Mitchell, Elizabeth A., Spatial phylogeography and correlation between genetic and geographical distance in *Trypanosoma cruzi* and its vectors of the genus *Triatoma* from Texas. Master of Science, August, 2013, (70) pp, (23) figures, (43) references

This study proposes to describe the phylogenetic relationship among *Triatoma* species that carry *Trypanosoma cruzi* and between *T. cruzi* isolates, present in Texas. Additionally, it evaluates any correlation between geographical location of *Triatoma* collection and genotype of *T. cruzi* isolates, and the distribution of *T. cruzi* genotypes relative to ecological variables potentially affecting maintenance in zoonotic cycles or affecting transmission rates.

T. cruzi and *Triatoma* spp. DNA was amplified by PCR, using primers targeting several conserved loci. The amplicons were sequenced and used as markers in establishing phylogenetic relationships. Spatial modeling was utilized to overlay vector and parasite geographic distributions with environmental factors in order to illustrate any potential associations.

SPATIAL PHYLOGEOGRAPHY AND CORRELATION BETWEEN GENETIC AND GEOGRAPHICAL DISTANCE IN *TRYPANOSOMA CRUZI* AND ITS VECTORS OF THE GENUS *TRIATOMA* FROM TEXAS

THESIS

Presented to the Graduate Council

of the Graduate School of Biomedical Sciences

University of North Texas

Health Science Center at Fort Worth

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

By

Elizabeth A. Mitchell, B.A.

Fort Worth, TX

August 2013

ACKNOWLEDGMENTS

First, I would like to thank the members of my committee as a whole for their time and service. Many thanks go to Michael Allen, who took me on as a student and provided sage advice as well as flexibility while I was still taking classes. I would also like to thank Arthur Eisenberg for all his many (important) signatures, for funding a portion of my training and for his firm commitment not only to his students but also the field of forensic science. There is more justice in the world because of Dr. Eisenberg and his work, and I am proud to say I'm a graduate of his program. I owe a debt of gratitude to Sahotra Sarkar for sharing with me his Renaissance brilliance and for taking time out of his busy schedule to teach me about modeling and *Triatoma* taxonomy, and to Chissa Rivaldi for making spatial modeling seem so easy, as well as the rest of the Sarkar Laboratory at UT Austin and all other specimen contributors for their roles in the collection and identification of my test samples. I would also like to thank Joseph Yuan for his service as my university member and Jonathan King for all his satire, advice, and for being so excited about Excel.

Additionally, there are a number of people without whom this thesis might not have been written, and to whom I am especially grateful. Bobby LaRue, Shantanu Shewale and Carey Davis managed to help maintain my sanity with their friendship, viewpoints from the other side, and the "occasional" tasty beverage. I also wish to express my sincerest gratitude to my friend and mentor, Peggy Billingsley. In addition to the integral part she played in this project, she honed my lab skills, gave me new ones, showed me how to expertly cut up insects full of less than luxurious things, and taught me all sorts of other "science-y" things before leaving me in the nest to fly on my own. Most importantly, I would like to thank Phillip Williamson and Meredith Turnbough, my friends, mentors, and unceasing supporters who freely gave me opportunities,

this awesome project, their time, and copious amounts of academic and emotional support over the past two years. They changed my direction and (patiently) kept me moving forward, and I honestly don't know what I would have done without them. I would also like to mention that Dr. Williamson did not have to provide guidance for this project to the extent that he did. However, he voluntarily and enthusiastically did so despite being extremely busy, and I appreciate this more than I can say.

Last but not least, I would like to thank my parents as well as the rest of my family, who have always had faith in me, even though they thought (and probably still do) that I might be in school forever. I appreciate their love, prayers, support, and all the "funny" little jokes.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AUC	area under curve
bp	basepair
BSA	bovine serum albumin
°C	degrees Celsius
COII-ND1	cytochrome oxidase subunit II + NADH dehydrogenase
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
GIS	geographic information system
GPS	global positioning system
km	kilometer
MaxEnt	Maximum Entropy Species Distribution Modeling
MEGA	Molecular Evolutionary Genetics Analysis
$MgCl_2$	magnesium chloride
µl	microliter
µM	micromolar
mM	millimolar
ml	milliliter
mm	millimeter
min	minute
M	molar
mtDNA	mitochondrial DNA
n	number
NaOCl	sodium hypochlorite
PCR	polymerase chain reaction

R. prolixus	*Rhodnius prolixus*
s	second
spp.	species
TAE	tris (hydroxymethyl) aminomethane, acetic acid, ethylenediamine tetraacetic acid
T. brucei	*Trypansoma brucei*
T. cruzi	*Trypanosoma cruzi*
T. dimidiata	*Triatoma dimidiata*
T. gerstaeckeri	*Triatoma gerstaeckeri*
T. indictiva	*Triatoma indictiva*
T. lecticularia	*Triatoma lecticularia*
T. protracta	*Triatoma protracta*
T. rubida	*Triatoma rubida*
T. sanguisuga	*Triatoma sanguisuga*
UV	ultraviolet
U	units
v	version

CHAPTER I

INTRODUCTION

Chagas disease, or American Trypanosomiasis, was identified in 1909 by a Brazilian physician for whom the disease was named, Carlos Chagas (1). Well-documented occurrence of Chagas disease is primarily concentrated in Latin America, and the disease is commonly associated with poor or rural populations (2). The World Health Organization now estimates that at least 10 million people are infected globally, resulting in greater than 15,000 deaths per year, with an additional 75-90 million people being susceptible to exposure (1, 3, 4). This number is in contrast to a much higher infection rate estimation of 16-18 million in 1990. The drop in prevalence was presumably due to initiation of mass vector and serological control programs in Latin American countries over the past 20 years (2). While the majority of cases of Chagas disease are concentrated in Latin America, it is not geographically restricted. There have been confirmed diagnoses in Europe, Japan, Australia, and North America (1). It is considered endemic in the United States with Texas being a primary region of growing epidemiological concern (2).

Trypansoma cruzi, a parasitic hemoflagellate protozoan of the class Kinetoplastida, is the etiological agent of Chagas disease. The infection begins with a short incubation period lasting about 2 weeks after exposure to *T. cruzi*. This is followed by an 8-12 week acute phase, during which high levels of parasites are circulating in the blood and can be found in all tissues (5, 6). Parasites can be seen in blood smears microscopically, and patients can be either asymptomatic or sometimes demonstrate vague symptoms (5). It is during this phase that potential for treatment with anti-parasitic drugs is highest. Following the acute phase, blood parasite levels

drop until they are not detectable microscopically, and ultimately, the development of the lifelong chronic phase begins (5). Patients may remain asymptomatic during all phases, leading to the potential for many to remain undiagnosed and untreated. Clinical presentations of Chagas disease vary based on what form the patient has. The indeterminate form is characterized by a lack of symptoms but the presence of *T. cruzi* antibodies in serological testing. The cardiac form includes sequelae of the heart, which is initially caused by chronic inflammation in cardiac tissue and other detrimental immune responses seemingly due to the presence of the *T. cruzi* parasites residing in cardiac tissue (5). Manifestations include cardiomyopathy, congestive heart failure, conduction abnormalities and complications resulting from cardiac lesions or scar tissue. The gastrointestinal form of Chagas disease occurs more rarely. This form eventually leads to megacolon or megaesophagus due to parasitic invasion of gastrointestinal tissues and the resulting injury to neurons located along the walls of the digestive tract (5). A Chagas patient could have either form of the disease, and rarely both, in varying degrees (7). The variability in histotropism is hypothesized to partially be the result of specific interactions between parasite ligands, such as surface glycoproteins, and both intra and extracellular host cell receptors, which is due to different expression levels of variant *T. cruzi* genotypes (6). Variability in histotropism seems to be influenced by geographic origin of infection, with origination of the gastrointestinal form being confined to areas south of the equator (7). Parasitic pathogenicity, morphology and interaction with host immune responses are also known to differ by region (7). For example, the cardiac form tends to be more severe south of the equator, and treatments are usually less effective than in northern countries (7). This geographic divergence is thought to be caused by intraspecific genetic variation among *T. cruzi* isolates, with distinct genotypes being located in particular regions (4, 5, 6).

T. cruzi can be transmitted to a significant number of mammalian hosts through blood transfusion, organ transplantation, congenital means, ingestion and via its hematophagous triatomine vector (Hemiptera: Reduviidae), with vector-borne transmission being the primary source of human infections (8). The parasites enter the vector through an infected blood meal and spend several days in the foregut before moving into the midgut, where they begin replicating extracellularly through binary fission (9). They then move into the hindgut, where they stop replicating and are released into the environment in the insect's urine and feces.

Infection of the mammalian host occurs when the vector carrying *T. cruzi* feeds on mammalian blood and then defecates close to the bite wound. The parasites in the fecal matter enter the host through the bite wound or nearby mucous membranes. Once inside the host, they replicate intracellularly (9). After erupting out the cell, they circulate freely in the blood and eventually invade other tissue cells.

The vectors of *T. cruzi* belong to the Triatominae family of insects, which can be classified into 130 species (8). Eleven Triatominae species have been found in the United States, and all but two are known to carry *T. cruzi* (8). Seven of the US species have been reported in Texas, all within the genus *Triatoma* (10). Texas and Arizona have the greatest number of species that could potentially interact with one another and their respective reservoirs, which could lead to modifications of the natural zoonotic cycle for reasons such as adaptations, migrations or competition (5). The location of Texas as a crossroads between both west and east *Triatoma* spp. distributions and between species of Mexico and the US, its size, as well as ecological variability, make Texas a primary region for *Triatoma* research. Each species in Texas has been found to carry *T. cruzi*, either in sylvatic (wild) or domestic/peridomestic settings (5, 10). Many of these species are known to display broad geographic distribution across the

state, with overlap occurring in range and across the ten ecoregions, or defined areas in Texas that share distinct ecological features, such as climate, vegetation, wildlife and land usage (8, 11). Intraspecific polymorphism has also been seen in *Triatoma* spp. of different regions as well as strong morphological similarities between different species, and potential reorganization of the current taxonomic classification among species in Texas has been proposed (5). There is also a need for genetic characterization and analysis of phylogenetic relationships between *Triatoma* spp. in Texas. This information could be compared with morphological similarities and differences between and among species which would aid in clarifying whether the taxonomy should be revised.

A 2009 study by Kjos, *et al*, specifically on Texas *Triatoma* spp., found that greater than 50% of the sample specimens tested positive for *T. cruzi*, and those vectors have spanned most geographic areas across the state (8, 10). *Triatoma gerstaeckeri*, *Triatoma lecticularia*, and *Triatoma sanguisuga* exhibit the largest distribution in Texas, and all three participate in *T. cruzi* peridomestic and domestic cycles which increase the risk of transmission to humans (10, 12). Previous studies have shown that the ability of the triatomine to transmit *T. cruzi* is not equal across species, and species with the highest parasite load are not always responsible for the highest number of transmissions to mammalian hosts (7). Transmission depends on how efficiently the potential vector assists in transforming the parasite to certain stages of its life cycle that are required for mammalian infectivity, so if efficiency is low, transmission rate could be low as well, even if parasite load is high (8). The transformation of *T. cruzi* within the vector is thought to be supported and possibly relies on biochemical exchanges between the parasitic cells and the membranes of cells lining the gut (9). In addition, variability has been noted in triatomine defecation behavior, both between and among species, and across various geographic

distributions (13). *Triatoma* spp. found in the United States are reported to delay defecation a longer period of time after feeding than species found elsewhere, and this is potentially a significant factor in autochthonous transmissibility risk (5). This difference in behavior could imply the existence of genetic variation between geographic locations.

There have been seven documented autochthonous cases of Chagas disease in the United States, with four occurring in Texas (5). Despite this low number, serologic studies in wild mammalian reservoirs have reported a high percentage of sylvatic cases, as well as a concerning number of peridomestic/domestic cases involving canine infections, many originating in Texas (8, 12, 14). Those studies also indicated heterogeneity in *T. cruzi* genotypes among various reservoir species (14). In the US, many sylvatic *T. cruzi* infections are reported to have been caused by isolates belonging to TcIV. This genotype has never been found in humans in North America, and a 2013 study by Roellig, *et al*, showed that the TcIV genotype exhibited a much higher degree of variability than isolates of the other 5 genotypes and has proved to be less virulent in mouse models (15). This could suggest that genetic variation among different *T. cruzi* isolates may result in both host specificity and intraspecific dissimilarities in virulence.

Detection of *T. cruzi* in human hosts and method reliability depend on many factors, such as phase of the disease, age of patient and how and where patient became infected. There is currently not a single test that can lead to a definitive positive or negative diagnosis. Common methods include polymerase chain reaction and various presumptive and confirmatory immunoassays (13). Sensitivity seems to vary with geographic origin of infection, potentially due to variant *T. cruzi* genotypes between locations (13). Genotypic differences between parasites could also lead to resulting variation in host immune response, and because many of the

serologic detection methods are based on factors involved in expected immune response, this could result in detection sensitivity differences (5).

Despite indications of a potentially growing Chagas disease risk in Texas, there have been few studies focused entirely on the heterogeneity of trypanosome or *Triatoma* genotypes found in Texas. The variability of genotypes and how each corresponds to the spatial distribution across Texas has not been assessed. The primary purpose of this study is to investigate whether a correlation exists between genotype, phylogenetic distance and geographic occurrence for both *T. cruzi* and *Triatoma* spp. occurring in Texas, and to determine if the presence and circulation of *T. cruzi* among the *Triatoma* populations is affected by ecological variables.

CHAPTER II

MATERIALS AND METHODS

Sample Collection

From June 2009 through August 2012, 300 triatomine specimens were field-collected from multiple locations and habitats in Texas, including domestic, peridomestic and sylvatic and/or mixed, and provided through various means, primarily by the field collection efforts of Dr. Sahotra Sarkar of the Department of Integrative Biology at the University of Texas. Each specimen was individually stored in ethanol, in a sealed tube. Taxonomic identification was performed in the Sarkar laboratory at the University of Texas using morphological keys as defined by Lent and Wygodzinsky (16). The majority of specimens were keyed as *T. gerstaeckeri* (*n*=237), with *T. sanguisuga* (*n*=23), *T. lecticularia* (*n*=11), *T. indictiva* (*n*=6), *T. protracta* (*n*=1), and *T. rubida* (*n*=1) being identified as well. The collection also included nymphs and other morphologically unidentifiable specimens (*n*=21).

DNA Extraction

All procedures involving triatomine bugs were performed in a Class II Type B biosafety cabinet with HEPA filtration that was sanitized with a 0.6% NaOCl solution prior to and after each use. The biosafety cabinet was also UV irradiated for 30 minutes prior to and after each use. The surface of each *Triatoma* specimen was also disinfected with three consecutive washes in a 0.6% NaOCl solution followed by three rinses in molecular-grade water. Each specimen was individually placed on a sterile petri dish for dissection and allowed to air dry (Figure 1).

T. cruzi Extraction

The posterior end of each body segment was removed using a sterile blade and forceps in order to isolate the hindgut (Figure 2). Each segment was then placed into a sterile 1.5 ml centrifuge tube and manually pulverized with a sterile pestle. The genetic material in each hindgut sample was extracted using the Purification of Total DNA from Animal Tissues (Spin Column) Protocol from the DNeasy® Blood and Tissue Kit, (Qiagen, Valencia, CA). All steps were performed according to the manufacturer's protocol.

Triatomine Extraction

One leg was removed from each specimen with sterilized forceps, cut into smaller pieces, and placed in a sterile 1.5 ml centrifuge tube. DNA was extracted using the E.Z.N.A® Mollusc DNA kit (Omega Bio-Tek, Norcross, GA). The manufacturer's protocol was followed as directed except that pulverization of each sample occurred in the lysis buffer and Proteinase K solution using a sterile pestle, and the initial incubation step at 60^0C was extended overnight.

