	0.03	9.07***	1.47	0.038	0.0216	A
Tc00.1047053503891.50	10	810	0.06	6.17**	1.54	0.053	0.0268	A
Tc00.1047053503909.76	12	612	0.47	1.09	1.27	0.023	0.0231	B
Tc00.1047053504013.40	25	804	0	NA	0.37	0.029	0.0451	A
Tc00.1047053504045.100	10	885	0	NA	Only 1 copy	0.019	0.0428	A

Table 2-3 continued

Tc00.1047053504057.80	10	855	0	NA	Only 1 copy	0.015	0.0150	ND [g]
Tc00.1047053504059.20	13	894	0.02	2.88	0.74	0.019	0.0316	ND [g]
Tc00.1047053506247.200	10	918	0.03	2.67	Only 1 copy	0.018	0.0080	ND [g]
Tc00.1047053506525.150	13	819	0	NA	2.06	0.019	0.0049	A
Tc00.1047053506529.310	27	726	0.01	13.97***	0.42	0.028	0.0501	A [g]
Tc00.1047053506739.20	11	807	0.33	1.41	Only 1 copy	0.029	0.0120	B [g]
Tc00.1047053507801.70	11	675	0.03	4.79	∞	0.021	0.0002	ND [g]
Tc00.1047053508153.540	13	786	0.02	3.41	6.35	0.030	0.0036	H [g]
Tc00.1047053508461.80	20	699	0	NA	-	0.017	0.0388	G
Tc00.1047053508719.70	24	708	0.19	1.16	0.33	0.020	0.0292	H [g]
Tc00.1047053509007.30	13	813	0.58	1.25	1.59	0.028	0.0166	E
Tc00.1047053509105.70	24	849	0.07	4.07***	0.34	0.037	0.0531	H [g]
Tc00.1047053509561.20	23	879	0.11	5.62***	0.95	0.045	0.0318	A
Tc00.1047053509967.50	11	591	0	NA	2.22	0.024	0.0103	C [g]
Tc00.1047053510101.480	12	828	0.01	3.57	2.19	0.025	0.0109	G [g]
Tc00.1047053510123.24	12	879	0.09	3.21	3.76	0.037	0.0098	A
Tc00.1047053510131.90	12	933	0.01	2.54	2.81	0.022	0.0051	B [g]
Tc00.1047053510765.50	13	813	0.29	2.00*	1.58	0.025	0.0145	A [g]

Table 2-3 continued

Tc00.1047053510877.190	8	453	0.49	1.58	Only 1 copy	0.048	0.014	A
Tc00.1047053510889.210	25	693	0.02	5.96*	1.01	0.030	0.0234	A
Tc00.1047053510889.310	23	762	0.16	2.13	0.82	0.022	0.0343	A
Tc00.1047053511153.124	12	510	0.29	1.69	5.32	0.034	0.0063	C [g]
Tc00.1047053511529.200	13	666	0.008	20.94*	3.51	0.035	0.0089	H [g]

Table 2-4. Bayesian estimates of divergence time (in mya) for different *T. cruzi* lineages. Times to the most recent common ancestor (tMRCA) are shown in mya. In parentheses are 95% HPD (highest posterior density) intervals.[a] tMRCA of *T. cruzi* and its two outgroups (*T. c. marinkellei*, *T. verspertilionis*). [b] tMRCA of extant *T. cruzi* lineages.[c] tMRCA of TcI (nuclear data: SO34, SC13; mtDNA: TEH cl2, CEPA EP, Vin C6, X10 cl1, SABP3, A80, A92, MA-V, OPS21 cl11, CUTIA cl1, 133 79 cl7, V121, 26 79, CUICA cl1, SO34 cl4, P209 cl1, 85/818, P0AC, Esquilo cl1, SC13). [d] tMRCA of strains SO34, SC13, CL35, EP225, CLA39-Haplotype1 and CL_Brener-Haplotype1. [e] tMRCA of strains Florida C16, CANIII, M6241, CM 17, EP 255, 86-1, SO3, EPP, PSC-O, Tulahuen, CL F11F5, VM V4, P63, 86/2036, P251, X9/3, XII0/8 and XI09/2. [f] tMRCA of TcII or TcIII and the respective closest haplotypes from both hybrid DTUs (TcV, TcVI).[g] tMRCA of strains Esmeraldo, X-300, CBB, MCV, MSC2, TU18, MVB.[h] The Relaxed Lognormal clock model for the mitochondrial data set did not converge even after combining 10 independent runs in Tracer. Therefore, the estimates from these analyses are not reliable and not shown here.

Nuclear loci (20 loci)							
Clock model	**Tryps [a]**	***T. cruzi* [b]**	**TcI [c]**	**TcI, TcIII-VI [d]**	**TcII-Hybrids [f] (TcV, TcVI)**	**TcIII-Hybrids [f] (TcV, TcVI)**	**Posterior Likelihood**
Strict	6.23 (4.7-7.7)	1.36 (1-1.7)	0.15 (0.09-0.2)	1.11 (0.8-1.4)	0.49 (0.3-0.6)	0.49 (0.3-0.6)	-57232.7477
Relaxed lognormal	8.3 (3.8-13.8)	2.18 (0.9-3.7)	0.25 (0.08-0.5)	1.69 (0.77-2.9)	0.8 (0.3-1.4)	0.73 (0.3-1.3)	-57183.9809
Mitochondrial loci (COII-ND1) [h]							
Clock Model	**Tryps [a]**	***T. cruzi* [b]**	**TcI [c]**	**TcI, TcIII-VI [e]**	**TcII [g]**		**Posterior Likelihood**
Strict	16.8 (11.1-23.1)	11.00 (7-15.2)	1.76 (1-2.6)	7.5 (4.7-10.5)	0.35 (0.1-0.66)		-4670.0313

Figures

Figure 2-1. Phylogenetic topologies obtained from the 32 analyzed loci. Number on top of each topology represents the number of times that particular topology was observed (Table 2-2). All internal branches shown had bootstrap support values >70%. The topologies are depicted with respect to the classification system that divides *T. cruzi* in two major lineages (Anon 1999), *T. cruzi I* (blue) and *T. cruzi II* (red), and the six major DTUs are labeled. Topology H is consistent with the current classification, and represents a history of divergence in which *T. cruzi I* and *II* are reciprocally monophyletic.

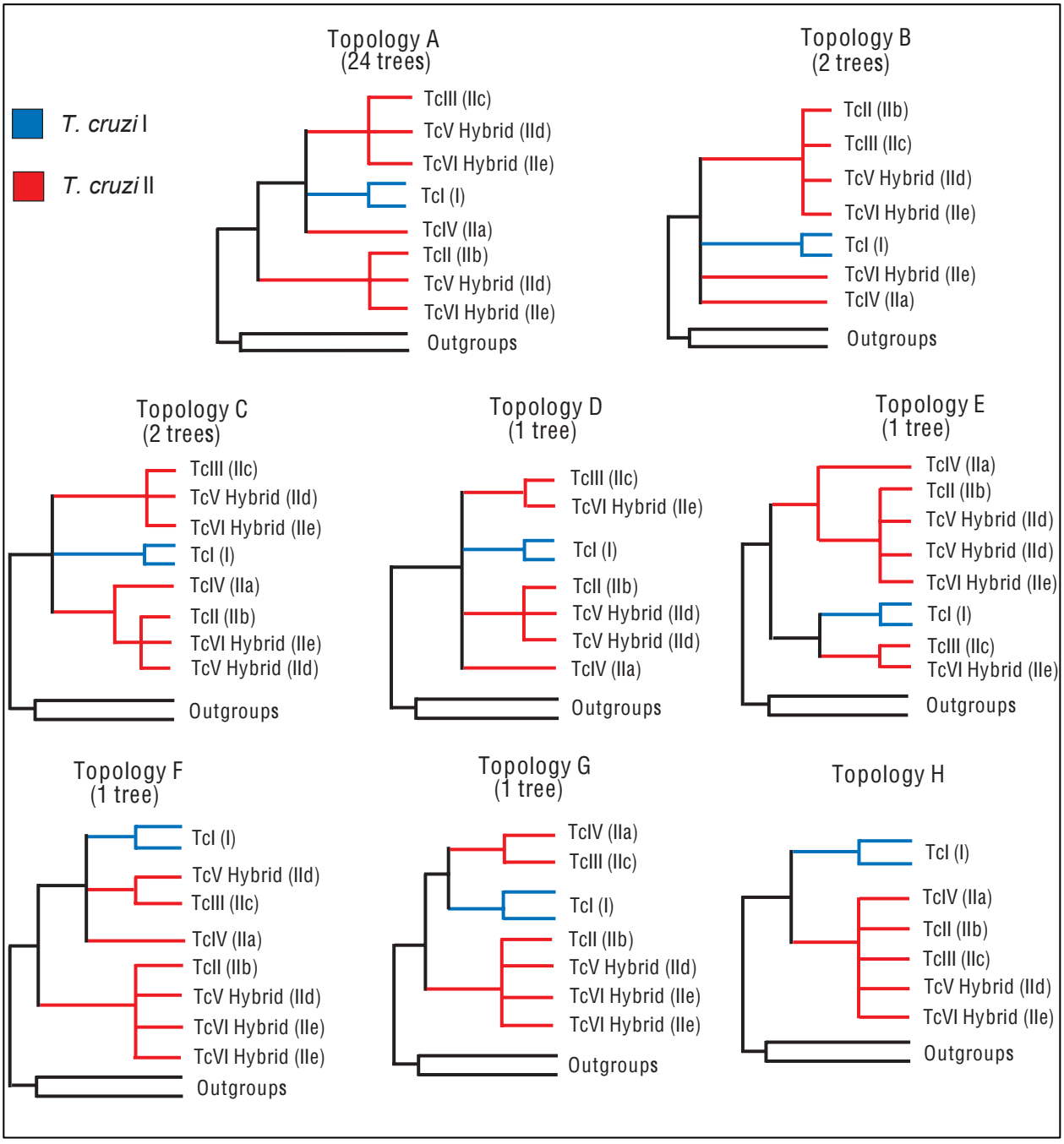


Figure 2-2. Maximum likelihood tree of concatenated data set. Data set consists of 31 nuclear loci and 1 mitochondrial region (COII-ND1), totaling 26,329 nucleotides per strain. Numbers above and below branches are Bootstrap (from ML analyses) and Bayesian support values, respectively. Taxon names represent the six major DTUs. Scale bar in number of substitutions per site.

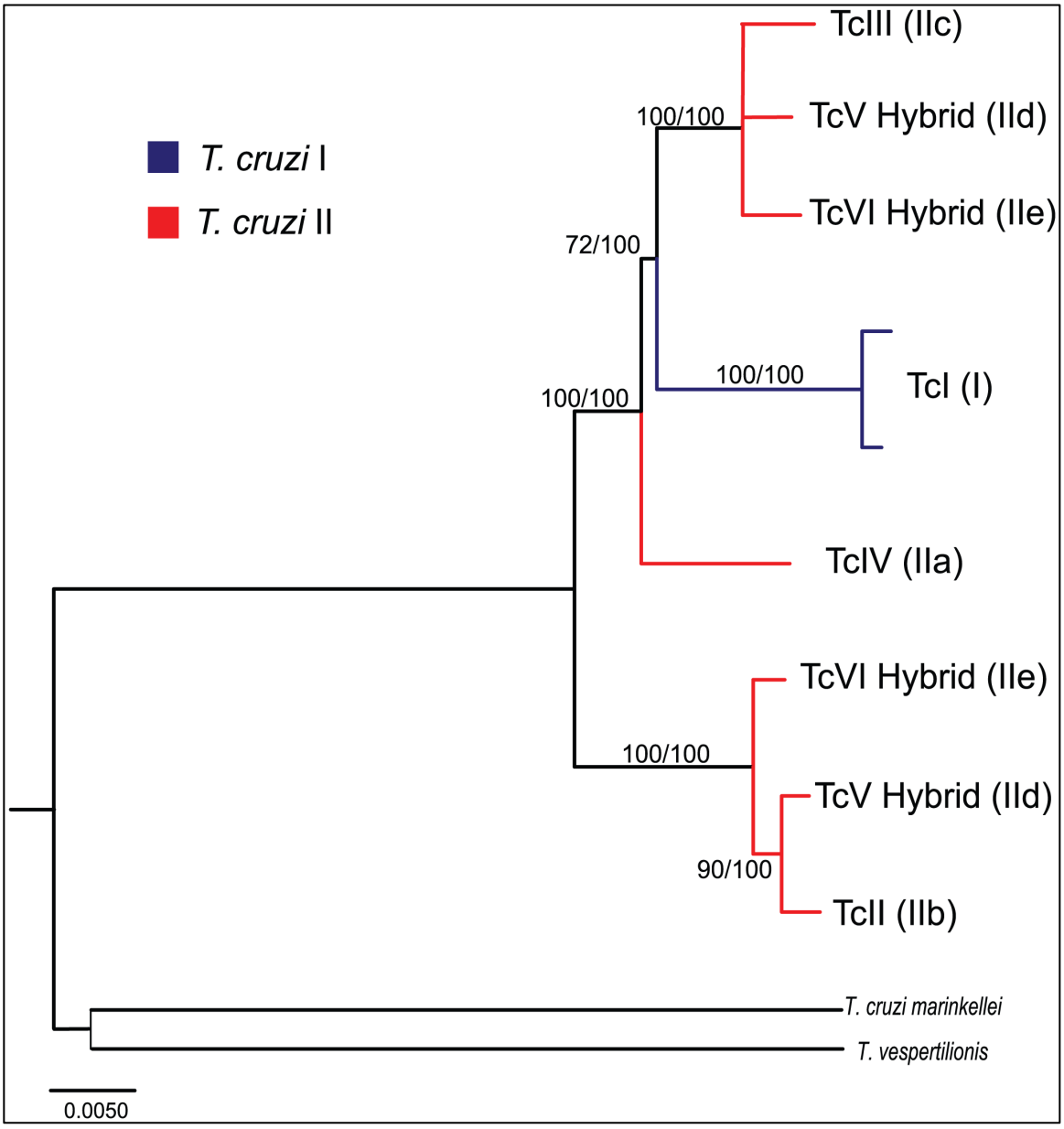
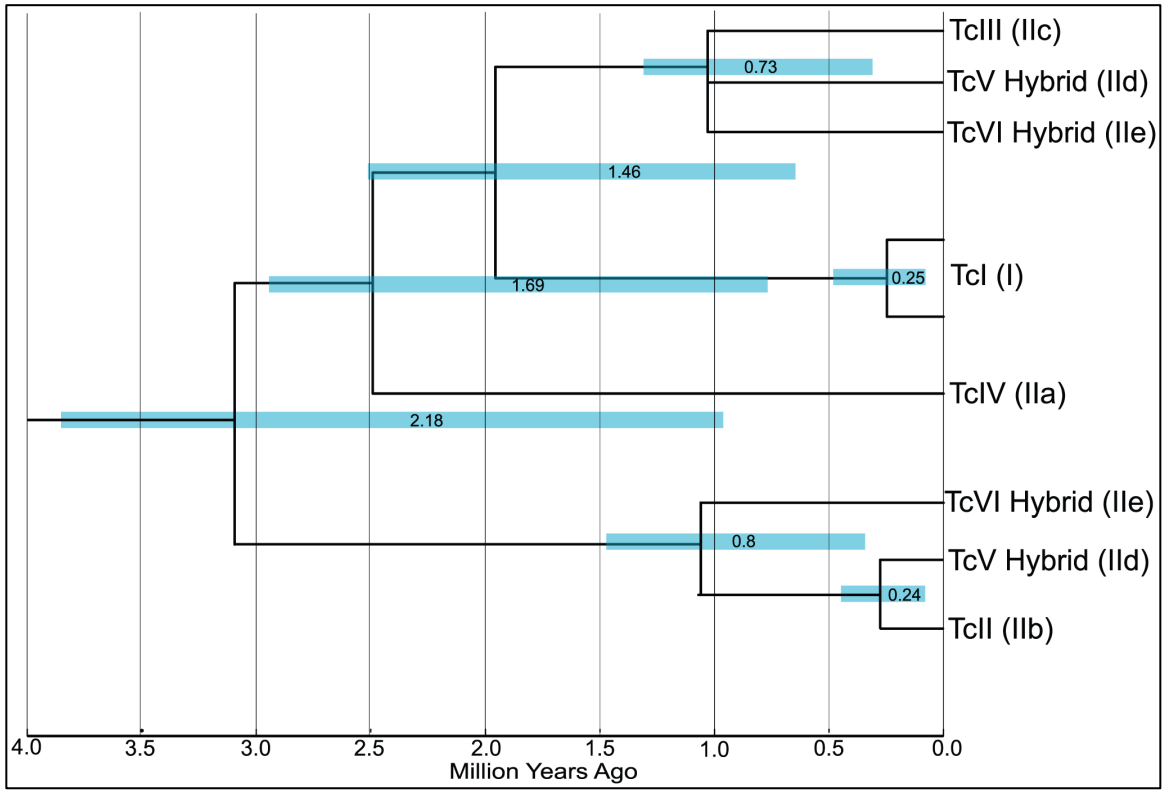


Figure 2-3. Divergence times for main DTU clades of *T. cruzi* using nuclear loci with the relaxed clock model. Data set consists of an alignment of 22 concatenated nuclear loci for which the molecular clock was not rejected (Table S2-4), and that had a homolog in *T. brucei*. Taxon names represent the six major DTUs. Scale bar in millions of years ago (mya).



Chapter 3

Description of a new *Trypanosoma cruzi* lineage from the United States reveals an introduction into North America during the Pleistocene, with evidence of a genetic exchange event with a coexisting lineage.

ABSTRACT

Trypanosoma cruzi is the causative agent of Chagas disease, a devastating parasitic disease endemic to Mexico, Central and South America. It is estimated that the disease causes 10,000 deaths each year and that 7.7 million people are infected with the parasite. Unfortunately, neither a vaccine against infection nor a completely effective treatment for chronic Chagas disease currently exists. The parasite has been isolated from mammals and their triatomine vectors throughout the endemic countries and in southern USA. In the USA, however, only a few autochthonous cases of vectorial transmission of *T. cruzi* have been reported to date. *T. cruzi* is a genetically highly diverse parasite, and many studies indicate that there is an association between this diversity and the variable clinical manifestations of Chagas disease. Currently, *T. cruzi* is divided into 6 groups termed TcI-VI, and a recent study described an additional lineage found in bats. In this study we report the description of a new lineage isolated from North American triatomines from Texas and Arizona, Tc North America (TcNA), by sequencing a mitochondrial locus and three nuclear loci. We additionally report that some of the strains within TcNA appear to have had genetic exchange events with the more commonly found lineage in the USA (TcI), thus supporting the notion that genetic exchange events in this mostly asexual parasite are not uncommon in nature. Time estimates suggest TcNA diverged 110-160 thousand years ago during the Pleistocene.

INTRODUCTION

Chagas disease is caused by the protozoan parasite *Trypanosoma cruzi*, which is transmitted to humans by blood-sucking insects of the family Reduviidae (Triatominae). This is the most important parasitic disease in the Americas, with about 7.7 million people infected, 108 million

people considered at risk, at least 1.7 million symptomatic cases, and an estimated 10,000 deaths each year throughout Mexico, Central and South America (Organization 2006; WHO 2011). In the USA, although both infected triatomines and wild mammalian reservoirs are plentiful [e.g. packrats, mice, armadillos, raccoons and opossums] only seven autochthonous cases of Chagas transmission have been reported to date, all in the southern half of the country (Dorn, et al. 2007; Herwaldt, et al. 2000; Ochs, et al. 1996; Schiffler, et al. 1984; Woody and Woody 1955). A recent study based on data from USA blood banks added sixteen additional cases of presumed autochthonous transmission (Cantey, et al. 2012).

T. cruzi is transmitted to humans when feces and/or urine from infected triatomines (which are deposited during feeding or shortly thereafter) contact the damaged skin or oral and/or eye mucous membranes. Clinical manifestations of the acute stage of parasite infection start 6-10 days after infection and last 1-2 months. Although treatable during the 30-40 day acute phase of infection, most cases are not diagnosed because this phase involves nonspecific symptoms (e.g. rushes of fever). Then the patient enters an asymptomatic phase characterized by a lifelong, low-grade parasitemia. Between 10-30% percent of the infected people develop illness of the cardiac, gastrointestinal and/or nervous systems, which can result in severe debilitation and ultimately death (Barrett, et al. 2003; CDC 2003; Prata 2001)**.**

The transmission cycles of *T. cruzi* are complex. The so-called "sylvatic" cycle involves transmission between non-domiciliated triatomines and their sylvatic mammalian hosts, while the "domestic" cycle involves transmission of the parasite between domiciliated triatomine species, humans, and domestic animals. Sylvatic and domestic transmission cycles may be separate or overlap, depending on the geographical areas (Miles, et al. 2003). In the USA,

transmission of *T. cruzi* mostly occurs within a sylvatic cycle, although triatomines occasionally invade human houses but appear to have a low capacity for domiciliation.

The variable clinical course of Chagas disease has been attributed to the host's response and especially to the genomic heterogeneity of *T. cruzi* (Zingales, et al. 2012). A new intraspecific nomenclature was recently proposed to account for the genetic diversity of *T. cruzi* (Flores-López and Machado 2011; Machado and Ayala 2001). Although the genetic diversity that has been described within *T. cruzi* could merit the division of the taxa into more than one species, the taxonomy of *T. cruzi* has always remained within the species level. Nomenclature of the parasite was recently changed from a system that divided the parasite into two main groups to a six-group division termed TcI-VI (Zingales, et al. 2012). The role played by genetic exchange events in the evolution of this mostly clonal parasite is evident in groups TcV-VI, since these lineages are the product of an ancient hybridization event that took place between the ancestors of groups TcII and TcIII (Machado and Ayala 2001). However, the frequency and role that these genetic exchange events play in nature is not completely understood. Regardless, the current classification scheme of the parasite is likely to change eventually, since the complete genetic diversity of the parasite has definitely not been thoroughly sampled. For instance, an additional lineage that infects bats in Central America termed TcBAT has been recently described (Pinto, et al. 2012). Therefore, as more geographical regions are studied, a more detailed view of the genetic diversity of the parasite will be revealed.

In particular, the much higher genetic diversity of strains found in South America points toward a South American origin of the parasite, with subsequent dispersion of the parasite into Central and North America (Stevens, et al. 1999). Our previous divergence analysis using a concatenated multi locus data set pointed to an origin of the major extant *T. cruzi* lineages in the

past 2 million years (Flores-Lopez and Machado 2011). A recent study proposed that after the dispersion of TcI strains into North America there was a reintroduction of a TcI lineage into South America (Zumaya-Estrada, et al. 2012). Therefore the history of the radiation and subsequent dispersion of *T. cruzi* across the Americas is a complex one. Additional studies need to be conducted in order to elucidate the historical biogeography of the parasite.

The correlation between the variable clinical symptoms of Chagas disease and the genetic background of *T. cruzi* is ambiguous. Some studies support the histotropism hypothesis, in which the genetic background of the parasite can have a large influence in the organs and tissues that become infected and therefore, in the severity of the symptoms observed during the chronic phase of the infection (Mantilla, et al. 2010; Vago, et al. 1996). Furthermore, certain symptoms are more commonly observed in certain geographic areas, where particular lineages of the parasite are frequently associated with the domestic cycle. For instance, in countries north of the Amazon basin, where the most benign forms of Chagas occur, TcI is the most common strain found in the domestic cycle. In contrast, in regions south of the Amazon basin, where the most severe clinical forms of Chagas disease occur, TcII-VI lineages are commonly found in the domestic cycle, while TcI is usually found in the sylvatic cycle (Breniere, et al. 1998; Coura, et al. 2002; Luquetti, et al. 1986; Yeo, et al. 2005). Thus, knowledge of the *T. cruzi* lineages circulating in a given geographical area is important for treatment of infected individuals.

The few studies that have attempted to characterize the genetic diversity of *T. cruzi* in the USA have shown that TcI appears to be the most prevalent lineage found in the country (Barnabe, et al. 2001; Clark and Pung 1994; Roellig, et al. 2008). A recent study that used sequence data described the presence of only two lineages of *T. cruzi* (TcI and TcIV) within the USA, each isolated from both mammals and triatomines (Roellig, et al. 2013). The authors suggest that the

lower genetic diversity observed among TcI samples, in comparison with that of TcIV, indicates that TcI was introduced to the USA more recently than TcIV. Additionally, they reported evidence of a genetic exchange event between TcI and TcIV. Although only a few cases of autochthonous Chagas disease transmission have been reported in the USA (see above), Chagas is becoming and important health issue in the USA owing to the presence of a significant number of blood donors seropositive for *T. cruzi* (Cantey, et al. 2012; Dorn, et al. 2007; Herwaldt, et al. 2000; Leiby 1997; Leiby 2004; Ochs, et al. 1996; Schiffler, et al. 1984; Stramer, et al. 2007; Woody and Woody 1955). Interestingly, some of the seropositive blood donors have never left the USA (Cantey, et al. 2012), indicating that autochthonous transmission maybe higher than previously thought. Moreover, parts of southern USA maybe at a greater risk of vectorial transmission than previously thought because human populations are continuously expanding into habitats where triatomines, many of them infected with *T. cruzi* (Reisenman, et al. 2010) (Sarkar, et al. 2010), readily feed on humans (Stevens, et al. 2012). Therefore, in order to better understand the genetic diversity of *T. cruzi* currently present in USA, we conducted a phylogenetic study of *T. cruzi* isolated from triatomine bugs from two southern USA states; Texas (which has already had an autochthonous case of Chagas) and Arizona, where triatomines are abundant and in contact with human beings (Beard, et al. 2003; Bern, et al. 2011; Kjos, et al. 2009).

MATERIALS AND METHODS

Collection and location of samples. Hand collection, dry ice traps as well as white light and UV traps were used for triatomine collected in Texas between 2009-2012. Triatomines from Arizona were collected near Tucson (32°13′18″ N, 110°55′35″ W). Triatomines were manually collected

in 2009 outside human houses, attracted by patio and porch lights, during the insect dispersal season (beginning of May through July) (Reisenman, et al. 2010). *Triatoma recurva* is one of the three triatomine species commonly found in the area, and it has been reported to be infected with *T. cruzi* in previous studies (Reisenman, et al. 2010). Collected insects were individually placed in 95% ethanol immediately after collection or upon death and stored at 4°C until analysis

DNA from *T. cruzi* strains that pertain to all major Discrete Typing Units (Table 3-1) were kindly provided by Dr. Christian Barnabé. DNA was extracted using the QIAGEN extraction kit. The locality of all collection sites and triatomine species is shown in Table 3-1 and Figure 3-1.

Molecular Methods. Triatomine guts were dissected and used for DNA extraction. Primers that target a mitochondrial region (1226 bp) encompassing the maxicircle-encoded genes cytochrome oxidase subunit II and NADH dehydrogenase subunit 1 (COII-NDI) (Machado and Ayala 2001), as well as a partial segment (727 bp) of a nuclear gene with unknown function (Tc00.1047053506529.310) found on chromosome 6 (Flores-Lopez and Machado 2011), were used to screen all samples. A region of the mismatch repair gene (MSH2) (Augusto-Pinto, et al. 2001) and a region of the Dihydrofolate reductase-thymidylate synthase (DHFR-TS) (Machado and Ayala 2002) gene were additionally amplified in the majority of samples from Texas. The following primers (5'-3') were used for PCR: for DHFR-TS region, (CGCTGTTTAAGATCC GNATGCC, CGCATAGTCAATGACCTCCATGTC); for the Tc00.1047053506529.310 region, (TTCTTTCAGGCTGCGATTTT, CGCTGTTTGGCTCATTTCTT); for the COII-NDI region, (GCTACTARTTCACTTTCACATTC, GCATAAATCCATGTAAGACMCCACA); for the MSH2 region, (ACAGTTTCTGTACTATATTG, AGGTGGATGGAATTGTATGC). Conditions for the PCR amplifications were: 35 cycles of a 30 second denaturation step at 94°C,

annealing at 55°C and 58°C for 30s for the nuclear and mitochondrial loci respectively, and extension at 68°C for 60 and 90s respectively. PCR primers were used for bidirectional sequencing on a 3730xl DNA Analyzer (Applied Biosystems). Sequences were edited using CodonCode Aligner (Codoncode Coorporation).

Alignment and phylogenetic analysis. Sequences were aligned using MUSCLE (Edgar 2004) and then manually checked. A Neighbor Joining (NJ) tree was reconstructed for each data set and each topology was used to estimate maximum likelihood parameters for different models of nucleotide substitution. The most appropriate nucleotide substitution model to analyze each locus was chosen using jModeltest (Posada 2008). Maximum likelihood (ML) trees with 100 pseudoreplicates of bootstrap were individually obtained for each locus using RAxML (Stamatakis, et al. 2008). MrBayes 3.1.2 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003) was used to conduct Bayesian analyses using the substitution models chosen by jModeltest (Posada 2008). We ran two independent simultaneous Markov Chain Monte Carlo runs with four chains each for 1,000,000 generations and sampled trees every 10 generations. If the standard deviation of split frequencies as not below 0.01 after analyses were done, the analyses were run for an additional 1,000,000 generations and were stopped after standard deviation of split frequencies were < 0.01. Parameters and corresponding trees were summarized after discarding the initial 25% of each chain as burnin.

Divergence time estimates. A molecular clock was enforced on each data set and the likelihood of the ML tree was compared to the tree without the molecular clock assumption to determine if any data set did not conform to the molecular clock. Divergence dates were estimated using Bayesian analysis in BEAST v1.5.3 (Drummond and Rambaut 2007). Both the strict and relaxed Lognormal clock models were used to estimate divergence times on the mitochondrial and the

concatenated nuclear loci data sets. A concatenated data set of the 3 nuclear loci and the mitochondrial alignments were independently used to estimate the divergence dates of the major *T. cruzi* lineages. All analyses were conducted without any topological constraints using the HKY substitution model with the gamma plus invariant sites as the site heterogeneity model, with 4 gamma categories, as well as partitioning of codons into 3 positions. All priors were set to default values, except for the divergence estimate between *T. cruzi* and *T. c. marinkellei*, which was set to 6.23 million years ago (estimates from Chapter 2). Times of divergence were obtained by converging 2 independent Markov Chain Monte Carlo (MCMC) runs in Tracer v1.5 (Drummond and Rambaut 2007). Each run had a chain length of 10 million, with sampling every 1000 chains. Time of divergence was estimated for the major crown nodes in the phylogeny (Table 3-2).

RESULTS

A new lineage of *T. cruzi* from southwest USA

The majority (75%, 38/51) of the Texas *T. cruzi* lineages isolated from our samples cluster within the TcI clade (Fig. 3-3), as expected based on results from previous studies (Clark and Pung 1994; Roellig, et al. 2013). Unfortunately, we were only able to amplify one nuclear locus (Tc00.1047053506529.310) from the samples collected in Arizona, of which only one sample clustered within the TcI clade (Fig S3-1). None of the samples isolated in this study appeared to be TcIV.

The majority of samples from Arizona (3/4) and 25% of the samples from Texas clustered within a new un described monophyletic *T. cruzi* clade that is strongly supported to be a sister taxa to TcIII and TcIV (Fig 3-3). We name this new *T. cruzi* lineage Tc North America

(TcNA). TcNA is different from TcBAT, since TcBAT appears to be a sister taxa to TcI (Pinto, et al. 2012). A sample from Gainesville (Florida) that was previously typed as a TcI strain, also clustered within the mitochondrial phylogenetic analysis within the TcNA clade (Fig 3-2), as it was previously reported (Machado and Ayala 2001). Additionally, a sample from southern Mexico appears to cluster within TcNA (unpublished data).

The newly described *T. cruzi* lineage is not specific to a particular species of insect vector. TcNA sequences were amplified from five North American triatomine species (*Triatoma gerstaeckeri*, *Triatoma. lecticularia*, *Triatoma indictiva*, *Triatoma sanguisuga* and *Triatoma recurva*), not including the sample from southern Mexico.

Genetic exchange events

Four samples (Tex 1, 16, 72 & 260) had conflicting phylogenies among the nuclear and mitochondrial loci (Fig 3-2, Fig 3-3, Fig S3-2). This phylogenetic incongruence has been typically observed in *T. cruzi* studies when a genetic exchange event has taken place among distinct lineages, a phenomena that is observed in the well-studied hybrid strains of the parasite (Machado and Ayala 2001). Most nuclear loci in the hybrid strains of *T. cruzi* (TcV & TVI) are heterozygous due to the double parental ancestry (TcII and TcIII) of their nuclear DNA, and therefore show heterozygous sites in their nuclear sequence chromatograms. Usually, molecular cloning of the PCR products is required in order to attain both haplotypes of these loci. Since the mitochondria have a uniparental mode of inheritance, the presence of heterozygous sites in a mitochondrial chromatogram indicates to the amplification of two distinct lineages of the species under study. In the case of *T. cruzi*, this would suggest that the triatomine vector or the mammalian host was co-infected with distinct lineages of the parasite. One of the Texas samples that showed phylogenetic incongruence between the mitochondrial and nuclear phylogenies (Tex

260) had a few sites that appeared to be ambiguous sites in the mitochondrial chromatogram. This suggests that the sample was probably infected with at least two distinct *T. cruzi* lineages (TcI and TcNA). The additional two samples from Texas showing phylogenetic incongruence (Tex 1 and Tex 72) did not appear to have ambiguous sites in their respective mitochondrial chromatograms. One of these samples (Tex 72) had several ambiguous sites in the MSH2 loci, which all corresponded to sites that appear to be polymorphic sites that distinguish TcI from TcNA (Fig 3-3). The reference sample from Florida also had conflicting phylogenies among the nuclear and the mitochondrial analysis, as was previously observed (Machado and Ayala 2001), with the exception that the mitochondrial inheritance appears to be from the TcNA, the opposite of what we found in the two Texas samples (Tex 1 and Tex 72) (Fig 3-3). Sample Tex 16 had a couple of heterozygous sites on the MSH2 loci and appeared as a TcI strain in this phylogeny (Fig. S3-3), whereas all other loci did not show any heterozygous sites and clustered this sample within TcNA. Thus, at this point, it is not clear if this isolate had a genetic exchange event between TcI and TcNA, or whether the insect had a double infection and the mitochondrial amplification only rescued one of the haplotypes.

Divergence analysis

The results from the divergence analysis pointed to slightly different times for the estimation of the start of divergence of TcNA (Table 3-2). Mitochondrial analysis estimated that the divergence of the TcI clade of North America diverged between 260 and 315 thousand years ago, whereas the estimate for the TcNA clade was between 120-160 thousand years ago. The concatenated nuclear data analysis estimated that the divergence of the TcI clade occurred around 70 thousand years ago, whereas the TcNA clade diverged between 110 thousand years ago. Both time estimates agree with the fact that the divergence occurred well before humans

arrived in the Americas, suggesting that the parasite entered North America before humans entered the Continent.

DISCUSSION

TcI in the USA

We confirm previous studies, which indicated that TcI is the most common lineage of *T. cruzi* in the USA. This is not surprising, given that TcI is also the most frequently reported lineage found in neighboring Mexico (Bosseno, et al. 2002; Espinoza, et al. 1998; Roellig, et al. 2008; Roellig, et al. 2013; Zumaya-Estrada, et al. 2012). The lineage of TcI is of special interest because it has been suggested that this lineage does not cause the severe "acute" clinical manifestations of Chagas, which are often observed in regions below the Amazon basin (Breniere, et al. 1998, Coura, et al. 2002; Luquetti, et al. 1986; Yeo, et al. 2005).

A recent study from USA (from both wildlife mammal reservoirs and triatomines) that include samples from a larger number of geographical areas reported the presence of TcIV (Roellig, et al. 2013) in samples from Florida, Georgia and Tennessee. In our study, which includes samples from Arizona and Texas only (except for one reference sample from Gainesville, Florida), we did not find any TcIV lineages. This suggests that the TcIV lineage maybe restricted to southeast areas of the USA. In contrast, TcI appears to have a much broader distribution, since it has been found in most states of the USA where the parasite has been isolated.

Recent phylogenetic studies that incorporate the large genetic diversity of TcI found across the Americas reported that TcI samples isolated from North America tend to form a distinct cluster nested within the TcI isolates from South America (Cura, et al. 2010; Llewellyn,

et al. 2009b), which suggests that the TcI isolates found in North America are derived from South America. The timing of the introduction is not clear; some authors suggest that the dispersal could have been facilitated by recent human migration (Lewis, et al. 2011). Our mitochondrial divergence analysis, in contrast, indicates that the divergence of the TcI clade from North America occurred between 250-310 thousand years ago (Table 3-2), suggesting that the introduction of *T. cruzi* into North America preceded the presence of humans in the Americas.

With sequencing efforts improving and expanded sampling of geographic regions the genetic diversity and evolutionary history of the parasite has become much clearer (Flores-Lopez and Machado 2011; Pinto, et al. 2012). The result from a recent study suggests that the movement of *T. cruzi* strains across the American continent is more complex than was previously thought (Zumaya-Estrada, et al. 2012). Thus, increasing the sampling to additional understudied areas (i.e. USA, Mexico and Central America), as well as screening a wider number of potential hosts, will increase our knowledge on the complete genetic diversity of the parasite present in nature.

Description of TcNA, a new *T. cruzi* lineage

The finding of a completely new un described clade of *T. cruzi* is not surprising considering the vast geographical area of North America, and the few phylogenetic studies of this nature in the area. The recent description of a new *T. cruzi* clade found in bats from Panama reflects this shortage of studies from certain geographical areas. It could also reflect the use of molecular markers that lack the resolution needed to identify such diversity, or the failure to use outgroups in the phylogenetic reconstruction, an error that has been shown to produce artificial phylogenetic groupings that do not represent the correct evolutionary history of the parasite (Flores-Lopez and Machado 2011).

A *T. cruzi* sample that was isolated from a *T. sanguisuga* specimen collected in Gainesville, Florida (Florida C16), and that has been used as a TcI reference strain in the past (kindly donated by Dr. Christian Barnabé), clustered with strong support within the TcNA in the mitochondrial analysis (Fig 3-2). Nuclear markers link this reference strain to the TcI clade (Fig 3-3). This event was previously described more than 10 years ago (Machado and Ayala 2001), but since there was only one TcNA sample in that original study the sample was basally positioned in the TcIII clade (Machado and Ayala 2001). This phylogenetically incongruent pattern is consistent with a genetic exchange event; such events are currently being described more often than before (Machado and Ayala 2001; Roellig, et al. 2013). Additionally, a sample isolated from a *T. dimidiata* specimen collected from the Yucatan peninsula in southern Mexico also clusters within TcNA (unpublished, data not shown). Together, these results obtained from the Florida and Southern Mexico samples, which cluster with TcNA, suggest that this new clade is not restricted to Texas and Arizona, and that the geographic distribution of this clade might be much broader than reported here. More sampling is needed in order to have a better understanding of the geographical distribution of this clade.

The timing of the entry of the ancestor of TcNA into the USA cannot be confidently estimated. The nuclear and mitochondrial analyses gave very different estimates, which is not surprising due to the faster evolutionary rate observed in mitochondria, compared to nuclear DNA, which tends to give older divergence estimates for mitochondrial analyses compared to nuclear analyses. However, both time estimates point towards a Pleistocene entry into the USA (Table 3-2, Fig. 3-4), well before humans inhabited the Americas.

Genetic exchange event

The potential genetic exchange event that took place between TcI and TcNA definitely seems plausible, due to the co-occurrence of these lineages in time and space. Both lineages were amplified in *Triatoma gerstaeckeri*, *Triatoma lecticularia*, *Triatoma indictiva* and *Triatoma sanguisuga*, which have geographic overlapping distributions. It is not known, however, how long these *T. cruzi* lineages have been in contact with each other co-occurring in the same niche. In spite of this, given the divergence estimates presented here and assuming that the divergences (Table 3-2) of TcNA and the USA TcI strains occurred in the USA, the likelihood of a genetic exchange events among these strains seem plausible. The biological implications for these genetic exchange events in nature have not been addressed.

CONCLUSIONS

Infection rates, lack of domiciliated vectors, socioeconomic factors as well as habitat conditions that are usually associated with Chagas are much lower in the USA than in Mexico, Central and South America. Nevertheless the epidemiological importance of TcNA remains to be studied. The virulence associated to this new clade is unknown, and further experimental and broader geographical sampling should help to determine the range and specific niche of TcNA as well as its potential pathogenicity for humans.

Tables

Table 3-1. Information for isolated samples from Texas and Arizona

Sample ID	Genus	Species	Longitude	Latitude	Location
Tex1	*Triatoma*	*gerstaeckeri*	-99.250	29.56812	Hondo, Texas
Tex2	*Triatoma*	*gerstaeckeri*	-99.250	29.56812	Hondo, Texas
Tex15	*Triatoma*	*sanguisuga*	-97.375	30.217	Elgin, Texas
Tex16	*Triatoma*	*sanguisuga*	-97.375	30.217	Elgin, Texas
Tex26	*Triatoma*	*indictiva*	-98.186	30.235	Hays County, Texas
Tex34	*Triatoma*	*gerstaeckeri*	-99.766	29.349	Uvalde County, Texas
Tex35	*Triatoma*	*gerstaeckeri*	-99.766	29.349	Uvalde County, Texas
Tex44	*Triatoma*	*gerstaeckeri*	-99.766	29.349	Uvalde County, Texas
Tex72	*Triatoma*	*lecticularia*	-99.179314	29.18498541	Medina County, Texas
Tex77	*Triatoma*	*gerstaeckeri*	-99.179314	29.18498541	Medina County, Texas
Tex78	*Triatoma*	*gerstaeckeri*	-99.179314	29.18498541	Medina County, Texas
Tex92	*Triatoma*	*gerstaeckeri*	-99.179314	29.18498541	Medina County, Texas
Tex96	*Triatoma*	*gerstaeckeri*	-99.179314	29.18498541	Medina County, Texas
Tex102	*Triatoma*	*gerstaeckeri*	-99.179314	29.18498541	Medina County, Texas
Tex103	*Triatoma*	*gerstaeckeri*	-99.179314	29.18498541	Medina County, Texas
Tex139	*Triatoma*	*gerstaeckeri*	98.2548052	26.09747629	519 8th St., Hidalgo, Texas, 78557
Tex173	*Triatoma*	*gerstaeckeri*	-97.227244	26.073822	02 E.A.C.P. 1765 hwy 100, Port Isabel, TX
Tex176	*Triatoma*	*gerstaeckeri*	-98.070297	26.530892	03 E.Q. La Sal del Rey, Hidalgo County, Texas
Tex177	*Triatoma*	*gerstaeckeri*	-98.070297	26.530892	4 E.Q. La Sal del Rey, Hidalgo County, Texas
Tex178	*Triatoma*	*gerstaeckeri*	-98.344465	26.398811	03 E.A.C.P. 22675 N. Moorefield Rd., Edinburg, TX, 78541
Tex192	*Triatoma*	*gerstaeckeri*	-98.350567	26.181283	Hidalgo County TX
Tex197	*Triatoma*	*gerstaeckeri*	-98.350567	26.181283	Hidalgo County TX

Table 3-1 continued

Tex199	*Triatoma*	*gerstaeckeri*	-98.350567	26.181283	Hidalgo County TX
Tex200	*Triatoma*	*gerstaeckeri*	-98.350567	26.181283	Hidalgo County TX
Tex205	*Triatoma*	*gerstaeckeri*	-98.350567	26.181283	Hidalgo County TX
Tex211	*Triatoma*	*gerstaeckeri*	-98.350567	26.181283	Hidalgo County TX
Tex217	*Triatoma*	*gerstaeckeri*	-98.350567	26.181283	Hidalgo County TX
Tex218	*Triatoma*	*gerstaeckeri*	-98.350567	26.181283	Hidalgo County TX
Tex228	*Triatoma*	*gerstaeckeri*	-98.350567	26.181283	Hidalgo County TX
Tex231	*Triatoma*	*gerstaeckeri*	-98.350567	26.1813	Hidalgo County TX
Tex233	*Triatoma*	*gerstaeckeri*	-98.350567	26.1813	Hidalgo County TX
Tex236	*Triatoma*	*gerstaeckeri*	-98.350567	26.1813	Hidalgo County TX
Tex253	*Triatoma*	*lecticularia*	-99.17651	28.58779	La Salle Cty, TX
Tex254	*Triatoma*	*lecticularia*	-99.17651	28.58779	La Salle Cty, TX
Tex255	*Triatoma*	*gerstaeckeri*	-99.177253	28.58596	La Salle Cty, TX
Tex256	*Triatoma*	*gerstaeckeri*	-99.177253	28.58596	La Salle Cty, TX
Tex260	*Triatoma*	*lecticularia*	-99.250555	29.578333	Medina County, Texas
Tex261	*Triatoma*	*gerstaeckeri*	-99.250555	29.578333	Medina County, Texas
Tex262	*Triatoma*	*gerstaeckeri*	99.2505556	29.586667	Medina County, Eagle Bluff Ranch, TX
Tex265	*Triatoma*	*gerstaeckeri*	-98.35055	26.18121667	Bastrop Cty, TX
Tex266	*Triatoma*	*gerstaeckeri*	-98.35055	26.18121667	Bastrop Cty, TX
Tex267	*Triatoma*	*gerstaeckeri*	-98.35055	26.18121667	Bastrop Cty, TX
Tex272	*Triatoma*	*gerstaeckeri*	-98.35055	26.18121667	Bastrop Cty, TX
Tex275	*Triatoma*	*gerstaeckeri*	-98.35055	26.18121667	Bastrop Cty, TX
Tex276	*Triatoma*	*gerstaeckeri*	-98.35055	26.1812167	Bastrop Cty, TX
Tex279	*Triatoma*	*gerstaeckeri*	-98.35055	26.18121667	Bastrop Cty, TX
Tex280	*Triatoma*	*gerstaeckeri*	-98.35055	26.18121667	Bastrop Cty, TX
Tex287	*Triatoma*	*gerstaeckeri*	-97.9563	30.146613	24 Concord Circle, Austin, Hays County, TX, 78737

Table 3-1 continued

Tex293	*Triatoma*	*gerstaeckeri*	-97.570179	30.102131	Bastrop Cty, TX
EW6	*Triatoma*	*gerstaeckeri*	No data	No data	Uvalde County, Texas
EW7	*Triatoma*	*gerstaeckeri*	No data	No data	Uvalde County, Texas
DF78 (09-371)	*Triatoma*	*recurva*	No data	No data	Rio Rico, AZ
DF79 (09-252)	*Triatoma*	*recurva*	No data	No data	Vail, AZ
DF80 (09-502)	*Triatoma*	*recurva*	No data	No data	Rio Rico, AZ
DF81 (09-160)	*Triatoma*	*recurva*	No data	No data	Tubac, AZ

Table 3-2 Bayesian estimates of divergence (time in millions of years) for the main *T. cruzi* lineages and TcNA. Nuclear data consisted of 2757 concatenated nucleotides, whereas mitochondrial data set consisted of 1161 nucleotides. Due to inconsistency of amplifying all 3 nuclear loci in each sample, for the nuclear analyses we only included isolates that had all three nuclear loci. Both analyses were calibrated with the divergence estimate of 6.23 million years ago (with 1.5 million years of standard deviation with a prior normal distribution for the tmrca) between *T. cruzi* and *T. c. marinkellei* that was estimated from our previous analysis (Flores-Lopez and Machado 2011). Parameters: HKY as DNA substitution model, gamma with invariant sites as the site heterogeneity model, with 4 gamma categories, data was partitioned into codons. Convergence could not be attained for the nuclear relaxed analysis thus data is not shown. TcI North Am: clade of TcI that is exclusively composed of samples from North America. TcI South Am: clade of TcI that is exclusively composed of samples from South America. Tex 16 was not included in the analysis due to conflicting phylogenies between nuclear trees (Fig S3-3).

Mitochondrial loci (COII-NDI)								
	T.cruzi	**TcII**	**TcIII-IV**	**TcI North Am.**	**TcI South Am.**	**TcI**	**TcNA**	**Posterior likelihood**
Strict	3.49 (2.09-4.8)	0.209 (0.07-0.36)	0.542 (0.28-0.83)	0.263 (0.11-0.41)	0.405 (0.21-0.61)	0.475 (0.26-0.71)	0.12 (0.03-0.21)	-4659.033
Relaxed	3.52 (1.36-5.8)	0.244 (0.047-0.51)	0.56 (0.17-1.03)	0.315 (0.086-0.59)	0.454 (0.13-0.81)	0.59 (0.19-1.07)	0.159 (0.032-0.34)	-4656.375
NUCLEAR (3 loci)								
	T.cruzi	**TcII**	**TcIII-IV (not informative)**	**TcI**	**TcNA**		**Posterior likelihood**	
Strict	1.63 (0.95-2.33)	0.43 (0.19-0.69)	1.15 (0.67-1.7)	0.073 (0.001-0.18)	0.11 (0.02-0.22)		-5124.638	

Figures

Figure 3-1. Map of sample location in the USA. Geographical location within the USA of the samples used in this study. TcI was also found in all TcNA sites from Texas.

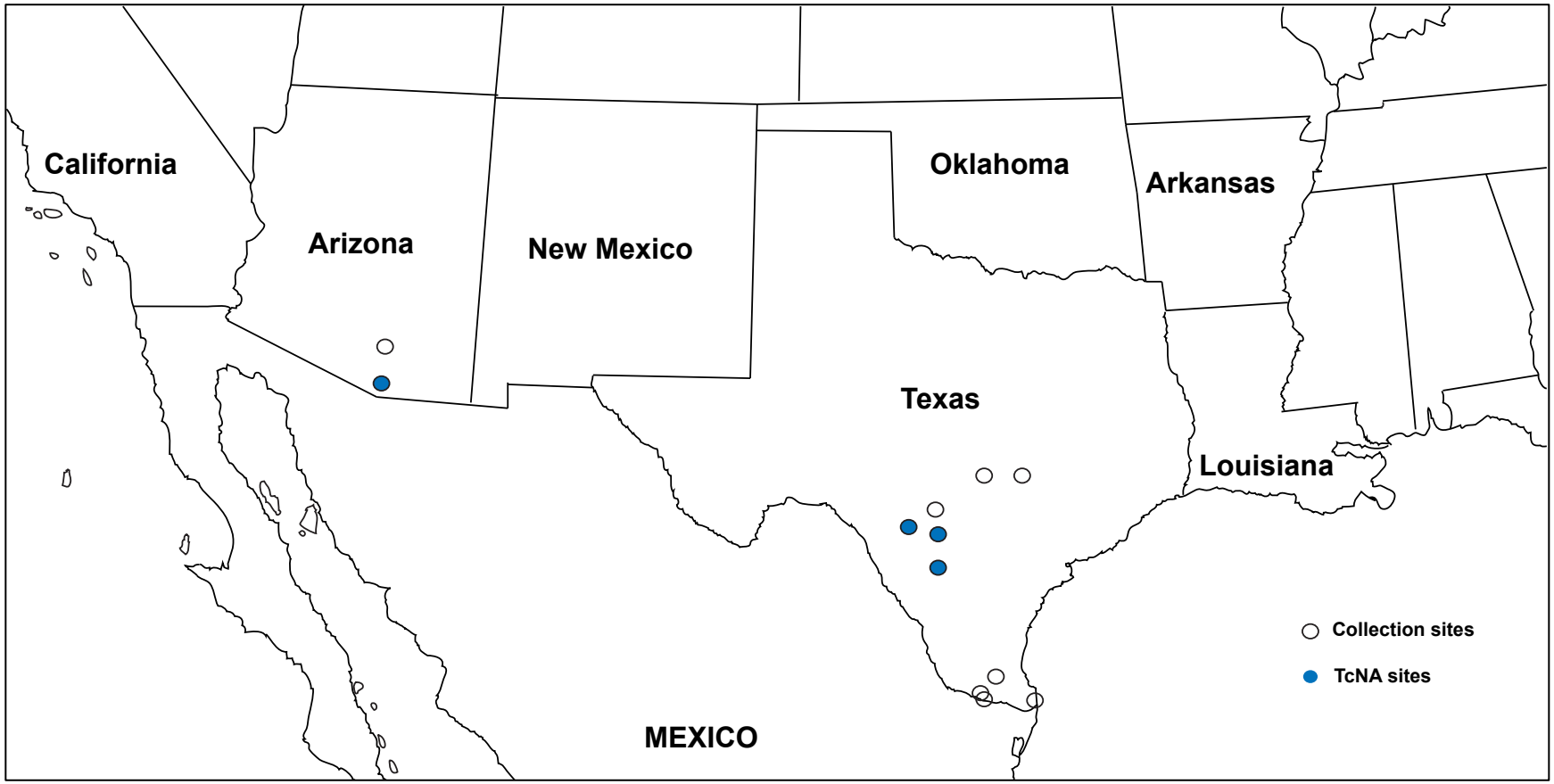


Figure 3-2. Bayesian mitochondrial phylogeny. Data set consists of a 1706 bp of the COII-NDI loci. Numbers above main nodes represent Bayesian posterior probabilities. All major *T. cruzi* clades (TcI-VI) are highlited in phylogeny (except TcBAT).

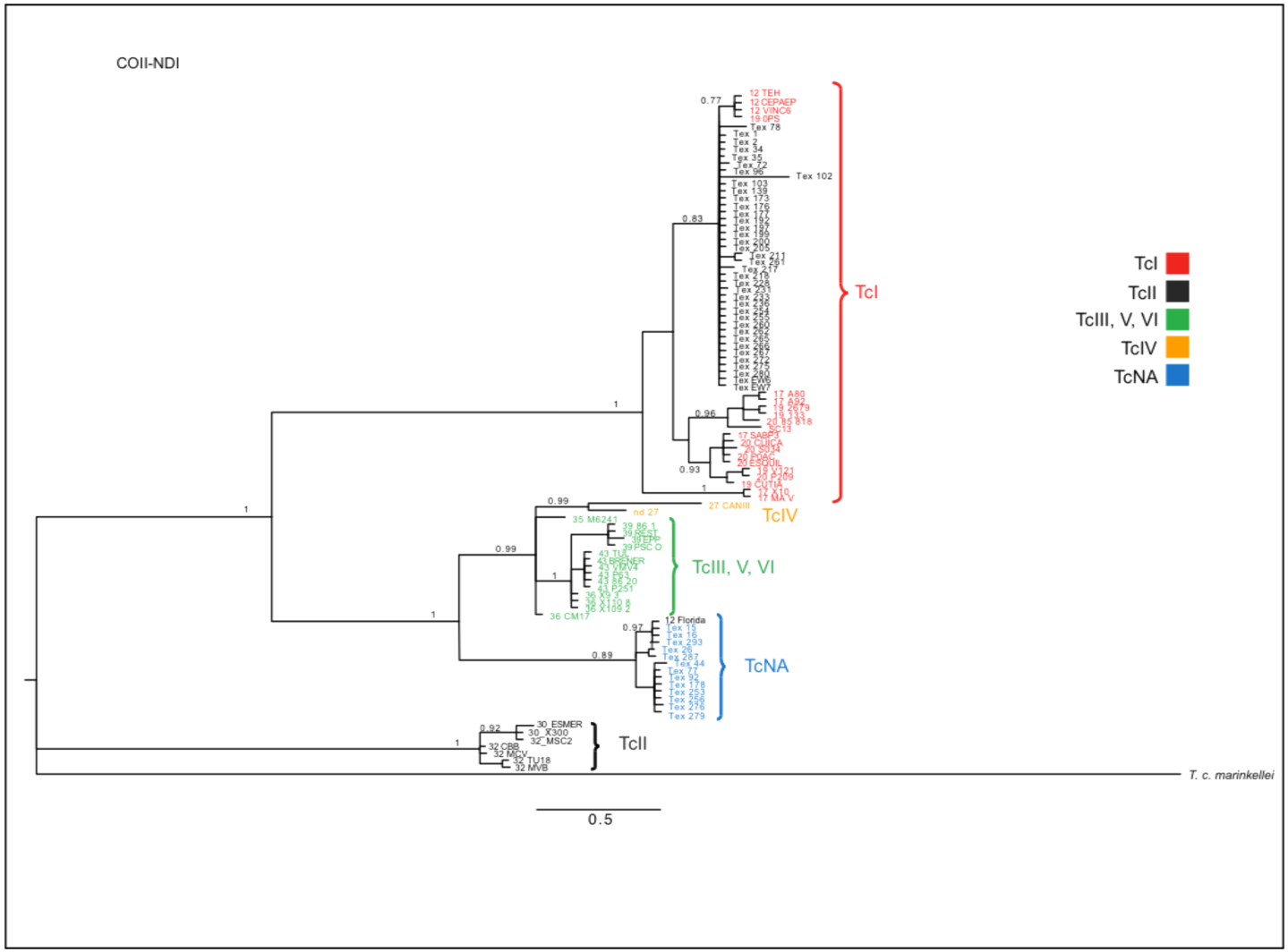


Figure 3-3. Bayesian nuclear phylogeny. Concatenated data set of all three nuclear loci was assembled for analysis (2757 bp). Maximum likelihood bootstraps higher than 70% are shown above major nodes. Bayesian posterior probabilities higher than 70% are shown below major nodes. All major *T. cruzi* clades (TcI-VI) are highlited in phylogeny (except TcBAT).

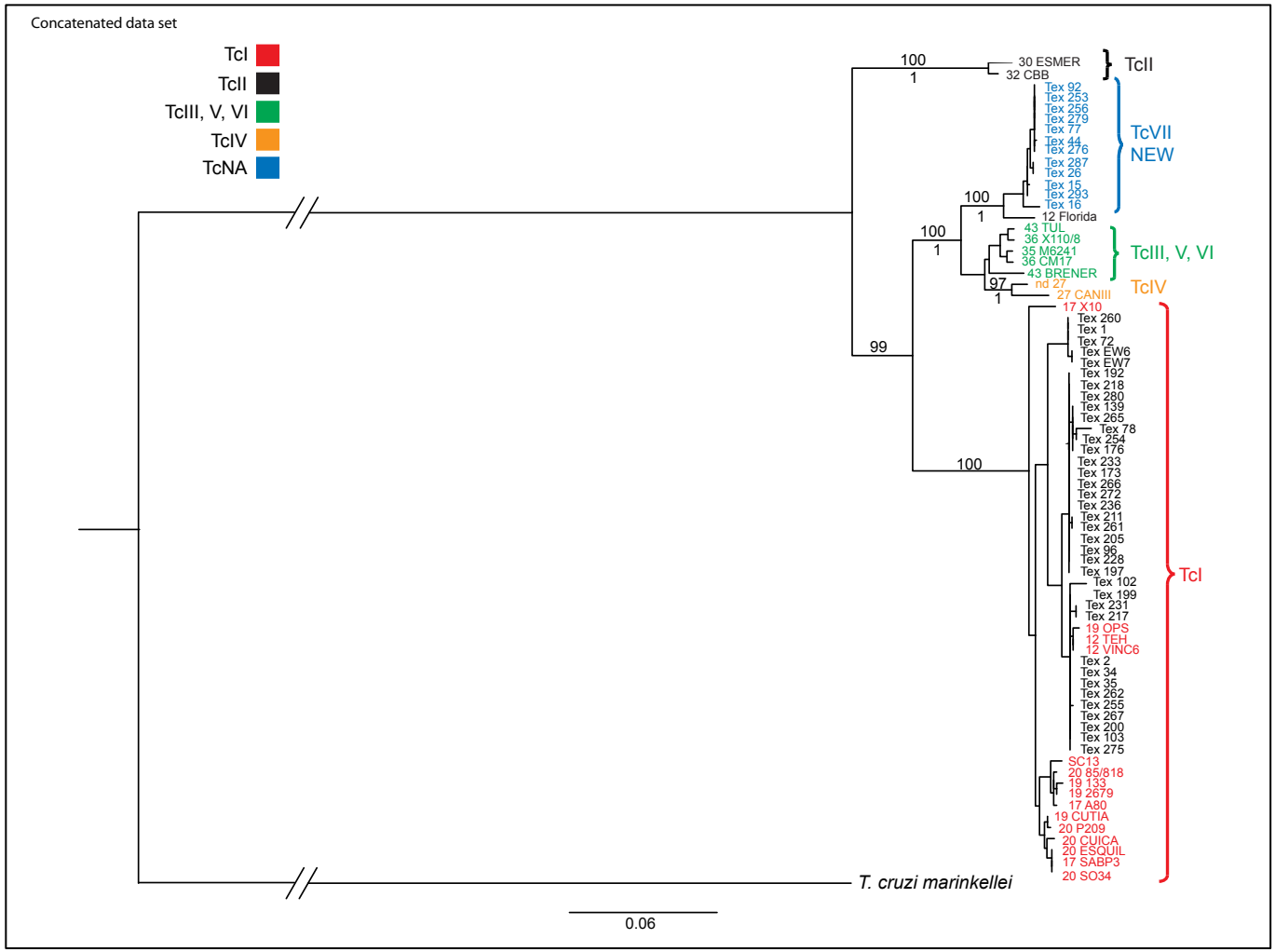
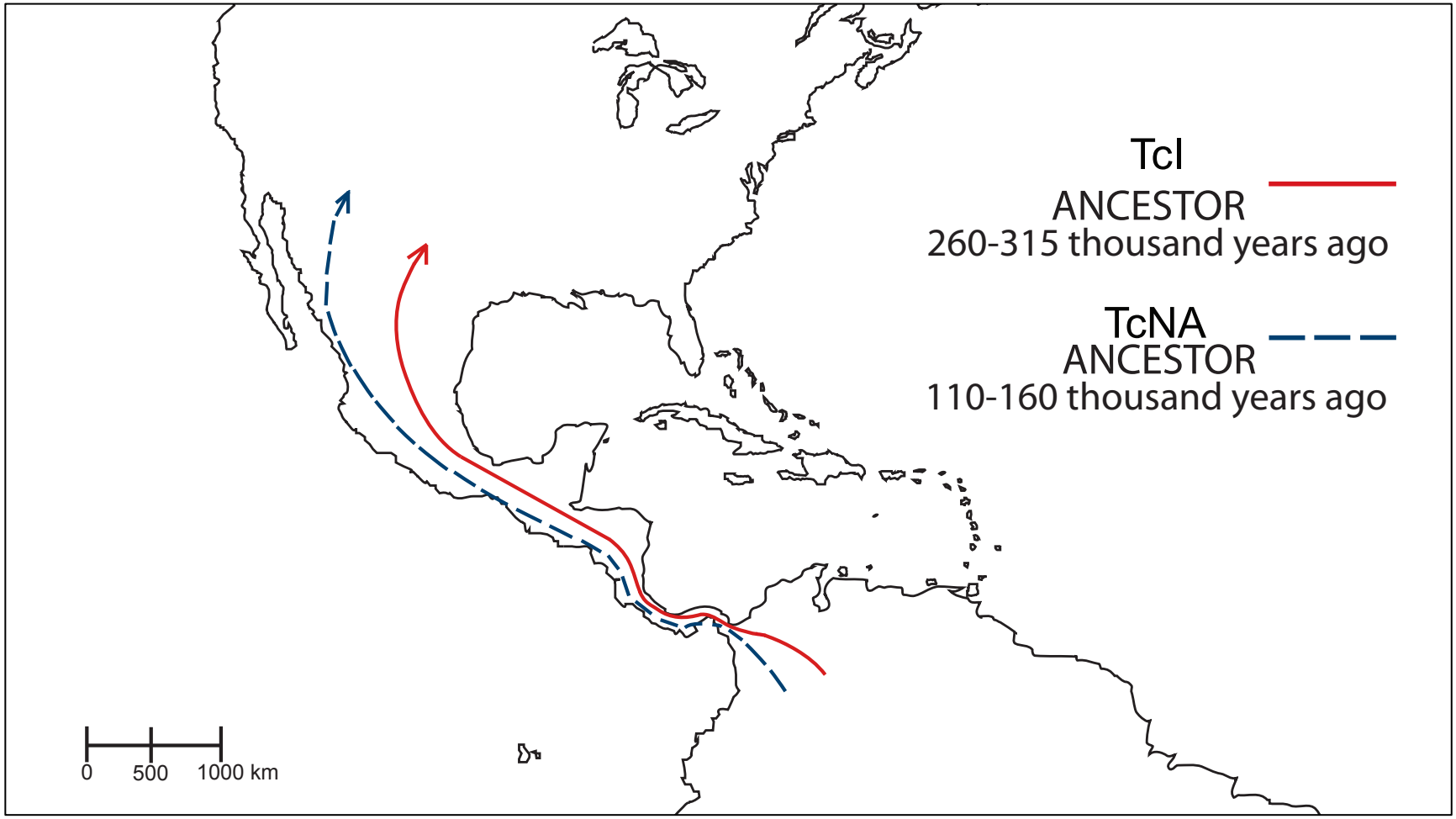


Figure 3-4. Ancestral TcI and TcNA entries from South America into North America. Bayesian time divergence estimates can be seen in Table 3-2.



Chapter 4

Positive selection has played a larger role in the evolution of *Trypanosoma cruzi* proteins than in the evolution of *Leishmania spp*. proteins

This chapter is being reviewed in PLoS NTD as of July 2013

ABSTRACT

Chagas disease and Leishmaniasis are neglected tropical diseases that together affect more than 30 million people worldwide. They are caused by two different groups of parasites: *Trypanosoma cruzi* and *Leishmania spp*., respectively. Currently there is neither a vaccine nor adequate drugs to control these human diseases. With the amount of genomic sequence data available on the rise it has become feasible to use comparative evolutionary tools to identify pathogen proteins that have experienced natural selection and have characteristics that are desirable for vaccine development. We show that natural selection has played a larger role in the evolution of *T. cruzi* proteins than in the evolution of *Leishmania spp*. proteins. We propose that this difference is the result of the greater versatility of *T. cruzi* in terms of mammal species it can infect, its cell tropism and its intracellular invasion mechanisms. We provide a list of proteins that could be tested for vaccine potential in both human diseases. Identification of a protein that was recently proposed as a Leishmaniasis vaccine using our evolutionary analyses confirms the importance that comparative evolutionary studies can have on the exploration of vaccine candidates.

INTRODUCTION

Natural selection is the most important evolutionary force driving the diversification of all living organisms. Comparative and population genetic analyses of orthologous DNA sequences are routinely used for inferring whether natural selection has shaped levels and patterns of intra and interspecific nucleotide divergence and diversity in natural populations (Fay 2011; Nielsen 2005; Oleksyk, et al. 2010). In comparative studies of orthologous protein-coding sequences, the action of natural selection is inferred from the value of the ratio of non-synonymous (dN) to synonymous (dS) substitutions (dN/dS or ω). Proteins or sections of proteins under purifying

(negative) selection (i.e. selectively constrained) have dN/dS < 1, while proteins or sections of proteins that have experienced positive selection have dN/dS > 1. However, as usually only a small proportion of codon sites are evolving under positive selection, averaging dN/dS over an entire protein is a very conservative approach for inferring natural selection. Consequently, methods that test for codon sites evolving under positive selection are more powerful and accurate (Swanson, et al. 2003; Yang and Swanson 2002).

Immune system elicitors, or antigens from pathogens that evolve rapidly to avoid recognition from the host immune system, constitute good examples of protein evolution driven by natural selection. Bioinformatic and evolutionary analyses focused on the use of dN/dS ratios have become powerful methods for identifying proteins evolving under positive selection driven by the arms race between pathogens and their host's immune defenses (Gu, et al. 2011; McCann, et al. 2012; Pacheco, et al. 2012; Petersen, et al. 2007; Soyer, et al. 2009; Xu, et al. 2011; Zhang, et al. 2011). Protein evolution analyses have thus become useful approaches for identifying candidate genes for vaccine development.

Chagas disease and Leishmaniasis are neglected tropical diseases that currently do not have any vaccine (Hotez, et al. 2007; Machado and Ayala 2002), making them excellent candidates for conducting *in silico* searches for protein-coding genes that can be targeted for vaccine or drug development. *Trypanosoma cruzi* is the etiological agent of Chagas Disease; it infects approximately 7.7 million people, and kills about 10 thousand people every year in Latin America (Rassi, et al. 2010; WHO 2011). Leishmaniasis, on the other hand, is caused by more than 20 species of the genus *Leishmania* and has a much wider geographic range, occurring in 88 different countries from Latin America, Asia, Africa and Europe. It is estimated that twelve million people are infected with *Leishmania*, with approximately 20,000-40,000 deaths per year

(Alvar, et al. 2012; Desjeux 2001). Both parasites are protozoans that belong to the class Kinetoplastea, an early-diverged branch in the eukaryotic tree of life (Simpson, et al. 2006). Members of this eukaryotic class have unique characteristics not found in other eukaryotes (Schmidt 2005) like post-transcriptional gene regulation (poli-cistronic mRNA modification and uracil insertion modification of mRNA), and the presence of an enlarged mitochondria (i.e. kinetoplast) with a unique chromosomal architecture composed of a few megacircles and thousands of concatenated mini circles.

The taxonomic classification of kinetoplastids has been a matter of debate due to the clonal reproductive nature of some of the taxa, and because the classification of the genetic diversity found within each taxonomic group has been treated differently (Kjos, et al. 2009; Rassi, et al. 2010). *Trypanosoma cruzi* has all its genetic diversity classified into strains rather than species like in *Leishmania* (see below) even though major lineages of this parasite are independent genetic lineages due predominant clonal reproduction (Kjos, et al. 2009), and even though nucleotide sequence divergence and estimated divergence times among *T. cruzi* strains can be as high as 5.4% (El-Sayed, et al. 2005a; Schmidt 2005) and 2.18 million years (Flores-Lopez and Machado 2011), respectively. On the other hand, the genus *Leishmania* has been divided in at least 30 different species. Despite that difference in taxonomic classification, a comparison between *Leishmania* species and *T. cruzi* strains is relevant not only because clonally reproducing *T. cruzi* strains represent independent genetic lineages, akin to genetically isolated *Leishmania* species, but also because there is overlap in genetic divergence levels among *T. cruzi* strains and among some *Leishmania* species (see below). Furthermore, there are many shared and yet distinct characteristics between the two groups that may have influenced their patterns of evolution. For example, both parasites have evolved to be digenetic, and both are thought to have

independently evolved intracellular evasion mechanisms within their mammal hosts (Sibley 2011). Further, both groups of parasites are believed to have had their evolutionary origins in South America (Lukes, et al. 2007; Yeo, et al. 2005; Zingales, et al. 2012), yet they now have different geographic ranges and have adapted to very different insect vectors and mammal hosts. Sylvatic *Leishmania* has been found mostly in rodents, dogs, foxes and jackals, whereas *T. cruzi* has a much wider range of mammalian hosts (> 100 hosts), including opossums, armadillos, raccoons, rodents, bats, dogs, primates and humans (Schmidt 2005). However, one of the most dramatic differences between the two taxa is their contrasting cell/tissue tropism. *Leishmania* strictly invade macrophages in the vertebrate host although neutrophiles have also been shown to play a role in the early invasion process, whereas *T. cruzi* has been found to be able to invade a much wider array of cell types (e.g. myocytes, macrophages, cardiomyocytes, nerve cells, etc.) (Fernandes and Andrews 2012; Moradin and Descoteaux 2012). In fact, *T. cruzi* appears to be able to invade any type of cell *in vitro* (Manso-Alves 2009).

Whether the major differences in mammalian host diversity, cell/tissue tropism and possibly cell invasion mechanisms, have had any impact on patterns of genetic diversification of immune system elicitors, antigens and proteins targeting host surface cell receptors in the two pathogens, has yet to be explored. Here we present evolutionary analyses of protein-coding genes from 4 strains of *T. cruzi* and 4 species of *Leishmania* whose genomes have been sequenced and annotated. We present results showing significant differences between the two taxa in the proportion of genes showing evidence of positive selection, and in the intensity of selection and proportion of sites that have experienced positive selection in those genes. Genes showing evidence of positive selection within each taxonomic group may be under diversifying

selection to evade the immune system and thus, depending on their functions, could represent viable candidates for the development of vaccines.

MATERIALS AND METHODS

Sequence data sets. We analyzed all available annotated genome sequences from *Trypanosoma cruzi* and *Leishmania spp*. Four *T. cruzi* have been sequenced, representing 4 of the 6 distinct typing units (DTUs) or lineages (TcI-TcVI) in to which the genetic diversity of this parasite is currently divided (Zingales, et al. 2009). The four sequenced genomes include the most divergent set of lineages of this species and thus provide a good representation of genetic divergence within *T. cruzi*. Two of the sequenced genomes (haplotypes Esmeraldo and "Non-Esmeraldo") were obtained during the sequencing of the genome strain of *T. cruzi* CL Brener (El-Sayed, et al. 2005a). This strain is a hybrid (Brisse, et al. 2003; Machado and Ayala 2001; Machado and Ayala 2002), the result of a hybridization event between two divergent lineages that took place ~0.4-0.8 million years ago (Flores-Lopez and Machado 2011), and thus represents three of the six lineages (DTUs) in which the genetic diversity of the parasite is divided: TcII (Esmeraldo), TcIII ("Non-Esmeraldo") and TcVI (the hybrid lineage). Two additional genome sequences from lineage TcI have been recently obtained (Sylvio and JRcl4) (Franzen, et al. 2011). The number of annotated protein coding genes in each sequenced genome is: 10,834 (Non-Esmeraldo), 10,342 (Esmeraldo), 7,755 (JRcl4) and 7,456 (Sylvio).

The genetic diversity of *Leishmania spp.* has been divided into more than 30 species (Schmidt 2005) of which 5 associated with humans have had their genomes sequenced: *Leishmania major*, responsible for cutaneous leishmaniasis (CL) in the Old World, *L. mexicana*

and *L. braziliensis*, both of which cause CL in the New World, and *L. infantum* and *L. donovani* which cause visceral leishmaniasis (VL). *Leishmania major* (strain Friedlin) was the first *Leishmania* species sequenced (Ivens, et al. 2005), followed by *L. infantum* and *L. braziliensis* (Peacock, et al. 2007). Annotated genome sequences of *L. mexicana* are not published but are available in TriTrypDB release 4.0 (http://tritrypdb.org/tritrypdb/). The annotated genome sequence of *L. donovani* is available in GeneDB (http://www.genedb.org/). However, this species has very low genetic divergence with *L. infantum* (~0.48% average nucleotide divergence) and for that reason it was not included in this study. The number of annotated protein coding genes in each of the 4 *Leishmania* genomes we included is the following: *L. major* (8,408), *L. infantum* (8,241), *L. braziliensis* (8,357) and *L. mexicana* (8,250).

Ortholog data sets. Reciprocal best hit blastx searches were conducted to find true orthologs within each taxonomic group. A blastx E-value of 10^{-5} was used as threshold for the orthologous search similarity criteria. The approach to identify orthologs using reciprocal best hit blastx searches is conservative due to its low false positive rate and medium false negative rate (Chen, et al. 2007). Within each taxonomic group we conducted reciprocal best-hit blastx searches for all possible pairwise strain comparisons, and selected ortholog pairs that matched across all pairwise comparisons. This conservative approach filtered out almost all proteins that form part of the largest protein families found in *T. cruzi* (i.e. trans-sialidases, mucins, MASP, surface glycoprotein gp63 protease) due to large sequence similarities among protein members of these large families.

The final datasets of putative orthologs consisted of 5,146 protein-coding genes in *Trypanosoma cruzi* and 7,439 protein-coding genes in *Leishmania spp*. *T. cruzi* strains were highly dissimilar in terms of the number of predicted protein-coding genes in each genome

compared to *Leishmania spp.*. The CL-Brener haplotypes had an average of 10,588 protein-coding genes per haplotype, while the TcI strains (Sylvio & JRcl4) had an average of 7,605 protein-coding genes per strain. The first *T. cruzi* genome sequenced study noted that approximately 50% of the predicted protein-coding genes were members of few very large protein families (El-Sayed, et al. 2005a). The significantly smaller number of protein-coding genes predicted in the TcI strains is mostly due to differences in the copy number of these large protein families (mostly trans-sialidases, mucins, MASP and gp63s protein families) (Franzen, et al. 2011). In contrast to *T. cruzi*, the number of predicted protein-coding genes in the genomes of *Leishmania* was very similar across species. The largest difference in number of protein-coding genes annotated among *Leishmania* species was 167, compared to approximately 3,000 among *T. cruzi* strains. The average *Leishmania* genome contained 8,314 predicted protein coding-genes, from which an ortholog data set of 7,439 protein coding-genes was constructed.

Alignment and selection analyses. Sequences from each ortholog data set were translated to amino acids with translatorx3.pl (Abascal, et al. 2010) and aligned with MUSCLE (Edgar 2004). Poorly aligned regions were removed using Gblocks (Castresana 2000). Removing poorly aligned regions is a conservative approach that has recently been shown to outweigh the costs of removing true positively selected sites from the analyses (Privman, et al. 2011). Thus, the true number of positively selected sites and/or proteins in our data set might actually be larger than what is presented here. A concatenated data set of ~1.75 million base pairs of aligned protein-coding sequence for each taxa was used to reconstruct phylogenetic relationships among the four *T. cruzi* strains and among the four *Leishmania* species using Maximum likelihood in PAUP (Swofford 1998) (Figure 4-1). The DNA substitution model for the phylogenetic reconstructions was selected for each taxonomic data set using jModeltest (Posada 2008). The phylogenies were

then input to PAML for conducting the positive selection analyses. Aligned data sets were back translated into nucleotides for the selection analysis in PAML (Yang 2007). Pairs of nested models, M7 (beta) versus M8 (beta & w) and M8 versus M8a (beta & ω=1), were compared using a likelihood ratio test (LRT). Significance of the LRT for M7 vs M8 was determined using 2 degrees of freedom. Since model M8a is not completely nested within M8, significance was determined by halving the p value from a chi-square test with one degree of freedom as previously suggested (Yang 2007).

Identification of putative false positives. To determine whether the number of taxa or levels of sequence divergence had a major influence in the observed trends in our results, the genomes of *L. braziliensis* and the JRcl4 strain of *T. cruzi* were excluded from the analyses. Those samples represent the most divergent *Leishmania* genome and one of the two least divergent *T. cruzi* genomes (Figure 4-1), and were removed to reduce the difference in the average levels of nucleotide divergence found between data sets (uncorrected p distance range in the 3 taxa data set: *Leishmania spp.* (0.053-0.078), *T. cruzi* (0.018-0.022); uncorrected p distance range in the 4 taxa data set: *Leishmania spp.* (0.053-0.175), *T. cruzi* (0.003-0.022))(Tables S4-5, S4-6). Proteins that showed evidence of positive selection in both the 3 and 4 taxa analyses were identified as true positives (TP), while proteins that appeared to be under positive selection in the 4 taxa dataset but not in the 3 taxa dataset were identified as false positives (FP) (Table S4-8). Statistical comparisons between TP and FP were performed to analyze potential biases leading to false positives. Tree length, proportion of sites predicted to be under positive selection and dS were compared.

Functional overrepresentation analyses. Functional annotations associated with *L. major* and the *T. cruzi* Non-Esmeraldo haplotype were used to determine if any molecular or biological

functions were overrepresented among genes showing evidence of positive selection. Both a comprehensive analysis (integrating all proteins with evidence of positive selection) and a conservative analysis (only using TP) were performed (Table 4-1 & Table S4-9). Unfortunately, between 50% and 68% of the protein coding genes in *T. cruzi* and *Leishmania majo*r have unknown functions based on genome annotations, and only 10% (~500 proteins) of *T. cruzi* proteins and 42% (3525 proteins) in *L. major* have a Gene Ontology term associated with them (www.genedb.org). Consequently, to conduct a more comprehensive functional over representation analysis, we additionally clustered all our protein data into Pfam clans (pfam.sanger.ac.uk/), a broader classification scheme that groups evolutionary related protein families into clans based on related structure, related function, and significant matching of the same sequence to databases of hidden Markov Models (HMMs) from different families and profile-profile comparisons (Finn, et al. 2006). Clustering our data set into Pfam clans allowed us to include many more proteins into the functional overrepresentation analyses. Statistically overrepresented Pfam clans were identified with GeneMerge (Castillo-Davis and Hartl 2003). All statistical overrepresentation tests were conducted with the protein lists predicted to be under positive selection from the M8 versus M8a model comparison.

Hypothetical clustering. Genes with unknown function (i.e. hypothetical genes) were grouped into protein clusters by conducting a self blastp search (E value < 10^{-10}) of all the hypothetical proteins from the Non-Esmeraldo genome or from the *L. major* genome. A numerical code (e.g. protein family 1, 2, 3, etc.) was given to all clusters, including proteins from clusters of size 1. GeneMerge was used to determine statistical overrepresentation of protein clusters that had evidence of sites under positive selection (based on the comparison of models M8 and M8a).

RESULTS

More evidence of positive selection in *Trypanosoma cruzi* than in *Leishmania spp.* Figure 4-1 shows the phylogenetic relationships among the *Leishmania* species and *T. cruzi* strains that were used in the selection analyses (See Methods). In *Leishmania spp.* our analysis identified 78 and 170 genes with evidence of positive selection using the M8 vs M8a or M7 vs M8 model tests respectively, representing 1.0% or 2.3% of the 7,439 ortholog data set from *Leishmania* (Table 4-1, $p<0.01$, Tables S4-1,S4-2). In contrast, in *T. cruzi,* a total of 614 and 628 protein-coding genes presented evidence of positive selection using the M8 vs M8a or M7 vs M8 model tests respectively (Table 4-1, $p<0.01$, Tables S4-3,S4-4), representing 11.9% or 12.2% of the 5,146 *T. cruzi* genes used in this study. The number of protein-coding genes showing evidence of positive selection is significantly higher in *T. cruzi* than in *Leishmania spp.* (M8 vs M8a: $\chi^2 = 693.33$ with Yates correction, $p = 8.43 \times 10^{-153}$; M7 vs M8: $\chi^2 = 503.86$ with Yates correction, $p = 1.37 \times 10^{-111}$) (Figure 4-2). Furthermore, the average dN/dS ratio in sites evolving under positive selection in proteins showing significant evidence of positive selection (Figure 4-3A) and the proportion of codon sites under selection in those proteins (Figure 4-3B) are significantly higher in *T. cruzi* than in *Leishmania spp* (Wilcoxon rank-sum test, $p<0.0001$).

To examine the effect of removing highly divergent codons from the alignment we performed the *T. cruzi* ortholog alignments without alignment editing by Gblocks (Castresana 2000), which removes highly divergent aligned regions. It is important to note that recent analyses (Privman, et al. 2011) show that the benefits of using aligner filters in studies of this nature outweigh the costs of removing true positively selected sites from the analyses. After using Gblocks the number of genes predicted to be under positive selection slightly increased from 628 to 796 (15.4%, M7 vs M8) and from 614 to 735 (14.2%, M8 vs M8a) in *T. cruzi*,

confirming that the reported results are conservative. Similar results were observed in *Leishmania spp.*, and thus the number of proteins predicted to be under positive selection in *T. cruzi* remained larger than those predicted in *Leishmania spp* (not shown).

Low divergence at synonymous sites (dS) could generate false positives with high dN/dS. However, *T. cruzi* proteins under positive selection had a significantly higher divergence at synonymous sites compared to proteins under purifying selection (t=2.19, p=0.0144), whereas the *Leishmania spp.* proteins under positive selection did not show any significant difference with those under purifying selection (t=0.29, p=0.38). Furthermore, the distribution of dN/dS values for all proteins when comparing the most divergent lineages of *Leishmania spp*. (*L. braziliensis vs L. mexicana*) or *T. cruzi* (Esmeraldo vs Non-Esmeraldo) is significantly different between the two taxa (Wilcoxon rank-sum text, $p < 0.0001$) (Figure 4-4). Therefore, it is unlikely that the results presented here are the result of artifacts generated by low divergence at synonymous sites in the *T. cruzi* dataset.

Estimating positive selection using 3 or 4 taxa is close to the minimum number of taxa recommended in order for the positive selection tests to have any power (Anisimova, et al. 2001; Wong, et al. 2004), although comparisons using 2 taxa have been regularly conducted (e.g. humans vs chimps (Nielsen, et al. 2005)). For that reason, and to reduce the difference in the average levels of nucleotide divergence found between data sets (see Methods and Tables S4-5,S4-6), we compared results of analyses conducted with 4 and 3 taxa (See Methods). After removing the genome of *L. braziliensis*, 4.3% (320 genes, models M8 vs M8a) and 4.81% (358 genes, models M7 vs M8) of the *Leishmania spp.* genes presented evidence of positive selection in the 7,439 ortholog dataset. After removing the JRcl4 strain of *T. cruzi*, 10.37% (534 genes, models M8 vs M8a) and 10.51% (541 genes, models M7 vs M8) of the genes showed evidence

of positive selection in the 5,146 ortholog *T. cruzi* dataset. Although the number of genes showing evidence of positive selection in the *Leishmania spp.* and *T. cruzi* 3 taxa datasets increased and decreased, respectively, the number of genes with evidence of positive selection was still significantly higher in *T. cruzi* (M8 vs M8a: $\chi^2 = 177.58$ with Yates correction, $p = 1.64 \times 10^{-40}$; M7 vs M8: $\chi^2 = 177.26$ with Yates correction, $p = 1.92 \times 10^{-40}$) (Table S4-7).

There were 357 "True Positives" (TP) and 257 "False Positives" (FP) in *T. cruzi* comparing M8 vs M8a ($p<0.01$, see Methods), and 25 TP and 53 FP in *Leishmania spp.* (Table S4-8). The number of TP is significantly higher in *T. cruzi* than in *Leishmania spp.* ($p= 1.59 \times 10^{-98}$). To analyze characteristics that TP could have in comparison to FP we compared tree length, the proportion of sites predicted to be under positive selection and the divergence at synonymous sites (dS) of the two most divergent taxa for each ortholog data set. The mean dS of TP and FP were not significantly different for either taxa. However the average tree length averages were significantly smaller in FP for *T. cruzi* ($t=2.66$, $p=0.0039$) and significantly higher in FP for *Leishmania* ($t=3.83$, $p<0.0001$). Further, in *T. cruzi* the proportion of sites predicted to be under selection in FP was significantly lower than that in TP ($t=4.27$, $p<0.0001$). In *Leishmania spp.* the opposite was found, FP had a significantly larger number of sites under selection than TP ($t=2.78$, p:0.003).

Functional overrepresentation of genes under positive selection. In *Leishmania spp.*, 5 functional categories were statistically overrepresented (Table 4-2). The most significant was glutathione peroxidase, followed by ATP binding cassette, iron superoxide dismutase, cysteine peptidase and zinc related functions (transporter and finger domain). Interestingly, proteins with those functions have been shown to play some role in the evolution of drug resistance (ATP binding cassette), in the interaction with the host immune system (cysteine peptidase), or have

been proposed as vaccine candidates for Leishmaniasis (iron superoxidase dismutase) (Daifalla, et al. 2012) (see Discussion). However, when only TP are included in the analyses glutathione peroxidase and zinc related functions are no longer significant (Table S4-9).

In *T. cruzi,* genes with mucin function were the most statistically over represented among those showing evidence of positive selection (Table 4-2). This is an important finding given the important role played by mucins in the interaction between *T. cruzi* and the host immune system (Buscaglia, et al. 2006) and because our strict orthologous data set filtering only captured 15 mucin orthologs out of the 2000 mucin associated proteins in the *T. cruzi* genome (El-Sayed, et al. 2005a). Amino acid permease was the second most statistically overrepresented function. These proteins could be interesting targets for vaccine development given their function: they are cell membrane proteins involved in the transport of amino acids into the cell. Third on the list is a group of splicing factors involved in mRNA editing. However, none of these three functions remained significantly overrepresented when only TP were included in the analyses.

The genes showing evidence of selection in either taxa are quite divergent between *T. cruzi* and *Leishmania spp*. Of the 170 genes predicted to be under positive selection in *Leishmania spp.* (M7 vs M8, $p<0.01$), 148 have at least one significant hit (E value $< 10^{-5}$) in the Non-Esmeraldo *T. cruzi* strain, but only 14 of those genes show at least 70% amino acid sequence identity between Non-Esmeraldo *T. cruzi* and *L. major*. Of the 628 proteins in *T. cruzi* predicted to be under positive selection (M7 vs M8, $p<0.01$), 505 had at least another significant blast hit in the *L. major* genome, but only 28 of these protein-coding genes show at least 70% amino acid sequence identity between Non-Esmeraldo *T. cruzi* and *L. major*.

Hypothetical genes under positive selection. Although the ortholog dataset of *T. cruzi* had a smaller number of hypothetical proteins than *Leishmania spp.* (3737 vs 5054), a significantly higher proportion of hypothetical proteins were predicted to be under positive selection in *T. cruzi* (M8 vs M8a: 572 vs 140, $\chi^2 = 8.87$, $p=2.89 \times 10^{-3}$; M7 vs M8: 585 vs 242, $\chi^2 = 24.72$, $p=6.6 \times 10^{-7}$). In an attempt to increase the power of the analysis for proteins with unknown function, all those proteins were clustered based on sequence similarity. Of the 3737 hypothetical proteins within the *T. cruzi* ortholog data set 461 clusters were formed after the sequence similarity cluster criteria (E value < 10^{-10}), whereas in the *Leishmania spp.* ortholog data set 830 clusters were formed. These results show, as expected, that most hypothetical proteins are not part of protein families, reducing the power of functional overrepresentation analyses. Of the clusters of hypothetical proteins recovered, only 3 clusters in *T. cruzi* and 2 clusters in *Leishmania spp.* were overrepresented among those under positive selection (Table 4-3).

DISCUSSION

The availability of annotated genome sequences from human pathogens has made feasible the application of bioinformatic methods in the identification of proteins with immunogenic properties that could become candidates for vaccine development. Recent studies have provided solid evidence that candidates for vaccine development can be identified by surveying parasite genomes for proteins in which a few amino acid sites have experienced high rates of amino acid substitution (consistent with the action of positive selection) while the rest of the protein is under strong purifying selection (Gu, et al. 2011; McCann, et al. 2012). The rationale behind this idea is that proteins with those characteristics have regions that are rapidly evolving because of their recognition by the host's immune system, but also have conserved regions under strong negative

selection that may have a critical role in the biology of the pathogen. Effective vaccines of long-term effectiveness would target those conserved regions rather than the rapidly changing regions of those proteins (Burton, et al. 2012).

We show that positive selection has played a larger role in the evolution of *T. cruzi* proteins than in the evolution of *Leishmania spp*. proteins. First, we report that a significantly larger fraction of protein-coding genes show evidence of positive selection in *Trypanosoma cruzi* than in *Leishmania spp* (Figure 4-2). Furthermore, we report that the average dN/dS of sites under positive selection and the proportion of sites under positive selection in genes showing significant signal of positive selection are significantly higher in *T. cruzi* than in *Leishmania spp.* (Figure 4-3). We show that those results are not due to artifacts of the analyses after evaluating the effects on the overall results of maintaining or removing highly divergent regions using GBlocks, of the level of sequence divergence at silent sites, and of changing the number of taxa analyzed. Although the overall level of divergence was smaller among *T. cruzi* strains (p-distance: 0.003-0.022 (4 taxa), 0.018-0.022 (3 taxa)) than among *Leishmania* species (p-distance: 0.053-0.175 (4 taxa), 0.053-0.078 (3 taxa)), a similar comparative genomic study in primates observed the opposite trend that we report: significantly fewer proteins were predicted to be under positive selection in the smaller data set (3 vs 7 species) with the less divergent sample of primate species (George, et al. 2011).

We propose that the larger number of hosts mammal species, the larger number of target cells and tissues it can infect, and the more diverse intracellular invasive developmental stages of *Trypanosoma cruzi* are the underlying reasons behind the observed difference in the number of proteins under positive selection observed between the two taxa. These major differences in the biology of the two groups of parasites likely influence the level of interaction between the

parasite, surface receptors of host target cells and/or the immune system of their different hosts, and can therefore influence patterns of protein evolution in immune elicitors of the parasite and in proteins involved in cell invasion. Species from the genus *Leishmania* are known to depend almost exclusively on macrophages for their intracellular survival mechanism within a relatively small number of mammal hosts (Schmidt 2005). Although there are some lizard-infecting species of *Leishmania* that live in the lumen of the cloacae, the intestine or in the bloodstream of lizards and do not infect macrophages, none of these species were included in this study, and there has been a debate among taxonomists about the placement of these species within the genus *Leishmania* or the subgenus Sauroleishmania (Noyes, et al. 1998). On the other hand *T. cruzi* has the ability to invade any type of cell in humans during the initial phases of infection, even though the parasite does tend to have tropism towards muscle and nerve cells (Schmidt 2005). The pathology of *T. cruzi* in its estimated 180 mammal reservoir species is not well known (2002; Noireau, et al. 2009). Currently, the dogma is that pathologies associated with Chagas disease are unique to humans and that other mammals are immune to the parasite. However, the lack of studies on the subject makes such a dogma inconclusive to say the least. It is unlikely that the capacity to invade multiple cell types is a characteristic unique to the interaction between *T. cruzi* and humans given the vast number and diversity of mammalian species *T. cruzi* is known to infect. The wider range of host cells and host species that *T. cruzi* can infect, combined with its more complex life cycle that exposes the parasite to more diverse intra- and extracellular environments, has exposed a larger number of its proteins to selective pressures during its evolution than it has in *Leishmania*. Proteins directly exposed to the immune system as well as proteins that directly interact with surface receptors of the wide range of cell types *T. cruzi* can invade should have been exposed to different levels of selection. Therefore, we

consider that the higher versatility of *T. cruzi* is the most likely reason for the significant differences in the fraction of proteins under positive selection (Figure 4-2) and the more intense levels of selection in those proteins (Figure 4-3) between the two taxa. We stress the fact that selection pressures have been exerted not just by humans but mostly by non-human hosts of these parasites given the short time of association of these parasites with humans.

Our study is the first to use comparative evolutionary analyses to generate a preliminary list of potentially immunogenic proteins that could have the potential of being adequate immune elicitors for Leishmaniasis and Chagas disease (Tables S4-1, S4-2, S4-3, S4-4). In *Leishmania spp.* the most interesting overrepresented function with genes under positive selection was iron superoxidase dismutase, since one gene from this protein family was recently proposed as a vaccine candidate for Leishmaniasis due to the protective role it induced in mice (Daifalla, et al. 2012). That result further reinforces the point that evolutionary approaches can play a major role in detecting immunogenic molecules. In addition, we found four additional overrepresented functions, some of which have been shown to play some role in the evolution of drug resistance (ATP binding cassette) or in the interaction with the host immune system (cysteine peptidase). ATP binding cassette was the second most significant overrepresented function. Members of this large protein family are known to be involved in the development of amphotericin-B drug resistance in *L. donovani* (Purkait, et al. 2012). This antifungal compound (amphotericin-B) is the main drug therapy employed by the WHO to treat visceral leishmaniasis. It would be of particular interest to determine if the sites predicted to be under positive selection in these proteins are directly involved in the development of amphotericin-B drug resistance. Further, cysteine peptidases are known to play a very important role in the manipulation of the host's immune response in *L. mexicana* by controlling the T-helper cell response, which has been

shown to determine the fate of the infection (Alexander and Bryson 2005; Mottram, et al. 2004). Interestingly the phylogenetic branches leading to *L. mexicana* in both ortholog data sets appear to be under positive selection (data not shown).

In *Trypanosoma cruzi* we found that proteins with molecular function mucin, amino acid permease, splicing factor and GTP binding were overrepresented among those showing signals of positive selection. Mucins stand out as logical vaccine candidates for *T. cruzi* since they are located in the cell membrane of the parasite and play very important roles in the mechanism of intracellular invasion (Buscaglia, et al. 2006). However none of these functions appeared to be statistically significant if only true positives were used in the analysis. In fact the only consistent biological function statistically significant after this correction were the proteins of unknown function (i.e. hypothetical proteins), suggesting that more research effort should be put into characterizing those proteins.

Tables

Table 4-1. The number of proteins under positive selection in *T. cruzi* and *Leishmania* spp. *T. cruzi* taxa set consists of Non-Esmeraldo and Esmeraldo haplotypes and Sylvio and Jrcl4 strains. The *Leishmania spp.* taxa set consists of *L. braziliensis, L. mexicana, L. major* and *L. infantum*. The conserved section represents the number of proteins with no evidence of positive selection by the LRT of models M8 vs M8a and M7 vs M8 in PAML using two p-value cuttoffs ($p<0.01$ and $p<0.05$). Note: [a] Numbers based on false discovery rate corrected q-values.

Taxonomic group	***T. cruzi***		***Leishmania spp.***	
Models compared	**Conserved (%)**	**Positive selection (%)**	**Conserved (%)**	**Positive selection (%)**
M8 vs M8a (p<0.05)	4369 (84.9%)	777 (15.1%)	7218 (97.0%)	221 (3.0%)
	4579 (89.0%) [a]	567 (11.0%) [a]	7427 (99.8%) [a]	12 (0.2%) [a]
M8 vs M8a (p<0.01)	4532 (88.1%)	614 (11.9%)	7361 (99.0%)	78 (1.0%)
	4682 (91.0%) [a]	464 (9.0%) [a]	7431 (99.9%) [a]	8 (0.1%) [a]
M7 vs M8 (p<0.05)	4347 (84.5%)	799 (15.5%)	7030 (94.5%)	409 (5.5%)
	4566 (88.7%) [a]	580 (11.3%) [a]	7418 (99.7%) [a]	21 (0.3%) [a]
M7 vs M8 (p<0.01)	4518 (87.8%)	628 (12.2%)	7269 (97.7%)	170 (2.3%)
	4667 (90.7%) [a]	479 (9.3%) [a]	7430 (99.8%) [a]	9 (0.2%) [a]

Table 4-2. Functional over representation of genes showing evidence of positive selection. Note: N: number of proteins of this function in ortholog data set. n: number of proteins of this function under positive selection under model M8 vs M8a. p: statistical significance estimated from GeneMerge. Gene codes are the gene codes for Non-Esmeraldo (*T. cruzi*) and *Leishmania major* found in Tritrypdb.org as of March 2012

Trypanosoma cruzi				
Predicted function	N	n	p	Gene names
Mucin	15	6	0.00924	Tc00.1047053504081.270, Tc00.1047053506615.50, Tc00.1047053506815.20, Tc00.1047053508873.10, Tc00.1047053508999.80, Tc00.1047053511643.50
Amino acid permease	6	3	0.03414	Tc00.1047053506831.20, Tc00.1047053506985.40, Tc00.1047053511649.100
Splicing factor 3a & TSR1	3	2	0.048129	Tc00.1047053507023.230, Tc00.1047053509607.30
GTP binding	7	3	0.0539	Tc00.1047053504105.210, Tc00.1047053507715.40, Tc00.1047053509321.19
Leishmania spp.				
Predicted function	N	n	p	Gene names
Glutathione peroxidase & synthetase	2	2	0.00043037	LmjF.36.3010, LmjF.14.0910
ATP-binding cassette protein subfamily A, D & G	34	5	0.00062779	LmjF.11.1270, LmjF.11.1290, LmjF.27.0970, LmjF.33.1860, LmjF.06.0090
Iron superoxide dismutase	5	2	0.0041293	LmjF.32.1820, LmjF.32.1830
Cysteine peptidase	7	2	0.0084367	LmjF.29.0820, LmjF.19.1420
Zinc finger domain & transporter	16	2	0.042656	LmjF.11.1080, LmjF.31.2390

Table 4-3. Statistically overrepresented hypothetical protein clusters under positive selection in *T. cruzi* and *Leishmania* spp.

Protein cluster	Contributing proteins	Population fraction	Study fraction	p-value	p-value (BF)*	dN/dS	% sites dN/dS>1	Ov. expr	Syntenic	% identity within cluster
*T.c.*1	Tc00.1047053506559.20 Tc00.1047053509039.10	2/7079	2/479	0.0045696	0.39299	25.44 12.64	0.014 0.034	Am, Epi	Chr34, Chr32	27% identity
T.c. 2	Tc00.1047053507031.130 Tc00.1047053509569.140	3/7079	2/479	0.013093	1	6.12 10.10	0.063 0.092	NA, NA	Chr40, Chr12	39.25%
T.c. 3	Tc00.1047053505999.170 Tc00.1047053508299.30 Tc00.1047053511287.4	9/7079	3/479	0.019016	1	30.08 24.15 21.38	0.005 0.116 0.023	NA, NA, NA	Chr9, Chr38, Chr40	48.25%
L. spp. 1	LmjF.09.1020 LmjF.32.0510	3/6031	2/102	0.000840	0.08404	10.849 5.9108	0.00557 0.03476		Chr 9, Chr32	33%
L. spp. 2	LmjF.29.1500 LmjF.11.0670	17/6031	2/102	0.032657	1	6.3739 392.16	0.01263 0.00450		Chr29, Chr11	32.25%

Figures

Figure 4-1. Phylogenies of *Leishmania* species and *T. cruzi* strains used in this study. The unrooted phylogenies used in PAML analyses are shown for A) *Leishmania spp*. B) *T. cruzi*. In each case, the phylogeny was reconstructed using a concatenated data set of 1.75 million base pairs of aligned sequence.

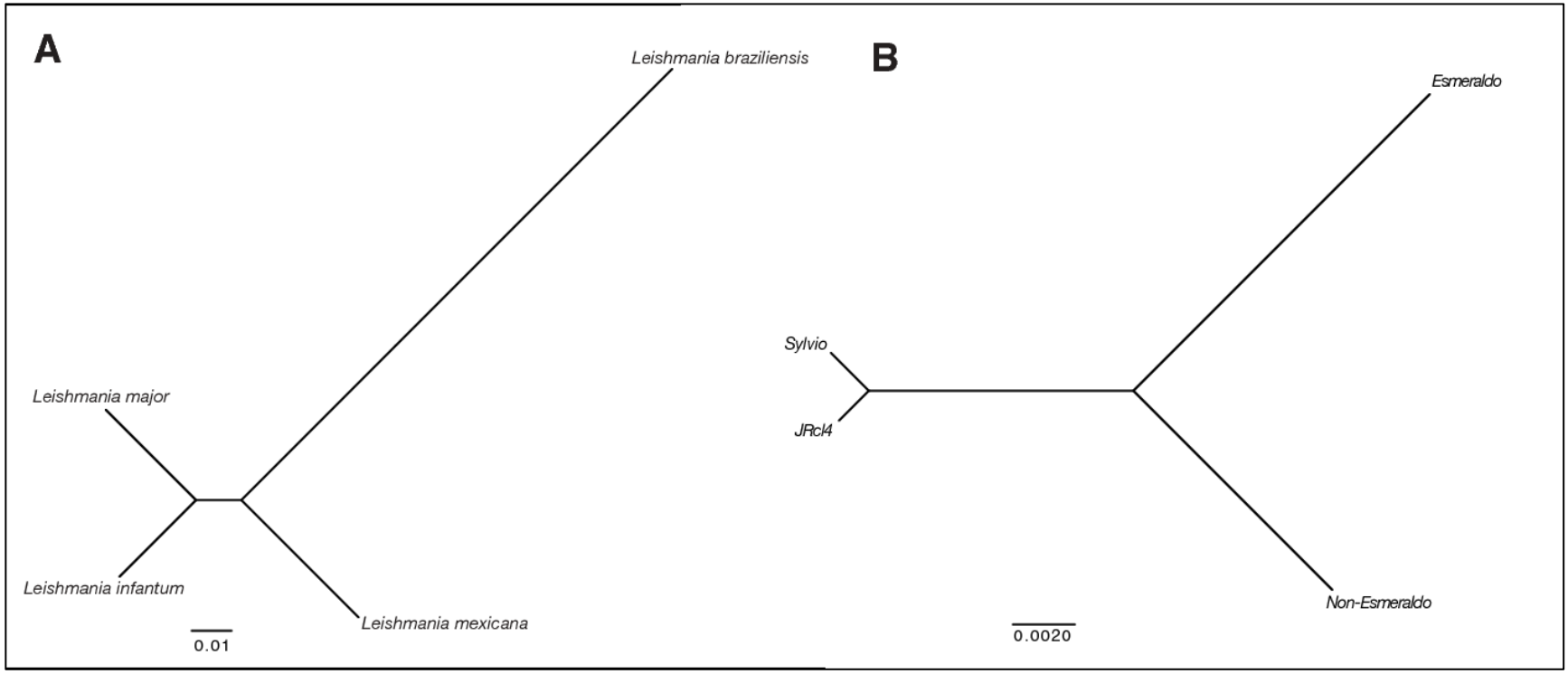


Figure 4-2. Difference in the number of proteins showing evidence of positive selection within each taxa. A) Data for codon models M7 versus M8 ($p<0.01$). B) Data for codon models M8 versus M8a ($p<0.01$). The observed difference is highly significant (M8 vs M8a: $\chi^2 = 693.33$, $p = 8.43 \times 10^{-153}$; M7 vs M8: $\chi^2 = 503.86$, $p = 1.37 \times 10^{-111}$).

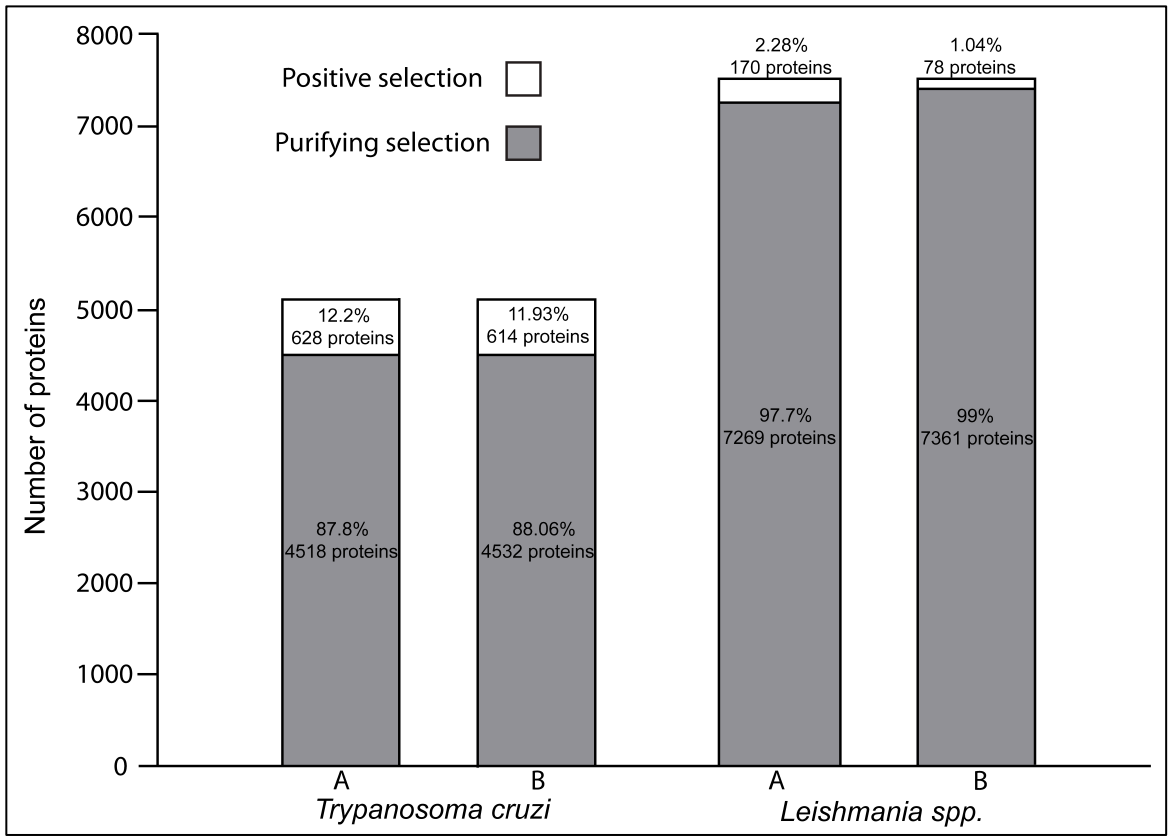


Figure 4-3A. dN/dS for sites predicted to be evolving under positive selection. Boxplot of dN/dS values in sites evolving under positive selection in all the proteins showing significant positive selection signal ($p < 0.05$) using the M8 vs M8a model tests. The two distributions are significantly different (Wilcoxon rank-sum test, $p<0.0001$). dN/dS values higher than 10 were removed from the analysis (difference between species remained highly significant without the removal of outliers $p<0.0001$). Boxes encompass the lower and upper quartiles, with the internal line representing the median and whiskers extending to the 2.5^{th} and 97.5^{th} percentiles.

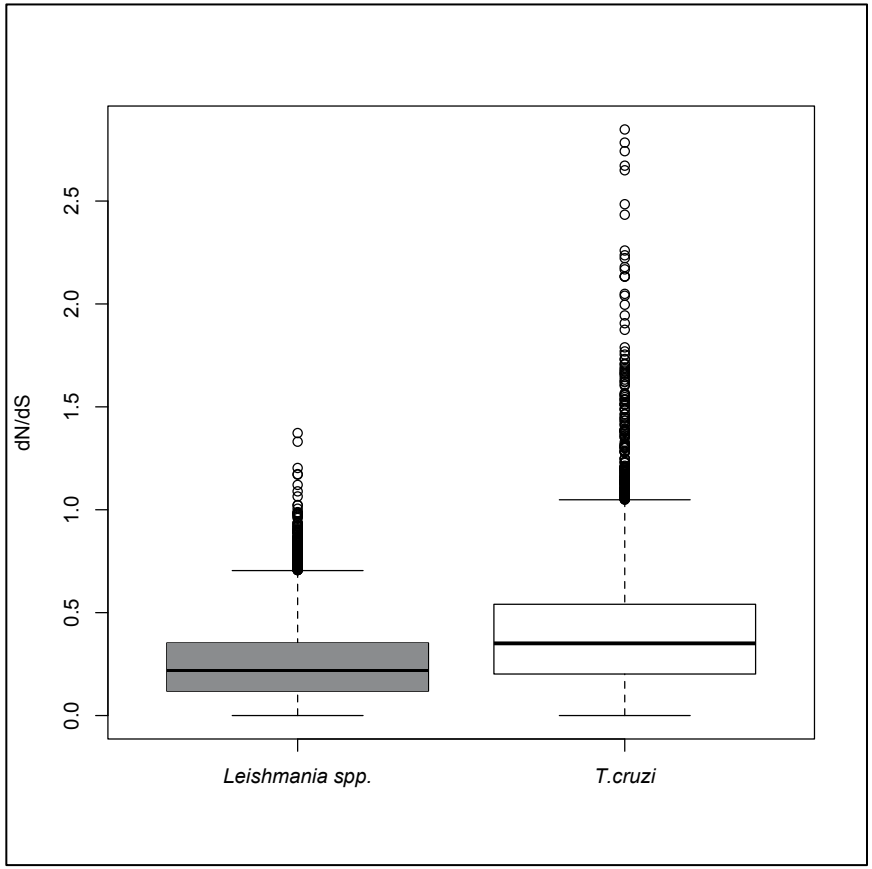


Figure 4-3B. Proportion of sites predicted to be under positive selection. Proportion of sites predicted to be under positive selection for all the proteins that had evidence of being under positive selection ($p < 0.05$) under codon models M8 vs M8a. The two distributions are significantly different (Wilcoxon rank-sum test, $p<0.0001$). Boxes encompass the lower and upper quartiles, with the internal line representing the median and whiskers extending to the 2.5^{th} and 97.5^{th} percentiles.

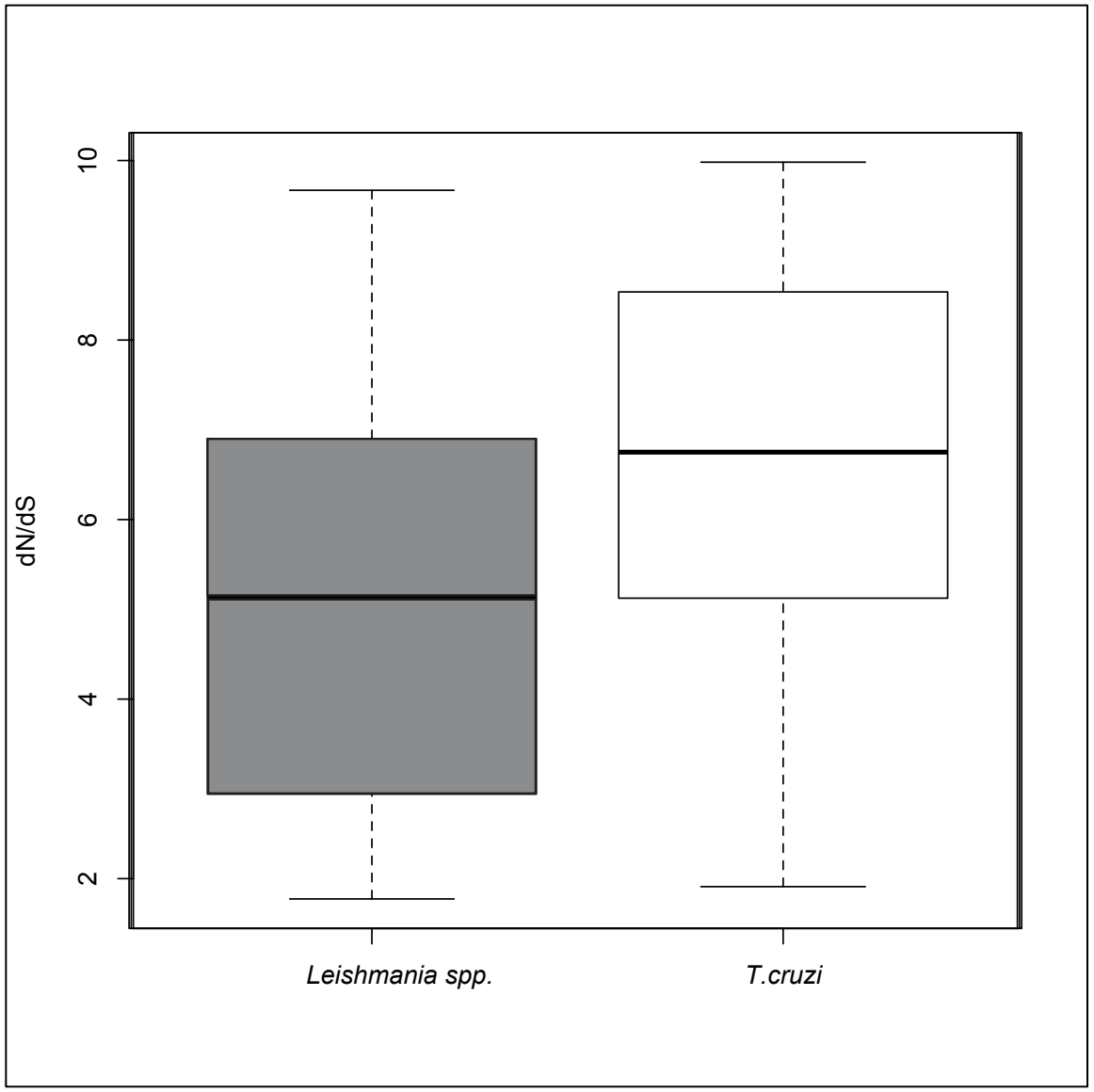
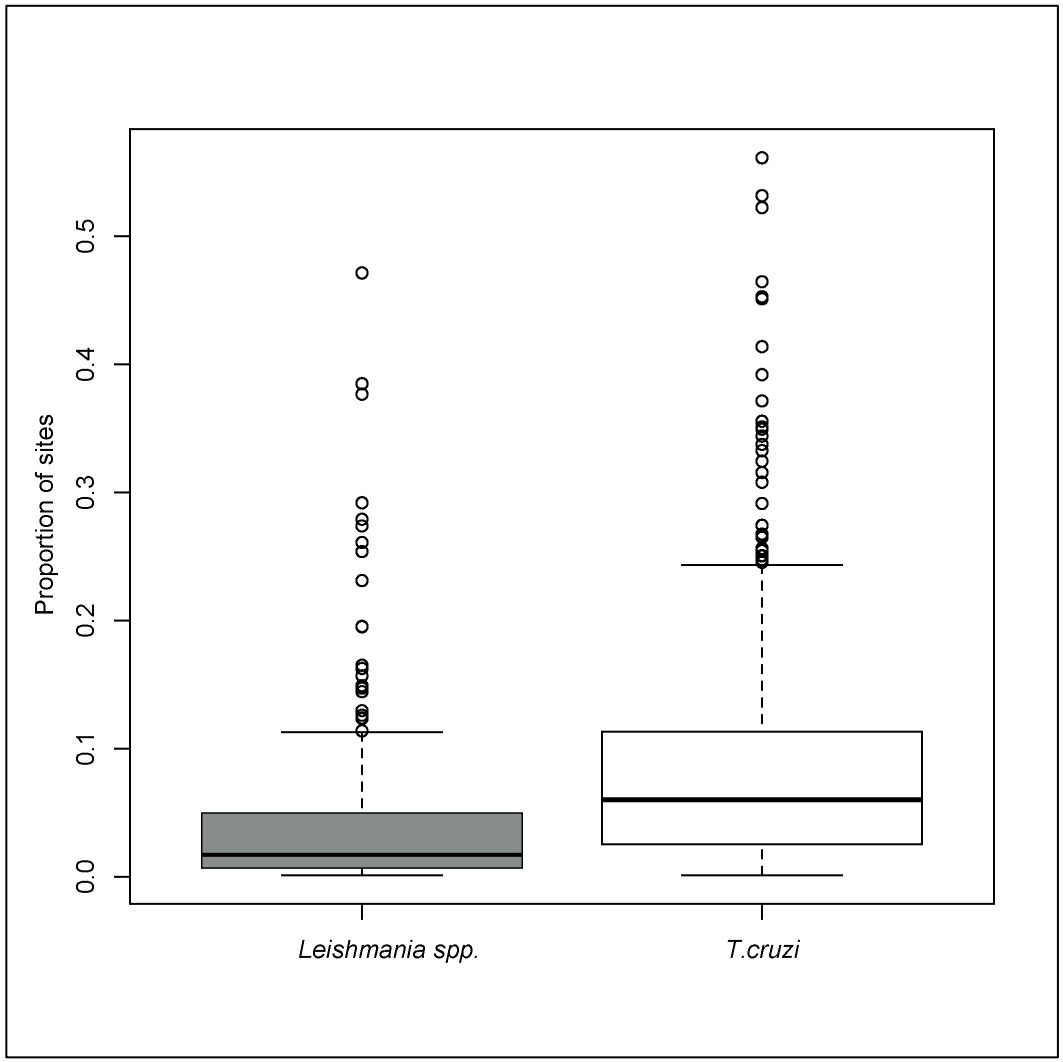


Figure 4-4. Comparison of dN/dS values for all protein coding genes. dN/dS values were estimated between the two most divergent lineages of *Leishmania spp*. (*L. braziliensis vs L. mexicana*) or *T. cruzi* (Esmeraldo vs Non-Esmeraldo), and dN/dS values higher than 3 were removed from the analyses. The two distributions are significantly different between taxa (Wilcoxon rank-sum test, $p < 0.0001$). If outliers with dN/dS > 3 are not removed the difference between species remains highly significant ($p<0.0001$). Boxes encompass the lower and upper quartiles, with the internal line representing the median and whiskers extending to the 2.5^{th} and 97.5^{th} percentiles.



Chapter 5

Genomic changes associated with the evolution of multicellular invasion ability and adaptation to multiple hosts in a pathogen: Protein family expansions in the parasitic lifestyle of *Trypanosoma cruzi*

ABSTRACT

Understanding the genomic events associated with the evolution of major biological innovations in a pathogen, like intracellular parasitism or the ability to invade many cell tissues has only been recently attainable via comparative genomic methods. Genomic sequencing efforts in the major human pathogenic kinetoplastids revealed an extensive expansion of a few surface protein families in *Trypanosoma cruzi* (the causative agent of Chagas disease). The sheer number of paralogs in these surface protein families constitutes a large fraction of all proteins thus far annotated in this parasite's genome, and represents the largest gene family expansions in all kinetoplastid pathogens sequenced thus far. With the use of duplicate age distributions we show that the expansions of those cell surface protein families in *T. cruzi* occurred recently and in rapid, concurrent, bursts. Convergent evolution of intracellular parasitism in another kinetoplastid (*Leishmania spp*.) suggests that major protein family duplication events were not required for the evolution of this trait. However, the time estimates and phylogenetic distribution of these gene family expansions, plus their putative functions, suggest that these massive gene family expansions were linked to the evolution of the parasite's capacity for invading multiple cell tissues (*T. cruzi* can invade any nucleated cell in the mammal hosts) and its adaption to invading multiple host species (*T. cruzi* has been isolated in more than 70 genera of mammals). Time estimates for the massive gene family expansions overlap with the time of evolution of hematophagy in the insect vectors. Given that hematophagy increased the vector's host range, we hypothesize that this event may have indirectly selected for those rapid gene family expansions in the ancestor of *T. cruzi*.

INTRODUCTION

The advent of the genomic era has provided fundamental data to study how a parasitic lifestyle can evolve. In particular, it is now possible to infer what genomic changes take place when: 1) a free-living organism makes the transition to a pathogenic lifestyle, 2) a pathogen evolves the ability to survive inside a cell (intracellular parasitism), 3) a pathogen evolves the capacity to invade different host cells, and/or 4) a pathogen acquires

the ability to invade multiple host species. Most of the genomic studies about pathogen evolution that have been recently completed have focused on addressing the first question, where genome size reduction, gene length reduction, loss of metabolic pathways, and species-specific protein expansions have been the most common traits found amongst a diverse array of pathogens (Glaser, et al. 2001; Heinz, et al. 2012; Nerima, et al. 2010; Razin 1997; Toft and Andersson 2010; Tsai, et al. 2013; Wernegreen 2005). In this study we address the last three questions using genomic data from *Trypanosoma cruzi* (the agent of Chagas disease) and closely related pathogens.

To explore the genomic changes associated with the evolution of several aspects of the evolution of a parasitic lifestyle we focused on *Trypanosoma cruzi* due to several characteristics of this parasite: 1) availability of an annotated genome sequence (El-Sayed, et al. 2005a), 2) the parasite is an obligate intracellular parasite during the mammal host stage, an adaptation that is not shared by the closely related *T. brucei*, the causative agent of sleeping sickness in the African continent, 3) the evolution of intracellular parasitism occurred more than once in the kinetoplastids (species of the genus *Leishmania* also adapted to an intracellular niche and became obligate intracellular parasites that infect mostly macrophages (Sibley 2011)), 4) *T. cruzi* has the remarkable ability to invade any nucleated cell within the mammal host, 5) *T. cruzi* can infect a diverse range of mammal hosts (the parasite has been isolated from more than 70 genera (Zingales, et al. 2012)) (Fig 5-1).

The *T. cruzi* genome sequencing project found that approximately 50% of the protein-coding genes of this parasite were composed of repetitive elements (El-Sayed, et al. 2005a). A large majority of the repetitive elements were part of surface proteins such as trans-sialidases (TSs), mucins, mucin associated surface proteins (MASPs), dispersed gene family 1 (DGF-1), and gp63 peptidases (De Pablos and Osuna 2012; El-Sayed, et al. 2005a). Higher sequence conservation observed within paralogs from the same strain haplotype, rather than between haplotypes, suggests these protein families are of recent origin (El-Sayed, et al. 2005a). When comparing the genome content of *T. cruzi* with other kinetoplastids, it became clear that most of these surface protein families were

unique to *T. cruzi*, and therefore must play roles that are unique to this parasite (El-Sayed, et al. 2005b).

One of the most prominent characteristics of *T. cruzi* is its ability to invade any type of nucleated cell within its mammal host. To effectively establish an intracellular infection in its host, the parasite needs to establish a stable interaction with matching receptors in the host cell. In order to accomplish such a feat the parasite had to evolve the ability to attach to numerous different types of host cell receptors in order for intracellular entry to ensue. This feat is accomplished with the parasite's ability to express a diverse array of surface proteins with cell-attachment functions (Epting, et al. 2010). Many of the functions of the surface protein families that are unique to *T. cruzi* (i.e. TSs, DGFs) have cell-attachment roles (De Pablos and Osuna 2012).

Gene duplication has been shown to be if not the most important, one of the most important mechanisms that increases an organism protein diversity (Ohno 1970). Gene duplicate age distributions have been shown to be the ideal method to describe ancestral genome duplication events as well as small-scale duplication events (SSDs) (Lynch and Conery 2000; Lynch and Conery 2003). Estimating duplicate age distributions in protozoan genomes allow us to address questions pertaining what genomic changes are associated with the evolution of certain pathogenic traits. Here we use a large set of annotated protozoan genomes to investigate what genomic transformations have ensued in kinetoplastid pathogens when they have evolved: 1) an obligate intracellular lifestyle, 2) the ability to invade different host cells, and 3) the ability to invade multiple host species.

MATERIALS AND METHODS

Data sets

We analyzed two genome data sets of *Trypanosoma cruzi (*haplotypes Esmeraldo and "Non-Esmeraldo"). Both haplotypes were obtained during the original genome-sequencing project of *T. cruzi* (El-Sayed, et al. 2005a). The genome sequences of *T. cruzi marinkellei, T. brucei*, two *Leishmania spp*. (*L. infantum* responsible for visceral

leishmaniasis (VL) and *L. major*, responsible for cutaneous leishmaniasis (CL) in the old world) and *Naegleria gruberi* were also used (Franzen, et al. 2012; Fritz-Laylin, et al. 2011; Ivens, et al. 2005; Peacock, et al. 2007). Blastall searches with an E-value of 10^{-10} were done on protein genomic data files to construct the paralog data sets. Both *T. cruzi* strains (non-Esmeraldo and Esmeraldo), *T. c. marinkellei*, *T. brucei, Leishmania infantum* and *Naegleria gruberi* data sets consisted of 3450, 2984, 4283, 2254, 2645 and 8024 paralog proteins respectively. In *T.* cruzi this represents approximately 29-34% of the protein coding genes annotated in the most current genome annotations (as of May 2013).

The non-Esmeraldo haplotype was used for the surface protein family analysis. The surface protein family *T. cruzi* data set consisted of: 245 mucins out of 502 found in the genome, 445 trans-sialidases out of 581 found in the genome, 171 DGF-1 proteins out of 186 found in the genome, and 400 MASPs out of 502 found in the genome. The *T. c. marinkellei* surface protein family data set consisted of: 757 trans-sialidases out of 858 found in the genome, 48 mucins out of 62 found in the genome, 166 MASPs out of 251 found in the genome, and 901 DGF-1 out of 1004 found in the genome. Annotated pseudogenes were not included in the analyses.