In all cases, the final elution of DNA was stored at -20^0C until future use. All pre-PCR and post-PCR handling of samples were carried out in separate, designated areas to prevent contamination. Certified DNA/RNAse free filter barrier tips were utilized to prevent aerosol contamination. PCR assays were performed in duplicate with appropriate controls.

Triatoma spp. PCR Amplification

The loci amplified for analysis include a region of the 16S mitochondrial ribosomal DNA (16S rDNA) approximately 400 bp in length, a 663 bp region of the cytochrome B gene (*cytB*), and a 279 bp region of the cytochrome c oxidase subunit I (COI) gene.

Figure 1: *Triatoma* specimens drying prior to dissection.

Figure 2: Posterior end of *Triatoma* specimen that was removed to isolate the hindgut, in testing for the presence of *T. cruzi*.

16S rDNA

The primers used for this amplification were: LRJ 12966 (5'-AAAAAAATTACGCTGTTATCCCTAAAGTAA-3') and LRN 13393 (5'-C(G/A)CCTGTTTAACAAAAACAT-3') (17). The PCR cycling conditions were as follows: an initial denaturation step at 94°C for 5 min, followed by 35 cycles consisting of denaturation at 94°C for 30 s, primer annealing at 47.5°C for 30 s, and extension at 72°C for 60 s, followed by a final extension step at 72°C for 5 min and a 4°C indefinite hold.

Cytochrome B

This region was isolated using the primers: CytB 7432F (5'-GGACG(A/T)GG(A/T)ATTTATTATGGATC-3') and CytB 7433R (5'-GC(A/T)CCAATTCA(A/G)GTTA(A/G)TAA-3') (18). The PCR cycling conditions were as follows: an initial denaturation step at 94°C for 3 min, followed by 35 cycles consisting of denaturation at 94°C for 30 s, primer annealing at 47°C for 30 s, and extension at 72°C for 60 s, followed by a final extension step at 72°C for 5 min and a 4°C indefinite hold.

Cytochrome C Oxidase Subunit I

This primer set included: A2442 (5'-GCTAATCATCTAAAAATTTTAATTCCTGTTGG-3') and S1718 (5'-GGAGGATTTGGAAATTGATTAGTTCC-3") (19). The PCR cycling conditions were as follows: an initial denaturation step at 94°C for 3 min, followed by 35 cycles consisting of denaturation at 94°C for 60 s, primer annealing at 46°C for 80 s, and extension at 72°C for 90 s, followed by a final extension step at 72°C for 5 min and a 4°C indefinite hold.

T. cruzi PCR Amplification

The loci amplified for analysis include an 875 bp region of the mismatch repair gene class 2 (*msh*2), a 1266 bp region partially containing the kinetoplast maxicircle-encoded genes, cytochrome oxidase subunit II (COII) and NADH dehydrogenase subunit I (ND1) (Figure 3), and a 727 bp region of the 1359 bp conserved hypothetical protein gene located on chromosome 6 (Tc00.1047053506529.310) (20).

Mismatch Repair Gene Class 2

The primers used for this region were: Tmuts 30 (5'-GACGAACTGATGGAACTGGA-3') and Tmuts 41 (5'-CAAACCAAACCCATCGTAAG-3') (21). The PCR cycling conditions were as follows: an initial denaturation step at 94°C for 3 min, followed by 30 cycles consisting of denaturation at 94°C for 30 s, primer annealing at 58°C for 60 s, and extension at 72°C for 2 min, followed by a final extension step at 72°C for 5 min and a 4°C indefinite hold.

Cytochrome Oxidase subunit II/NADH Dehydrogenase Subunit I

This region was amplified using the following primers: ND1.3A (5'-GCTACTARTTCACTTTCACATTC-3') and COII.2A (5'-GCATAAATCCATGTAAGACMCCACA-3') (22). In most cases, a third primer was needed to sequence the middle portion of the targeted area. This primer was COII.A400 (5'-CTCCTATTACAACCAATAAACATC-3') (22). The PCR cycling conditions were as follows: an initial denaturation step at 94°C for 3 min, followed by 37 cycles consisting of denaturation at 94°C for 30 s, primer annealing at 53°C for 60 s, and extension at 72°C for 2 min, followed by a final extension step at 72°C for 5 min and a 4°C indefinite hold.

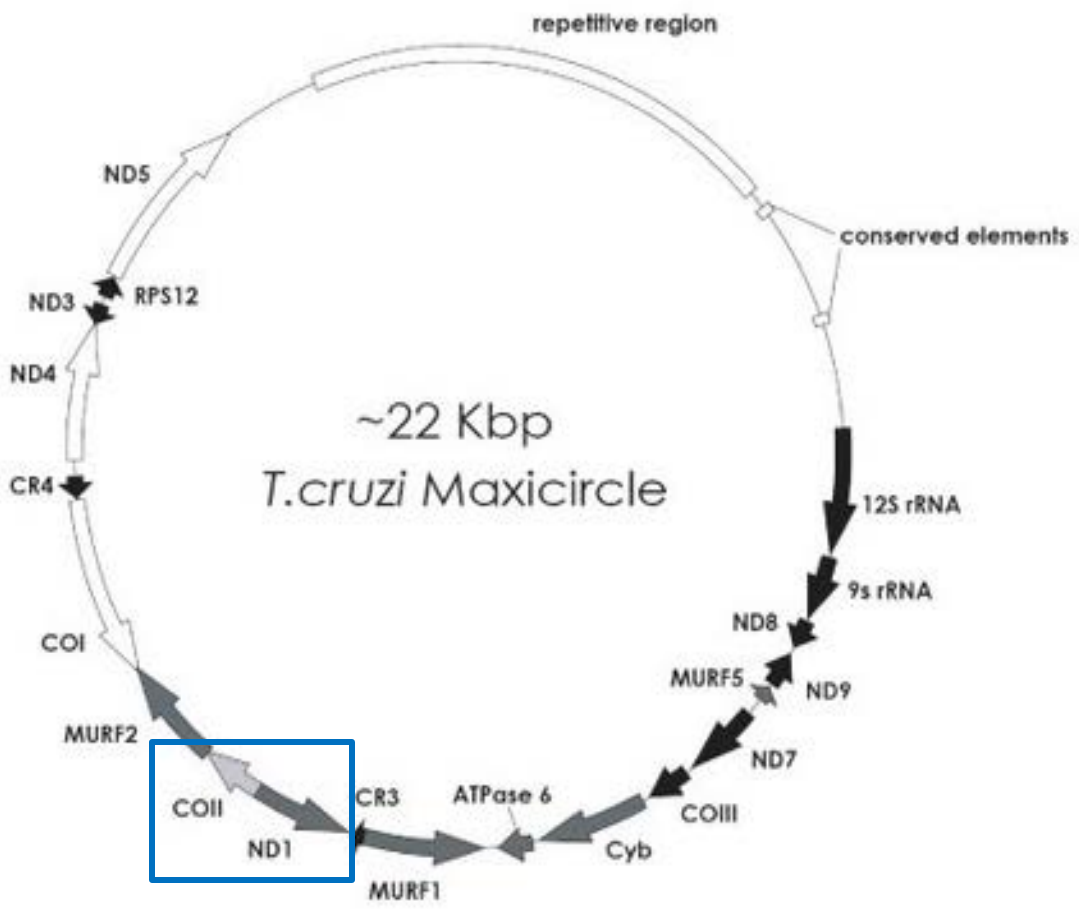

Figure 3: *T. cruzi* maxicircle DNA with COII and ND1 gene regions outlined in the blue box (23).

Hypothetical Protein Gene (Chromosome 6)

This primer set included: TcNucF (5'-TTTCTTTCAGGCTGCGATTTT-3') and TcNucR (5'-CGCTGTTTGGCTCATTTCTT-3') (20). The PCR cycling conditions were as follows: an initial denaturation step at 94°C for 3 min, followed by 34 cycles consisting of denaturation at 94°C for 30 s, primer annealing at 55°C for 30 s, and extension at 72°C for 60 s, followed by a final extension step at 72°C for 6 min and a 4°C indefinite hold.

PCR amplification was performed in 0.2 ml PCR tubes, using either a BioRad T100™ or BioRad C1000™ thermal cycler. For each sample, a standard PCR master mix solution was made which contained the following per 25 µl reaction volume: 5.75 µl molecular grade water, 2.5 µl 10X ThermoPol™ buffer (containing $MgCl_2$), 2.0 µl dNTP mix (2.5 mM each), 1 µl each of 5 µM forward and reverse primers, 2.5 µl 10X BSA, and 0.25 µl Taq DNA polymerase (5 U/µl). Ten µl of template DNA from each extracted sample was added to bring the total reaction volume to 25 µl.

The master mix solution was the same for all primer pairs except the COII-ND1 set, which was found to require an additional amount of $MgCl_2$ for optimal amplification. In the case of COII-ND1 reactions, a 10X buffer without magnesium was used and 1.5 µl of 25 mM $MgCl_2$ solution was added per 25 µl reaction volume. Water volume was decreased in accordance with the additional $MgCl_2$. In all cases, both a negative and positive control was run for each set of reactions per primer pair.

Gel Electrophoresis

Amplicons were separated electrophoretically using a 1.5% TAE agarose gel with 1X TAE buffer. Gels were stained with GelRed™ nucleic acid stain (Phenix, Candler, NC) and

reaction products were visualized under UV light with an EC3 transilluminator imaging system (UVP BioImaging Systems, Upland, CA).

Sequencing and Capillary Electrophoresis

Any amplification reaction that produced a band of equal size to the equivalent positive control (corresponding to the same approximate bp lengths of each loci listed above) were enzymatically treated to remove unincorporated dNTPs and primers by adding 4.0 µl of 1 U/µl ExoSAP-IT (Affymetrix, Santa Clara, CA). This was performed prior to using the products as template for cycle sequencing reactions. Products were incubated at 37°C for 15 min followed by 80°C for 15 min. Chain termination reactions were set up using BigDye® Terminator chemistry (Life Technologies, Carlsbad, CA), according to the following protocol, per 15 µl reaction volume: 6.0 µl molecular grade water, 5.0 µl BetterBuffer BigDye® dilution buffer (Gel Company, San Francisco, CA), 1.0 µl BigDye® v.3.1 Ready Reaction solution, 1.0 µl 5 µM primer (forward or reverse), and 2 µl PCR product. Thermal cycling parameters for the reactions were as follows: 96°C for 3 min, 25 cycles of 96°C for 15 s, 50°C for 10 s, and 60°C for 4 minutes. All products were stored at 4°C until use.

Following chain termination reactions, unincorporated primers were eliminated by spinning samples through Performa DTR Gel Filtration Columns (Edge Biosystems, Gathersburg, MD), according to the manufacturer's recommendations. Samples were then dried in a vacuum centrifuge at 45°C for approximately 15-25 minutes. The dried samples were resuspended in a 25.0 µl Hi-Di™ Formamide (Life Technologies, Carlsbad, CA) solution. Each sample was transferred to a 0.2 ml PCR tube and heat denatured at 95°C for 3 min. This was followed by immediate cooling on an ice block for 3 additional min. The BigDye chain

termination products were detected on a PRISM® 310 or 3130*xl* Genetic Analyzers (Applied Biosystems) through capillary electrophoresis. The sequences were analyzed using Sequencher v4.8 software (GeneCodes, Ann Arbor, MI) and aligned to published reference sequences obtained from the published GenBank database (http://blast.ncbi.nlm.nih.gov/).

Phylogenetic Analysis

Phylogenetic relationships were investigated among the groups of *T. cruzi* and *Triatoma* spp. samples that produced positive sequence results for all three respective genetic loci. This was performed using MEGA v5.1 software (Center for Evolutionary Medicine and Informatics, Tempe, AZ). Sequences were imported into MEGA and aligned by codons using Muscle. Maximum likelihood trees, using both concatenated and individual gene methods, were constructed with sample sequences that were aligned to published *T. cruzi* and *Triatoma* spp. reference sequences. A bootstrap test of 1000 replicates was set for each tree, and all models were set to disregard the 3rd (wobble) position of a codon (24, 25). Both the concatenated and individual trees for *Triatoma* spp. were created using the Tamura 3-parameter model. The concatenated *T. cruzi* tree used the Tamura 3-parameter model, and the individual *T. cruzi* trees were constructed with a Kimura 2-parameter model. The selected models were recommended by MEGA based on test performances of each possible model for the specific data sets. Published outgroup sequences were obtained from GenBank and included for each locus represented on a tree. For all *Triatoma* loci, *Rhodnius prolixus* reference sequences represented the outgroups. For the *T. cruzi* loci, corresponding *Trypanosoma brucei* sequences were used as outgroups for the COII-ND1 and *msh2* loci. There was not an available *T. brucei* reference sequence for the hypothetical protein locus, so a known sequence for *Trypanosoma cruzi marinkellei* (a distant bat strain) was used as an outgroup.

Geospatial Analysis

Biogeography of *Triatoma* spp. and *T. cruzi* genotype was examined using species distribution models in order to extrapolate potential ecological risk of Chagas disease based on known presence of both the pathogen and the vector (26, 27). This was performed using a maximum entropy algorithm created by MaxEnt v3.3.3k software, to combine geographic location (input as GPS coordinates) of each species of triatomine collected as well as the collection location of each triatomine found to be infected with *T. cruzi*, with certain environmental parameters, in order to quantify probability of spatial distribution and exposure based on ecological suitability in a specific defined geographical area (Figure 4), (10, 26). The environmental parameter data consisted of an altitude layer and 15 Bioclim layers downloaded from the global climate database, WorldClim, with a 30 arc-second (~1 km) resolution (www.worldclim.org) (10, 28).The Bioclim layers used are listed in Table 1. There are 19 Bioclim layers available; however, only 15 were used in this study due to discontinuities or artifacts present in the layer data for Texas that were potentially created during the construction of the layers (10). The Bioclim variables not included were the mean temperatures of the wettest, driest, warmest, and coldest quarters in Texas. Maxent settings were programmed according to published recommendations (29, 30), with a bootstrap of 100 replications and a test:training ratio of 40:60, which allows 60% of the data to be used in creating the models and 40% to be used in testing them (10, 31). Probability of the presence of a particular *Triatoma* spp. or *T. cruzi*-infected triatomine was represented by Maxent's relative suitability output. This is a value between 0 and 1, calculated for each model that indicates the occurrence probability for each cell in the region of interest (10).

The models were visualized using ArcGIS software. The MaxEnt output data was imported into ArcMap v10.1, and converted into the necessary raster file format. The ArcMap spatial analyst masking tool was used to confine the map data to the geographic region of Texas. This was performed by applying a Texas shapefile obtained from a GIS reference database (http://www.landsat.com/texas-free-gis-data.html) that input into ArcMap the geographic coordinates making up the Texas state border.