Alignment, selection and timing analyses

Sequences from each paralog data set were translated to amino acids with translatorx3.pl (Abascal, et al. 2010) and aligned with MUSCLE (Edgar 2004). Poorly aligned regions were removed using Gblocks (Castresana 2000). Customized scripts were used to format paralog alignments for their use in PAML (Yang 2007)**.** An approximate method that takes into account transition/transversion and codon bias was used to estimate the rate of synonymous substitutions (dS) (Yang and Nielsen 2000). Synonymous Codon Usage Order (SCUO) was used to estimate codon bias in each protein family (Wan, et al. 2004; Wan and Zhou 2003). To estimate the time of the surface protein expansions we solved the equation $D= \mu/2T$ equation, where D represents the divergence between the paralog pairwise comparison (measured as dS), μ represents the neutral mutation rate per site per year based on *T. brucei* estimates (Lynch 2010) and T represents the divergence time.

RESULTS

Duplicate age distribution

The duplicate age distributions of two distinct *T. cruzi* haplotypes (Esmeraldo and non-Esmeraldo) revealed the presence of what appears to be three distinct peaks within the dS ≤ 1 range (Fig 5-2A). Two additional much older and less pronounced peaks appear at around dS values of 1.85 and 3.9 (Fig 5-2A). In order to assess the relative appearance of these gene duplication events we analyzed the sister taxa of *T. cruzi* (*T. c. marinkellei*), a bat infecting *T. cruzi* species that had it's genome recently sequenced (Franzen, et al. 2012), as well as *T. brucei* the causative agent of sleeping sickness in Africa (Berriman, et al. 2005). The same 3 recent gene duplicate peaks were present in the *T. c. marinkellei* data set (Fig 5-2B), although the more ancient peak at around dS: 0.9 is not as pronounced in *T. c. marinkellei* as it is in *T. cruzi*. All three peaks were absent in the more distantly related *T. brucei* species (Fig 5-2C), however the more ancient gene duplication events observed at dS value of 1.85 and 3.9 are shared between all three species (Fig 5-2 C-D).

To evaluate the phylogenetic distribution of the ancestral gene duplication events observed at dS: 1.85 and 3.9 we analyzed several other genomes of kinetoplastids as well as members of the Excavata super kingdom (Fig 5-2D). Both peaks were present in most taxa analyzed, however the precise location of the peaks did not fit as nicely as the peaks that are unique in *T. cruzi spp*. Since these peaks are over the saturation point of dS ≥ 1 the timing of these events is unreliable. However the presence of these ancestral duplication events might be a useful taxonomic character (unpublished data).

Surface protein expansions.

Since surface protein families are known to comprise a large fraction of the described protein coding genes in *T. cruzi*, and appear to be unique to *T. cruzi spp*. we decided to analyze each of these major surface protein families individually (TS, Mucins, MASP's, DGF's and gp63 peptidases). Once our data set was divided into protein families it became quickly visible that these families (except gp 63) by themselves could explain the most recent peaks observed in the gene duplication distributions (Fig 5-3A). In fact it became clear that the three peaks were actually composed of 6 different peaks. A trans-

sialidase expansion was almost fully responsible for the peak at dS: 0.9, a second much smaller peak at dS: 0.6 was caused by a MASP expansion. A larger third peak at dS: 0.48 was caused by a family expansion of mucins. Trans-sialidases had a second significant gene family expansion at dS of 0.32. The fifth and largest peak observed was caused entirely by a large gene duplication expansion that occurred in DGF's at dS: 0.22. Finally, a second small gene duplication expansion took place in MASP's at dS: 0.11. The protein family analysis of the closely related bat infecting species of *T. c. marinkellei* revealed that the trans-sialidase and DGF protein family gene duplication expansions were the only events responsible for the peaks that coincide with *T. cruzi* protein expansions (Fig 5-3B). As observed in *T. cruzi*, the trans-sialidases in *T. c. marinkellei* went through two distinct waves of gene duplication, however there were no Mucin or MASP's expansions.

Since trans-sialidase genes in *T. cruzi* are thought to have many distinct crucial functions in the biology of the parasite, we assessed the composition of the two distinct trans-sialidase peaks observed in both *T. cruzi* species. A phylogenetic analyses of the proteins found in each peak did not distinguish each peak as unique, which means the peaks do not represent gene duplication expansions that are functionally unique from each other (Fig S5-1). However we found that in both peaks, there was a large amount (~70%) of proteins with a conserved motif (VTVxNVxLYNR) that has been shown to bind to host cell cytokeratins, consequently allowing a stable connection between the parasite and the mammal host cell. This motif has been reported to be partly responsible for the organ tropism observed in *T. cruzi* (El-Sayed, et al. 2005a; Magdesian, et al. 2001; Tonelli, et al. 2010).

The location of the gene duplication events for each surface protein family in *T. cruzi* was not random; we found there were certain chromosomes that were overrepresented for each surface protein peak observed in the duplication age distribution (Table S5-1). Both independent analyses of the protein family expansions in *T. cruzi* and *T. c. marinkellei* as well as the chromosome over-representation analysis (Table S5-1) confirm that these gene duplication events were caused by small-scale duplications (SSD's) rather than whole genome duplication events. These SSD's occurred between the divergence of the triatomine vectored trypanosomes and the ancestor of the *T. cruzi spp*.

clade. However the more ancestral peaks that are shared with *T. brucei* could in fact be due to an ancestral genome duplication (Fig 5-2C) (manuscript in progress).

Timing duplication events.

In order to determine when these events occurred we used the estimated mutation rate of *T. brucei* ($1.65x10^{-9}$ per base per generation) (Lynch 2010), with the estimated 10 generations per year. Time estimations using 10 generations per year: the most recent MASP peak (dS=0.11) was estimated to occur approximately 3.3 mya, the DGF-1 peak (dS=0.22) was estimated to occur approximately 6.6 mya, the most recent TS peak (dS=0.32) was estimated to occur approximately 9.6 mya, the mucin peak (dS=0.48) was estimated to occur approximately 14.5 mya, the older MASP peak (dS=0.6) was estimated to occur approximately 18.1 mya, and the oldest TS peak (dS=0.9) was estimated to occur approximately 27.2 mya.

In *T. cruzi* the average number of generations per year is not known, but since it is well known that the intracellular developmental stage of the parasite goes through 9 rounds of cellular division before exiting the host cell to invade other cells, the number of generations per year is likely to be higher than 10 on an average year (Dvorak and Hyde 1973). Therefore we additionally analyzed the data using a much higher number of generations (20). Time estimations using 20 generations per year: the most recent MASP peak (dS=0.11) was estimated to occur approximately 1.6 mya, the DGF-1 peak (dS=0.22) was estimated to occur approximately 3.3 mya, the most recent TS peak (dS=0.32) was estimated to occur approximately 4.8 mya, the mucin peak (dS=0.48) was estimated to occur approximately 7.2 mya, the older MASP peak (dS=0.6) was estimated to occur approximately 9 mya, and the oldest TS peak (dS=0.9) was estimated to occur approximately 13.6 mya.

The number of generations per year in our calculations is highly important since the higher number of cell divisions per year that the parasite has will result in a more recent estimate for the SSD's. The divergence measured as the number of synonymous substitutions per site per year (dS) assumes a clockwise rate (Lynch 2010).

The specific phylogenetic distribution of the SSD's described here is unknown. At the time when our analyses were done the only species of the Hemiptera transmitted trypanosomes that had their genome sequence available were 4 strains of *T. cruzi* and one strain of *T. c. marinkellei*, thus at this point we only know they occurred sometime between the split from the *T. brucei* clade and before the *T. cruzi* lineage divergence (Fig 5-1). Until genome sequences become available for additional Trypanosomatid species we cannot determine with certainty the phylogenetic distribution of these major gene duplication events.

Codon bias.

Of the four protein families that we analyzed separately the DGF proteins were the only surface protein family in our data set that had a significantly higher Synonymous Codon Usage Order (SCUO) than the rest of the *T. cruzi* proteins in the data set (t-test <0.001). Trans-sialidases, Mucins and MASP had significantly smaller SCUO values compared to the rest of the data set in *T. cruzi*. The burst of duplication of the DGF family is the most recent large duplication event based on synonymous site divergences (Fig 5-3A). Given the well known inverse relationship between codon bias and substitution rate (Sharp and Li 1987) it is likely that the burst of duplication of these gene family was not a separate event but rather occurred concurrently with the most recent expansions of trans-sialidases and mucins (Fig 5-3A).

DISCUSSION

The use of duplicate age distributions is the most used and practical method of revealing ancestral gene duplication events, whether for whole genome duplications (WGD) or for smaller scale duplication events (i.e. protein family expansions). Their use has revealed the importance gene duplication events have played in the biology of many taxa (Van de Peer, et al. 2009). The most studied examples have been described in many distinct families of plants, which have had ancestral WGD that are thought to have contributed to major radiation events (Cui, et al. 2006; Soltis and Soltis 2009). In organisms that have evolved into a pathogenic lifestyle, the reduction in genome size due to the loss of certain

biological functions that the pathogen can acquire from its host, as well as the expansion of certain protein families involved in cell-attachment, and/or immunogenic functions has been common (Heinz, et al. 2012; Toft and Andersson 2010). With the use of the sequenced genomes available in Kinetoplastids, we constructed duplicate age distributions and describe how the expansions of only four protein families explain the major peaks observed in this parasite duplication age distribution (Fig 5-3A). Based on the phylogenetic distribution and biological function of the proteins involved, we hypothesize these events were linked to the evolution of the parasite's ability to invade any nucleated cell in their mammalian hosts as well as the adaptation to invade many species of mammals.

Surface protein expansions.

The trans-sialidase (TS) protein family had two major expansions in the duplicate age distribution (Fig 5-3A). The TS family constitutes the largest protein family in all kinetoplastids sequenced thus far. TS are not unique to *T. cruzi*. There are 9 copies currently annotated in the *T. brucei* genome. The name trans-sialidase comes from the function of certain members of this family to transfer sialic acid from the surface of host cells. The parasites are unable to synthesize the molecule, which confers them protection from galactosyl antibodies (Pereira, et al. 1980). However many other functions are thought to be associated with the large protein family in *T. cruzi*, where the TS's went through a very dramatic expansion (e.g. in the Non-Esmeraldo genome there are 277 putative TS's and 304 TS's pseudogenes currently annotated compared to the 9 copies found in *T. brucei* as of May 2013). In *T. cruzi* TS are thought to have functions that are crucial for intracellular parasitism and the ability to invade multiple types of cells; like anitiapoptotic mechanisms, securing a connection with the host cell receptors, and immune evasion (Beucher and Norris 2008; Chuenkova and Pereira 2000; de Melo-Jorge and PereiraPerrin 2007; El-Sayed, et al. 2005a; Norris, et al. 1991; Tonelli, et al. 2010). A more careful inspection at the TS's present in both peaks (Fig 5-2A) revealed that a majority of the proteins (~70%) in each peak had a TS protein motif (VTVxNVxLYNR) that has been shown to bind to host cell membrane receptors. That connection is crucial for the parasite to establish a stable bond with the host cell membrane, therefore enabling

the parasite to invade host cells. Furthermore, this protein motif attaches strongly to heart, esophagus, and other tissues, thus resembling some of the tissue tropism observed in chronic Chagas (El-Sayed, et al. 2005a; Magdesian, et al. 2001; Tonelli, et al. 2010). The more ancestral of the two peaks in TS's is not as pronounced in *T. c. marinkellei* as it is observed in *T. cruzi*. The reason for more copies of TS to have been retained in *T. cruzi* is currently unknown, but we can speculate it is linked to the parasite ability to infect many species of mammals and/or immune evasion mechanisms. Once there is a better understanding of all the functions of the members of this large surface protein family a more testable hypothesis to explain the difference in copy number between species can be formulated.

A substantial gene family expansion was also observed in the surface expressed protein family known as mucins, specifically a sub-family of mucins (TcMUCII proteins) that are mostly expressed in the mammal cell invasion stage. Mucins have been described to play two primordial functions: 1) protect the parasite from both the host and vector's immune response, 2) guarantee the cell anchorage point so the parasite can invade the host cells (Buscaglia, et al. 2004; Buscaglia, et al. 2006). The bat restricted *T. c. marinkellei* has a significantly smaller number of mucins annotated in its genome (see methods), which is probably why the mucin gene expansion is not observed (Fig 5-3B). In order for *T. cruzi* to have evolved the ability to invade a large number of insects and mammal species, it seems logical that the parasite would need a much larger arsenal of mucins available. We speculate that the large mucin gene expansion observed in *T. cruzi* is associated with the parasite's adaptation to invade a much wider range of vectors and hosts than the closely related bat-infecting *T. c. marinkellei*. This same hypothesis probably explains the smaller gene expansions observed in the recently discovered mucin associated surface proteins (MASPs) (El-Sayed, et al. 2005a) (Fig 5-3A). Just as it is observed in mucins, there are fewer MASPs annotated in *T. c. marinkellei* than in *T. cruzi*. Although little is known about the biological role of this large protein family, a unique characteristic of these proteins is the formation of chimeras with members of the TS and mucin protein motifs (a trait that can influence the position of the MASP peaks). The large sequence diversity within this protein family suggests they might play an important role providing the parasite with a source of antigens and thus aiding in immune evasion.

Of the four major surface protein families in our study, only the dispersed gene family 1 (DGF-1) had a significantly higher codon bias usage compared to the *T. cruzi* data set. Strong codon bias can decrease substitution rate (Sharp and Li 1987). Therefore, the large DGF-1 protein family expansion observed in both *T. cruzi* and *T. c. marinkellei* is probably older than our dS based estimates. Due to epidermal growth factors and the presence of certain protein domains it is thought that these proteins may have a function similar to integrins. Integrins are cell surface receptors that facilitate cell-cell attachment and cell matrix interactions. The significant gene duplication enlargement of DGF-1 is therefore probably linked to the adaptation of the parasite to invade multiple cell types in the mammal hosts.

Estimating Time of protein family expansions.

The precise timing of each event can be attained with confidence when certain criteria and assumptions are met: 1) the event has not surpassed the saturation rate of synonymous substitutions ($ds \leq 1$), 2) there is a linear accumulation of synonymous substitutions through time (molecular clock), 3) the mutation rate of the taxa in question is known, and 4) the average number of generations per year is known. In our study, unfortunately both the number of generations per year and the mutation rate of *T. cruzi* are not known, so we used the mutation rate of *T. brucei* (the closest species with this information). The mutation rate of *T. cruzi* is likely to be very similar to the rate of *T. brucei*, however the number of generations per year greatly affects our time estimates. Therefore, we cannot estimate with confidence the timing of such events. Nevertheless the phylogenetic distribution and functions of each major surface protein family expansion allows us to build a hypothesis which links these events with the evolution of intracellular parasitism and the adaptation of the parasite to invade many types of cells in the mammal host.

Intracellularity, cell tropism and species range.

The evolution of an intracellular niche in kinetoplastids has been a very successful evolutionary trait. The large number of species that have been described in this phylogenetic clade with an intracellular parasitism niche support this idea (Lima, et al.

2012). *T. cruzi* has the additional remarkable ability to invade any nucleated cell in it's mammal host and has been found in more than 70 different genera of mammals thus far (Zingales, et al. 2012). The genomic footprints of such adaptations have only been recently addressed with the use of genomic data. Comparative genomic approaches have been a very powerful method to address these questions (Heinz, et al. 2012; Wernegreen 2005). Since intracellular parasitism also evolved in *Leishmania spp*. we had a very valuable data set to explore these questions. The *Leishmania spp.* analysis did not show any recent protein family expansions, only what appears to be a very ancient gene duplication event that is shared by members of the Excavata kingdom (Fig 5-2D) (although recent modeling studies have described how older gene duplication events (ds > 2) can cause different age duplication events to cluster in the same oversaturated dS value (Vanneste, et al. 2013)). *Leishmania spp*. have a very narrow intracellular parasitism niche, since they mostly only infect macrophages. Thus, the significant different cell tropisms observed between *T. cruzi* spp. and *Leishmania spp.* allowed us to compare the gene duplication age distributions between two closely related taxa that independently evolved the same trait. The lack of major gene duplications in the *Leishmania spp*. suggests that the evolution of intracellularity did not require the occurrence of major gene duplication events in either taxa. This result is in agreement with previous studies addressing the evolution of intracellular parasitism (Heinz, et al. 2012; Wernegreen 2005).

T. cruzi has been found in a wide range of mammal's species (> 70 genera of mammals) and has the outstanding ability to invade any nucleated cell in the mammal host (Fernandes and Andrews 2012; Zingales, et al. 2012). The vast gene duplication expansions that are observed in the duplicate age distributions strongly suggest these events played a crucial role in the biology of the trypanosomes due to the fact that gene duplications are usually removed from genomes because of their deleterious effects, leaving only copies that are advantageous for the taxa in question (Lynch and Conery 2000; Lynch and Conery 2003). Therefore, it is likely that a vast majority of those gene copies have been maintained in the parasite's genome by selection. Both the phylogenetic distribution and biological functions of the proteins responsible for these major gene

duplication expansions strongly suggest these genomic adaptations are linked to the evolution of these traits.

Role of hematophagy.

The question of when or why these gene duplication events took place at such a rapid rate and in what appears to be the same evolutionary time point cannot be answered at this moment with certainty. The latest studies suggest that the ancestral parasite was found in bats (Hamilton, et al. 2012; Poinar 2005). An amber embedded fossil of a Trypanosome infected Triatomine with evidence of what appears to be a hematophagous meal from a bat places the presence of hematophagy in Triatomines at least as far back as 15-40 mya (Poinar 2005). This time interval overlaps with our time estimations for the surface protein family expansions. The evolution of hematophagy in Triatomine bugs would have exposed the parasite to potential new hosts. As the Triatomine bugs diverged and started to adapt to feed on distinct mammal taxa the parasites evolved: 1) into obligate intracellular parasites, 2) the ability to invade many distinct cell tissues, and 3) the ability to infect many distinct species of mammals. The later two adaptations are the traits we believe are responsible for the gene duplications that significantly increased the number of protein coding genes in the genome of this parasite and consequently had a huge impact on the genomic changes associated with such traits. Immune evasion might also be an important force responsible for some of the expansions (e.g. the TS are thought to play a big role in immune evasion).

Once the genomic sequences of other kinetoplastids parasites (we are in the process of sequencing *T. dionisii*, *T. rangeli* and *T. vespertilionis*) are available we will be able to phylogenetically determine more precisely the moment of these gene duplication events that had such an impact on the genome of these pathogenic protozoa. A comparative genomic study will be particularly interesting once the genome of *T. rangeli* is sequenced, due to this parasite's vast species range (Guhl and Vallejo 2003). From our study it is clear that the evolution of certain biological traits that led to the remarkable adaptability of the parasite required major genomic modifications in the form of large surface protein family expansions.

Figures

Figure 5-1. Evolutionary relationships among kinetoplastids. Branch lengths do not represent true evolutionary distances. Evolutionary relationships were obtained from (Lima, et al. 2012; Parfrey, et al. 2011). All traits placed in branches are only placed in branches were all the descendant species have been confirmed to posses the traits in question. The evolution of intracellular parasitism in *T. rangeli* has been a matter of debate, therefore we do not place this trait earlier in the cladogram (Grisard, et al. 2010; Guhl and Vallejo 2003; Stoco, et al. 2012). The ability to invade practically any nucleated cell in the mammal host has not been thoroughly studied, therefore the placement of this trait might change once more studies are performed on the subject.

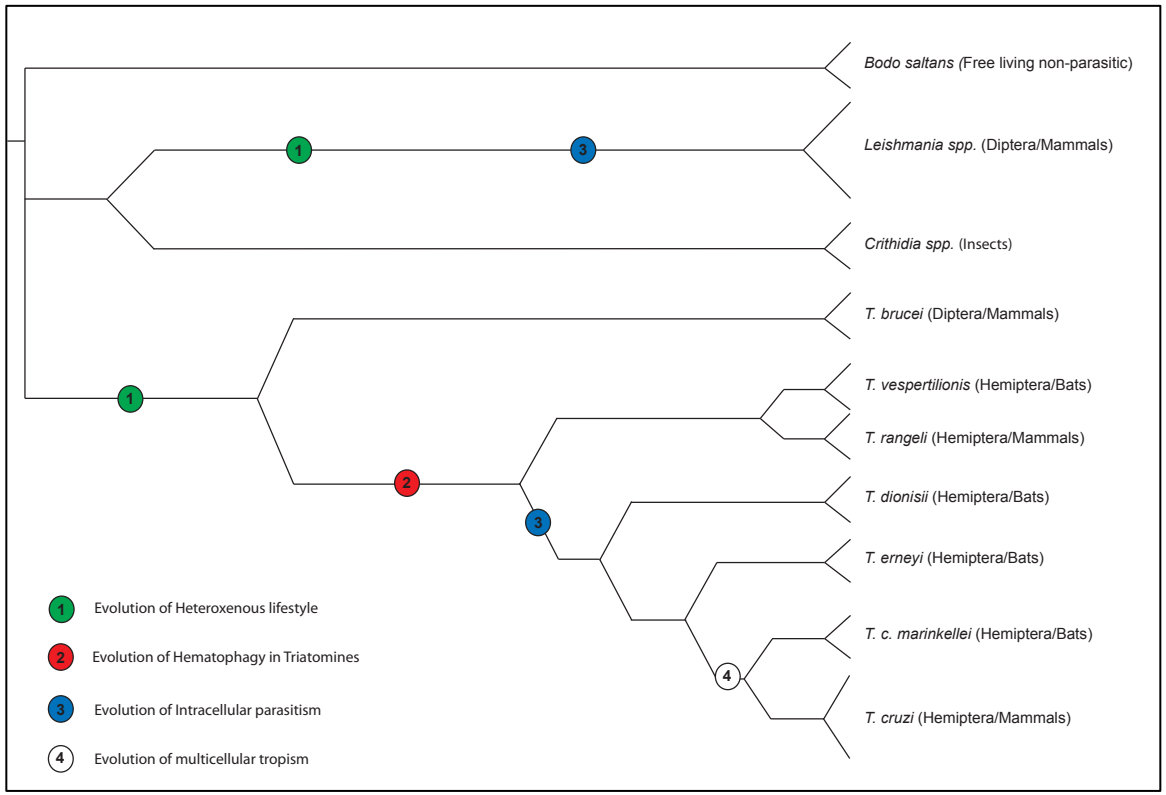


Figure 5-2. Age distribution of duplicate pairs of genes. dS: synonymous substitutions. A. Both *T. cruzi* haplotype data sets are showed. B. Non-Esmeraldo *T. cruzi* haplotype and *T. cruzi marinkellei*. C. *T. cruzi* versus *T. brucei*.

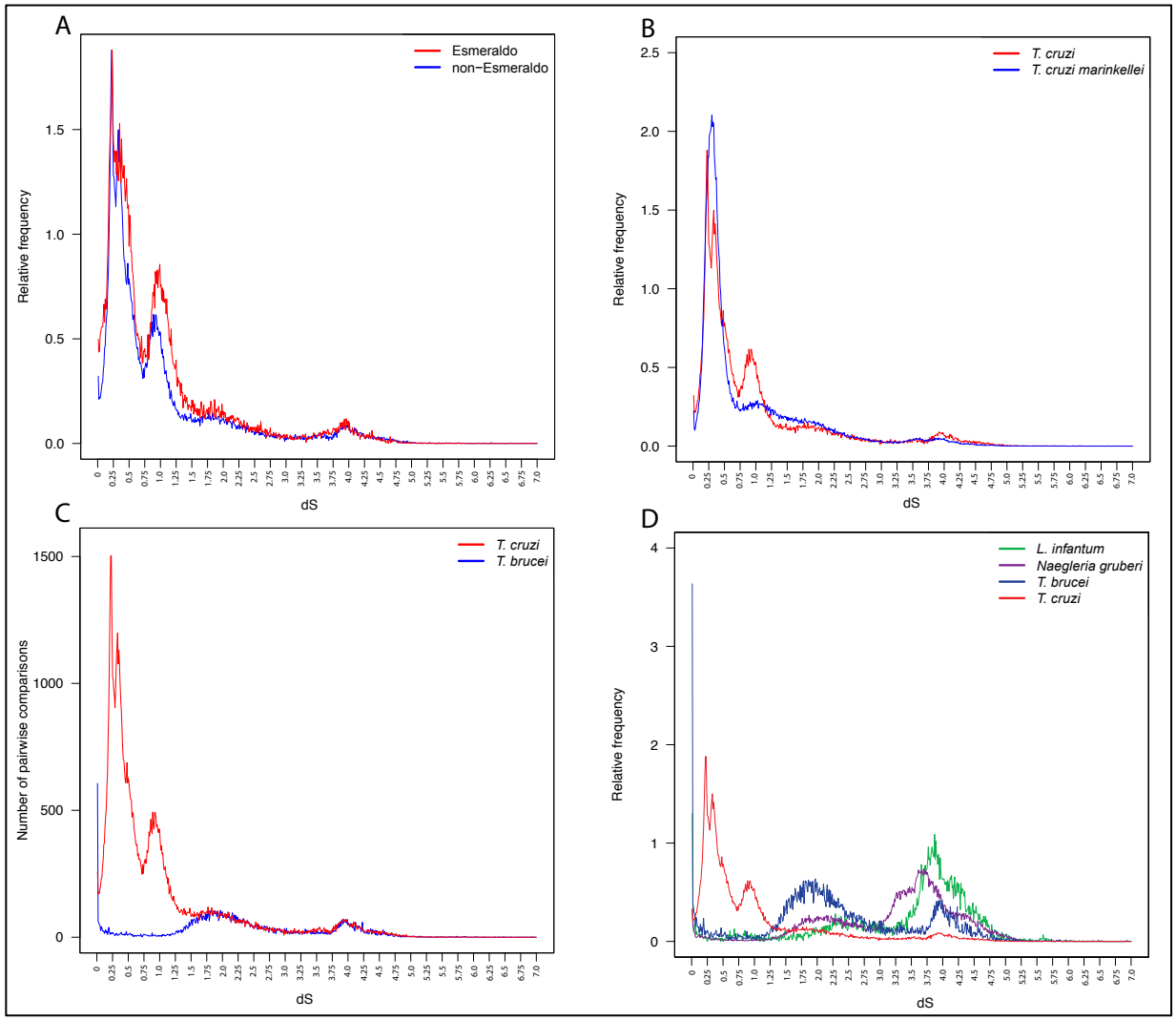
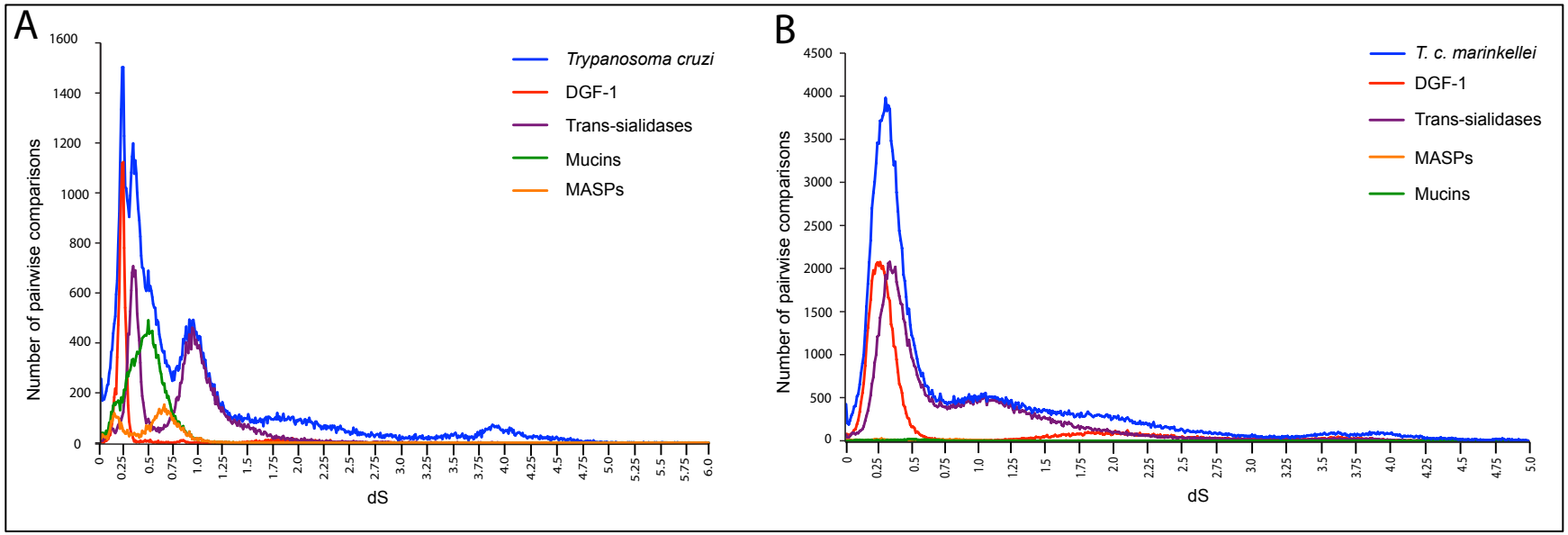


Figure 5-3. Age distribution of duplicate pairs of genes in the surface protein families. A. Non-Esmeraldo *T. cruzi* haplotype. B. *T. cruzi marinkellei*. Time estimations using 10 generations per year: the most recent MASP peak (dS=0.11) was estimated to occur approximately 3.3 mya, the DGF-1 peak (dS=0.22) was estimated to occur approximately 6.6 mya, the most recent TS peak (dS=0.32) was estimated to occur approximately 9.6 mya, the mucin peak (dS=0.48) was estimated to occur approximately 14.5 mya, the older MASP peak (dS=0.6) was estimated to occur approximately 18.1 mya, and the oldest TS peak (dS=0.9) was estimated to occur approximately 27.2 mya. Time estimations using 20 generations per year: the most recent MASP peak (dS=0.11) was estimated to occur approximately 1.6 mya, the DGF-1 peak (dS=0.22) was estimated to occur approximately 3.3 mya, the most recent TS peak (dS=0.32) was estimated to occur approximately 4.8 mya, the mucin peak (dS=0.48) was estimated to occur approximately 7.2 mya, the older MASP peak (dS=0.6) was estimated to occur approximately 9 mya, and the oldest TS peak (dS=0.9) was estimated to occur approximately 13.6 mya.



Chapter 6. Overview and Conclusions to Dissertation

Overview

Chapter 2: Phylogenetic relationships among *Trypanosoma cruzi* lineages

This chapter is a description of the evolutionary history of the major lineages that have been described in *T. cruzi*. The motivation to do this work came from the previous indication that the classification system of the parasite that was being used at the moment did not reflect the true evolutionary history of the parasite. Such an artifact would not allow proper evaluations of the association between the genetic background of the parasite with the diverse pathologies observed in Chagas. Our results confirmed that the previous nomenclature used to divide *T. cruzi* genetic diversity in two main lineages was erroneous, and it is demonstrated that any previous studies that attempted to link the genetic background of the parasite with a given pathology have to be reevaluated.

In chapter 2 we also addressed the fact that there is only evidence for one ancient hybridization event that took place between the ancestors of lineages TcII and TcIII, which gave rise to lineages TcV and TcVI. Previous studies had proposed an additional older hybridization event between the ancestors of TcI and TcII. However we found no evidence to support that second hybridization event.

Several studies have proposed humans have played an important role in the evolution and dispersal of *T. cruzi* major lineages across the Americas. Our estimates of the time of divergence of the major lineages of *T. cruzi* do not support this anthropogenic hypothesis. The number of informative sites used in our study remains to be the largest data set used thus far to reconstruct the evolutionary history of the parasite, therefore we

have confidence the anthropogenic hypothesis concerning the evolution of *T. cruzi* has no support.

Chapter 3: Description of a new *Trypanosoma cruzi* lineage from the United States

The work in chapter 3 portrays the description of a new phylogenetic clade of *T. cruzi* that appears to be unique to North America (TcNA). Our divergence estimates of the node of the crown of this clade (i.e. the most recent common ancestor) suggest a Pleistocene origin. Since this lineage appears to be unique to North America, our divergence estimates most likely are correlated with the time this lineage entered into North America. Furthermore, evidence for genetic exchange events in some of the isolates from the USA reinforces the importance these events play in the biology of the parasite.

Chapter 4: Positive selection has played a larger role in the evolution of *Trypanosoma cruzi* proteins than in the evolution of *Leishmania spp*. proteins

Chapter 4 describes evolutionary analyses of protein-coding genes from 4 annotated genomes of *T. cruzi* and 4 annotated genomes of *Leishmania spp.* Results show that positive selection has played a larger role in the evolution of *T. cruzi*. We report that a significantly larger fraction of protein-coding genes have been under positive selection in *T. cruzi* than in *Leishmania spp*. Furthermore, the intensity of positive selection (as estimated by the average value of dN/dS) is also significantly higher in *T. cruzi* than in *Leishmania spp.* We suggest that the greater versatility of *T. cruzi* in its host range, cell tropism and cell invasion mechanisms can explain the observed differences.

Any pathogen gene experiencing positive selection might be under strong diversifying selection in order to evade the immune system. Therefore, this approach can be useful for identifying candidate immunogenic genes for the development of a vaccine. In *Leishmania spp.* the top functional categories that were statistically overrepresented among those under positive selection have been shown to play some role in the evolution of drug resistance (ATP binding cassette), in the interaction with the host immune system (cysteine peptidase), or have been proposed as vaccine candidates for Leishmaniasis (iron superoxidase dismutase). In *T. cruzi,* hypothetical proteins, mucin, and amino acid permease were the top overrepresented functional categories with genes with signal of positive selection. We provide a list of proteins that could be tested for vaccine potential in both human diseases.

Chapter 5: Protein family expansions in the parasitic lifestyle of *Trypanosoma cruzi*

With the use of gene duplicate age distributions; we describe the recent expansions of a few cell surface protein families. The chromosomal locations of each surface protein expansion suggest these gene duplications are not due to a whole genome duplication event. The phylogenetic placement, time estimates and functions of the protein family expansions are most likely associated with the evolution of *T. cruzi* to invade multiple cell tissues and its adaptation to invade multiple host species. The time estimates for these gene duplication events overlap with the evolution of hematophagy in the insect vectors. Thus, we suggest that the evolution of hematophagy was tightly associated with these evolutionary events. The evolution of hematophagy in the insect vectors exposed the ancestral parasite to a vast range of potential new hosts. In turn, this could have contributed to the adaptation of the parasite to invade multiple cell tissues.

FUTURE DIRECTIONS

Concerning the evolutionary history of the parasite (Chapter 2) in conjunction with the description of a new phylogenetic clade that appears to be unique to North America (Chapter 3), once more under-sampled geographic areas and mammal hosts are studied, a more realistic view of the genetic diversity of the parasite currently circulating will be revealed. This in turn will allow for a more accurate study of the evolutionary history of the parasite. With DNA sequencing costs decreasing, many genome sequence projects of additional kinetoplastid species will be likely to be completed. This will allow for larger data sets to be assembled. Even though the enlargement in data sets does not guarantee the recovery of the true evolutionary history of any group of taxa, this will definitely allow for more accurate interpretations to be made.

Larger data sets will provide more power to quantify the role genetic exchange events have played in the evolutionary history of the parasite. The results from chapters 2 and 3 describe how genetic exchange events have played a very important role in the evolutionary history of the parasite. More specifically in chapter 2, we demonstrate evidence for only one major ancient genetic exchange event among the ancestral lineages of the parasite. This ancestral genetic exchange event produced two of the six extant classified lineages of *T. cruzi*.

In order to accurately estimate the timing of the divergence of a taxa, the use of multiple fossil records are usually recommended to be used as calibration points. There is almost a complete void of fossils for this group of parasites. Therefore even if more accurate mutation rates are employed, as long as there is a lack of fossils that can be used as calibration points, the timing of divergence in *T. cruzi* will continue to rely on the use

of vicariance events (the separation of Africa from South America is used as the vicariance event that separated *T. cruzi* from *T. brucei*).

The comparison of patterns of positive selection in *T. cruzi* and *Leishmania spp.* genes can be greatly improved by increasing the number of lineages used and, most importantly, by comparing data sets with similar divergence times. One way to achieve this is by sequencing the genomes of close relatives of *T. cruzi*. In collaboration with Dr. Machado, there are plans to sequence three additional closely related *Trypanosoma* species in the near future, as well as many other strains of *T. cruzi*. This potential improved data set should reduce the number of false positives and possibly even uncover additional proteins in *T. cruzi* that have the desirable traits a vaccine candidate should have. Even with an improved dataset the number of proteins predicted to be under positive selection should still be higher in *T. cruzi* than in *Leishmania spp.* as a result of the greater versatility of *T. cruzi* in terms of mammal species it can infect, but most likely due to its remarkably diverse cell tropism. Moreover, as more functional studies are conducted on the many proteins of unknown function currently annotated in *T. cruzi* genome, a better understanding of why such proteins might be evolving under positive selection should be achievable. This in turn might help prioritize the selection of proteins to be used as vaccine candidates.

In similar terms as those described above, the inferences described in chapter 5 would greatly benefit from the addition of species into the data set. Adding more species that diverged between *T. cruzi* and *T. brucei* will allow an improvement in regards to the phylogenetic placement of the surface protein family expansions. This will become possible once we sequence additional species of kinetoplastids. Further, more functional

studies on the protein families involved in the gene duplication events will be necessary to reinforce the hypothesis that these massive gene duplication events played a major role on the evolution of multicellular tropism and adaptation to invade a large range of species. For example, there have been many distinct functions assigned to the trans-sialidase protein family; therefore, an increased understanding on the many functions found in this immense protein family will allow for better hypothesis testing.

CONCLUSION

The description of the evolutionary history of *Trypanosoma cruzi* will be of great value for future studies attempting to address if there is any relationship between parasite genetic background and pathology. The detection of a new phylogenetic clade of the parasite circulating in the USA needs to be addressed by future studies that should investigate the pathology (if any) and natural habitat of this lineage. The description of uncharacterized genetic exchange events in some of the lineages isolated in the USA reinforces the importance genetic exchange events have played in the evolution of the parasite. We expect the lists of proteins from both *T. cruzi* and *Leishmania spp.* that are under purifying selection but have codon sites predicted to be under positive selection can be tested in the near future for their potential as vaccine candidates. Finally, the description of the very recent surface protein family expansions we uncovered in the genomes of *T. cruzi* and *T. cruzi marinkellei* will contribute to increase our knowledge of the genomic modifications that are associated with the evolution of pathogenicity and in particular with the adaptation of these parasites to invade multiple cell types.

Appendices

Table S2-1. Additional *Trypanosoma cruzi* strains used for some of the loci.

[a] Discrete typing unit (DTU) (1999; Brisse, et al. 2000a; Miles, et al. 1978; Tibayrenc and Breniere et al. 1988; Zingales, et al. 2009).

Strains	**DTU** [a]	**Zymodeme** [b]	**Isoenzyme types** [c]	**1999 classification** [d]	**2009 nomenclature** [e]
133 79 cl7, CUICA cl1, TEH cl2 cl92, CEPA EP, Vin C6, FLORIDA C16, X10 cl1, SABP3, A80, A92, MA-V, OPS21 cl11, Esquilo cl1, CUTIA cl1, V121, 2679, P209 cl1, 85/818, P0AC	I	Z1	12,17,19,20	TcI	TcI
TU18 cl2, ESMERALDO cl3, X-300, MSC2, MCV, MVB cl8	IIb	Z2	30,32	TcII	TcII
M5631 cl5, CM 17, X110/8, X9/3, X109/2	IIc	Z2	36	TcII	TcIII
CANIII cl1	IIa	Z3	27	TcII	TcIV
EPP, PSC-O, 86-1	IId	Z2	39	TcII	TcV
P251, TULAHUEN cl2, P63 cl1, 86/2036, VMV4	IIe	Z2	43	TcII	TcVI

Table S2-2. Amplified nuclear loci and PCR primers. Locus ID of *T. brucei* homolog was obtained from tritrypdb.org. [a]
Sequence data from Machado & Ayala (Machado and Ayala 2001)

Locus ID	Sequenced (bp)	Predicted function	Chromosome location	Location in Chr. Gene length	Primer sequence (5'-3')	Homologue in *T. brucei*
HSP70	508	Intergenic region	Chr 32	699686 - 700540 (-) Length: 855 bp	AGGGTGATCAGCAGAAGCAG CGCAAACGACGAGCGAACAT	N/A
Tc00.1047053503885.80	946	Not known	Chr 26	163788 - 164993 (+) Length: 1206 bp	ACAATCGATGTGCTTGACGA CAGTACGAGCCCGAGACATC	Tb927.8.6320
Tc00.1047053503891.50	813	Not known	Chr 20	75320 - 76489 (-) Length: 1170 bp	CACACCGTCTCTTCCCACTT CCGCTATGTCCATTTCACCT	Tb927.10.4750
Tc00.1047053503909.76	614	Ferric reductase transmembrane protein	Chr 32	556434 - 557156 (+) Length: 723 bp	GGAGCAACCGCATCTTTTAC TACGATACGCCAAAGTACGC	Tb927.6.3320
Tc00.1047053504013.40	805	Serine acetyltransferase	Chr34	465693 - 466718 (-) Length: 1026 bp	TCGAAGTCATTCGGAAGTCA TGGCGTAGATGGTCACTCTG	Not found
Tc00.1047053504045.100	886	Not known	Chr 40	1854961 - 1856415 (-) Length: 1455 bp	GCAGCGGCAGTTCTTTATTC AGCCTTTCGCTCATTCTCAA	Tb927_03_v4
Tc00.1047053504057.80	858	Not known	Chr 34	417310 - 418677 (-) Length: 1368 bp	ATTACGCCCTTTGTCCAGTG GACGGGACAAGAAAGATCCA	Tb927.4.1590
Tc00.1047053504059.20	896	Endomembrane protein, putative	Chr 14	465730 - 467526 (-) Length: 1797 bp	TGAGGGAGGAATTGGTTGAG TGCACCAAATCCAAATGAAA	Tb11.02.0960
Tc00.1047053506247.200	920	Beta-adaptin, role inferred from homology	Chr 37	133811 - 136708 (+) Length: 2898 bp	TGAGTCATTACAGCGCAAGG TCTTCACTGGCTTCCTCGTT	Tb927.10.8040
Tc00.1047053506525.150	821	Not known	Chr 40	593462 - 594415 (+) Length: 954 bp	GCCGCTGATACGGACAAG CAAGTCAGAGACGGTGTCAGG	Tb927.10.14310
Tc00.1047053506529.310	727	Not known	Chr 6	97318 - 98676 (-) Length: 1359 bp	TTCTTTCAGGCTGCGATTTT CGCTGTTTGGCTCATTTCTT	Tb927.1.4220
Tc00.1047053506739.20	810	Not known	Chr 3	25655 - 27589 (-) Length: 1935 bp	AGCTAAGCACACTCGCCAAT CAATCTCTCGAGCCGTTCTC	Tb927.5.1500
Tc00.1047053507801.70	677	Protein kinase	Chr 23	535126 - 535959 (+) Length: 834 bp	AAAGAGTTGCCGTCAAGGTG CATGGGTGTTCCAATGACTG	Tb927.2.5230
Tc00.1047053508153.540	774	Not known	Chr 36	699363 - 700391 (+) Length: 1029 bp	GCATTCGAGGAGAGAACGAG GCGCTCTCAGAAGCAAAGTT	Tb927.3.3060
Tc00.1047053508461.80	838	Prostaglandin F2 alpha synthase	Chr 39	1187987 - 1189126 (-) Length: 1140 bp	TCGGATTCCTGCCTATTTTG TGTTTGCATTTTCCCACTGA	Not found
Tc00.1047053508719.70	709	Not known	Chr 37	375185 - 376402 (+) Length: 1218 bp	AAAATTGTCCATGCGAGTCC CACCAAATCCTTGCGTTTCT	Tb927.10.8940

Table S2-2 continued

Tc00.1047053509007.30	815	Not known	Chr 31	573767 - 574690 (+) Length: 924 bp	CTTCCACGATGCGCTACAG GGAGCACACAATCTCCTTCC	Tb927.8.7810
Tc00.1047053509105.70	897	thiol-dependent reductase 1	Chr 37	769449 - 770786 (-) Length: 1338 bp	ATGGTCGTTCCATTTCTTGC TAAGCCACTCCTTGGTGGAC	Not found
Tc00.1047053509561.20	880	Flagellum-adhesion glycoprotein	Chr 12	285842 - 287581 (-) Length: 1740 bp	CACCCTTGCCGGTAGTAAAA TCTATCTGGCGGAAATACGG	Tb927.8.4110
Tc00.1047053509967.50	595	Not known	Chr 10	184622 - 185329 (+) Length: 708 bp	TCTTGACATCGGGAGTAGCC ACACCAAAACACTTGGCACA	Tb927.10.1240
Tc00.1047053510101.480	829	Not known	Chr 27	190063 - 191427 (-) Length: 1365 bp	CAGCGCATTGAAGATTGTGT TGCATCAACTGAAGGTCTGC	Tb11.02.4410
Tc00.1047053510123.24	880	Not known	Chr 20	372476 - 373429 (+) Length: 954 bp	ATGCATGCAGGAACACAAAT TTCACTCGTACTGGGTGTCG	Tb927.10.12030
Tc00.1047053510131.90	936	Not known	Chr 30	340360 - 342003 (+) Length: 1644 bp	ATCGACATGGAACTCGAACC GTACTCCTCCGTGACCCAAA	Not found
Tc00.1047053510765.50	817	Not known	Chr 39	1780396 - 1781763 (+) Length: 1368 bp	TTGTGTTGCTAAGGCACTGG AATGAGACCCTCGCAAAGAA	Tb11.12.0004
Tc00.1047053510877.190	453	Not known	Chr 34	493531 - 494328 (-) Length: 798 bp	TCTGGACTCGTACGTCTACCC GGGACGTCCGTTCACGTAT	Tb927.4.1910
Tc00.1047053510889.210	693	Not known	Chr 6	154383 - 156290 (-) Length: 1908 bp	ATGGAATTGGAGCAAGAACG GGTAAAAGCCGCATCAGAAA	Tb927.1.3840
Tc00.1047053510889.310	763	Not known	Chr 6	193929 - 196025 (+) Length: 2097 bp	GTTTGGGCAACACGAAAGAT TGATGTCTGCTTGGAACCTG	Tb927.1.3450
Tc00.1047053511153.124	513	Not known	Chr 27	412720 - 413271 (+) Length: 552 bp	CGTCTTTGGGATTTCTGTCC GGTGTCAGGCTGGTC CTCT	Not found
Tc00.1047053511529.200	667	Not known	Chr 35	170438 - 171232 (-) Length: 795 bp	GTGAGGCGCGAAGAAAATAC TACGAAACGTTGCCGTCAG	Not found
Tc00.1047053503555.30 (TR) [a]	1290	Trypanothione reductase	Chr 37	713055 - 714533 (-) Length: 1479 bp	ACTGGAGGCTGCTTGGAACGC GGATGCACACCRATRGTGTTGT	Tb927.10.10390
Tc00.1047053509153.90 (DHFR-TS) [a]	1473	Dihydrofolate reductase-thymidylate synthase	Chr 27	718463 - 720028 (+) Length: 1566 bp	CGCTGTTTAAGATCCGNATGCC CGCATAGTCAATGACCTCCATGTC	Tb927.7.5480

Table S2-3. Results of the Shimodaira-Hasegawa tests. [a] –ln likelihood value for the phylogeny reconstructed under the Maximum Likelihood criteria with no topological constraints or when TcI and TcII are constrained to be reciprocally monophyletic. [b] The most plausible tree according to the Shimodaira-Hasegawa test. *significant at p≤0.05, **significant at p≤0.001.

Gene ID	ML tree [a]	TcI-TcII [a]	S-H Best tree [b]
COII-ND1	3185.44688	3203.15949	ML tree*
Tc00.1047053503555.30	2435.87661	2670.43397	ML tree**
Tc00.1047053509153.90	2907.60509	3622.74370	ML tree**
HSP70	1365.70184	1631.03893	ML tree**
Tc00.1047053503885.80	2012.54516	2277.85677	ML tree**
Tc00.1047053503891.50	1859.67704	2020.17966	ML tree**
Tc00.1047053503909.76	1312.97442	1441.75777	ML tree**
Tc00.1047053504013.40	1738.01299	1986.07561	ML tree**
Tc00.1047053504045.100	1895.77443	1959.24750	ML tree*
Tc00.1047053504057.80	1616.35925	1676.26866	ML tree*
Tc00.1047053504059.20	1879.71822	2021.47943	ML tree**
Tc00.1047053506247.200	1778.40746	1825.83641	ML tree**
Tc00.1047053506525.150	1628.37847	1809.03975	ML tree**
Tc00.1047053506529.310	1560.80452	1707.20155	ML tree**
Tc00.1047053506739.20	1812.77614	1976.33301	ML tree**
Tc00.1047053507801.70	1416.29930	1521.99612	ML tree**
Tc00.1047053508153.540	1584.79227	1701.44882	ML tree*
Tc00.1047053508461.80	1264.19040	1336.38017	ML tree*
Tc00.1047053508719.70	1367.05726	1493.09884	ML tree*
Tc00.1047053509007.30	1721.93006	1894.18590	ML tree**
Tc00.1047053509105.70	1818.85886	2086.09222	ML tree**
Tc00.1047053509561.20	1908.84944	2141.94614	ML tree**
Tc00.1047053509967.50	1304.46342	1380.39366	ML tree*
Tc00.1047053510101.480	1655.11244	1775.00741	ML tree**
Tc00.1047053510123.24	2074.22941	2285.24245	ML tree**
Tc00.1047053510131.90	2029.45289	2168.68034	ML tree**
Tc00.1047053510765.50	1752.33457	1965.21006	ML tree**
Tc00.1047053510877.190	973.63644	1005.17473	ML tree*
Tc00.1047053510889.210	1500.68706	1723.67076	ML tree**
Tc00.1047053510889.310	1451.22617	1670.24522	ML tree**
Tc00.1047053511153.124	1151.05222	1302.29934	ML tree**
Tc00.1047053511529.200	1376.77905	1586.65525	ML tree**

Table S2-4. LRT of Molecular clock on genes that had a homolog in *T. brucei*. [a] Number of sequences.[b] Likelihood Ratio Test (p value; estimated value from chi-square distribution with df=s-2, where s is the number of taxa). * Molecular clock rejected.

Gene	N[a]	-ln L Enforced clock	-ln L No clock	LRT[b]
COII-ND1	9	3176.531	3171.313	10.436 (0.16)
TR	12	3541.111	3536.922	8.37 (0.59)
DHFR-TS	12	2367.433	2355.957	22.95 (0.01)*
Tc00.1047053503885.80	13	3086.302	3079.941	12.72 (0.31)
Tc00.1047053503891.50	11	3154.729	3147.585	14.28 (0.11)
Tc00.1047053504045.100	11	2918.118	2914.197	7.84 (0.55)
Tc00.1047053504057.80	11	2470.922	2465.348	11.14 (0.26)
Tc00.1047053504059.20	14	2932.983	2919.348	27.27 (0.007)*
Tc00.1047053506247.200	11	2679.311	2676.382	5.85 (0.75)
Tc00.1047053506525.150	14	2748.648	2743.028	11.24 (0.5)
Tc00.1047053506529.310	14	2296.608	2286.840	19.53 (0.07)
Tc00.1047053506739.20	9	2696.520	2691.508	10.02 (0.18)
Tc00.1047053507801.70	13	4355.910	4350.086	11.64 (0.39)
Tc00.1047053508153.540	14	2793.654	2784.699	17.91 (0.11)
Tc00.1047053508719.70	14	2152.143	2147.313	9.66 (0.64)
Tc00.1047053509007.30	14	2701.612	2696.047	11.13 (0.51)
Tc00.1047053509561.20	13	2980.772	2970.446	20.65 (0.03)*
Tc00.1047053509967.50	12	2287.766	2283.898	7.73 (0.65)
Tc00.1047053510101.480	13	2510.700	2503.038	15.32 (0.16)
Tc00.1047053510123.24	13	3060.619	3057.453	6.33 (0.85)

Table S2-4 continued

Tc00.1047053510765.50	14	2976.456	2971.093	10.72 (0.55)
Tc00.1047053510877.190	9	1957.833	1955.794	4.07 (0.77)
Tc00.1047053510889.210	14	2150.871	2145.585	10.57 (0.56)
Tc00.1047053510889.310	12	2383.305	2379.224	8.162 (0.61)

Figure S2-1. Individual Maximum likelihood trees of each amplified locus. The most appropriate substitution model to analyze each locus was chosen using Modeltest 3.7 (Posada and Crandall 1998). All trees were obtained for each locus using ML heuristic searches in PAUP* 4.0b10 (Swofford 1998) using the tree bisection-reconnection (TBR) branch swapping algorithm. Bootstrap support values were obtained by ML analyses of 100 pseudoreplicates of each dataset. Strains (codes used in the figure): M6241 cl6 (CL35), CL Brener (Genome), SO3 cl5 (CLA39), EP255, SC13, SO34 (SO34 cl4), CBB cl3 (CBB32).