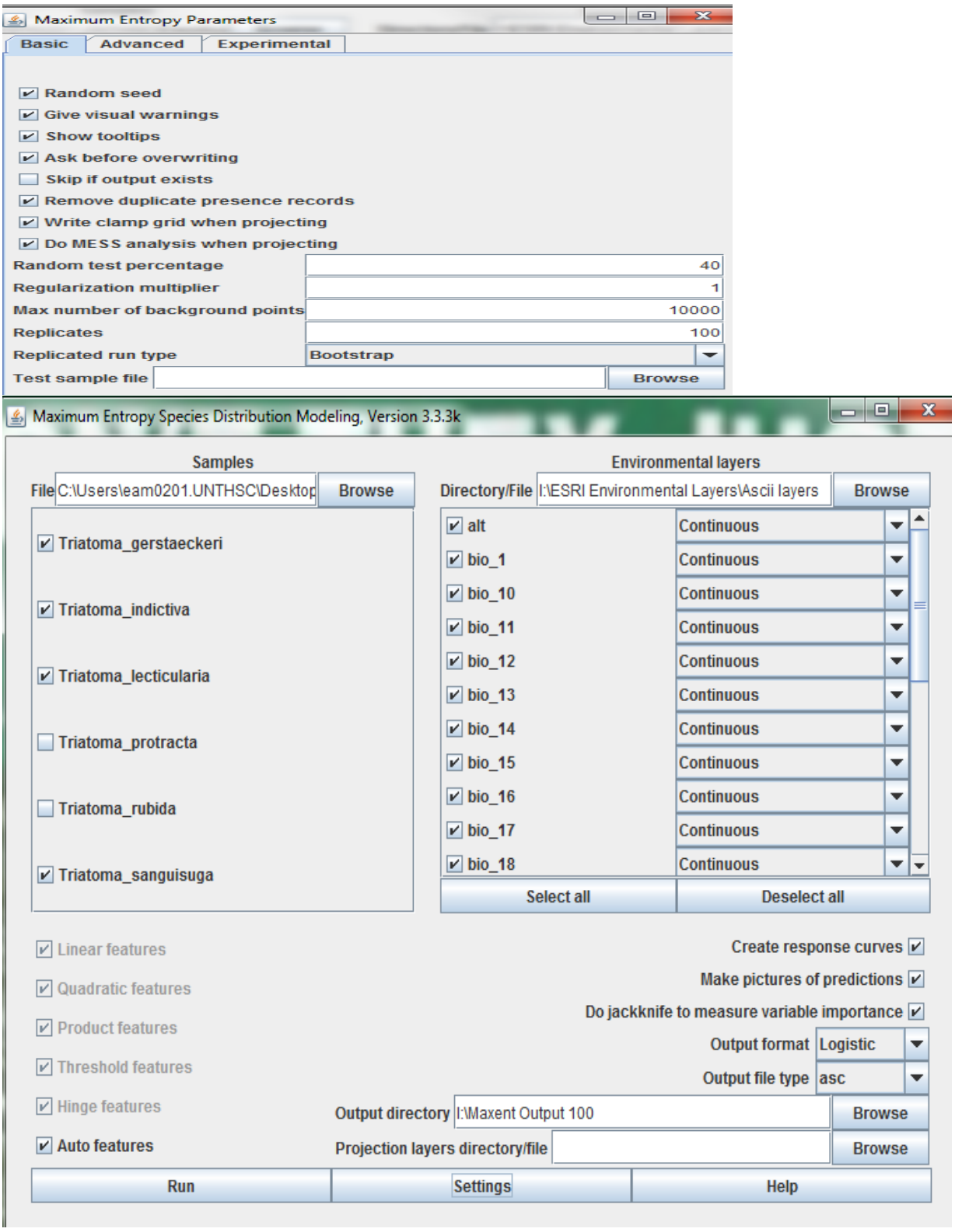

Figure 4: Example of MaxEnt v3.3.3k parameter settings

Environmental Parameters	
Bio 1	Annual Mean Temperature
Bio 2	Mean Diurnal Range (Mean of monthly (max temp – min temp))
Bio 3	Isothermality
Bio 4	Temperature Seasonality (standard deviation*100)
Bio 5	Max Temp of Warmest Month
Bio 6	Min Temp of Coldest Month
Bio 7	Temperature Annual Range
Bio 12	Annual Precipitation
Bio 13	Precipitation of Wettest Month
Bio 14	Precipitation of Driest Month
Bio 15	Precipitation of Seasonality
Bio 16	Precipitation of Wettest Quarter
Bio 17	Precipitation of Driest Quarter
Bio 18	Precipitation of Warmest Quarter
Bio 19	Precipitation of Coldest Quarter

Table 1: Bioclim variables used as environmental parameters for distribution modeling.

CHAPTER III

RESULTS AND DISCUSSION

Triatoma spp. DNA Amplification and Sequencing

DNA was extracted from all *Triatoma* specimens morphologically identified as non-*T. gerstaeckeri* by the Sarkar laboratory (n=43) and approximately 25% (n=60) of *T. gerstaeckeri* specimens. Only a subset of *the T. gerstaeckeri* specimens were tested since the species is the most abundant in Texas and morphological identification is straight-forward. The subset of 60 was chosen randomly to confirm that all the specimens identified as *T. gerstaeckeri* were genetically the same. All 3 gene markers, *16S rDNA*, *cytB*, and COI, were amplified in 41 of the *Triatoma* DNA samples. Only two of the loci amplified in 21 of the samples, and only 1 locus amplified in 42 of the samples. A summary of *Triatoma* DNA testing can be found in Table 2. PCR products were obtained from 100% of the *Triatoma* DNA extracts tested. Products were sequenced, and the results were compared to sequences available in GenBank. However, informative comparisons were restricted by the limited amount of published sequence data for the various *Triatoma* spp. at the time of this analysis. The insufficient amount of available reference data was the basis for sequencing the *Triatoma* DNA, as it was necessary to not only confirm the morphological identification for the completion of this analysis, but also to establish data that could be used for future reference. Sequences obtained from this study will be submitted to GenBank in an effort to provide a more standardized set of data for analysis of *Triatoma* spp for future research.

Several *Triatoma* spp. commonly found in Texas and the US were not represented in GenBank at all, including *T. gerstaeckeri* and *T. indictiva* (since the completion of this study, 3

cytB isolates identified as *T. gerstaeckeri* and originating from Mexico were published March 2013 by Espinoza, *et al* (32). For both *T. sanguisuga* and *T. lecticularia*, *cytB* and 16S rDNA reference sequences were available, but there was no data for COI. If no data was present in GenBank for a particular species or gene fragment, the query sequence results provide alignments to the most similar data available. In the case of COI, most samples partially aligned to *Triatoma dimidiata* or *Panstrongylus megistus* (Reduviids found outside the US and known vectors of *T. cruzi*) (33) with maximum identity values all below 80%. The one exception for this locus was sequence data for *T. rubida,* and the one *T. rubida* sample in this study provided a maximum identity value of 98%. All *T. sanguisuga* and all but one *T. lecticularia* sample(s) aligned to *cytB* and 16S rDNA reference sequences originating from the same respective species. The samples from the one *T. protracta* specimen compared to published *T. protracta* reference data as expected. All six *T. indictiva cytB* and 16S rDNA samples aligned with *T. sanguisuga* with maximum identity values of 96-99% and an average (E) value score of 1e-169.

The genetic similarities between *T. sanguisuga* and *T. indictiva* suggest that the current taxonomy needs to be revisited and potentially revised. The species are presently separated based on reported differences in morphology. *T. sanguisuga* generally tends to measure 20 mm or longer in length (16). However, this is known to vary based on geography. *T. sanguisuga* specimens in south Texas and south Florida are more likely to be 16-19 mm in length (Figure 5) (16). *T. indictiva* usually ranges from 18-22 mm in length (16). Other differences include color patterns. The pronotum (dorsal surface of the first thoracic segment) of *T. indictiva* is all black, and the dorsal surface of the connexium (flattened margin surrounding the abdomen) has distinctly narrow, reddish bands on each posterior segment (Figure 6) (16). In contrast, the pronotum of *T. sanguisuga* is dark with a red or yellow band extending from the collar, along the

lateral sides of the pronotum, down to regions of the posterior border, and the dorsal connexium contains wide orange-red to yellow bands that range in size from ¼ to ½ of each segment (Figure 7) (16).

In 1944, it was suggested that *T. indictiva* was a possible subspecies of *T. sanguisuga*, but the distinctions in morphology have precluded any integration of the two species (16). The majority of *Triatoma* spp. can be differentiated by morphological features. However, there have been previous examples where two species appeared identical but were later found to be genetically divergent (*T. sordida*), in addition to cases where different *Triatoma* spp. that exhibited variant morphology were reclassified into one species based on similar genetic characterization (33). *T. infestans* is an example of the latter type of taxonomic revision, where initial separate speciation was due to color variations of specimens found in different regions of South America. However, once studies revealed that the "different" species could successfully reproduce with each other and genetic characterization proved them to be identical, the individual species were reclassified as one species with pronounced phenotypic variation (33).

In light of the similar sequence data generated by this study for *T. sanguisuga* and *T. indictiva*, further studies, including larger sample size, additional molecular markers, and interspecies reproduction capabilities, are needed to investigate the extent of those similarities, in order to further clarify whether taxonomic revision is appropriate.

	# Collected	16S rDNA Amplified	COI Amplified	*cytB* Amplified	Confirmed *T.cruzi* Positive (%)
T.gerstaeckeri	237	55	29	18	101 (43%)
T.lecticularia	11	9	9	6	6 (55%)
T.sanguisuga	23	18	18	17	6 (26%)
T.indictiva	6	6	6	6	2 (33%)
T.protracta	1	1	1	1	0 (0%)
T.rubida	1	1	1	1	0 (0%)
Nymphs	17	9	1	0	0 (0%)
Unknown	4	1	0	0	0 (0%)

Table 2: Summary of *Triatoma* spp. DNA amplification

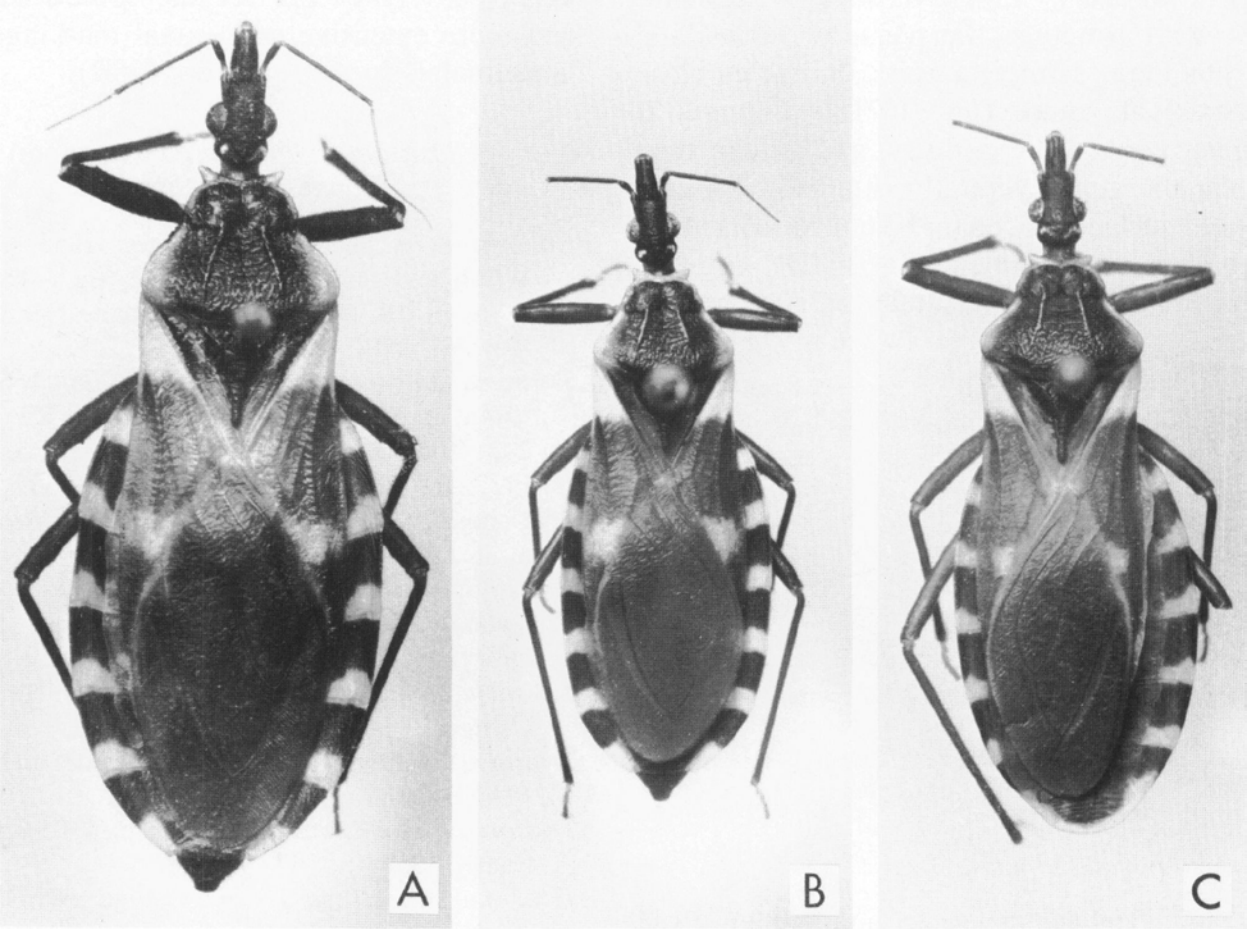

Figure 5: *Triatoma sanguisuga* indicating geographic size variation. A. Female from South Carolina. B. Female from Lake Placid, Florida. C. Male from Duval County, Texas. (16)

Figure 6: *Triatoma indictiva,* male from Arizona. (16)

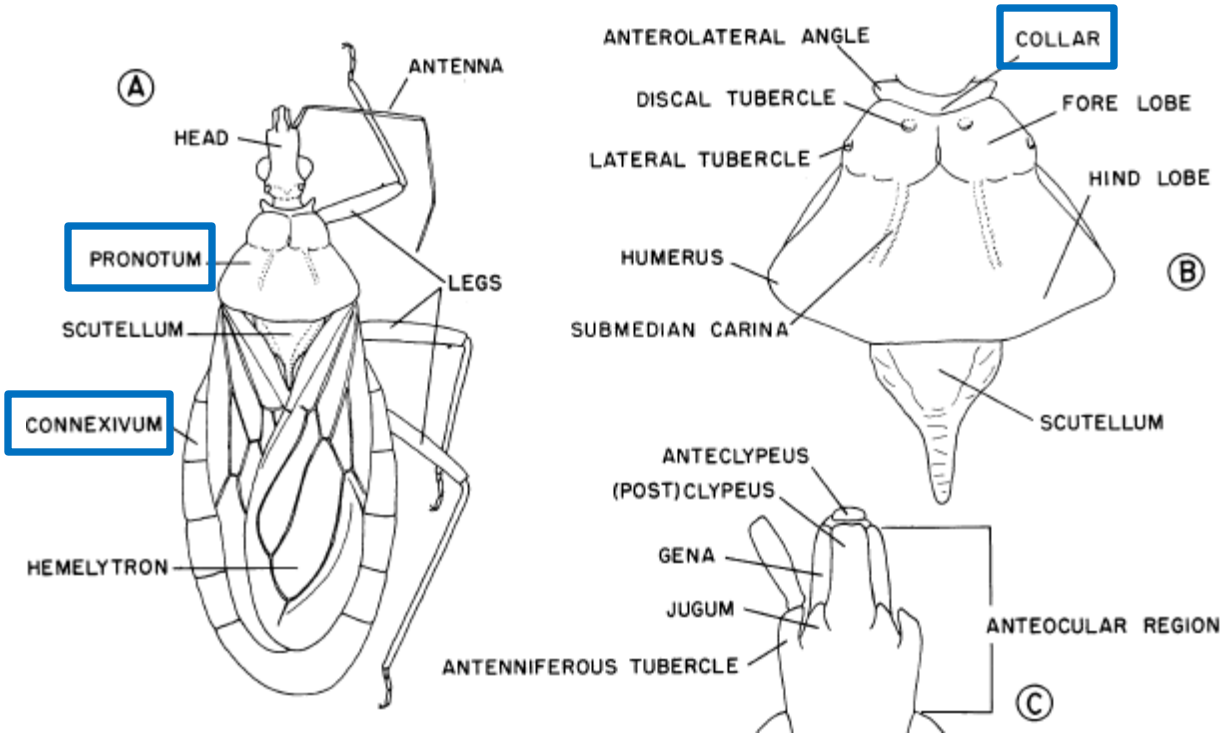

Figure 7: *Triatoma* spp. dorsal anatomy. A. Connexium and pronotum. B. Close up of pronotum. (16)

T. cruzi DNA Amplification and Sequencing

All 300 *Triatoma* specimens were tested for the presence of *T. cruzi* DNA through the targeting of 3 specific *T. cruzi* gene fragments: *msh2*, COII-ND1 and a hypothetical protein gene on chromosome 6 (Tc00.1047053506529.310) by PCR amplification using validated primer sets. One hundred fifty samples did not produce bands with any of the three loci's primer sets and were considered to be negative for *T. cruzi*. Thirty-five samples produced one or more bands that matched the equivalent band size to amplifications using known *T. cruzi* DNA. Faint bands did not consistently produce informative, high-quality sequence data. After this determination, the samples that produced faint bands were no longer included in downstream procedures to produce study data as a cost-saving measure. One hundred fifteen samples produced well-defined bands on a gel for one or more loci. Thus, gel results indicated that the proportion of *Triatoma* samples infected with *T. cruzi* could be as high as 50%. However, the bands that did not produce quality sequence data could not be verified as *T. cruzi* DNA and alternatively could have been non-specific amplification products. Therefore only 38% of triatomine specimens were confirmed by sequence data to be infected with *T. cruzi*. Subsequently, high yield and specific amplification products from 115 samples were sequenced and compared to known *T. cruzi* sequences in GenBank. Eighty-eight of the 115 samples, produced sequence data at all three *T. cruzi* loci, while 27 samples produced strong bands for one or two loci only. The *msh2* primer set produced more high-quality sequence results than either the COII-ND1 or TcNuc primer sets. This indicates the *msh2* primer binding sites might exhibit more uniformity between various *T. cruzi* isolates. The COII-ND1 primers produced the least number of informative sequence results, and this locus often required a third primer to obtain any sequence at all for the middle section of the amplification product. However, in some cases, the samples that did not

produce a COII-ND1 product did generate products with at least one of the primers sets for the other two loci.

The lack of successful amplification for this primer set could indicate more variability in the COII-ND1 priming regions, resulting in less effectiveness in amplifying the region. Moreover, the COII-ND1 primer set included both a forward and reverse primer with one degenerate position each, R (A/G) and M (A/C) respectively (34). This potentially contributed to the lower frequency of informative sequence data generated for this locus as well. In general, the more degenerate positions located on a primer, the higher the chance for more than one amplification product, which could lead to further problems downstream. The same degenerate primers were used to sequence the COII-ND1 amplification products, and this combined with more than one PCR product would result in a low primer concentration available for the correct template as well as potential for the additional, non-specific products to be sequenced, creating mixed sequence data. Low primer concentration could also result in shorter sequence reads due to the fact that less long extension products are generated prior to chain terminations. The non-informative COII-ND1 sequence data generated in this study often consisted of short, uninterpretable sequences or had electropherograms that appeared mixed with multiple peaks per base pair. The additional use of non-degenerate primer designed to prime the region in the middle of the amplicon did produce better quality data in many samples, but there were still a considerable number of samples that either gave poor quality sequences or sequences that were too short to substantiate identity.

Comparison of the *T. cruzi* sequence data with reference sequences in GenBank was not informative. Although there are a considerable number of sequences available for comparison, most contributors have not utilized a standardized set of loci with which phylogenetic analysis

was consistently performed. This study endeavors to improve upon this. Initially, many *T. cruzi* loci were tested, but the 3 selected loci were determined to produce the most consistent application, and therefore would present more consistent data for analysis from any sample.

Phylogenetic Analysis

In order to analyze phylogenetic relationships evenly between samples and gene markers, only those samples that produced high-quality sequence data with good depth of coverage for all three *Triatoma* or *T. cruzi* loci were utilized. Relationships among samples were evaluated for each individual gene by the construction of individual gene trees, as well as on a consensus level between individual samples by constructing concatenated trees. The individual approach produced similar trees to the consensus approach in most cases.

Phylogenetic Analysis of 16S rDNA, *cytB*, and COII-ND1 Genes

The consensus tree (Figure 8) reflected distinct genetic relationships between individual *Triatoma* spp. with the exception of one *T. lecticularia* sample and the *T. sanguisuga* and *T. indictiva* samples. All *T. sanguisuga* and *T. indictiva* samples were grouped together in one clade. The *T. gerstaeckeri* samples were also grouped together with some divergence indicated within the group, and the *T. lecticularia* samples were grouped together with the exception of one sample which was found to have a closer relationship to *T. gerstaeckeri*. Both the individual trees for the COI and 16S rDNA genes also reflected this same pattern and grouped all *T. gerstaeckeri* samples together, all *T. lecticularia* samples together and all *T. sanguisuga* and *T. indictiva* samples together into their own clade (Figures 9, 10). The individual tree representing *cytB* gene marker denoted slightly different relationships between samples than the previous individual trees or the consensus tree (Figure 11). This tree reflected a slightly more

homologous relationship between different species for some of the samples. This is not unexpected, considering the high level of conservation exhibited by *cytB* gene. Most of the *T. gerstaeckeri* and *T. lecticularia* samples grouped with the expected reference sequence, and all but one of the *T. sanguisuga* and *T. indictiva* samples grouped together as well; however, this tree produced a very distinct clade consisting of both *T. gerstaeckeri*, and *T. lecticularia* and one *T. indictiva* sample, in addition to 2 individual *T. gerstaeckeri* samples included with the *T. sanguisuga* and *T. indictiva* combined group. This could suggest that the *cytB* locus is more conserved between species than the other two loci. The bootstrap support values associated with the distinctive groupings, however, were all very low, making this particular tree difficult to interpret based on the sample size.

All three individual gene trees and the consensus tree signified that *T. sanguisuga* and *T. indictiva* are very closely related phylogenetically and gave further evidence that these two species are potentially the homogeneous genetically despite obvious phenotypic differences. In all cases, the outgroup *R. prolixus*, was shown to be a common ancestor but distant enough in relation to be a distinct outgroup.

Phylogenetic Analysis of *msh2*, COII-ND1, and Chromosome 6 Hypothetical Protein Genes

The *T. cruzi* consensus tree indicated that most samples were closely related to one another with only a small amount of variation (Figure 12). The consensus tree produced two distinct clades of *T. cruzi* samples with a smaller degree of distance revealed within the clades. The concatenated reference sequence represented the CL Brener strain, and it was grouped with the larger of the two clades. The smaller clade consisted of 9 sample sequences, and both clades produced a bootstrap value of 95, which denoted high reproducibility. The individual gene trees

were all organized differently from each other, with the COII-ND1 tree being almost identical in structure as the consensus tree with bootstrap support values of 100 for both clades (Figure 13). The tree representing the hypothetical protein gene showed virtually no phylogenetic divergence between samples, indicating all samples were either the same strain of *T. cruzi* or the particular gene is highly conserved between strains (Figure 14). The *msh2* tree was very similar in structure to the hypothetical protein tree (Figure 15). However, this tree exhibited a small subgroup of three sample sequences, more distantly related to the predominant clade. Bootstrap support values for this tree were low, however, with the exception of the small subgroup which produced a boostrap support of 100.

Considering the discrepancy between individual gene trees and the consensus tree, it is difficult to conclude with any certainty the phylogenetic relationships between the *T. cruzi* based on the individual loci sequences alone. However, this is not entirely unexpected taking into account the known heterogeneity among *T. cruzi* isolates. Also this study analyzed two nuclear gene markers and one mitochondrial marker. While nuclear genes are generally thought to evolve more slowly than mtDNA, the mitochondrial genes that code for the electron transport chain (including COII-ND1) are highly conserved as well (35). Additionally, *T. cruzi* is capable of both sexual and asexual reproduction (36, 37). *T. cruzi* population studies have estimated that clonal replication occurs at a much higher frequency than sexual reproduction (38). If accurate, this would indicate a low level of nuclear recombination which would suggest reduced variability between generations for nuclear DNA. These factors reveal the complex nature of phylogenetic analysis between individual *T. cruzi* isolates and demonstrate the necessity of further studies into the population biology and molecular genetics of *T. cruzi*.

The consensus tree concatenates all three gene sequences and analyzes each as a single sequence (39). It does not take into account significant differences in rates of evolution between genes. However, Gadagkar, *et al*, found that the concatenated approach to phylogenetic analysis generated a much more accurate tree compared to the individual gene approach. This is due in part to the longer sequence reads it creates for each sample which decreases variation in evolutionary distance estimation (40). Additionally, the concatenated approach allows for a higher resolution between internal branches due to the increased number of evolutionary alterations on those branches which diminishes the effects of slow rate of evolution, rapid divergence times, and short gene sequence size on small internal tree branches (40). The consensus tree and two of the three individual trees suggest some degree of heterogeneity in the phylogenetic relationships among *T. cruzi* isolates found in Texas.

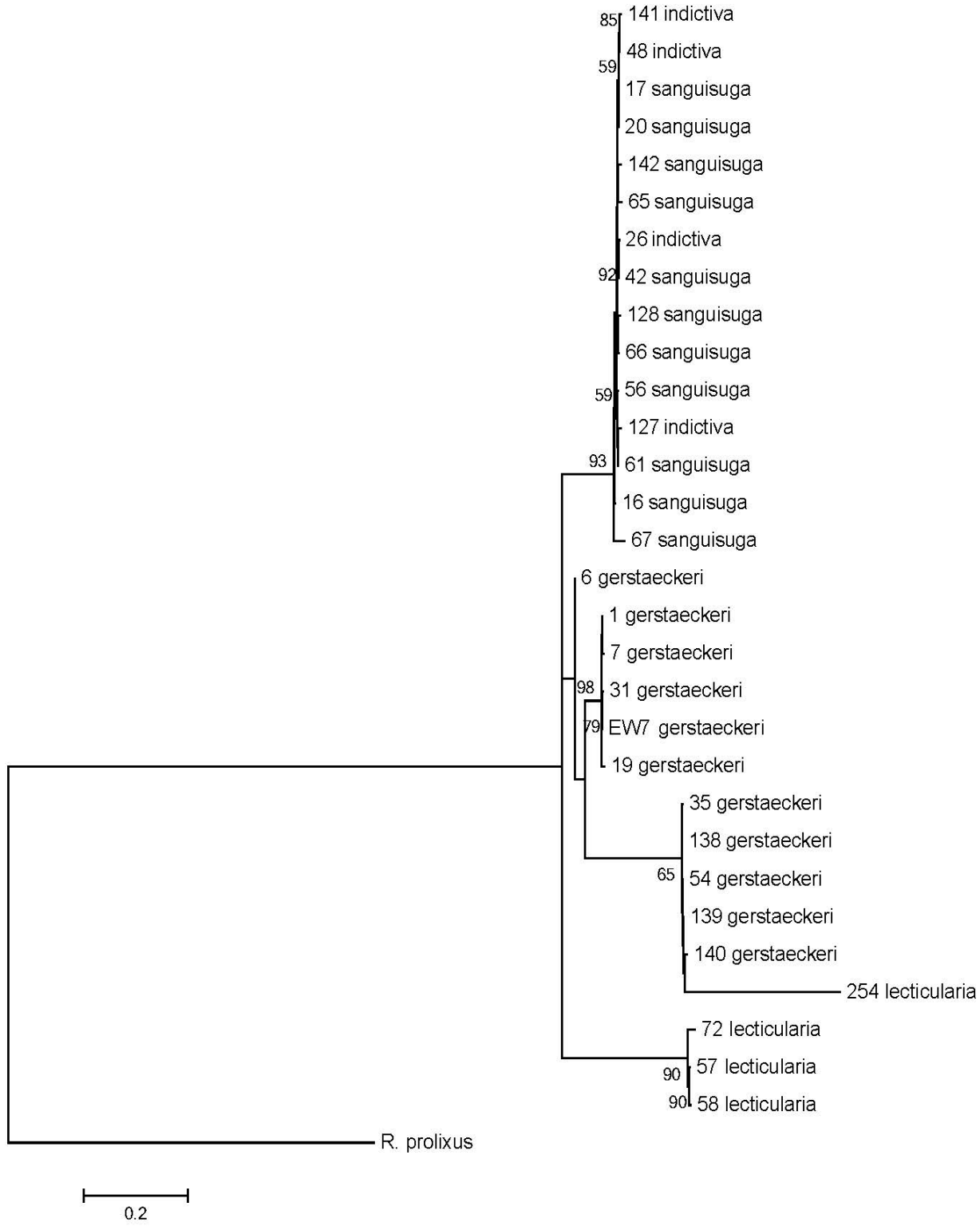

Figure 8: Concatenated maximum likelihood tree for *Triatoma* spp.

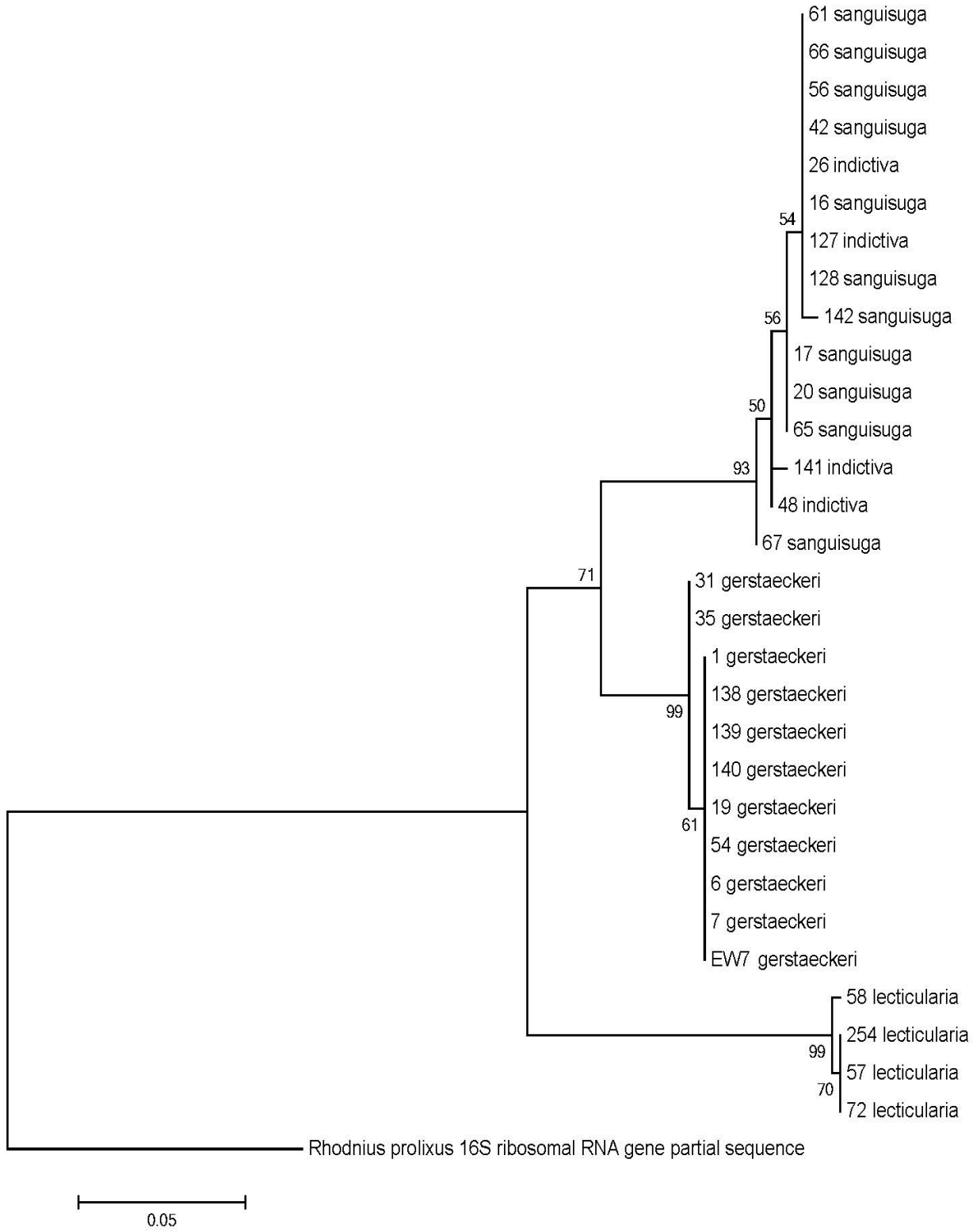

Figure 9: Maximum likelihood tree representing *Triatoma* spp. 16S rDNA locus.