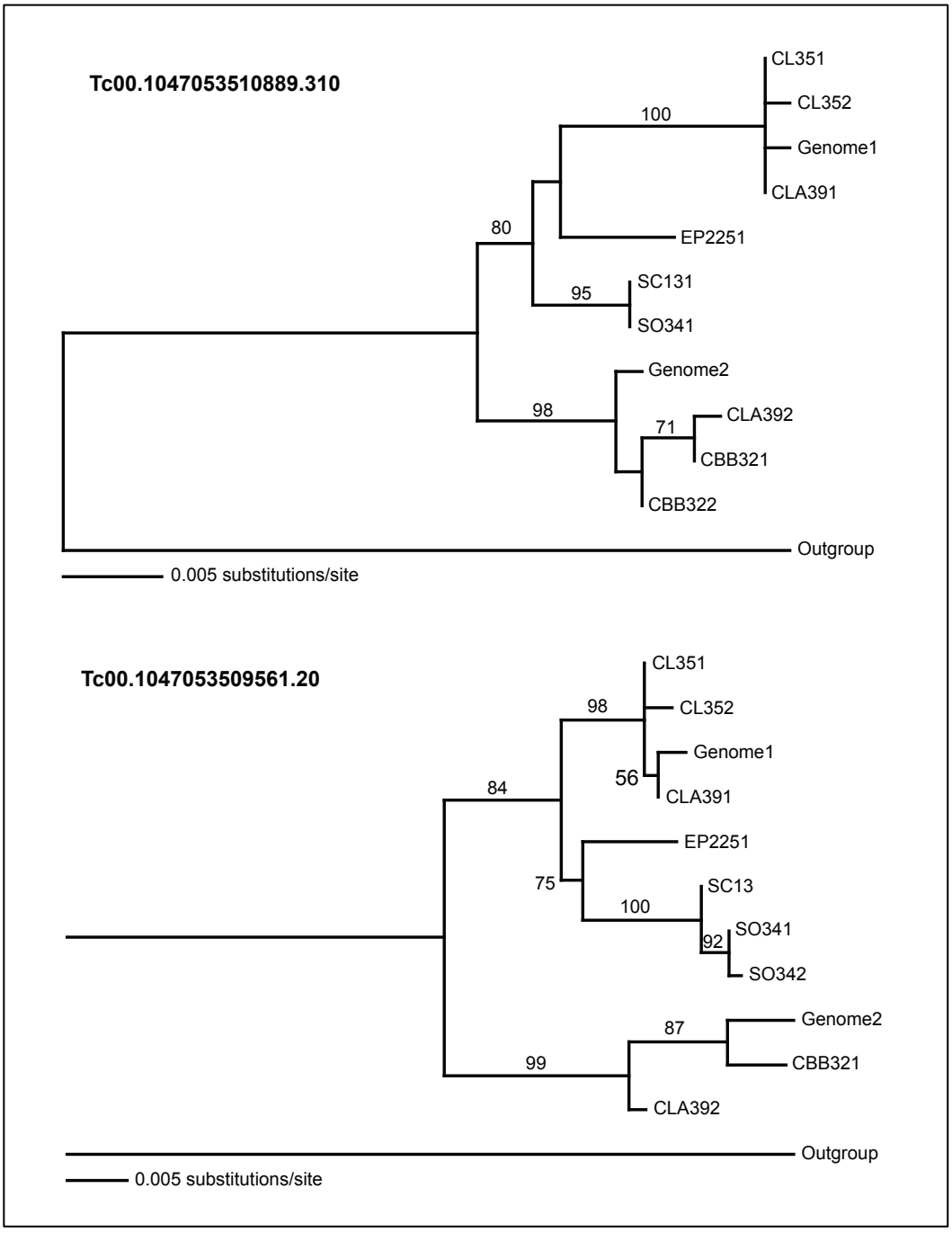


Figure S2-1 continued

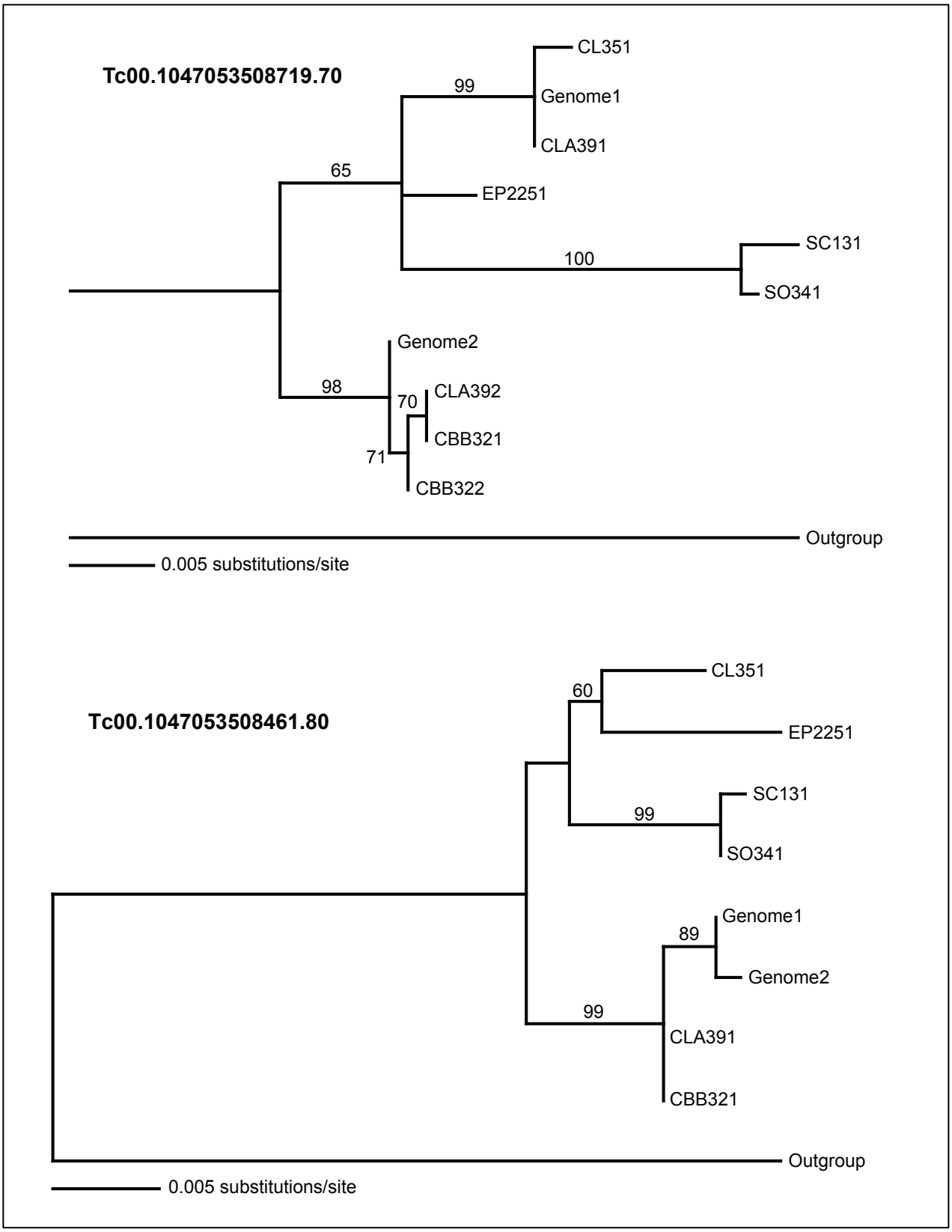


Figure S2-1 continued

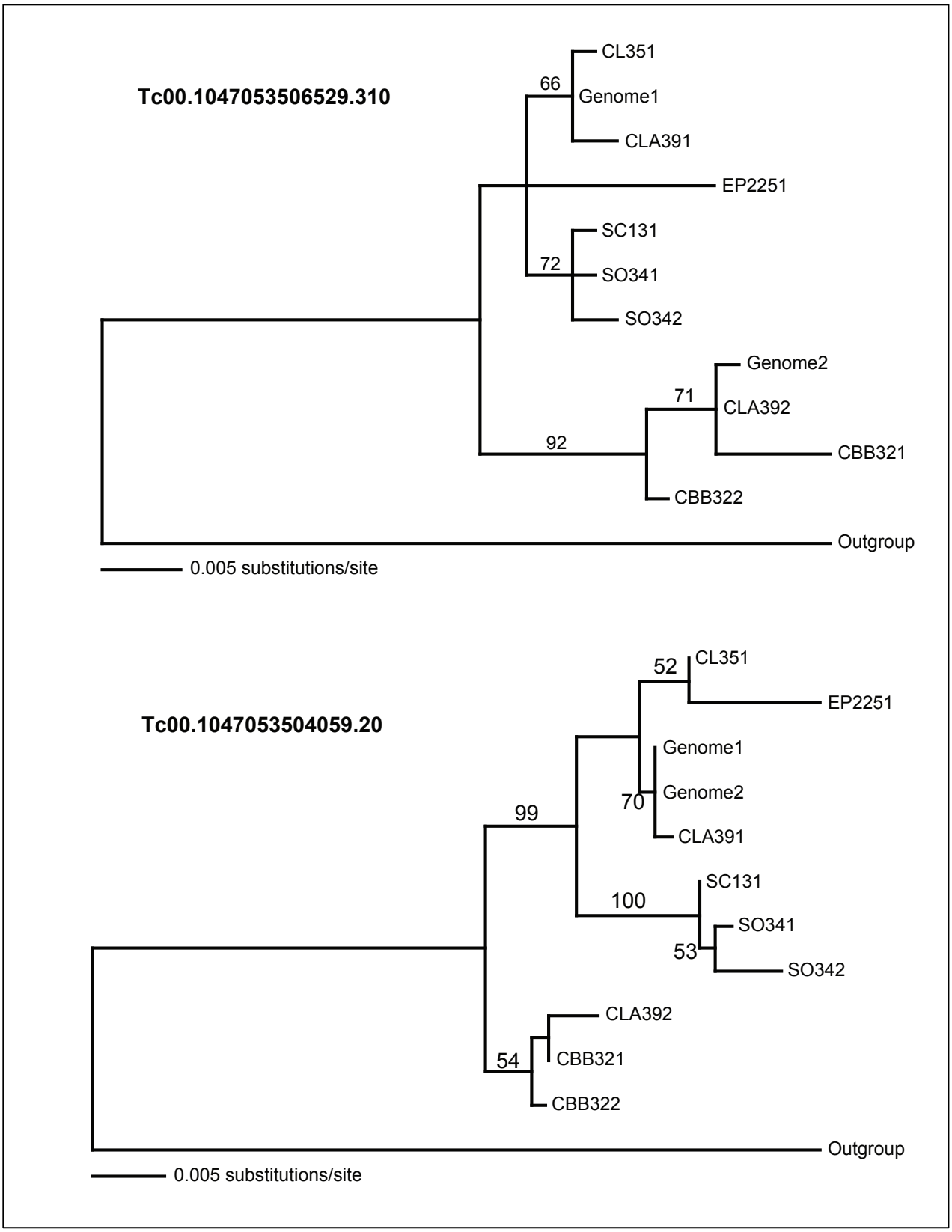


Figure S2-1 continued

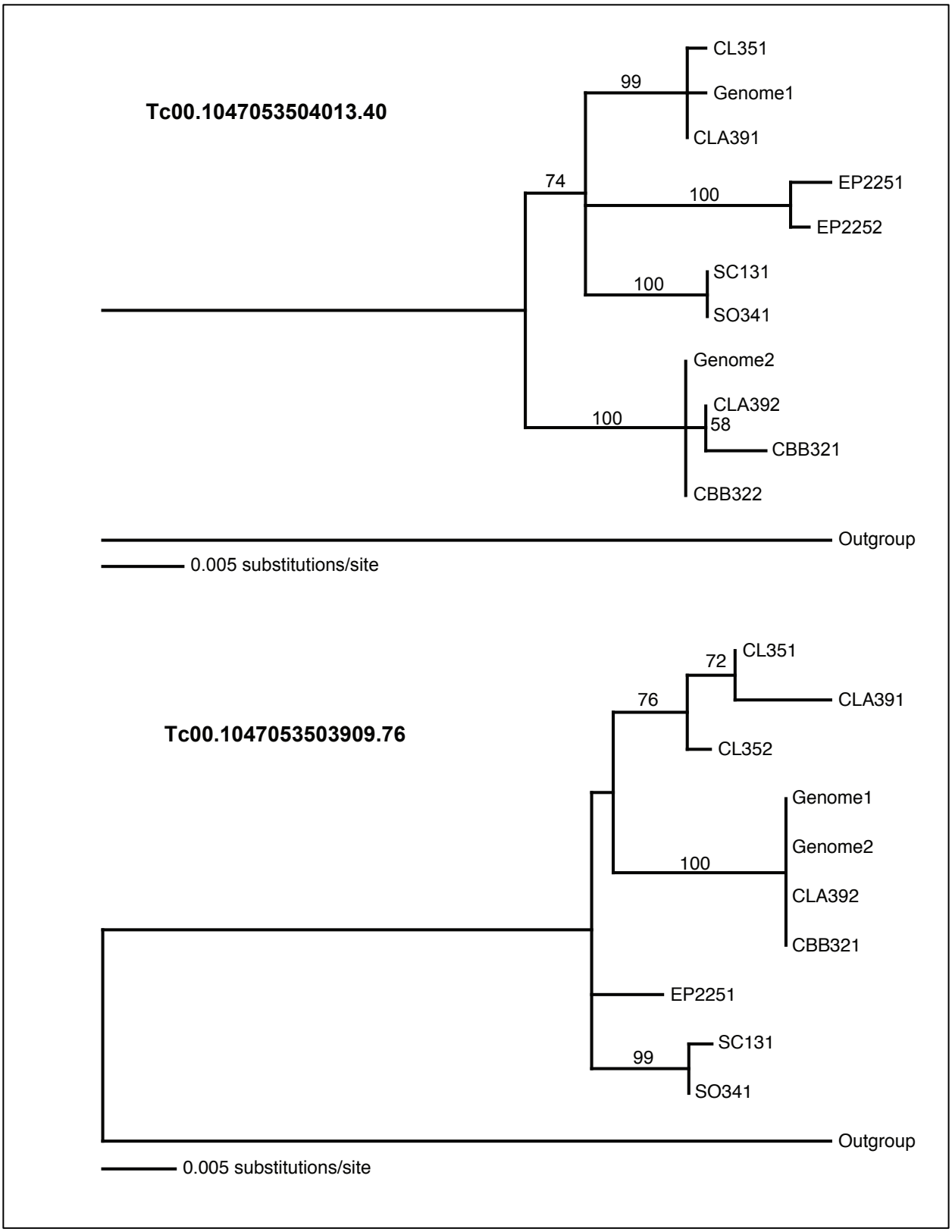


Figure S2-1 continued

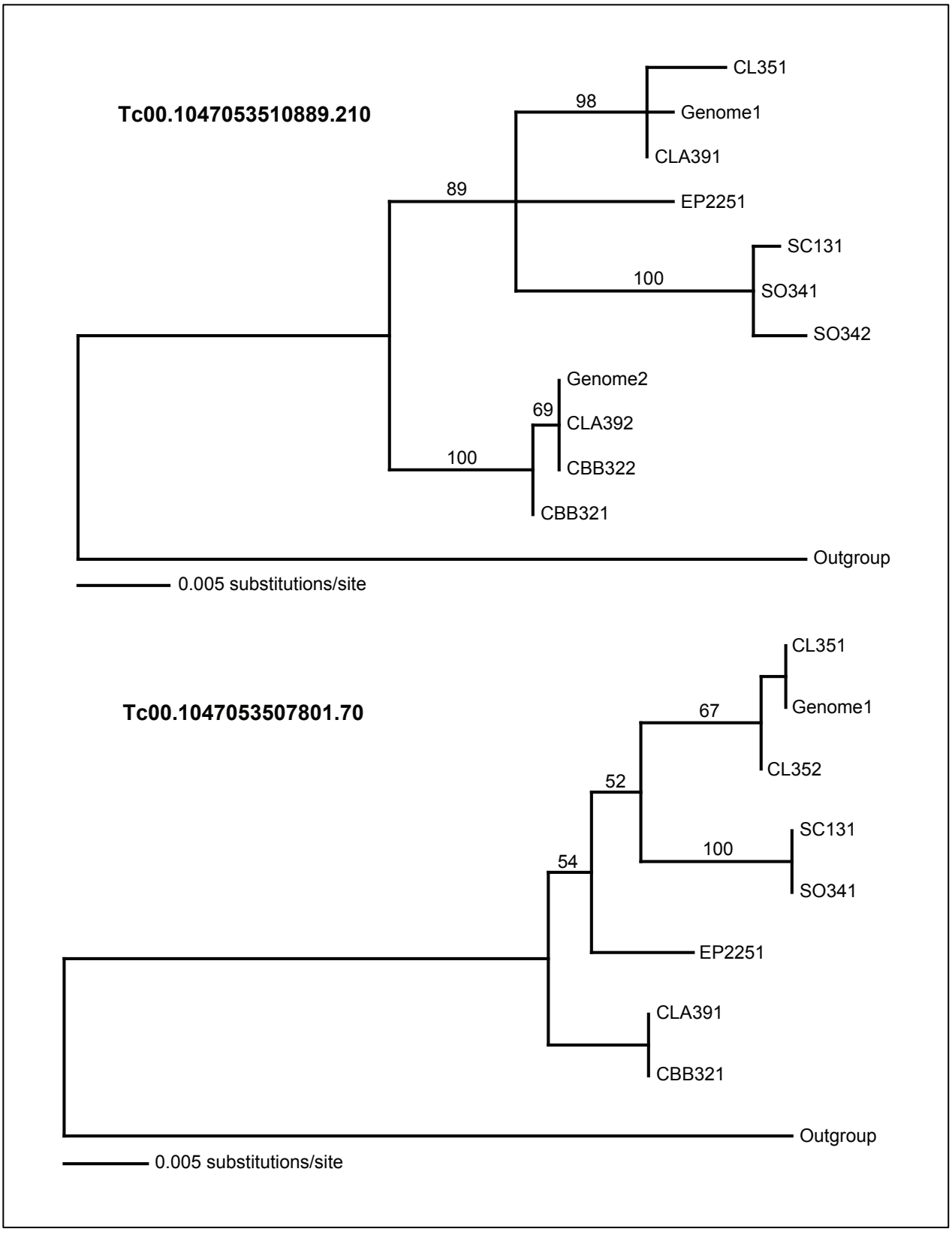


Figure S2-1 continued

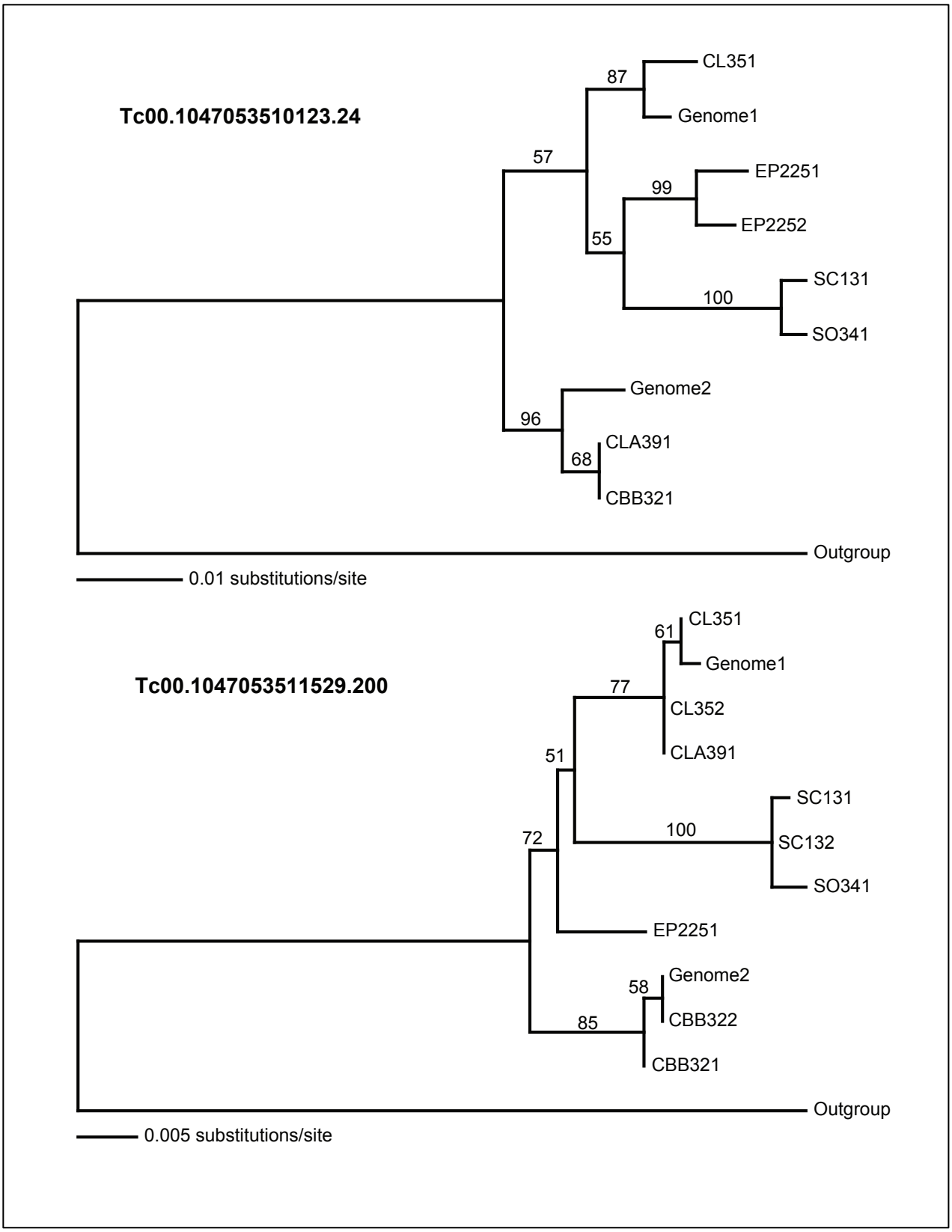


Figure S2-1 continued

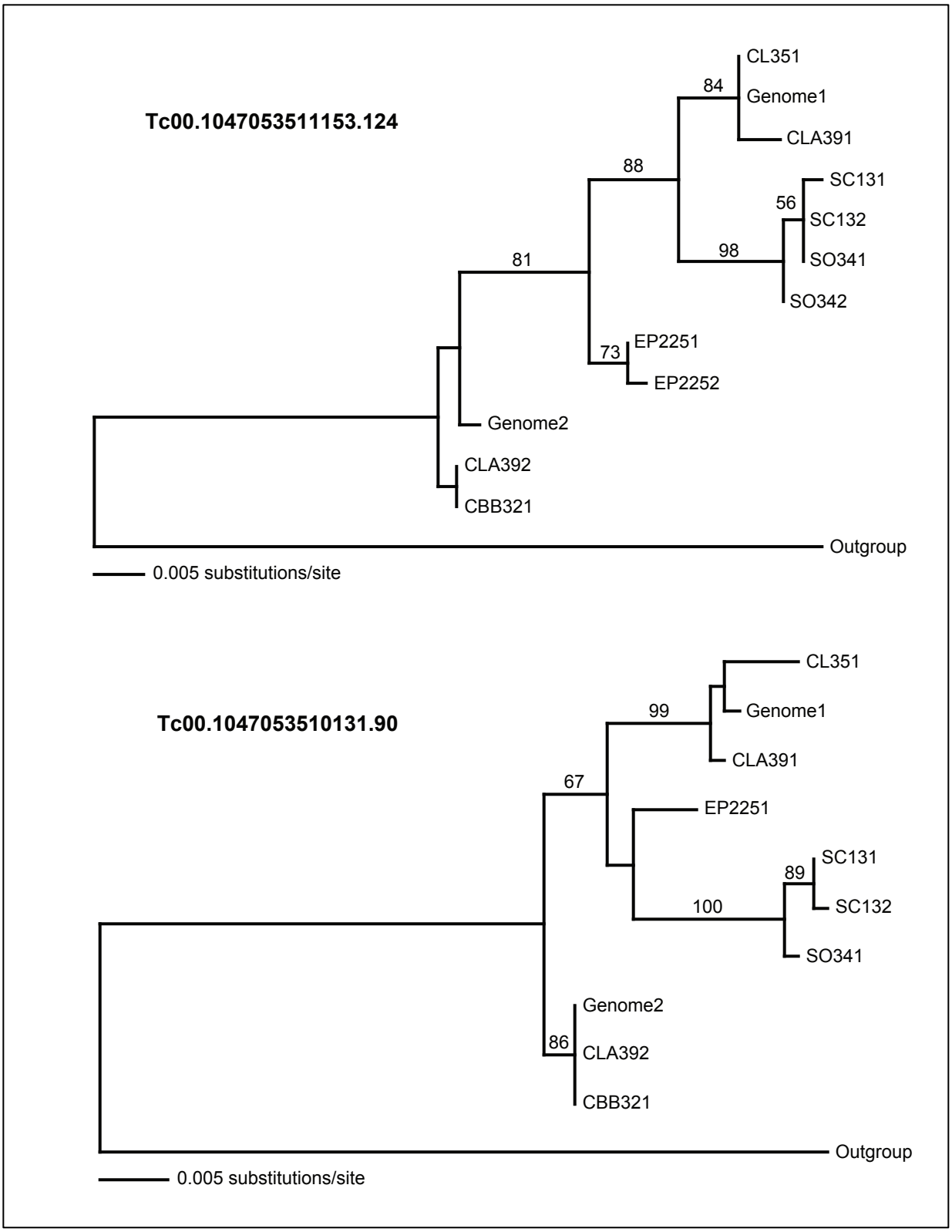


Figure S2-1 continued

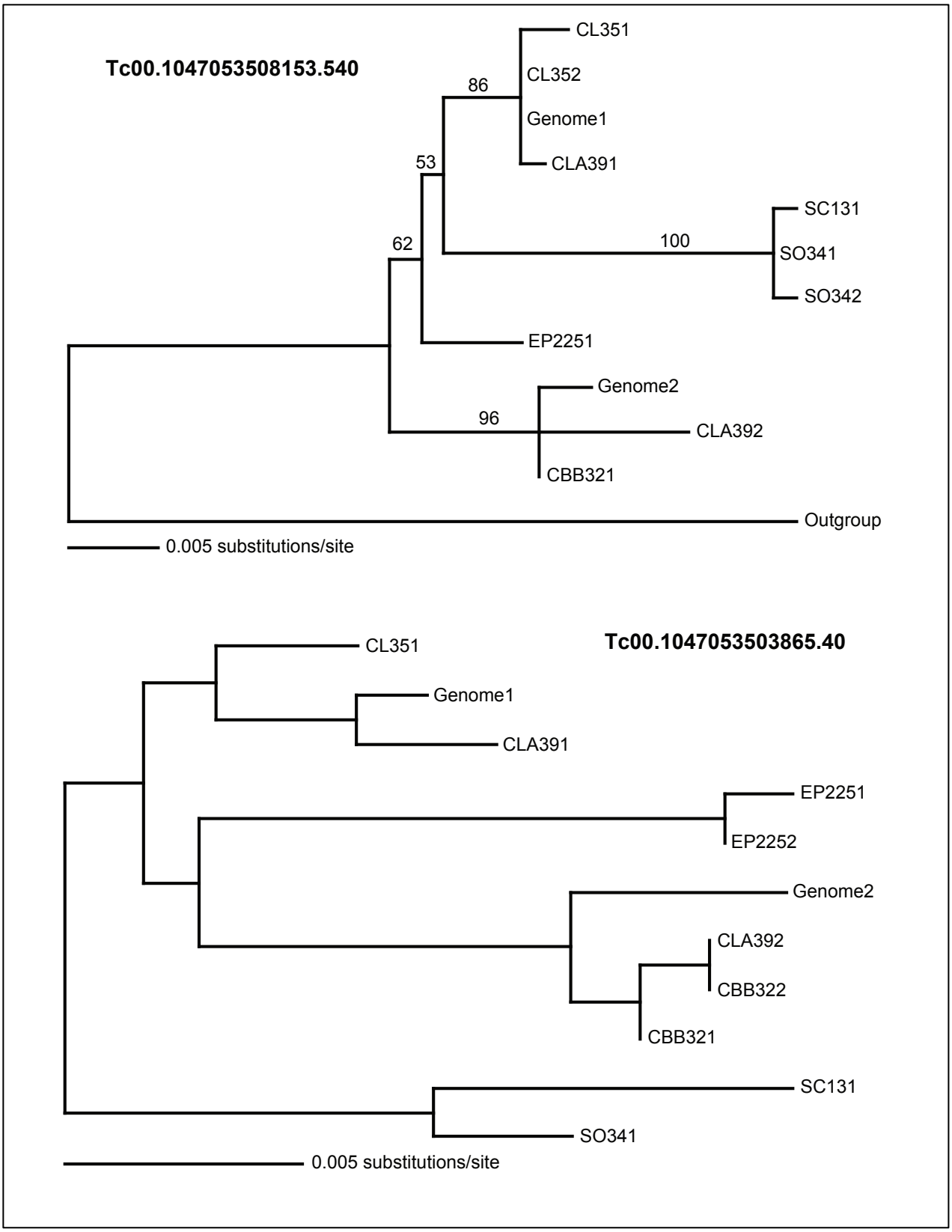


Figure S2-1 continued

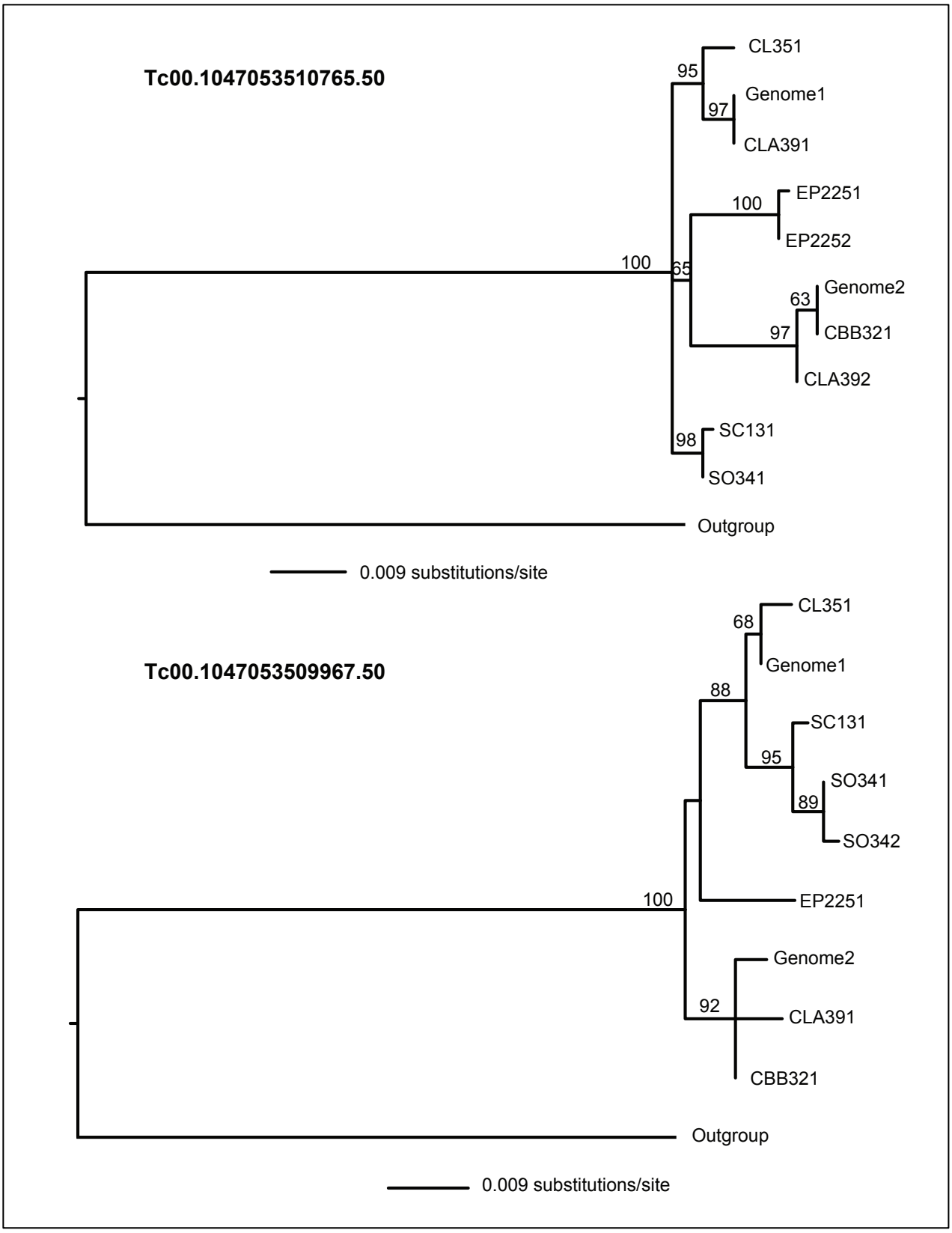


Figure S2-1 continued

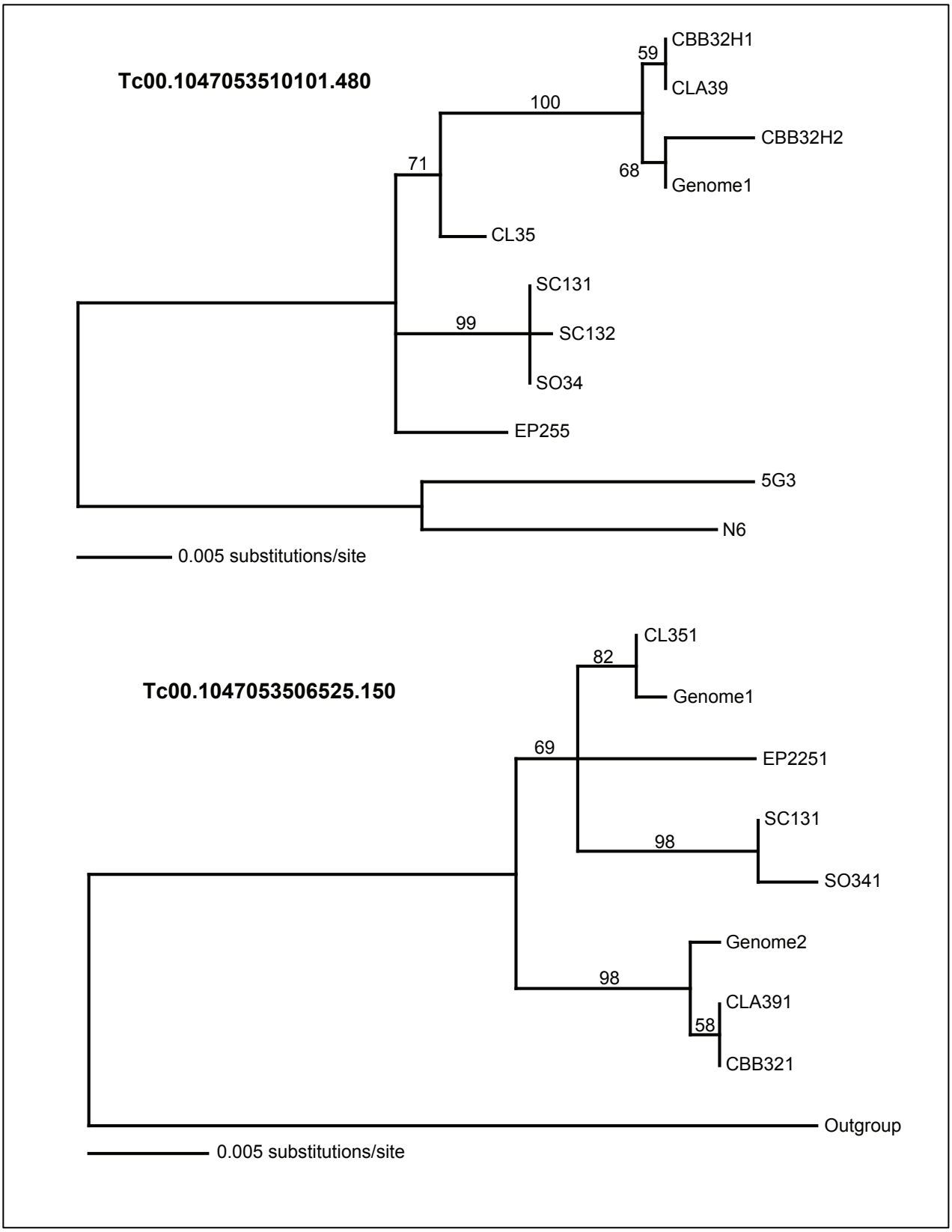


Figure S2-1 continued

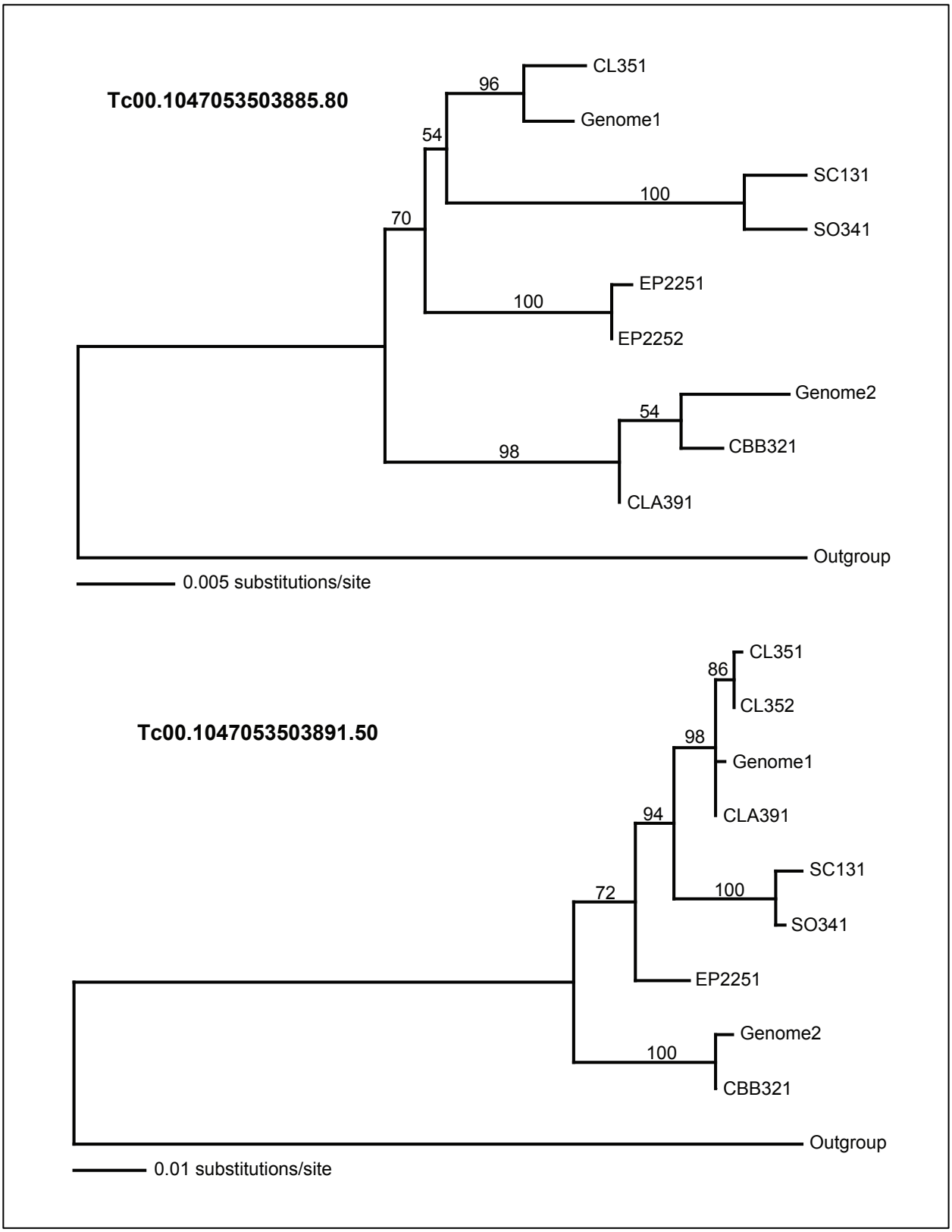


Figure S2-1 continued

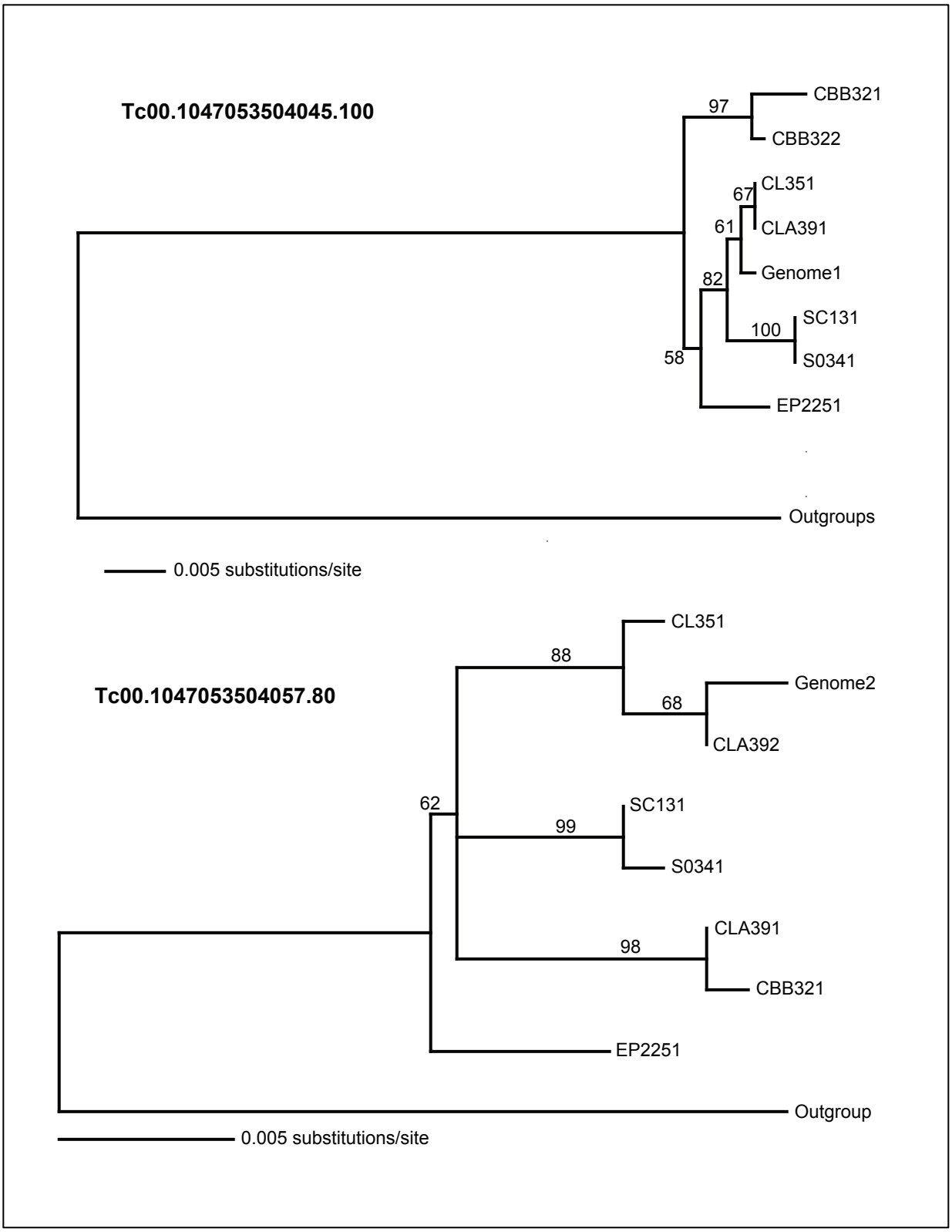


Figure S2-1 continued

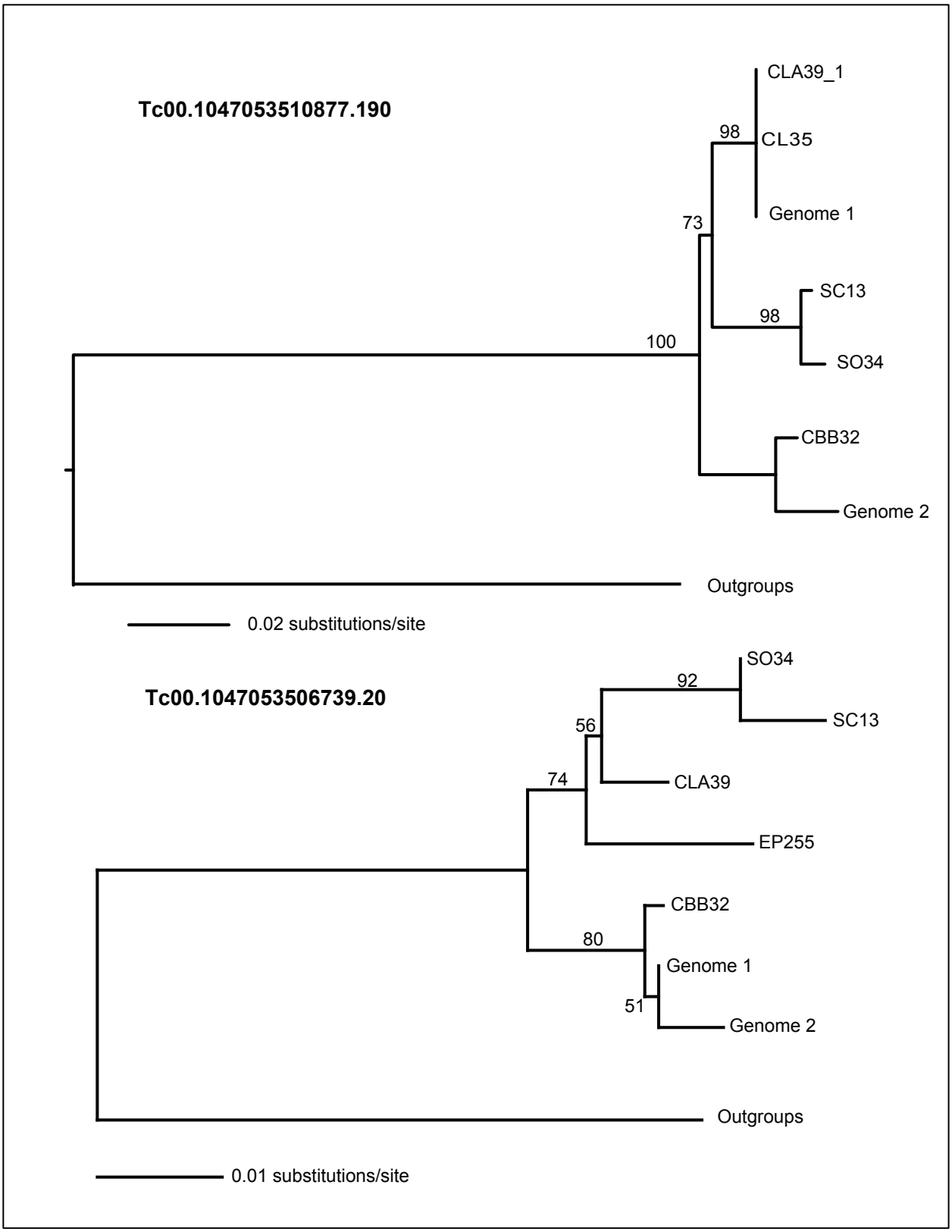


Figure S2-1 continued

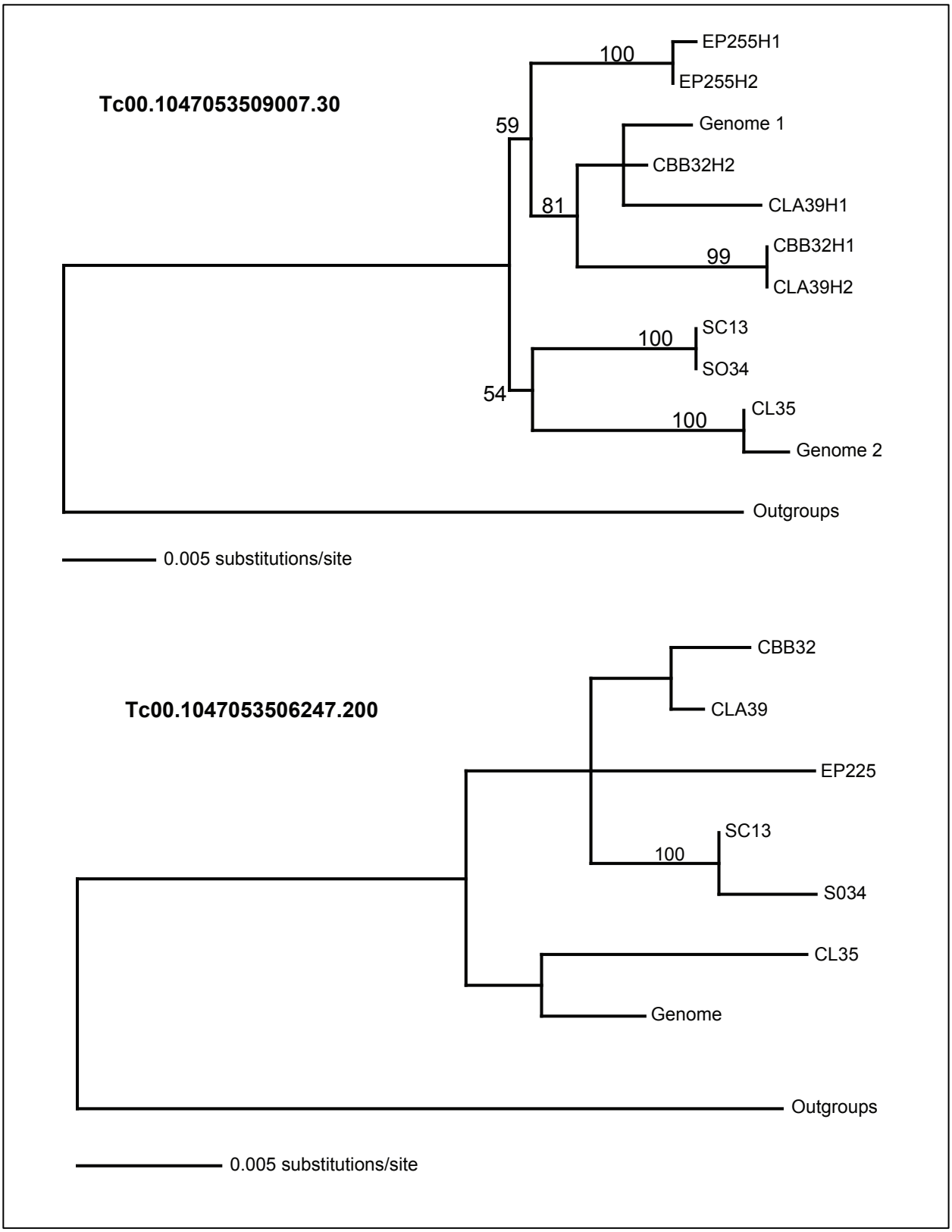


Figure S2-1 continued

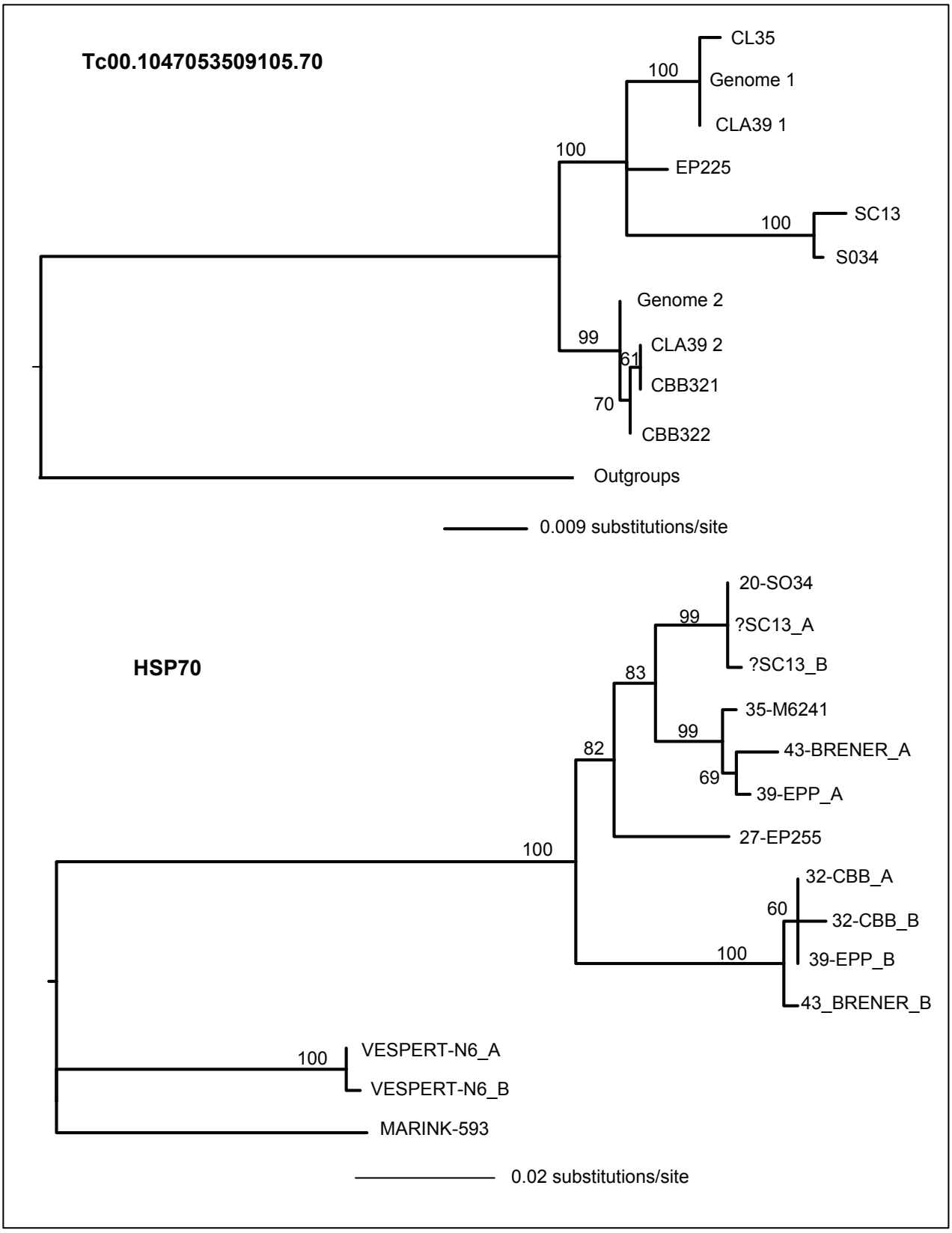

Tc00.1047053509105.70
CL35
100
Genome 1
CLA39 1
100
EP225
100
SC13
S034
Genome 2
99
CLA39 2
61
CBB321
70
CBB322
Outgroups
0.009 substitutions/site
HSP70
20-SO34
99
?SC13_A
?SC13_B
83
35-M6241
99
43-BRENER_A
82
69
39-EPP_A
27-EP255
100
32-CBB_A
60
32-CBB_B
100
39-EPP_B
43_BRENER_B
100
VESPERT-N6_A
VESPERT-N6_B
MARINK-593
0.02 substitutions/site

Figure S2-2. Divergence times for main DTU clades of *T. cruzi* using nuclear loci with the strict clock model. Data set consists of concatenated nuclear loci (22) for which the molecular clock was not rejected (Table S4), and had a homolog in *T. brucei*. Codes (Strains): TcI (SO34 cl4 & SC13), TcIV (EP 255), TcII (CBB cl3), TcIII (M6241 cl6), TcV Hybrid (SO3 cl5), TcVI Hybrid (CL Brener) (See Table 2-1). Scale bar in millions of years ago.

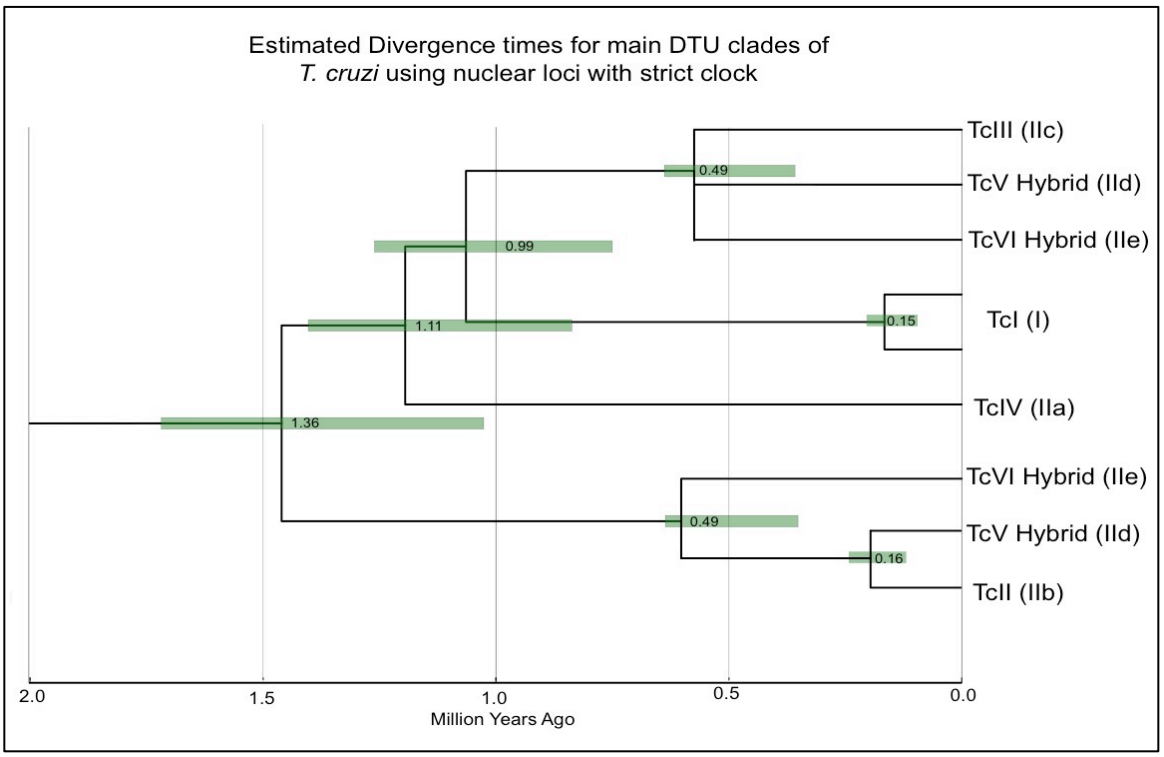


Figure S3-1. Maximum likelihood phylogenetic tree of Tc00.1047053506529.310 nuclear loci. Maximum likelihood bootstraps higher than 70% are shown above major nodes.

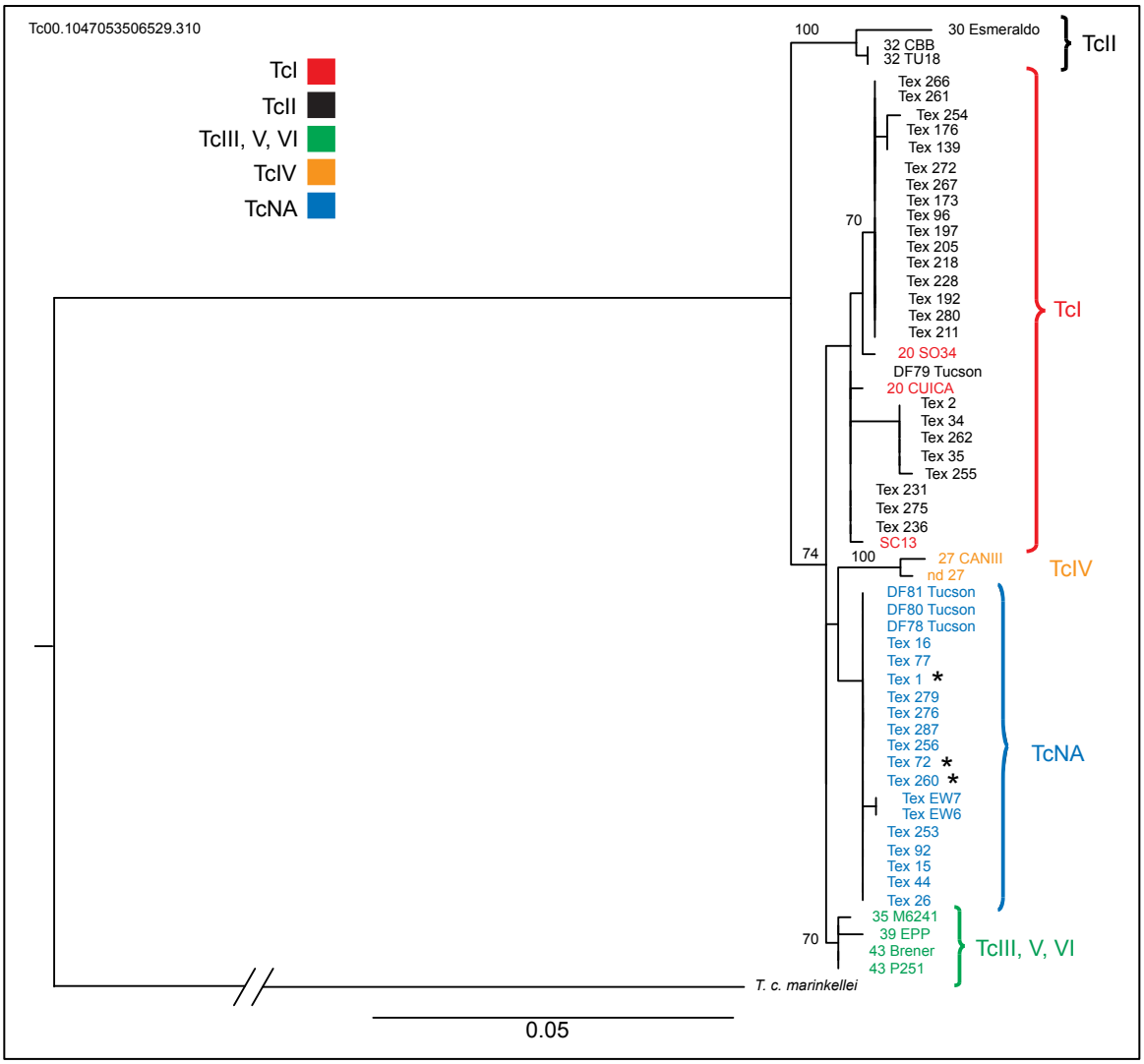


Figure S3-2. Maximum likelihood phylogenetic tree of DHFR nuclear loci. Maximum likelihood bootstraps higher than 70% are shown above major nodes.

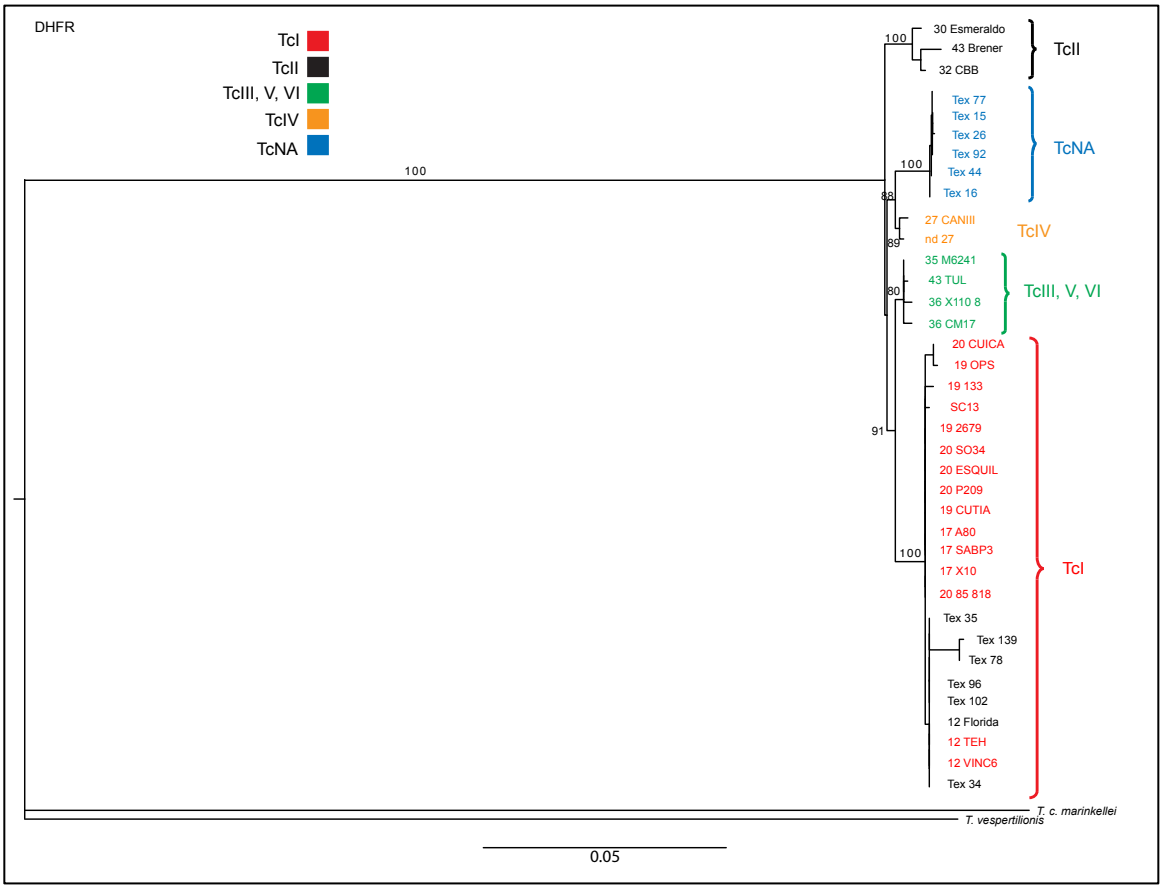


Figure S3-3. Maximum likelihood phylogenetic tree of MSH2 nuclear loci. Maximum likelihood bootstraps higher than 70% are shown above major nodes.

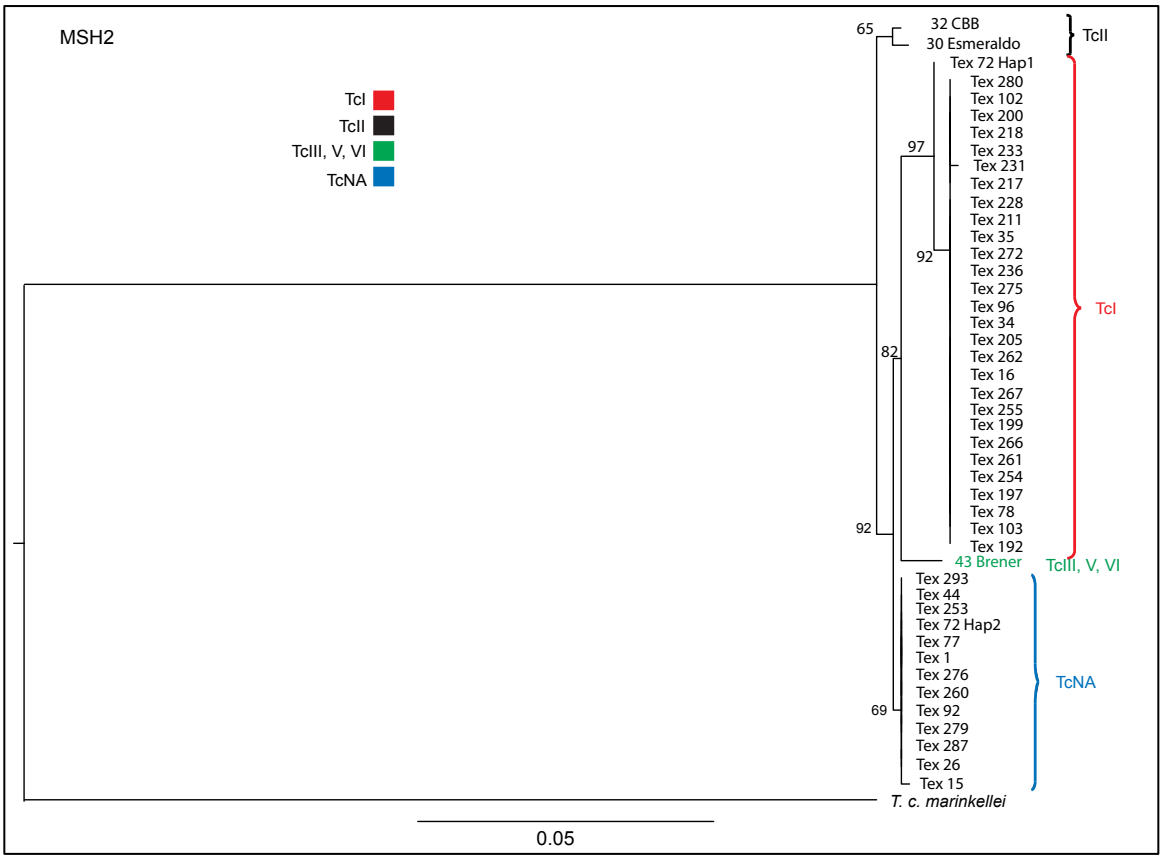


Table S4-1. *L.major* gene codes for orthologous protein predicted to be under PS with models M8 versus M8a with 4 and 3 taxa.