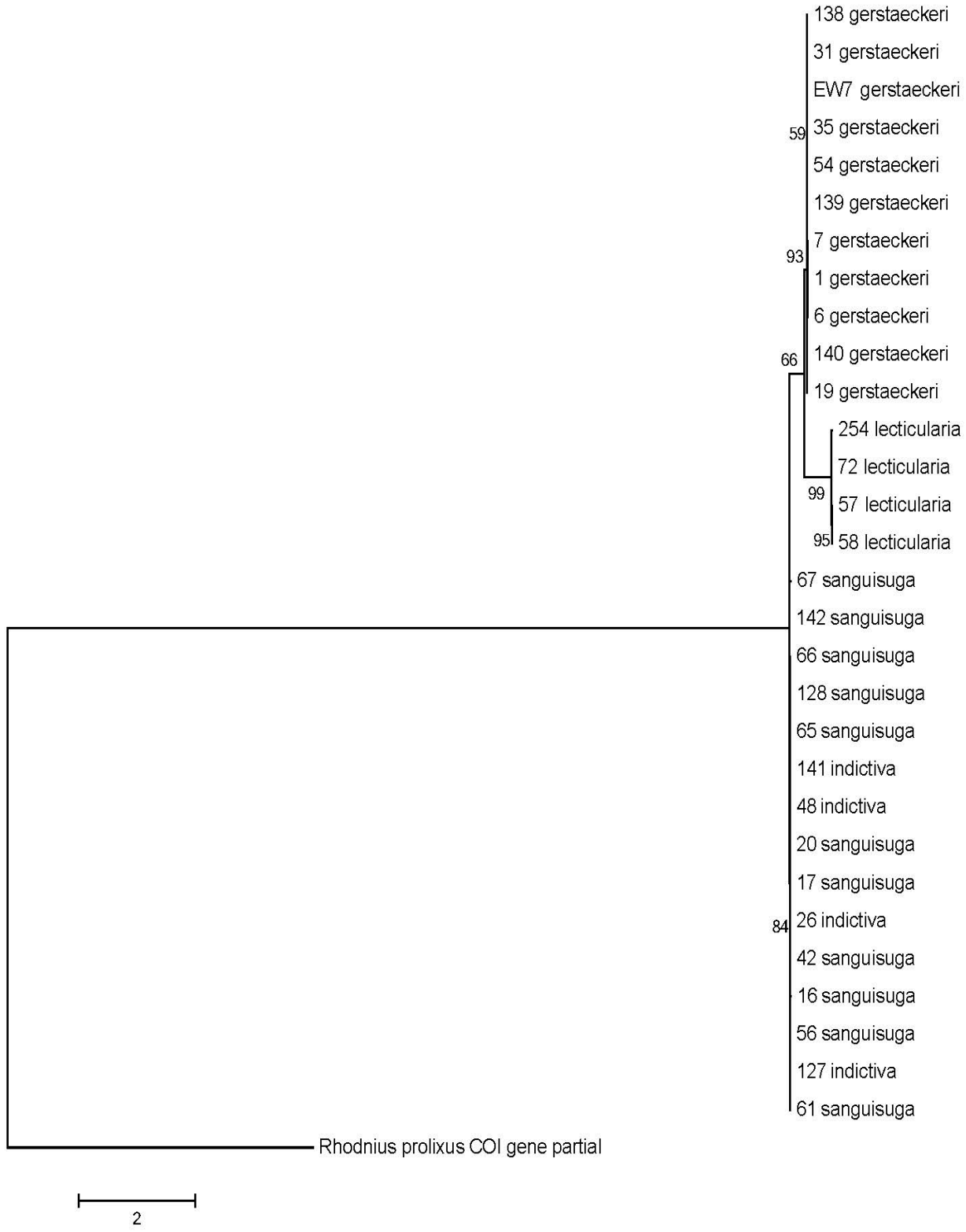

Figure 10: Maximum likelihood tree representing *Triatoma* spp. COI locus.

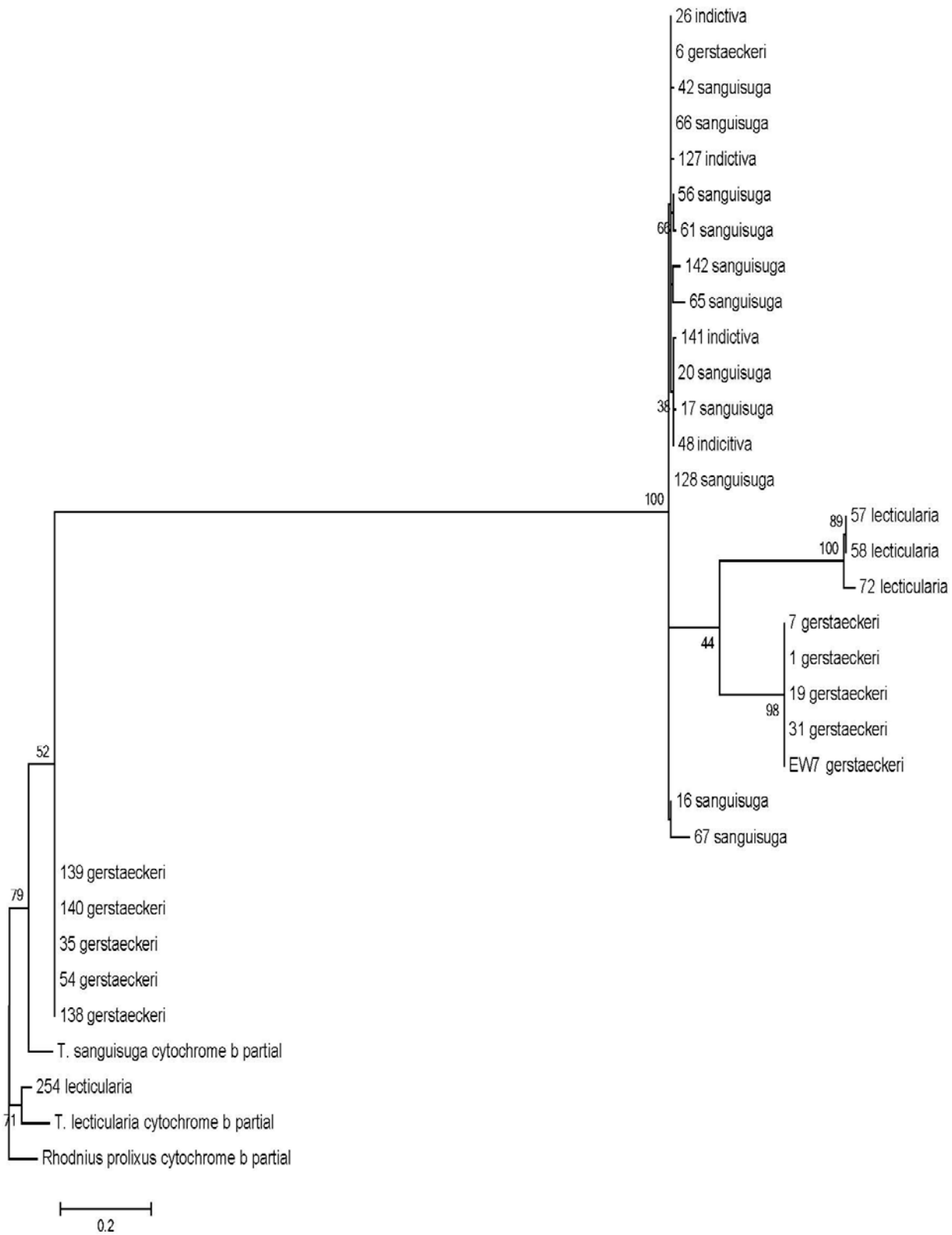

Figure 11: Maximum likelihood tree representing *Triatoma* spp. *cytB* locus.

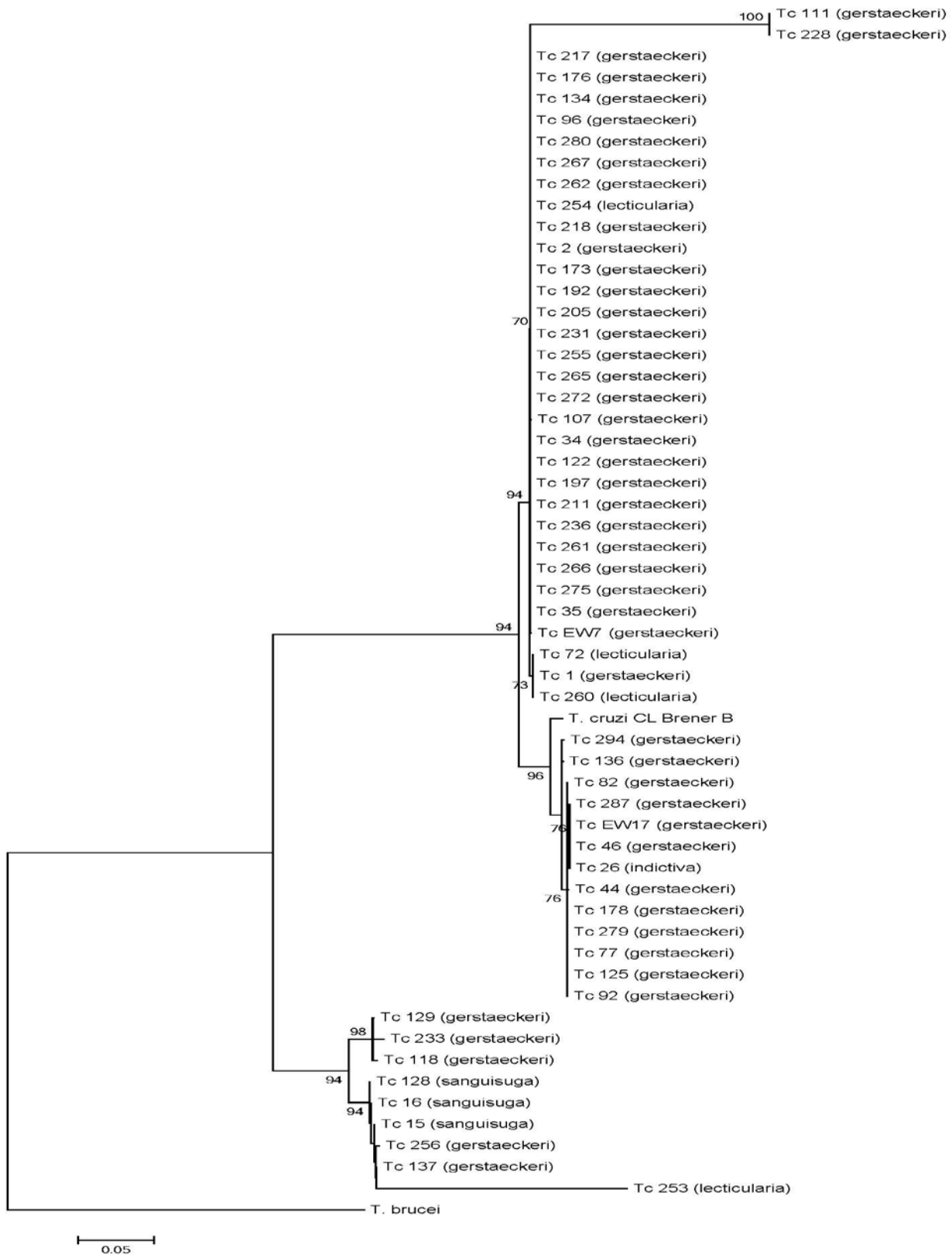

Figure 12: Concatenated maximum likelihood tree for *T. cruzi*.

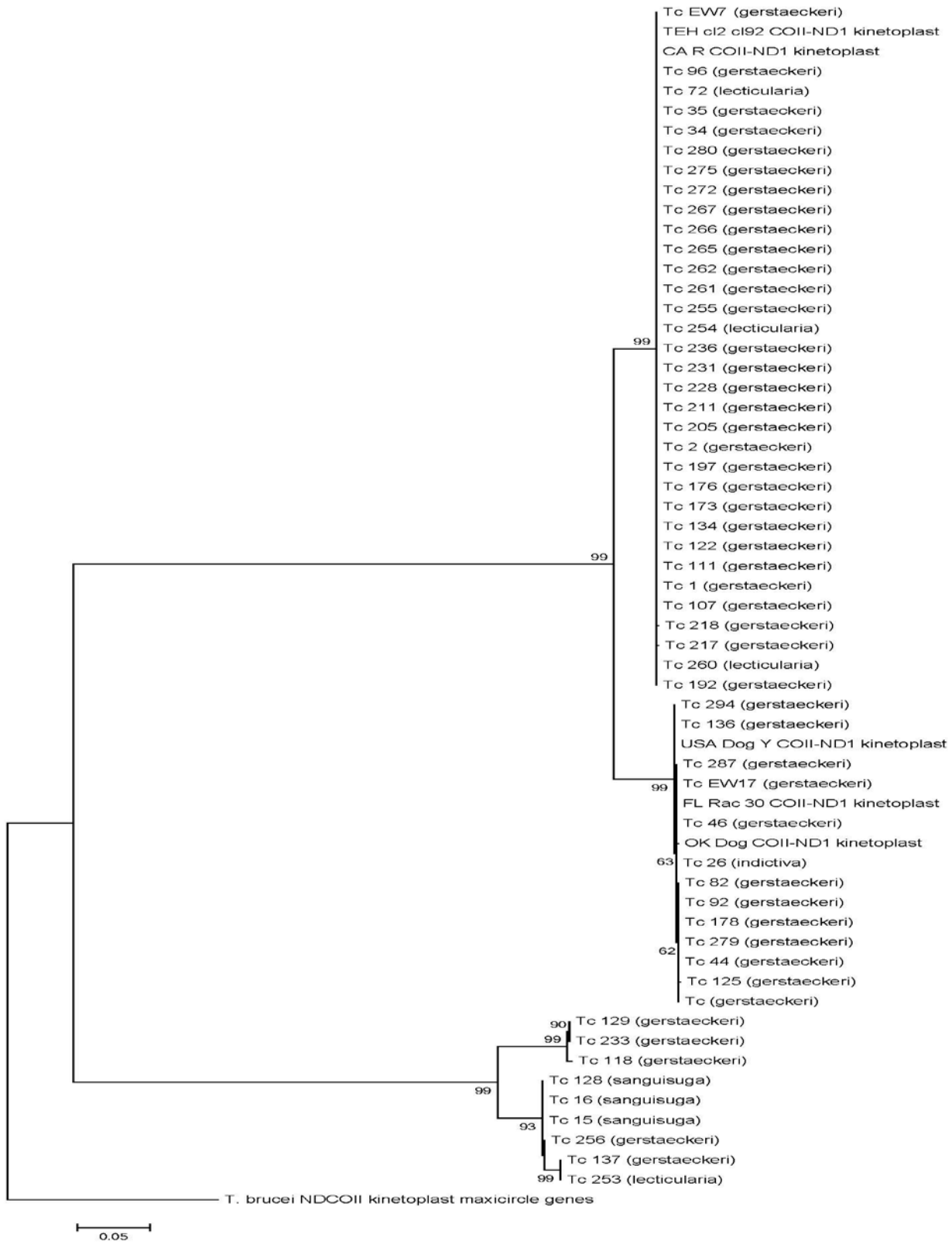

Figure 13: Maximum likelihood tree representing *T. cruzi* COII-ND1 locus. The species of *Triatoma* the sample originated from is noted in parenthesis.

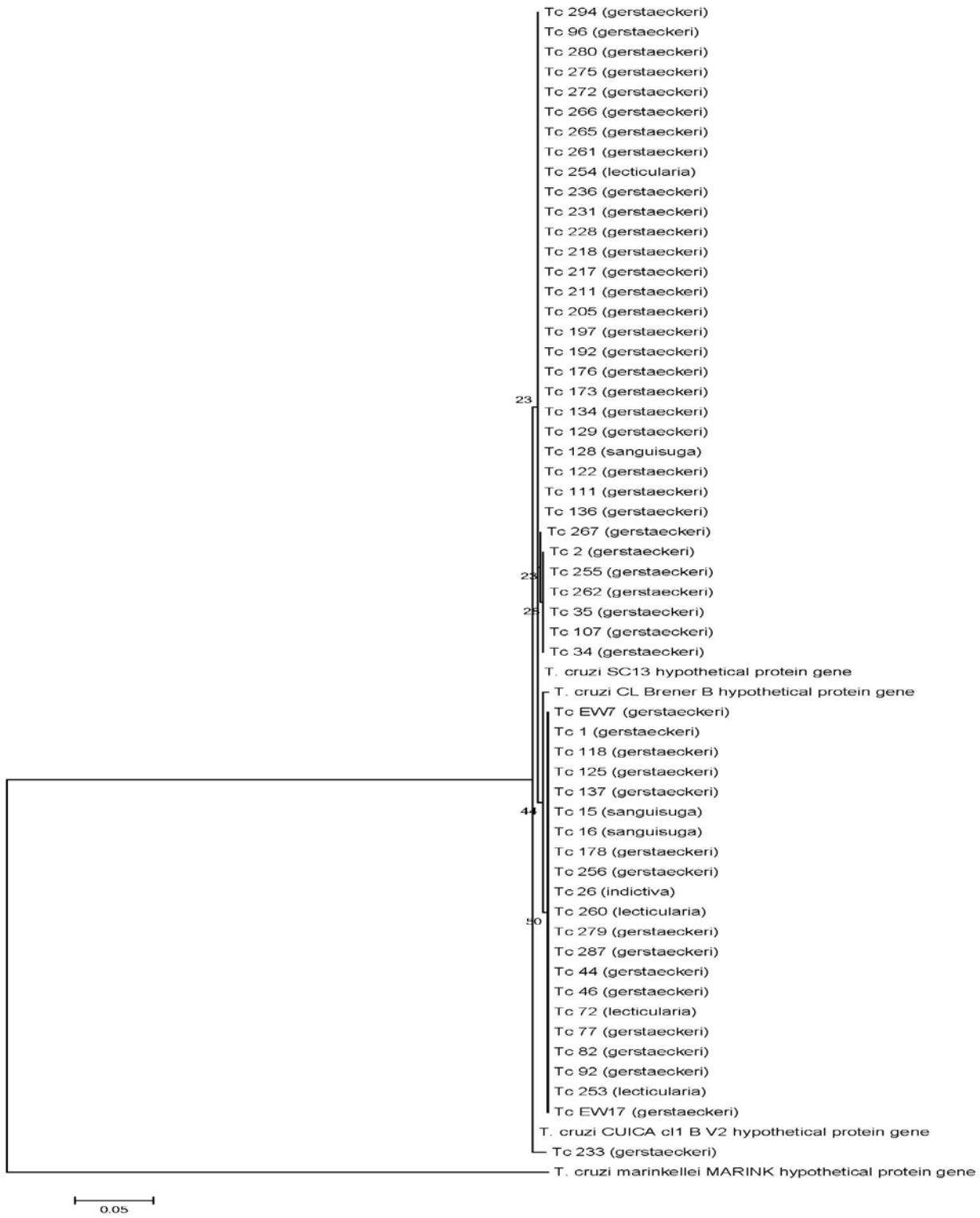

Figure 14: Maximum likelihood tree representing *T. cruzi* hypothetical protein gene locus. The species of *Triatoma* the sample originated from is noted in parenthesis.

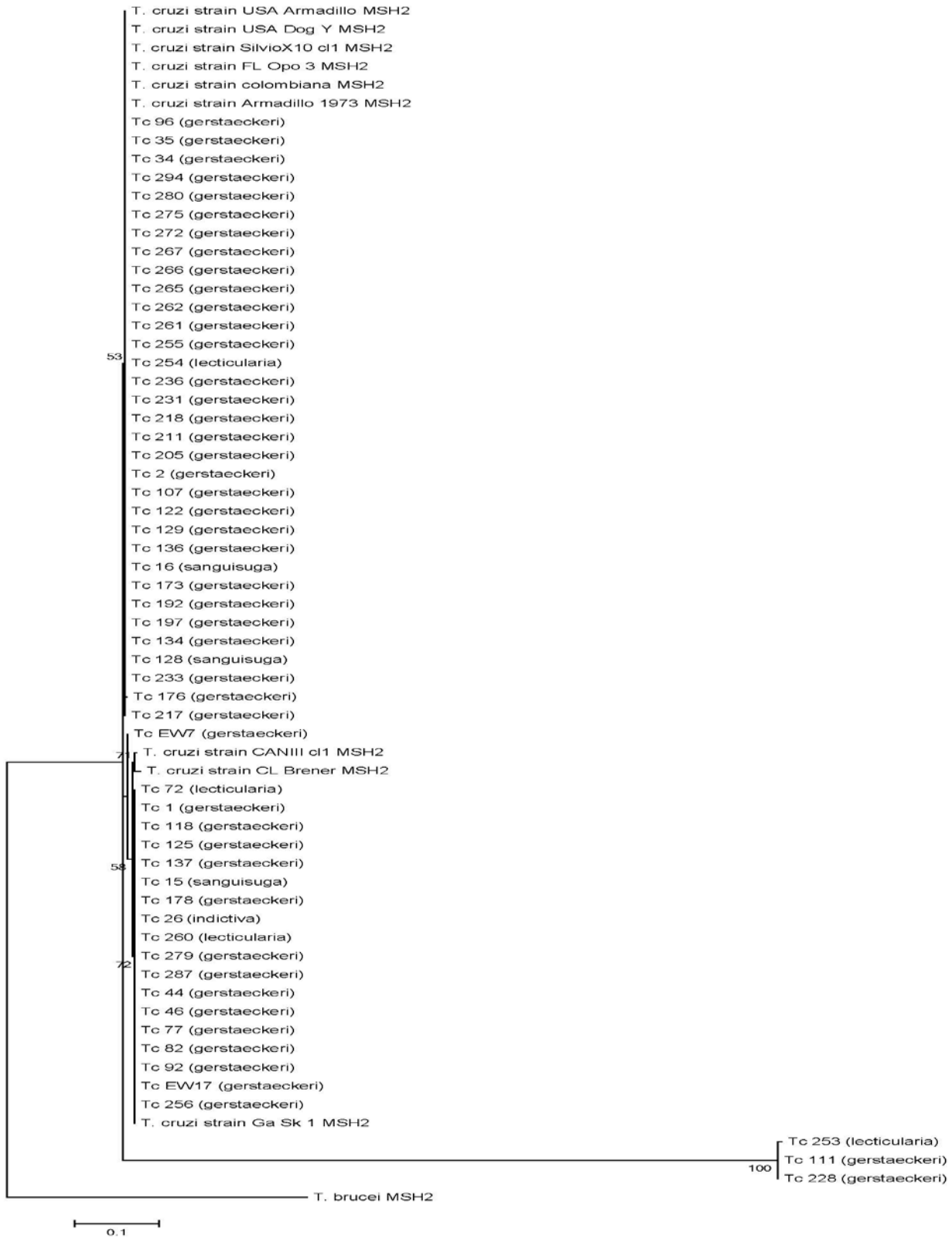

Figure 15: Maximum likelihood tree representing *T. cruzi msh2* locus. The species of *Triatoma* the sample originated from is noted in parenthesis.

Geospatial Distribution Modeling

GPS coordinates for all triatomine collection locations combined with environmental variable data from WorldClim were entered into MaxEnt to create species distribution models for *T. gerstaeckeri*, *T. sanguisuga*, *T. lecticularia*, and *T. indictiva*. Models were not created for *T. protracta* and *T. rubida* because each was only represented by a single specimen in the collection. An additional model was run for the distribution of *T. cruzi* infected triatomines based on the geographic collection site and the same environmental variables. The model included all *Triatoma* spp. infected with *T. cruzi* as a single group. Since the majority of specimens collected, and thus found to be infected with *T. cruzi*, were *T. gerstaeckeri*, the number of non-*gerstaeckeri* specimens infected with *T. cruzi* was too low to create reliable and informative models for each individual *Triatoma* spp (10). The Maximum Entropy method was chosen because it makes occurrence predictions relative to background environmental parameters (in this case, 15 Bioclim layers and altitude), in contrast to estimating distribution based on geographic presence only (41). The environmental layers were designed specifically to be recognized by ArcGIS software, which was utilized to generate maps reflecting the MaxEnt distribution probabilities.

Each *Triatoma* species distribution model produced similar yet distinct maps with individual occurrence probabilities for each cell on the map. Occurrence probabilities ranged between zero (no probability) and one (highest probability). The probability ranges were represented in ArcMap by a color wheel spanning from dark to light. Performance of each model was assessed using the average AUC values, or area under the receiver operating characteristic (ROC) curve. Each time MaxEnt runs a distribution model, it calculates a response curve and the corresponding AUC value. The AUC value represents how well the performance

of a particular model compares to another model. The closer the value is to one, the better the model functions in predicting reliable probability distributions (10). Probability distributions computed at random would have an AUC value of 0.5, so this number or below would imply less than optimal performance of the model (29). All AUC values for each *Triatoma* spp. model and the *T. cruzi* model were above 0.9 for each of the 100 bootstrap replication runs. This indicates performance of each model was optimal in its predictions. MaxEnt also determines which environmental variables are most important in influencing prediction probabilities in 2 ways. It evaluates how well a model performs using each variable by itself along with model performance using all variables except the one being tested to estimate how much the model performance decreases in the absence of that one variable (41).

Though each *Triatoma* spp. model produced somewhat overlapping presence probabilities, certain unique differences were evident. However, all models predicted a high distribution probability for south Texas. The model for *T. gerstaeckeri* indicated the highest distribution probability was concentrated in south Texas, extending up the middle of the state towards the north central border (Figure 16). The model representing the distribution of *T. cruzi* infected triatomines also followed a pattern that clearly paralleled the former model. The *T. cruzi* model was distinctly confined to the southern and middle sections of Texas, however, with very low probability shown in the far western and eastern sections and the panhandle (Figure 17). MaxEnt determined the environmental variable that increased model prediction performance the most when used in isolation to be Bioclim layer 5, which is maximum temperature of the warmest month. This indicates that the maximum temperature of the warmest month had the most influence on distribution probability if all other variables were not considered. The variable that decreased model performance the most when omitted was Bioclim

layer 3, or isothermality, which represents how even temperature is in a given area over a period of time. This suggests that though isothermality is not the most important variable on its own, it is the most significant when all other variables are utilized in estimating prediction probabilities. The models for *T. lecticularia, T. sanguisuga, and T. indictiva* showed a probability pattern similar to each other, where the highest occurrence predictions were predominantly concentrated in the eastern part of the state, ranging from the southeast coast to the far northeast/north central border (Figures 18, 19, 20).

The similarities between the distribution models for *T. sanguisuga* and *T. indictiva* correspond to results from the phylogenetic analyses previously discussed. If *T. sanguisuga* and *T. indictiva* have a very close phylogenetic relationship or are potentially equivalent genetically, as the maximum likelihood trees suggested, it could be presumed that they have similar or overlapping geographical or ecological ranges. However, since *T. sanguisuga* and *T. indictiva* appear to share genetics but not morphology, the notion of adaptation to divergent environments being a significant contributing factor to the morphological dissimilarity would be much less likely since their distribution overlaps.

The common pattern of highest probability being most concentrated in the south and central portions of Texas, extending out to the north and east supports findings of the 2010 Texas Chagas disease ecological risk analysis by Sarkar, *et al*, which was based on *Triatoma* occurrence data records in Texas over several decades (10). The relative ecological risk in that study was calculated at the county level using *Triatoma* presence probability estimations as described above (Figure 21) (10). Since the level of ecological risk in that study was based on the same methods for determining distribution probabilities and produced a comparable pattern

of results as this study, it is reasonable to infer the ecological risk might be similar when applied to *T. cruzi* presence in Texas.