4 taxa analysis		3 taxa analysis	
Gene code	**q value**	**Gene code**	**q value**
LmjF.10.0470	0.000119474	LmjF.35.0520	8.65E-64
LmjF.27.0970	0.000181745	LmjF.26.2170	1.11E-53
LmjF.25.0980	0.000289732	LmjF.30.1240	3.29E-38
LmjF.35.1140	0.001100642	LmjF.07.0830	5.51E-33
LmjF.08.0640	0.002988036	LmjF.14.1120	2.06E-32
LmjF.34.2700	0.003170225	LmjF.31.1490	5.96E-31
LmjF.36.3180	0.004186148	LmjF.02.0300	1.17E-26
LmjF.34.3630	0.007910492	LmjF.28.3010	2.65E-26
LmjF.11.1290	0.02310107	LmjF.17.0180	3.05E-26
LmjF.25.0670	0.04679435	LmjF.14.1110	1.26E-20
LmjF.25.1960	0.04889889	LmjF.35.0500	1.32E-19
LmjF.08.0850	0.04889889	LmjF.31.1440	2.36E-19
LmjF.10.0610	0.06188664	LmjF.34.2410	6.17E-18
LmjF.16.0720	0.07394614	LmjF.04.0760	1.58E-17
LmjF.27.0290	0.07810346	LmjF.22.1320	8.10E-17
LmjF.25.1950	0.1041355	LmjF.32.2220	6.61E-16
LmjF.05.0490	0.1262489	LmjF.05.0240	1.26E-15
LmjF.30.1410	0.154447	LmjF.24.0450	8.22E-15
LmjF.15.1080	0.154447	LmjF.32.1040	1.08E-14
LmjF.10.0530	0.178643	LmjF.11.0070	3.95E-14
LmjF.31.2730	0.1856987	LmjF.14.1440	6.84E-14
LmjF.01.0340	0.1904062	LmjF.28.3030	2.26E-13
LmjF.15.0870	0.1904062	LmjF.35.0740	4.73E-12
LmjF.27.1160	0.256059	LmjF.34.1920	1.19E-11
LmjF.18.1650	0.262165	LmjF.15.0195	1.00E-10
LmjF.14.1430	0.2905996	LmjF.36.6290	3.96E-10
LmjF.30.2850	0.3303373	LmjF.22.1210	6.52E-10
LmjF.24.0300	0.3642688	LmjF.29.2560	6.52E-10
LmjF.36.1870	0.480054	LmjF.26.2030	6.52E-10
LmjF.26.1360	0.480054	LmjF.11.1220	9.78E-10
LmjF.35.4810	0.480054	LmjF.31.2520	1.15E-09
LmjF.33.0530	0.480054	LmjF.36.0110	3.76E-09
LmjF.26.0950	0.480054	LmjF.21.0620	5.13E-09
LmjF.35.0460	0.4974024	LmjF.28.1430	1.10E-08
LmjF.24.0340	0.4974024	LmjF.17.0080	2.53E-08
LmjF.04.0460	0.4974024	LmjF.17.0200	1.15E-07
LmjF.06.0090	0.4974024	LmjF.02.0420	1.17E-07

Table S4-1 continued

LmjF.19.0520	0.4974024	LmjF.34.2560	1.86E-07
LmjF.09.0320	0.4992333	LmjF.03.0270	4.56E-07
LmjF.36.0950	0.5188991	LmjF.23.0220	5.51E-07
LmjF.31.2760	0.5188991	LmjF.35.3770	8.04E-07
LmjF.24.0310	0.5191967	LmjF.21.1170	1.49E-06
LmjF.14.0910	0.564447	LmjF.10.0470	1.49E-06
LmjF.30.1610	0.564447	LmjF.16.1660	1.93E-06
LmjF.28.2830	0.564447	LmjF.26.1665	3.59E-06
LmjF.25.2050	0.564447	LmjF.08.0390	4.02E-06
LmjF.35.4550	0.575855	LmjF.13.0850	9.50E-06
LmjF.32.0320	0.6001579	LmjF.20.1020	1.48E-05
LmjF.35.3630	0.6166127	LmjF.36.1380	1.71E-05
LmjF.32.1820	0.6166127	LmjF.14.0740	1.79E-05
LmjF.36.4180	0.6166127	LmjF.28.1070	4.08E-05
LmjF.26.1210	0.6166127	LmjF.36.1370	5.20E-05
LmjF.36.2570	0.6254072	LmjF.36.1395	5.35E-05
LmjF.34.0970	0.6432122	LmjF.31.2350	5.35E-05
LmjF.35.0840	0.6432122	LmjF.08.0640	6.86E-05
LmjF.10.0960	0.6432122	LmjF.05.1160	0.000135087
LmjF.25.0690	0.6485901	LmjF.09.0980	0.000170623
LmjF.02.0100	0.6508502	LmjF.23.0640	0.00033404
LmjF.13.0510	0.6615638	LmjF.07.0120	0.000407181
LmjF.16.1050	0.6711036	LmjF.15.1270	0.000452758
LmjF.30.0430	0.6711036	LmjF.34.1720	0.000585434
LmjF.06.1200	0.6887683	LmjF.04.0860	0.000585434
LmjF.33.1290	0.6887683	LmjF.17.0350	0.00061026
LmjF.28.2410	0.6914201	LmjF.26.1460	0.0007039
LmjF.11.0670	0.7004571	LmjF.31.0450	0.0007039
LmjF.36.2240	0.7036367	LmjF.32.2270	0.000720389
LmjF.36.0580	0.7036367	LmjF.31.2540	0.000807373
LmjF.25.1590	0.7175936	LmjF.08.0820	0.000986358
LmjF.02.0430	0.7202588	LmjF.15.0530	0.001133838
LmjF.30.1300	0.7202588	LmjF.04.0830	0.001333777
LmjF.29.2090	0.7202588	LmjF.36.1840	0.001471398
LmjF.31.0920	0.7202588	LmjF.35.1140	0.001471398
LmjF.05.0100	0.8017501	LmjF.30.2010	0.001610629
LmjF.30.2880	0.8736232	LmjF.01.0800	0.00170958
LmjF.36.0660	0.8736232	LmjF.27.1580	0.001882698
LmjF.19.1345	0.8736232	LmjF.21.0460	0.002189177
LmjF.29.0820	0.8736232	LmjF.15.1480	0.002283609

Table S4-1 continued

LmjF.27.0670	0.8866383	LmjF.19.1340	0.002331598
LmjF.28.0650	0.9348933	LmjF.13.0980	0.00245517
LmjF.09.0240	0.9695123	LmjF.12.0150	0.002642252
LmjF.26.2420	0.9695123	LmjF.36.5250	0.002705668
LmjF.31.2780	0.9695123	LmjF.35.1450	0.002705668
LmjF.35.2160	0.9695123	LmjF.31.0040	0.003066038
LmjF.33.0110	0.986604	LmjF.04.0620	0.003995976
LmjF.06.1080	1	LmjF.32.0715	0.004380654
LmjF.34.0140	1	LmjF.09.0760	0.005741259
LmjF.34.2670	1	LmjF.11.1240	0.005854037
LmjF.01.0810	1	LmjF.07.0180	0.006257686
LmjF.22.0410	1	LmjF.19.1365	0.006586592
LmjF.31.3110	1	LmjF.26.1560	0.007532029
LmjF.06.0530	1	LmjF.08.0270	0.007653778
LmjF.08.0800	1	LmjF.04.0700	0.007897944
LmjF.28.1300	1	LmjF.34.0500	0.008426663
LmjF.13.0680	1	LmjF.08.0310	0.008630766
LmjF.24.1470	1	LmjF.11.0660	0.008678039
LmjF.15.0730	1	LmjF.26.1430	0.0100614
LmjF.22.0220	1	LmjF.31.0800	0.01018534
LmjF.23.0450	1	LmjF.22.0940	0.0103779
LmjF.15.0160	1	LmjF.09.0220	0.01038884
LmjF.35.3660	1	LmjF.07.0340	0.01051552
LmjF.16.0820	1	LmjF.35.4840	0.010655
LmjF.30.1500	1	LmjF.21.0630	0.010655
LmjF.02.0620	1	LmjF.24.1790	0.0116249
LmjF.14.0990	1	LmjF.28.2840	0.0116249
LmjF.36.6760	1	LmjF.23.0900	0.01298018
LmjF.29.0260	1	LmjF.13.0670	0.01340405
LmjF.13.0420	1	LmjF.33.1240	0.01420733
LmjF.27.0500	1	LmjF.23.1170	0.01453126
LmjF.05.0285	1	LmjF.19.0940	0.01453126
LmjF.01.0020	1	LmjF.28.1170	0.01547316
LmjF.11.1080	1	LmjF.12.0590	0.01654568
LmjF.28.1460	1	LmjF.26.1950	0.0183551
LmjF.28.2370	1	LmjF.33.2960	0.01937963
LmjF.23.0540	1	LmjF.02.0470	0.02078963
LmjF.04.0240	1	LmjF.32.3690	0.02079568
LmjF.01.0230	1	LmjF.07.0740	0.02146531
LmjF.26.2340	1	LmjF.33.2020	0.02183339

Table S4-1 continued

LmjF.26.2330	1	LmjF.34.1680	0.02183339
LmjF.30.3530	1	LmjF.21.0530	0.02183339
LmjF.08.0320	1	LmjF.16.0720	0.02337308
LmjF.34.4620	1	LmjF.25.1570	0.02440284
LmjF.19.1420	1	LmjF.19.0985	0.02468857
LmjF.31.2390	1	LmjF.06.0080	0.02468857
LmjF.30.0230	1	LmjF.14.0340	0.02468857
LmjF.09.1020	1	LmjF.13.0780	0.02477814
LmjF.27.1500	1	LmjF.20.0790	0.02477814
LmjF.18.0680	1	LmjF.33.0650	0.02522151
LmjF.34.0080	1	LmjF.10.1100	0.02897935
LmjF.10.0630	1	LmjF.33.1290	0.03161935
LmjF.34.0740	1	LmjF.35.0030	0.03490508
LmjF.35.2060	1	LmjF.10.1080	0.03511907
LmjF.33.1860	1	LmjF.16.0500	0.03576102
LmjF.07.0150	1	LmjF.06.0830	0.03576102
LmjF.36.3010	1	LmjF.35.1160	0.03576102
LmjF.26.2480	1	LmjF.24.0360	0.03595607
LmjF.10.1260	1	LmjF.04.0250	0.03696703
LmjF.32.0510	1	LmjF.20.1070	0.03746277
LmjF.03.0380	1	LmjF.29.0820	0.04317757
LmjF.09.0780	1	LmjF.19.0520	0.04518344
LmjF.31.1445	1	LmjF.36.0910	0.04518344
LmjF.33.1890	1	LmjF.07.1060	0.04518344
LmjF.16.0700	1	LmjF.29.0110	0.04531431
LmjF.26.1410	1	LmjF.07.0730	0.04639491
LmjF.07.0730	1	LmjF.12.0470	0.05012202
LmjF.30.2410	1	LmjF.12.0190	0.05050072
LmjF.32.3880	1	LmjF.18.1420	0.05149808
LmjF.11.0610	1	LmjF.36.5610	0.05183224
LmjF.17.1160	1	LmjF.27.0490	0.05245799
LmjF.17.1310	1	LmjF.14.1100	0.05626867
LmjF.35.0310	1	LmjF.05.0210	0.05871332
LmjF.29.1500	1	LmjF.10.0500	0.05871332
LmjF.18.1390	1	LmjF.18.0850	0.05884466
LmjF.30.0050	1	LmjF.01.0740	0.05951544
LmjF.31.1000	1	LmjF.04.0840	0.05951544
LmjF.09.0300	1	LmjF.05.0690	0.05960021
LmjF.18.1260	1	LmjF.23.0355	0.06041771
LmjF.20.1380	1	LmjF.28.0440	0.06047312

Table S4-1 continued

LmjF.14.0590	1	LmjF.02.0500	0.06287885
LmjF.32.3610	1	LmjF.05.0270	0.06379769
LmjF.35.4220	1	LmjF.16.1090	0.06563315
LmjF.17.1240	1	LmjF.30.0600	0.06663484
LmjF.32.1830	1	LmjF.31.2370	0.06663484
LmjF.33.3160	1	LmjF.02.0340	0.06806765
LmjF.35.5020	1	LmjF.28.2445	0.06957417
LmjF.17.0730	1	LmjF.31.3020	0.06968746
LmjF.32.2490	1	LmjF.07.0170	0.07054921
LmjF.24.0580	1	LmjF.06.0980	0.07175721
LmjF.15.0195	1	LmjF.17.1440	0.07496785
LmjF.23.1130	1	LmjF.13.1470	0.07496785
LmjF.07.1035	1	LmjF.07.0890	0.07496785
LmjF.28.0330	1	LmjF.31.1800	0.0750717
LmjF.13.1500	1	LmjF.22.0640	0.07544393
LmjF.33.2090	1	LmjF.36.2020	0.0767676
LmjF.11.1270	1	LmjF.11.0670	0.07716011
LmjF.32.3720	1	LmjF.36.4720	0.07806017
LmjF.31.1310	1	LmjF.20.0470	0.07810844
LmjF.16.1060	1	LmjF.20.0770	0.08043836
LmjF.25.1640	1	LmjF.15.0660	0.08043836
LmjF.36.4120	1	LmjF.31.3190	0.08043836
LmjF.28.2970	1	LmjF.23.0210	0.08043836
LmjF.07.1160	1	LmjF.05.0820	0.08043836
LmjF.17.1020	1	LmjF.22.1460	0.08054826
LmjF.11.0170	1	LmjF.32.1820	0.08155762
LmjF.27.2130	1	LmjF.13.1400	0.08156344
LmjF.28.2080	1	LmjF.34.3570	0.08156344
LmjF.28.0970	1	LmjF.32.3670	0.08218019
LmjF.18.1300	1	LmjF.05.0770	0.08218019
LmjF.27.2360	1	LmjF.10.1310	0.08287496
LmjF.33.2940	1	LmjF.30.0230	0.08420245
LmjF.35.3470	1	LmjF.14.1180	0.0871529
LmjF.28.0080	1	LmjF.13.1410	0.08789552
LmjF.36.1300	1	LmjF.30.2060	0.08789552
LmjF.30.1090	1	LmjF.20.0040	0.08814285
LmjF.15.0420	1	LmjF.04.1030	0.08867569
LmjF.18.0290	1	LmjF.31.2800	0.0889179
LmjF.36.6970	1	LmjF.27.1160	0.08994086
LmjF.28.1480	1	LmjF.27.0970	0.09049902

Table S4-1 continued

LmjF.14.1180	1	LmjF.24.0680	0.09049902
LmjF.30.3070	1	LmjF.02.0700	0.09188452
LmjF.23.1700	1	LmjF.19.0260	0.09188452
LmjF.15.1190	1	LmjF.05.0380	0.092169
LmjF.04.0270	1	LmjF.01.0640	0.09460169
LmjF.14.0090	1	LmjF.36.4180	0.095947
LmjF.29.0620	1	LmjF.05.0300	0.09945559
LmjF.32.1900	1	LmjF.35.2810	0.09945559
LmjF.04.0260	1	LmjF.26.1420	0.1025905
LmjF.31.1140	1	LmjF.04.0500	0.1034096
LmjF.35.4650	1	LmjF.34.4020	0.1034096
LmjF.23.1510	1	LmjF.18.1090	0.1040825
LmjF.13.1570	1	LmjF.26.2320	0.1040825
LmjF.26.1810	1	LmjF.04.1110	0.1040825
LmjF.17.0340	1	LmjF.35.0670	0.1041795
LmjF.27.1120	1	LmjF.07.0280	0.1086335
LmjF.35.2400	1	LmjF.21.0125	0.1101394
LmjF.35.1060	1	LmjF.10.0610	0.113496
LmjF.32.2120	1	LmjF.25.2050	0.113496
LmjF.23.0940	1	LmjF.19.1420	0.1135026
LmjF.23.0355	1	LmjF.22.1440	0.1135026
LmjF.28.2140	1	LmjF.13.0920	0.1181491
LmjF.32.1495	1	LmjF.21.0825	0.1203274
LmjF.28.2940	1	LmjF.36.0880	0.1234281
		LmjF.24.0480	0.1234281
		LmjF.29.0500	0.1262803
		LmjF.20.1610	0.1265076
		LmjF.05.0670	0.1265076
		LmjF.25.0450	0.1283487
		LmjF.12.0320	0.1283487
		LmjF.11.0340	0.1283487
		LmjF.19.0220	0.1293272
		LmjF.09.0890	0.1345429
		LmjF.18.1180	0.1348085
		LmjF.19.1030	0.1409708
		LmjF.10.0220	0.1415342
		LmjF.29.1480	0.1425483
		LmjF.31.1270	0.14923
		LmjF.30.0430	0.1496027
		LmjF.33.1035	0.1523405

Table S4-1 continued

		LmjF.04.0460	0.1542269
		LmjF.26.1110	0.1542269
		LmjF.22.0330	0.1542269
		LmjF.34.1170	0.1572722
		LmjF.16.0620	0.1587867
		LmjF.32.2400	0.1587867
		LmjF.11.0640	0.1589821
		LmjF.15.1180	0.1624332
		LmjF.11.0840	0.1624332
		LmjF.18.0680	0.1643436
		LmjF.36.2570	0.1643436
		LmjF.16.1015	0.1643436
		LmjF.04.0430	0.1656972
		LmjF.33.0760	0.1656972
		LmjF.35.3520	0.1656972
		LmjF.15.1080	0.1726695
		LmjF.15.0050	0.1745113
		LmjF.30.2330	0.1745113
		LmjF.34.2780	0.175779
		LmjF.27.0510	0.1764074
		LmjF.21.1520	0.1764074
		LmjF.28.1630	0.1764074
		LmjF.34.2530	0.1766071
		LmjF.10.1110	0.1766071
		LmjF.16.0390	0.1766071
		LmjF.27.1295	0.1766071
		LmjF.07.0380	0.1766071
		LmjF.36.0840	0.1766071
		LmjF.34.0190	0.1766071
		LmjF.06.0760	0.1766071
		LmjF.09.0880	0.1789538
		LmjF.36.6760	0.1789538
		LmjF.25.1590	0.1803056
		LmjF.23.1180	0.1805665
		LmjF.02.0010	0.1806983
		LmjF.08.0850	0.1835082
		LmjF.20.0140	0.1835082
		LmjF.30.0820	0.1835082
		LmjF.36.2240	0.1835082
		LmjF.13.0900	0.1847601

Table S4-1 continued

		LmjF.02.0040	0.1856053
		LmjF.36.1300	0.1856053
		LmjF.03.0800	0.1868163
		LmjF.29.1520	0.1868163
		LmjF.10.0320	0.1868163
		LmjF.34.2700	0.193422
		LmjF.23.1430	0.200726
		LmjF.19.0630	0.200726
		LmjF.36.0010	0.200726
		LmjF.18.1050	0.200726
		LmjF.26.1210	0.200726
		LmjF.32.0800	0.2016564
		LmjF.13.0700	0.2046248
		LmjF.26.1960	0.2079719
		LmjF.05.0710	0.2087308
		LmjF.28.1800	0.2139954
		LmjF.31.2760	0.2144825
		LmjF.33.3030	0.2151094
		LmjF.17.0440	0.2186892
		LmjF.36.1955	0.2186892
		LmjF.29.1570	0.2186892
		LmjF.25.0780	0.2186892
		LmjF.16.0350	0.2191454
		LmjF.31.0440	0.2205673
		LmjF.36.3530	0.2215994
		LmjF.26.1820	0.2219753
		LmjF.34.0780	0.2219753
		LmjF.10.0810	0.2245861
		LmjF.20.1250	0.2253129
		LmjF.03.0820	0.2253129
		LmjF.10.0380	0.2273351
		LmjF.34.3690	0.2282825
		LmjF.08.0350	0.2282825
		LmjF.25.0470	0.2282825
		LmjF.13.1380	0.2282825
		LmjF.11.0160	0.2282825
		LmjF.34.0140	0.2283571
		LmjF.16.0310	0.2283571
		LmjF.33.1100	0.2305346
		LmjF.19.1080	0.2311108

Table S4-1 continued

		LmjF.21.0370	0.2312245
		LmjF.18.1320	0.2312245
		LmjF.36.1870	0.2312245
		LmjF.29.2430	0.2411953
		LmjF.06.0190	0.2430142
		LmjF.17.1140	0.2430142
		LmjF.14.0910	0.2430142
		LmjF.24.2120	0.2430142
		LmjF.16.0440	0.2432084
		LmjF.11.1290	0.2432084
		LmjF.24.0900	0.2440772
		LmjF.36.2590	0.2453452
		LmjF.02.0100	0.2486428
		LmjF.31.1130	0.2490072
		LmjF.33.1660	0.2491497
		LmjF.28.0650	0.2517129
		LmjF.30.1610	0.2517129
		LmjF.31.1640	0.2531583
		LmjF.24.1550	0.2531583
		LmjF.07.0160	0.2544962
		LmjF.19.1050	0.2544962
		LmjF.29.2090	0.2544962
		LmjF.30.1410	0.2551377
		LmjF.06.0090	0.2581357
		LmjF.31.1330	0.2581357
		LmjF.24.0800	0.2585067
		LmjF.22.1590	0.2621053
		LmjF.30.2160	0.2645244
		LmjF.34.3930	0.2666138
		LmjF.35.0290	0.2667653
		LmjF.29.0930	0.2701703
		LmjF.13.0200	0.2703751
		LmjF.21.1410	0.270805
		LmjF.10.0630	0.270805
		LmjF.07.0760	0.270805
		LmjF.17.1090	0.2733273
		LmjF.25.0670	0.2733273
		LmjF.25.1820	0.2771534
		LmjF.28.1300	0.2771534
		LmjF.22.1100	0.2796266

Table S4-1 continued

		LmjF.02.0150	0.282165
		LmjF.21.0270	0.291854
		LmjF.15.1200	0.2925472
		LmjF.17.0990	0.2981698
		LmjF.21.0550	0.2981698
		LmjF.27.1630	0.2981698
		LmjF.21.1570	0.2985863
		LmjF.31.3030	0.2985863
		LmjF.33.2120	0.2985863
		LmjF.06.0810	0.2985863
		LmjF.28.1930	0.2985863
		LmjF.15.1380	0.2998965
		LmjF.31.1730	0.300367
		LmjF.13.0510	0.303849
		LmjF.35.2360	0.3063062
		LmjF.28.2690	0.3082825
		LmjF.03.0850	0.3095207
		LmjF.21.0120	0.3095207
		LmjF.09.0410	0.3095207
		LmjF.32.3190	0.3095207
		LmjF.31.2460	0.3173527
		LmjF.04.0600	0.3199622
		LmjF.30.1500	0.3233941
		LmjF.26.2420	0.3284365
		LmjF.36.0470	0.3295448
		LmjF.32.1830	0.331535
		LmjF.27.1750	0.331535
		LmjF.30.1080	0.33182
		LmjF.09.1240	0.3335656
		LmjF.24.1470	0.3335656
		LmjF.23.1035	0.3335656
		LmjF.21.1040	0.343092
		LmjF.32.0370	0.3434249
		LmjF.21.0410	0.3434249
		LmjF.17.1230	0.3454187
		LmjF.10.0280	0.345661
		LmjF.32.3610	0.3484299
		LmjF.34.0970	0.3500829
		LmjF.16.0170	0.3500829
		LmjF.15.0620	0.3500829

Table S4-1 continued

		LmjF.15.0520	0.3500829
		LmjF.06.0820	0.3508019
		LmjF.30.2350	0.3574161
		LmjF.21.0823	0.3574161
		LmjF.17.0240	0.358294
		LmjF.29.1020	0.358294
		LmjF.26.1230	0.358294
		LmjF.02.0410	0.358294
		LmjF.15.1280	0.358294
		LmjF.26.0980	0.3586711
		LmjF.05.0790	0.3586711
		LmjF.31.0310	0.359272
		LmjF.10.0960	0.3596
		LmjF.31.2730	0.3596
		LmjF.04.0780	0.3596
		LmjF.36.0660	0.3596
		LmjF.33.2190	0.3622167
		LmjF.26.0770	0.3622167
		LmjF.26.2350	0.3622167
		LmjF.22.1310	0.3622167
		LmjF.31.0650	0.3622167
		LmjF.26.0610	0.3623241
		LmjF.21.0861	0.371372
		LmjF.10.0800	0.3720785
		LmjF.23.1320	0.3720785
		LmjF.31.0110	0.3737712
		LmjF.36.0400	0.3737712
		LmjF.24.2280	0.3737712
		LmjF.34.3270	0.3748983
		LmjF.23.0540	0.3769099
		LmjF.02.0310	0.3769099
		LmjF.36.1820	0.3769099
		LmjF.05.0640	0.3795878
		LmjF.26.2520	0.3796398
		LmjF.34.0530	0.3796398
		LmjF.20.1360	0.3796398
		LmjF.19.0590	0.3796398
		LmjF.36.4280	0.3796398
		LmjF.13.0500	0.3796398
		LmjF.18.0800	0.3796398

Table S4-1 continued

		LmjF.34.3750	0.3800186
		LmjF.33.1490	0.3843917
		LmjF.35.1150	0.3843917
		LmjF.13.1570	0.3843917
		LmjF.36.2210	0.3843917
		LmjF.26.0650	0.3892681
		LmjF.13.0100	0.3902105
		LmjF.28.2240	0.3944238
		LmjF.20.1570	0.3944238
		LmjF.22.0840	0.3956957
		LmjF.35.1490	0.3990776
		LmjF.17.0190	0.402472
		LmjF.26.2160	0.402472
		LmjF.12.1170	0.4028949
		LmjF.14.1210	0.4028949
		LmjF.20.1410	0.4050234
		LmjF.06.0030	0.4052636
		LmjF.32.3500	0.4079406
		LmjF.31.3110	0.4079406
		LmjF.15.0160	0.4079406
		LmjF.13.0970	0.4099589
		LmjF.36.3930	0.4108001
		LmjF.32.2000	0.4113494
		LmjF.10.0200	0.4113494
		LmjF.09.1315	0.4115539
		LmjF.21.0865	0.4115539
		LmjF.30.2850	0.4118694
		LmjF.07.0050	0.4118694
		LmjF.24.0310	0.4118694
		LmjF.24.1435	0.4118694
		LmjF.03.0450	0.4118694
		LmjF.24.0580	0.4118694
		LmjF.27.2140	0.4136037
		LmjF.29.1880	0.4162462
		LmjF.33.0180	0.4163927
		LmjF.07.1020	0.4163927
		LmjF.05.0260	0.4163927
		LmjF.34.4180	0.4164324
		LmjF.07.0920	0.4164324
		LmjF.31.2390	0.4167382

Table S4-1 continued

		LmjF.31.1300	0.4167382
		LmjF.08.0920	0.4173248
		LmjF.13.0210	0.4173248
		LmjF.31.2140	0.4173248
		LmjF.32.0470	0.4225298
		LmjF.16.0380	0.4234589
		LmjF.35.4000	0.4256495
		LmjF.04.0240	0.4264688
		LmjF.13.0960	0.4280456
		LmjF.18.1230	0.4280456
		LmjF.19.1310	0.4280456
		LmjF.35.4220	0.4299486
		LmjF.36.0950	0.4320463
		LmjF.28.2170	0.4327014
		LmjF.15.0640	0.4332016
		LmjF.16.1560	0.4342424
		LmjF.14.0705	0.4388074
		LmjF.15.0825	0.4417585
		LmjF.11.0730	0.4446094
		LmjF.27.1030	0.445442
		LmjF.26.1150	0.447361
		LmjF.32.0270	0.447361
		LmjF.33.2930	0.4500384
		LmjF.24.1130	0.4525868
		LmjF.09.0940	0.4525868
		LmjF.33.2650	0.4525868
		LmjF.33.1140	0.4525868
		LmjF.24.1270	0.4525868
		LmjF.24.1040	0.4549113
		LmjF.29.0530	0.458018
		LmjF.12.0290	0.458018
		LmjF.28.1120	0.4587893
		LmjF.29.1330	0.4625318
		LmjF.26.0710	0.464584
		LmjF.15.0880	0.4649935
		LmjF.18.0570	0.4658842
		LmjF.04.0510	0.4680189
		LmjF.07.0800	0.4804654
		LmjF.35.3620	0.480736
		LmjF.20.0690	0.480736

Table S4-1 continued

		LmjF.19.1005	0.480736
		LmjF.32.0580	0.480736
		LmjF.33.0530	0.4815519
		LmjF.24.0440	0.4815519
		LmjF.24.0690	0.4825625
		LmjF.15.0930	0.4828256
		LmjF.35.1350	0.4849831
		LmjF.23.0790	0.4849831
		LmjF.15.1020	0.4849831
		LmjF.23.0840	0.4849831
		LmjF.36.3180	0.4849831
		LmjF.30.2050	0.4849831
		LmjF.09.0350	0.4849831
		LmjF.35.4150	0.4863001
		LmjF.21.0853	0.4885513
		LmjF.03.0510	0.497314
		LmjF.36.6850	0.4978512
		LmjF.16.1460	0.498157
		LmjF.09.0740	0.4999754
		LmjF.05.0700	0.5018047
		LmjF.22.0470	0.5032371
		LmjF.32.3860	0.5038439
		LmjF.06.0995	0.5043447
		LmjF.31.2115	0.5043447
		LmjF.17.1210	0.5043447
		LmjF.34.0690	0.5043536
		LmjF.22.0670	0.5078703
		LmjF.35.4810	0.5078703
		LmjF.17.0330	0.5078703
		LmjF.16.1340	0.508197
		LmjF.21.1860	0.5100882
		LmjF.05.0650	0.5105173
		LmjF.34.2765	0.5105376
		LmjF.23.0450	0.5105376
		LmjF.34.2070	0.5149099
		LmjF.20.1460	0.5152908
		LmjF.27.1210	0.5152908
		LmjF.08.0930	0.5181656
		LmjF.27.0950	0.5194574
		LmjF.36.1590	0.5194574

Table S4-1 continued

		LmjF.16.0700	0.5234723
		LmjF.19.0580	0.5234723
		LmjF.35.2030	0.5234723
		LmjF.35.3990	0.5234723
		LmjF.22.1230	0.5234723
		LmjF.26.2060	0.5236295
		LmjF.10.0980	0.5252944
		LmjF.36.2180	0.527178
		LmjF.30.2880	0.5281012
		LmjF.19.1060	0.5294405
		LmjF.05.0620	0.5306749
		LmjF.35.0310	0.5308565
		LmjF.32.0970	0.5308565
		LmjF.35.4990	0.5308565
		LmjF.27.1130	0.5308565
		LmjF.25.0020	0.5308565
		LmjF.36.6300	0.5308565
		LmjF.32.1190	0.5308565
		LmjF.17.0660	0.5308565
		LmjF.34.0740	0.5308565
		LmjF.16.0930	0.5309592
		LmjF.14.0370	0.536209
		LmjF.27.0610	0.536209
		LmjF.35.5320	0.5372555
		LmjF.26.1540	0.5372555
		LmjF.30.0410	0.5372555
		LmjF.07.0150	0.5377787
		LmjF.35.4650	0.5377787
		LmjF.32.0780	0.5397998
		LmjF.24.2020	0.540509
		LmjF.16.1260	0.5410313
		LmjF.12.1190	0.5415302
		LmjF.23.0750	0.542555
		LmjF.31.2890	0.5429135
		LmjF.36.0640	0.543105
		LmjF.33.2170	0.543105
		LmjF.36.4220	0.548038
		LmjF.16.0770	0.5486002
		LmjF.14.0430	0.5486002
		LmjF.26.1360	0.5486002

Table S4-1 continued

		LmjF.34.3230	0.5515195
		LmjF.28.1000	0.557892
		LmjF.06.0670	0.5603899
		LmjF.12.0270	0.5603899
		LmjF.25.1510	0.5603899
		LmjF.23.1420	0.5603899
		LmjF.08.0440	0.5607722
		LmjF.09.0530	0.564269
		LmjF.14.0590	0.564269
		LmjF.03.0600	0.5647057
		LmjF.33.0560	0.5647057
		LmjF.30.1870	0.5647057
		LmjF.15.0350	0.5647057
		LmjF.16.0810	0.5647057
		LmjF.09.0390	0.5647057
		LmjF.34.2990	0.5647057
		LmjF.15.0420	0.5647057
		LmjF.13.1010	0.5647057
		LmjF.20.0850	0.5647057
		LmjF.33.0850	0.5647057
		LmjF.35.0830	0.5647057
		LmjF.20.0020	0.5647057
		LmjF.33.0550	0.5742573
		LmjF.25.1400	0.5765782
		LmjF.06.0660	0.5765782
		LmjF.16.0670	0.5801348
		LmjF.29.1230	0.5809645
		LmjF.04.0730	0.5831587
		LmjF.27.2160	0.5833228
		LmjF.35.2560	0.5844233
		LmjF.32.3490	0.5858776
		LmjF.03.0530	0.5880054
		LmjF.23.0940	0.5880054
		LmjF.07.0670	0.5880054
		LmjF.30.1560	0.5881545
		LmjF.25.1680	0.5881545

Table S4-2. *L.major* gene codes for orthologous protein predicted to be under PS with models M7 versus M8 with 4 and 3 taxa.

4 taxa analysis		3 taxa analysis	
Gene codes	**q value**	**Gene codes**	**q value**
LmjF.10.0470	5.09E-06	LmjF.26.2170	2.75E-76
LmjF.27.0970	5.64E-06	LmjF.35.0520	8.73E-65
LmjF.34.3630	0.000356601	LmjF.14.1120	3.41E-48
LmjF.11.1290	0.000377619	LmjF.30.1240	5.98E-41
LmjF.36.3180	0.000597379	LmjF.28.3010	1.21E-34
LmjF.35.1140	0.000788393	LmjF.07.0830	4.88E-34
LmjF.34.2700	0.000790628	LmjF.31.1490	3.86E-31
LmjF.08.0640	0.001430566	LmjF.14.1110	6.75E-28
LmjF.25.0670	0.004561686	LmjF.02.0300	1.31E-27
LmjF.25.1960	0.01078953	LmjF.17.0180	2.04E-26
LmjF.25.1950	0.02181324	LmjF.35.0500	2.61E-20
LmjF.10.0610	0.02829523	LmjF.31.1440	1.44E-19
LmjF.34.0460	0.03143893	LmjF.34.2410	5.08E-18
LmjF.08.0850	0.03489024	LmjF.04.0760	1.43E-17
LmjF.16.0720	0.03783461	LmjF.22.1320	1.44E-17
LmjF.05.0490	0.03783461	LmjF.24.0450	3.72E-17
LmjF.24.0300	0.04220504	LmjF.32.2220	3.85E-16
LmjF.14.1430	0.04228484	LmjF.05.0240	4.39E-16
LmjF.27.1160	0.04228484	LmjF.11.0070	1.30E-15
LmjF.01.0340	0.04242834	LmjF.32.1040	1.03E-14
LmjF.27.0290	0.04487772	LmjF.14.1440	6.81E-14
LmjF.26.0950	0.05010234	LmjF.28.3030	7.96E-14
LmjF.22.0220	0.05096955	LmjF.34.1920	2.02E-13
LmjF.24.0340	0.05096955	LmjF.35.0740	2.63E-12
LmjF.15.1080	0.06080953	LmjF.29.2560	5.07E-11
LmjF.31.2730	0.06080953	LmjF.15.0195	8.41E-11
LmjF.15.0870	0.06594067	LmjF.36.6290	3.80E-10
LmjF.10.0530	0.06655726	LmjF.31.2520	4.46E-10
LmjF.08.0800	0.06997175	LmjF.22.1210	4.46E-10
LmjF.06.0090	0.0709078	LmjF.17.0080	4.89E-10
LmjF.26.1360	0.07145113	LmjF.26.2030	6.04E-10
LmjF.18.1650	0.07145113	LmjF.11.1220	8.20E-10
LmjF.09.0320	0.0733774	LmjF.36.0110	3.06E-09
LmjF.05.0100	0.08090197	LmjF.21.0620	3.89E-09
LmjF.30.1410	0.08332955	LmjF.28.1430	8.67E-09
LmjF.23.0450	0.1055784	LmjF.02.0420	6.25E-08
LmjF.33.2090	0.1074549	LmjF.17.0200	9.30E-08

Table S4-2 continued

LmjF.27.0670	0.1074549	LmjF.34.2560	1.57E-07
LmjF.30.1610	0.1074549	LmjF.23.0220	1.89E-07
LmjF.23.0540	0.1074549	LmjF.03.0270	2.11E-07
LmjF.11.1270	0.1229166	LmjF.35.3770	7.08E-07
LmjF.30.2850	0.1296878	LmjF.16.1660	9.40E-07
LmjF.22.0410	0.1296878	LmjF.21.1170	1.10E-06
LmjF.27.0500	0.1296878	LmjF.10.0470	1.34E-06
LmjF.31.0090	0.1296878	LmjF.08.0390	2.22E-06
LmjF.32.1820	0.1296878	LmjF.26.1665	3.48E-06
LmjF.33.0530	0.1323101	LmjF.13.0850	7.57E-06
LmjF.04.0460	0.1357489	LmjF.20.1020	1.03E-05
LmjF.35.0460	0.1357489	LmjF.14.0740	1.27E-05
LmjF.31.0920	0.1357489	LmjF.36.1380	1.33E-05
LmjF.31.3110	0.1357489	LmjF.36.1370	3.09E-05
LmjF.24.1470	0.140751	LmjF.31.2350	3.14E-05
LmjF.14.0910	0.1553485	LmjF.28.1070	3.67E-05
LmjF.25.2050	0.1555919	LmjF.08.0640	4.73E-05
LmjF.34.2190	0.1610511	LmjF.36.1395	5.10E-05
LmjF.29.0620	0.1679471	LmjF.34.1720	6.16E-05
LmjF.36.1870	0.1712277	LmjF.05.1160	8.71E-05
LmjF.29.1520	0.1712277	LmjF.32.2270	9.57E-05
LmjF.28.2410	0.1716276	LmjF.31.0450	0.000108908
LmjF.35.4550	0.1907465	LmjF.09.0980	0.000135789
LmjF.30.3530	0.1957056	LmjF.07.0120	0.00018565
LmjF.35.3630	0.2051999	LmjF.23.0640	0.000242624
LmjF.02.0430	0.2051999	LmjF.17.0350	0.000242624
LmjF.31.2760	0.2081473	LmjF.15.1270	0.000295889
LmjF.33.1290	0.2113994	LmjF.04.0860	0.000295889
LmjF.01.0810	0.2113994	LmjF.34.0500	0.000322159
LmjF.35.4810	0.2113994	LmjF.26.1460	0.000502524
LmjF.19.0520	0.2113994	LmjF.08.0820	0.000502524
LmjF.28.2830	0.2163137	LmjF.31.2540	0.00052708
LmjF.09.0240	0.2163137	LmjF.15.0530	0.000716018
LmjF.36.2570	0.2195823	LmjF.36.5250	0.000771756
LmjF.28.1300	0.2334883	LmjF.04.0830	0.000956516
LmjF.16.1050	0.2356624	LmjF.36.1840	0.000991151
LmjF.36.0950	0.2399613	LmjF.31.0040	0.001113193
LmjF.02.0100	0.2399613	LmjF.01.0800	0.001121402
LmjF.24.0310	0.2411523	LmjF.27.1580	0.001121402
LmjF.34.4620	0.2472599	LmjF.30.2010	0.001160265

Table S4-2 continued

LmjF.34.0970	0.2472599	LmjF.14.1100	0.001160265
LmjF.34.2820	0.2472599	LmjF.11.1240	0.00122356
LmjF.13.0420	0.2472599	LmjF.35.1140	0.001290426
LmjF.35.2400	0.2472599	LmjF.35.1450	0.001567085
LmjF.36.4180	0.2484701	LmjF.12.0150	0.001618679
LmjF.32.3610	0.2484701	LmjF.13.0980	0.001624377
LmjF.28.1370	0.2484701	LmjF.21.0460	0.001905865
LmjF.30.1300	0.2484701	LmjF.32.0715	0.001921103
LmjF.32.2120	0.2601915	LmjF.19.1340	0.001921103
LmjF.06.1200	0.2627156	LmjF.15.1480	0.002005235
LmjF.19.1420	0.2627156	LmjF.04.0620	0.002454933
LmjF.19.0260	0.2657762	LmjF.26.1950	0.0031416
LmjF.34.2990	0.2674232	LmjF.21.0630	0.004628852
LmjF.29.2090	0.2677387	LmjF.07.0180	0.004664017
LmjF.25.1590	0.2677387	LmjF.09.0760	0.005321477
LmjF.06.1080	0.2677387	LmjF.19.1365	0.005321477
LmjF.36.2240	0.2710272	LmjF.26.1560	0.006017677
LmjF.32.1020	0.2761112	LmjF.31.0800	0.006590622
LmjF.06.0530	0.2761112	LmjF.08.0270	0.006590622
LmjF.32.0320	0.2772133	LmjF.04.0700	0.006590622
LmjF.13.1570	0.277546	LmjF.08.0310	0.007221576
LmjF.14.0990	0.277546	LmjF.24.1790	0.007221576
LmjF.13.0680	0.277546	LmjF.12.0590	0.007311368
LmjF.36.5240	0.2795803	LmjF.13.0780	0.007530507
LmjF.31.2350	0.2796246	LmjF.26.1430	0.007530507
LmjF.30.0430	0.2799937	LmjF.11.0660	0.007601277
LmjF.35.0840	0.2799937	LmjF.22.0940	0.008600488
LmjF.36.1300	0.2852239	LmjF.35.4840	0.009150474
LmjF.13.0510	0.3021365	LmjF.09.0220	0.009312928
LmjF.26.1210	0.3041999	LmjF.07.0340	0.009449692
LmjF.31.1140	0.3068776	LmjF.33.1240	0.01025195
LmjF.10.0960	0.3329689	LmjF.13.0670	0.01027738
LmjF.27.2130	0.3382187	LmjF.23.0900	0.01081895
LmjF.33.2940	0.3406297	LmjF.28.2840	0.01086646
LmjF.01.0020	0.3424252	LmjF.21.0530	0.01189833
LmjF.30.2880	0.3424252	LmjF.28.1170	0.01212213
LmjF.36.0580	0.3433198	LmjF.23.1170	0.01236649
LmjF.16.1360	0.3452593	LmjF.34.1680	0.01317632
LmjF.09.1020	0.3463119	LmjF.33.2960	0.01325317
LmjF.30.2410	0.3488295	LmjF.19.0940	0.01344579

Table S4-2 continued

LmjF.07.1035	0.3488295	LmjF.32.3690	0.01399029
LmjF.01.0470	0.3528252	LmjF.25.1570	0.01399029
LmjF.26.2420	0.3528252	LmjF.07.0740	0.01492615
LmjF.35.3620	0.3528252	LmjF.02.0470	0.01528745
LmjF.34.0140	0.3528252	LmjF.20.0790	0.01610156
LmjF.35.2060	0.3540982	LmjF.14.0340	0.01867811
LmjF.23.1510	0.3540982	LmjF.19.0985	0.01901393
LmjF.01.0230	0.3540982	LmjF.06.0080	0.01947568
LmjF.11.0670	0.3642013	LmjF.33.2020	0.01995106
LmjF.35.5020	0.3642013	LmjF.33.0650	0.01995106
LmjF.36.0660	0.3642013	LmjF.16.0720	0.02007041
LmjF.02.0620	0.3682171	LmjF.06.0980	0.02287523
LmjF.30.0230	0.3683898	LmjF.07.0890	0.02287523
LmjF.32.2490	0.371739	LmjF.10.1080	0.0254061
LmjF.09.0780	0.371739	LmjF.07.0280	0.02587162
LmjF.33.1860	0.3725297	LmjF.24.0360	0.02675167
LmjF.33.0110	0.3779936	LmjF.10.1100	0.02675167
LmjF.35.3660	0.3787511	LmjF.33.1290	0.02675167
LmjF.29.0820	0.3787511	LmjF.10.0500	0.02675167
LmjF.35.2160	0.3852711	LmjF.06.0830	0.02680268
LmjF.30.0050	0.3852711	LmjF.36.2020	0.02745435
LmjF.10.1290	0.3901678	LmjF.27.0490	0.0287769
LmjF.15.0420	0.3924152	LmjF.35.1160	0.0287769
LmjF.34.0080	0.396034	LmjF.16.0500	0.02970578
LmjF.29.0260	0.396034	LmjF.24.0680	0.03047654
LmjF.28.0080	0.3977383	LmjF.35.0030	0.03087001
LmjF.28.0330	0.3977383	LmjF.36.4720	0.03146711
LmjF.10.1260	0.3977383	LmjF.05.0380	0.03174699
LmjF.36.5110	0.400182	LmjF.20.1070	0.03220237
LmjF.28.0650	0.400182	LmjF.04.0250	0.03298347
LmjF.30.2350	0.4027358	LmjF.12.0470	0.03316116
LmjF.31.1130	0.4064166	LmjF.19.0520	0.03316116
LmjF.18.0680	0.4181006	LmjF.36.0910	0.03647771
LmjF.31.2810	0.4181006	LmjF.29.0820	0.03649232
LmjF.25.1640	0.420277	LmjF.28.0440	0.03649232
LmjF.19.1345	0.420277	LmjF.29.0110	0.03649232
LmjF.31.2780	0.4214853	LmjF.07.0730	0.03722633
LmjF.17.1020	0.4214853	LmjF.18.1420	0.03803239
LmjF.15.0730	0.4216139	LmjF.07.1060	0.03969126
LmjF.33.1890	0.4216139	LmjF.26.1420	0.04125014

Table S4-2 continued

LmjF.16.0820	0.4216139	LmjF.05.0270	0.04180795
LmjF.31.2390	0.42175	LmjF.12.0190	0.0422012
LmjF.11.0330	0.4255116	LmjF.04.0840	0.0422012
LmjF.31.1000	0.4344633	LmjF.36.5610	0.0422012
LmjF.28.2370	0.4344633	LmjF.14.1180	0.04338856
LmjF.27.1500	0.4344633	LmjF.23.0355	0.04338856
LmjF.21.1555	0.4344633	LmjF.18.0850	0.04675087
LmjF.20.0370	0.4344633	LmjF.30.2060	0.05004168
LmjF.23.1520	0.4344633	LmjF.05.0210	0.05006716
LmjF.26.2340	0.4344633	LmjF.02.0340	0.05006716
LmjF.26.2330	0.4344633	LmjF.11.0670	0.05006716
LmjF.32.0510	0.4344633	LmjF.05.0690	0.05006716
LmjF.15.0160	0.4344633	LmjF.30.0230	0.05006716
LmjF.18.1300	0.4444833	LmjF.13.1470	0.05006716
LmjF.14.0280	0.4510443	LmjF.02.0500	0.05006716
LmjF.35.0310	0.460549	LmjF.16.1015	0.05076715
LmjF.28.1460	0.4621054	LmjF.01.0740	0.05088691
LmjF.07.0050	0.4621054	LmjF.20.0770	0.05088691
LmjF.11.1080	0.4621054	LmjF.04.1030	0.05088691
LmjF.25.2040	0.4768415	LmjF.22.0640	0.0525995
LmjF.18.1260	0.4768415	LmjF.28.2445	0.05422712
LmjF.04.0240	0.4768415	LmjF.23.0210	0.05423194
LmjF.04.0260	0.4773065	LmjF.20.0470	0.05468694
LmjF.20.0260	0.4773065	LmjF.07.0170	0.05468694
LmjF.26.0170	0.4773065	LmjF.17.1440	0.05511964
LmjF.34.2670	0.4773065	LmjF.16.1090	0.05659446
LmjF.13.1500	0.480722	LmjF.16.1460	0.05659446
LmjF.14.1180	0.485186	LmjF.09.0890	0.05664265
LmjF.32.3880	0.485186	LmjF.30.0600	0.05706459
LmjF.18.1390	0.4913587	LmjF.31.2370	0.05771443
LmjF.30.1500	0.4972968	LmjF.05.0770	0.06009372
LmjF.32.3720	0.4991152	LmjF.31.3020	0.06072328
LmjF.35.3470	0.5011603	LmjF.15.0660	0.06121001
LmjF.28.1480	0.5065605	LmjF.01.0640	0.06135438
LmjF.26.2410	0.5065605	LmjF.10.1310	0.0630185
LmjF.03.0410	0.5065605	LmjF.05.0300	0.06408722
LmjF.20.1210	0.5065605	LmjF.31.1800	0.06408722
LmjF.26.2360	0.5102011	LmjF.27.0970	0.06545844
LmjF.25.1720	0.5102011	LmjF.32.1820	0.06815668
LmjF.06.1070	0.5102011	LmjF.31.2800	0.06848791

Table S4-2 continued

LmjF.30.1010	0.5102011	LmjF.31.3190	0.06974063
LmjF.29.1500	0.5132972	LmjF.35.2810	0.06974063
LmjF.14.0590	0.5132972	LmjF.13.1400	0.0704071
LmjF.26.1410	0.5132972	LmjF.24.0480	0.0704071
LmjF.34.0070	0.5160926	LmjF.35.0670	0.0704071
LmjF.34.2180	0.5194222	LmjF.18.1090	0.0704071
LmjF.07.0150	0.5199964	LmjF.22.1440	0.0704071
LmjF.26.0720	0.5199964	LmjF.13.1410	0.07067657
LmjF.22.1110	0.5199964	LmjF.05.0820	0.07067657
LmjF.32.1830	0.5415607	LmjF.22.1460	0.07079563
LmjF.08.0320	0.5468655	LmjF.34.3570	0.07186746
LmjF.16.0700	0.5469278	LmjF.04.0430	0.07186746
LmjF.31.1445	0.5529632	LmjF.32.3670	0.07228826
LmjF.05.0285	0.5553818	LmjF.27.0510	0.07228826
LmjF.36.6970	0.5567363	LmjF.19.1420	0.072824
LmjF.30.1090	0.5571963	LmjF.04.1110	0.07286309
LmjF.09.1390	0.5571963	LmjF.02.0700	0.07286309
LmjF.09.1240	0.5571963	LmjF.20.0040	0.07423723
LmjF.11.0610	0.5571963	LmjF.27.1160	0.07471386
LmjF.17.1310	0.5608729	LmjF.19.0260	0.0748059
LmjF.35.2530	0.5608729	LmjF.04.0460	0.07505296
LmjF.23.1700	0.5608729	LmjF.36.0880	0.07539919
LmjF.24.1270	0.5608729	LmjF.34.0690	0.07614417
LmjF.35.1060	0.5608729	LmjF.21.0125	0.07704435
LmjF.28.2350	0.5614735	LmjF.25.2050	0.0794048
LmjF.36.6760	0.5650534	LmjF.36.4180	0.0794048
LmjF.03.0380	0.5673229	LmjF.10.0610	0.0794048
LmjF.10.0630	0.5673229	LmjF.13.0920	0.08052671
LmjF.36.5690	0.5684935	LmjF.09.0410	0.08052671
LmjF.34.3930	0.5712674	LmjF.34.4020	0.0815784
LmjF.36.4120	0.5712674	LmjF.15.0825	0.0815784
LmjF.35.3210	0.581182	LmjF.25.0450	0.0825892
LmjF.15.1190	0.5875899	LmjF.21.0825	0.0844547
LmjF.31.0630	0.5875899	LmjF.08.0850	0.0844547
LmjF.09.1330	0.588734	LmjF.31.1270	0.08677899
LmjF.01.0690	0.588734	LmjF.26.2320	0.08692463
LmjF.19.0680	0.588734	LmjF.04.0500	0.0873066
LmjF.17.0730	0.588734	LmjF.36.0010	0.0908041
LmjF.26.2480	0.589132	LmjF.36.0840	0.0944003
LmjF.07.1050	0.5939961	LmjF.20.1610	0.09486053

Table S4-2 continued

LmjF.18.0510	0.594363	LmjF.11.0340	0.09729583
LmjF.28.2140	0.596922	LmjF.32.2400	0.09729583
LmjF.35.5210	0.5982909	LmjF.10.0220	0.09729583
LmjF.36.3010	0.6102356	LmjF.34.1170	0.0984648
LmjF.05.0120	0.6288415	LmjF.26.1110	0.098945
LmjF.17.1160	0.6412238	LmjF.19.1030	0.09930516
LmjF.25.0150	0.6439833	LmjF.29.1520	0.09996278
LmjF.29.2800	0.6546406	LmjF.10.1110	0.1025486
LmjF.07.0730	0.6546406	LmjF.34.2780	0.1025486
LmjF.24.0580	0.6651221	LmjF.16.0390	0.1031522
LmjF.36.2020	0.6743283	LmjF.18.1320	0.1031522
LmjF.36.1730	0.6743283	LmjF.18.0680	0.1055278
LmjF.11.0170	0.6776443	LmjF.12.0320	0.1070737
LmjF.36.2280	0.6776443	LmjF.15.1180	0.1072536
LmjF.28.2080	0.6776443	LmjF.29.1480	0.1072536
LmjF.29.1600	0.6776443	LmjF.11.0640	0.1072536
LmjF.34.0740	0.6789998	LmjF.18.1180	0.1087126
LmjF.12.0240	0.6789998	LmjF.29.0500	0.1087126
LmjF.35.4220	0.6789998	LmjF.19.0220	0.1087126
LmjF.30.2190	0.6789998	LmjF.05.0670	0.1094883
LmjF.32.1970	0.6798186	LmjF.27.1750	0.111015
LmjF.14.0320	0.6798186	LmjF.20.1250	0.1110958
LmjF.15.1380	0.6798186	LmjF.15.1080	0.1142466
LmjF.15.0300	0.6798186	LmjF.25.1590	0.1180806
LmjF.28.2970	0.6798186	LmjF.34.0190	0.1183596
LmjF.25.0690	0.6798186	LmjF.16.1425	0.120886
LmjF.34.3700	0.6871335	LmjF.31.1130	0.1273487
LmjF.23.1630	0.6871335	LmjF.25.0470	0.1274108
LmjF.30.1730	0.6871335	LmjF.30.2330	0.1286146
LmjF.28.1620	0.6871335	LmjF.15.0050	0.1286146
LmjF.14.0430	0.6871335	LmjF.32.3610	0.1288108
LmjF.27.2360	0.6871335	LmjF.22.0330	0.1292878
LmjF.30.1260	0.7130696	LmjF.16.0620	0.1292878
LmjF.17.0610	0.7142256	LmjF.30.0430	0.1302805
LmjF.33.3160	0.7142256	LmjF.33.1035	0.1317957
LmjF.32.1495	0.7148753	LmjF.36.1870	0.1341067
LmjF.17.0490	0.7150052	LmjF.36.2240	0.1341067
LmjF.23.0710	0.7152843	LmjF.03.0800	0.1342091
LmjF.20.0060	0.7152843	LmjF.11.0840	0.1351639
LmjF.30.3070	0.7262054	LmjF.13.0700	0.1368964

Table S4-2 continued

LmjF.35.3520	0.7262054	LmjF.27.1295	0.1386974
LmjF.20.1610	0.7296743	LmjF.16.0170	0.1386974
LmjF.35.5330	0.7452653	LmjF.31.1640	0.1386974
LmjF.25.1443	0.7452653	LmjF.18.1050	0.1386974
LmjF.28.1970	0.7527775	LmjF.34.2530	0.1386974
LmjF.29.0060	0.7527775	LmjF.07.0380	0.1386974
LmjF.18.0740	0.7527775	LmjF.33.0760	0.1386974
LmjF.36.3430	0.7527775	LmjF.21.1520	0.1395099
LmjF.20.1380	0.7527775	LmjF.12.0210	0.1396249
LmjF.30.2600	0.7527775	LmjF.36.2570	0.1396249
LmjF.14.0090	0.7527775	LmjF.36.3530	0.1396249
LmjF.36.2030	0.7527775	LmjF.06.0760	0.1396249
LmjF.28.0970	0.7527775	LmjF.17.1230	0.1396249
LmjF.36.5920	0.7527775	LmjF.15.1200	0.1396249
LmjF.09.0300	0.7527775	LmjF.21.1860	0.1396249
LmjF.04.0270	0.7527775	LmjF.35.3520	0.1396249
LmjF.01.0250	0.7527775	LmjF.23.1180	0.1406352
LmjF.36.6850	0.7527775	LmjF.09.0880	0.1406352
LmjF.17.0580	0.7592217	LmjF.02.0040	0.1407477
LmjF.32.3300	0.7592217	LmjF.30.0820	0.1420613
LmjF.19.0220	0.7624741	LmjF.20.0140	0.1424177
LmjF.24.2280	0.7624741	LmjF.02.0010	0.1481026
LmjF.18.0290	0.7624741	LmjF.21.0370	0.149214
LmjF.17.0340	0.7793114	LmjF.26.1210	0.149214
LmjF.05.0660	0.7793114	LmjF.28.1630	0.1510204
LmjF.16.1060	0.7793114	LmjF.34.0780	0.1522948
LmjF.19.0340	0.7793114	LmjF.19.0630	0.1522948
LmjF.15.0195	0.7793114	LmjF.16.0440	0.1532739
LmjF.34.3440	0.7793114	LmjF.13.0900	0.1548474
LmjF.08.1180	0.7793114	LmjF.24.0900	0.1548474
LmjF.20.0470	0.7793114	LmjF.16.0350	0.1548474
LmjF.19.0450	0.7793114	LmjF.36.6760	0.155595
LmjF.29.1020	0.7795998	LmjF.10.0320	0.155595
LmjF.07.1160	0.77973	LmjF.11.0160	0.155595
LmjF.06.0230	0.77973	LmjF.30.0050	0.155595
LmjF.17.1240	0.7996371	LmjF.24.2120	0.155595
LmjF.26.1440	0.8468581	LmjF.23.1430	0.1562686
LmjF.26.0220	0.8394426	LmjF.21.0270	0.1624902
LmjF.27.1130	0.8470732	LmjF.36.1300	0.1624902
LmjF.32.2240	0.8842463	LmjF.17.0440	0.1624902

Table S4-2 continued

LmjF.27.0690	0.8470732	LmjF.10.0380	0.1624902
LmjF.21.1360	0.8169426	LmjF.29.2460	0.1624902
LmjF.27.0840	0.8470732	LmjF.29.2090	0.1624902
LmjF.35.4350	0.9008967	LmjF.36.2590	0.1624902
LmjF.36.2610	0.903273	LmjF.34.3690	0.1643934
LmjF.23.1690	0.8196239	LmjF.11.1290	0.1645173
LmjF.23.1130	0.8196239	LmjF.33.3030	0.1671646
LmjF.26.0210	0.8340746	LmjF.34.2700	0.1671646
LmjF.29.0350	0.8531788	LmjF.32.0800	0.1676894
LmjF.27.0400	0.8470732	LmjF.21.1570	0.1700154
LmjF.25.1090	0.8340746	LmjF.26.1960	0.1704112
LmjF.35.0770	0.9008967	LmjF.29.2430	0.1704112
LmjF.29.2710	0.8774517	LmjF.29.1570	0.1704112
LmjF.07.0200	0.7894539	LmjF.26.1820	0.1704112
LmjF.14.0730	0.7918536	LmjF.10.0200	0.1704112
LmjF.31.1310	0.8842463	LmjF.25.0780	0.170544
LmjF.27.1030	0.8470732	LmjF.13.1570	0.1726051
LmjF.12.0300	0.7894539	LmjF.34.0140	0.1727638
LmjF.32.2000	0.8842463	LmjF.28.1800	0.1747957
LmjF.10.0200	0.7894539	LmjF.05.0710	0.1798881
LmjF.35.4650	0.9008967	LmjF.31.0440	0.1803458
LmjF.26.0540	0.8394426	LmjF.31.2760	0.1803458
LmjF.28.2770	0.8490793	LmjF.17.0990	0.1807673
LmjF.36.4230	0.904571	LmjF.13.1380	0.1813421
LmjF.04.0340	0.7894539	LmjF.10.0810	0.1815918
LmjF.26.0180	0.8340746	LmjF.03.0820	0.1831731
LmjF.26.1810	0.8470732	LmjF.36.1955	0.1832427
LmjF.22.0350	0.8176893	LmjF.21.1040	0.1855409
LmjF.31.3150	0.8842463	LmjF.08.0350	0.1858036
LmjF.15.1020	0.7974741	LmjF.33.1100	0.1868812
LmjF.32.2080	0.8842463	LmjF.07.0160	0.1875352
LmjF.19.1410	0.7996371	LmjF.19.1080	0.1926047
LmjF.20.0150	0.816745	LmjF.30.1410	0.1956605
LmjF.32.1400	0.8842463	LmjF.31.0310	0.1956605
LmjF.05.1190	0.7894539	LmjF.30.1610	0.1978288
LmjF.32.3310	0.8842463	LmjF.02.0100	0.1989378
LmjF.32.1110	0.8842463	LmjF.06.0090	0.1992354
LmjF.10.0760	0.7894539	LmjF.06.0810	0.2000136
LmjF.36.4780	0.904571	LmjF.16.0310	0.201471
LmjF.05.0630	0.7894539	LmjF.31.1330	0.2057838

Table S4-2 continued

LmjF.22.1620	0.8182243	LmjF.28.0650	0.2066691
LmjF.32.1900	0.8842463	LmjF.32.1830	0.2089883
LmjF.35.1600	0.9008967	LmjF.15.0620	0.2089883
LmjF.03.0170	0.7891677	LmjF.06.0190	0.2089883
LmjF.14.0705	0.7901633	LmjF.13.0200	0.2089883
LmjF.36.0535	0.9008967	LmjF.28.1300	0.2089883
LmjF.24.1140	0.8340746	LmjF.32.0370	0.2094268
LmjF.36.5880	0.904571	LmjF.14.0910	0.211028
LmjF.36.2210	0.903273	LmjF.30.1500	0.2135792
LmjF.24.1350	0.8340746	LmjF.17.1140	0.213911
LmjF.29.2070	0.8739053	LmjF.30.2350	0.213911
LmjF.34.4180	0.9008967	LmjF.31.3030	0.2177496
LmjF.35.4710	0.9008967	LmjF.26.2420	0.2177496
LmjF.25.0980	0.8340746	LmjF.15.1380	0.2177496
LmjF.28.2220	0.8470732	LmjF.30.0410	0.2177496
LmjF.15.0050	0.7974741	LmjF.31.2460	0.2177496
LmjF.32.0190	0.8842463	LmjF.15.0520	0.2186526
LmjF.29.0590	0.8689591	LmjF.33.1660	0.2189806
LmjF.34.3800	0.9008967	LmjF.19.1050	0.2199167
LmjF.27.1120	0.8470732	LmjF.24.1470	0.2213772
LmjF.33.2600	0.9008967	LmjF.35.0290	0.2223571
LmjF.20.1740	0.8169426	LmjF.24.0800	0.2228014
LmjF.31.0140	0.8842463	LmjF.15.0640	0.2228014
LmjF.34.3330	0.9008967	LmjF.24.1550	0.2229657
LmjF.32.3620	0.8860313	LmjF.02.0150	0.2229657
LmjF.36.2500	0.903273	LmjF.31.1730	0.2229657
LmjF.33.1600	0.9008967	LmjF.21.1410	0.2246242
LmjF.16.1425	0.7974741	LmjF.35.4000	0.2246242
LmjF.19.1220	0.7996371	LmjF.32.3190	0.2246242
LmjF.19.1480	0.8004119	LmjF.25.0670	0.2255948
LmjF.35.3840	0.9008967	LmjF.31.0110	0.2279359
LmjF.29.1150	0.8739053	LmjF.10.0630	0.2302302
LmjF.30.1050	0.8826799	LmjF.30.2160	0.2309306
LmjF.33.0780	0.9002384	LmjF.34.3930	0.2309306
LmjF.36.1150	0.903273	LmjF.28.1930	0.2309306
LmjF.23.0940	0.8196239	LmjF.10.0960	0.2309306
LmjF.21.1670	0.8169426	LmjF.15.1280	0.2319784
LmjF.13.1620	0.7894539	LmjF.21.0550	0.2326462
LmjF.14.0980	0.7955899	LmjF.14.0705	0.2326462
LmjF.17.1120	0.7996371	LmjF.29.0930	0.2329678

Table S4-2 continued

LmjF.18.1140	0.7996371	LmjF.27.1630	0.2329678
LmjF.30.0805	0.8775546	LmjF.33.2930	0.2335104
LmjF.30.1350	0.8842463	LmjF.17.1090	0.2342387
LmjF.18.1490	0.7996371	LmjF.07.0760	0.2372982
LmjF.33.0560	0.894007	LmjF.25.1820	0.2384176
LmjF.28.2940	0.8531788	LmjF.21.0120	0.2385902
LmjF.36.6900	0.9058592	LmjF.33.2120	0.2387357
LmjF.26.2350	0.8470732	LmjF.23.0540	0.2390269
LmjF.23.0355	0.8196239	LmjF.26.0650	0.2396738
LmjF.12.1240	0.7894539	LmjF.22.1100	0.2427332
LmjF.30.2780	0.8842463	LmjF.03.0850	0.2430502
LmjF.24.1080	0.82604	LmjF.36.3180	0.2464159
		LmjF.34.4180	0.24961
		LmjF.33.1490	0.2501498
		LmjF.25.1750	0.2513878
		LmjF.22.0840	0.2528998
		LmjF.34.0970	0.2554125
		LmjF.28.2690	0.2578151
		LmjF.35.2360	0.2655308
		LmjF.13.0500	0.269147
		LmjF.13.0510	0.269147
		LmjF.04.0600	0.2729548
		LmjF.02.0410	0.2751668
		LmjF.04.0240	0.2764953
		LmjF.12.1170	0.2764953
		LmjF.07.1020	0.2777101
		LmjF.26.1230	0.2784534
		LmjF.31.2390	0.2799494
		LmjF.30.1080	0.2810495
		LmjF.23.1320	0.2830949
		LmjF.19.0590	0.2832942
		LmjF.05.0790	0.2844464
		LmjF.21.0823	0.2844464
		LmjF.02.0310	0.284608
		LmjF.23.1035	0.2879214
		LmjF.21.0861	0.2880685
		LmjF.36.0470	0.2894585
		LmjF.24.2280	0.2894585
		LmjF.26.0770	0.2894585
		LmjF.36.0950	0.29

Table S4-2 continued

		LmjF.36.3930	0.2928502
		LmjF.31.2730	0.2939976
		LmjF.10.0280	0.2947688
		LmjF.21.0410	0.2950669
		LmjF.17.0190	0.2967061
		LmjF.17.0240	0.2967106
		LmjF.36.0400	0.3001981
		LmjF.04.0780	0.3043203
		LmjF.07.0920	0.3047993
		LmjF.06.0820	0.3047993
		LmjF.29.1020	0.3076295
		LmjF.05.0640	0.3076295
		LmjF.26.1150	0.3098209
		LmjF.06.0960	0.3128362
		LmjF.27.0500	0.3128362
		LmjF.27.1130	0.3128362
		LmjF.07.0800	0.3131545
		LmjF.22.1310	0.3131545
		LmjF.35.4150	0.3132195
		LmjF.26.2160	0.3143599
		LmjF.36.1820	0.317143
		LmjF.07.0050	0.3172636
		LmjF.15.0160	0.3172636
		LmjF.26.0980	0.3172636
		LmjF.23.0450	0.3172636
		LmjF.36.0660	0.3172636
		LmjF.26.2520	0.3187473
		LmjF.33.2190	0.3187473
		LmjF.26.2350	0.3207152
		LmjF.31.0650	0.3210782
		LmjF.20.1360	0.3214049
		LmjF.26.0610	0.3231416
		LmjF.11.0730	0.3231416
		LmjF.03.0450	0.3231416
		LmjF.33.1140	0.3231416
		LmjF.18.0800	0.3231416
		LmjF.18.0570	0.3231893
		LmjF.34.3750	0.3231893
		LmjF.10.0800	0.3231893
		LmjF.34.3270	0.3239699