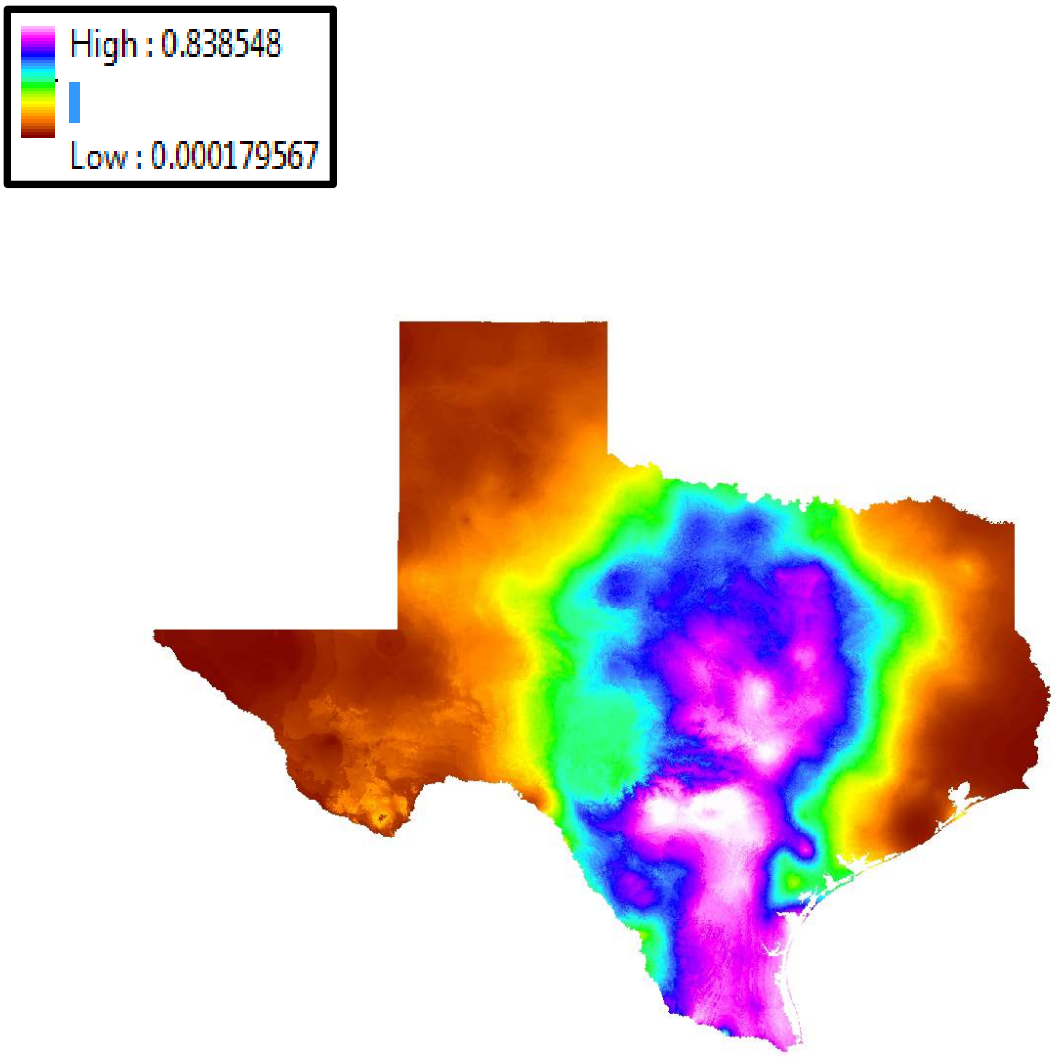

Figure 16: Species distribution map for *Triatoma gerstaeckeri*

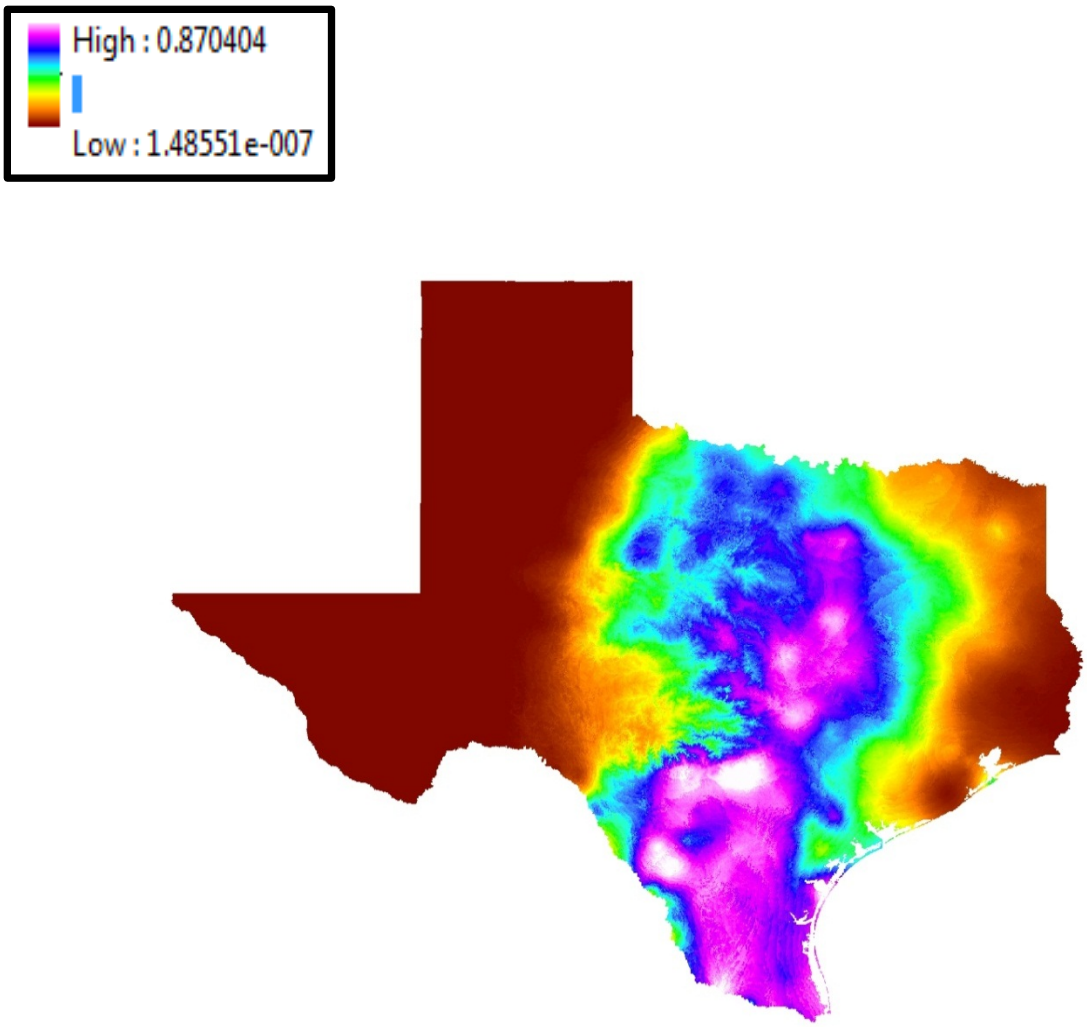

Figure 17: Probability distribution map for triatomine vectors infected with *T. cruzi*

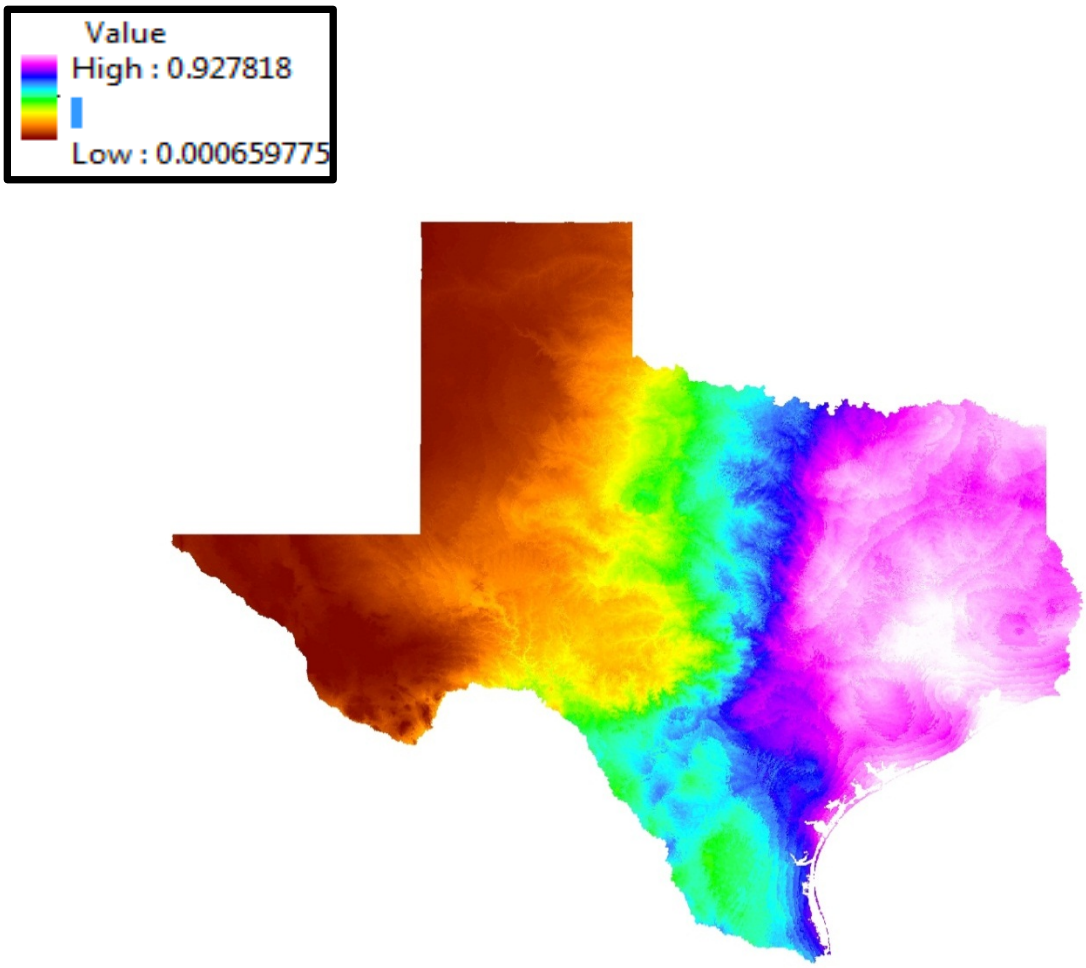

Figure 18: Species distribution map for *Triatoma lecticularia*

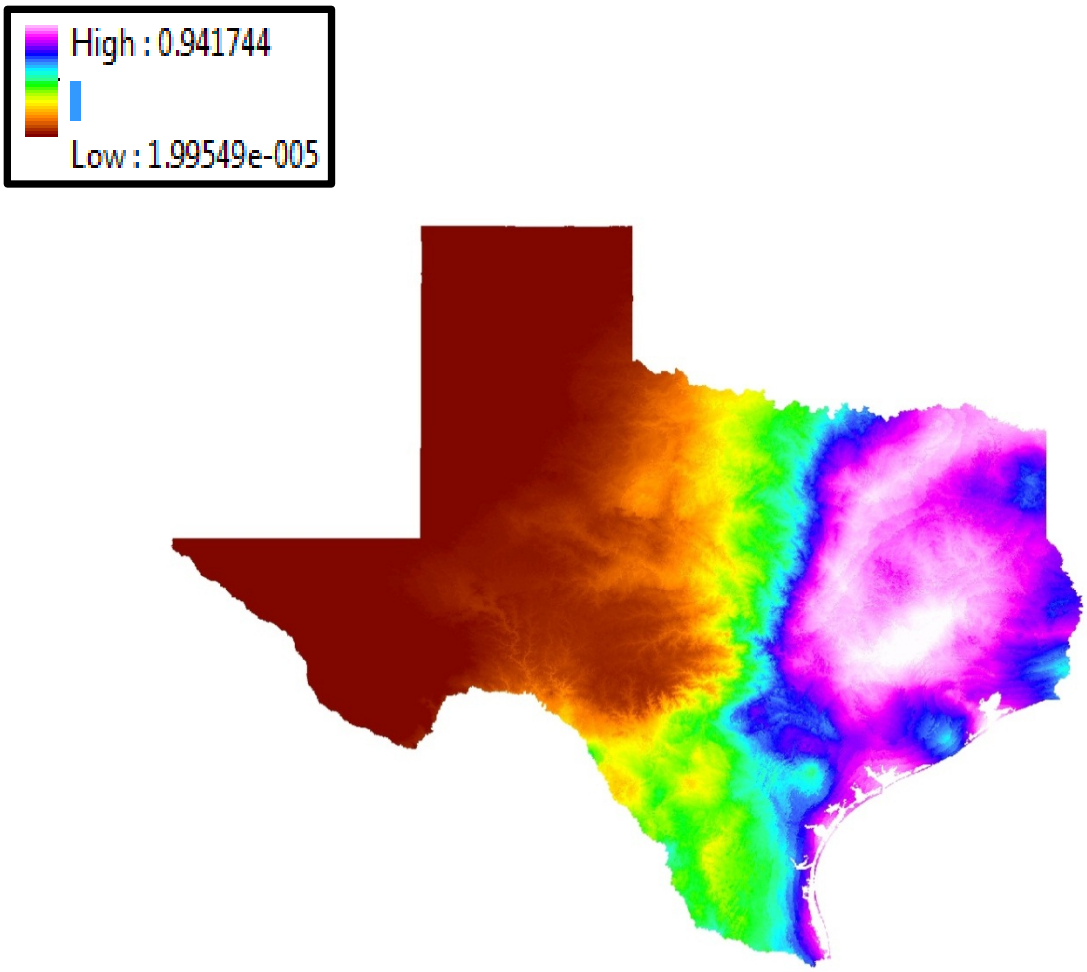

Figure 19: Species distribution map for *Triatoma sanguisuga*

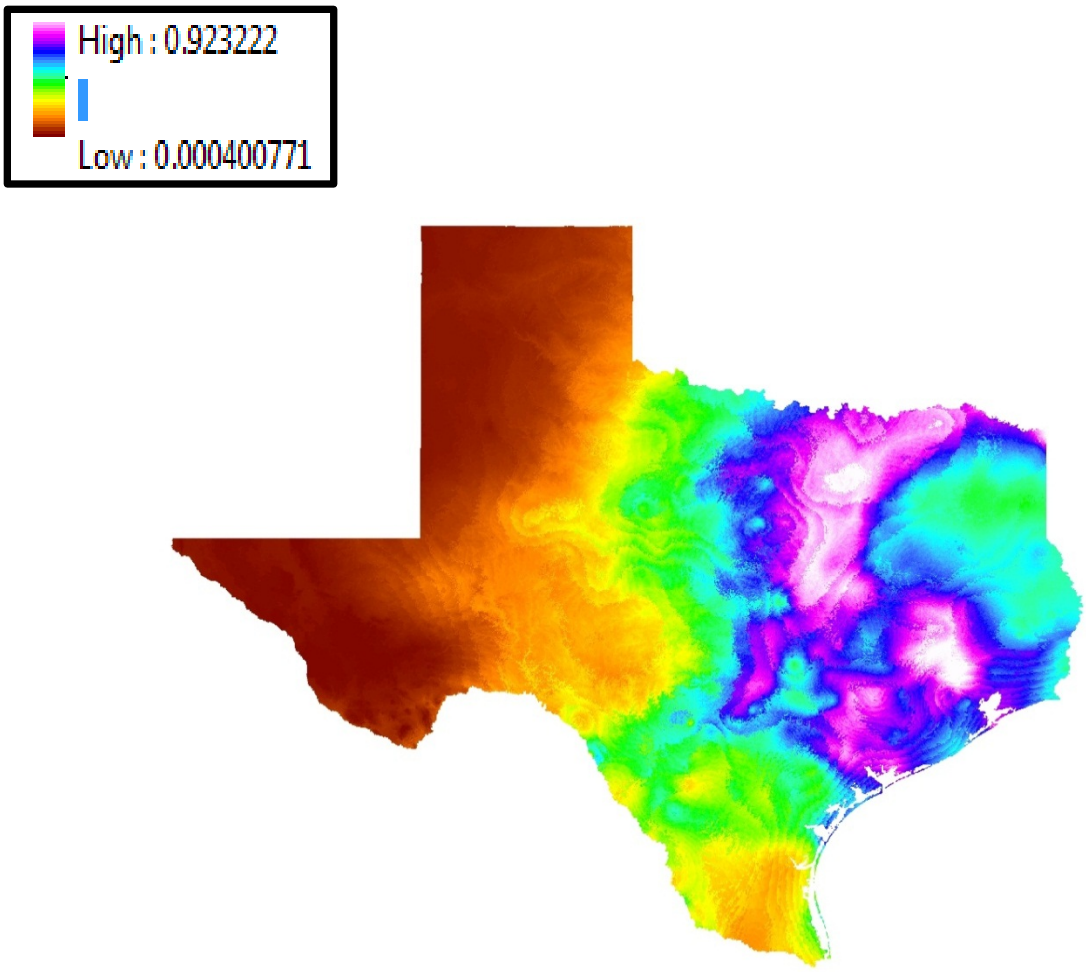

Figure 20: Species distribution map for *Triatoma indictiva*

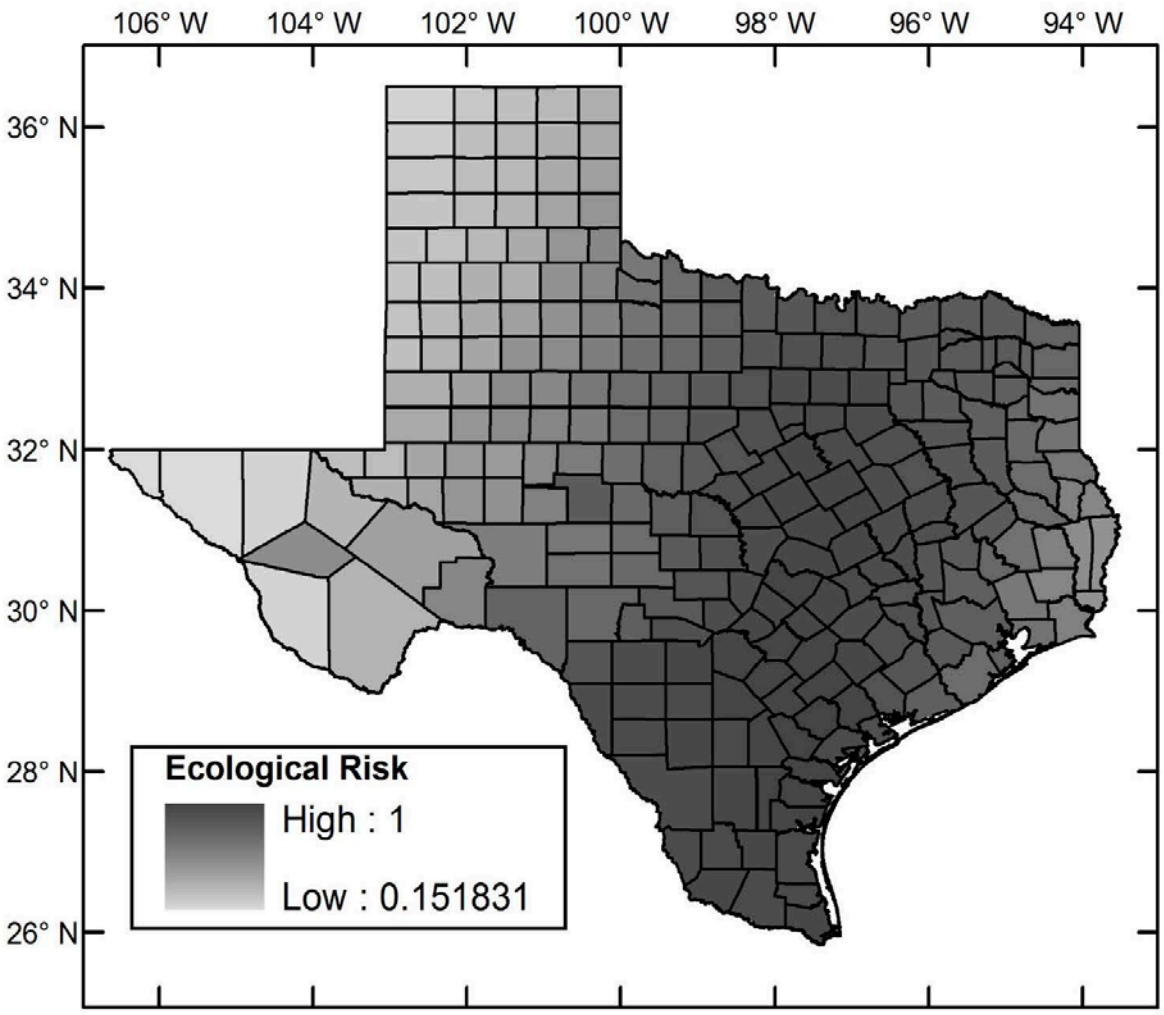

Figure 21: Ecological risk map for Chagas disease at the county level (10). Darker shaded areas indicate high ecological risk for Chagas while lighter areas indicate low risk.