Table S4-2 continued

		LmjF.32.0780	0.324834
		LmjF.20.1570	0.3265586
		LmjF.32.0470	0.3267355
		LmjF.22.0470	0.3279013
		LmjF.03.0510	0.3302979
		LmjF.09.0740	0.3330786
		LmjF.05.0260	0.3330786
		LmjF.34.0530	0.3330786
		LmjF.34.3230	0.3330786
		LmjF.36.4280	0.3344027
		LmjF.35.1150	0.3360771
		LmjF.33.0180	0.3376898
		LmjF.31.3110	0.3376898
		LmjF.32.0580	0.3383974
		LmjF.36.2210	0.342517
		LmjF.21.0865	0.348056
		LmjF.06.0030	0.348056
		LmjF.32.3500	0.348056
		LmjF.13.0100	0.3488284
		LmjF.27.1030	0.3488284
		LmjF.32.0270	0.348963
		LmjF.29.1880	0.348963
		LmjF.28.1120	0.348963
		LmjF.28.2240	0.3494274
		LmjF.14.1210	0.3494274
		LmjF.26.0710	0.3503995
		LmjF.20.0690	0.351099
		LmjF.30.2850	0.351099
		LmjF.35.1490	0.351099
		LmjF.36.6850	0.3556841
		LmjF.31.1300	0.358204
		LmjF.20.1410	0.3586163
		LmjF.16.0380	0.3591116
		LmjF.13.0970	0.3592179
		LmjF.07.0670	0.3592179
		LmjF.09.0350	0.3592179
		LmjF.34.2400	0.361256
		LmjF.24.0310	0.3620619
		LmjF.09.1315	0.3639718
		LmjF.32.2000	0.3645749

Table S4-2 continued

		LmjF.33.0560	0.3652024
		LmjF.28.2170	0.3661094
		LmjF.24.0580	0.3661167
		LmjF.24.2020	0.3680641
		LmjF.20.1460	0.3687096
		LmjF.24.1435	0.368983
		LmjF.27.2140	0.3709901
		LmjF.29.0530	0.3727126
		LmjF.32.3490	0.3727126
		LmjF.15.1020	0.3727126
		LmjF.16.1560	0.3727126
		LmjF.13.0960	0.3727126
		LmjF.05.0700	0.3727126
		LmjF.35.0830	0.3727126
		LmjF.18.1230	0.3727126
		LmjF.31.2140	0.3727126
		LmjF.07.0150	0.3728541
		LmjF.08.0920	0.3728541
		LmjF.16.1340	0.3731841
		LmjF.13.0210	0.3732605
		LmjF.19.1310	0.3732605
		LmjF.12.1190	0.374567
		LmjF.26.1570	0.374567
		LmjF.04.0310	0.374567
		LmjF.35.4220	0.374937
		LmjF.19.1160	0.3766605
		LmjF.04.0510	0.3766605
		LmjF.30.2050	0.3766605
		LmjF.24.1270	0.3800626
		LmjF.35.4990	0.3862245
		LmjF.22.1230	0.3862245
		LmjF.33.2650	0.386525
		LmjF.35.5320	0.3880033
		LmjF.23.1510	0.3887165
		LmjF.23.1420	0.3887165
		LmjF.23.0840	0.3887165
		LmjF.33.0530	0.3887165
		LmjF.25.0020	0.3887165
		LmjF.05.0650	0.3887165
		LmjF.36.4220	0.3887165

Table S4-2 continued

		LmjF.24.1220	0.3887165
		LmjF.12.0290	0.3887165
		LmjF.32.0970	0.3887165
		LmjF.14.0430	0.3900813
		LmjF.03.0600	0.3900813
		LmjF.26.1540	0.3907381
		LmjF.27.0610	0.3913705
		LmjF.14.1060	0.3963063
		LmjF.19.1005	0.396387
		LmjF.24.1040	0.3966204
		LmjF.29.1330	0.3977807
		LmjF.19.1060	0.3977807
		LmjF.21.0853	0.3977807
		LmjF.24.1130	0.3977807
		LmjF.24.0440	0.4008024
		LmjF.25.1680	0.4008024
		LmjF.35.3620	0.4008024
		LmjF.16.0810	0.4094787
		LmjF.15.0880	0.4104265
		LmjF.06.0670	0.4104265
		LmjF.10.0980	0.4109087
		LmjF.33.0550	0.4110145
		LmjF.25.1490	0.4117223
		LmjF.26.1360	0.4117223
		LmjF.36.1440	0.4117223
		LmjF.35.4650	0.4117223
		LmjF.34.2765	0.4117223
		LmjF.31.2115	0.4142041
		LmjF.24.0690	0.4142041
		LmjF.35.1350	0.4142041
		LmjF.23.0790	0.4142041
		LmjF.28.0500	0.4142041
		LmjF.06.0995	0.4156691
		LmjF.17.0330	0.4156691
		LmjF.35.2030	0.4156691
		LmjF.20.0020	0.4158176
		LmjF.28.1000	0.4159652
		LmjF.34.0460	0.4169746
		LmjF.35.4810	0.4169746
		LmjF.15.0930	0.4185297

Table S4-2 continued

		LmjF.32.1020	0.4225295
		LmjF.36.6300	0.4234424
		LmjF.31.2570	0.4234424
		LmjF.17.1210	0.4247185
		LmjF.32.3860	0.4247185
		LmjF.13.0810	0.4314807
		LmjF.34.2990	0.4336539
		LmjF.34.2070	0.4353573
		LmjF.09.0390	0.4353573
		LmjF.36.4550	0.4438747
		LmjF.35.2560	0.4440189
		LmjF.26.2060	0.4444452
		LmjF.09.1500	0.4474846
		LmjF.36.0640	0.4474846
		LmjF.22.0670	0.4480976
		LmjF.36.1590	0.4480976
		LmjF.08.0930	0.4480976
		LmjF.36.2030	0.4480976
		LmjF.03.0530	0.4480976
		LmjF.16.0700	0.4487559
		LmjF.17.0660	0.4553087
		LmjF.27.0950	0.4566859
		LmjF.30.2260	0.4566859
		LmjF.05.0620	0.4566859
		LmjF.08.0440	0.4566859
		LmjF.06.0660	0.4585932
		LmjF.27.1210	0.4589107
		LmjF.31.2890	0.4589107
		LmjF.23.0940	0.4589107
		LmjF.19.0580	0.4591838
		LmjF.28.2480	0.4598953
		LmjF.34.0740	0.4598953
		LmjF.36.5760	0.460813
		LmjF.13.1010	0.4614732
		LmjF.30.1560	0.4614732
		LmjF.35.3990	0.4614732
		LmjF.33.2170	0.4614732
		LmjF.35.0310	0.4672349
		LmjF.25.0320	0.4679589
		LmjF.30.2880	0.4682189

Table S4-2 continued

		LmjF.16.1260	0.4682189
		LmjF.25.0980	0.4682189
		LmjF.36.2180	0.4682189
		LmjF.36.4680	0.4694584
		LmjF.11.0310	0.4694584
		LmjF.32.1190	0.4717114
		LmjF.16.0770	0.4724749
		LmjF.29.2260	0.4724749
		LmjF.10.0790	0.4724749
		LmjF.16.0930	0.4741339
		LmjF.08.0320	0.4749879
		LmjF.14.0370	0.4752254
		LmjF.04.1180	0.4759726
		LmjF.25.1510	0.4759726
		LmjF.32.1620	0.4766943
		LmjF.15.0420	0.4818688
		LmjF.23.0750	0.4821782
		LmjF.25.1900	0.4883912
		LmjF.31.1470	0.488411
		LmjF.36.0270	0.4916832
		LmjF.29.2100	0.4916832
		LmjF.26.0820	0.4916832
		LmjF.09.0530	0.4966717
		LmjF.09.1240	0.4989487
		LmjF.30.1870	0.4999029
		LmjF.09.0003	0.4999029
		LmjF.31.2100	0.4999029
		LmjF.32.2580	0.5029974
		LmjF.18.0440	0.5035011
		LmjF.12.0270	0.5035011
		LmjF.15.0350	0.5042519
		LmjF.08.0780	0.5049392
		LmjF.36.0860	0.505351
		LmjF.10.0900	0.505351
		LmjF.12.0330	0.505351
		LmjF.33.0670	0.505351
		LmjF.14.0590	0.505351
		LmjF.35.1640	0.505351
		LmjF.24.0260	0.505351
		LmjF.07.0310	0.505351

Table S4-2 continued

		LmjF.29.1230	0.505351
		LmjF.17.1290	0.505351
		LmjF.33.0850	0.505351
		LmjF.36.2790	0.5079446
		LmjF.36.4740	0.5079446
		LmjF.20.0850	0.5103344
		LmjF.16.0670	0.5103344
		LmjF.30.3660	0.5103344
		LmjF.25.1400	0.5125706
		LmjF.12.0300	0.5198614
		LmjF.36.4610	0.5198614
		LmjF.04.0730	0.5227156
		LmjF.08.0870	0.5227156
		LmjF.34.3010	0.5242152
		LmjF.26.1390	0.5274605
		LmjF.17.1010	0.5274605
		LmjF.22.0410	0.5275893
		LmjF.04.0400	0.5278637
		LmjF.26.0690	0.5280486
		LmjF.09.0690	0.5280486
		LmjF.19.0950	0.5280486
		LmjF.33.3040	0.5280486
		LmjF.02.0570	0.5321363
		LmjF.08.0810	0.5331574

Table S4-3. *T. cruzi* (Non-Esmeraldo) gene codes for orthologous proteins predicted to be under PS with models M8 versus M8a with 4 and 3 taxa.

4 taxa analysis		3 taxa analysis	
Gene codes	**q value**	**Gene codes**	**q value**
Tc00.1047053509791.70	3.08268E-35	Tc00.1047053509023.200	4.53E-40
Tc00.1047053509023.200	5.62407E-34	Tc00.1047053511287.10	2.82E-31
Tc00.1047053508799.240	3.44201E-30	Tc00.1047053509229.40	1.22E-27
Tc00.1047053508299.30	1.46886E-29	Tc00.1047053510359.80	1.22E-27
Tc00.1047053508895.40	1.60214E-29	Tc00.1047053508045.4	2.15E-27
Tc00.1047053508175.370	1.08904E-28	Tc00.1047053506989.70	2.08E-24
Tc00.1047053506989.70	1.12622E-26	Tc00.1047053506477.30	5.02E-22
Tc00.1047053506821.120	2.15792E-26	Tc00.1047053509799.140	6.30E-22
Tc00.1047053507601.90	1.41724E-25	Tc00.1047053511807.60	1.92E-21
Tc00.1047053503573.9	2.19656E-25	Tc00.1047053508865.20	4.30E-21
Tc00.1047053508045.4	6.70457E-24	Tc00.1047053506415.10	2.56E-20
Tc00.1047053505183.40	2.22322E-23	Tc00.1047053507491.50	6.87E-20
Tc00.1047053510359.80	8.73005E-23	Tc00.1047053508355.70	4.66E-19
Tc00.1047053509745.60	3.27462E-22	Tc00.1047053509719.20	1.61E-18
Tc00.1047053508699.10	2.67668E-21	Tc00.1047053508153.1010	2.70E-18
Tc00.1047053511807.60	2.17566E-20	Tc00.1047053506957.80	3.69E-18
Tc00.1047053506959.90	5.08206E-20	Tc00.1047053509051.50	4.11E-18
Tc00.1047053511287.10	4.24467E-19	Tc00.1047053505183.20	4.82E-18
Tc00.1047053509229.40	5.48578E-19	Tc00.1047053507895.160	9.49E-18
Tc00.1047053506247.30	2.10108E-18	Tc00.1047053507059.80	1.84E-17
Tc00.1047053507519.100	2.82574E-18	Tc00.1047053509353.50	2.51E-17
Tc00.1047053506833.30	5.01809E-18	Tc00.1047053507771.60	4.13E-17
Tc00.1047053509641.50	6.52278E-18	Tc00.1047053511285.80	7.20E-17
Tc00.1047053509607.30	2.74919E-17	Tc00.1047053509429.230	8.97E-17
Tc00.1047053509857.40	2.74919E-17	Tc00.1047053511865.50	2.24E-16
Tc00.1047053503809.158	2.97919E-17	Tc00.1047053508141.60	2.24E-16
Tc00.1047053506195.270	3.89442E-17	Tc00.1047053509641.50	3.00E-16
Tc00.1047053503847.60	1.60463E-16	Tc00.1047053506619.90	4.14E-16
Tc00.1047053511285.80	1.92054E-16	Tc00.1047053510729.100	4.26E-16
Tc00.1047053509617.80	2.23852E-16	Tc00.1047053504021.50	5.77E-16
Tc00.1047053510889.130	2.63101E-16	Tc00.1047053506725.80	9.18E-16
Tc00.1047053507491.50	3.22343E-16	Tc00.1047053509207.110	1.84E-15
Tc00.1047053509695.100	1.02443E-15	Tc00.1047053508891.4	2.31E-15
Tc00.1047053510885.80	1.41695E-15	Tc00.1047053510039.80	2.33E-15
Tc00.1047053508709.10	1.57485E-15	Tc00.1047053507601.60	3.98E-15
Tc00.1047053510121.40	1.7186E-15	Tc00.1047053510579.119	1.04E-14
Tc00.1047053508613.20	1.91953E-15	Tc00.1047053510735.50	1.24E-14

Table S4-3 continued

Tc00.1047053511385.100	2.05122E-15	Tc00.1047053504131.140	1.65E-14
Tc00.1047053511153.120	1.00736E-14	Tc00.1047053506459.280	1.80E-14
Tc00.1047053511239.120	1.39725E-14	Tc00.1047053503999.40	2.95E-14
Tc00.1047053508153.1010	1.56677E-14	Tc00.1047053511859.80	5.36E-14
Tc00.1047053509799.140	4.15436E-14	Tc00.1047053509875.180	6.62E-14
Tc00.1047053506425.130	1.67181E-13	Tc00.1047053508543.180	7.01E-14
Tc00.1047053506459.280	1.69235E-13	Tc00.1047053509605.20	1.18E-13
Tc00.1047053503879.119	1.76964E-13	Tc00.1047053508299.30	1.48E-13
Tc00.1047053510121.140	2.62902E-13	Tc00.1047053508895.60	1.66E-13
Tc00.1047053508543.10	1.18358E-12	Tc00.1047053508899.130	3.74E-13
Tc00.1047053508891.4	1.42172E-12	Tc00.1047053509569.140	3.78E-13
Tc00.1047053511461.50	1.56196E-12	Tc00.1047053509637.27	6.24E-13
Tc00.1047053509719.20	2.10473E-12	Tc00.1047053508059.10	9.80E-13
Tc00.1047053508175.50	2.10833E-12	Tc00.1047053511031.40	1.16E-12
Tc00.1047053507771.60	2.2648E-12	Tc00.1047053507895.140	1.49E-12
Tc00.1047053507023.90	3.89517E-12	Tc00.1047053504105.230	1.84E-12
Tc00.1047053506615.50	4.01812E-12	Tc00.1047053507895.164	3.11E-12
Tc00.1047053511865.50	4.04768E-12	Tc00.1047053511283.140	3.13E-12
Tc00.1047053506989.150	4.24403E-12	Tc00.1047053506959.80	3.13E-12
Tc00.1047053508175.90	4.24403E-12	Tc00.1047053510305.70	3.96E-12
Tc00.1047053510735.50	5.62324E-12	Tc00.1047053509745.60	3.96E-12
Tc00.1047053511231.14	7.13711E-12	Tc00.1047053503823.40	9.69E-12
Tc00.1047053509429.230	1.0354E-11	Tc00.1047053508577.160	9.69E-12
Tc00.1047053511859.80	1.26065E-11	Tc00.1047053506151.10	1.05E-11
Tc00.1047053509029.140	1.31984E-11	Tc00.1047053504253.10	1.28E-11
Tc00.1047053511385.70	1.31984E-11	Tc00.1047053504021.90	1.39E-11
Tc00.1047053509569.140	1.35801E-11	Tc00.1047053510667.110	1.61E-11
Tc00.1047053509733.80	1.69386E-11	Tc00.1047053511757.10	1.61E-11
Tc00.1047053509637.27	2.3383E-11	Tc00.1047053504797.80	1.77E-11
Tc00.1047053510895.20	2.39083E-11	Tc00.1047053506615.50	2.75E-11
Tc00.1047053504105.230	3.09455E-11	Tc00.1047053508415.30	5.17E-11
Tc00.1047053510039.80	3.14293E-11	Tc00.1047053508645.40	5.61E-11
Tc00.1047053511653.10	3.84106E-11	Tc00.1047053506957.100	5.67E-11
Tc00.1047053504021.90	3.89974E-11	Tc00.1047053511521.40	7.45E-11
Tc00.1047053509207.110	5.21664E-11	Tc00.1047053510575.130	7.92E-11
Tc00.1047053509353.50	5.3604E-11	Tc00.1047053507715.40	7.92E-11
Tc00.1047053509799.120	6.44546E-11	Tc00.1047053509229.120	9.05E-11
Tc00.1047053509395.59	6.8815E-11	Tc00.1047053510665.40	1.05E-10
Tc00.1047053508543.100	7.21332E-11	Tc00.1047053510339.70	1.36E-10

Table S4-3 continued

Tc00.1047053510667.110	9.1498E-11	Tc00.1047053511661.10	1.49E-10
Tc00.1047053427091.10	1.16378E-10	Tc00.1047053509671.168	1.59E-10
Tc00.1047053508899.130	1.24709E-10	Tc00.1047053511263.30	1.66E-10
Tc00.1047053507895.164	1.24906E-10	Tc00.1047053508593.110	1.86E-10
Tc00.1047053509671.168	1.31594E-10	Tc00.1047053509799.120	2.20E-10
Tc00.1047053507771.80	1.33491E-10	Tc00.1047053503729.10	4.68E-10
Tc00.1047053507895.140	1.42042E-10	Tc00.1047053506195.270	4.98E-10
Tc00.1047053511031.40	1.42042E-10	Tc00.1047053507003.24	6.09E-10
Tc00.1047053503823.40	1.55118E-10	Tc00.1047053506425.70	6.09E-10
Tc00.1047053504253.10	1.95531E-10	Tc00.1047053503411.10	6.09E-10
Tc00.1047053508593.110	2.73195E-10	Tc00.1047053511501.10	6.65E-10
Tc00.1047053506679.190	3.19461E-10	Tc00.1047053507777.10	8.56E-10
Tc00.1047053506957.100	3.78422E-10	Tc00.1047053506195.260	1.11E-09
Tc00.1047053506275.20	4.48546E-10	Tc00.1047053511237.100	1.30E-09
Tc00.1047053508059.10	4.85112E-10	Tc00.1047053511209.40	1.73E-09
Tc00.1047053506959.80	9.59815E-10	Tc00.1047053509797.60	1.73E-09
Tc00.1047053503411.10	1.27538E-09	Tc00.1047053504059.20	1.83E-09
Tc00.1047053509505.40	1.38159E-09	Tc00.1047053510967.30	1.90E-09
Tc00.1047053508413.40	1.46045E-09	Tc00.1047053510579.110	2.27E-09
Tc00.1047053510339.70	2.15081E-09	Tc00.1047053506989.150	2.69E-09
Tc00.1047053506619.90	2.74097E-09	Tc00.1047053509053.80	2.73E-09
Tc00.1047053508415.30	2.90435E-09	Tc00.1047053506629.200	2.87E-09
Tc00.1047053509229.120	3.01145E-09	Tc00.1047053507771.80	3.46E-09
Tc00.1047053504021.50	3.05761E-09	Tc00.1047053503917.14	4.01E-09
Tc00.1047053509877.80	4.04807E-09	Tc00.1047053507649.40	4.73E-09
Tc00.1047053510435.20	4.06333E-09	Tc00.1047053506301.30	5.28E-09
Tc00.1047053506151.10	5.1135E-09	Tc00.1047053508153.480	5.95E-09
Tc00.1047053509049.10	5.11695E-09	Tc00.1047053507969.30	6.08E-09
Tc00.1047053511321.26	5.4346E-09	Tc00.1047053508891.40	6.99E-09
Tc00.1047053507513.100	5.62601E-09	Tc00.1047053509207.40	7.42E-09
Tc00.1047053510329.10	5.89345E-09	Tc00.1047053509791.170	7.89E-09
Tc00.1047053510655.60	6.38859E-09	Tc00.1047053504175.20	8.23E-09
Tc00.1047053511521.40	6.73529E-09	Tc00.1047053506511.30	8.38E-09
Tc00.1047053508895.60	6.91684E-09	Tc00.1047053511755.60	8.80E-09
Tc00.1047053506297.340	7.33852E-09	Tc00.1047053508507.20	1.18E-08
Tc00.1047053506477.60	7.68185E-09	Tc00.1047053506789.310	1.47E-08
Tc00.1047053511237.100	9.28058E-09	Tc00.1047053507023.230	1.65E-08
Tc00.1047053507057.40	9.49549E-09	Tc00.1047053507773.30	1.65E-08
Tc00.1047053506985.40	1.07765E-08	Tc00.1047053509331.160	3.01E-08
Tc00.1047053507823.39	1.12491E-08	Tc00.1047053511855.20	3.04E-08

Table S4-3 continued

Tc00.1047053510155.130	1.19377E-08	Tc00.1047053506389.70	3.05E-08
Tc00.1047053510665.40	1.3724E-08	Tc00.1047053506195.230	3.25E-08
Tc00.1047053504113.30	1.46443E-08	Tc00.1047053511859.20	3.61E-08
Tc00.1047053507023.70	1.77716E-08	Tc00.1047053506145.10	3.61E-08
Tc00.1047053503987.20	1.7845E-08	Tc00.1047053505999.120	4.37E-08
Tc00.1047053507715.40	1.7845E-08	Tc00.1047053511277.480	4.37E-08
Tc00.1047053503729.10	1.95872E-08	Tc00.1047053508911.70	4.95E-08
Tc00.1047053508543.180	2.5636E-08	Tc00.1047053510435.20	6.34E-08
Tc00.1047053509663.20	2.67882E-08	Tc00.1047053504003.60	6.57E-08
Tc00.1047053508355.70	2.72649E-08	Tc00.1047053503719.39	7.09E-08
Tc00.1047053510729.100	2.82812E-08	Tc00.1047053508741.340	7.73E-08
Tc00.1047053506301.30	2.89359E-08	Tc00.1047053509065.9	7.73E-08
Tc00.1047053507969.30	2.89359E-08	Tc00.1047053511511.30	9.48E-08
Tc00.1047053508899.20	3.04982E-08	Tc00.1047053506831.20	9.53E-08
Tc00.1047053504157.120	3.07279E-08	Tc00.1047053511657.60	1.10E-07
Tc00.1047053511277.480	3.63618E-08	Tc00.1047053507601.10	1.13E-07
Tc00.1047053503995.20	3.97514E-08	Tc00.1047053503505.14	1.19E-07
Tc00.1047053507609.40	4.80274E-08	Tc00.1047053506559.20	1.47E-07
Tc00.1047053510121.50	4.82728E-08	Tc00.1047053511817.134	1.57E-07
Tc00.1047053511649.100	6.65881E-08	Tc00.1047053511465.20	1.63E-07
Tc00.1047053511283.140	6.67216E-08	Tc00.1047053508891.60	1.68E-07
Tc00.1047053506303.10	9.41795E-08	Tc00.1047053510439.20	1.69E-07
Tc00.1047053507087.30	9.95215E-08	Tc00.1047053511649.110	1.69E-07
Tc00.1047053511263.30	1.04164E-07	Tc00.1047053511857.40	1.91E-07
Tc00.1047053503657.70	1.05424E-07	Tc00.1047053503879.20	2.12E-07
Tc00.1047053510823.30	1.05424E-07	Tc00.1047053509791.50	2.34E-07
Tc00.1047053509875.180	1.16914E-07	Tc00.1047053506319.60	2.62E-07
Tc00.1047053504069.4	1.19073E-07	Tc00.1047053503987.20	3.13E-07
Tc00.1047053504797.80	1.21035E-07	Tc00.1047053510887.14	3.36E-07
Tc00.1047053507649.40	1.37134E-07	Tc00.1047053507771.90	3.37E-07
Tc00.1047053508737.50	1.47597E-07	Tc00.1047053509799.10	3.68E-07
Tc00.1047053508141.60	1.60029E-07	Tc00.1047053510885.10	3.96E-07
Tc00.1047053510969.10	2.25378E-07	Tc00.1047053503465.10	5.30E-07
Tc00.1047053503823.104	2.86666E-07	Tc00.1047053504105.210	7.46E-07
Tc00.1047053509331.160	2.92978E-07	Tc00.1047053506587.20	8.09E-07
Tc00.1047053511817.20	2.93119E-07	Tc00.1047053509747.50	8.44E-07
Tc00.1047053416511.9	3.13932E-07	Tc00.1047053506851.20	1.31E-06
Tc00.1047053504175.20	3.14497E-07	Tc00.1047053506479.80	1.31E-06
Tc00.1047053504059.20	3.15461E-07	Tc00.1047053503995.20	1.43E-06
Tc00.1047053506195.230	3.69064E-07	Tc00.1047053507775.10	1.63E-06

Table S4-3 continued

Tc00.1047053503521.80	4.93842E-07	Tc00.1047053506475.105	1.66E-06
Tc00.1047053511655.10	6.11897E-07	Tc00.1047053508707.160	1.66E-06
Tc00.1047053507539.20	6.55693E-07	Tc00.1047053503939.80	1.66E-06
Tc00.1047053506319.60	7.54812E-07	Tc00.1047053508641.320	1.67E-06
Tc00.1047053509683.40	7.59413E-07	Tc00.1047053511855.80	1.80E-06
Tc00.1047053510967.30	7.90392E-07	Tc00.1047053509437.30	1.80E-06
Tc00.1047053508741.340	8.68251E-07	Tc00.1047053503703.70	2.14E-06
Tc00.1047053506297.200	8.78359E-07	Tc00.1047053510329.10	2.14E-06
Tc00.1047053511209.40	8.80411E-07	Tc00.1047053511521.30	2.15E-06
Tc00.1047053506511.30	9.56108E-07	Tc00.1047053509437.40	2.41E-06
Tc00.1047053511855.20	1.05213E-06	Tc00.1047053506831.10	2.53E-06
Tc00.1047053510439.20	1.09684E-06	Tc00.1047053506315.10	2.56E-06
Tc00.1047053511661.10	1.15239E-06	Tc00.1047053510969.10	2.92E-06
Tc00.1047053509791.50	1.16516E-06	Tc00.1047053508441.70	3.29E-06
Tc00.1047053507889.20	1.19129E-06	Tc00.1047053507057.40	3.30E-06
Tc00.1047053508601.90	1.19129E-06	Tc00.1047053511031.30	3.35E-06
Tc00.1047053510657.10	1.19129E-06	Tc00.1047053504827.21	3.60E-06
Tc00.1047053504741.4	1.28832E-06	Tc00.1047053509911.10	3.89E-06
Tc00.1047053507023.40	1.28832E-06	Tc00.1047053511505.20	4.15E-06
Tc00.1047053507663.40	1.31005E-06	Tc00.1047053503823.104	4.41E-06
Tc00.1047053507775.10	1.31005E-06	Tc00.1047053510265.90	4.71E-06
Tc00.1047053503791.30	1.31337E-06	Tc00.1047053507601.90	4.72E-06
Tc00.1047053507769.20	1.31337E-06	Tc00.1047053509733.80	4.88E-06
Tc00.1047053508153.480	1.37668E-06	Tc00.1047053508693.70	5.11E-06
Tc00.1047053509399.150	1.59931E-06	Tc00.1047053509237.130	5.39E-06
Tc00.1047053510519.140	1.7489E-06	Tc00.1047053510303.299	5.44E-06
Tc00.1047053506725.80	1.88575E-06	Tc00.1047053506725.60	5.49E-06
Tc00.1047053509321.19	1.9522E-06	Tc00.1047053503879.119	5.95E-06
Tc00.1047053508891.40	2.30807E-06	Tc00.1047053504069.4	7.44E-06
Tc00.1047053508507.20	2.45827E-06	Tc00.1047053510667.100	7.44E-06
Tc00.1047053508911.60	2.47488E-06	Tc00.1047053511165.80	7.51E-06
Tc00.1047053508911.70	2.58954E-06	Tc00.1047053510305.9	8.30E-06
Tc00.1047053507623.60	2.63636E-06	Tc00.1047053504259.19	8.43E-06
Tc00.1047053508645.40	2.63636E-06	Tc00.1047053504113.30	8.45E-06
Tc00.1047053506145.10	2.9001E-06	Tc00.1047053509715.80	9.17E-06
Tc00.1047053509715.140	2.94222E-06	Tc00.1047053511517.20	1.08E-05
Tc00.1047053508799.200	3.18704E-06	Tc00.1047053507889.20	1.09E-05
Tc00.1047053508909.300	3.18704E-06	Tc00.1047053511153.90	1.14E-05
Tc00.1047053507765.130	3.39605E-06	Tc00.1047053509207.130	1.14E-05
Tc00.1047053511237.20	3.51301E-06	Tc00.1047053508873.10	1.14E-05

Table S4-3 continued

Tc00.1047053508857.140	3.53491E-06	Tc00.1047053509791.70	1.17E-05
Tc00.1047053506425.70	3.62096E-06	Tc00.1047053511385.100	1.35E-05
Tc00.1047053510725.70	3.82058E-06	Tc00.1047053509059.50	1.35E-05
Tc00.1047053506831.20	3.93747E-06	Tc00.1047053511297.10	1.38E-05
Tc00.1047053503999.40	4.5365E-06	Tc00.1047053510131.50	1.46E-05
Tc00.1047053510303.299	4.87806E-06	Tc00.1047053507017.60	1.54E-05
Tc00.1047053509399.90	5.04852E-06	Tc00.1047053507539.20	1.61E-05
Tc00.1047053508661.40	5.70129E-06	Tc00.1047053509793.20	1.62E-05
Tc00.1047053511521.30	5.73743E-06	Tc00.1047053506629.210	1.62E-05
Tc00.1047053507513.80	5.89859E-06	Tc00.1047053507103.20	1.62E-05
Tc00.1047053511511.30	5.95788E-06	Tc00.1047053506989.120	1.62E-05
Tc00.1047053508461.400	6.05467E-06	Tc00.1047053414243.20	1.62E-05
Tc00.1047053506175.60	6.76437E-06	Tc00.1047053508645.50	1.69E-05
Tc00.1047053506327.20	7.36764E-06	Tc00.1047053508965.39	1.91E-05
Tc00.1047053507801.140	7.63708E-06	Tc00.1047053507513.90	1.98E-05
Tc00.1047053503891.90	7.68623E-06	Tc00.1047053506725.40	1.99E-05
Tc00.1047053506559.20	8.13794E-06	Tc00.1047053504105.180	2.04E-05
Tc00.1047053508153.1100	8.15791E-06	Tc00.1047053511807.80	2.08E-05
Tc00.1047053511269.20	8.24627E-06	Tc00.1047053508525.10	2.10E-05
Tc00.1047053503987.30	8.46948E-06	Tc00.1047053509331.120	2.32E-05
Tc00.1047053511293.60	8.5154E-06	Tc00.1047053511277.630	2.32E-05
Tc00.1047053507023.230	9.61102E-06	Tc00.1047053511033.29	2.39E-05
Tc00.1047053511643.50	9.62557E-06	Tc00.1047053507663.40	2.79E-05
Tc00.1047053506009.80	9.89627E-06	Tc00.1047053511025.20	2.86E-05
Tc00.1047053509437.40	1.02202E-05	Tc00.1047053508525.30	2.90E-05
Tc00.1047053503917.14	1.06104E-05	Tc00.1047053508625.50	3.09E-05
Tc00.1047053511859.20	1.11201E-05	Tc00.1047053508231.174	3.10E-05
Tc00.1047053511509.50	1.13133E-05	Tc00.1047053503781.90	3.13E-05
Tc00.1047053510305.70	1.14181E-05	Tc00.1047053511031.10	3.14E-05
Tc00.1047053504131.140	1.18363E-05	Tc00.1047053506629.140	3.39E-05
Tc00.1047053506587.50	1.21845E-05	Tc00.1047053507023.40	3.65E-05
Tc00.1047053509429.160	1.22697E-05	Tc00.1047053511521.27	3.75E-05
Tc00.1047053509567.30	1.2962E-05	Tc00.1047053506855.260	3.99E-05
Tc00.1047053508693.70	1.6371E-05	Tc00.1047053504235.9	4.05E-05
Tc00.1047053511817.134	1.65402E-05	Tc00.1047053507063.150	4.18E-05
Tc00.1047053511661.80	1.66296E-05	Tc00.1047053507949.40	4.63E-05
Tc00.1047053506989.120	1.83121E-05	Tc00.1047053511261.60	4.79E-05
Tc00.1047053504949.39	1.93875E-05	Tc00.1047053509901.170	5.15E-05
Tc00.1047053506821.160	1.93875E-05	Tc00.1047053510155.30	5.25E-05
Tc00.1047053511629.20	1.98211E-05	Tc00.1047053510121.50	5.34E-05

Table S4-3 continued

Tc00.1047053509053.80	2.18717E-05	Tc00.1047053508175.20	5.45E-05
Tc00.1047053503719.39	2.57594E-05	Tc00.1047053508269.70	5.65E-05
Tc00.1047053503823.110	2.8304E-05	Tc00.1047053507601.140	5.74E-05
Tc00.1047053506475.105	2.90449E-05	Tc00.1047053506815.20	5.80E-05
Tc00.1047053509797.60	3.011E-05	Tc00.1047053509317.50	6.03E-05
Tc00.1047053503431.30	3.05239E-05	Tc00.1047053509429.130	6.09E-05
Tc00.1047053506629.200	3.28638E-05	Tc00.1047053507823.39	6.13E-05
Tc00.1047053507601.10	3.38772E-05	Tc00.1047053511719.20	6.39E-05
Tc00.1047053509237.70	3.49787E-05	Tc00.1047053507517.50	6.77E-05
Tc00.1047053511825.30	3.51318E-05	Tc00.1047053506575.9	7.49E-05
Tc00.1047053509745.70	3.64531E-05	Tc00.1047053508919.140	8.88E-05
Tc00.1047053508441.70	3.77701E-05	Tc00.1047053506195.120	8.97E-05
Tc00.1047053504231.20	3.90604E-05	Tc00.1047053509005.70	9.08E-05
Tc00.1047053507777.10	3.93022E-05	Tc00.1047053506833.30	9.08E-05
Tc00.1047053511909.70	4.10153E-05	Tc00.1047053511817.20	0.000108852
Tc00.1047053508667.20	4.31173E-05	Tc00.1047053503647.40	0.000108936
Tc00.1047053506871.50	4.34039E-05	Tc00.1047053509565.9	0.000116961
Tc00.1047053507031.29	4.71318E-05	Tc00.1047053508153.570	0.000116961
Tc00.1047053506679.240	4.80631E-05	Tc00.1047053510039.40	0.000130834
Tc00.1047053506575.9	5.0189E-05	Tc00.1047053508601.130	0.000139577
Tc00.1047053511859.30	5.0189E-05	Tc00.1047053506835.60	0.000139577
Tc00.1047053503923.30	5.42872E-05	Tc00.1047053508409.150	0.000145532
Tc00.1047053508175.170	5.46972E-05	Tc00.1047053509429.200	0.000145532
Tc00.1047053506855.300	5.58984E-05	Tc00.1047053511655.10	0.00014952
Tc00.1047053509911.110	5.70913E-05	Tc00.1047053511151.50	0.000150291
Tc00.1047053509507.49	5.92046E-05	Tc00.1047053472777.30	0.000151846
Tc00.1047053511385.40	6.31611E-05	Tc00.1047053504157.40	0.000159902
Tc00.1047053508707.160	6.73216E-05	Tc00.1047053503431.30	0.000162824
Tc00.1047053510339.20	6.7332E-05	Tc00.1047053506201.170	0.000168346
Tc00.1047053509911.10	7.50983E-05	Tc00.1047053510835.20	0.000171037
Tc00.1047053509611.10	8.33464E-05	Tc00.1047053511661.30	0.000189168
Tc00.1047053507949.40	9.41483E-05	Tc00.1047053510659.219	0.000189168
Tc00.1047053509207.150	9.66367E-05	Tc00.1047053503703.80	0.000189168
Tc00.1047053509571.60	0.000103668	Tc00.1047053510629.450	0.000199606
Tc00.1047053511657.60	0.000104114	Tc00.1047053511261.80	0.00020443
Tc00.1047053506855.260	0.00010629	Tc00.1047053506821.120	0.000206122
Tc00.1047053507017.64	0.000106961	Tc00.1047053508911.60	0.000226686
Tc00.1047053509045.10	0.000107553	Tc00.1047053509719.56	0.000230555
Tc00.1047053509057.20	0.000117416	Tc00.1047053506303.160	0.00024063
Tc00.1047053506725.60	0.000120778	Tc00.1047053507609.70	0.000249368

Table S4-3 continued

Tc00.1047053506831.10	0.000129212	Tc00.1047053508547.130	0.000263599
Tc00.1047053509429.290	0.000130587	Tc00.1047053511671.140	0.000271677
Tc00.1047053509065.36	0.000148406	Tc00.1047053506009.40	0.000306911
Tc00.1047053506297.270	0.000153545	Tc00.1047053503823.10	0.000318412
Tc00.1047053510579.110	0.000154555	Tc00.1047053510657.160	0.000318412
Tc00.1047053503911.30	0.000154558	Tc00.1047053511167.60	0.00032504
Tc00.1047053509395.100	0.000154558	Tc00.1047053506821.160	0.000325893
Tc00.1047053511017.40	0.000154558	Tc00.1047053507515.140	0.000327809
Tc00.1047053508641.320	0.000157287	Tc00.1047053506855.320	0.000346811
Tc00.1047053507771.90	0.000157321	Tc00.1047053510819.20	0.000351277
Tc00.1047053509715.104	0.000159505	Tc00.1047053510903.20	0.000353014
Tc00.1047053506303.140	0.000181949	Tc00.1047053508859.80	0.000373768
Tc00.1047053504105.200	0.000186434	Tc00.1047053504741.60	0.000381222
Tc00.1047053506195.120	0.000196344	Tc00.1047053504105.170	0.000439257
Tc00.1047053508355.420	0.00019746	Tc00.1047053509719.30	0.00044169
Tc00.1047053508525.10	0.000210764	Tc00.1047053507649.20	0.000462823
Tc00.1047053509005.70	0.000211443	Tc00.1047053509567.30	0.000468817
Tc00.1047053511725.40	0.000211443	Tc00.1047053505071.100	0.000562378
Tc00.1047053506855.120	0.000211682	Tc00.1047053506681.40	0.000568292
Tc00.1047053510885.10	0.000227425	Tc00.1047053508059.4	0.000585373
Tc00.1047053511857.59	0.000242829	Tc00.1047053506629.40	0.000596176
Tc00.1047053511855.80	0.000247321	Tc00.1047053511657.70	0.00059723
Tc00.1047053509065.9	0.000260329	Tc00.1047053503919.60	0.000599437
Tc00.1047053507649.20	0.000281857	Tc00.1047053511301.20	0.000628229
Tc00.1047053510681.30	0.00029014	Tc00.1047053508409.170	0.00062944
Tc00.1047053511523.30	0.000295476	Tc00.1047053508177.20	0.000659724
Tc00.1047053506009.40	0.000306409	Tc00.1047053511269.40	0.000690935
Tc00.1047053507071.180	0.000308905	Tc00.1047053503733.80	0.00070214
Tc00.1047053506195.260	0.00031254	Tc00.1047053506199.30	0.000707193
Tc00.1047053511277.630	0.000319463	Tc00.1047053508027.120	0.000761482
Tc00.1047053509429.200	0.000322148	Tc00.1047053508029.20	0.000776649
Tc00.1047053508707.270	0.000322486	Tc00.1047053511863.20	0.000776649
Tc00.1047053503703.80	0.000323165	Tc00.1047053507769.30	0.000779763
Tc00.1047053506479.80	0.000324377	Tc00.1047053508613.20	0.00078441
Tc00.1047053510899.19	0.000333411	Tc00.1047053509509.30	0.000786485
Tc00.1047053510657.160	0.000334423	Tc00.1047053508881.10	0.000810724
Tc00.1047053510579.10	0.000334762	Tc00.1047053506247.470	0.000872531
Tc00.1047053503879.20	0.00033637	Tc00.1047053507513.40	0.00091804
Tc00.1047053504105.180	0.00034768	Tc00.1047053511825.30	0.00093174
Tc00.1047053510899.50	0.000349549	Tc00.1047053508915.9	0.000986522

Table S4-3 continued

Tc00.1047053504827.100	0.000354596	Tc00.1047053503721.30	0.001019341
Tc00.1047053504013.70	0.000359021	Tc00.1047053509849.10	0.001071323
Tc00.1047053508153.550	0.000359021	Tc00.1047053509161.149	0.001098142
Tc00.1047053508501.240	0.000359021	Tc00.1047053511245.200	0.001099849
Tc00.1047053511719.20	0.000359021	Tc00.1047053506789.334	0.001176777
Tc00.1047053503703.70	0.000383362	Tc00.1047053508799.200	0.001332092
Tc00.1047053511505.10	0.000387313	Tc00.1047053504021.30	0.001345092
Tc00.1047053511649.110	0.000397801	Tc00.1047053509789.30	0.001345092
Tc00.1047053508177.20	0.000400646	Tc00.1047053503797.4	0.001408235
Tc00.1047053506835.60	0.000406871	Tc00.1047053508677.160	0.001456029
Tc00.1047053509733.60	0.000406871	Tc00.1047053505183.110	0.001479679
Tc00.1047053509611.140	0.00041154	Tc00.1047053506475.100	0.001560164
Tc00.1047053503797.4	0.000422614	Tc00.1047053511649.169	0.001614814
Tc00.1047053507601.60	0.000438218	Tc00.1047053506789.110	0.001725239
Tc00.1047053508231.130	0.000458471	Tc00.1047053508799.80	0.001766178
Tc00.1047053506629.40	0.000464472	Tc00.1047053507029.41	0.001788931
Tc00.1047053509901.170	0.000477904	Tc00.1047053508319.40	0.001788931
Tc00.1047053503939.100	0.000478302	Tc00.1047053509875.204	0.001857023
Tc00.1047053510155.160	0.00047849	Tc00.1047053509693.90	0.002008133
Tc00.1047053503505.14	0.000503345	Tc00.1047053508899.20	0.002026998
Tc00.1047053509793.20	0.000503345	Tc00.1047053506855.80	0.002026998
Tc00.1047053510131.60	0.000547322	Tc00.1047053508349.30	0.002026998
Tc00.1047053509207.130	0.000557025	Tc00.1047053508547.20	0.002049524
Tc00.1047053511261.60	0.000573878	Tc00.1047053510723.20	0.002055223
Tc00.1047053507029.41	0.000660484	Tc00.1047053503781.40	0.002069755
Tc00.1047053510889.190	0.000752385	Tc00.1047053508177.64	0.002257807
Tc00.1047053504003.60	0.000756422	Tc00.1047053504797.110	0.002282732
Tc00.1047053506193.20	0.000756422	Tc00.1047053508403.10	0.002288179
Tc00.1047053511151.50	0.000757254	Tc00.1047053508899.90	0.002317767
Tc00.1047053505999.70	0.000768677	Tc00.1047053508661.40	0.002397135
Tc00.1047053508059.4	0.0007753	Tc00.1047053504131.170	0.002414489
Tc00.1047053503647.40	0.000789687	Tc00.1047053508429.20	0.002463757
Tc00.1047053506315.10	0.000794135	Tc00.1047053510731.50	0.002464815
Tc00.1047053510155.30	0.000801106	Tc00.1047053509215.13	0.002525202
Tc00.1047053509789.30	0.000837248	Tc00.1047053506337.4	0.002525202
Tc00.1047053507017.60	0.000860558	Tc00.1047053508299.50	0.002534635
Tc00.1047053508817.170	0.000901111	Tc00.1047053503809.158	0.00279873
Tc00.1047053509509.30	0.000911631	Tc00.1047053511857.59	0.00285064
Tc00.1047053510657.20	0.000943553	Tc00.1047053510575.160	0.002904524
Tc00.1047053508153.1020	0.000964947	Tc00.1047053506241.190	0.002915805

Table S4-3 continued

Tc00.1047053509399.180	0.000989749	Tc00.1047053508909.200	0.003090307
Tc00.1047053507513.40	0.001034397	Tc00.1047053508547.80	0.003182073
Tc00.1047053503617.15	0.001060345	Tc00.1047053511245.100	0.003277472
Tc00.1047053510575.160	0.001072456	Tc00.1047053504175.30	0.003277664
Tc00.1047053503791.20	0.001084403	Tc00.1047053509045.10	0.003302467
Tc00.1047053506247.470	0.001135762	Tc00.1047053509213.140	0.003335194
Tc00.1047053509747.50	0.001176893	Tc00.1047053509399.160	0.003335194
Tc00.1047053509051.50	0.001237102	Tc00.1047053503995.30	0.003378617
Tc00.1047053508153.230	0.001300763	Tc00.1047053510889.30	0.003548303
Tc00.1047053511465.20	0.001345486	Tc00.1047053506445.90	0.003554739
Tc00.1047053426897.10	0.001379102	Tc00.1047053508909.80	0.003782506
Tc00.1047053506855.320	0.001416178	Tc00.1047053508153.860	0.004146935
Tc00.1047053507615.73	0.001466201	Tc00.1047053509617.40	0.004167595
Tc00.1047053511717.174	0.001466201	Tc00.1047053503635.50	0.004189561
Tc00.1047053508741.250	0.001635669	Tc00.1047053506999.160	0.004205557
Tc00.1047053507049.10	0.001732572	Tc00.1047053506425.130	0.004511938
Tc00.1047053509437.30	0.001737085	Tc00.1047053508199.90	0.004594215
Tc00.1047053506479.67	0.001758246	Tc00.1047053508819.20	0.004596244
Tc00.1047053511267.30	0.00191329	Tc00.1047053509877.10	0.004706125
Tc00.1047053511509.40	0.001947014	Tc00.1047053511819.50	0.004744545
Tc00.1047053507609.70	0.00197618	Tc00.1047053509001.40	0.004990306
Tc00.1047053511661.120	0.002180089	Tc00.1047053509717.50	0.005015149
Tc00.1047053507007.9	0.002239728	Tc00.1047053507515.100	0.006077528
Tc00.1047053509875.204	0.002259036	Tc00.1047053506559.200	0.006129172
Tc00.1047053511261.130	0.002287367	Tc00.1047053503843.30	0.006202613
Tc00.1047053511505.20	0.002287367	Tc00.1047053511261.130	0.00632248
Tc00.1047053506507.20	0.002395872	Tc00.1047053509057.20	0.006426737
Tc00.1047053511287.4	0.002418383	Tc00.1047053507159.41	0.006426737
Tc00.1047053508153.570	0.002535611	Tc00.1047053511671.100	0.006561385
Tc00.1047053511903.110	0.002588039	Tc00.1047053503925.60	0.00699018
Tc00.1047053510889.170	0.002641826	Tc00.1047053507209.30	0.00699018
Tc00.1047053506629.210	0.002690558	Tc00.1047053509877.80	0.006996401
Tc00.1047053508415.10	0.002723512	Tc00.1047053506957.23	0.007000153
Tc00.1047053510691.95	0.002723512	Tc00.1047053510899.50	0.007209251
Tc00.1047053503721.30	0.002746767	Tc00.1047053508409.300	0.007588778
Tc00.1047053506445.90	0.002805524	Tc00.1047053508231.220	0.008008363
Tc00.1047053508677.90	0.002898344	Tc00.1047053509671.90	0.00820752
Tc00.1047053414243.20	0.002911741	Tc00.1047053507049.160	0.008224425
Tc00.1047053508915.9	0.003018641	Tc00.1047053510519.149	0.008464525
Tc00.1047053509065.40	0.003116566	Tc00.1047053510973.9	0.008464525

Table S4-3 continued

Tc00.1047053511245.100	0.003143484	Tc00.1047053511647.20	0.0086631
Tc00.1047053510903.50	0.003155171	Tc00.1047053511865.40	0.008701388
Tc00.1047053503733.80	0.003166431	Tc00.1047053511297.30	0.008858946
Tc00.1047053508693.190	0.0031949	Tc00.1047053507615.60	0.009091489
Tc00.1047053510889.290	0.003298867	Tc00.1047053509611.80	0.009556749
Tc00.1047053508899.90	0.00332481	Tc00.1047053503473.20	0.009566181
Tc00.1047053507023.30	0.003330586	Tc00.1047053509671.10	0.009785528
Tc00.1047053508909.80	0.003381961	Tc00.1047053511761.80	0.009854165
Tc00.1047053508175.340	0.00343417	Tc00.1047053510659.28	0.009854426
Tc00.1047053510299.30	0.00343417	Tc00.1047053511033.20	0.01005929
Tc00.1047053511025.20	0.003440956	Tc00.1047053509399.180	0.01053218
Tc00.1047053511863.20	0.003536887	Tc00.1047053510967.10	0.01070826
Tc00.1047053506815.20	0.003554961	Tc00.1047053503635.40	0.01073911
Tc00.1047053508577.130	0.003575952	Tc00.1047053506585.50	0.01074709
Tc00.1047053506727.50	0.003615061	Tc00.1047053506629.100	0.01084397
Tc00.1047053509429.170	0.003727099	Tc00.1047053510945.60	0.01113162
Tc00.1047053507857.30	0.00376163	Tc00.1047053511277.130	0.01124126
Tc00.1047053509849.10	0.003812669	Tc00.1047053511817.10	0.01160228
Tc00.1047053510835.20	0.003886583	Tc00.1047053507519.100	0.01184027
Tc00.1047053510183.40	0.004063375	Tc00.1047053507001.20	0.01201085
Tc00.1047053506577.130	0.004088255	Tc00.1047053510305.20	0.0122026
Tc00.1047053503939.80	0.004093169	Tc00.1047053508643.40	0.01237947
Tc00.1047053508677.160	0.004102947	Tc00.1047053510423.50	0.01245682
Tc00.1047053508859.30	0.004571821	Tc00.1047053506145.30	0.01250057
Tc00.1047053509237.130	0.004571821	Tc00.1047053503565.20	0.01268701
Tc00.1047053503823.10	0.004578991	Tc00.1047053510645.20	0.01269146
Tc00.1047053507053.50	0.004824789	Tc00.1047053506679.270	0.01269797
Tc00.1047053508231.174	0.004849801	Tc00.1047053511277.483	0.01282749
Tc00.1047053508891.20	0.004849801	Tc00.1047053511163.50	0.0132422
Tc00.1047053507007.30	0.004977516	Tc00.1047053506195.200	0.01342537
Tc00.1047053506621.14	0.005016972	Tc00.1047053506419.20	0.01370273
Tc00.1047053507023.170	0.005165644	Tc00.1047053506337.40	0.01386398
Tc00.1047053503995.30	0.005356525	Tc00.1047053507615.73	0.01536061
Tc00.1047053511649.140	0.005356525	Tc00.1047053510657.20	0.0159062
Tc00.1047053509601.110	0.005716316	Tc00.1047053509399.120	0.0159062
Tc00.1047053508175.70	0.005802498	Tc00.1047053509569.120	0.01596489
Tc00.1047053504021.30	0.00588545	Tc00.1047053509029.140	0.01604169
Tc00.1047053509213.140	0.006026301	Tc00.1047053506475.130	0.01614558
Tc00.1047053504741.60	0.006160257	Tc00.1047053511755.119	0.01631139
Tc00.1047053511657.50	0.006242964	Tc00.1047053506475.110	0.01721431

Table S4-3 continued

Tc00.1047053508881.10	0.006314523	Tc00.1047053506009.90	0.01764214
Tc00.1047053509001.30	0.006314523	Tc00.1047053510729.130	0.017949
Tc00.1047053505183.110	0.00643408	Tc00.1047053504191.10	0.01821131
Tc00.1047053511755.119	0.006722057	Tc00.1047053506679.10	0.01861895
Tc00.1047053504185.9	0.006725864	Tc00.1047053507317.30	0.01881718
Tc00.1047053508199.90	0.006725864	Tc00.1047053507765.80	0.01996597
Tc00.1047053508027.120	0.006743844	Tc00.1047053506509.60	0.02045771
Tc00.1047053506559.180	0.006875324	Tc00.1047053509437.70	0.02045771
Tc00.1047053509671.90	0.00699876	Tc00.1047053506679.280	0.02052122
Tc00.1047053508547.130	0.007105927	Tc00.1047053509065.40	0.02078982
Tc00.1047053509059.50	0.007105927	Tc00.1047053509105.110	0.02169097
Tc00.1047053511033.29	0.007414603	Tc00.1047053506241.60	0.0224464
Tc00.1047053506855.60	0.007624352	Tc00.1047053509179.100	0.02322593
Tc00.1047053511167.60	0.007626323	Tc00.1047053505939.60	0.02367755
Tc00.1047053509399.130	0.007681318	Tc00.1047053506175.50	0.02388339
Tc00.1047053506175.50	0.007874719	Tc00.1047053509569.30	0.02432095
Tc00.1047053510153.10	0.007892605	Tc00.1047053508577.130	0.02442864
Tc00.1047053504105.170	0.007969034	Tc00.1047053509065.36	0.02554681
Tc00.1047053509065.30	0.007969034	Tc00.1047053510301.80	0.02565409
Tc00.1047053506921.10	0.008103392	Tc00.1047053510423.40	0.0256576
Tc00.1047053509569.170	0.00820589	Tc00.1047053504013.70	0.02618121
Tc00.1047053511859.60	0.008255607	Tc00.1047053511257.110	0.02660378
Tc00.1047053511277.115	0.008300466	Tc00.1047053510121.150	0.02661198
Tc00.1047053509877.10	0.008702938	Tc00.1047053507771.20	0.02821371
Tc00.1047053509567.60	0.009026995	Tc00.1047053509507.30	0.02848184
Tc00.1047053508909.200	0.009388635	Tc00.1047053507615.130	0.0286894
Tc00.1047053511671.100	0.009583532	Tc00.1047053504105.190	0.02963042
Tc00.1047053503635.50	0.009798777	Tc00.1047053507007.50	0.03239992
Tc00.1047053510571.10	0.01006833	Tc00.1047053503559.90	0.03475599
Tc00.1047053508177.64	0.0102304	Tc00.1047053507623.50	0.03527632
Tc00.1047053510039.40	0.01023995	Tc00.1047053507589.30	0.0369234
Tc00.1047053508409.150	0.01027838	Tc00.1047053510355.250	0.03756998
Tc00.1047053510645.20	0.01033661	Tc00.1047053508851.59	0.03770802
Tc00.1047053506855.80	0.01051055	Tc00.1047053511027.20	0.03772586
Tc00.1047053508859.80	0.01051055	Tc00.1047053504069.60	0.03830393
Tc00.1047053507615.60	0.01053589	Tc00.1047053503891.60	0.03856928
Tc00.1047053508859.100	0.01059585	Tc00.1047053507887.30	0.03856928
Tc00.1047053505939.60	0.01065815	Tc00.1047053510221.20	0.04072744
Tc00.1047053503823.90	0.01133726	Tc00.1047053507717.20	0.04226459
Tc00.1047053511165.20	0.01133726	Tc00.1047053511763.19	0.04400537

Table S4-3 continued

Tc00.1047053503559.90	0.01146565	Tc00.1047053506435.20	0.04403863
Tc00.1047053508873.10	0.01157572	Tc00.1047053503431.100	0.04416881
Tc00.1047053506337.4	0.01172519	Tc00.1047053506789.249	0.04442178
Tc00.1047053509161.149	0.01174144	Tc00.1047053511287.130	0.04484752
Tc00.1047053509215.13	0.01176061	Tc00.1047053509429.240	0.04493816
Tc00.1047053506475.30	0.01253526	Tc00.1047053511277.160	0.04537006
Tc00.1047053509001.40	0.01300524	Tc00.1047053506789.150	0.0463874
Tc00.1047053510723.20	0.01306404	Tc00.1047053503923.20	0.04700766
Tc00.1047053509693.90	0.01347333	Tc00.1047053509237.120	0.04835118
Tc00.1047053511033.20	0.01347333	Tc00.1047053506679.240	0.0485835
Tc00.1047053503773.20	0.01401201	Tc00.1047053503939.100	0.04871068
Tc00.1047053508711.10	0.01401201	Tc00.1047053509331.50	0.04900964
Tc00.1047053506583.50	0.01413138	Tc00.1047053474937.9	0.0500653
Tc00.1047053504131.80	0.01440737	Tc00.1047053504157.120	0.0514822
Tc00.1047053505071.100	0.01440737	Tc00.1047053509429.170	0.05155475
Tc00.1047053506789.110	0.01475927	Tc00.1047053509875.250	0.0521564
Tc00.1047053509683.60	0.01489797	Tc00.1047053506435.134	0.05548612
Tc00.1047053507317.20	0.01516316	Tc00.1047053510435.40	0.05548612
Tc00.1047053511807.50	0.01569722	Tc00.1047053508707.70	0.05581635
Tc00.1047053410199.4	0.01606775	Tc00.1047053507895.170	0.05585666
Tc00.1047053504105.210	0.01698474	Tc00.1047053503773.20	0.0567614
Tc00.1047053508231.220	0.01741098	Tc00.1047053507025.50	0.0567614
Tc00.1047053506475.110	0.01773155	Tc00.1047053510265.14	0.05700862
Tc00.1047053511807.80	0.01887979	Tc00.1047053410199.4	0.05714994
Tc00.1047053505171.70	0.01948702	Tc00.1047053506195.90	0.05857532
Tc00.1047053506195.90	0.01948702	Tc00.1047053509039.10	0.05868355
Tc00.1047053472777.30	0.02042673	Tc00.1047053506587.50	0.05868355
Tc00.1047053511027.20	0.02072606	Tc00.1047053511165.20	0.05982372
Tc00.1047053510661.40	0.02108642	Tc00.1047053504021.149	0.0599388
Tc00.1047053506999.160	0.02121988	Tc00.1047053508741.380	0.06108992
Tc00.1047053507001.20	0.02149135	Tc00.1047053503809.130	0.06126052
Tc00.1047053508547.20	0.02149135	Tc00.1047053506835.99	0.06130142
Tc00.1047053508501.270	0.02157523	Tc00.1047053511167.100	0.06134892
Tc00.1047053508741.130	0.0219661	Tc00.1047053508297.41	0.0641931
Tc00.1047053511017.60	0.02231083	Tc00.1047053506629.170	0.06591697
Tc00.1047053508461.70	0.02281343	Tc00.1047053510041.30	0.06652514
Tc00.1047053504209.10	0.02294696	Tc00.1047053505999.24	0.06863082
Tc00.1047053508355.124	0.02393433	Tc00.1047053506459.270	0.06868647
Tc00.1047053509693.120	0.0241067	Tc00.1047053507951.150	0.07046405
Tc00.1047053510977.9	0.02411675	Tc00.1047053509331.20	0.07072723

Table S4-3 continued

Tc00.1047053503939.120	0.02427904	Tc00.1047053509713.30	0.0727486
Tc00.1047053510305.20	0.02490963	Tc00.1047053509857.50	0.07431266
Tc00.1047053504575.10	0.02498384	Tc00.1047053510359.290	0.07657531
Tc00.1047053503809.75	0.02515304	Tc00.1047053507023.70	0.07919138
Tc00.1047053510729.130	0.02547256	Tc00.1047053509617.70	0.07919138
Tc00.1047053505169.21	0.02584172	Tc00.1047053509399.100	0.07957133
Tc00.1047053507221.20	0.02602673	Tc00.1047053503593.70	0.08225736
Tc00.1047053508795.10	0.02667363	Tc00.1047053505163.60	0.08225736
Tc00.1047053510355.250	0.02703651	Tc00.1047053511237.50	0.08415928
Tc00.1047053509875.250	0.02729666	Tc00.1047053507027.59	0.08479235
Tc00.1047053508027.70	0.02735462	Tc00.1047053509337.15	0.08836633
Tc00.1047053506389.70	0.02747714	Tc00.1047053507823.10	0.09054502
Tc00.1047053508059.20	0.02832995	Tc00.1047053503733.40	0.0907272
Tc00.1047053509719.56	0.0284042	Tc00.1047053511151.11	0.09157776
Tc00.1047053511847.20	0.02860659	Tc00.1047053511209.50	0.091844
Tc00.1047053506559.200	0.02892066	Tc00.1047053507769.20	0.09384534
Tc00.1047053509337.19	0.02908614	Tc00.1047053509001.20	0.09730712
Tc00.1047053507209.60	0.02941852	Tc00.1047053504827.150	0.09783356
Tc00.1047053507769.30	0.02992241	Tc00.1047053504231.10	0.1001587
Tc00.1047053506419.20	0.03037961	Tc00.1047053460125.10	0.1042056
Tc00.1047053506585.50	0.03063945	Tc00.1047053504243.49	0.1052739
Tc00.1047053503891.60	0.0307593	Tc00.1047053510941.3	0.1068451
Tc00.1047053508409.300	0.03086165	Tc00.1047053508741.370	0.1111228
Tc00.1047053506241.190	0.03103456	Tc00.1047053506509.30	0.1141053
Tc00.1047053507021.110	0.03211255	Tc00.1047053507241.20	0.1141595
Tc00.1047053508153.400	0.03404032	Tc00.1047053506913.10	0.1145779
Tc00.1047053511821.40	0.03491618	Tc00.1047053508501.270	0.1155674
Tc00.1047053509237.120	0.0355378	Tc00.1047053503939.130	0.1161984
Tc00.1047053511649.39	0.0355378	Tc00.1047053504575.10	0.1162597
Tc00.1047053507825.40	0.03594839	Tc00.1047053505183.10	0.1167761
Tc00.1047053508707.60	0.03602386	Tc00.1047053509857.10	0.1182501
Tc00.1047053507765.80	0.03636043	Tc00.1047053503703.60	0.1200884
Tc00.1047053508643.40	0.03636043	Tc00.1047053510039.109	0.1215024
Tc00.1047053511293.90	0.03639193	Tc00.1047053509017.40	0.1244278
Tc00.1047053511671.140	0.03746005	Tc00.1047053511289.100	0.1244452
Tc00.1047053508999.80	0.03832261	Tc00.1047053506581.60	0.124949
Tc00.1047053508299.70	0.03865377	Tc00.1047053510183.40	0.1249921
Tc00.1047053511287.130	0.03894773	Tc00.1047053503923.30	0.1282534
Tc00.1047053507517.40	0.04163083	Tc00.1047053507849.60	0.1285049
Tc00.1047053507771.20	0.04163083	Tc00.1047053508741.360	0.1291135