CHAPTER IV

CONCLUSION

Several different *Triatoma* spp. have previously been reported in Texas, and records indicate the presence of *T. cruzi* found in wild mammals and domestic pets (8, 12, 42, 43). This study provided further confirmation that various *Triatoma* spp. can be found in multiple regions in Texas, and they are capable of carrying T. *cruzi*, with a conservative estimation of infection being 38%. DNA from specimens morphologically identified as belonging to six different *Triatoma* spp. was tested. Four *Triatoma* spp. were found to be carrying *T. cruzi* (all but *T. protracta* and *T. rubida*), but the specimen collection only included a small number of non-*T. gerstaeckeri* specimens. *T. gerstaeckeri* is by far the most common of all *Triatoma* spp. collected for this study.

There was some limitation in comparison with well documented *Triatoma* references due to the lack of pertinent sequence data in GenBank. However, for *T. lecticularia*, *T. sanguisuga*, *T. protracta*, and *T. rubida*, there were sufficient reference sequences accessible in GenBank, and the results of those comparisons suggested genotypic variability coincided with the phenotypic variation between each species. This suggested that the morphological identification of each of those samples was reliable.

In the case of *T. gerstaeckeri*, where sufficient reference data was not available but taxonomic keys are very distinct, each sample sequence aligned together. Sequences for *T. indictiva* aligned with the *T. sanguisuga* sample sequences as well as *T. sanguisuga* reference sequences in GenBank, signifying that the two species may be more genetically similar than evidenced by morphological comparison. This finding conveys the necessity of further studies

involving *T. sanguisuga* and *T. indictiva* to resolve the taxonomic issues. This finding also implies that standard morphological identification for *T. indictiva* may not be reliable and also provides some support for the revision of current taxonomic classification. The similarities between the *T. sanguisuga* and *T. indictiva* probability distributions provide further support of this hypothesis.

Phylogenetic analysis of *T. cruzi* isolates was also limited by the lack of well characterized comparators in GenBank and only a few references with all three loci used in this study, but analysis revealed that while there was some variation present among isolates, it only occurred in a small subset of samples. This finding supports the current general consensus found in the literature about the lack of variability among *T. cruzi* isolates found in the US (15).

Geospatial analysis was used to explore the possibility that variation between *Triatoma* spp. and among *T. cruzi* isolates can be correlated with geographic occurrence. The areas where triatomine bugs are most likely to occur were predominantly in south and central Texas, which supports previous studies (5, 10). The environmental variables determined to have the most importance when predicting occurrence were maximum temperature of the warmest month and isothermality, which signifies that areas of Texas distinguished by warm temperatures with minimal fluctuations are predicted to have the highest probability that *Triatoma* infected with *T. cruzi* can be found. This model was also concordant with previous risk analyses performed at the county level (10) and suggested further studies specifically in the high risk areas may provide further insight relative to the risk of Chagas transmission.

It was not possible to evaluate any association between distribution probability of a particular *Triatoma* species and a high frequency of testing positive for *T. cruzi* because the

sample collection included only a small number of non-*T. gerstaeckeri* samples. The low number of non-*T. gerstaeckeri* specimens also precluded the ability to run accurate distribution models for any individual species carrying *T. cruzi* other than *T. gerstaeckeri*. As a result, the *T. cruzi* model had to be run with all species combined and was just able to predict the probability for a *T. cruzi*-infected *Triatoma*, in general. This also meant it was not possible to evaluate if the presence of *T. cruzi* is dependent upon the particular *Triatoma* species or environmental variables which may affect the maintenance of *T. cruzi* in the enzootic cycle.

The analyses performed in this study provided some useful information on the distribution of *T. cruzi* and the infection of *Triatoma* spp. in Texas. However, in order to fully interpret the data and the results, there are still several issues that must still be addressed. The current *Triatoma* taxonomy needs to be revisited and potentially revised in accordance with the new genetic data described here. A larger sample size including additional areas distributed across the state to corroborate evidence of high genetic similarity between Texas *T. cruzi* isolates and to what extent those may cause veterinary or human disease. Additionally, studies utilizing animal models are necessary to investigate factors involving transmission of *T. cruzi* to mammalian hosts as well as competency of *Triatoma* vectors in Texas. Often, *T. cruzi* infections are reported in domestic animals, so a comparison of the infection locations indicated in the veterinary reports to the species distribution models predicting risk probability would be informative. Additionally, more information is needed regarding documented reports of Chagas in humans, so it could be determined if those cases arose predominantly in areas of high risk. These future studies are necessary to obtain more information on the status of Chagas disease in Texas and to be able to more accurately interpret the data that has already been obtained, so that

true epidemiological risk can be determined and the most appropriate and efficient prevention and control methods can be put in place.

REFERENCES

1. *Nature* 2010. Vol **465**, No. 7301 Suppl, S4-S5.

2. Noireau, F., Diosque, P., & Jansen, A. M. 2009. *Trypanosoma cruzi*: adaptation to its vectors and its hosts. *Vet Res*, **40(**2), 26-26.

3. WHO - Media Centre: Chagas Disease (American Trypanosomiasis) [Internet] [cited 12/1/2012]. Available from: http://www.who.int/mediacentre/factsheets/fs340/en/index.html

4. Bellini MF, Silistino-Souza R, Varella-Garcia M, Vilela de Azeredo-Oliveira MT, Silva AE. 2012. Biologic and Genetics Aspects of Chagas Disease at Endemic Areas. *J Trop Med*, 2012.

5. Bern C, Kjos S, Yabsley MJ, Montgomery SP. 2011. *Trypanosoma cruzi* and Chagas' Disease in the United States. *Clin Microbiol Rev* **24**(4): 655-681.

6. Manoel-Caetano FDS, Silva AE. 2007. Implications of genetic variability of *Trypanosoma cruzi* for the pathogenesis of Chagas disease. *Cad Saude Publica* **23**(10): 2263-2274.

7. Coura JR, Borges-Pereira J. 2012. Chagas disease. What is known and what should be improved: a systemic review. *Revista da Sociedade Brasileira de Medicina Tropical* **45**(3): 286-286.

8. Kjos SA, Snowden KF, Olson JK. 2009. Biogeography and *Trypanosoma cruzi* Infection Prevalence of Chagas Disease Vectors in Texas, USA. *Vector-Borne and Zoonotic Diseases* **9**(1): 41-49.

9. Teixeira, ARL, Gomes C, Lozzi SP, Hecht MM, de Cássia Rosa A, Monteiro PS, Bussacos AC, Nitz N, and McManus C. 2009. Environment, interactions between *Trypanosoma cruzi* and its host, and health. *Cadernos de Saúde Pública* **25**: S32-S44.

10. Sarkar S, Stavana SE, Frank DM, Rivaldi C-L, Sissel B, Sanchez-Cordero V. 2010. Chagas Disease Risk in Texas. *PLoS Negl Trop Dis* **4**(10): e836.

11. EPA – Western Ecology Division – Ecoregions of Texas [Internet] [cited 12/5/2012]. Available from: http://www.epa.gov/wed/pages/ecoregions/tx_eco.htm

12. Beard CB, Pye G, Steurer FJ, Rodriguez R, Campman R, Peterson AT, Ramsey J, Wirtz RA, Robinson LE. 2003. Chagas Disease in a Domestic Transmission Cycle in Southern Texas, USA. *Emerging Infectious Diseases* **9**(1): 103-105.

13. Piron M, Verges M, Munoz J, Casamitjana N, Sanz S, Maymo RM, Hernandez JM, Puig L, Portus M, Gascon J, Sauleda S. 2008. Seroprevalence of *Trypansoma cruzi* infection in at-risk blood donors in Catalonia (Spain). *Transfusion* **48**(9):1862-1868.

14. Brown EL, Roellig DM, Gompper ME, Monell RJ, Wenning KM, Gabriel MW, Yabsley MJ. 2010. Seroprevalence of *Trypanosoma cruzi* Among Eleven Potential Reservoir Species from Six States Across the Southern United States. *Vector-Borne and Zoonotic Diseases* **10**(8):757-763.

15. Roellig, D. M., Savage, M. Y., Fujita, A. W., Barnabé, C., Tibayrenc, M., Steurer, F. J., & Yabsley, M. J. 2013. Genetic Variation and Exchange in *Trypanosoma cruzi* Isolates from the United States. *PloS one*, **8**(2), e56198.

16. Lent H, Wygodzinsky P. 1979. Revision of the triatominae (Hemiptera, Reduviidae), and their significance as vectors of Chagas' disease. *Bulletin of the American Museum of Natural History* **163**(3): 123-520.

17. Lyman DF, Monteiro FA, Escalante AA, Cordon-Rosales C, Wesson DM, Dujardin JP, Beard CB. 1999. Mitochondrial DNA sequence variation among triatomine vectors of Chagas' disease. *Am J of Trop Med Hyg* **60**: 377-386.

18. Monteiro FA, Barrett TV, Fitzpatrick S, Cordon-Rosales C, Feliciangeli D, Beard CB. 2003. Molecular phylogeography of the Amazonian Chagas disease vectors *Rhodnius prolixus* and *R. robustus*. *Molecular Ecology* **12**(4): 997-1006.

19. Scataglini MA, Lanteri AA, Confalonieri VA. 2006. Diversity of boll weevil populations in South America: a phylogeographic approach. *Genetica* **126**(3): 353-368.

20. Flores-López 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**(8); e1272.

21. Augusto-Pinto L, Teixeira SM, Pena SD, Machado CR. 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**: 117-126.

22. Machado CA, Ayala FJ. 2001. Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi*. *Proceedings of the National Academy of Sciences* **98**(13): 7396-7401.

23. Westenberger, SJ., Cerqueira GC, El-Sayed NM, Zingales B, Campbell DA,Sturm NR. 2006. *Trypanosoma cruzi* mitochondrial maxicircles display species-and strain-specific variation and a conserved element in the non-coding region. *BMC Genomics*, **7**(1), 60.

24. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* **28**: 2731-2739.

25. Hall, BG. 2013. Building Phylogenetic Trees from Molecular Data with MEGA. *Mol Biol Evol.* 2013.

26. Illoldi-Rangel P, Rivaldi C-L, Sissel B, Fryxell RT, Gordillo-Perez G, Rodriguez-Moreno A, Williamson P, Montiel-Parra G, Sanchez-Cordero V, Sarkar S. 2012. Species

Distribution Models and Ecological Suitability Analysis for Potential Tick Vectors of Lyme Disease in Mexico. *J Trop Med*. 2012.

27. Kidd, DM, Ritchie, MG. 2006. Phylogeographic information systems: putting the geography into phylogeography. *J Biogeogr* **33**:1851-1865.

28. Hijmans R, Cameron S, Parra J, Jones P, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *Int J Clim* **25**:1965–1978.

29. Colorado State University - International Biological Information Systems Tutorials – A MaxEnt Model v.3.3.3e Tutorial (ArcGIS v10) [Internet] [cited 4/1/2013] Available from: http://ibis.colostate.edu/WebContent/WS/ColoradoView/TutorialsDownloads/A_Maxent_Model_v7.pdf

30. Phillips, SJ, Anderson, RP, Schapire, RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**:231-259.

31. Phillips, SJ, Dudik, M. 2008. Modeling of species distributions with MaxEnt: new extensions and a comprehensive evaluation. *Ecography* **31**:161-175.

32. Espinoza B, Martínez-Ibarra JA, Villalobos G, De La Torre P, Laclette JP, Martínez-Hernández F. 2013. Genetic variation of North American Triatomines (Insecta: Hemiptera: Reduviidae): initial divergence between species and populations of Chagas disease vector. *Am J Trop Med Hyg* **88**(2):275-284.

33. Gourbière S, Dorn P, Tripet F, Dumonteil E. 2012. Genetics and evolution of triatomines: from phylogeny to vector control. *Heredity* **108**(3):190-202.

34. University of Wisconsin – Bioweb – Primer Design [Internet] [cited 5/23/2013] Available from: http://bioweb.uwlax.edu/genweb/molecular/seq_anal/primer_design/primer_design.htm

35. Gonçalves RL, Barreto RF, Polycarpo CR, Gadelha FR, Castro SL, Oliveira MF. 2011. A comparative assessment of mitochondrial function in epimastigotes and bloodstream trypomastigotes of *Trypanosoma cruzi*. *J Bioenerg Biomembr* **43**(6):651-661.

36. Carrasco HJ, Frame IA, Valente SA, Miles MA. 1996. Genetic exchange as a possible source of genomic diversity in sylvatic populations of *Trypanosoma cruzi*. *Am J Trop Med Hyg* **54**(4):418-424.

37. de Freitas JM, Augusto-Pinto L, Pimenta JR, Bastos-Rodrigues L, Gonçalves VF, Teixeira SM, Chiari E, Junqueira AC, Fernandes O, Macedo AM, Machado CR, Pena SD. 2006. Ancestral genomes, sex, and the population structure of *Trypanosoma cruzi*. *PLoS Pathog* **2**(3):e24.

38. Higo H, Miura S, Horio M, Mimori T, Hamano S, Agatsuma T, Yanagi T, Cruz-Reyes A, Uyema N, Rojas de Arias A, Matta V, Akahane H, Hirayama K, Takeuchi T, Tada I, Himeno K. 2004. Genotypic variation among lineages of *Trypanosoma cruzi* and its geographic aspects. *Parasitology International* **53**:337-344.

39. Pacheco RC, Moraes-Filho J, Marcili A, Richtzenhain LJ, Szabo MPJ, Catroxo MHB, Bouyer DH, Labruna MB. 2011. *Rickettsia monteiroi* sp. nov., Infecting the Tick *Amblyomma incisum* in Brazil. *Appl Environ Microbiol* **77**(15):5207-5211.

40. Gadagkar SR, Rosenberg MS, Kumar S. 2005. Inferring Species Phylogenies from Multiple Genes: Concatenated Sequence Tree Versus Consensus Gene Tree. *J Exp Zool* **304**(B):64-74.

41. Syracuse University – Plant Ecology and Evolution Research Group – Presence-only modeling with MaxEnt [Internet] [cited 5/20/2013] Available from: http://plantecology.syr.edu/fridley/bio793/maxent.html

42. Packchanian, A. 1942. Reservoir Hosts of Chagas' Disease in the State of Texas Natural Infection of Nine-Banded Armadillo (*Dasypus novemcinctus texanus*), House Mice (Mus musculus), Opossum (Didelphis virginiana), and Wood Rats (*Neotoma micropus micropus),* with *Trypanosoma cruzi* in the State of Texas. *Am J Trop Med Hyg* **1**(6): 623-631.

43. Kjos, SA, Snowden KF, Craig, TM, Lewis B, Ronald N, Olson JK. 2008. Distribution and characterization of canine Chagas disease in Texas. *Vet Parasitol* **152:** 249-256.