Table S4-3 continued

Tc00.1047053507913.39	0.04225072	Tc00.1047053510329.260	0.1347565
Tc00.1047053511277.483	0.04255834	Tc00.1047053511859.60	0.1357153
Tc00.1047053503565.20	0.04471603	Tc00.1047053508741.130	0.1383916
Tc00.1047053508657.10	0.04527747	Tc00.1047053504131.100	0.1400189
Tc00.1047053509029.90	0.0459752	Tc00.1047053509999.30	0.1441045
Tc00.1047053511633.79	0.04657357	Tc00.1047053508699.60	0.144773
Tc00.1047053510667.100	0.04659344	Tc00.1047053510299.50	0.1494974
Tc00.1047053508153.860	0.04854455	Tc00.1047053503993.10	0.1539756
Tc00.1047053405737.14	0.04864506	Tc00.1047053506975.69	0.1544189
Tc00.1047053509207.40	0.04882193	Tc00.1047053505163.30	0.154479
Tc00.1047053507519.154	0.04935256	Tc00.1047053507851.30	0.1566195
Tc00.1047053506679.280	0.05004826	Tc00.1047053511459.50	0.1566195
Tc00.1047053507623.50	0.05004826	Tc00.1047053511817.280	0.156798
Tc00.1047053511865.40	0.05005741	Tc00.1047053510731.120	0.1572984
Tc00.1047053506195.200	0.05104407	Tc00.1047053506999.20	0.1572984
Tc00.1047053507057.4	0.05341678	Tc00.1047053510299.40	0.1574789
Tc00.1047053503431.100	0.05391191	Tc00.1047053507623.101	0.1579924
Tc00.1047053504131.100	0.05397148	Tc00.1047053509437.50	0.1637152
Tc00.1047053504253.30	0.05712918	Tc00.1047053511807.50	0.1657003
Tc00.1047053506679.270	0.05847359	Tc00.1047053506841.20	0.1691957
Tc00.1047053509999.140	0.05978926	Tc00.1047053446067.9	0.1694623
Tc00.1047053510945.60	0.06202291	Tc00.1047053510665.20	0.1702124
Tc00.1047053506629.170	0.06260438	Tc00.1047053506459.250	0.1702124
Tc00.1047053510659.28	0.06385461	Tc00.1047053508059.24	0.1764858
Tc00.1047053509179.100	0.0640603	Tc00.1047053508711.10	0.1780291
Tc00.1047053511763.19	0.06494676	Tc00.1047053503537.18	0.1798834
Tc00.1047053505163.60	0.06594734	Tc00.1047053504741.240	0.1899019
Tc00.1047053510303.160	0.06602121	Tc00.1047053504257.60	0.195341
Tc00.1047053507159.41	0.06605781	Tc00.1047053506855.310	0.1963278
Tc00.1047053507849.70	0.06815853	Tc00.1047053508507.24	0.1963278
Tc00.1047053511823.14	0.06964891	Tc00.1047053507087.60	0.1963278
Tc00.1047053508303.44	0.06993942	Tc00.1047053506905.20	0.1974676
Tc00.1047053506815.40	0.07273798	Tc00.1047053508859.100	0.1984359
Tc00.1047053510301.80	0.07340717	Tc00.1047053510155.115	0.2017372
Tc00.1047053507023.20	0.07369944	Tc00.1047053511463.20	0.2024183
Tc00.1047053511903.60	0.07375977	Tc00.1047053509875.240	0.2035114
Tc00.1047053507951.150	0.07409007	Tc00.1047053508859.40	0.203841
Tc00.1047053504081.270	0.074173	Tc00.1047053511463.10	0.203841
Tc00.1047053511277.20	0.074205	Tc00.1047053508349.10	0.2052039
Tc00.1047053507025.50	0.07496687	Tc00.1047053511649.140	0.2099816

Table S4-3 continued

Tc00.1047053507317.30	0.07496687	Tc00.1047053510339.80	0.2099816
Tc00.1047053511517.20	0.0764194	Tc00.1047053503625.20	0.2099816
Tc00.1047053509901.150	0.07673473	Tc00.1047053506303.70	0.2186482
Tc00.1047053506145.30	0.07675121	Tc00.1047053508999.260	0.2201392
Tc00.1047053506725.40	0.07675121	Tc00.1047053509487.20	0.2244582
Tc00.1047053505999.170	0.07733452	Tc00.1047053504209.10	0.2303355
Tc00.1047053509791.170	0.07733452	Tc00.1047053507801.179	0.2305716
Tc00.1047053506459.250	0.07750422	Tc00.1047053509049.20	0.2305716
Tc00.1047053431849.20	0.07789587	Tc00.1047053511511.150	0.2328667
Tc00.1047053503719.30	0.07861848	Tc00.1047053510131.60	0.2343593
Tc00.1047053511261.80	0.07861848	Tc00.1047053506247.30	0.2388672
Tc00.1047053509507.30	0.07969366	Tc00.1047053506177.50	0.2390935
Tc00.1047053508895.20	0.079879	Tc00.1047053511671.50	0.2409288
Tc00.1047053510329.260	0.07990946	Tc00.1047053509399.140	0.2409288
Tc00.1047053503733.40	0.0806332	Tc00.1047053510303.70	0.2441976
Tc00.1047053506679.10	0.0806332	Tc00.1047053509029.70	0.2446829
Tc00.1047053507087.80	0.0806332	Tc00.1047053511815.170	0.2452842
Tc00.1047053507513.90	0.08080364	Tc00.1047053509733.50	0.24773
Tc00.1047053508919.140	0.08190328	Tc00.1047053506991.19	0.2477418
Tc00.1047053504259.19	0.08202782	Tc00.1047053511483.50	0.2512036
Tc00.1047053511165.80	0.08294058	Tc00.1047053511289.80	0.253429
Tc00.1047053509163.60	0.08493055	Tc00.1047053510659.130	0.2548801
Tc00.1047053511167.100	0.08864218	Tc00.1047053507765.110	0.2557228
Tc00.1047053503923.20	0.0887965	Tc00.1047053509459.70	0.2557228
Tc00.1047053507895.170	0.08912198	Tc00.1047053506297.80	0.2579381
Tc00.1047053511817.100	0.08980407	Tc00.1047053509901.100	0.2581336
Tc00.1047053511483.50	0.0900472	Tc00.1047053508175.90	0.2586493
Tc00.1047053506835.99	0.09103792	Tc00.1047053508153.920	0.2617797
Tc00.1047053504131.170	0.09210818	Tc00.1047053511245.140	0.2625766
Tc00.1047053510131.20	0.0935569	Tc00.1047053506241.100	0.2627926
Tc00.1047053507887.30	0.0940556	Tc00.1047053506789.220	0.263613
Tc00.1047053509505.80	0.09645185	Tc00.1047053510689.60	0.2649749
Tc00.1047053509039.10	0.09738568	Tc00.1047053509719.26	0.2649749
Tc00.1047053504243.49	0.1022733	Tc00.1047053511857.20	0.2660517
Tc00.1047053507823.10	0.1023784	Tc00.1047053511515.9	0.2674401
Tc00.1047053503891.80	0.1028743	Tc00.1047053506195.210	0.2683001
Tc00.1047053508707.110	0.1028743	Tc00.1047053504259.10	0.2799724
Tc00.1047053506913.10	0.1030807	Tc00.1047053509461.20	0.2830231
Tc00.1047053507515.140	0.1076309	Tc00.1047053505171.70	0.2830231
Tc00.1047053511151.90	0.1076309	Tc00.1047053504003.50	0.2840204

Table S4-3 continued

Tc00.1047053506839.10	0.109345	Tc00.1047053503559.80	0.2854019
Tc00.1047053510819.20	0.1096099	Tc00.1047053508737.70	0.2854019
Tc00.1047053510653.10	0.1096939	Tc00.1047053506583.50	0.2893419
Tc00.1047053503703.90	0.1114884	Tc00.1047053504003.40	0.2976194
Tc00.1047053509459.50	0.1117178	Tc00.1047053511151.60	0.2999498
Tc00.1047053511209.50	0.1117178	Tc00.1047053503413.4	0.2999498
Tc00.1047053505939.40	0.1140296	Tc00.1047053504157.80	0.3030452
Tc00.1047053503465.10	0.1180433	Tc00.1047053506201.84	0.3034734
Tc00.1047053507209.40	0.1180433	Tc00.1047053509571.60	0.3037242
Tc00.1047053511507.60	0.1184874	Tc00.1047053505939.40	0.3037242
Tc00.1047053511237.50	0.119476	Tc00.1047053507645.70	0.3085625
Tc00.1047053507717.20	0.1212803	Tc00.1047053509991.100	0.3092493
Tc00.1047053506475.40	0.1217476	Tc00.1047053503939.70	0.3125917
Tc00.1047053511459.50	0.1217476	Tc00.1047053510655.120	0.3260752
Tc00.1047053508693.150	0.1217977	Tc00.1047053511153.60	0.3294389
Tc00.1047053506435.134	0.123583	Tc00.1047053511517.99	0.3387316
Tc00.1047053510731.50	0.1259636	Tc00.1047053437575.18	0.3441314
Tc00.1047053510121.150	0.1265525	Tc00.1047053511277.140	0.3471758
Tc00.1047053506009.90	0.1282169	Tc00.1047053506593.50	0.3507948
Tc00.1047053509253.40	0.1300999	Tc00.1047053511837.50	0.3509268
Tc00.1047053510359.290	0.1302538	Tc00.1047053506195.140	0.3524277
Tc00.1047053506581.60	0.1304161	Tc00.1047053509065.170	0.3560697
Tc00.1047053511321.40	0.1305721	Tc00.1047053503999.20	0.3577073
Tc00.1047053508533.30	0.1322586	Tc00.1047053511323.60	0.3615339
Tc00.1047053511153.100	0.1382718	Tc00.1047053511291.70	0.3615339
Tc00.1047053503781.40	0.1456186	Tc00.1047053509941.140	0.3637755
Tc00.1047053507991.120	0.1460754	Tc00.1047053507049.190	0.3649475
Tc00.1047053510221.20	0.1478918	Tc00.1047053507801.160	0.3654883
Tc00.1047053507019.10	0.1488727	Tc00.1047053511519.9	0.3666345
Tc00.1047053511291.50	0.1488727	Tc00.1047053510725.50	0.3667956
Tc00.1047053506819.10	0.1551389	Tc00.1047053510131.90	0.3788398
Tc00.1047053511817.10	0.1551389	Tc00.1047053511269.60	0.3810701
Tc00.1047053510731.120	0.156042	Tc00.1047053503965.30	0.3810701
Tc00.1047053506509.60	0.1582426	Tc00.1047053508409.240	0.3811065
Tc00.1047053510799.9	0.1583611	Tc00.1047053511903.190	0.3840318
Tc00.1047053504231.10	0.159985	Tc00.1047053510655.110	0.3859837
Tc00.1047053503539.20	0.1611118	Tc00.1047053506491.20	0.3924485
Tc00.1047053506999.20	0.1611118		
Tc00.1047053508891.60	0.1613599		
Tc00.1047053508741.390	0.1634318		

Table S4-3 continued

Tc00.1047053504157.80	0.1645433		
Tc00.1047053507031.130	0.1645433		
Tc00.1047053510329.220	0.1655652		
Tc00.1047053510665.20	0.166488		
Tc00.1047053509569.120	0.1673633		
Tc00.1047053509683.10	0.1690067		
Tc00.1047053503929.40	0.1698899		
Tc00.1047053505163.30	0.1698899		
Tc00.1047053505999.24	0.1698899		
Tc00.1047053509331.20	0.1759448		
Tc00.1047053508645.50	0.1776568		
Tc00.1047053509999.30	0.1805559		
Tc00.1047053474937.9	0.1828999		
Tc00.1047053511291.40	0.1873194		
Tc00.1047053507031.150	0.1895072		
Tc00.1047053504105.190	0.197591		
Tc00.1047053508661.70	0.197591		
Tc00.1047053509029.80	0.197591		
Tc00.1047053509047.50	0.197591		
Tc00.1047053508455.30	0.1979171		
Tc00.1047053511247.18	0.1981606		
Tc00.1047053509561.9	0.200847		
Tc00.1047053508699.60	0.2039818		
Tc00.1047053508741.370	0.206566		
Tc00.1047053511245.140	0.2107187		
Tc00.1047053504191.10	0.2116099		
Tc00.1047053508349.10	0.213975		
Tc00.1047053507589.30	0.2165933		
Tc00.1047053507023.110	0.2189303		
Tc00.1047053507851.30	0.2210209		
Tc00.1047053508819.20	0.2210209		
Tc00.1047053510357.130	0.2210209		
Tc00.1047053506789.310	0.2236893		
Tc00.1047053510329.320	0.2303732		
Tc00.1047053506297.130	0.2309315		
Tc00.1047053509055.60	0.2309315		
Tc00.1047053506789.220	0.23108		
Tc00.1047053508263.30	0.2316695		
Tc00.1047053507049.160	0.2317279		
Tc00.1047053507087.60	0.2339151		

Table S4-3 continued

Tc00.1047053511647.20	0.2361549		
Tc00.1047053507715.90	0.2369657		
Tc00.1047053510943.70	0.2369657		
Tc00.1047053509621.10	0.2415464		
Tc00.1047053511517.99	0.2433731		
Tc00.1047053504003.50	0.2435074		
Tc00.1047053506009.100	0.2438434		
Tc00.1047053507801.179	0.2438434		
Tc00.1047053503843.30	0.2490993		
Tc00.1047053506479.110	0.2490993		
Tc00.1047053511807.40	0.2501496		
Tc00.1047053506851.20	0.2523022		
Tc00.1047053505183.10	0.2523022		
Tc00.1047053507017.130	0.2523022		
Tc00.1047053506489.59	0.2549258		
Tc00.1047053509399.160	0.2557208		
Tc00.1047053509875.240	0.2579		
Tc00.1047053509733.50	0.2584277		
Tc00.1047053506175.10	0.2602137		
Tc00.1047053509857.10	0.2633495		
Tc00.1047053511281.60	0.2682359		
Tc00.1047053510689.60	0.2724182		
Tc00.1047053509487.20	0.2732843		
Tc00.1047053506727.70	0.2732843		
Tc00.1047053509063.10	0.273379		
Tc00.1047053509109.114	0.2735688		
Tc00.1047053503809.130	0.2760834		
Tc00.1047053506275.60	0.2763509		
Tc00.1047053508859.40	0.277089		
Tc00.1047053503559.80	0.277089		
Tc00.1047053511463.10	0.277089		
Tc00.1047053507849.60	0.277089		
Tc00.1047053508029.20	0.2793889		
Tc00.1047053509713.30	0.2820969		
Tc00.1047053506435.210	0.2835839		
Tc00.1047053510661.100	0.2851166		
Tc00.1047053509565.9	0.2960058		
Tc00.1047053508961.10	0.2967474		
Tc00.1047053437575.18	0.3023165		
Tc00.1047053511727.129	0.3034761		

Table S4-3 continued

Tc00.1047053506435.120	0.3034761		
Tc00.1047053506561.10	0.304595		
Tc00.1047053506435.280	0.304595		
Tc00.1047053511263.50	0.3084086		
Tc00.1047053509287.110	0.3087669		
Tc00.1047053511287.80	0.3089775		
Tc00.1047053504003.40	0.3102417		
Tc00.1047053508741.380	0.310414		
Tc00.1047053507049.190	0.310414		
Tc00.1047053509001.20	0.310414		
Tc00.1047053506575.60	0.3110749		
Tc00.1047053510339.80	0.3120428		
Tc00.1047053509461.20	0.3141565		
Tc00.1047053511819.50	0.3144478		
Tc00.1047053508153.1110	0.315844		
Tc00.1047053509053.120	0.3205837		
Tc00.1047053511289.80	0.3223723		
Tc00.1047053510131.90	0.3225838		

Table S4-4. *T.cruzi* gene codes for orthologous protein predicted to be under PS with models M7 versus M8 with 4 and 3 taxa.

4 taxa analysis		3 taxa analysis	
Gene codes	**q value**	**Gene codes**	**q value**
Tc00.1047053509791.70	2.53E-35	Tc00.1047053509023.200	1.17331E-40
Tc00.1047053509023.200	2.59E-34	Tc00.1047053511287.10	1.32698E-31
Tc00.1047053508895.40	1.51E-29	Tc00.1047053509229.40	1.22982E-29
Tc00.1047053508299.30	1.87E-29	Tc00.1047053510359.80	4.09425E-29
Tc00.1047053508175.370	2.21E-29	Tc00.1047053508045.4	4.00221E-28
Tc00.1047053509745.60	1.55E-27	Tc00.1047053506989.70	1.88232E-24
Tc00.1047053506821.120	2.08E-27	Tc00.1047053506477.30	4.94532E-22
Tc00.1047053506989.70	5.20E-27	Tc00.1047053509799.140	5.16886E-22
Tc00.1047053507601.90	1.72E-25	Tc00.1047053511807.60	1.67429E-21
Tc00.1047053503573.9	2.62E-25	Tc00.1047053508865.20	3.67187E-21
Tc00.1047053508045.4	8.13E-24	Tc00.1047053506415.10	1.59478E-20
Tc00.1047053505183.40	2.39E-23	Tc00.1047053507059.80	5.10425E-20
Tc00.1047053510359.80	5.05E-23	Tc00.1047053507491.50	5.26335E-20
Tc00.1047053508699.10	9.22E-22	Tc00.1047053508355.70	6.1666E-20
Tc00.1047053506959.90	1.13E-20	Tc00.1047053506957.80	3.76778E-19
Tc00.1047053511807.60	1.98E-20	Tc00.1047053509719.20	1.11532E-18
Tc00.1047053509641.50	4.19E-20	Tc00.1047053508153.1010	2.26133E-18
Tc00.1047053509229.40	2.22E-19	Tc00.1047053509051.50	2.83285E-18
Tc00.1047053511287.10	4.33E-19	Tc00.1047053505183.20	3.05527E-18
Tc00.1047053506833.30	7.74E-19	Tc00.1047053510729.100	4.13844E-18
Tc00.1047053503809.158	1.35E-18	Tc00.1047053507895.160	7.19182E-18
Tc00.1047053506247.30	1.79E-18	Tc00.1047053509353.50	1.31853E-17
Tc00.1047053507519.100	3.06E-18	Tc00.1047053507771.60	3.60741E-17
Tc00.1047053507491.50	9.18E-18	Tc00.1047053511285.80	6.63678E-17
Tc00.1047053506195.270	2.09E-17	Tc00.1047053509429.230	7.34906E-17
Tc00.1047053509857.40	2.09E-17	Tc00.1047053508141.60	1.80807E-16
Tc00.1047053509607.30	2.40E-17	Tc00.1047053511865.50	1.94713E-16
Tc00.1047053508799.240	3.03E-17	Tc00.1047053509641.50	2.01E-16
Tc00.1047053503847.60	3.98E-17	Tc00.1047053506619.90	2.31337E-16
Tc00.1047053509617.80	1.23E-16	Tc00.1047053504021.50	5.66779E-16
Tc00.1047053511285.80	2.03E-16	Tc00.1047053506725.80	5.85779E-16
Tc00.1047053510889.130	2.78E-16	Tc00.1047053508891.4	1.41024E-15
Tc00.1047053509695.100	9.29E-16	Tc00.1047053509207.110	1.73729E-15
Tc00.1047053508709.10	9.42E-16	Tc00.1047053510039.80	2.25324E-15
Tc00.1047053508613.20	1.25E-15	Tc00.1047053507601.60	3.86391E-15
Tc00.1047053510885.80	1.39E-15	Tc00.1047053510579.119	9.30233E-15
Tc00.1047053506615.50	1.44E-15	Tc00.1047053510735.50	9.30233E-15

Table S4-4 continued

Tc00.1047053510121.40	1.80E-15	Tc00.1047053504131.140	9.42641E-15
Tc00.1047053511385.100	1.81E-15	Tc00.1047053506459.280	1.53197E-14
Tc00.1047053511153.120	4.55E-15	Tc00.1047053503999.40	2.91982E-14
Tc00.1047053511239.120	1.42E-14	Tc00.1047053506959.80	3.14966E-14
Tc00.1047053508153.1010	1.62E-14	Tc00.1047053506615.50	3.14966E-14
Tc00.1047053509799.140	4.46E-14	Tc00.1047053511859.80	3.43924E-14
Tc00.1047053506459.280	1.09E-13	Tc00.1047053510305.70	3.49552E-14
Tc00.1047053503879.119	1.30E-13	Tc00.1047053509875.180	6.16926E-14
Tc00.1047053506425.130	1.66E-13	Tc00.1047053508543.180	6.54711E-14
Tc00.1047053510121.140	2.58E-13	Tc00.1047053509605.20	7.46655E-14
Tc00.1047053511385.70	3.79E-13	Tc00.1047053511031.40	1.19332E-13
Tc00.1047053507023.90	5.63E-13	Tc00.1047053508895.60	1.19332E-13
Tc00.1047053508543.10	6.06E-13	Tc00.1047053508299.30	1.22053E-13
Tc00.1047053508891.4	7.28E-13	Tc00.1047053509569.140	3.24702E-13
Tc00.1047053511461.50	1.03E-12	Tc00.1047053508899.130	3.37819E-13
Tc00.1047053508175.50	1.58E-12	Tc00.1047053509637.27	4.13537E-13
Tc00.1047053509671.168	1.77E-12	Tc00.1047053508059.10	4.19609E-13
Tc00.1047053507771.60	1.88E-12	Tc00.1047053507895.140	1.34967E-12
Tc00.1047053509719.20	2.07E-12	Tc00.1047053511283.140	1.44134E-12
Tc00.1047053511865.50	3.50E-12	Tc00.1047053504105.230	1.70391E-12
Tc00.1047053506989.150	3.99E-12	Tc00.1047053507895.164	2.67279E-12
Tc00.1047053507053.209	4.12E-12	Tc00.1047053509745.60	3.82882E-12
Tc00.1047053508175.90	4.24E-12	Tc00.1047053503823.40	6.54386E-12
Tc00.1047053510735.50	4.50E-12	Tc00.1047053506151.10	6.72304E-12
Tc00.1047053511231.14	6.60E-12	Tc00.1047053506957.100	8.89791E-12
Tc00.1047053509569.140	6.65E-12	Tc00.1047053508577.160	9.22263E-12
Tc00.1047053507053.50	6.90E-12	Tc00.1047053504021.90	1.19282E-11
Tc00.1047053508593.110	7.71E-12	Tc00.1047053504253.10	1.21904E-11
Tc00.1047053511859.80	8.66E-12	Tc00.1047053511757.10	1.46253E-11
Tc00.1047053509429.230	9.65E-12	Tc00.1047053510667.110	1.49807E-11
Tc00.1047053510039.80	1.12E-11	Tc00.1047053504797.80	1.70071E-11
Tc00.1047053509029.140	1.17E-11	Tc00.1047053510575.130	2.95936E-11
Tc00.1047053510667.110	1.61E-11	Tc00.1047053508415.30	3.67002E-11
Tc00.1047053509637.27	1.63E-11	Tc00.1047053508645.40	5.03444E-11
Tc00.1047053509733.80	1.63E-11	Tc00.1047053511521.40	6.61205E-11
Tc00.1047053510895.20	1.74E-11	Tc00.1047053507715.40	7.77088E-11
Tc00.1047053504105.230	1.93E-11	Tc00.1047053509229.120	8.43844E-11
Tc00.1047053507969.30	2.33E-11	Tc00.1047053510665.40	9.9623E-11
Tc00.1047053509395.59	2.95E-11	Tc00.1047053511501.10	1.0235E-10
Tc00.1047053511653.10	3.75E-11	Tc00.1047053510339.70	1.05073E-10

Table S4-4 continued

Tc00.1047053504021.90	3.80E-11	Tc00.1047053511661.10	1.45469E-10
Tc00.1047053509207.110	4.32E-11	Tc00.1047053509671.168	1.55947E-10
Tc00.1047053507771.80	4.40E-11	Tc00.1047053511263.30	1.58931E-10
Tc00.1047053508543.100	5.06E-11	Tc00.1047053508593.110	1.71462E-10
Tc00.1047053509353.50	5.11E-11	Tc00.1047053509799.120	2.15724E-10
Tc00.1047053509799.120	6.15E-11	Tc00.1047053506425.70	2.58179E-10
Tc00.1047053427091.10	8.77E-11	Tc00.1047053503729.10	3.36266E-10
Tc00.1047053503823.40	1.08E-10	Tc00.1047053507003.24	3.36266E-10
Tc00.1047053507895.164	1.08E-10	Tc00.1047053510967.30	4.71177E-10
Tc00.1047053508899.130	1.17E-10	Tc00.1047053506195.270	4.71177E-10
Tc00.1047053507895.140	1.30E-10	Tc00.1047053503411.10	5.79459E-10
Tc00.1047053511031.40	1.42E-10	Tc00.1047053508153.480	5.79459E-10
Tc00.1047053504253.10	1.97E-10	Tc00.1047053507649.40	6.34988E-10
Tc00.1047053506957.100	2.99E-10	Tc00.1047053507777.10	7.79249E-10
Tc00.1047053506679.190	3.24E-10	Tc00.1047053506195.260	9.12783E-10
Tc00.1047053506275.20	4.53E-10	Tc00.1047053511237.100	1.05445E-09
Tc00.1047053508059.10	4.88E-10	Tc00.1047053509331.160	1.20536E-09
Tc00.1047053506303.10	4.99E-10	Tc00.1047053504059.20	1.51145E-09
Tc00.1047053506959.80	8.63E-10	Tc00.1047053509797.60	1.53141E-09
Tc00.1047053503411.10	1.25E-09	Tc00.1047053509791.170	1.53535E-09
Tc00.1047053508413.40	1.27E-09	Tc00.1047053511209.40	1.58805E-09
Tc00.1047053510655.60	1.31E-09	Tc00.1047053510579.110	1.90243E-09
Tc00.1047053509505.40	1.33E-09	Tc00.1047053506301.30	1.96654E-09
Tc00.1047053506619.90	1.35E-09	Tc00.1047053509053.80	1.98766E-09
Tc00.1047053511237.100	1.36E-09	Tc00.1047053506989.150	2.03835E-09
Tc00.1047053506301.30	1.66E-09	Tc00.1047053506629.200	2.17348E-09
Tc00.1047053506297.340	1.70E-09	Tc00.1047053507771.80	3.0761E-09
Tc00.1047053510339.70	1.92E-09	Tc00.1047053503917.14	3.64534E-09
Tc00.1047053508461.400	1.93E-09	Tc00.1047053509207.40	5.23156E-09
Tc00.1047053509229.120	2.41E-09	Tc00.1047053507773.30	5.46736E-09
Tc00.1047053508415.30	2.54E-09	Tc00.1047053511277.480	5.55713E-09
Tc00.1047053506151.10	2.77E-09	Tc00.1047053507969.30	5.55713E-09
Tc00.1047053511321.26	2.87E-09	Tc00.1047053506511.30	6.13344E-09
Tc00.1047053504021.50	2.91E-09	Tc00.1047053508891.40	6.1463E-09
Tc00.1047053509877.80	3.39E-09	Tc00.1047053504175.20	7.82535E-09
Tc00.1047053511283.140	3.58E-09	Tc00.1047053511755.60	8.22239E-09
Tc00.1047053510435.20	3.70E-09	Tc00.1047053507023.230	8.22239E-09
Tc00.1047053511277.480	3.71E-09	Tc00.1047053508507.20	8.2579E-09
Tc00.1047053506477.60	3.84E-09	Tc00.1047053506789.310	1.4177E-08
Tc00.1047053507513.100	3.84E-09	Tc00.1047053508741.340	2.10981E-08

Table S4-4 continued

Tc00.1047053509049.10	4.63E-09	Tc00.1047053506389.70	2.37684E-08
Tc00.1047053507057.40	5.23E-09	Tc00.1047053511855.20	2.65898E-08
Tc00.1047053508895.60	5.30E-09	Tc00.1047053506195.230	2.82972E-08
Tc00.1047053510329.10	5.44E-09	Tc00.1047053506145.10	3.22666E-08
Tc00.1047053511521.40	5.44E-09	Tc00.1047053511859.20	3.5255E-08
Tc00.1047053507823.39	8.03E-09	Tc00.1047053508911.70	3.58491E-08
Tc00.1047053506985.40	9.19E-09	Tc00.1047053505999.120	4.25864E-08
Tc00.1047053504113.30	1.02E-08	Tc00.1047053510435.20	6.27208E-08
Tc00.1047053503987.20	1.09E-08	Tc00.1047053503719.39	6.29067E-08
Tc00.1047053510155.130	1.15E-08	Tc00.1047053504003.60	6.35809E-08
Tc00.1047053506175.20	1.26E-08	Tc00.1047053509065.9	7.63647E-08
Tc00.1047053509663.20	1.26E-08	Tc00.1047053506831.20	8.32411E-08
Tc00.1047053510665.40	1.29E-08	Tc00.1047053503505.14	8.4526E-08
Tc00.1047053507023.70	1.36E-08	Tc00.1047053506319.60	8.4526E-08
Tc00.1047053503729.10	1.42E-08	Tc00.1047053511511.30	8.69664E-08
Tc00.1047053510729.100	1.43E-08	Tc00.1047053511817.134	9.35447E-08
Tc00.1047053503791.30	1.56E-08	Tc00.1047053507601.10	1.02546E-07
Tc00.1047053507715.40	1.59E-08	Tc00.1047053511657.60	1.06412E-07
Tc00.1047053508543.180	2.14E-08	Tc00.1047053506559.20	1.23157E-07
Tc00.1047053508355.70	2.16E-08	Tc00.1047053511465.20	1.45614E-07
Tc00.1047053510121.50	2.71E-08	Tc00.1047053508891.60	1.60329E-07
Tc00.1047053508899.20	2.91E-08	Tc00.1047053510439.20	1.62146E-07
Tc00.1047053504157.120	2.99E-08	Tc00.1047053511649.110	1.67523E-07
Tc00.1047053507087.30	3.77E-08	Tc00.1047053503987.20	1.80513E-07
Tc00.1047053503995.20	3.79E-08	Tc00.1047053511857.40	1.88442E-07
Tc00.1047053511237.20	4.36E-08	Tc00.1047053503879.20	2.07617E-07
Tc00.1047053507609.40	4.68E-08	Tc00.1047053509791.50	2.28031E-07
Tc00.1047053510519.140	4.85E-08	Tc00.1047053504827.21	2.48473E-07
Tc00.1047053509399.150	4.88E-08	Tc00.1047053510885.10	2.7472E-07
Tc00.1047053504175.20	4.97E-08	Tc00.1047053509799.10	2.7472E-07
Tc00.1047053511817.20	5.94E-08	Tc00.1047053510887.14	2.97684E-07
Tc00.1047053504059.20	5.96E-08	Tc00.1047053507771.90	3.29992E-07
Tc00.1047053511649.100	6.32E-08	Tc00.1047053506851.20	3.45698E-07
Tc00.1047053504797.80	6.86E-08	Tc00.1047053503465.10	4.08802E-07
Tc00.1047053503657.70	7.60E-08	Tc00.1047053504105.210	7.33807E-07
Tc00.1047053511263.30	8.97E-08	Tc00.1047053506587.20	7.86429E-07
Tc00.1047053510823.30	9.92E-08	Tc00.1047053506475.105	7.93005E-07
Tc00.1047053509875.180	1.09E-07	Tc00.1047053509747.50	8.22672E-07
Tc00.1047053504069.4	1.09E-07	Tc00.1047053507775.10	8.58852E-07
Tc00.1047053508141.60	1.10E-07	Tc00.1047053506479.80	9.22794E-07

Table S4-4 continued

Tc00.1047053507649.40	1.25E-07	Tc00.1047053511855.80	1.00362E-06
Tc00.1047053508737.50	1.30E-07	Tc00.1047053503703.70	1.19727E-06
Tc00.1047053507765.130	1.47E-07	Tc00.1047053503939.80	1.25028E-06
Tc00.1047053508799.200	1.61E-07	Tc00.1047053510969.10	1.33771E-06
Tc00.1047053510969.10	1.64E-07	Tc00.1047053509437.30	1.34616E-06
Tc00.1047053508153.480	1.87E-07	Tc00.1047053503995.20	1.36228E-06
Tc00.1047053503823.104	2.66E-07	Tc00.1047053508707.160	1.50162E-06
Tc00.1047053509331.160	2.66E-07	Tc00.1047053508641.320	1.61814E-06
Tc00.1047053416511.9	3.01E-07	Tc00.1047053510329.10	1.66664E-06
Tc00.1047053506195.230	3.31E-07	Tc00.1047053508693.70	1.71428E-06
Tc00.1047053503521.80	3.69E-07	Tc00.1047053511521.30	1.82054E-06
Tc00.1047053511661.10	4.24E-07	Tc00.1047053510303.299	1.83948E-06
Tc00.1047053504741.4	5.54E-07	Tc00.1047053507057.40	1.95451E-06
Tc00.1047053508741.340	5.88E-07	Tc00.1047053509437.40	2.12584E-06
Tc00.1047053511655.10	5.88E-07	Tc00.1047053508441.70	2.17133E-06
Tc00.1047053507539.20	6.25E-07	Tc00.1047053506831.10	2.17813E-06
Tc00.1047053509683.40	6.58E-07	Tc00.1047053506315.10	2.47124E-06
Tc00.1047053506511.30	6.90E-07	Tc00.1047053510265.90	2.78129E-06
Tc00.1047053510967.30	6.90E-07	Tc00.1047053509237.130	2.78582E-06
Tc00.1047053506319.60	7.08E-07	Tc00.1047053509733.80	2.95693E-06
Tc00.1047053510657.10	7.38E-07	Tc00.1047053511031.30	2.96404E-06
Tc00.1047053511209.40	8.06E-07	Tc00.1047053509911.10	3.7627E-06
Tc00.1047053511855.20	8.19E-07	Tc00.1047053511505.20	4.033E-06
Tc00.1047053506297.200	8.25E-07	Tc00.1047053503823.104	4.28728E-06
Tc00.1047053508601.90	8.68E-07	Tc00.1047053507601.90	4.40928E-06
Tc00.1047053509791.50	1.03E-06	Tc00.1047053506725.60	5.07955E-06
Tc00.1047053510439.20	1.03E-06	Tc00.1047053503879.119	5.83748E-06
Tc00.1047053507889.20	1.11E-06	Tc00.1047053510667.100	5.90129E-06
Tc00.1047053507023.40	1.25E-06	Tc00.1047053504069.4	6.15114E-06
Tc00.1047053507663.40	1.28E-06	Tc00.1047053511165.80	6.75645E-06
Tc00.1047053507775.10	1.28E-06	Tc00.1047053510131.50	6.95656E-06
Tc00.1047053507769.20	1.28E-06	Tc00.1047053510305.9	7.25395E-06
Tc00.1047053509321.19	1.62E-06	Tc00.1047053504259.19	7.44778E-06
Tc00.1047053506725.80	1.75E-06	Tc00.1047053504113.30	8.27913E-06
Tc00.1047053508507.20	1.84E-06	Tc00.1047053509715.80	8.40912E-06
Tc00.1047053508891.40	1.90E-06	Tc00.1047053508873.10	8.94364E-06
Tc00.1047053508911.70	1.92E-06	Tc00.1047053509207.130	9.7922E-06
Tc00.1047053507623.60	1.98E-06	Tc00.1047053511517.20	9.97309E-06
Tc00.1047053508909.300	2.06E-06	Tc00.1047053507889.20	1.07262E-05
Tc00.1047053508645.40	2.32E-06	Tc00.1047053511153.90	1.09002E-05

Table S4-4 continued

Tc00.1047053508911.60	2.32E-06	Tc00.1047053511277.630	1.0908E-05
Tc00.1047053506425.70	2.71E-06	Tc00.1047053509791.70	1.15519E-05
Tc00.1047053506145.10	2.83E-06	Tc00.1047053511297.10	1.26875E-05
Tc00.1047053509715.140	2.89E-06	Tc00.1047053507017.60	1.28134E-05
Tc00.1047053508857.140	3.36E-06	Tc00.1047053414243.20	1.29687E-05
Tc00.1047053507031.29	3.61E-06	Tc00.1047053511385.100	1.29687E-05
Tc00.1047053510725.70	3.64E-06	Tc00.1047053509059.50	1.29687E-05
Tc00.1047053506831.20	3.66E-06	Tc00.1047053507513.90	1.31909E-05
Tc00.1047053503999.40	4.11E-06	Tc00.1047053506815.20	1.32808E-05
Tc00.1047053510303.299	4.49E-06	Tc00.1047053509793.20	1.41589E-05
Tc00.1047053508153.1100	4.76E-06	Tc00.1047053506989.120	1.44135E-05
Tc00.1047053509399.90	4.78E-06	Tc00.1047053506629.210	1.47092E-05
Tc00.1047053511521.30	4.96E-06	Tc00.1047053507539.20	1.47092E-05
Tc00.1047053507513.80	5.21E-06	Tc00.1047053507103.20	1.49393E-05
Tc00.1047053511511.30	5.28E-06	Tc00.1047053508525.10	1.55347E-05
Tc00.1047053507023.230	5.29E-06	Tc00.1047053508645.50	1.59096E-05
Tc00.1047053511269.20	5.41E-06	Tc00.1047053508965.39	1.73978E-05
Tc00.1047053506327.20	5.55E-06	Tc00.1047053508911.60	1.74276E-05
Tc00.1047053508661.40	5.55E-06	Tc00.1047053506725.40	1.94161E-05
Tc00.1047053503987.30	5.80E-06	Tc00.1047053504105.180	1.95226E-05
Tc00.1047053506175.60	6.11E-06	Tc00.1047053511807.80	2.04274E-05
Tc00.1047053504131.140	6.34E-06	Tc00.1047053509331.120	2.12933E-05
Tc00.1047053503891.90	7.27E-06	Tc00.1047053509429.130	2.16123E-05
Tc00.1047053507801.140	7.50E-06	Tc00.1047053511033.29	2.23137E-05
Tc00.1047053509437.40	7.60E-06	Tc00.1047053511031.10	2.24476E-05
Tc00.1047053506559.20	8.00E-06	Tc00.1047053508625.50	2.42903E-05
Tc00.1047053511293.60	8.46E-06	Tc00.1047053508525.30	2.49201E-05
Tc00.1047053511817.134	8.68E-06	Tc00.1047053507663.40	2.70963E-05
Tc00.1047053511643.50	8.88E-06	Tc00.1047053511521.27	2.75656E-05
Tc00.1047053506009.80	9.64E-06	Tc00.1047053511025.20	2.76625E-05
Tc00.1047053511509.50	1.00E-05	Tc00.1047053506855.260	2.82015E-05
Tc00.1047053503917.14	1.05E-05	Tc00.1047053503781.90	2.95334E-05
Tc00.1047053511859.20	1.05E-05	Tc00.1047053508231.174	2.99156E-05
Tc00.1047053510305.70	1.06E-05	Tc00.1047053510155.30	2.99156E-05
Tc00.1047053506587.50	1.09E-05	Tc00.1047053509901.170	3.01826E-05
Tc00.1047053511385.40	1.16E-05	Tc00.1047053506629.140	3.2589E-05
Tc00.1047053509567.30	1.16E-05	Tc00.1047053507023.40	3.29725E-05
Tc00.1047053509429.160	1.18E-05	Tc00.1047053504235.9	3.79979E-05
Tc00.1047053506989.120	1.43E-05	Tc00.1047053507063.150	3.79979E-05
Tc00.1047053510681.30	1.43E-05	Tc00.1047053507949.40	3.93177E-05

Table S4-4 continued

Tc00.1047053511661.80	1.58E-05	Tc00.1047053510121.50	4.38884E-05
Tc00.1047053508693.70	1.60E-05	Tc00.1047053511261.60	4.5054E-05
Tc00.1047053509237.70	1.60E-05	Tc00.1047053508175.20	4.54031E-05
Tc00.1047053511629.20	1.73E-05	Tc00.1047053511817.20	4.76036E-05
Tc00.1047053506821.160	1.80E-05	Tc00.1047053507823.39	4.97883E-05
Tc00.1047053504949.39	1.85E-05	Tc00.1047053510039.40	4.99937E-05
Tc00.1047053503823.110	1.87E-05	Tc00.1047053509317.50	5.06081E-05
Tc00.1047053506475.105	1.90E-05	Tc00.1047053507601.140	5.22826E-05
Tc00.1047053509053.80	1.99E-05	Tc00.1047053508269.70	5.41786E-05
Tc00.1047053503719.39	2.01E-05	Tc00.1047053511719.20	5.61506E-05
Tc00.1047053506297.270	2.05E-05	Tc00.1047053507517.50	6.45565E-05
Tc00.1047053509057.20	2.29E-05	Tc00.1047053506575.9	6.93894E-05
Tc00.1047053506629.200	2.33E-05	Tc00.1047053506835.60	7.39464E-05
Tc00.1047053509507.49	2.34E-05	Tc00.1047053509005.70	7.63879E-05
Tc00.1047053506871.50	2.64E-05	Tc00.1047053506195.120	7.93061E-05
Tc00.1047053509745.70	2.68E-05	Tc00.1047053508919.140	8.14633E-05
Tc00.1047053509797.60	2.81E-05	Tc00.1047053503647.40	8.17367E-05
Tc00.1047053503431.30	2.89E-05	Tc00.1047053506833.30	8.83879E-05
Tc00.1047053504231.20	2.89E-05	Tc00.1047053509429.200	0.000103357
Tc00.1047053507601.10	2.98E-05	Tc00.1047053508601.130	0.000114371
Tc00.1047053507777.10	3.14E-05	Tc00.1047053509565.9	0.000114371
Tc00.1047053511825.30	3.25E-05	Tc00.1047053508153.570	0.000114371
Tc00.1047053511909.70	3.40E-05	Tc00.1047053510629.450	0.000114371
Tc00.1047053508441.70	3.49E-05	Tc00.1047053472777.30	0.000127663
Tc00.1047053508667.20	4.17E-05	Tc00.1047053510835.20	0.000129645
Tc00.1047053506575.9	4.37E-05	Tc00.1047053510659.219	0.000131402
Tc00.1047053508175.170	4.52E-05	Tc00.1047053508409.150	0.000133435
Tc00.1047053506679.240	4.66E-05	Tc00.1047053504157.40	0.000136122
Tc00.1047053506855.300	4.75E-05	Tc00.1047053511655.10	0.000137446
Tc00.1047053511859.30	4.90E-05	Tc00.1047053511151.50	0.00014
Tc00.1047053509611.10	5.05E-05	Tc00.1047053506303.160	0.000146429
Tc00.1047053503923.30	5.24E-05	Tc00.1047053503431.30	0.000153029
Tc00.1047053509911.110	5.58E-05	Tc00.1047053503703.80	0.000153029
Tc00.1047053508707.160	5.66E-05	Tc00.1047053511261.80	0.000154287
Tc00.1047053506855.120	6.17E-05	Tc00.1047053506201.170	0.000155869
Tc00.1047053510339.20	6.28E-05	Tc00.1047053511671.140	0.000158424
Tc00.1047053510899.19	6.42E-05	Tc00.1047053511661.30	0.00017182
Tc00.1047053506303.140	6.54E-05	Tc00.1047053506821.120	0.000185154
Tc00.1047053506855.260	6.92E-05	Tc00.1047053509719.56	0.000191086
Tc00.1047053509911.10	6.94E-05	Tc00.1047053507609.70	0.000247785

Table S4-4 continued

Tc00.1047053507949.40	7.21E-05	Tc00.1047053508547.130	0.000254905
Tc00.1047053509429.290	7.39E-05	Tc00.1047053506821.160	0.000266504
Tc00.1047053509611.140	8.96E-05	Tc00.1047053503823.10	0.000293156
Tc00.1047053511657.60	9.20E-05	Tc00.1047053511167.60	0.000296328
Tc00.1047053509207.150	9.38E-05	Tc00.1047053506009.40	0.000302506
Tc00.1047053507017.64	9.43E-05	Tc00.1047053510657.160	0.000308173
Tc00.1047053509571.60	9.81E-05	Tc00.1047053507515.140	0.000311749
Tc00.1047053509045.10	0.000105412	Tc00.1047053510903.20	0.000337503
Tc00.1047053508707.270	0.000111187	Tc00.1047053506855.320	0.000341452
Tc00.1047053504827.100	0.000113187	Tc00.1047053510819.20	0.000349153
Tc00.1047053506195.260	0.000113187	Tc00.1047053508177.20	0.000370383
Tc00.1047053506725.60	0.000113187	Tc00.1047053508859.80	0.000371777
Tc00.1047053510579.110	0.00011995	Tc00.1047053504741.60	0.000373765
Tc00.1047053506831.10	0.00012504	Tc00.1047053503919.60	0.000382952
Tc00.1047053509395.100	0.000127325	Tc00.1047053507649.20	0.000406656
Tc00.1047053511017.40	0.000134881	Tc00.1047053504105.170	0.000413925
Tc00.1047053511277.630	0.000135733	Tc00.1047053508881.10	0.000413925
Tc00.1047053503911.30	0.000144732	Tc00.1047053509719.30	0.000422676
Tc00.1047053509065.36	0.000146125	Tc00.1047053509567.30	0.000429916
Tc00.1047053507771.90	0.00014927	Tc00.1047053506629.40	0.000441566
Tc00.1047053508641.320	0.000150001	Tc00.1047053508059.4	0.00044785
Tc00.1047053509715.104	0.00015629	Tc00.1047053505071.100	0.000506124
Tc00.1047053506195.120	0.000158876	Tc00.1047053506681.40	0.00054128
Tc00.1047053508153.1020	0.00016658	Tc00.1047053511301.20	0.00054979
Tc00.1047053504105.200	0.000179814	Tc00.1047053511657.70	0.000588951
Tc00.1047053508355.420	0.000189102	Tc00.1047053508613.20	0.000607889
Tc00.1047053508525.10	0.000193853	Tc00.1047053508409.170	0.000623215
Tc00.1047053506479.80	0.000194126	Tc00.1047053511825.30	0.000643332
Tc00.1047053509005.70	0.000201467	Tc00.1047053511269.40	0.000644716
Tc00.1047053511855.80	0.000202169	Tc00.1047053509161.149	0.000680837
Tc00.1047053511725.40	0.000203659	Tc00.1047053503733.80	0.000685618
Tc00.1047053510885.10	0.000206559	Tc00.1047053506199.30	0.000687569
Tc00.1047053509509.30	0.000232691	Tc00.1047053509509.30	0.000712635
Tc00.1047053511523.30	0.000232691	Tc00.1047053508029.20	0.000718584
Tc00.1047053511151.50	0.000232947	Tc00.1047053507513.40	0.000739421
Tc00.1047053511857.59	0.000233421	Tc00.1047053508027.120	0.000739421
Tc00.1047053503879.20	0.00024267	Tc00.1047053507769.30	0.000743676
Tc00.1047053511903.110	0.000244663	Tc00.1047053506247.470	0.000747967
Tc00.1047053509065.9	0.000250706	Tc00.1047053511863.20	0.000748266
Tc00.1047053510657.160	0.00025476	Tc00.1047053506789.110	0.0007957

Table S4-4 continued

Tc00.1047053507649.20	0.000263222	Tc00.1047053506789.334	0.000951463
Tc00.1047053507071.180	0.000273441	Tc00.1047053508915.9	0.000976347
Tc00.1047053504013.70	0.000280271	Tc00.1047053503721.30	0.000985975
Tc00.1047053510579.10	0.000283295	Tc00.1047053511245.200	0.001019421
Tc00.1047053508501.240	0.000285125	Tc00.1047053509849.10	0.001050012
Tc00.1047053506009.40	0.000290765	Tc00.1047053504021.30	0.001053735
Tc00.1047053509429.200	0.000292651	Tc00.1047053504131.170	0.001071281
Tc00.1047053503703.80	0.000293104	Tc00.1047053508153.860	0.001123638
Tc00.1047053511505.10	0.000293104	Tc00.1047053508799.200	0.001123638
Tc00.1047053511719.20	0.00031905	Tc00.1047053508799.80	0.001185457
Tc00.1047053506835.60	0.000323649	Tc00.1047053509789.30	0.001233325
Tc00.1047053510899.50	0.000326113	Tc00.1047053503797.4	0.001292258
Tc00.1047053508153.550	0.00032859	Tc00.1047053510575.160	0.001319373
Tc00.1047053504105.180	0.000331578	Tc00.1047053508429.20	0.00132694
Tc00.1047053509793.20	0.000332622	Tc00.1047053506241.190	0.00139328
Tc00.1047053508177.20	0.000348102	Tc00.1047053508677.160	0.00139328
Tc00.1047053508231.130	0.000348102	Tc00.1047053505183.110	0.001443316
Tc00.1047053503703.70	0.000357799	Tc00.1047053506475.100	0.001524046
Tc00.1047053503505.14	0.000360516	Tc00.1047053511649.169	0.001550165
Tc00.1047053511649.110	0.000364925	Tc00.1047053507029.41	0.001620506
Tc00.1047053510131.60	0.000377728	Tc00.1047053510723.20	0.001644332
Tc00.1047053503797.4	0.000378466	Tc00.1047053509693.90	0.001681248
Tc00.1047053509733.60	0.00039401	Tc00.1047053503781.40	0.001696129
Tc00.1047053507601.60	0.00039631	Tc00.1047053508319.40	0.001716682
Tc00.1047053510155.160	0.000430955	Tc00.1047053509875.204	0.001755525
Tc00.1047053506629.40	0.000448792	Tc00.1047053508547.20	0.001790204
Tc00.1047053509901.170	0.000448792	Tc00.1047053508349.30	0.001894126
Tc00.1047053503939.100	0.000450343	Tc00.1047053508899.20	0.001894126
Tc00.1047053509207.130	0.000544035	Tc00.1047053506445.90	0.001894126
Tc00.1047053511261.60	0.000544035	Tc00.1047053506855.80	0.001894878
Tc00.1047053505999.70	0.00055582	Tc00.1047053509215.13	0.001977074
Tc00.1047053508059.4	0.000566207	Tc00.1047053508299.50	0.002004044
Tc00.1047053504003.60	0.000583096	Tc00.1047053510731.50	0.002132236
Tc00.1047053507029.41	0.000633786	Tc00.1047053508403.10	0.002154205
Tc00.1047053509051.50	0.000650867	Tc00.1047053508177.64	0.002187121
Tc00.1047053510657.20	0.000654204	Tc00.1047053504797.110	0.00221805
Tc00.1047053510889.190	0.000654204	Tc00.1047053508899.90	0.002253711
Tc00.1047053507017.60	0.000667316	Tc00.1047053508661.40	0.002291691
Tc00.1047053503647.40	0.000685067	Tc00.1047053506337.4	0.002414728
Tc00.1047053506193.20	0.000690572	Tc00.1047053511261.130	0.002470496

Table S4-4 continued

Tc00.1047053509399.180	0.000751146	Tc00.1047053503809.158	0.002565265
Tc00.1047053510155.30	0.000756922	Tc00.1047053511857.59	0.002804403
Tc00.1047053506315.10	0.000781409	Tc00.1047053508909.200	0.002951727
Tc00.1047053509789.30	0.000785976	Tc00.1047053508547.80	0.002984649
Tc00.1047053508817.170	0.000893437	Tc00.1047053504175.30	0.003170549
Tc00.1047053510575.160	0.00100101	Tc00.1047053510889.30	0.003188494
Tc00.1047053507513.40	0.001004861	Tc00.1047053511245.100	0.003213294
Tc00.1047053503617.15	0.001013711	Tc00.1047053503995.30	0.003227049
Tc00.1047053506815.20	0.001056955	Tc00.1047053509045.10	0.003245517
Tc00.1047053503791.20	0.001084197	Tc00.1047053509213.140	0.003252684
Tc00.1047053506247.470	0.001094895	Tc00.1047053510899.50	0.003252684
Tc00.1047053509747.50	0.001095812	Tc00.1047053509399.160	0.003279026
Tc00.1047053511465.20	0.001138283	Tc00.1047053506425.130	0.003550226
Tc00.1047053507615.73	0.001152356	Tc00.1047053508819.20	0.003678698
Tc00.1047053509875.204	0.001159897	Tc00.1047053508909.80	0.003726442
Tc00.1047053511509.40	0.001201107	Tc00.1047053510967.10	0.003934171
Tc00.1047053506855.320	0.001240371	Tc00.1047053503635.50	0.004076669
Tc00.1047053426897.10	0.001250277	Tc00.1047053509617.40	0.004100928
Tc00.1047053511261.130	0.001250782	Tc00.1047053506999.160	0.004112821
Tc00.1047053508153.230	0.001268621	Tc00.1047053511819.50	0.004506551
Tc00.1047053511267.30	0.001389063	Tc00.1047053508199.90	0.004517798
Tc00.1047053511717.174	0.001445335	Tc00.1047053509001.40	0.004569222
Tc00.1047053511661.120	0.001454424	Tc00.1047053509877.10	0.004644258
Tc00.1047053506507.20	0.001471789	Tc00.1047053509717.50	0.004680572
Tc00.1047053510691.95	0.001515418	Tc00.1047053506585.50	0.005007545
Tc00.1047053507007.30	0.001552728	Tc00.1047053511671.100	0.005067052
Tc00.1047053507049.10	0.001579716	Tc00.1047053509057.20	0.005716146
Tc00.1047053508741.250	0.001594991	Tc00.1047053506559.200	0.005870023
Tc00.1047053506479.67	0.001599458	Tc00.1047053503843.30	0.005931113
Tc00.1047053509437.30	0.001599458	Tc00.1047053507515.100	0.005938979
Tc00.1047053506727.50	0.001711295	Tc00.1047053507209.30	0.005955933
Tc00.1047053510903.50	0.001893423	Tc00.1047053503925.60	0.005967737
Tc00.1047053507609.70	0.001941756	Tc00.1047053509877.80	0.00613233
Tc00.1047053507007.9	0.0020188	Tc00.1047053507159.41	0.006260491
Tc00.1047053511287.4	0.002174838	Tc00.1047053506679.270	0.006839654
Tc00.1047053508153.570	0.002223633	Tc00.1047053506957.23	0.006915822
Tc00.1047053511505.20	0.002247579	Tc00.1047053511033.20	0.006969645
Tc00.1047053414243.20	0.002350109	Tc00.1047053508409.300	0.007303574
Tc00.1047053510889.170	0.00240225	Tc00.1047053507049.160	0.007336156
Tc00.1047053508415.10	0.002540574	Tc00.1047053508231.220	0.007715335

Table S4-4 continued

Tc00.1047053506629.210	0.00265476	Tc00.1047053510659.28	0.007766807
Tc00.1047053503721.30	0.002684263	Tc00.1047053510423.50	0.00793889
Tc00.1047053506445.90	0.002684263	Tc00.1047053510519.149	0.007966751
Tc00.1047053508677.160	0.0027942	Tc00.1047053509671.90	0.008042841
Tc00.1047053511863.20	0.002808818	Tc00.1047053509065.36	0.008121068
Tc00.1047053509065.40	0.002815746	Tc00.1047053511647.20	0.008136378
Tc00.1047053508677.90	0.002855598	Tc00.1047053510973.9	0.008287173
Tc00.1047053510183.40	0.00292257	Tc00.1047053511865.40	0.008433615
Tc00.1047053508915.9	0.002943799	Tc00.1047053511297.30	0.008655356
Tc00.1047053511245.100	0.002953097	Tc00.1047053507615.60	0.008831045
Tc00.1047053504021.30	0.002986715	Tc00.1047053511761.80	0.008973546
Tc00.1047053503939.80	0.003062202	Tc00.1047053503635.40	0.009071992
Tc00.1047053508693.190	0.003062202	Tc00.1047053509399.180	0.009122833
Tc00.1047053506621.14	0.003076168	Tc00.1047053503473.20	0.009188611
Tc00.1047053510889.290	0.003089557	Tc00.1047053511817.10	0.009188611
Tc00.1047053507857.30	0.003089642	Tc00.1047053509611.80	0.009232557
Tc00.1047053503733.80	0.003098844	Tc00.1047053509671.10	0.009249214
Tc00.1047053508231.174	0.003100743	Tc00.1047053506629.100	0.01007351
Tc00.1047053508899.90	0.003187873	Tc00.1047053510945.60	0.01058326
Tc00.1047053507023.30	0.003203491	Tc00.1047053511277.130	0.01058326
Tc00.1047053508577.130	0.003268199	Tc00.1047053511277.483	0.01089325
Tc00.1047053508859.30	0.003268199	Tc00.1047053507001.20	0.0112054
Tc00.1047053508909.80	0.003268199	Tc00.1047053510645.20	0.01122729
Tc00.1047053510299.30	0.003268199	Tc00.1047053507317.30	0.01122729
Tc00.1047053510835.20	0.003268199	Tc00.1047053510305.20	0.01157503
Tc00.1047053511025.20	0.003289381	Tc00.1047053507519.100	0.01157681
Tc00.1047053508175.340	0.003329181	Tc00.1047053509029.140	0.01157681
Tc00.1047053509429.170	0.003388617	Tc00.1047053506337.40	0.01158269
Tc00.1047053508881.10	0.003483287	Tc00.1047053508643.40	0.01191061
Tc00.1047053509849.10	0.00372907	Tc00.1047053506145.30	0.01223645
Tc00.1047053511859.60	0.003876952	Tc00.1047053511163.50	0.01235445
Tc00.1047053506577.130	0.004032	Tc00.1047053503565.20	0.01244967
Tc00.1047053508175.70	0.004244159	Tc00.1047053506679.10	0.01313602
Tc00.1047053507023.170	0.004315139	Tc00.1047053506195.200	0.01319704
Tc00.1047053503823.10	0.004435117	Tc00.1047053506419.20	0.01348185
Tc00.1047053508891.20	0.004435117	Tc00.1047053507615.73	0.01408203
Tc00.1047053509237.130	0.004435117	Tc00.1047053510657.20	0.0145385
Tc00.1047053503995.30	0.004999679	Tc00.1047053511755.119	0.0147019
Tc00.1047053511657.50	0.004999679	Tc00.1047053509399.120	0.01484567
Tc00.1047053509213.140	0.005045535	Tc00.1047053509179.100	0.01490929

Table S4-4 continued

Tc00.1047053511649.140	0.005068858	Tc00.1047053509569.120	0.01496905
Tc00.1047053508153.400	0.005215436	Tc00.1047053506475.130	0.01504905
Tc00.1047053506559.180	0.005225967	Tc00.1047053510729.130	0.01629092
Tc00.1047053509601.110	0.005316634	Tc00.1047053506009.90	0.01654873
Tc00.1047053511755.119	0.005388372	Tc00.1047053504191.10	0.01654873
Tc00.1047053506175.50	0.005565471	Tc00.1047053506475.110	0.01659624
Tc00.1047053509215.13	0.005898388	Tc00.1047053506509.60	0.0176189
Tc00.1047053504741.60	0.006046156	Tc00.1047053507765.80	0.01781196
Tc00.1047053511033.20	0.006046156	Tc00.1047053506679.280	0.01807827
Tc00.1047053509001.30	0.006249237	Tc00.1047053509437.70	0.02032092
Tc00.1047053505183.110	0.006359511	Tc00.1047053509065.40	0.02062371
Tc00.1047053509059.50	0.006437967	Tc00.1047053509105.110	0.02106138
Tc00.1047053504185.9	0.006561277	Tc00.1047053510121.150	0.02150974
Tc00.1047053509065.30	0.006561277	Tc00.1047053504013.70	0.02184042
Tc00.1047053508027.120	0.006568702	Tc00.1047053506241.60	0.02191077
Tc00.1047053508199.90	0.006604081	Tc00.1047053505939.60	0.02344555
Tc00.1047053508547.130	0.006647422	Tc00.1047053506175.50	0.02366682
Tc00.1047053511167.60	0.006714048	Tc00.1047053510301.80	0.02366682
Tc00.1047053509671.90	0.006921485	Tc00.1047053509569.30	0.02366682
Tc00.1047053506855.60	0.006929833	Tc00.1047053508577.130	0.02418661
Tc00.1047053511033.29	0.006977205	Tc00.1047053511257.110	0.02418661
Tc00.1047053503773.20	0.007341735	Tc00.1047053510423.40	0.02420549
Tc00.1047053504105.170	0.007341735	Tc00.1047053510355.250	0.02606999
Tc00.1047053509399.130	0.00737667	Tc00.1047053507771.20	0.02748179
Tc00.1047053508409.150	0.007724029	Tc00.1047053508543.160	0.0275594
Tc00.1047053510153.10	0.007734502	Tc00.1047053509507.30	0.02782154
Tc00.1047053506921.10	0.007797195	Tc00.1047053507615.130	0.02806269
Tc00.1047053511277.115	0.007903559	Tc00.1047053507007.50	0.02817771
Tc00.1047053509569.170	0.008142324	Tc00.1047053503559.90	0.02878444
Tc00.1047053506789.110	0.008150297	Tc00.1047053504105.190	0.02926731
Tc00.1047053506583.50	0.008472971	Tc00.1047053503431.100	0.03067565
Tc00.1047053508909.200	0.008503487	Tc00.1047053503773.20	0.03192255
Tc00.1047053509567.60	0.008503487	Tc00.1047053511511.60	0.03299258
Tc00.1047053509877.10	0.008564555	Tc00.1047053508851.59	0.03372719
Tc00.1047053511671.100	0.008885348	Tc00.1047053507623.50	0.03428118
Tc00.1047053506195.90	0.008956436	Tc00.1047053511027.20	0.03505868
Tc00.1047053503559.90	0.009296117	Tc00.1047053507589.30	0.03584063
Tc00.1047053509001.40	0.009370222	Tc00.1047053507887.30	0.03663114
Tc00.1047053503635.50	0.009419079	Tc00.1047053511277.160	0.03680298
Tc00.1047053508177.64	0.009710869	Tc00.1047053503891.60	0.03693704

Table S4-4 continued

Tc00.1047053510571.10	0.009818706	Tc00.1047053504069.60	0.03758908
Tc00.1047053510645.20	0.00995744	Tc00.1047053506789.150	0.03798008
Tc00.1047053506855.80	0.01001235	Tc00.1047053509429.170	0.03823902
Tc00.1047053510039.40	0.01007409	Tc00.1047053510221.20	0.03903417
Tc00.1047053508859.80	0.01020296	Tc00.1047053507717.20	0.03949198
Tc00.1047053507615.60	0.01041043	Tc00.1047053506435.20	0.03958613
Tc00.1047053508859.100	0.01044253	Tc00.1047053511763.19	0.04109546
Tc00.1047053505939.60	0.01047821	Tc00.1047053511287.130	0.04147702
Tc00.1047053511165.20	0.01070506	Tc00.1047053506789.249	0.04377601
Tc00.1047053509693.90	0.01078055	Tc00.1047053509429.240	0.04377997
Tc00.1047053503823.90	0.01120002	Tc00.1047053509237.120	0.04515393
Tc00.1047053508873.10	0.01143388	Tc00.1047053506679.240	0.04517768
Tc00.1047053506337.4	0.01155315	Tc00.1047053503923.20	0.04620717
Tc00.1047053509161.149	0.01160726	Tc00.1047053508707.70	0.04720796
Tc00.1047053508355.124	0.01188393	Tc00.1047053509875.250	0.04724392
Tc00.1047053510723.20	0.01215938	Tc00.1047053503939.100	0.04760855
Tc00.1047053506475.30	0.01230344	Tc00.1047053509331.50	0.04795216
Tc00.1047053504131.80	0.01231161	Tc00.1047053474937.9	0.04910799
Tc00.1047053505071.100	0.0125362	Tc00.1047053510265.14	0.04910799
Tc00.1047053509683.60	0.01295738	Tc00.1047053507025.50	0.0494249
Tc00.1047053509287.110	0.01318319	Tc00.1047053507895.170	0.04942634
Tc00.1047053508711.10	0.01334234	Tc00.1047053509039.10	0.04946674
Tc00.1047053507221.20	0.01412344	Tc00.1047053506195.90	0.04978982
Tc00.1047053507317.20	0.01508661	Tc00.1047053504157.120	0.04979235
Tc00.1047053511807.50	0.01508661	Tc00.1047053507023.70	0.05041877
Tc00.1047053507589.30	0.01520051	Tc00.1047053510435.40	0.05047607
Tc00.1047053506475.110	0.01539594	Tc00.1047053506435.134	0.05250352
Tc00.1047053410199.4	0.01585204	Tc00.1047053410199.4	0.05596918
Tc00.1047053504105.210	0.01585204	Tc00.1047053506587.50	0.05598545
Tc00.1047053508231.220	0.01675376	Tc00.1047053511165.20	0.05632162
Tc00.1047053506389.70	0.01689774	Tc00.1047053503809.130	0.05771627
Tc00.1047053511807.80	0.01715801	Tc00.1047053508741.380	0.05847718
Tc00.1047053508461.70	0.01797993	Tc00.1047053504021.149	0.05932417
Tc00.1047053505171.70	0.01800008	Tc00.1047053506835.99	0.06086205
Tc00.1047053510355.250	0.01833895	Tc00.1047053511167.100	0.06086205
Tc00.1047053510729.130	0.01833895	Tc00.1047053510041.30	0.06286061
Tc00.1047053508741.130	0.01886924	Tc00.1047053506459.270	0.06341835
Tc00.1047053472777.30	0.0198754	Tc00.1047053508297.41	0.06354871
Tc00.1047053507001.20	0.0198754	Tc00.1047053505163.60	0.06411811
Tc00.1047053508501.270	0.0198754	Tc00.1047053506629.170	0.06498833

Table S4-4 continued

Tc00.1047053506999.160	0.02027333	Tc00.1047053509713.30	0.06552576
Tc00.1047053511027.20	0.02036355	Tc00.1047053505999.24	0.06583559
Tc00.1047053509901.150	0.02038028	Tc00.1047053507951.150	0.06927539
Tc00.1047053510661.40	0.02038028	Tc00.1047053509331.20	0.06993677
Tc00.1047053508547.20	0.02064999	Tc00.1047053509617.70	0.07026944
Tc00.1047053511017.60	0.02117681	Tc00.1047053509857.50	0.07340206
Tc00.1047053504209.10	0.0227297	Tc00.1047053509399.100	0.07375317
Tc00.1047053510977.9	0.02344562	Tc00.1047053510359.290	0.07471444
Tc00.1047053503939.120	0.023697	Tc00.1047053503593.70	0.07796791
Tc00.1047053508027.70	0.023697	Tc00.1047053503733.40	0.07885029
Tc00.1047053509693.120	0.023697	Tc00.1047053507823.10	0.0812446
Tc00.1047053510305.20	0.023697	Tc00.1047053511237.50	0.08351791
Tc00.1047053508795.10	0.02422554	Tc00.1047053506841.20	0.08355336
Tc00.1047053503809.75	0.02462931	Tc00.1047053507027.59	0.08396327
Tc00.1047053504575.10	0.02462931	Tc00.1047053511209.50	0.08514605
Tc00.1047053505169.21	0.0253175	Tc00.1047053509337.15	0.08661605
Tc00.1047053506241.190	0.0253175	Tc00.1047053510039.109	0.08810398
Tc00.1047053509719.56	0.0256187	Tc00.1047053460125.10	0.0890126
Tc00.1047053507519.154	0.02618708	Tc00.1047053511151.11	0.09055346
Tc00.1047053509337.19	0.02620511	Tc00.1047053507769.20	0.09291382
Tc00.1047053509875.250	0.02620511	Tc00.1047053504827.150	0.09417634
Tc00.1047053507769.30	0.02738268	Tc00.1047053509001.20	0.09618409
Tc00.1047053508999.80	0.02738571	Tc00.1047053504231.10	0.09669768
Tc00.1047053508059.20	0.02759818	Tc00.1047053506913.10	0.09745495
Tc00.1047053511847.20	0.02759818	Tc00.1047053511289.100	0.1015052
Tc00.1047053506559.200	0.02838149	Tc00.1047053506509.30	0.1031098
Tc00.1047053507209.60	0.02838149	Tc00.1047053504243.49	0.103134
Tc00.1047053506585.50	0.02868049	Tc00.1047053510941.3	0.1039216
Tc00.1047053507021.110	0.0290084	Tc00.1047053503939.130	0.1044048
Tc00.1047053509029.90	0.02932921	Tc00.1047053508741.370	0.106864
Tc00.1047053508299.70	0.02975448	Tc00.1047053504575.10	0.1094027
Tc00.1047053506419.20	0.02995366	Tc00.1047053507241.20	0.109592
Tc00.1047053503891.60	0.03028013	Tc00.1047053508501.270	0.1120602
Tc00.1047053508409.300	0.03050623	Tc00.1047053505183.10	0.1123785
Tc00.1047053511633.79	0.03207283	Tc00.1047053503703.60	0.1125382
Tc00.1047053511649.39	0.03374451	Tc00.1047053511859.60	0.1137298
Tc00.1047053509237.120	0.03384439	Tc00.1047053510183.40	0.1137298
Tc00.1047053511821.40	0.03399346	Tc00.1047053509857.10	0.1155231
Tc00.1047053511671.140	0.0341915	Tc00.1047053510731.120	0.1158845
Tc00.1047053507825.40	0.03533772	Tc00.1047053509017.40	0.1167599

Table S4-4 continued

Tc00.1047053507765.80	0.03542569	Tc00.1047053503923.30	0.1167599
Tc00.1047053508707.60	0.03564615	Tc00.1047053506581.60	0.119882
Tc00.1047053508643.40	0.0356711	Tc00.1047053509437.50	0.1205076
Tc00.1047053511293.90	0.0356711	Tc00.1047053508999.260	0.12145
Tc00.1047053511287.130	0.03606332	Tc00.1047053510329.260	0.1259141
Tc00.1047053508153.860	0.03913473	Tc00.1047053508741.360	0.1259141
Tc00.1047053507913.39	0.0397337	Tc00.1047053507849.60	0.1261091
Tc00.1047053507771.20	0.040683	Tc00.1047053508741.130	0.1273924
Tc00.1047053510667.100	0.04107147	Tc00.1047053510339.80	0.1284085
Tc00.1047053507517.40	0.04127154	Tc00.1047053506281.134	0.1288334
Tc00.1047053511277.483	0.04142607	Tc00.1047053509999.30	0.1343226
Tc00.1047053507057.4	0.04167505	Tc00.1047053504131.100	0.1353382
Tc00.1047053510659.28	0.04170062	Tc00.1047053508699.60	0.1380369
Tc00.1047053508657.10	0.04318012	Tc00.1047053510299.50	0.1432573
Tc00.1047053503431.100	0.04379145	Tc00.1047053506999.20	0.1432573
Tc00.1047053503565.20	0.04402461	Tc00.1047053505163.30	0.1447841
Tc00.1047053506195.200	0.04636995	Tc00.1047053503993.10	0.1477606
Tc00.1047053509179.100	0.04669909	Tc00.1047053504741.240	0.1483497
Tc00.1047053509207.40	0.04703493	Tc00.1047053507623.101	0.1509826
Tc00.1047053405737.14	0.04806731	Tc00.1047053506975.69	0.1509826
Tc00.1047053507623.50	0.04875565	Tc00.1047053511459.50	0.1509826
Tc00.1047053506679.280	0.04904206	Tc00.1047053511817.280	0.1530923
Tc00.1047053511865.40	0.04969418	Tc00.1047053507851.30	0.1532532
Tc00.1047053507025.50	0.05001846	Tc00.1047053510665.20	0.1542121
Tc00.1047053511823.14	0.05274116	Tc00.1047053510299.40	0.1544279
Tc00.1047053508543.160	0.05315992	Tc00.1047053508175.90	0.1549799
Tc00.1047053509105.110	0.05315992	Tc00.1047053511483.50	0.1559294
Tc00.1047053504131.100	0.05330773	Tc00.1047053446067.9	0.1565659
Tc00.1047053510301.80	0.0535402	Tc00.1047053511807.50	0.1575984
Tc00.1047053504253.30	0.05390762	Tc00.1047053506459.250	0.1651151
Tc00.1047053509999.140	0.05402959	Tc00.1047053508711.10	0.1651151
Tc00.1047053506679.270	0.05422053	Tc00.1047053508059.24	0.1679797
Tc00.1047053509507.30	0.05907921	Tc00.1047053509719.26	0.1679797
Tc00.1047053507317.30	0.06017827	Tc00.1047053506241.100	0.1732479
Tc00.1047053511763.19	0.06017827	Tc00.1047053503537.18	0.1735695
Tc00.1047053511321.40	0.06062037	Tc00.1047053506855.310	0.1741105
Tc00.1047053506629.170	0.06076395	Tc00.1047053507087.60	0.1783076
Tc00.1047053510945.60	0.06076395	Tc00.1047053511857.20	0.1784276
Tc00.1047053507159.41	0.06413181	Tc00.1047053508859.100	0.1850054
Tc00.1047053508303.44	0.06424098	Tc00.1047053506905.20	0.185308

Table S4-4 continued

Tc00.1047053505163.60	0.06465183	Tc00.1047053511463.20	0.1899251
Tc00.1047053510303.160	0.06491512	Tc00.1047053504257.60	0.190451
Tc00.1047053431849.20	0.06527971	Tc00.1047053508507.24	0.1915328
Tc00.1047053511247.18	0.065559	Tc00.1047053511463.10	0.1924368
Tc00.1047053506725.40	0.06643045	Tc00.1047053503625.20	0.1941898
Tc00.1047053508919.140	0.06666374	Tc00.1047053509901.100	0.1945271
Tc00.1047053507849.70	0.0666895	Tc00.1047053510155.115	0.1945271
Tc00.1047053508455.30	0.06678393	Tc00.1047053509875.240	0.1988977
Tc00.1047053506459.250	0.06787142	Tc00.1047053508859.40	0.1988977
Tc00.1047053511517.20	0.06872054	Tc00.1047053508349.10	0.2011441
Tc00.1047053504259.19	0.06875971	Tc00.1047053510659.130	0.2017596
Tc00.1047053503809.149	0.07021418	Tc00.1047053511649.140	0.2029471
Tc00.1047053506145.30	0.07049326	Tc00.1047053509487.20	0.2136507
Tc00.1047053506815.40	0.07049326	Tc00.1047053506303.70	0.2143574
Tc00.1047053507023.20	0.07049326	Tc00.1047053511511.150	0.2162628
Tc00.1047053504081.270	0.07084879	Tc00.1047053509459.70	0.2180139
Tc00.1047053503733.40	0.07117062	Tc00.1047053509049.20	0.2199811
Tc00.1047053507513.90	0.07134797	Tc00.1047053506177.50	0.2210912
Tc00.1047053511903.60	0.07164571	Tc00.1047053504209.10	0.2242386
Tc00.1047053507951.150	0.07196308	Tc00.1047053507801.179	0.2253973
Tc00.1047053511277.20	0.07214432	Tc00.1047053510131.60	0.2277898
Tc00.1047053508895.20	0.07360822	Tc00.1047053504157.80	0.2278661
Tc00.1047053503719.30	0.07431747	Tc00.1047053506247.30	0.2294245
Tc00.1047053505999.170	0.07431747	Tc00.1047053511289.80	0.2299474
Tc00.1047053509791.170	0.07564305	Tc00.1047053511815.170	0.2319835
Tc00.1047053511261.80	0.07736229	Tc00.1047053509029.70	0.2333037
Tc00.1047053510329.260	0.0775317	Tc00.1047053511671.50	0.2336754
Tc00.1047053507823.10	0.07830839	Tc00.1047053509399.140	0.2336754
Tc00.1047053506679.10	0.07864824	Tc00.1047053509733.50	0.2349917
Tc00.1047053511165.80	0.0791687	Tc00.1047053506297.80	0.2366305
Tc00.1047053507087.80	0.07941238	Tc00.1047053506195.210	0.2369958
Tc00.1047053511817.100	0.08086071	Tc00.1047053510303.70	0.2369958
Tc00.1047053504243.49	0.08177265	Tc00.1047053506991.19	0.239144
Tc00.1047053509039.10	0.08177265	Tc00.1047053507765.110	0.2425638
Tc00.1047053505939.40	0.08246427	Tc00.1047053508153.920	0.2514276
Tc00.1047053509163.60	0.08246427	Tc00.1047053511903.110	0.2545431
Tc00.1047053504131.170	0.08409944	Tc00.1047053510689.60	0.2577532
Tc00.1047053503923.20	0.08442142	Tc00.1047053511245.140	0.2577532
Tc00.1047053511671.50	0.08449926	Tc00.1047053506789.220	0.2585882
Tc00.1047053507895.170	0.08469826	Tc00.1047053511515.9	0.2631007

Table S4-4 continued

Tc00.1047053511167.100	0.08469826	Tc00.1047053510655.120	0.2658859
Tc00.1047053503703.90	0.08493921	Tc00.1047053504259.10	0.2697561
Tc00.1047053506835.99	0.08753943	Tc00.1047053505171.70	0.2792578
Tc00.1047053511483.50	0.08753943	Tc00.1047053509461.20	0.28015
Tc00.1047053507887.30	0.08980064	Tc00.1047053504003.50	0.2804446
Tc00.1047053510131.20	0.09078102	Tc00.1047053503559.80	0.2804446
Tc00.1047053509505.80	0.09340124	Tc00.1047053508737.70	0.2804446
Tc00.1047053511209.50	0.09958649	Tc00.1047053506201.84	0.2804446
Tc00.1047053503891.80	0.09999567	Tc00.1047053511151.60	0.2844638
Tc00.1047053508707.110	0.09999567	Tc00.1047053506583.50	0.2852452
Tc00.1047053507515.140	0.1014392	Tc00.1047053504003.40	0.2898956
Tc00.1047053506913.10	0.1014705	Tc00.1047053503413.4	0.2932715
Tc00.1047053510653.10	0.1027136	Tc00.1047053509991.100	0.2965596
Tc00.1047053510121.150	0.103349	Tc00.1047053509571.60	0.300919
Tc00.1047053511151.90	0.1046398	Tc00.1047053505939.40	0.300919
Tc00.1047053506839.10	0.1047412	Tc00.1047053507645.70	0.3018978
Tc00.1047053506435.134	0.105528	Tc00.1047053503939.70	0.3041738
Tc00.1047053510819.20	0.1074219	Tc00.1047053510725.50	0.3142863
Tc00.1047053507717.20	0.1076631	Tc00.1047053503999.20	0.3159607
Tc00.1047053510359.290	0.1076631	Tc00.1047053509941.140	0.3230579
Tc00.1047053509459.50	0.1098214	Tc00.1047053511153.60	0.3236205
Tc00.1047053508693.150	0.1124999	Tc00.1047053511903.190	0.3278555
Tc00.1047053507209.40	0.1128052	Tc00.1047053437575.18	0.3327837
Tc00.1047053511507.60	0.1130109	Tc00.1047053511517.99	0.3335313
Tc00.1047053503465.10	0.1153079	Tc00.1047053511837.50	0.3378755
Tc00.1047053511459.50	0.1161065	Tc00.1047053507031.80	0.3378755
Tc00.1047053510329.220	0.116265	Tc00.1047053506195.140	0.3395552
Tc00.1047053508533.30	0.1171494	Tc00.1047053511277.140	0.3407236
Tc00.1047053511237.50	0.1171591	Tc00.1047053506593.50	0.343493
Tc00.1047053506475.40	0.1172726	Tc00.1047053509065.170	0.3454999
Tc00.1047053507991.120	0.1177264	Tc00.1047053511323.60	0.3474318
Tc00.1047053506509.60	0.1195954	Tc00.1047053511269.60	0.3500264
Tc00.1047053509029.80	0.1202992	Tc00.1047053511519.9	0.3505102
Tc00.1047053510731.50	0.1235911	Tc00.1047053511291.70	0.3514007
Tc00.1047053506009.90	0.1250893	Tc00.1047053507801.160	0.3514007
Tc00.1047053509253.40	0.1250893	Tc00.1047053507991.120	0.3532359
Tc00.1047053509683.10	0.1250893	Tc00.1047053503965.30	0.3539283
Tc00.1047053506581.60	0.1282344	Tc00.1047053507049.190	0.3580622
Tc00.1047053511817.10	0.1316439	Tc00.1047053510655.110	0.3711734
Tc00.1047053510339.80	0.1325705	Tc00.1047053510661.100	0.3715875

Table S4-4 continued

Tc00.1047053503781.40	0.1326968	Tc00.1047053510131.90	0.372436
Tc00.1047053511153.100	0.1331947	Tc00.1047053508409.240	0.3748401
Tc00.1047053504157.80	0.1366288	Tc00.1047053506491.20	0.3882509
Tc00.1047053508547.110	0.1382533		
Tc00.1047053508741.390	0.1423538		
Tc00.1047053505999.24	0.1428451		
Tc00.1047053510221.20	0.1432452		
Tc00.1047053506819.10	0.1445713		
Tc00.1047053510799.9	0.1445713		
Tc00.1047053511291.50	0.1445713		
Tc00.1047053507019.10	0.1452956		
Tc00.1047053506999.20	0.1472268		
Tc00.1047053509331.20	0.1499112		
Tc00.1047053510731.120	0.150047		
Tc00.1047053506851.20	0.1519171		
Tc00.1047053503539.20	0.1526292		
Tc00.1047053504003.50	0.1543761		
Tc00.1047053504231.10	0.1543761		
Tc00.1047053507031.130	0.1543761		
Tc00.1047053508739.90	0.1543761		
Tc00.1047053510665.20	0.1545736		
Tc00.1047053508891.60	0.1572735		
Tc00.1047053509569.120	0.1594665		
Tc00.1047053505163.30	0.164383		
Tc00.1047053503929.40	0.167022		
Tc00.1047053509999.30	0.1708057		
Tc00.1047053508645.50	0.1724923		
Tc00.1047053474937.9	0.1787948		
Tc00.1047053509561.9	0.1810009		
Tc00.1047053510889.231	0.1818489		
Tc00.1047053506297.130	0.182152		
Tc00.1047053511291.40	0.1838418		
Tc00.1047053507049.160	0.1851554		
Tc00.1047053507031.150	0.1856923		
Tc00.1047053509399.160	0.1863648		
Tc00.1047053508661.70	0.1881015		
Tc00.1047053504105.190	0.1896966		
Tc00.1047053508819.20	0.1896966		
Tc00.1047053507851.30	0.1900979		
Tc00.1047053510357.130	0.1909141		

Table S4-4 continued

Tc00.1047053509047.50	0.1916575		
Tc00.1047053508741.370	0.1956628		
Tc00.1047053508699.60	0.1986586		
Tc00.1047053504191.10	0.1987174		
Tc00.1047053511245.140	0.2026829		
Tc00.1047053506999.180	0.205048		
Tc00.1047053510903.70	0.205048		
Tc00.1047053507715.90	0.2082088		
Tc00.1047053508349.10	0.2085529		
Tc00.1047053511153.124	0.2101601		
Tc00.1047053507023.110	0.2135353		
Tc00.1047053503559.80	0.2151116		
Tc00.1047053511281.60	0.2166901		
Tc00.1047053509055.60	0.216846		
Tc00.1047053507087.60	0.2175421		
Tc00.1047053506789.310	0.217586		
Tc00.1047053511647.20	0.2183864		
Tc00.1047053506789.220	0.2188664		
Tc00.1047053510329.320	0.2194336		
Tc00.1047053508263.30	0.2200926		
Tc00.1047053510943.70	0.2328339		
Tc00.1047053506489.59	0.2331157		
Tc00.1047053506009.100	0.2353106		
Tc00.1047053509621.10	0.2363086		
Tc00.1047053509799.10	0.2363086		
Tc00.1047053511517.99	0.2363086		
Tc00.1047053503843.30	0.2388683		
Tc00.1047053507801.179	0.2388683		
Tc00.1047053506415.10	0.239857		
Tc00.1047053506275.60	0.2399179		
Tc00.1047053506479.110	0.2412609		
Tc00.1047053511807.40	0.2448354		
Tc00.1047053509713.30	0.2450408		
Tc00.1047053506447.19	0.2453732		
Tc00.1047053507017.130	0.2459421		
Tc00.1047053509487.20	0.2459421		
Tc00.1047053509733.50	0.2459421		
Tc00.1047053505183.10	0.2460266		
Tc00.1047053509063.10	0.250384		
Tc00.1047053509857.10	0.2504053		

Table S4-4 continued

Tc00.1047053511165.50	0.2504053		
Tc00.1047053509875.240	0.2507121		
Tc00.1047053510689.60	0.2522087		
Tc00.1047053506175.10	0.2525569		
Tc00.1047053506727.70	0.2592953		
Tc00.1047053508029.20	0.2592953		
Tc00.1047053506435.120	0.2596776		
Tc00.1047053511263.50	0.261079		
Tc00.1047053503809.130	0.2644189		
Tc00.1047053509109.114	0.2644189		
Tc00.1047053506435.210	0.2659219		
Tc00.1047053508961.10	0.2677824		
Tc00.1047053511819.50	0.2677824		
Tc00.1047053508859.40	0.2698404		
Tc00.1047053507849.60	0.2707383		
Tc00.1047053511463.10	0.2707383		
Tc00.1047053506561.10	0.2737619		
Tc00.1047053506435.280	0.2755057		
Tc00.1047053510661.100	0.2789215		
Tc00.1047053508851.70	0.2808799		
Tc00.1047053509053.120	0.2809384		
Tc00.1047053509331.120	0.2827275		
Tc00.1047053511021.40	0.2862378		
Tc00.1047053437575.18	0.2868516		
Tc00.1047053509565.9	0.2868516		
Tc00.1047053511287.80	0.2948189		
Tc00.1047053511727.129	0.2956532		
Tc00.1047053506575.60	0.2959326		
Tc00.1047053508741.380	0.2979361		
Tc00.1047053460125.10	0.2992174		
Tc00.1047053508153.1110	0.2999716		
Tc00.1047053504003.40	0.302089		
Tc00.1047053506989.30	0.302089		
Tc00.1047053507049.190	0.302089		
Tc00.1047053508669.10	0.302089		
Tc00.1047053508699.50	0.302089		
Tc00.1047053509001.20	0.302089		
Tc00.1047053509461.20	0.302089		
Tc00.1047053510131.90	0.3102782		
Tc00.1047053508707.70	0.313182		

Table S4-4 continued

Tc00.1047053511289.80	0.3143557		
Tc00.1047053508999.260	0.3175551		

Table S4-5. Nucleotide divergences among *T. cruzi* strains (haplotypes). Upper right values: average Nei & Gojobori dS values based on alignments of 4,921 orthologous genes. Left lower values: Uncorrected nucleotide divergence (p distance) based on 1,750,000 aligned nucleotides.

	Esmeraldo	**Non-Esmeraldo**	**Sylvio**	**JRcl4**
Esmeraldo	-	0.0477	0.073	0.072
Non-Esmeraldo	0.022	-	0.057	0.060
Sylvio	0.023	0.018	-	0.008
JRcl4	0.022	0.018	0.004	-

Table S4-6. Nucleotide divergences among *Leishmania spp.* species. Upper right values: average Nei & Gojobori dS values based on alignments of 7,439 orthologous genes. Left lower values: Uncorrected nucleotide divergence (p distance) based on 1,750,000 aligned nucleotides.

	L. major	***L. infantum***	***L. mexicana***	***L. braziliensis***
L. major	-	0.113	0.172	0.444
L. infantum	0.053	-	0.162	0.436
L. mexicana	0.078	0.073	-	0.448
L. braziliensis	0.174	0.171	0.175	-

Table S4-7. The number of proteins under positive selection in *T. cruzi* and *Leishmania* spp. based on analyses of 3 taxa. *T. cruzi* taxa set consists of Non-Esmeraldo and Esmeraldo haplotypes and Sylvio strain. The *Leishmania spp.* taxa set consists of *L. mexicana, L. major* and *L. infantum*. The conserved section represents the number of proteins with no evidence of positive selection by the LRT of models M7 vs M8 and M8 vs M8a in PAML with $p<0.05$ and $p<0.01$. Note: * False discovery rate corrected q-values.

Taxonomic group	***T. cruzi***		***Leishmania spp.***	
Models compared	Conserved (%)	Positive selection (%)	Conserved (%)	Positive selection (%)
M8 vs M8a ($p<0.05$)	4473 (86.9%) *4658 (90.51%)	673 (13.1%) *488 (9.48%)	6806 (91.5%) *7296 (98.07%)	633 (8.5%) *143 (1.92%)
M8 vs M8a ($p<0.01$)	4612 (89.6%) *4741 (92.12%)	534 (10.4%) *405 (7.87%)	7119 (95.7%) *7344 (98.72%)	320 (4.3%) *95 (1.27%)
M7 vs M8 ($p<0.05$)	4466 (86.8%) *4645 (90.26%)	680 (13.2%) *501 (9.73%)	6738 (90.6%) *7275 (97.79%)	701 (9.4%) *164 (2.2%)
M7 vs M8 ($p<0.01$)	4605 (89.5%) *4732 (91.95%)	541 (10.5%) *414 (8.04%)	7081 (95.2%) *7332 (98.56%)	358 (4.8%) *107 (1.43%)

Table S4-8. The number of putative false positives. False Positives (FP): proteins predicted to be under positive selection based on the 4 taxa analysis but not in the 3 taxa analysis ($p < 0.05$). True Positives (TP): proteins predicted to be under positive selection in both 3 and 4 taxa analyses ($p < 0.05$).

Taxonomic group	***T. cruzi***			***Leishmania spp.***		
Models compared	Conserved (%)	Positive selection		Conserved (%)	Positive selection	
		False positives (%)	True positives (%)		False positives (%)	True positives (%)
M8 vs M8a	4369 (84.9%)	310 (6.0%)	467 (9.1%)	7218 (97.0%)	140 (1.9%)	81 (1.1%)
M7 vs M8	4347 (84.5%)	321 (6.2%)	478 (9.3%)	7030 (94.5%)	285 (3.8%)	124 (1.7%)

Table S4-9. Functional over representation of True Positives. N: number of proteins of this function in ortholog data set. n: number of proteins of this function under positive selection under model M8 vs M8a. p: statistical significance estimated from GeneMerge. Gene codes are the gene codes for Non-Esmeraldo (*T. cruzi)* and *Leishmania major* found in Tritrypdb.org as of March 2012.

Trypanosoma cruzi				
Predicted function	N	n	p	Gene codes
Hypothetical	3496	307	0.02524	Not shown
GTP binding	7	2	0.1008	Tc00.1047053504105.210, Tc00.1047053507715.40,
Mucin	15	3	0.1218	Tc00.1047053506615.50, Tc00.1047053506815.20, Tc00.1047053508873.10,
Leishmania spp.				
Predicted function	N	n	p	Gene codes
Iron superoxide dismutase	5	2	0.00057	LmjF.32.1820, LmjF.32.1830
Cysteine peptidase	7	2	0.001204	LmjF.29.0820, LmjF.19.1420
ATP-binding cassette protein subfamily A, D & G	34	3	0.002221	LmjF.11.1290, LmjF.27.0970, LmjF.06.0090
Amastin like	51	2	0.05884	LmjF.08.0850, LmjF.34.0970

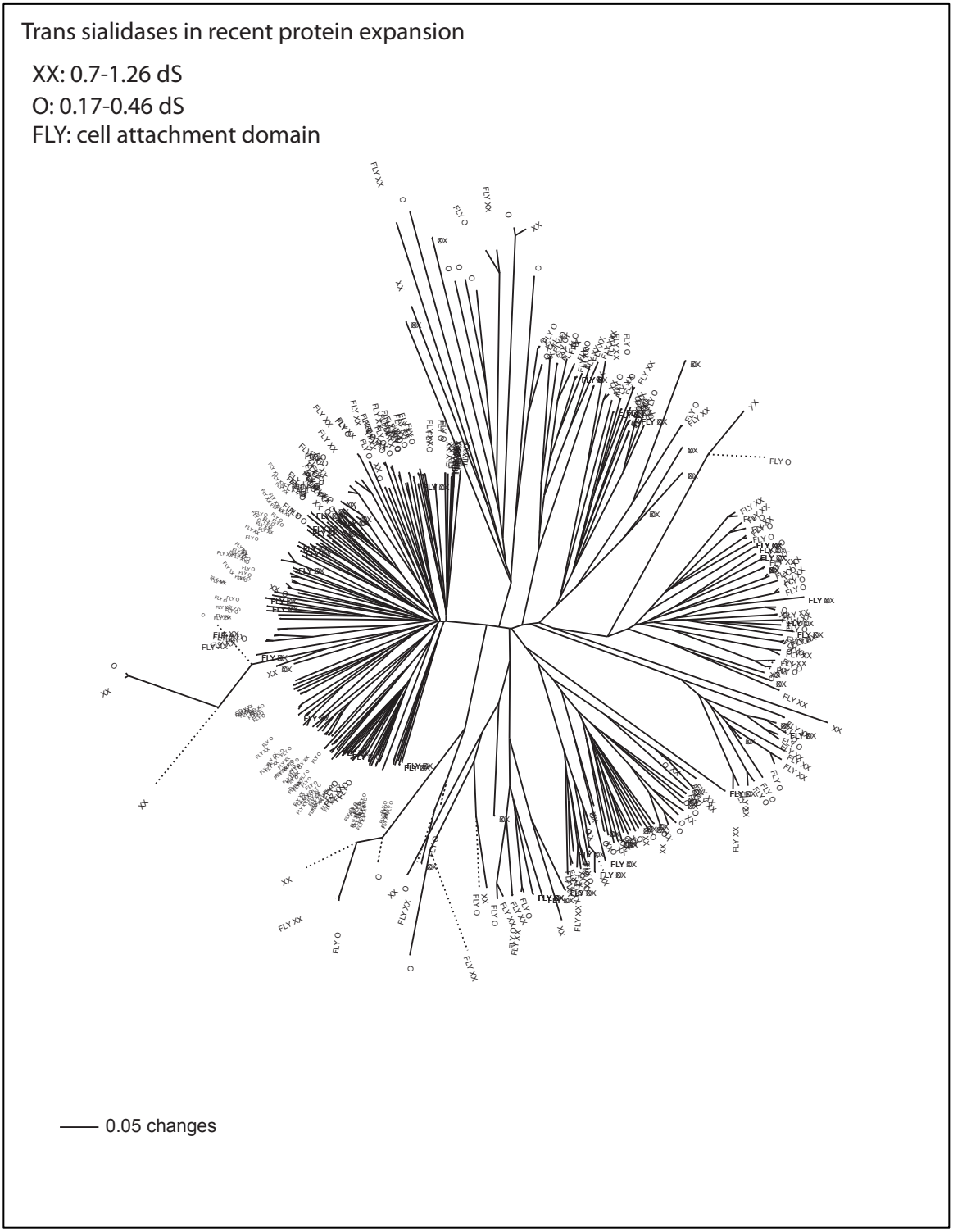
Table S5-1. Overrepresented chromosomal locations for the major protein family expansions. Significant P-values are shown. GeneMergre was used to estimate overrepresented location of paralogs represented within each surface protein family expansion.

Protein surface family	**Trans-sialidases**		**MUCINS**	**MASP's**		**DGF**
Chromosome	dS: 0.17 - 0.46	dS: 0.7 - 1.26	dS: 0.1 - 0.7	dS: 0.02 – 0.23	dS: 0.35 – 0.8	dS: 0.13 – 0.22
1						$1.6x10^{-50}$
2						$2.25x10^{-34}$
3						
4	$1.06x10^{-17}$			$1.44x10^{-11}$		
5						$3.68x10^{-08}$
6		$6.58x10^{-46}$				
7						$9.72x10^{-13}$
8						$6.83x10^{-96}$
9						
10		$2.54x10^{-40}$				
11						
12		$1.65x10^{-65}$				
13						
14						$1.02x10^{-30}$
15	$3.58x10^{-05}$	$5.8x10^{-05}$			$1.45x10^{-12}$	$1.5x10^{-11}$
16	$1.84x10^{-07}$		$1.24x10^{-03}$		$6.78x10^{-07}$	
17						$1.43x10^{-14}$
18	$5.2x10^{-09}$		$4.77x10^{-180}$		$2.72x10^{-75}$	
19						$1.61x10^{-17}$
20		$1.01x10^{-16}$				
21			$5.63x10^{-31}$		$1.19x10^{-06}$	
22						$4.98x10^{-15}$
23	$1.45x10^{-06}$					
24			$9.13x10^{-274}$			$8.33x10^{-21}$
25	$1.03x10^{-40}$	$1.61x10^{-156}$				
26						$2.09x10^{-88}$
27						

Table S5-1 continued

28				$2.3x10^{-143}$	$3.03x10^{-29}$	$1.93x10^{-190}$
29	$2.41x10^{-08}$		$< 1x10^{-500}$			
30						
31						$3.78x10^{-83}$
32						$4.2x10^{-04}$
33			$5.97x10^{-171}$			
34						
35		$1.7x10^{-124}$				
36						
37		$2.61x10^{-06}$				
38	$4.78x10^{-101}$	$1.67x10^{-08}$		$9.16x10^{-83}$	$3.73x10^{-227}$	
39		$7.02x10^{-05}$				
40		$3.15x10^{-06}$				
41	$4.41x10^{-276}$	$3.21x10^{-12}$	$6.03x10^{-36}$	$9.54x10^{-36}$	$1.11x10^{-14}$	

Figure S5-1. Neighbor Joining tree of TS's found in both family expansions. FLY domain is the TS protein motif (VTVxNVxLYNR) that has been shown to bind to host cell membrane receptors (El-Sayed, et al. 2005a; Magdesian, et al. 2001; Tonelli, et al. 2010).



BIBLIOGRAPHY

1999 Recommendations from a satellite meeting. Memorias do Instituto Oswaldo Cruz 94 Suppl 1:429-32.

2002 Control of Chagas disease. World Health Organ Tech Rep Ser 905:i-vi, 1-109, back cover.

Abascal, F., R. Zardoya, and M. J. Telford

2010 TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res 38(Web Server issue):W7-13.

Alexander, J., and K. Bryson

2005 T helper (h)1/Th2 and Leishmania: paradox rather than paradigm. Immunology letters 99(1):17-23.

Alvar, J., et al.

2012 Leishmaniasis Worldwide and Global Estimates of Its Incidence. PLoS One 7(5).

Andrade, S. G. and Magalhaes J.B.

1997 Biodemes and zymodemes of Trypanosoma cruzi strains: correlations with clinical data and experimental pathology. Revista da Sociedade Brasileira de Medicina Tropical 30:27-35.

Anisimova, M., J. P. Bielawski, and Z. Yang

2001 Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol Biol Evol 18(8):1585-92.

Anon

1999 Recommendations from a satellite meeting. Memorias do Instituto Oswaldo Cruz 94(Suppl 1):429-32.

Augusto-Pinto, L., et al.

2001 Molecular cloning and characterization of the DNA mismatch repair gene class 2 from the Trypanosoma cruzi. Gene 272(1-2):323-33.

Augusto-Pinto, L., et al.

2003 Single-nucleotide polymorphisms of the Trypanosoma cruzi MSH2 gene support the existence of three phylogenetic lineages presenting differences in mismatch-repair efficiency. Genetics 164(1):117-26.

Barnabe, C., S. Brisse, and M. Tibayrenc

2000 Population structure and genetic typing of Trypanosoma cruzi, the agent of Chagas disease: a multilocus enzyme electrophoresis approach. Parasitology 120 (Pt 5):513-26.

Barnabe, C., et al.

2001 Trypanosoma cruzi: a considerable phylogenetic divergence indicates that the agent of Chagas disease is indigenous to the native fauna of the United States. Experimental parasitology 99(2):73-9.

Barrett, M. P., et al.

2003 The trypanosomiases. Lancet 362(9394):1469-80.

Beard, C. B., et al.

2003 Chagas disease in a domestic transmission cycle, southern Texas, USA. Emerging infectious diseases 9(1):103-5.

Bern, C., et al.

2011 Trypanosoma cruzi and Chagas' Disease in the United States. Clinical microbiology reviews 24(4):655-81.

Berriman, M., et al.

2005 The genome of the African trypanosome Trypanosoma brucei. Science 309(5733):416-22.

Beucher, M., and K. A. Norris

2008 Sequence diversity of the Trypanosoma cruzi complement regulatory protein family. Infection and immunity 76(2):750-8.

Bosseno, M. F., et al.

2002 Predominance of Trypanosoma cruzi lineage I in Mexico. Journal of clinical microbiology 40(2):627-32.

Briones, M. R. S., et al.

1999 The evolution of two Trypanosoma cruzi subgroups inferred from rRNA genes can be correlated with the interchange of American mammalian faunas in the Cenozoic and has implications to pathogenicity and host specificity. Molecular and Biochemical Parasitology 104(2):219-232.

Brisse, S., C. Barnabe, and M. Tibayrenc

2000 Identification of six Trypanosoma cruzi phylogenetic lineages by random amplified polymorphic DNA and multilocus enzyme electrophoresis. International journal for parasitology 30(1):35-44.

Burton, D. R., et al.

2012 Broadly neutralizing antibodies present new prospects to counter highly antigenically diverse viruses. Science 337(6091):183-6.

Buscaglia, C. A., et al.

2004 The surface coat of the mammal-dwelling infective trypomastigote stage of Trypanosoma cruzi is formed by highly diverse immunogenic mucins. The Journal of biological chemistry 279(16):15860-9.

Buscaglia, C. A., et al.

2006 Trypanosoma cruzi surface mucins: host-dependent coat diversity. Nature reviews. Microbiology 4(3):229-36.

Bush, R. M., et al.

1999 Predicting the evolution of human influenza A. Science 286(5446):1921-5.

Cantey, P. T., et al.

2012 The United States Trypanosoma cruzi Infection Study: evidence for vector-borne transmission of the parasite that causes Chagas disease among United States blood donors. Transfusion 52(9):1922-30.

Castillo-Davis, C. I., and D. L. Hartl

2003 GeneMerge--post-genomic analysis, data mining, and hypothesis testing. Bioinformatics 19(7):891-2.

Castoe, T. A., et al.

2009 Evidence for an ancient adaptive episode of convergent molecular evolution. Proc. Natl. Acad. Sci. USA 106(22):8986-8991.

Castresana, J.

2000 Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17(4):540-52.

CDC

2003 Chagas Disease: Insect vectors and human health. Report of the scientific working group meeting.

Chen, F., et al.

2007 Assessing performance of orthology detection strategies applied to eukaryotic genomes. PLoS One 2(4):e383.

Chuenkova, M. V., and M. A. Pereira

2000 A trypanosomal protein synergizes with the cytokines ciliary neurotrophic factor and leukemia inhibitory factor to prevent apoptosis of neuronal cells. Molecular biology of the cell 11(4):1487-98.

Clark, C. G., and O. J. Pung

1994 Host specificity of ribosomal DNA variation in sylvatic Trypanosoma cruzi from North America. Molecular and biochemical parasitology 66(1):175-9.

Coura, J. R., and J. C. Dias

2009 Epidemiology, control and surveillance of Chagas disease - 100 years after its discovery. Mem Inst Oswaldo Cruz 104:31-40.

Cui, L., et al.

2006 Widespread genome duplications throughout the history of flowering plants. Genome research 16(6):738-49.

Daifalla, N. S., A. G. Bayih, and L. Gedamu

2012 Leishmania donovani recombinant iron superoxide dismutase B1 protein in the presence of TLR-based adjuvants induces partial protection of BALB/c mice against Leishmania major infection. Experimental parasitology.

de Freitas, J. M., et al.

2006 Ancestral genomes, sex, and the population structure of Trypanosoma cruzi. PLoS Pathog 2(3):e24.

de Melo-Jorge, M., and M. PereiraPerrin

2007 The Chagas' disease parasite Trypanosoma cruzi exploits nerve growth factor receptor TrkA to infect mammalian hosts. Cell host & microbe 1(4):251-61.

De Pablos, L. M., and A. Osuna

2012 Multigene families in Trypanosoma cruzi and their role in infectivity. Infection and immunity 80(7):2258-64.

Desjeux, P.

2001 The increase in risk factors for leishmaniasis worldwide. Trans R Soc Trop Med Hyg 95(3):239-43.

Dorn, P. L., et al.

2007 Autochthonous transmission of Trypanosoma cruzi, Louisiana. Emerging infectious diseases 13(4):605-7.

Drummond, A. J., and A. Rambaut

2007 BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:214.

Drummond, AJ, et al.

2007 A Rough Guide to beast 1.4. New Zealand: University of Auckland.

Dvorak, J. A., and T. P. Hyde

1973 Trypanosoma cruzi: interaction with vertebrate cells in vitro. 1. Individual interactions at the cellular and subcellular levels. Experimental parasitology 34(2):268-83.

Edgar, R. C.

2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32(5):1792-7.

El-Sayed, N. M., et al.

2005a The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease. Science 309(5733):409-15.

El-Sayed, N. M., et al.

2005b Comparative genomics of trypanosomatid parasitic protozoa. Science 309(5733):404-9.

Epting, C. L., B. M. Coates, and D. M. Engman

2010 Molecular mechanisms of host cell invasion by Trypanosoma cruzi. Experimental parasitology 126(3):283-91.

Espinoza, B., et al.

1998 Genotype and virulence correlation within Mexican stocks of Trypanosoma cruzi isolated from patients. Acta tropica 70(1):63-72.

Farris, J. S., et al.

1994 Testing significance of congruence. Cladistics 10:315–319.

Fay, J. C.

2011 Weighing the evidence for adaptation at the molecular level. Trends Genet 27(9):343-9.
Felsenstein, J.

1988 Phylogenies from molecular sequences: inference and reliability. Annu. Rev. Genet. 22:521-65.
Fernandes, M. C., and N. W. Andrews

2012 Host cell invasion by Trypanosoma cruzi: a unique strategy that promotes persistence. FEMS microbiology reviews 36(3):734-47.
Finn, R. D., et al.

2006 Pfam: clans, web tools and services. Nucleic Acids Res 34(Database issue):D247-51.
Flores-Lopez, C. A., and C. A. Machado

2011 Analyses of 32 loci clarify phylogenetic relationships among Trypanosoma cruzi lineages and support a single hybridization prior to human contact. PLoS Negl Trop Dis 5(8):e1272.
Franzen, O., et al.

2011 Shotgun sequencing analysis of Trypanosoma cruzi I Sylvio X10/1 and comparison with T. cruzi VI CL Brener. PLoS Negl Trop Dis 5(3):e984.
Franzen, O., et al.

2012 Comparative genomic analysis of human infective Trypanosoma cruzi lineages with the bat-restricted subspecies T. cruzi marinkellei. BMC genomics 13:531.
Garten, R. J., et al.

2009 Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. Science 325(5937):197-201.
Gaunt, M., and M. Miles

2000 The ecotopes and evolution of triatomine bugs (triatominae) and their associated trypanosomes. Memorias do Instituto Oswaldo Cruz 95(4):557-65.
George, R. D., et al.

2011 Trans genomic capture and sequencing of primate exomes reveals new targets of positive selection. Genome Res 21(10):1686-94.
Gilbert, C., et al.

2010 A role for host-parasite interactions in the horizontal transfer of transposons across phyla. Nature 464(7293):1347-50.
Glaser, P., et al.

2001 Comparative genomics of Listeria species. Science 294(5543):849-52.
Gorla, D. E., J. P. Dujardin, and C. J. Schofield

1997 Biosystematics of Old World Triatominae. Acta tropica 63(2-3):127-40.
Grisard, E. C., et al.

2010 Transcriptomic analyses of the avirulent protozoan parasite Trypanosoma rangeli. Molecular and biochemical parasitology 174(1):18-25.

Gu, M., et al.

2011 Positive selection in the hemagglutinin-neuraminidase gene of Newcastle disease virus and its effect on vaccine efficacy. Virology journal 8:150.

Guhl, F., and G. A. Vallejo

2003 Trypanosoma (Herpetosoma) rangeli Tejera, 1920: an updated review. Memorias do Instituto Oswaldo Cruz 98(4):435-42.

Hamilton, P. B., M. M. Teixeira, and J. R. Stevens

2012 The evolution of Trypanosoma cruzi: the 'bat seeding' hypothesis. Trends in parasitology 28(4):136-41.

Hay, W. W., et al.

1999 Alternative global Cretaceous paleogeography. *In* Evolution of the Cretaceous ocean-climate system. E. Barrera and C.C. Johnson, eds. Pp. 1-47. Boulder, CO: Geological Society of America.

Heinz, E., et al.

2012 The genome of the obligate intracellular parasite Trachipleistophora hominis: new insights into microsporidian genome dynamics and reductive evolution. PLoS pathogens 8(10):e1002979.

Herwaldt, B. L., et al.

2000 Use of polymerase chain reaction to diagnose the fifth reported US case of autochthonous transmission of Trypanosoma cruzi, in Tennessee, 1998. The Journal of infectious diseases 181(1):395-9.

Higo, H., et al.

2004 Genotypic variation among lineages of Trypanosoma cruzi and its geographic aspects. Parasitol Int 53(4):337-44.

Hotez, P. J., et al.

2007 Control of neglected tropical diseases. N Engl J Med 357(10):1018-27.

Ivens, A. C., et al.

2005 The genome of the kinetoplastid parasite, Leishmania major. Science 309(5733):436-42.

Kjos, S. A., K. F. Snowden, and J. K. Olson

2009 Biogeography and Trypanosoma cruzi infection prevalence of Chagas disease vectors in Texas, USA. Vector borne and zoonotic diseases 9(1):41-50.

Koffi, M., et al.

2009 Population genetics of Trypanosoma brucei gambiense, the agent of sleeping sickness in Western Africa. Proc Natl Acad Sci U S A 106(1):209-14.

Laurent, J. P., et al.

1997 Impact of clonal evolution on the biological diversity of Trypanosoma cruzi. Parasitology 114 (Pt 3):213-8.

Leiby, D.

1997 Blood banking (Safety): Seroepidemiology of Trypanosoma cruzi, etiological agent of Chagas' Disease, in US blood donors. Blood Weekly 8(December):8-20.

—

2004 Present status of studies on Trypanosoma cruzi in U.S. Blood Donors.

Lewis, M. D., et al.

2011 Recent, independent and anthropogenic origins of Trypanosoma cruzi hybrids. Plos Neglected Tropical Diseases 5(10):e1363.

Lima, L., et al.

2012 Evolutionary insights from bat trypanosomes: morphological, developmental and phylogenetic evidence of a new species, Trypanosoma (Schizotrypanum) erneyi sp. nov., in African bats closely related to Trypanosoma (Schizotrypanum) cruzi and allied species. Protist 163(6):856-72.

Llewellyn, M. S., et al.

2009 Genome-scale multilocus microsatellite typing of Trypanosoma cruzi discrete typing unit I reveals phylogeographic structure and specific genotypes linked to human infection. PLoS Pathog 5(5):e1000410.

Lukes, J., et al.

2007 Evolutionary and geographical history of the Leishmania donovani complex with a revision of current taxonomy. Proc Natl Acad Sci U S A 104(22):9375-80.

Lynch, M., and J. S. Conery

2000 The evolutionary fate and consequences of duplicate genes. Science 290(5494):1151-5.

—

2003 The evolutionary demography of duplicate genes. Journal of structural and functional genomics 3(1-4):35-44.

Machado, C. A., and F. J. Ayala

2002 Sequence variation in the dihydrofolate reductase-thymidylate synthase (DHFR-TS) and trypanothione reductase (TR) genes of Trypanosoma cruzi. Molecular and biochemical parasitology 121(1):33-47.

Maddison, D. R., Donoghue M.J. and Maddison D.R.

1984 Outgroup analysis and Parsimony. Syst. Zool. 33(1):83-103.

Magdesian, M. H., et al.

2001 Infection by Trypanosoma cruzi. Identification of a parasite ligand and its host cell receptor. The Journal of biological chemistry 276(22):19382-9.

Manso-Alves, M.J. & Arruda-Mortara, R.

2009 A century of research: what have we learned about the interactions of Trypanosoma cruzi with host cells? Mem. Inst. Osw. Cruzi 104:76-88.

Mantilla, J. C., et al.

2010 Mixed infection of Trypanosoma cruzi I and II in a Colombian cardiomyopathic patient. Human pathology 41(4):610-3.
McCann, H. C., et al.

2012 Identification of innate immunity elicitors using molecular signatures of natural selection. Proc Natl Acad Sci U S A 109(11):4215-20.
Miles, M. A., M. D. Feliciangeli, and A. R. de Arias

2003 American trypanosomiasis (Chagas' disease) and the role of molecular epidemiology in guiding control strategies. BMJ 326(7404):1444-8.
Miles, M. A., et al.

1978 Isozymic heterogeneity of Trypanosoma cruzi in the first autochthonous patients with Chagas' disease in Amazonian Brazil. Nature 272(5656):819-21.
Monteiro, F. A., et al.

2003 Molecular phylogeography of the Amazonian Chagas disease vectors Rhodnius prolixus and R. robustus. Mol Ecol 12(4):997-1006.
Montilla, M., et al.

2002 Isoenzyme clustering of Trypanosomatidae Colombian populations. American Journal of Tropical Medicine and Hygiene 66(4):394-400.
Moradin, N., and A. Descoteaux

2012 Leishmania promastigotes: building a safe niche within macrophages. Frontiers in cellular and infection microbiology 2:121.
Mottram, J. C., G. H. Coombs, and J. Alexander

2004 Cysteine peptidases as virulence factors of Leishmania. Current opinion in microbiology 7(4):375-81.
Nei, M.

1987 Molecular Evolutionary Genetics. New York: Columbia University Press.
Nerima, B., D. Nilsson, and P. Maser

2010 Comparative genomics of metabolic networks of free-living and parasitic eukaryotes. BMC genomics 11:217.
Nielsen, R.

2005 Molecular signatures of natural selection. Annu Rev Genet 39:197-218.
Nielsen, R., et al.

2005 A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS biology 3(6):e170.
Noireau, F., P. Diosque, and A. M. Jansen

2009 Trypanosoma cruzi: adaptation to its vectors and its hosts. Veterinary research 40(2):26.
Norris, K. A., et al.

1991 Characterization of a Trypanosoma cruzi C3 binding protein with functional and genetic similarities to the human complement regulatory protein, decay-accelerating factor. Journal of immunology 147(7):2240-7.

Noyes, H. A., et al.

1998 Leishmania (sauroleishmania): a comment on classification. Parasitology today 14(4):167.

Nunes, L. R., M. R. de Carvalho, and G. A. Buck

1997 Trypanosoma cruzi strains partition into two groups based on the structure and function of the spliced leader RNA and rRNA gene promoters. Mol Biochem Parasitol 86(2):211-24.

Ochs, D. E., et al.

1996 Postmortem diagnosis of autochthonous acute chagasic myocarditis by polymerase chain reaction amplification of a species-specific DNA sequence of Trypanosoma cruzi. The American journal of tropical medicine and hygiene 54(5):526-9.

Ohno, S

1970 Evolution by gene duplication.

Oleksyk, T. K., M. W. Smith, and S. J. O'Brien

2010 Genome-wide scans for footprints of natural selection. Philos Trans R Soc Lond B Biol Sci 365(1537):185-205.

Organization, Pan American Health

2006 Estimación cuantitativa de la enfermedad de Chagas en las Américas.

Pacheco, M. A., et al.

2012 Evidence of purifying selection on merozoite surface protein 8 (MSP8) and 10 (MSP10) in Plasmodium spp. Infect Genet Evol 12(5):978-86.

Parfrey, L. W., et al.

2011 Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proceedings of the National Academy of Sciences of the United States of America 108(33):13624-9.

Pereira, M. E., et al.

1980 Lectin receptors as markers for Trypanosoma cruzi. Developmental stages and a study of the interaction of wheat germ agglutinin with sialic acid residues on epimastigote cells. The Journal of experimental medicine 152(5):1375-92.

Petersen, L., et al.

2007 Genes under positive selection in Escherichia coli. Genome Res 17(9):1336-43.

Pinto, C. M., et al.

2012 TcBat a bat-exclusive lineage of Trypanosoma cruzi in the Panama Canal Zone, with comments on its classification and the use of the 18S rRNA gene for lineage identification. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases 12(6):1328-32.

Poinar, G., Jr.

2005 Triatoma dominicana sp. n. (Hemiptera: Reduviidae: Triatominae), and Trypanosoma antiquus sp. n. (Stercoraria: Trypanosomatidae), the first fossil evidence of a triatomine-trypanosomatid vector association. Vector borne and zoonotic diseases 5(1):72-81.

Posada, D.

2008 jModelTest: phylogenetic model averaging. Mol Biol Evol 25(7):1253-6.

Prata, A.

2001 Clinical and epidemiological aspects of Chagas disease. The Lancet infectious diseases 1(2):92-100.

Privman, E., O. Penn, and T. Pupko

2011 Improving the performance of positive selection inference by filtering unreliable alignment regions. Mol Biol Evol 29(1):1-5.

Purkait, B., et al.

2012 Mechanism of amphotericin B resistance in clinical isolates of Leishmania donovani. Antimicrobial agents and chemotherapy 56(2):1031-41.

Rambaut, A.

2002 Se-Al. Sequence alignment editor. Oxford, UK: University of Oxford.

Rassi, A., Jr., A. Rassi, and J. A. Marin-Neto

2010 Chagas disease. Lancet 375(9723):1388-402.

Razin, S.

1997 Comparative genomics of mycoplasmas. Wiener klinische Wochenschrift 109(14-15):551-6.

Reisenman, C. E., et al.

2010 Infection of kissing bugs with Trypanosoma cruzi, Tucson, Arizona, USA. Emerging infectious diseases 16(3):400-5.

Revollo, S., et al.

1998 Trypanosoma cruzi: impact of clonal evolution of the parasite on its biological and medical properties. Exp Parasitol 89(1):30-9.

Robello, C., et al.

2000 Evolutionary relationships in Trypanosoma cruzi: molecular phylogenetics supports the existence of a new major lineage of strains. Gene 246(1-2):331-8.

Roellig, D. M., et al.

2008 Molecular typing of Trypanosoma cruzi isolates, United States. Emerging infectious diseases 14(7):1123-5.

Roellig, D. M., et al.

2013 Genetic variation and exchange in Trypanosoma cruzi isolates from the United States. PLoS One 8(2):e56198.

Rozen, D. E., and H. Skaletsky

2000 Primer3 on the WWW for general users and for biologist programmers. *In* Bioinformatics Methods and Protocols: Methods in Molecular Biology. S. Krawetz and S. Misener, eds. Pp. 365-386. Totowa, NJ: Humana Press.

Sarkar, S., et al.

2010 Chagas disease risk in Texas. Plos Neglected Tropical Diseases 4(10).

Schiffler, R. J., et al.

1984 Indigenous Chagas' disease (American trypanosomiasis) in California. JAMA : the journal of the American Medical Association 251(22):2983-4.

Schmidt, G.D. & Roberts, L.S.

2005 Foundations of Parasitology.

Schofield, C.

2000 *Trypanosoma cruzi* : the vector-parasite paradox. Memorias do Instituto Oswaldo Cruz 95(4):535-44.

Sharp, P. M., and W. H. Li

1987 The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. Molecular biology and evolution 4(3):222-30.

Shimodaira, H., and M. Hasegawa

1999 Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16(8):1114-1116.

Simpson, A. G., Y. Inagaki, and A. J. Roger

2006 Comprehensive multigene phylogenies of excavate protists reveal the evolutionary positions of "primitive" eukaryotes. Mol Biol Evol 23(3):615-25.

Soltis, P. S., and D. E. Soltis

2009 The role of hybridization in plant speciation. Annual review of plant biology 60:561-88.

Souto, R. P., et al.

1996 DNA markers define two major phylogenetic lineages of Trypanosoma cruzi. Molecular and biochemical parasitology 83(2):141-52.

Souto, R. P., and B. Zingales

1993 Sensitive detection and strain classification of Trypanosoma cruzi by amplification of a ribosomal RNA sequence. Molecular and biochemical parasitology 62(1):45-52.

Soyer, Y., et al.

2009 Genome wide evolutionary analyses reveal serotype specific patterns of positive selection in selected Salmonella serotypes. BMC Evol Biol 9:264.

Stamatakis, A., P. Hoover, and J. Rougemont

2008 A rapid bootstrap algorithm for the RAxML Web servers. Systematic biology 57(5):758-71.

Stevens, L., et al.

2012 Vector blood meals and Chagas disease transmission potential, United States. Emerging infectious diseases 18(4):646-9.

Stoco, P. H., et al.

2012 Trypanosoma rangeli expresses a beta-galactofuranosyl transferase. Experimental parasitology 130(3):246-52.

Stramer, S. L., et al.

2007 Blood screening for Chagas Disease, United States. Morbidity and Mortality Weekly Report 23:141-143.

Sturm, N. R., and D. A. Campbell

2010 Alternative lifestyles: the population structure of Trypanosoma cruzi. Acta Trop 115(1-2):35-43.

Sturm, N. R., et al.

2003 Evidence for multiple hybrid groups in Trypanosoma cruzi. Int J Parasitol 33(3):269-79.

Subileau, M., et al.

2009 Trypanosoma cruzi: new insights on ecophylogeny and hybridization by multigene sequencing of three nuclear and one maxicircle genes. Exp Parasitol 122(4):328-37.

Thompson, J. D., D. G. Higgins, and T. J. Gibson

1994 CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22(22):4673-4680.

Tibayrenc, M.

1995 Population genetics of parasitic protozoa and other microorganisms. Advances in parasitology 36:47-115.

—

2010 Modelling the Transmission of Trypanosoma cruzi: The Need for an Integrated Genetic Epidemiological and Population Genomics Approach. *In* Modelling Parasite Transmission and Control. E. Michael and R.C. Spear, eds. Pp. 200-211. Advances in Experimental Medicine and Biology, Vol. 673. Austin, TX: Landes.

Tibayrenc, M., and F. J. Ayala

1988 Isozyme variability in *Trypanosoma cruzi*, the agent of Chagas' disease: Genetical, taxonomical, and epidemiological significance. Evolution 42(2):277-292.

Tibayrenc, M., and S. F. Breniere

1988 Trypanosoma cruzi: major clones rather than principal zymodemes. Memorias do Instituto Oswaldo Cruz 83 Suppl 1:249-55.

Tibayrenc, M., et al.

1993 Genetic characterization of six parasitic protozoa: parity between random-primer DNA typing and multilocus enzyme electrophoresis. Proc. Natl. Acad. Sci. USA 90(4):1335-9.

Tibayrenc, M., et al.

1986 Natural populations of *Trypanosoma cruzi*, the agent of Chagas disease, have a complex multiclonal structure. Proc. Natl. Acad. Sci. USA 83(1):115-9.

Toft, C., and S. G. Andersson

2010 Evolutionary microbial genomics: insights into bacterial host adaptation. Nature reviews. Genetics 11(7):465-75.

Tomazi, L., et al.

2009 Haplotype distribution of five nuclear genes based on network genealogies and Bayesian inference indicates that Trypanosoma cruzi hybrid strains are polyphyletic. Genet Mol Res 8(2):458-76.

Tonelli, R. R., et al.

2010 Role of the gp85/trans-sialidases in Trypanosoma cruzi tissue tropism: preferential binding of a conserved peptide motif to the vasculature in vivo. Plos Neglected Tropical Diseases 4(11):e864.

Tsai, I. J., et al.

2013 The genomes of four tapeworm species reveal adaptations to parasitism. Nature 496(7443):57-63.

Vago, A. R., et al.

1996 Kinetoplast DNA signatures of Trypanosoma cruzi strains obtained directly from infected tissues. The American journal of pathology 149(6):2153-9.

Van de Peer, Y., S. Maere, and A. Meyer

2009 The evolutionary significance of ancient genome duplications. Nature reviews. Genetics 10(10):725-32.

Vanneste, K., Y. Van de Peer, and S. Maere

2013 Inference of genome duplications from age distributions revisited. Molecular biology and evolution 30(1):177-90.

Wan, X-F., et al.

2004 Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes. BMC Evol. Biol. 4(19).

Wan, X-F., and J. Zhou

2003 A new informatics method measuring synonymous codon usage bias. Intelligent engineering systems through artificial neural networks 13.

Weatherly, D. B., C. Boehlke, and R. L. Tarleton

2009 Chromosome level assembly of the hybrid Trypanosoma cruzi genome. BMC genomics 10:255.

Wernegreen, J. J.

2005 For better or worse: genomic consequences of intracellular mutualism and parasitism. Current opinion in genetics & development 15(6):572-83.

Westenberger, S. J., et al.

2005 Two hybridization events define the population structure of Trypanosoma cruzi. Genetics 171(2):527-43.

WHO

2011 Chagas disease (American trypanosomiasis).

Wilson, D. J.

2012 Insights from genomics into bacterial pathogen populations. PLoS pathogens 8(9):e1002874.

Wong, W. S., et al.

2004 Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics 168(2):1041-51.

Woody, N. C., and H. B. Woody

1955 American trypanosomiasis (Chagas' disease); first indigenous case in the United States. Journal of the American Medical Association 159(7):676-7.

Xu, Z., H. Chen, and R. Zhou

2011 Genome-wide evidence for positive selection and recombination in Actinobacillus pleuropneumoniae. BMC Evol Biol 11:203.

Yang, Z.

2007 PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24(8):1586-91.

Yang, Z., and R. Nielsen

2000 Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Molecular biology and evolution 17(1):32-43.

Yeo, M., et al.

2005 Origins of Chagas disease: Didelphis species are natural hosts of Trypanosoma cruzi I and armadillos hosts of Trypanosoma cruzi II, including hybrids. Int J Parasitol 35(2):225-33.

Zhang, Y., et al.

2011 Genes under positive selection in Mycobacterium tuberculosis. Comput Biol Chem 35(5):319-22.

Zingales, B., et al.

2012 The revised Trypanosoma cruzi subspecific nomenclature: rationale, epidemiological relevance and research applications. Infect Genet Evol 12(2):240-53.

Zingales, B., Souto, R.P., Mangia, R.H., Lisboa, C.V., Campbell, D.A., Coura, J.R., Jansen, A. and Fernandes, O.

1998 Molecular epidemiology of American trypanosomiasis in Brazil based on dimorphisms of rRNA and mini-exon gene sequences. International Journal for Parasitology 28:105-112.

Zingales, B., et al.

1999 Epidemiology, biochemistry and evolution of *Trypanosoma cruzi* lineages based on ribosomal RNA sequences. Mem Inst Oswaldo Cruz 94(Suppl 1):159-64.

Zumaya-Estrada, F. A., et al.

2012 North American import? Charting the origins of an enigmatic Trypanosoma cruzi domestic genotype. Parasites & vectors 5:226.